

# **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 Statistical and mathematical models can help fill many gaps in our knowledge about species in-  
3 teractions. Predictive and null models are two complementary types of models that have been  
4 developed in network ecology for this purpose. On one hand, predictive models can partially  
5 alleviate the Eltonian shortfall, which describes our current lack of knowledge on food webs  
6 and other ecological networks (Hortal et al. 2015). A variety of such models have recently  
7 been developed using machine learning and other statistical tools, most of which are presented  
8 in Strydom et al. (2021). On the other hand, null models help us identify potential ecological  
9 mechanisms that drive species interactions. They do so by comparing empirical data with an  
10 unbiased distribution of measures generated using a set of rules that exclude the mechanism of  
11 interest (Fortuna and Bascompte 2006; Delmas et al. 2019). Both types of models are frequently  
12 topological, i.e. they often predict the adjacency matrix or specific measures of network struc-  
13 ture without taking into account species' identity. According to Strydom et al. (2021), these  
14 topological models could be used to make better predictions of pairwise species interactions by  
15 constraining the space of feasible networks.

16 The principle of maximum entropy (MaxEnt) is a statistical and topological model that can be  
17 used for both of these purposes, i.e. to make predictions of network structure and to better un-  
18 derstand processes shaping ecological networks. Food-web properties that can be derived using  
19 MaxEnt are varied and pertain to different elements of the network (i.e. at the species (node),  
20 the interaction (edge) or the community (network) levels). Because MaxEnt is a method of find-  
21 ing probability distributions that are least biased given partial knowledge about a system, these  
22 properties need to be represented probabilistically. Otherwise, some data transformation might  
23 be needed. For example, at the species level, MaxEnt can be used to predict the distribution  
24 of trophic levels among species, as well as the distribution of species' vulnerability (number of  
25 predators) and generality (number of preys). By contrast, at the interaction level, predictions  
26 can be made on the distribution of interaction strengths in weighted food webs. At the commu-  
27 nity level, it can give us a probability distribution of potential networks or of specific measures  
28 of their emerging structure. The applicability and performance of this method depend on the  
29 ecological information available and on our capacity to find the set of state variables that best

30 represent natural systems and to translate them into appropriate statistical constraints. MaxEnt  
31 has been used in a wide range of disciplines, from thermodynamics to chemistry and biology  
32 (Martyushev and Seleznev 2006). It has also been proven useful in ecology, e.g. in species dis-  
33 tribution models (Phillips, Anderson, and Schapire 2006) and macroecological models (Harte  
34 et al. 2008; Harte and Newman 2014). The probability distribution of the variable of interest  
35 with the maximum entropy value is unique and least biased under the set of constraints used.  
36 These constraints are built using state variables, i.e. variables that represent the macrostate of  
37 the system. Having a validated maximum entropy model for the system at hand allows us to  
38 make rigorous predictions using a minimal amount of data, as well as helping us describe the  
39 most important factors driving that system.

40 Despite its extensive use in graph and network theories (e.g., Park and Newman 2004; van der  
41 Hoorn, Lippner, and Krioukov 2018), MaxEnt has in comparison been little used in network  
42 ecology. Like other real system networks, food webs and other ecological networks are repre-  
43 sented mathematically as graphs. However, the very nature of food webs (directed simple graphs  
44 frequently having self-loops) makes the mathematical optimization of maximum entropy graph  
45 models more complicated than with many other types of (non-ecological) networks. MaxEnt has  
46 nevertheless been used to predict the degree distribution of bipartite ecological networks from  
47 the number of species and the number of interactions (Williams 2011) and to predict interaction  
48 strengths between species pairs using their relative abundances within an optimal transportation  
49 theory regularized with information entropy (Stock, Poisot, and De Baets 2021). However, to  
50 the best of our knowledge, MaxEnt has never been used to predict food-web structure directly,  
51 even though food webs are among the most documented and widespread ecological networks.

52 In this contribution, we used two complementary approaches to predict the structure of food  
53 webs using the principle of maximum entropy. We then compared our predictions against em-  
54 pirical data and null and neutral models commonly used in network ecology. The first approach  
55 consists in deriving constrained probability distributions of given network properties directly.  
56 We derived the joint degree distribution (the joint probability distribution that a species has a  
57 given number of prey and predators in a food web) of maximum entropy using only the number  
58 of species  $S$  and the number of interactions  $L$  as state variables. Then, we predicted the degree

59 distribution of maximum entropy directly from the joint degree distribution since the first is the  
60 sum of the marginal distributions of the second (a species' degree is the sum of its in and out-  
61 degrees). Because of the scarcity of empirical data on the number of interactions in ecological  
62 networks, we present a method to predict  $L$  from  $S$  (Box 1), thus allowing the prediction of the  
63 joint degree distribution from  $S$  solely. In turn, the second approach consists in finding, using  
64 different constraints, the adjacency matrix of maximum entropy from which network properties  
65 can be measured. To do so, we used a flexible and heuristic approach based on simulated an-  
66 nealing (an optimization algorithm) to find networks *close* to maximum entropy. As discussed  
67 above, our choice of algorithm stands from the very nature of food webs (i.e. simple directed  
68 networks allowing self-loops) that makes the analytical derivation of a maximum entropy graph  
69 model difficult. We first built our type I MaxEnt network model constrained by the connectance  
70 of the network (i.e. the ratio  $L/S^2$ ). A comparison of this model against empirical data indi-  
71 cated that connectance alone was not sufficient to predict many aspects of network structure.  
72 For this reason, we built our type II MaxEnt network model, which instead uses the whole joint  
73 degree sequence as a constraint. Overall, we found that this second model was much better at  
74 predicting food-web structure than the one based on connectance.

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

76 The principle of maximum entropy is a mathematical method of finding probability distribu-  
77 tions, strongly rooted in statistical mechanics and information theory (Jaynes 1957a, 1957b;  
78 Harremoës and Topsøe 2001). Starting from a set of constraints given by prior knowledge of a  
79 system (i.e. what we call state variables), this method helps us find least-biased probability dis-  
80 tributions subject to the constraints. These probability distributions are guaranteed to be unique  
81 given our prior knowledge and represent the most we can say about a system without making  
82 more assumptions. For example, if the only thing we know about a biological community is  
83 its average number of individuals per species, the least-biased inference we could make on its  
84 species abundance distribution is the exponential distribution (Frank and Smith 2011; Harte  
85 and Newman 2014). However, this does not imply that this distribution will be the best fit to  
86 empirical data. The challenge is to find the right set of constraints that would best reproduce

87 distributions found in nature.

