

Maximum entropy modelling of food-web structure

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

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

Correspondance to:

Francis Banville — francis.banville@umontreal.ca

Abstract: The principle of maximum entropy is a rigorous mathematical method of finding constrained probability distributions that has been proven useful in many ecological modelling problems. However, despite its broad application in graph and network theory, it has seldom been used to model ecological networks. Here we show how the (joint) degree distribution of maximum entropy can be directly derived using the number of species and the number of links in terrestrial and aquatic food webs. We also present a heuristic and flexible approach of finding the network of maximum entropy based on simulating annealing and SVD entropy. We built two of these network-level models using constraints given by the connectance and the joint degree sequence, respectively. All maximum entropy models were compared against open access food-web data and null and neutral models commonly used in network ecology. We found that the maximum entropy network model constrained by the joint degree sequence was a good predictor of food-web structure, including nestedness and motifs distribution. Overall, our results suggest that many properties of ecological networks are mainly driven by the joint degree distribution and statistical phenomena.

1 Introduction

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

15 The principle of maximum entropy (MaxEnt) is a statistical and topological model that can be
16 used for both of these purposes, i.e. to make predictions of network structure and to better un-
17 derstand processes shaping ecological networks. This mathematical method, briefly presented
18 in Box 1, has been used in a wide range of disciplines, from thermodynamics to chemistry and
19 biology (Martyushev and Seleznev 2006). It has also been proven useful in ecology, e.g. in
20 species distribution models (Phillips, Anderson, and Schapire 2006) and macroecological mod-
21 els (Harte et al. 2008; Harte and Newman 2014). As discussed in Box 1, maximizing a measure
22 of entropy ensures that the derived probability distributions are unique and least biased under
23 the set of constraints used. These constraints are built using state variables, i.e. variables that
24 represent the macrostate of the system. The challenge is to find the set of state variables that best
25 represent natural systems and to translate them into appropriate statistical constraints. Having
26 a validated maximum entropy model for the system at hand allows us to make rigorous predic-
27 tions using a minimal amount of data, as well as helping us describe the most important factors
28 driving that system.

29 Despite its extensive use in graph and network theories (e.g., Park and Newman 2004; van der
30 Hoorn, Lippner, and Krioukov 2018), MaxEnt has in comparison been little used in network
31 ecology. The very nature of ecological networks (directed simple graphs frequently having self-
32 loops) makes the mathematical optimization of maximum entropy graph models more com-
33 plicated than with many other types of (non-ecological) networks. MaxEnt has nevertheless
34 been used to predict the degree distribution of bipartite ecological networks from the number
35 of species and the number of interactions (Williams 2011) and to predict interaction strengths
36 between species pairs using their relative abundances within an optimal transportation theory
37 regularized with information entropy (Stock, Poisot, and De Baets 2021). However, to the best of
38 our knowledge, MaxEnt has never been used to predict food-web structure directly, even though
39 food webs are among the most documented and widespread ecological networks.

40 In this contribution, we used two complementary approaches to predict the structure of food
41 webs using the principle of maximum entropy. We then compared our predictions against em-
42 pirical data and null and neutral models commonly used in network ecology. The first approach
43 consists in deriving constrained probability distributions of given network properties directly.
44 We derived the joint degree distribution (a probability distribution) of maximum entropy using
45 only the number of species S and the number of interactions L as state variables. Then, we pre-
46 dicted the degree distribution of maximum entropy directly from the joint degree distribution
47 since the first is the sum of the marginal distributions of the second (a species' degree is the
48 sum of its in and out-degrees). Because of the scarcity of empirical data on the number of links
49 in ecological networks, in Box 2 we present a method to predict L from S , thus allowing the
50 prediction of the joint degree distribution from S solely. In turn, the second approach consists
51 in finding, using different constraints, the adjacency matrix of maximum entropy from which
52 network properties can be measured. To do so, we used a flexible and heuristic approach based
53 on simulating annealing to find networks *close* to maximum entropy. As discussed above, our
54 choice of algorithm stands from the very nature of food webs (i.e., simple directed networks
55 allowing self-loops) that makes the analytical derivation of a maximum entropy graph model
56 difficult. We first built our type I MaxEnt network model constrained by the connectance of
57 the network (i.e., the ratio L/S^2). A comparison of this model against empirical data indicated

58 that connectance alone was not sufficient to predict many aspects of network structure. For this
59 reason, we built our type II MaxEnt network model, which instead uses the whole joint degree
60 sequence as a constraint. Overall, we found that this second model was much better at predicting
61 food-web structure than the one based on connectance.

62 **Box 1 - The principle of maximum entropy: A primer for ecologists**

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

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

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

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

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

89 where g_i is the mathematical formulation of the constraint i and c_i , its value. Note that F is just
 90 Shannon's entropy to which we added terms that each sums to zero ($g_i = c_i$). F is maximized by
 91 setting to 0 its partial derivative with respect to $p(n)$. We will show how this can be done when
 92 we derive the joint degree distribution analytically from the number of species and the number
 93 of links in food webs.

94 In this contribution, we also use the SVD entropy as a measure of entropy, which is an applica-
 95 tion of Shannon's entropy to the relative non-zero singular values of a truncated singular value
 96 decomposition (t-SVD; Strydom, Dalla Riva, and Poisot 2021) of a food web's Boolean ad-
 97 jacency matrix. This measure also satisfies all four properties above-mentioned, while being a
 98 proper measure of the internal complexity of food webs (Strydom, Dalla Riva, and Poisot 2021).

99 We measured SVD entropy as follows:

$$J = - \sum_{i=1}^R s_i \log s_i, \quad (3)$$

100 where s_i are the relative singular values ($s_i = \sigma_i / \sum_{i=1}^R \sigma_i$, where σ_i are the singular values).
 101 Following Strydom, Dalla Riva, and Poisot (2021), we standardized this measure with the rank
 102 R of the matrix (i.e., $J / \ln(R)$) to account for the difference in dimensions between networks
 103 (Pielou's evenness; Pielou 1975). In a following section, we will show how SVD entropy can
 104 be used to predict a network of maximum entropy (i.e., of maximum complexity) heuristically.

