

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

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

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

Correspondance to:

Francis Banville — francis.banville@umontreal.ca

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

1 Introduction

2 The constrained structure of ecological networks

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

116 Analytical and heuristic approaches

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

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

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

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

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

145 Analytical maximum entropy models

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

We tested our analytical MaxEnt model against open food-web data queried from three different sources and integrated into what we call our *complete dataset*. These sources include (1) terrestrial and aquatic food webs sampled globally and archived on the ecological interactions database Mangal; (2) freshwater stream food webs from New Zealand; and (3) aquatic food webs from Tuesday Lake (Michigan, United States). First, all food webs archived on mangal.io

165 (Poisot et al. 2016; Banville, Vissault, and Poisot 2021) were directly queried from the database
 166 ($N = 235$). Most ecological networks archived on Mangal are multilayer networks, i.e. net-
 167 works that describe different types of interactions. We kept all networks whose interactions
 168 were mainly of predation and herbivory types, and removed the largest network ($S = 714$)
 169 for computational efficiency reasons. Then, to this set we added food webs from two different
 170 sources: the New Zealand dataset ($N = 21$; J. P. Pomeranz et al. 2018) and the Tuesday Lake
 171 dataset ($N = 2$; Cohen, Jonsson, and Carpenter 2003). Our complete dataset thus contained
 172 a total of 257 food webs. All code and data to reproduce this article are available at the Open
 173 Science Framework (OSF.IO/KT4GS). Data cleaning, simulations and analyses were conducted
 174 in Julia v1.8.0.

175 **Joint degree distribution**

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

183 We want to maximize Shannon's entropy

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

184 subject to the following constraints:

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

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

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

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

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

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

194 where λ_1 , λ_2 , and λ_3 are the Lagrange multipliers. The probability distribution that maximizes
 195 entropy is obtained by finding these values. As pointed out above, F is just Shannon's entropy
 196 to which we added terms that each sums to zero (our constraints). F is maximized by setting to
 197 0 its partial derivative with respect to $p(k_{in}, k_{out})$. Because the derivative of a constant is zero,
 198 this gives us:

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

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

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

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

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

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

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

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

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

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

205 We solved eq. 12 and eq. 13 numerically using the Julia library JuMP.jl v0.21.8 (Dunning,
 206 Huchette, and Lubin 2017). JuMP.jl supports nonlinear optimization problems by providing
 207 exact second derivatives that increase the accuracy and performance of its solvers. The estimated
 208 values of λ_2 and λ_3 can be substituted in eq. 10 to have a more workable expression for the joint
 209 degree distribution.

210 We assessed the empirical support of this expression using all food webs in our complete dataset.
 211 First, we predicted the joint degree distribution of maximum entropy for each of these food webs,
 212 i.e. using their number of species and number of interactions as state variables. Then, we sam-
 213 pled one realization of the joint degree sequence for each network using the probabilities given
 214 by the joint degree distribution of maximum entropy, while fixing the total number of interac-
 215 tions. This gave us a random realization of the number of prey and predators for each species in
 216 each network. We standardized the predicted k_{out} and k_{in} of each species by the total number of
 217 species in their network to generate relative values, which can be compared across networks. In
 218 fig. 1 (left panels), we show the relationship between these relative k_{out} and k_{in} obtained from the
 219 joint degree distributions of maximum entropy (bottom panel) and this relationship using em-
 220 pirical values (top panel). We observe that our model predicts higher values of generality and
 221 vulnerability compared to empirical food webs (i.e. relative values of k_{out} and k_{in} both closer to

222 1) for many species. In other words, our model predicts that species that have many predators
223 also have more prey than what is observed empirically (and conversely). This is not surprising,
224 given that our model did not include biological factors preventing generalist predators from hav-
225 ing many prey. Nevertheless, with the exception of these generalist species, MaxEnt adequately
226 predicts that most species have low generality and vulnerability values.

227 Examining the difference between predicted and empirical values for each species gives a slightly
228 different perspective (right panel of fig. 1). To do so, we must first associate each of our predic-
229 tions to a specific species in a network in order to make that comparison. Indeed, our predicted
230 joint degree sequences have the same number of species (elements) as their empirical counter-
231 parts, but they are species agnostic. In other words, instead of predicting a value for each species
232 directly, we predicted the entire joint degree sequence without taking into account species' iden-
233 tity. The challenge is thus to adequately associate predictions with empirical data. In fig. 1, we
234 present these differences when species are ordered by their total degree in their respective net-
235 works (i.e. by the sum of their in and out-degrees). This means that the species with the highest
236 total degree in its network will be associated with the highest prediction, and so forth. Doing so,
237 we see that species predicted to have a higher number of predators than what is observed gen-
238 erally have a lower number of prey than what is observed (and conversely). This is because the
239 difference in total degree ($k_{out} + k_{in}$) between predictions and empirical data is minimized when
240 species are ranked by their total degree (i.e. the average deviation of the sum of relative k_{out} and
241 k_{in} is close to 0 across all species). This result thus shows that the difference between predicted
242 and empirical total degrees is low for most species when ordered by their total degrees. There
243 are no apparent biases towards in or out degrees. In fig. S1, we show how these differences
244 change when species are instead ordered by their out-degrees (left panel) and in-degrees (right
245 panel), respectively.