88 The least-biased probability distribution given the constraints used is the one with the highest  
89 entropy among all probability distributions that satisfy these constraints. Entropy is a measure of  
90 the average amount of information given by the outcome of a random variable. Many measures  
91 of entropy have been developed in physics (Beck 2009), but only a fraction of them could be used  
92 as an optimization measure with the principle of maximum entropy. According to Beck (2009)  
93 and Khinchin (2013), a measure of entropy  $H$  should satisfy four properties in the discrete case:  
94 (1) it should be a function of a probability distribution  $p(n)$  only; (2) it should be maximized  
95 when  $p(n)$  is uniform; (3) it should not be influenced by outcomes with a null probability; and  
96 (4) it should be independent of the order of information acquisition. The Shannon's entropy  
97 (Shannon 1948)

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

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

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

104 where  $g_i$  is the mathematical formulation of the constraint  $i$  and  $c_i$ , its value. Note that  $F$  is just  
105 Shannon's entropy to which we added terms that each sums to zero ( $g_i = c_i$ ).  $F$  is maximized  
106 by setting to 0 its partial derivative with respect to  $p(n)$ . In this contribution, we show how this  
107 can be done by deriving the joint degree distribution analytically from the number of species  
108 and the number of interactions in food webs.

109 **Data and code**

110 We tested our analytical and heuristic MaxEnt models against open food-web data queried from  
111 three different sources and integrated into what we call our *complete dataset*. These sources  
112 include (1) terrestrial and aquatic food webs sampled globally and archived on the ecologi-  
113 cal interactions database Mangal; (2) freshwater stream food webs from New Zealand; and (3)  
114 aquatic food webs from Tuesday Lake (Michigan, United States). First, all food webs archived  
115 on `mangal.io` (Poisot et al. 2016; Banville, Vissault, and Poisot 2021) were directly queried  
116 from the database ( $n = 235$ ). Most ecological networks archived on Mangal are multilayer net-  
117 works, i.e. networks that describe different types of interactions. We kept all networks whose  
118 interactions were mainly of predation and herbivory types, and removed the largest network  
119 ( $S = 714$ ) for computational efficiency reasons. Then, to this set we added food webs from two  
120 different sources: the New-Zealand dataset ( $n = 21$ ; Pomeranz et al. 2018) and the Tuesday  
121 lake dataset ( $n = 2$ ; Cohen, Jonsson, and Carpenter 2003). Of these two datasets, 19 networks  
122 had data on species' relative abundances that were used in the neutral model (Box 2). These 19  
123 networks constitute our *abundance dataset*, which is a subset of our complete dataset.

124 All code and data to reproduce this article are available at the Open Science Framework. Data  
125 cleaning, simulations and analyses were conducted in Julia v1.5.4.

126 **Analytical models: Measures of maximum entropy**

127 **Joint degree distribution**

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

134 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>135</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>136</sup> The first constraint  $g_1$  is our normalizing constraint, whereas the other two ( $g_2$  and  $g_3$ ) fix the  
<sup>137</sup> average of the marginal distributions of  $k_{in}$  and  $k_{out}$  to the linkage density  $L/S$ . It is important  
<sup>138</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>139</sup> without any further constraints, we expect the joint degree distribution of maximum entropy to  
<sup>140</sup> be a symmetric probability distribution with regards to  $k_{in}$  and  $k_{out}$ . However, this does not mean  
<sup>141</sup> that the joint degree *sequence* will be symmetric, since the joint degree sequence is essentially  
<sup>142</sup> a random realization of its probabilistic counterpart.

<sup>143</sup> The joint probability distribution of maximum entropy given these constraints is found using the  
<sup>144</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>145</sup> where  $\lambda_1$ ,  $\lambda_2$ , and  $\lambda_3$  are the Lagrange multipliers. The probability distribution that maximizes  
<sup>146</sup> entropy is obtained by finding these values. As pointed out in the introduction,  $F$  is just Shan-  
<sup>147</sup> non's entropy to which we added terms that each sums to zero (our constraints).  $F$  is maximized  
<sup>148</sup> by setting to 0 its partial derivative with respect to  $p(k_{in}, k_{out})$ . Because the derivative of a con-  
<sup>149</sup> stant is zero, this gives us:

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

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

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

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

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

152 where  $Z = e^{1+\lambda_1}$  is called the partition function. The partition function ensures that probabilities

153 sum to 1 (our normalization constraint). It can be expressed in terms of  $\lambda_2$  and  $\lambda_3$  as follows:

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

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

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

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

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

161 We predicted the joint degree distribution of maximum entropy for each food web in our com-

plete dataset, i.e. using their numbers of species and numbers of interactions as state variables.  
We then sampled one realization of the degree sequence for each network using the probabilities given by the joint degree distribution. In fig. 1 (left panels), we show the relationship between  $k_{out}$  and  $k_{in}$  standardized by the number of species in their networks, for empirical and maximum entropy joint degree distributions. We see that our model predicts higher values of generality and vulnerability compared to empirical food webs (i.e. relative values of  $k_{out}$  and  $k_{in}$  closer to 1). However, plotting the difference between predicted and empirical values for each species gives a different perspective. The right panel of fig. 1 presents these differences when species are ordered by their total degree in their network (i.e. by the sum of their in and out-degrees). Indeed, our predicted joint degree sequences have the same number of species 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 to associate predictions with empirical data. If we associate each node according to their rank in total degrees, we see that species predicted to be have a higher generality (number of preys) generally have a lower vulnerability (number of preys) than what is observed (and conversely). This is because the difference in total degree ( $k_{out} + k_{in}$ ) is minimized when species are ordered by their total degrees in both predicted and empirical sets (i.e. the average sum of relative  $k_{out}$  and  $k_{in}$  is close to 0 across all species). This result shows that the differences between predicted and empirical degrees (in and out) are low for most species when ordered by their total degrees and that there is 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.]