105 **Testing MaxEnt models**

106 **Data**

107 We tested our MaxEnt models (both approaches) against open food-web data queried from three
108 different sources and integrated into what we call our *complete dataset*. First, all food webs
109 archived on `mangal.io` were directly queried from the database ($n = 235$). Most ecological
110 networks archived on Mangal are multilayer networks, i.e. networks that describe different types
111 of interactions. We kept all networks whose interactions were mainly of predation and herbivory
112 types, and removed the largest network ($S = 714$) for computational efficiency reasons. Then,
113 to this set we added food webs from two different sources: the New-Zealand dataset ($n = 21$;
114 Pomeranz et al. 2018) and the Tuesday lake dataset ($n = 2$; Cohen, Jonsson, and Carpenter
115 2003). Of these two datasets, 19 networks had data on species' relative abundances that were
116 used in the neutral model presented in a following subsection. These networks are part of what
117 we call our *abundance dataset*, which is a subset of our complete dataset.

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

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

121 Our maximum entropy network models (second approach only) were compared with two topo-
122 logical null models. The first is the type I null model of Fortuna and Bascompte (2006), in which
123 the probability that a species i predares on another species j is given by

$$p(i \rightarrow j) = \frac{L}{S^2}. \quad (4)$$

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

$$p(i \rightarrow j) = \frac{1}{2} \left(\frac{k_{in}(j)}{S} + \frac{k_{out}(i)}{S} \right), \quad (5)$$

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

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

136 Neutral model

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

$$p(i \rightarrow j) \propto \frac{n_i}{N} \times \frac{n_j}{N}, \quad (6)$$

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

143 Analytical models: Measures of maximum entropy

144 Joint degree distribution

145 The joint degree distribution $p(k_{in}, k_{out})$ is a joint discrete probability distribution describing
146 the probability that a species has k_{in} predators and k_{out} preys, with k_{in} and $k_{out} \in [0, S]$. Basal
147 species (e.g., plants) have a k_{out} of 0, whereas top predators have a k_{in} of 0. In contrast, the

148 maximum number of preys and predators a species can have is set by the number of species S
 149 in the food web. Here we show how the joint degree distribution of maximum entropy can be
 150 obtained given knowledge of S and L .

151 We want to maximize Shannon's entropy

$$H = - \sum_{k_{in}=0}^S \sum_{k_{out}=0}^S p(k_{in}, k_{out}) \log p(k_{in}, k_{out}) \quad (7)$$

152 subject to the following constraints:

$$g_1 = \sum_{k_{in}=0}^S \sum_{k_{out}=0}^S p(k_{in}, k_{out}) = 1; \quad (8)$$

$$g_2 = \sum_{k_{in}=0}^S \sum_{k_{out}=0}^S k_{in} p(k_{in}, k_{out}) = \langle k_{in} \rangle = \frac{L}{S}; \quad (9)$$

$$g_3 = \sum_{k_{in}=0}^S \sum_{k_{out}=0}^S k_{out} p(k_{in}, k_{out}) = \langle k_{out} \rangle = \frac{L}{S}. \quad (10)$$

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

160 The joint probability distribution of maximum entropy given these constraints is found using the
 161 method of Lagrange multipliers. To do so, we seek to maximize the following expression:

$$F = H - \lambda_1(g_1 - 1) - \lambda_2 \left(g_2 - \frac{L}{S} \right) - \lambda_3 \left(g_3 - \frac{L}{S} \right), \quad (11)$$

162 where λ_1 , λ_2 , and λ_3 are the Lagrange multipliers. The probability distribution that maximizes

¹⁶³ entropy is obtained by finding these values. As pointed out in Box 1, F is just Shannon's entropy
¹⁶⁴ to which we added terms that each sums to zero (our constraints). F is maximized by setting to
¹⁶⁵ 0 its partial derivative with respect to $p(k_{in}, k_{out})$. Because the derivative of a constant is zero,
¹⁶⁶ this gives us:

$$\frac{\partial H}{\partial p(k_{in}, k_{out})} = \lambda_1 \frac{\partial g_1}{\partial p(k_{in}, k_{out})} + \lambda_2 \frac{\partial g_2}{\partial p(k_{in}, k_{out})} + \lambda_3 \frac{\partial g_3}{\partial p(k_{in}, k_{out})}. \quad (12)$$

¹⁶⁷ Evaluating the partial derivatives with respect to $p(k_{in}, k_{out})$, we obtain:

$$-\log p(k_{in}, k_{out}) - 1 = \lambda_1 + \lambda_2 k_{in} + \lambda_3 k_{out}. \quad (13)$$

¹⁶⁸ Then, solving eq. 13 for $p(k_{in}, k_{out})$, we obtain:

$$p(k_{in}, k_{out}) = \frac{e^{-\lambda_2 k_{in} - \lambda_3 k_{out}}}{Z}, \quad (14)$$

¹⁶⁹ where $Z = e^{1+\lambda_1}$ is called the partition function. The partition function ensures that probabilities
¹⁷⁰ sum to 1 (our normalization constraint). It can be expressed in terms of λ_2 and λ_3 as follows:

$$Z = \sum_{k_{in}=0}^S \sum_{k_{out}=0}^S e^{-\lambda_2 k_{in} - \lambda_3 k_{out}}. \quad (15)$$

¹⁷¹ After substituting $p(k_{in}, k_{out})$ in eq. 9 and eq. 10, we get a nonlinear system of two equations and
¹⁷² two unknowns:

$$\frac{1}{Z} \sum_{k_{in}=0}^S \sum_{k_{out}=0}^S k_{in} e^{-\lambda_2 k_{in} - \lambda_3 k_{out}} = \frac{L}{S}; \quad (16)$$

$$\frac{1}{Z} \sum_{k_{in}=0}^S \sum_{k_{out}=0}^S k_{out} e^{-\lambda_2 k_{in} - \lambda_3 k_{out}} = \frac{L}{S}. \quad (17)$$

¹⁷³ We solved eq. 16 and eq. 17 numerically using the Julia library JuMP.jl v0.21.8 (Dunning,
¹⁷⁴ Huchette, and Lubin 2017). JuMP.jl supports nonlinear optimization problems by providing

175 exact second derivatives that increase the accuracy and performance of its solvers. The estimated
176 values of λ_2 and λ_3 can be substituted in eq. 14 to have a more workable expression for the joint
177 degree distribution.