246 [Figure 1 about here.]

247 Another way to evaluate the empirical support of the predicted joint degree sequences is to com-
248 pare their shape with the ones of empirical food webs. We can describe the shape of a joint degree
249 sequence by comparing its marginal distributions with one another. To do so, we calculated the

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

266 [Figure 2 about here.]

267 **Degree distribution**

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

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

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

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

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

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

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

279 One aspect of the degree distribution that informs us of its ecological realism is the number of
 280 isolated species it predicts. As MacDonald, Banville, and Poisot (2020) pointed out, the size of
 281 food webs should at least be of $S - 1$ interactions, since a lower number would yield isolated
 282 species, i.e. species without any predators or prey. Because non-basal species must eat to survive,
 283 isolated species could indicate that other species are missing or they could simply be removed
 284 from the food web. In fig. S2, we show that the degree distribution of maximum entropy, given
 285 S and L , gives very low probabilities that a species will be isolated in its food web (i.e. having
 286 $k = 0$) above the $S - 1$ threshold. However, under our purely information-theoretic model,
 287 the probability that a species is isolated is quite high when the total number of interactions is
 288 below $S - 1$. Moreover, the expected proportion of isolated species rapidly declines by orders
 289 of magnitude with increasing numbers of species and interactions. This supports the ecological
 290 realism of the degree distribution of maximum entropy derived above. Nevertheless, ecologists
 291 wanting to model a system without allowing isolated species could simply change the lower limit
 292 of k to 1 in eq. 15 and solve the resulting equation numerically.

293 Box 1 - Working with predicted numbers of interactions

294 Our analytical MaxEnt models require information on the number of species and the number of
 295 interactions. However, since the later is rarely measured empirically, ecologists might need to

296 use predictive models to estimate the total number of interactions in a food web before using
297 MaxEnt. Here we illustrate how this can be done by combining both models sequentially.

298 We used the flexible links model of MacDonald, Banville, and Poisot (2020) to predict the
299 number of interactions from the number of species. The flexible links model, in contrast to other
300 predictive models of the number of interactions, incorporates meaningful ecological constraints
301 into the prediction of L , namely the minimum $S - 1$ and maximum S^2 numbers of interactions
302 in food webs. It estimates the proportion of the $S^2 - (S - 1)$ *flexible links* that are realized.
303 More precisely, this model states that the number of *realized* flexible links (or interactions)
304 L_{FL} in a food web represents the number of realized interactions above the minimum (i.e. $L =$
305 $L_{FL} + S - 1$) and is obtained from a beta-binomial distribution with $S^2 - (S - 1)$ trials and
306 parameters $\alpha = \mu e^\phi$ and $\beta = (1 - \mu) e^\phi$:

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

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

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

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

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

317 The flexible links model is a generative model, i.e. it can generate plausible values of the pre-

318 dictated variable. We thus simulated 1000 values of L for different values of S using the joint
 319 posterior distribution of our model parameters (eq. 17), and calculated the mean degree for each
 320 simulated value. The resulting distributions are shown in the left panel of fig. 3 for three different
 321 values of species richness. In the right panel of fig. 3, we show how the probability distribution
 322 for the mean degree constraints can be used to generate a distribution of maximum entropy de-
 323 gree distributions, since each simulated value of mean degree generates a different maximum
 324 entropy degree distribution (eq. 14 and eq. 15).

[Figure 3 about here.]

326 Box 2 - Corresponding null and neutral models

327 Null models (types I and II)

328 The predictions of our heuristic maximum entropy models were compared against two topologi-
 329 cal null models. These null models use the same ecological information as our heuristic models
 330 and thus constitute an adequate baseline for comparison. The first is the type I null model of
 331 Fortuna and Bascompte (2006), in which the probability that a species i preda-
 332 tes on another species j is given by

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

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

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

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

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

346 Neutral model

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

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

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

353 Heuristic maximum entropy models

354 With the analytical approach, we showed how important measures of food-web structure (e.g.,
355 the degree distribution and the joint degree distribution) can be derived with the principle of
356 maximum entropy using minimal knowledge about a biological community. This type of mod-
357 els, although useful to make least-biased predictions of many network properties, can be hard
358 to apply for other measures. Indeed, there are dozens of measures of network structure (Del-
359 mas et al. 2019) and many are not directly calculated with mathematical equations, but are
360 instead estimated algorithmically. Moreover, the applicability of this method to empirical sys-
361 tems is limited by the state variables we can actually measure and use. Here, we propose a more

362 flexible method to predict many measures of network structure simultaneously, i.e. by finding
363 heuristically the network configuration having maximum entropy given partial knowledge of its
364 emerging structure.