We can describe the shape of the joint degree distribution by comparing its marginal distributions with one another. Doing so allows us to evaluate how much the shape of the joint degree distribution of maximum entropy differs from empirical data. We plotted the Kullback–Leibler (KL) divergence between in and out-degree sequences to compare the symmetry of empirical and maximum entropy joint degree sequences fig. 2. As we expected, our model predicted more

190 similar in-degree and out-degree distributions than empirical data (shown by a lower KL diver-  
 191 gence). However, this difference decreased with connectance right panel of 2. This suggests  
 192 that other ecological constraints might be needed to account for the asymmetry of the joint de-  
 193 gree distribution, especially for networks with a lower connectance. This might be due to the  
 194 fact that food webs with low connectance are typically harder to predict than food webs with  
 195 high connectance. Indeed, in low connectance systems, what makes two species interact might  
 196 be more important for predictions than in high connectance systems, in which what prevents  
 197 species from interacting might be more meaningful. This implies that more ecological informa-  
 198 tion might be needed in food webs with fewer numbers of interactions because more nuances  
 199 determine interactions compared to non-interactions.

200 [Figure 2 about here.]

## 201 **Degree distribution**

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

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

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

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

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

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

215 This can be solved numerically using the constraint of the average degree  $\langle k \rangle = \frac{2L}{S}$  of a species.

216 Note that the mean degree is twice the value of the linkage density, because every link must be  
217 counted twice when we add in and out-degrees together.

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

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

221 In this section, we showed how important measures of food-web structure, namely the degree  
222 distribution and the joint degree distribution, could be derived with the principle of maximum  
223 entropy using minimal knowledge on a biological community. This type of models, although  
224 useful to make least-biased predictions on many network properties, can be hard to apply for  
225 other measures. Indeed, there are dozens of measures of network structure (Delmas et al. 2019)  
226 and many are not calculated with mathematical equations, but with algorithms. Moreover, the  
227 applicability of this method to empirical systems is limited by the state variables we can actu-  
228 ally measure and use. In the next section, we propose a more flexible method to predict many  
229 measures of network structure simultaneously, i.e. by finding heuristically the configuration of  
230 the network having maximum entropy given partial knowledge of its emerging structure.

## 231 **Box 1 - Working with predicted numbers of interactions**

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

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

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

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

246 We fitted the flexible links model on all food webs in our complete dataset, and estimated the  
 247 parameters of eq. 16 using a Hamiltonian Monte Carlo sampler with static trajectory (1 chain  
 248 and 3000 iterations):

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

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

254 The flexible links model is a generative model, i.e. it can generate plausible values of the pre-  
 255 dicted variable. We thus simulated 1000 values of  $L$  for different values of  $S$  using the joint  
 256 posterior distribution of our model parameters (eq. 17), and calculated the mean degree for

257 each simulated values. The resulting distributions are shown in the left panel of fig. 3 for three  
258 different values of species richness. In the right panel of fig. 3, we show how the probability  
259 distribution for the mean degree constraints can be used to generate a distribution of maximum  
260 entropy degree distributions, since each simulated value of mean degree generates a different  
261 maximum entropy degree distribution (eq. 14 and eq. 15).

262 [Figure 3 about here.]

## 263 **Box 2 - Corresponding null and neutral models**

### 264 **Null models (types I and II)**

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

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

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

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

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

274 We predicted both types of null networks for all empirical networks in our complete dataset  
275 ( $n = 257$ ). We converted all probabilistic networks to Boolean networks by generating 100  
276 random Boolean networks for each of these probabilistic webs. Then, we counted the number

277 of times each interaction was sampled, and kept the  $L$  entries that were drawn the most amount  
278 of time, with  $L$  given by the number of interactions in each food web. This ensured that the  
279 resulting null networks had the same number of interactions as their empirical counterparts.

280 **Neutral model**

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

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

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

287 **Heuristic models: Networks of maximum entropy**

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

289 The principle of maximum entropy can be applied on the network itself if we decompose its  
290 adjacency matrix into a non-zero vector of relative values. This is a necessary step when work-  
291 ing with food webs, which are frequently expressed as a matrix  $A = [a_{ij}]$  of Boolean values  
292 representing the presence ( $a_{ij} = 1$ ) or absence ( $a_{ij} = 0$ ) of an interaction between two species  $i$   
293 and  $j$ . Knowing one or many properties of a food web of interest (e.g., its number of species and  
294 number of interactions), we can simulate its adjacency matrix randomly by using these known  
295 ecological information to constrain the space of potential networks. The entropy of this hypo-  
296 thetical matrix can then be measured after decomposing it into appropriate values. Simulating  
297 a series of networks until we find the one having the highest entropy allows us to search for the  
298 most complex food-web configuration given the ecological constraints used. This configuration

299 is the least biased one considering the information we had. In other words, the most we can  
300 say about a network's adjacency matrix, without making more assumptions than the ones given  
301 by our incomplete knowledge of its structure, is the one of maximum entropy. Generating the  
302 most complex network that corresponds to this structure allows us to explore more easily other  
303 properties of food webs under MaxEnt.

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

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

311 where  $s_i$  are the relative singular values of the adjacency matrix ( $s_i = \sigma_i / \sum_{i=1}^R \sigma_i$ , where  $\sigma_i$   
312 are the singular values). Note that the distribution of relative singular values is analogous to  
313 a probability distribution, with  $0 < s_i < 1$  and  $\sum s_i = 1$ . Thus, this measure also satisfies  
314 all four properties of an appropriate entropy measure (see introduction), while being a proper  
315 measure of the internal complexity of food webs (Strydom, Dalla Riva, and Poisot 2021). Fol-  
316 lowing Strydom, Dalla Riva, and Poisot (2021), we standardized this measure with the rank  $R$  of  
317 the matrix (i.e.  $J / \ln(R)$ ) to account for the difference in dimensions between networks (*sensu*  
318 Pielou's evenness; Pielou 1975). We used SVD entropy to predict the network configuration  
319 of maximum entropy (i.e. of maximum complexity) heuristically given a predetermined set of  
320 constraints.

### 321 **MaxEnt network models (types I and II)**

322 We built two types of MaxEnt network models: one based on connectance (type I MaxEnt net-  
323 work model) and the other based on the joint degree sequence (type II MaxEnt network model).

324 They are based on the same constraints as the types I and II null models presented above. For  
325 both models, we used a simulated annealing algorithm with 4 chains, 2000 steps and an initial  
326 temperature of 0.2. For each chain, we first generated one random Boolean matrix with the same  
327 order (number of species) as empirical webs, while maintaining the total number of interactions  
328 (type I MaxEnt network model) or rows and columns sums (type II MaxEnt network model).  
329 These are our initial configurations. Then, we swapped interactions sequentially while main-  
330 taining the original connectance or the joint degree sequence for types I and II MaxEnt network  
331 model, respectively. Configurations with a higher SVD entropy than the previous one in the  
332 chain were always accepted, whereas they were accepted with a probability conditional to a de-  
333 creasing temperature when lower. The final configuration with the highest SVD entropy among  
334 the four chains constitute our estimated MaxEnt network. Even though we decided to work with  
335 point estimates, it is possible to have a (non MaxEnt) probability distribution of networks when  
336 working with the entire chains (or part of the chains). For each network in our complete and  
337 abundance datasets, we estimated their configuration with maximum entropy using both types  
338 of MaxEnt network models.