178 We predicted the joint degree distribution of maximum entropy for each food web in our com-
179 plete dataset, i.e. using their numbers of species and numbers of links as state variables. We then
180 sampled one realization of the degree sequence for each network using the probabilities given
181 by the joint degree distribution. In fig. 1 (left panels), we show the relationship between k_{out} and
182 k_{in} standardized by the number of species in their networks, for empirical and maximum entropy
183 joint degree distributions. We see that our model predicts a greater number of generalist species
184 than empirical food webs (i.e., relative values of k_{out} and k_{in} closer to 1). However, plotting the
185 difference between predicted and empirical values for each species gives a different perspective.
186 The right panel of fig. 1 presents these differences when species are ordered by their total degree
187 in their network (i.e., by the sum of their in and out-degrees). Indeed, our predicted joint degree
188 sequences have the same number of species as their empirical counterparts, but they are species
189 agnostic; in other words, instead of predicting a value for each species directly, we predicted
190 the entire joint degree sequence without taking into account species' identity. When we asso-
191 ciate predictions and empirical data according to their rank in total degrees, we see that species
192 predicted to be have a higher generality (number of preys) generally have a lower vulnerability
193 (number of preys) than what is observed (and conversely). In fig. S1, we show how these differ-
194 ences change when species are instead ordered by their out-degrees (left panel) and in-degrees
195 (right panel), respectively.

196 [Figure 1 about here.]

197 We plotted the Kullback–Leibler divergence between in and out-degree distributions to compare
198 the symmetry of empirical and maximum entropy joint degree sequences fig. 2. As we expected,
199 our model predicted more similar in-degree and out-degree distributions than empirical data.
200 However, this difference decreased with connectance right panel of 2. Overall, this suggests that
201 other ecological constraints might be needed to account for the asymmetry of the joint degree
202 distribution, especially for networks with a lower connectance.

203

[Figure 2 about here.]

204 **Degree distribution**

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

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

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

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

$$p(k) = \frac{e^{-\lambda_2 k}}{Z}, \quad (18)$$

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

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

$$\frac{1}{Z} \sum_{k=0}^S k e^{-\lambda_2 k} = \frac{2L}{S} \quad (19)$$

220 The numerical solution is identical to the one we obtained using the joint degree distribution
 221 as an intermediate. Ecologists wanting to model a system without considering isolated species

222 could simply change the lower limit of k to 1 and solve the resulting equation numerically.

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

233 Box 2 - Working with predicted numbers of links

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

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

$$L_{FL} \sim \text{BB}(S^2 - (S - 1), \mu e^\phi, (1 - \mu)e^\phi), \quad (20)$$

246 where μ is the average probability across food webs that a flexible link is realized, and ϕ is the
247 concentration parameter around μ .

248 We fitted the flexible links model on all food webs in our complete dataset, and estimated the
249 parameters of eq. 20 using a Hamiltonian Monte Carlo sampler with static trajectory (1 chain
250 and 3000 iterations):

$$[\mu, \phi | \mathbf{L}, \mathbf{S}] \propto \prod_{i=1}^m \text{BB}(L_i - (S_i - 1) | S_i^2 - (S_i - 1)), \mu e^\phi, (1 - \mu) e^\phi \times \text{B}(\mu | 3, 7) \times \mathcal{N}(\phi | 3, 0.5), \quad (21)$$

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

256 The flexible links model is a generative model, i.e. it can generate plausible values of the pre-
257 dicted variable. We thus simulated 1000 values of L for different values of S using the joint
258 posterior distribution of our model parameters, and calculated the mean degree for each simu-
259 lated values. The resulting distributions are shown in the left panel of fig. 3 for three different
260 values of species richness. In the right panel of fig. 3, we show how the probability distribution
261 for the mean degree constraints can be used to generate a distribution of maximum entropy de-
262 gree distributions, since each simulated value of mean degree generates a different maximum
263 entropy degree distribution.

264 [Figure 3 about here.]

265 **Heuristical models: Networks of maximum entropy**

266 **MaxEnt network models (types I and II)**

267 We define networks of maximum entropy as the configuration of the adjacency matrix with the
268 highest SVD entropy under a set of constraints. As mentioned in Box 1, we used the SVD
269 entropy as our measure of entropy since it has been shown to be a reliable measure of food-web
270 complexity (Strydom, Dalla Riva, and Poisot 2021), in addition to having the required properties
271 of a proper measure of information entropy. We thus seek to find the network with the highest
272 complexity, or randomness, that exactly reproduces specified constraints on its structure. Our
273 method is in contrast with maximum entropy graph models that predict a probability distribution
274 on networks under soft or hard constraints Cimini et al. (2019). We believe our approach to be
275 more flexible, easier to compute, while allowing direct comparisons of empirical food webs with
276 more complex networks with similar structure.

277 We built two types of MaxEnt network models: one based on connectance (type I MaxEnt net-
278 work model) and the other based on the joint degree sequence (type II MaxEnt network model).
279 They are based on the same constraints as the types I and II null models presented above. For
280 both models, we used a simulating annealing algorithm with 4 chains, 2000 steps and an initial
281 temperature of 0.2. For each chain, we first generated one random Boolean matrix with the same
282 order (number of species) as empirical webs, while maintaining the total number of interactions
283 (type I MaxEnt network model) or rows and columns sums (type II MaxEnt network model).
284 These are our initial configurations. Then, we swapped interactions sequentially while main-
285 taining the original connectance or the joint degree sequence for types I and II MaxEnt network
286 model, respectively. Configurations with a higher SVD entropy than the previous one in the
287 chain were always accepted, whereas they were accepted with a probability conditional to a de-
288 creasing temperature when lower. The final configuration with the highest SVD entropy among
289 the four chains constitute our estimated MaxEnt network. Even though we decided to work with
290 point estimates, it is possible to have a (non MaxEnt) probability distribution of networks when
291 working with the entire chains after burn-in. For each network in our complete and abundance
292 datasets, we estimated their configuration with maximum entropy using both types of MaxEnt

293 network models.

294 **Structure of MaxEnt networks**

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

306 Overall, we found that models based on the joint degree sequence (type II null and MaxEnt
307 network models) reproduced the structure of empirical networks much better than the ones based
308 on connectance and the neutral model of relative abundances. On average, the type II MaxEnt
309 network model was better at predicting nestedness (0.62 ± 0.08) than its corresponding null model
310 (0.73 ± 0.05 ; empirical networks: 0.63 ± 0.09), as well as the proportion of cannibal species.
311 However, the type II null model was better at predicting network diameter and average maximum
312 similarity between species pairs. Predictions were similar between both type II models for the
313 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. ρ : nestedness measured by the spectral radius of the adjacency matrix. *maxtl*: maximum trophic level. *diam*: network diameter. *MxSim*: average maximum similarity between species pairs. *Cannib*: proportion of cannibal species (self loops). *Omniv*: proportion of omnivorous species. *entropy*: SVD entropy.

model	rho	maxtl	diam	MxSim	Cannib	Omniv	entropy
null 1	-0.167	0.980	1.428	-0.502	2.007	1.493	0.056
MaxEnt 1	-0.226	0.831	1.274	-0.524	1.982	1.863	0.106
null 2	0.160	-0.125	0.016	0.007	1.078	0.559	-0.023
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. ρ : nestedness measured by the spectral radius of the adjacency matrix. *maxtl*: maximum trophic level. *diam*: network diameter. *MxSim*: average maximum similarity between species pairs. *Cannib*: proportion of cannibal species (self loops). *Omniv*: proportion of omnivorous species. *entropy*: SVD entropy.

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

³¹⁴ The picture slightly changes when we consider another important property of ecological networks, i.e. their motifs profile. We measured the proportion of three-species motifs, which can

316 be considered as simple building blocks of ecological networks (Milo et al. 2002; Stouffer et al.
317 2007). Motifs are the backbone of complex ecological networks from which network structure
318 is built upon and play a crucial role in community dynamics and assembly (Bastolla et al. 2009;
319 Stouffer and Bascompte 2011). In fig. 4, we show that the motifs profile of networks generated
320 using the type II MaxEnt network model was very close to the one of empirical data, and that
321 they made better predictions than the type II null model. This is also shown in fig. 5, where we
322 see that relationships between motifs pairs of empirical food webs are very similar to the ones
323 of the type II MaxEnt networks. Overall, these results suggest that the type II MaxEnt network
324 model can reproduce many aspects of network structure. This highlights the importance of the
325 joint degree sequence in shaping ecological networks.

326 [Figure 4 about here.]

327 [Figure 5 about here.]

328 Moreover, we found that empirical networks are close to their maximum entropy considering a
329 fixed joint degree sequence (fig. S3). Empirical networks had an SVD entropy of 0.89 ± 0.04 ,
330 compared to type II MaxEnt networks which had values of 0.94 ± 0.03 . The relationship between
331 the SVD entropy of empirical food webs and their maximum entropy configuration is plotted in
332 the last panel of fig. 6. We also found no relationship between the difference in SVD entropy
333 between empirical and type II MaxEnt networks and species richness, the number of links, or
334 connectance (fig. S4). Similarly, we found no correlation between the difference in SVD entropy
335 and the Jaccard distance of empirical and type II MaxEnt networks.

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

344 Conclusion

345 The principle of maximum entropy is a robust mathematical method of finding least-biased
346 probability distributions that have some desired properties given by prior knowledge on a sys-
347 tem. We first used MaxEnt to predict the joint degree distribution and the degree distribution
348 of maximum entropy given known numbers of species and of interactions. We found that the
349 resulting joint degree distributions were more symmetric than the ones of empirical food webs,
350 which suggests that other constraints might be needed to improve those predictions. We also
351 used MaxEnt to predict networks of maximum entropy with a specified structure. These net-
352 works are the most complex, or random, given the constraints used. Likewise, we found that
353 knowledge of species richness and of the number of links were not sufficient to reproduce many
354 aspects of network structure found in nature. However, a model based on the entire joint degree
355 sequence, i.e. on the number of preys and predators for each species, gave more convincing re-
356 sults. Our type II MaxEnt network model yielded better or similar predictions than the type II
357 null model, also based on the joint degree sequence, for most measures considered including
358 the motifs profile. This suggests that the joint degree sequence drives many aspects of network
359 structure. Indeed, considering our findings that empirical networks are close to maximum en-
360 tropy for a given joint degree sequence, our results suggest that food-web topology is configured
361 almost entirely randomly around these marginal numbers of predators and preys.

362 Our method and results could be used for different purposes. First, they could be used as first-
363 order approximations of network structure when only state variables are known. This could
364 prove useful when predicting network structure at large spatial scales, where few ecological in-
365 formation is known at that scale. Second, they could be used as informative priors in Bayesian
366 analyses of the structure of ecological networks. Third, they could be used to make better predic-
367 tions of pairwise species interactions by constraining the space of feasible networks, as discussed
368 in Strydom et al. (2021). Finally, they could be used as alternative null models of ecological net-
369 works to better understand ecological mechanisms driving food-web structure. In that case, our

370 model might need to be slightly adapted to give a probability distribution of Boolean networks
371 (in contrast with point estimates of maximum entropy networks).

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

379 Acknowledgments

380 We acknowledge that this study was conducted on land within the traditional unceded territory
381 of the Saint Lawrence Iroquoian, Anishinabewaki, Mohawk, Huron-Wendat, and Omàmiwinini-
382 wak nations. This work was supported by the Institute for Data Valorisation (IVADO) and the
383 NSERC BIOS² CREATE program.

384 References

- 385 Albert, R., and A. L. Barabasi. 2002. “Statistical Mechanics of Complex Networks.” *Reviews*
386 *of Modern Physics* 74 (1): 47–97. <https://doi.org/10.1103/RevModPhys.74.47>.
- 387 Bascompte, J., P. Jordano, C. J. Melian, and J. M. Olesen. 2003. “The Nested Assembly of
388 Plant-Animal Mutualistic Networks.” *Proceedings of the National Academy of Sciences*
389 *of the United States of America* 100 (16): 9383–87. <https://doi.org/10.1073/pnas.1633576100>.
- 390 Bastolla, Ugo, Miguel A. Fortuna, Alberto Pascual-Garcia, Antonio Ferrera, Bartolo Luque, and
391 Jordi Bascompte. 2009. “The Architecture of Mutualistic Networks Minimizes Competition
392 and Increases Biodiversity.” *Nature* 458 (7241): 1018–U91. <https://doi.org/10.1038/nature07950>.