### 339 Structure of MaxEnt networks

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

351 Overall, we found that models based on the joint degree sequence (type II null and MaxEnt

352 network models) reproduced the structure of empirical networks much better than the ones based  
 353 on connectance and the neutral model of relative abundances. On average, the type II MaxEnt  
 354 network model was better at predicting nestedness ( $0.62 \pm 0.08$ ) than its corresponding null  
 355 model ( $0.73 \pm 0.05$ ; empirical networks:  $0.63 \pm 0.09$ ), as well as the proportion of cannibal  
 356 species. This might in part be due because nestedness was calculated using the spectral radius  
 357 of the adjacency matrix, which leverages information on the network itself just like the MaxEnt  
 358 network models. However, the type II null model was better at predicting network diameter and  
 359 average maximum similarity between species pairs. Predictions were similar between both type  
 360 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 all 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

361 The picture slightly changes when we consider another important property of ecological net-  
 362 works, i.e. their motifs profile. We measured the proportion of three-species motifs, which can  
 363 be considered as simple building blocks of ecological networks (Milo et al. 2002; Stouffer et al.  
 364 2007). Motifs are the backbone of complex ecological networks from which network structure  
 365 is built upon and play a crucial role in community dynamics and assembly (Bastolla et al. 2009;  
 366 Stouffer and Bascompte 2011). In fig. 4, we show that the motifs profile of networks generated  
 367 using the type II MaxEnt network model was very close to the one of empirical data, and that  
 368 they made better predictions than the type II null model. This is also shown in fig. 5, where we  
 369 see that relationships between motifs pairs of empirical food webs are very similar to the ones  
 370 of the type II MaxEnt networks. Overall, these results suggest that the type II MaxEnt network  
 371 model can reproduce many aspects of network structure. This highlights the importance of the  
 372 joint degree sequence in shaping ecological networks.

373 [Figure 4 about here.]

374 [Figure 5 about here.]

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

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

390 [Figure 6 about here.]

## 391 Discussion

392 The principle of maximum entropy is a robust mathematical method of finding least-biased  
393 probability distributions that have some desired properties given by prior knowledge on a sys-  
394 tem. We first used MaxEnt to predict the joint degree distribution and the degree distribution  
395 of maximum entropy given known numbers of species and of interactions. We found that the  
396 resulting joint degree distributions were more symmetric than the ones of empirical food webs,  
397 which suggests that other constraints might be needed to improve those predictions. We also  
398 used MaxEnt to predict networks of maximum entropy with a specified structure. These net-  
399 works are the most complex, or random, given the constraints used. Likewise, we found that  
400 knowledge of species richness and of the number of interactions were not sufficient to reproduce  
401 many aspects of network structure found in nature. However, a model based on the entire joint

402 degree sequence, i.e. on the number of preys and predators for each species, gave more con-  
403 vincing results. Our type II MaxEnt network model yielded better or similar predictions than  
404 the type II null model, also based on the joint degree sequence, for most measures considered  
405 including the motifs profile. This suggests that the joint degree sequence drives many aspects of  
406 network structure. Indeed, considering our findings that empirical networks are close to max-  
407 imum entropy for a given joint degree sequence, our results suggest that food-web topology is  
408 configured almost entirely randomly around these marginal numbers of predators and preys.

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

## 418 Alternative MaxEnt models

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

430 Another method generating adjacency matrices are maximum entropy graph models, which pre-  
431 dict a probability distribution of networks under soft or hard constraints (e.g., Park and Newman  
432 2004; Cimini et al. 2019). Under hard constraints, every network with a non-zero probability  
433 exactly satisfies the constraints on its structure. This is in contrast with soft constraints, which  
434 require that networks satisfy them on average (i.e. many networks with a non-zero probability do  
435 not have the exact structure set by the constraints). Maximum entropy graph models are helpful  
436 because they can provide probability distributions for many network properties, by measuring  
437 the structure of all adjacency matrices with a non-zero probability. However, we consider that  
438 our approach based on simulated annealing is more flexible and more computationally efficient.  
439 Indeed, many measures of food-web structure are hard to translate into mathematical constraints.  
440 Moreover, because food webs are directed networks with self-loops, it makes the mathematical  
441 derivation of maximum entropy graph models difficult. We believe that identifying heuristically  
442 what really constrains the topology of food webs is a useful first step before attempting to derive  
443 the mathematical formulation of a maximum entropy graph model adapted to food webs.

## 444 Applications

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

456 **Conclusion**

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

464 **Acknowledgments**

465 We acknowledge that this study was conducted on land within the traditional unceded territory  
466 of the Saint Lawrence Iroquoian, Anishinabewaki, Mohawk, Huron-Wendat, and Omàmiwinini-  
467 wak nations. This work was supported by the Institute for Data Valorisation (IVADO) and the  
468 Natural Sciences and Engineering Research Council of Canada (NSERC) Collaborative Re-  
469 search and Training Experience (CREATE) program, through the Computational Biodiversity  
470 Science and Services (BIOS<sup>2</sup>) program.

471 **References**

- 472 Albert, R., and A. L. Barabasi. 2002. “Statistical Mechanics of Complex Networks.” *Reviews*  
473 *of Modern Physics* 74 (1): 47–97. <https://doi.org/10.1103/RevModPhys.74.47>.
- 474 Banville, Francis, Steve Vissault, and Timothée Poisot. 2021. “Mangal.jl and EcologicalNet-  
475 works.jl: Two Complementary Packages for Analyzing Ecological Networks in Julia.” *Jour-*  
476 *nal of Open Source Software* 6 (61): 2721. <https://doi.org/10.21105/joss.02721>.
- 477 Bascompte, J., P. Jordano, C. J. Melian, and J. M. Olesen. 2003. “The Nested Assembly of  
478 Plant-Animal Mutualistic Networks.” *Proceedings of the National Academy of Sciences*  
479 *of the United States of America* 100 (16): 9383–87. <https://doi.org/10.1073/pnas.1633576100>.