- 395 Beck, Christian. 2009. "Generalised Information and Entropy Measures in Physics." *Contemporary Physics* 50 (4): 495–510. <https://doi.org/10.1080/00107510902823517>.
- 396
- 397 Cimini, Giulio, Tiziano Squartini, Fabio Saracco, Diego Garlaschelli, Andrea Gabrielli, and
398 Guido Caldarelli. 2019. "The Statistical Physics of Real-World Networks." *Nature Reviews
399 Physics* 1 (1): 58–71. <https://doi.org/10.1038/s42254-018-0002-6>.
- 400 Cohen, Joel E., Tomas Jonsson, and Stephen R. Carpenter. 2003. "Ecological Community
401 Description Using the Food Web, Species Abundance, and Body Size." *Proceedings of
402 the National Academy of Sciences* 100 (4): 1781–86. [https://doi.org/10.1073/pnas.
403 232715699](https://doi.org/10.1073/pnas.232715699).
- 404 Delmas, Eva, Mathilde Besson, Marie Hélène Brice, Laura A. Burkle, Giulio V. Dalla Riva,
405 Marie Josée Fortin, Dominique Gravel, et al. 2019. "Analysing Ecological Networks of
406 Species Interactions." *Biological Reviews*. <https://doi.org/10.1111/brv.12433>.
- 407 Dunning, Iain, Joey Huchette, and Miles Lubin. 2017. "JuMP: A Modeling Language for
408 Mathematical Optimization." *SIAM Review* 59 (2): 295–320. [https://doi.org/10.1137/
409 15M1020575](https://doi.org/10.1137/15M1020575).
- 410 Fortuna, M. A., and J. Bascompte. 2006. "Habitat Loss and the Structure of Plant-Animal
411 Mutualistic Networks." *Ecology Letters* 9 (3): 278–83. [https://doi.org/10.1111/j.1461-0248.2005.00868.x](https://doi.org/10.1111/j.
412 1461-0248.2005.00868.x).
- 413 Fortuna, M. A., Daniel B. Stouffer, Jens M. Olesen, Pedro Jordano, David Mouillot, Boris R.
414 Krasnov, Robert Poulin, and Jordi Bascompte. 2010. "Nestedness Versus Modularity in
415 Ecological Networks: Two Sides of the Same Coin?" *Journal of Animal Ecology* 79 (4):
416 811–17. <https://doi.org/10.1111/j.1365-2656.2010.01688.x>.
- 417 Frank, S. A., and E. Smith. 2011. "A Simple Derivation and Classification of Common Prob-
418 ability Distributions Based on Information Symmetry and Measurement Scale." *Journal of
419 Evolutionary Biology* 24 (3): 469–84. [https://doi.org/10.1111/j.1420-9101.2010.02204.x](https://doi.org/10.1111/j.1420-9101.2010.
420 02204.x).
- 421 Harremoës, Peter, and Flemming Topsøe. 2001. "Maximum Entropy Fundamentals." *Entropy*
422 3 (3): 191–226. <https://doi.org/10.3390/e3030191>.

- 423 Harte, J., and Erica A. Newman. 2014. “Maximum Information Entropy: A Foundation for
424 Ecological Theory.” *Trends in Ecology & Evolution* 29 (7): 384–89. <https://doi.org/10.1016/j.tree.2014.04.009>.
- 425
- 426 Harte, J., T. Zillio, E. Conlisk, and A. B. Smith. 2008. “Maximum Entropy and the State-
427 Variable Approach to Macroecology.” *Ecology* 89 (10): 2700–2711. <https://doi.org/10.1890/07-1369.1>.
- 428
- 429 Hoorn, Pim van der, Gabor Lippner, and Dmitri Krioukov. 2018. “Sparse Maximum-Entropy
430 Random Graphs with a Given Power-Law Degree Distribution.” *Journal of Statistical Physics*
431 173 (3): 806–44. <https://doi.org/10.1007/s10955-017-1887-7>.
- 432
- 433 Hortal, Joaquín, Francesco de Bello, José Alexandre F. Diniz-Filho, Thomas M. Lewinsohn,
434 Jorge M. Lobo, and Richard J. Ladle. 2015. “Seven Shortfalls That Beset Large-Scale
435 Knowledge of Biodiversity.” *Annual Review of Ecology, Evolution, and Systematics* 46 (1):
523–49. <https://doi.org/10.1146/annurev-ecolsys-112414-054400>.
- 436 Jaynes, E. T. 1957a. “Information Theory and Statistical Mechanics.” *Physical Review* 106 (4):
437 620–30. <https://doi.org/10.1103/PhysRev.106.620>.
- 438 ———. 1957b. “Information Theory and Statistical Mechanics. II.” *Physical Review* 108 (2):
439 171–90. <https://doi.org/10.1103/PhysRev.108.171>.
- 440 Khinchin, A. Ya. 2013. *Mathematical Foundations of Information Theory*. Courier Corpora-
441 tion.
- 442 MacDonald, Arthur Andrew Meahan, Francis Banville, and Timothée Poisot. 2020. “Revisiting
443 the Links-Species Scaling Relationship in Food Webs.” *Patterns* 0 (0). <https://doi.org/10.1016/j.patter.2020.100079>.
- 444
- 445 Martyushev, L. M., and V. D. Seleznev. 2006. “Maximum Entropy Production Principle in
446 Physics, Chemistry and Biology.” *Physics Reports-Review Section of Physics Letters* 426
447 (1): 1–45. <https://doi.org/10.1016/j.physrep.2005.12.001>.
- 448 Milo, R., S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon. 2002. “Net-
449 work Motifs: Simple Building Blocks of Complex Networks.” *Science* 298 (5594): 824–27.
450 <https://doi.org/10.1126/science.298.5594.824>.

- 451 Park, Juyong, and M. E. J. Newman. 2004. “Statistical Mechanics of Networks.” *Physical*
452 *Review E* 70 (6): 066117. <https://doi.org/10.1103/PhysRevE.70.066117>.
- 453 Phillips, Steven J., Robert P. Anderson, and Robert E. Schapire. 2006. “Maximum Entropy
454 Modeling of Species Geographic Distributions.” *Ecological Modelling* 190 (3): 231–59.
455 <https://doi.org/10.1016/j.ecolmodel.2005.03.026>.
- 456 Pielou, Evelyn C. 1975. “Ecological Diversity.” In. 574.524018 P5.
- 457 Pomeranz, Justin Page, Ross M. Thompson, Timothée Poisot, Jon S. Harding, and Justin P. F.
458 Pomeranz. 2018. “Data from: Inferring Predator-Prey Interactions in Food Webs.” Dryad.
459 <https://doi.org/10.5061/DRYAD.K59M37F>.
- 460 Shannon, C. E. 1948. “A Mathematical Theory of Communication.” *The Bell System Technical*
461 *Journal* 27 (3): 379–423. <https://doi.org/10.1002/j.1538-7305.1948.tb01338.x>.
- 462 Staniczenko, Phillip P. A., Jason C. Kopp, and Stefano Allesina. 2013. “The Ghost of Nested-
463 ness in Ecological Networks.” *Nature Communications* 4 (1): 1391. <https://doi.org/10.1038/ncomms2422>.
- 464 Stock, Michiel, Timothée Poisot, and Bernard De Baets. 2021. “Optimal Transportation Theory
465 for Species Interaction Networks.” *Ecology and Evolution* 11 (9): 3841–55. <https://doi.org/10.1002/ece3.7254>.
- 466 Stouffer, Daniel B., and Jordi Bascompte. 2011. “Compartmentalization Increases Food-Web
467 Persistence.” *Proceedings of the National Academy of Sciences* 108 (9): 3648–52. <https://doi.org/10.1073/pnas.1014353108>.
- 468 Stouffer, Daniel B., Juan Camacho, Wenxin Jiang, and Luis A. Nunes Amaral. 2007. “Evidence
469 for the Existence of a Robust Pattern of Prey Selection in Food Webs.” *Proceedings of the*
470 *Royal Society B-Biological Sciences* 274 (1621): 1931–40. <https://doi.org/10.1098/rspb.2007.0571>.
- 471 Strydom, Tanya, Michael D. Catchen, Francis Banville, Dominique Caron, Gabriel Dansereau,
472 Philippe Desjardins-Proulx, Norma R. Forero-Muñoz, et al. 2021. “A Roadmap Towards
473 Predicting Species Interaction Networks (across Space and Time).” *Philosophical Trans-*

- 478 *actions of the Royal Society B: Biological Sciences* 376 (1837): 20210063. <https://doi.org/10.1098/rstb.2021.0063>.
- 480 Strydom, Tanya, Giulio V. Dalla Riva, and Timothee Poisot. 2021. “SVD Entropy Reveals the
481 High Complexity of Ecological Networks.” *Frontiers in Ecology and Evolution* 9: 623141.
482 <https://doi.org/10.3389/fevo.2021.623141>.
- 483 Williams, Richard J. 2011. “Biology, Methodology or Chance? The Degree Distributions of
484 Bipartite Ecological Networks.” *PLOS ONE* 6 (3): e17645. <https://doi.org/10.1371/journal.pone.0017645>.
- 486 Williams, Richard J., and Neo D. Martinez. 2000. “Simple Rules Yield Complex Food Webs.”
487 *Nature* 404 (6774): 180–83. <https://doi.org/10.1038/35004572>.

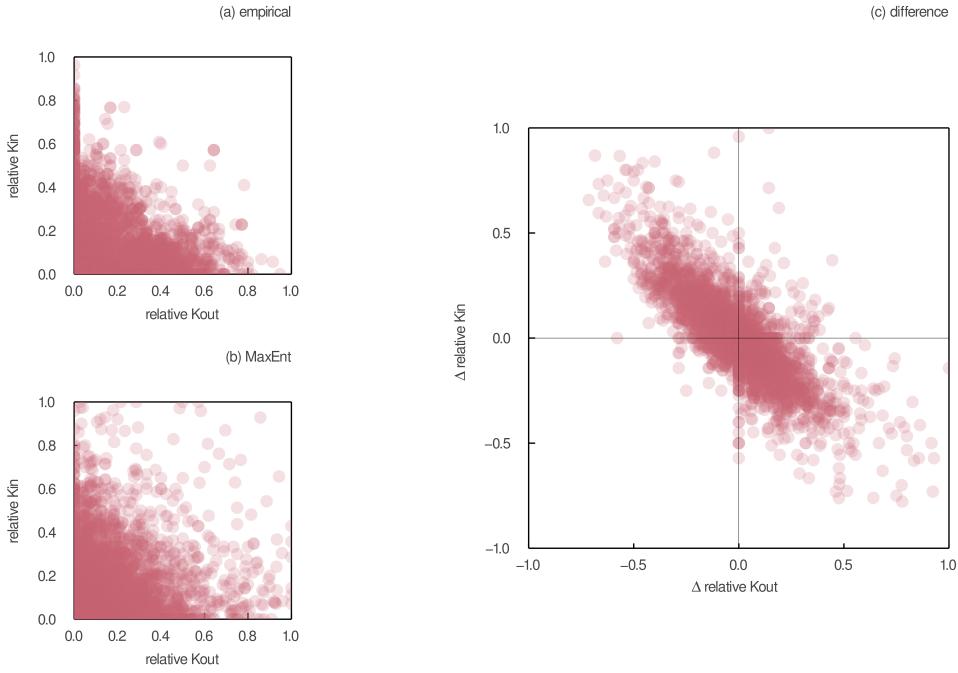


Figure 1: Relative number of predators (k_{in}) as a function of the relative number of preys (k_{out}) for each species in (a) empirical and (b) maximum entropy joint degree sequences. Empirical networks include all food webs archived on Mangal, as well as the New-Zealand and Tuesday lake datasets (our 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.