- 481 Bastolla, Ugo, Miguel A. Fortuna, Alberto Pascual-Garcia, Antonio Ferrera, Bartolo Luque, and  
482 Jordi Bascompte. 2009. “The Architecture of Mutualistic Networks Minimizes Competition  
483 and Increases Biodiversity.” *Nature* 458 (7241): 1018–U91. <https://doi.org/10.1038/nature07950>.
- 485 Beck, Christian. 2009. “Generalised Information and Entropy Measures in Physics.” *Contemporary Physics* 50 (4): 495–510. <https://doi.org/10.1080/00107510902823517>.
- 487 Cimini, Giulio, Tiziano Squartini, Fabio Saracco, Diego Garlaschelli, Andrea Gabrielli, and  
488 Guido Caldarelli. 2019. “The Statistical Physics of Real-World Networks.” *Nature Reviews Physics* 1 (1): 58–71. <https://doi.org/10.1038/s42254-018-0002-6>.
- 490 Cirtwill, Alyssa R., Anna Eklof, Tomas Roslin, Kate Wootton, and Dominique Gravel. 2019.  
491 “A Quantitative Framework for Investigating the Reliability of Empirical Network Construc-  
492 tion.” *Methods in Ecology and Evolution* 10 (6): 902–11. <https://doi.org/10.1111/2041-210X.13180>.
- 494 Cohen, Joel E., Tomas Jonsson, and Stephen R. Carpenter. 2003. “Ecological Community  
495 Description Using the Food Web, Species Abundance, and Body Size.” *Proceedings of  
496 the National Academy of Sciences* 100 (4): 1781–86. <https://doi.org/10.1073/pnas.232715699>.
- 498 Delmas, Eva, Mathilde Besson, Marie Hélène Brice, Laura A. Burkle, Giulio V. Dalla Riva,  
499 Marie Josée Fortin, Dominique Gravel, et al. 2019. “Analysing Ecological Networks of  
500 Species Interactions.” *Biological Reviews*. <https://doi.org/10.1111/brv.12433>.
- 501 Dunning, Iain, Joey Huchette, and Miles Lubin. 2017. “JuMP: A Modeling Language for  
502 Mathematical Optimization.” *SIAM Review* 59 (2): 295–320. <https://doi.org/10.1137/15M1020575>.
- 504 Fortuna, M. A., and J. Bascompte. 2006. “Habitat Loss and the Structure of Plant-Animal  
505 Mutualistic Networks.” *Ecology Letters* 9 (3): 278–83. <https://doi.org/10.1111/j.1461-0248.2005.00868.x>.
- 507 Fortuna, M. A., Daniel B. Stouffer, Jens M. Olesen, Pedro Jordano, David Mouillot, Boris R.  
508 Krasnov, Robert Poulin, and Jordi Bascompte. 2010. “Nestedness Versus Modularity in

- 509 Ecological Networks: Two Sides of the Same Coin?” *Journal of Animal Ecology* 79 (4):  
510 811–17. <https://doi.org/10.1111/j.1365-2656.2010.01688.x>.
- 511 Frank, S. A., and E. Smith. 2011. “A Simple Derivation and Classification of Common Prob-  
512 ability Distributions Based on Information Symmetry and Measurement Scale.” *Journal of*  
513 *Evolutionary Biology* 24 (3): 469–84. <https://doi.org/10.1111/j.1420-9101.2010.02204.x>.
- 515 Harremoës, Peter, and Flemming Topsøe. 2001. “Maximum Entropy Fundamentals.” *Entropy*  
516 3 (3): 191–226. <https://doi.org/10.3390/e3030191>.
- 517 Harte, J., and Erica A. Newman. 2014. “Maximum Information Entropy: A Foundation for  
518 Ecological Theory.” *Trends in Ecology & Evolution* 29 (7): 384–89. <https://doi.org/10.1016/j.tree.2014.04.009>.
- 520 Harte, J., T. Zillio, E. Conlisk, and A. B. Smith. 2008. “Maximum Entropy and the State-  
521 Variable Approach to Macroecology.” *Ecology* 89 (10): 2700–2711. <https://doi.org/10.1890/07-1369.1>.
- 523 Hoorn, Pim van der, Gabor Lippner, and Dmitri Krioukov. 2018. “Sparse Maximum-Entropy  
524 Random Graphs with a Given Power-Law Degree Distribution.” *Journal of Statistical Physics*  
525 173 (3): 806–44. <https://doi.org/10.1007/s10955-017-1887-7>.
- 526 Hortal, Joaquín, Francesco de Bello, José Alexandre F. Diniz-Filho, Thomas M. Lewinsohn,  
527 Jorge M. Lobo, and Richard J. Ladle. 2015. “Seven Shortfalls That Beset Large-Scale  
528 Knowledge of Biodiversity.” *Annual Review of Ecology, Evolution, and Systematics* 46 (1):  
529 523–49. <https://doi.org/10.1146/annurev-ecolsys-112414-054400>.
- 530 Jaynes, E. T. 1957a. “Information Theory and Statistical Mechanics.” *Physical Review* 106 (4):  
531 620–30. <https://doi.org/10.1103/PhysRev.106.620>.
- 532 ———. 1957b. “Information Theory and Statistical Mechanics. II.” *Physical Review* 108 (2):  
533 171–90. <https://doi.org/10.1103/PhysRev.108.171>.
- 534 Khinchin, A. Ya. 2013. *Mathematical Foundations of Information Theory*. Courier Corpora-  
535 tion.

- 536 MacDonald, Arthur Andrew Meahan, Francis Banville, and Timothée Poisot. 2020. “Revisiting  
537 the Links-Species Scaling Relationship in Food Webs.” *Patterns* 0 (0). <https://doi.org/10.1016/j.patter.2020.100079>.
- 538
- 539 Martyushev, L. M., and V. D. Seleznev. 2006. “Maximum Entropy Production Principle in  
540 Physics, Chemistry and Biology.” *Physics Reports-Review Section of Physics Letters* 426  
541 (1): 1–45. <https://doi.org/10.1016/j.physrep.2005.12.001>.
- 542 Milo, R., S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon. 2002. “Net-  
543 work Motifs: Simple Building Blocks of Complex Networks.” *Science* 298 (5594): 824–27.  
544 <https://doi.org/10.1126/science.298.5594.824>.
- 545 Park, Juyong, and M. E. J. Newman. 2004. “Statistical Mechanics of Networks.” *Physical  
546 Review E* 70 (6): 066117. <https://doi.org/10.1103/PhysRevE.70.066117>.
- 547 Phillips, Steven J., Robert P. Anderson, and Robert E. Schapire. 2006. “Maximum Entropy  
548 Modeling of Species Geographic Distributions.” *Ecological Modelling* 190 (3): 231–59.  
549 <https://doi.org/10.1016/j.ecolmodel.2005.03.026>.
- 550 Pielou, Evelyn C. 1975. “Ecological Diversity.” In. 574.524018 P5.
- 551 Poisot, Timothee, Benjamin Baiser, Jennifer A. Dunne, Sonia Kefi, Francois Massol, Nico-  
552 las Mouquet, Tamara N. Romanuk, Daniel B. Stouffer, Spencer A. Wood, and Dominique  
553 Gravel. 2016. “Mangal - Making Ecological Network Analysis Simple.” *Ecography* 39 (4):  
554 384–90. <https://doi.org/10.1111/ecog.00976>.
- 555 Pomeranz, Justin Page, Ross M. Thompson, Timothée Poisot, Jon S. Harding, and Justin P. F.  
556 Pomeranz. 2018. “Data from: Inferring Predator-Prey Interactions in Food Webs.” Dryad.  
557 <https://doi.org/10.5061/DRYAD.K59M37F>.
- 558 Shannon, C. E. 1948. “A Mathematical Theory of Communication.” *The Bell System Technical  
559 Journal* 27 (3): 379–423. <https://doi.org/10.1002/j.1538-7305.1948.tb01338.x>.
- 560 Staniczenko, Phillip P. A., Jason C. Kopp, and Stefano Allesina. 2013. “The Ghost of Nested-  
561 ness in Ecological Networks.” *Nature Communications* 4 (1): 1391. <https://doi.org/10.1038/ncomms2422>.
- 562

- 563 Stock, Michiel, Timothée Poisot, and Bernard De Baets. 2021. “Optimal Transportation Theory  
564 for Species Interaction Networks.” *Ecology and Evolution* 11 (9): 3841–55. <https://doi.org/10.1002/ece3.7254>.
- 565
- 566 Stouffer, Daniel B., and Jordi Bascompte. 2011. “Compartmentalization Increases Food-Web  
567 Persistence.” *Proceedings of the National Academy of Sciences* 108 (9): 3648–52. <https://doi.org/10.1073/pnas.1014353108>.
- 568
- 569 Stouffer, Daniel B., Juan Camacho, Wenxin Jiang, and Luis A. Nunes Amaral. 2007. “Evidence  
570 for the Existence of a Robust Pattern of Prey Selection in Food Webs.” *Proceedings of the  
571 Royal Society B-Biological Sciences* 274 (1621): 1931–40. <https://doi.org/10.1098/rspb.2007.0571>.
- 572
- 573 Strydom, Tanya, Michael D. Catchen, Francis Banville, Dominique Caron, Gabriel Dansereau,  
574 Philippe Desjardins-Proulx, Norma R. Forero-Muñoz, et al. 2021. “A Roadmap Towards  
575 Predicting Species Interaction Networks (across Space and Time).” *Philosophical Trans-  
576 actions of the Royal Society B: Biological Sciences* 376 (1837): 20210063. <https://doi.org/10.1098/rstb.2021.0063>.
- 577
- 578 Strydom, Tanya, Giulio V. Dalla Riva, and Timothee Poisot. 2021. “SVD Entropy Reveals the  
579 High Complexity of Ecological Networks.” *Frontiers in Ecology and Evolution* 9: 623141.  
580 <https://doi.org/10.3389/fevo.2021.623141>.
- 581
- 582 Williams, Richard J. 2011. “Biology, Methodology or Chance? The Degree Distributions of  
583 Bipartite Ecological Networks.” *PLOS ONE* 6 (3): e17645. <https://doi.org/10.1371/journal.pone.0017645>.
- 584
- 585 Williams, Richard J., and Neo D. Martinez. 2000. “Simple Rules Yield Complex Food Webs.”  
*Nature* 404 (6774): 180–83. <https://doi.org/10.1038/35004572>.

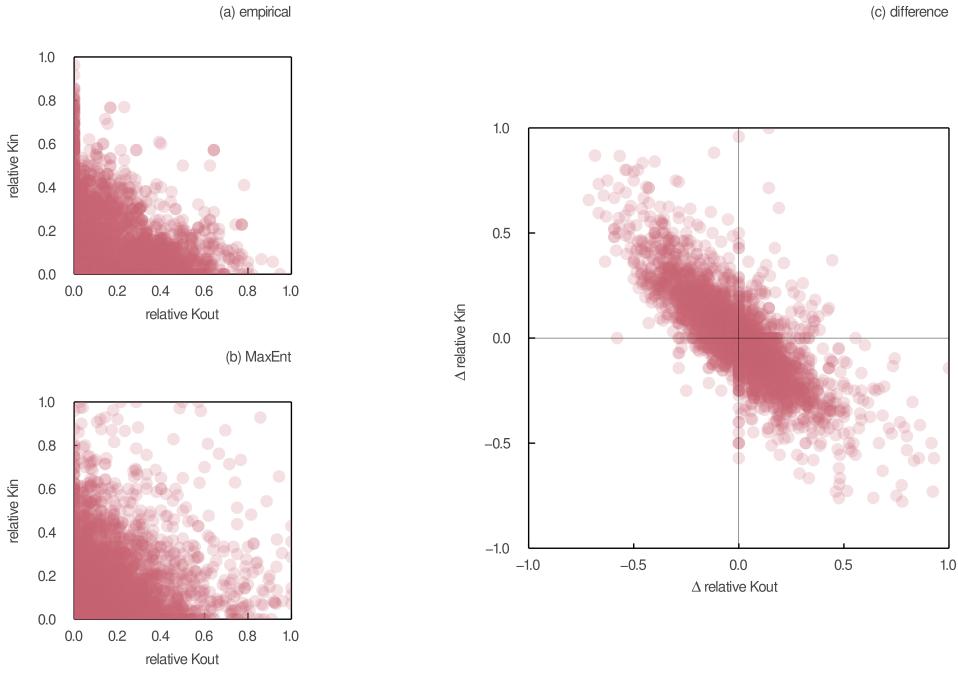


Figure 1: Relative number of predators ( $k_{in}$ ) as a function of the relative number of preys ( $k_{out}$ ) for each species in (a) empirical and (b) maximum entropy joint degree sequences. Empirical networks include all food webs archived on Mangal, as well as the New-Zealand and Tuesday lake datasets (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. (c) Difference between predicted and empirical values when species are ordered according to their total degrees. In all panels, each dot corresponds to a single species in one of the network of our complete dataset.