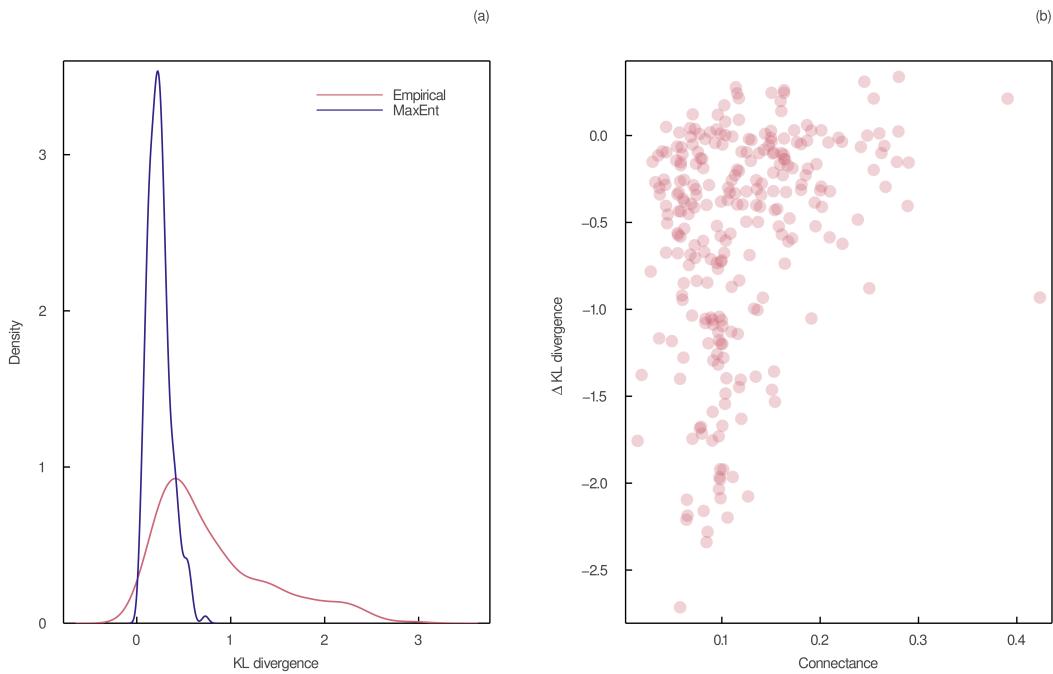


Figure 2: (a) Probability density of KL divergence between in and out-degree sequences of empirical and maximum entropy joint degree sequences. (b) Difference between the KL divergence of empirical and predicted networks as a function of connectance. In both panels, empirical networks include all food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

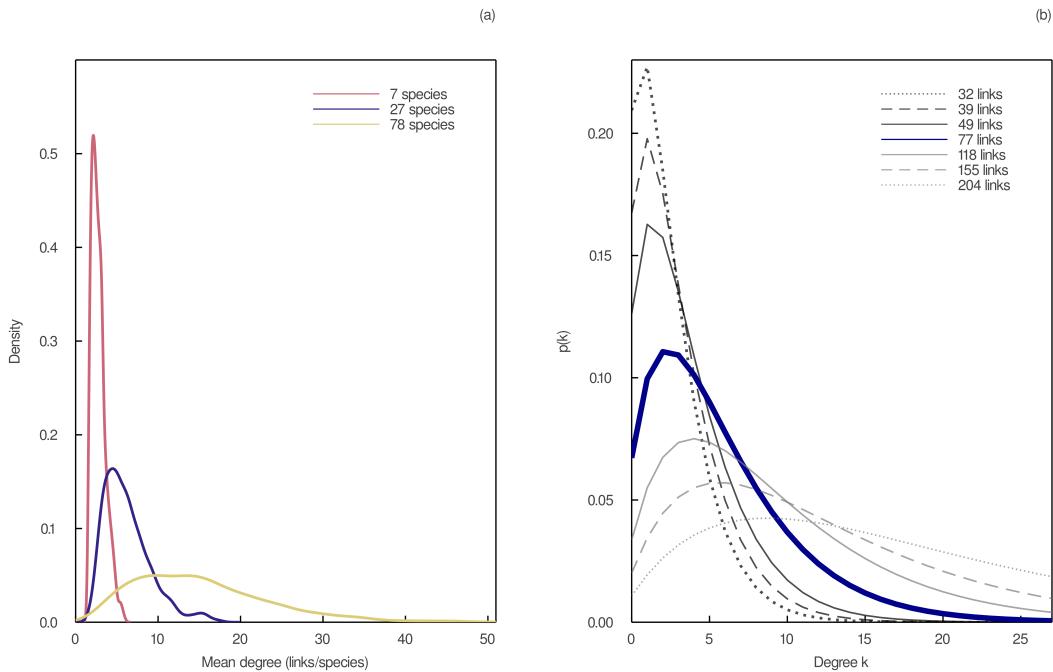


Figure 3: (a) Probability density of the predicted mean degree for different values of species richness. The number of links was predicted using the flexible links model fitted to all empirical networks in our complete dataset. (b) Degree distributions of maximum entropy for a network of 27 species and different numbers of links. The numbers of links correspond to the lower and upper bounds of the 67, 89, and 97 percentile intervals (PI), as well as the median, of the counterfactuals of the flexible links model.

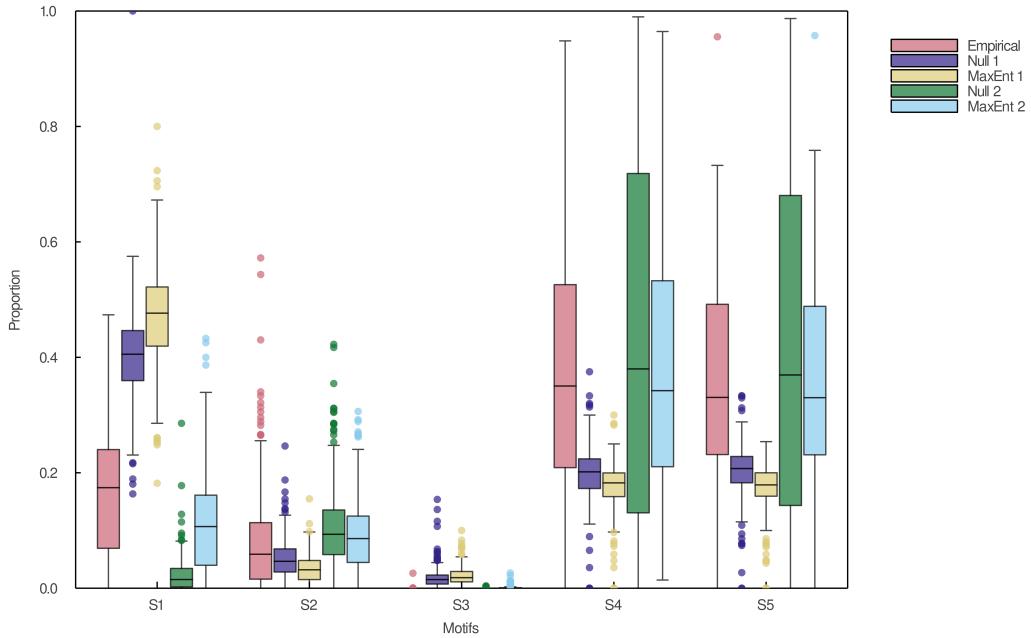


Figure 4: Motifs profile of empirical, maximum entropy, and null food webs. Empirical networks include all food webs archived on Mangal, as well as the New-Zealand and Tuesday lake datasets (our complete dataset). Maximum entropy networks were estimated using a simulating annealing algorithm to find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