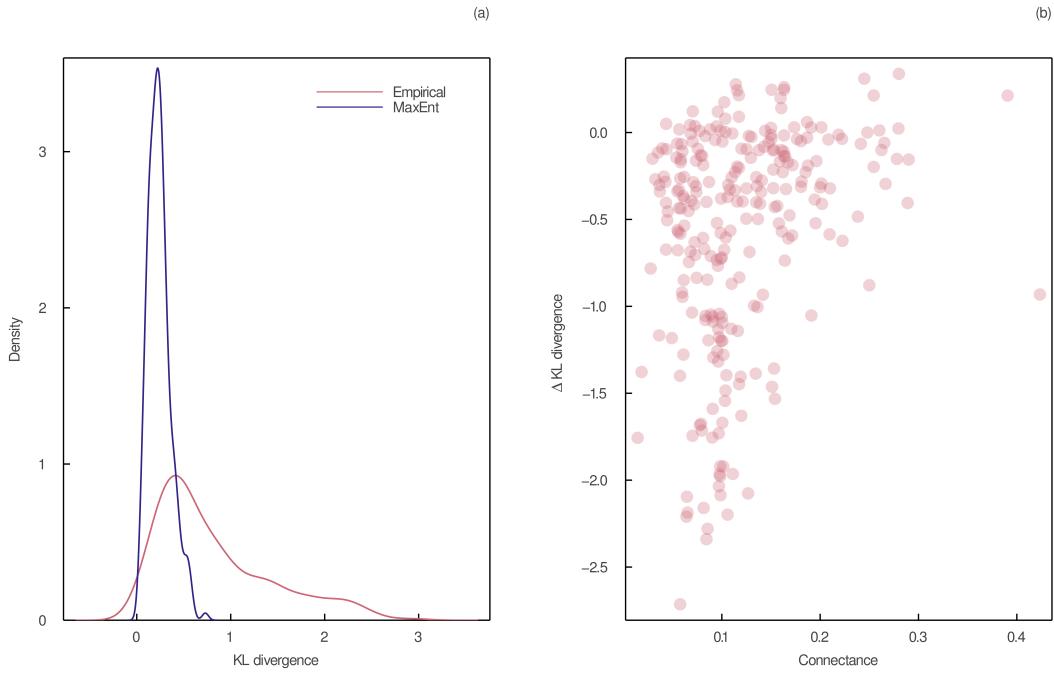


Figure 2: (a) Probability density of KL divergence between in and out-degree sequences of empirical and maximum entropy joint degree sequences. (b) Difference between the KL divergence of empirical and predicted networks as a function of connectance. In both panels, empirical networks include all food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

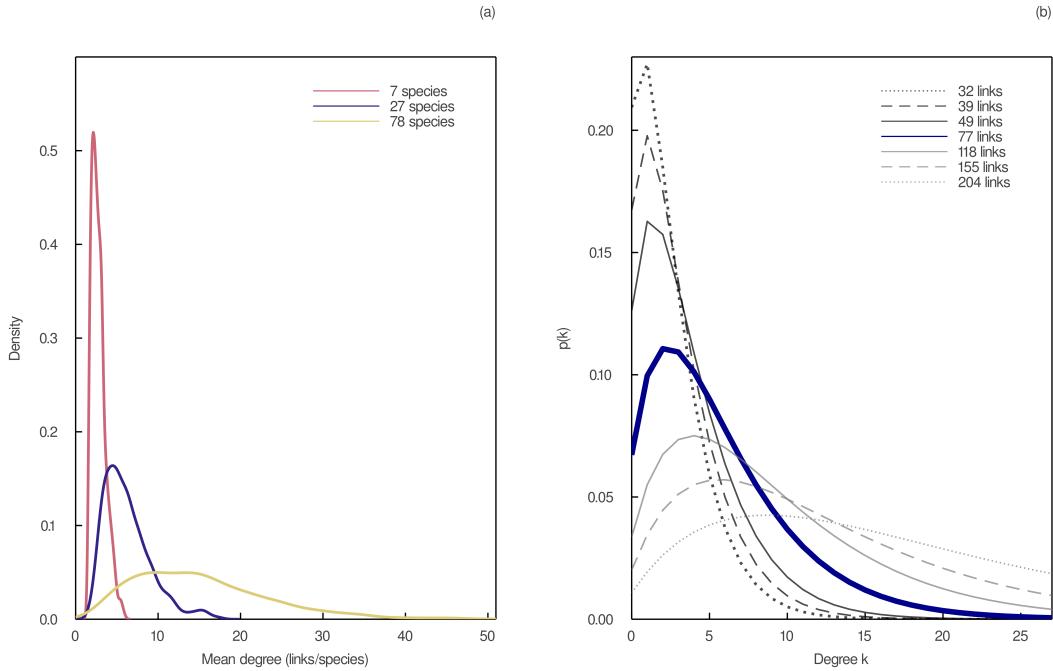


Figure 3: (a) Probability density of the predicted mean degree for different values of species richness. The number of interactions was predicted using the flexible links model fitted to all empirical networks in our complete dataset (eq. 17). (b) Degree distributions of maximum entropy for a network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

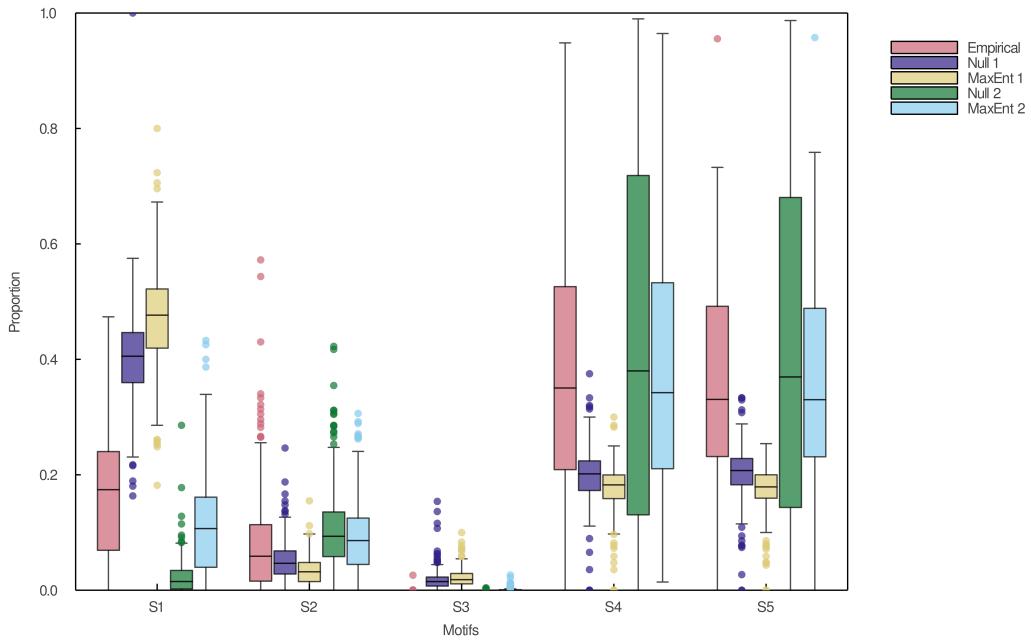


Figure 4: Motifs profile of empirical, maximum entropy, and null food webs. Empirical networks include all food webs archived on Mangal, as well as the New-Zealand and Tuesday lake datasets (our complete dataset). Maximum entropy networks were estimated using a simulated annealing algorithm to find the 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).

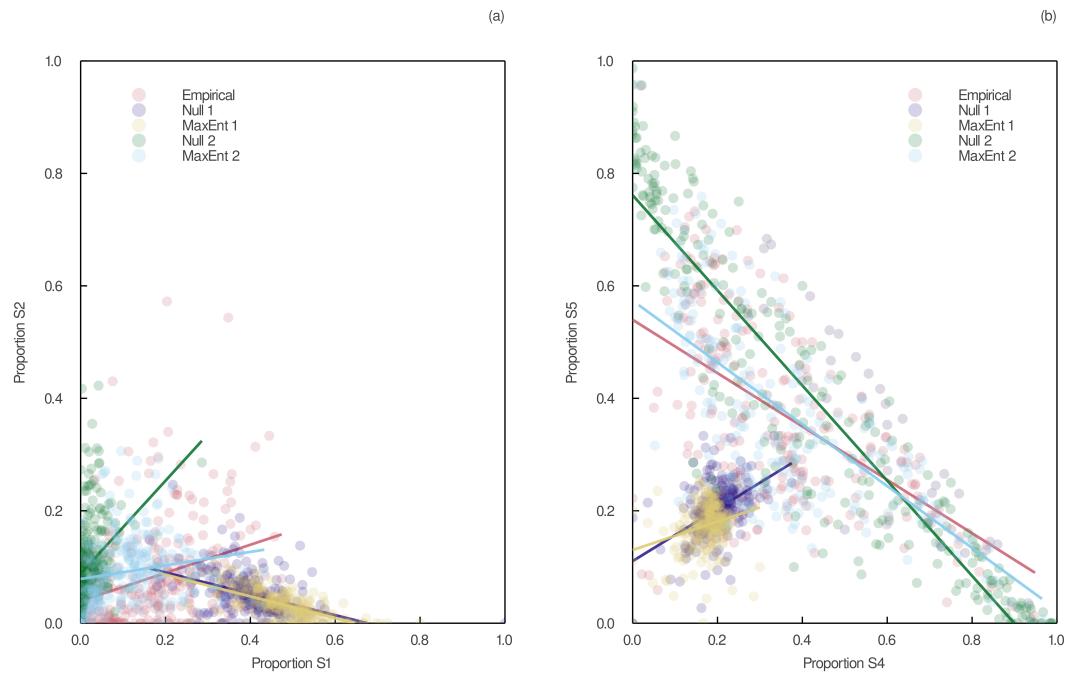


Figure 5: Relationship between motifs proportions of empirical, maximum entropy, and null food webs. Empirical networks include all food webs archived on Mangal, as well as the New-Zealand and Tuesday lake datasets (our complete dataset). Maximum entropy networks were estimated using a simulated annealing 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).

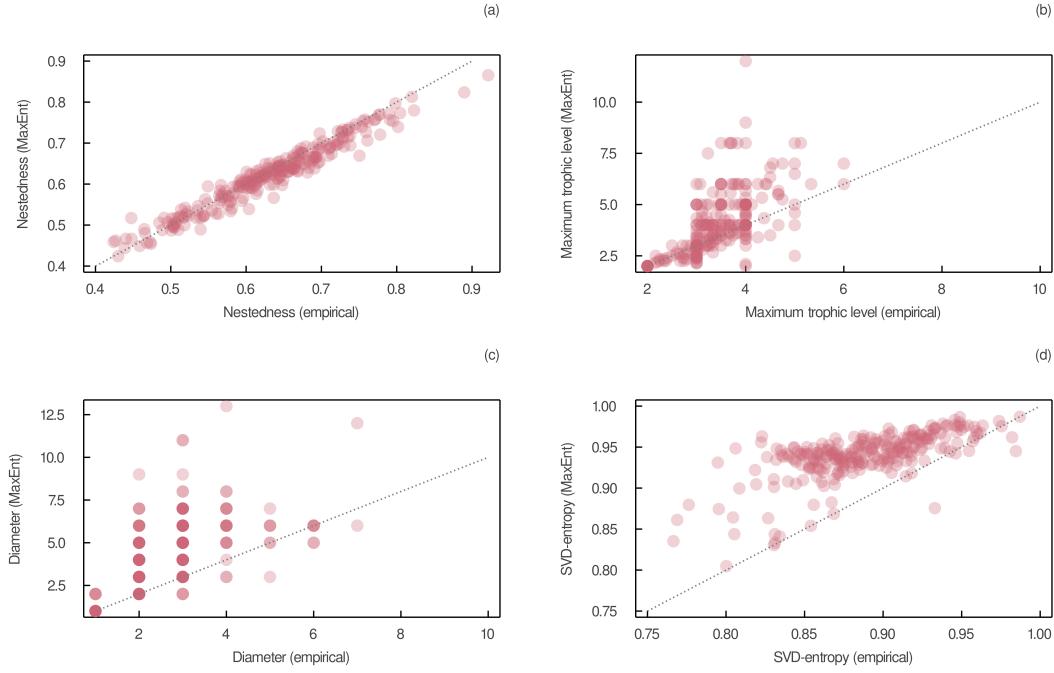


Figure 6: Relationship between the structure of empirical and maximum entropy food webs. Empirical networks include all food webs archived on Mangal, as well as the New-Zealand and Tuesday lake datasets (our complete dataset). Maximum entropy networks were estimated using a simulated annealing algorithm 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.