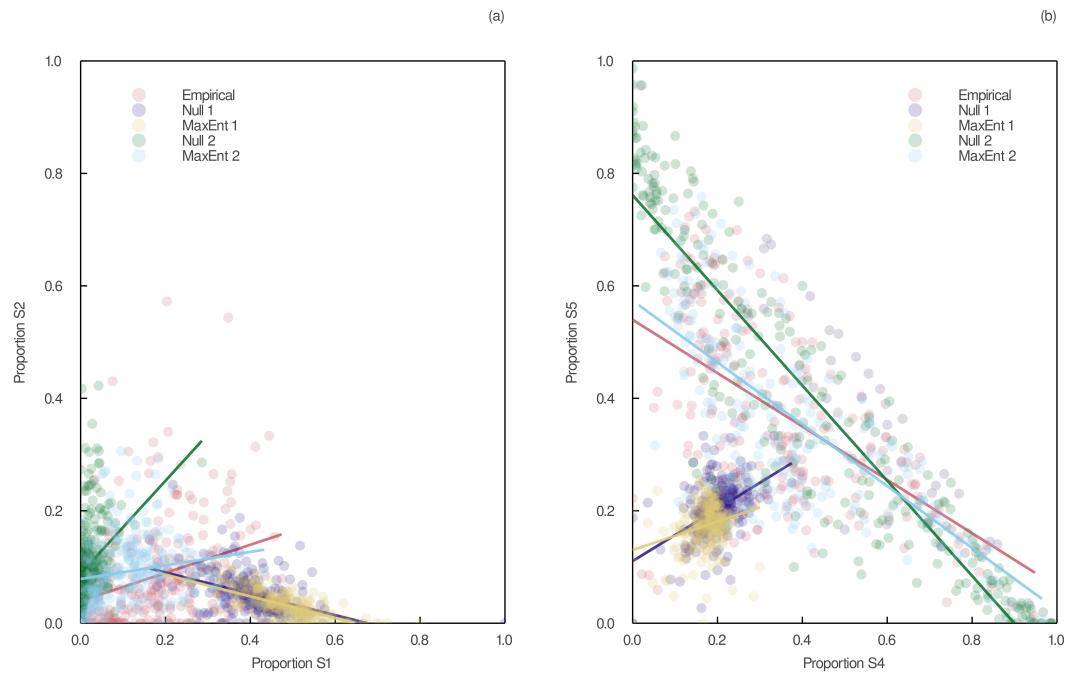


Figure 5: Relationship between motifs proportions of empirical, maximum entropy, and null food webs. Empirical networks include all food webs archived on Mangal, as well as the New-Zealand and Tuesday lake datasets (our complete dataset). Maximum entropy networks were estimated using a simulating annealing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

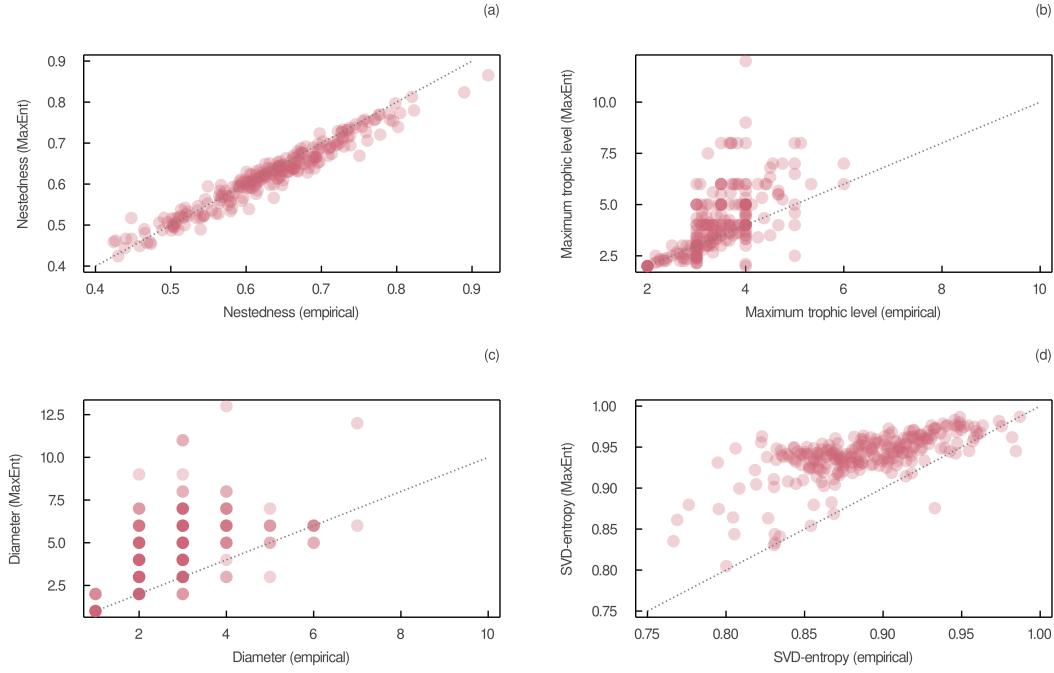


Figure 6: Relationship between the structure of empirical and maximum entropy food webs. Empirical networks include all food webs archived on Mangal, as well as the New-Zealand and Tuesday lake datasets (our complete dataset). Maximum entropy networks were estimated using a simulating annealing algorithm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