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

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

381 Shannon's entropy can only be calculated on conventional probability distributions such as the
382 joint degree distribution. This is an issue when working with the adjacency matrix of ecological
383 networks. For this reason, we need to use another measure of entropy if we want to predict a
384 network's configuration directly using MaxEnt. We used the SVD entropy as our measure of
385 entropy, which is an application of Shannon's entropy to the relative non-zero singular values
386 of a truncated singular value decomposition (t-SVD; Strydom, Dalla Riva, and Poisot 2021) of
387 a food web's Boolean adjacency matrix. We measured SVD entropy as follows:

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

388 where s_i are the relative singular values of the adjacency matrix ($s_i = \sigma_i / \sum_{i=1}^R \sigma_i$, where σ_i
 389 are the singular values). Note that the distribution of relative singular values is analogous to
 390 a probability distribution, with $0 < s_i < 1$ and $\sum s_i = 1$. This measure also satisfies all four
 391 properties of an appropriate entropy measure above-mentioned, while being a proper measure of
 392 the internal complexity of food webs (Strydom, Dalla Riva, and Poisot 2021). Following Stry-
 393 dom, Dalla Riva, and Poisot (2021), we standardized this measure with the rank R of the matrix
 394 (i.e. $J / \ln(R)$) to account for the difference in dimensions between networks (*sensu* Pielou's
 395 evenness; Pielou 1975).

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

397 We used SVD entropy to predict the network configuration of maximum entropy (i.e. of max-
 398 imum complexity) heuristically given different constraints for all food webs in our complete
 399 dataset. We built two types of heuristic MaxEnt models that differ on the constraint used. The
 400 type I heuristic MaxEnt model is based on connectance, whereas the type II heuristic MaxEnt
 401 model is based on the joint degree sequence. These models are thus based on the same con-
 402 straints as the types I (Fortuna and Bascompte 2006) and II (Bascompte et al. 2003) null models
 403 (Box 2) frequently used to generate random networks topologically. This allows direct com-
 404 parison of the performance of null and heuristic MaxEnt models in reproducing the emerging
 405 structure of empirical food webs.

406 For each network in our complete dataset, we estimated their configuration of maximum entropy
 407 given each of these constraints. For both types of heuristic MaxEnt models, we used a simu-
 408 lated annealing algorithm with 4 chains, 2000 steps and an initial temperature of 0.2. For each
 409 food web, we first generated one random Boolean matrix per chain while fixing the number of
 410 species. We also maintained the total number of interactions (i.e. the sum of all elements in the
 411 matrix) in the type I MaxEnt model and the joint degree sequence (i.e. the rows and columns
 412 sums) in the type II MaxEnt model. These were our initial configurations. Then, we swapped

413 interactions sequentially while maintaining the original connectance or joint degree sequence.
414 Configurations with a higher SVD entropy than the previous one in the chain were always ac-
415 cepted, whereas they were accepted with a probability conditional to a decreasing temperature
416 and the difference in SVD entropy when lower. The final configuration with the highest SVD
417 entropy among the four chains constitute the estimated maximum entropy configuration of a
418 food web given the constraint used.

419 **Structure of MaxEnt food webs**

420 We measured various properties of these configurations of maximum entropy to evaluate how
421 well they reproduce the structure of sampled food webs. Specifically, we evaluated their nest-
422 edness ρ , their maximum trophic level $maxtl$, their network diameter $diam$, their average maxi-
423 mum similarity between species pairs $MxSim$ (Williams and Martinez 2000), their proportion
424 of cannibal species $Cannib$, their proportion of omnivorous species $Omniv$, their SVD entropy,
425 and their motifs profile. Nestedness indicates how much the diet of specialist species is a subset
426 of the one of generalists (Delmas et al. 2019) and was measured using the spectral radius of
427 the adjacency matrix (Staniczenko, Kopp, and Allesina 2013). In turn, the network diameter
428 represents the longest of the shortest paths between all species pair (Albert and Barabasi 2002).
429 Further, cannibal species are species that can eat individuals of their own species (i.e. species
430 having self loops), whereas omnivorous species can prey on different trophic levels (Williams
431 and Martinez 2000). Finally, a motifs profile represents the proportion of three-species motifs
432 (subnetworks), which can be considered as simple building blocks of ecological networks (Milo
433 et al. 2002; Stouffer et al. 2007). All of these properties are relatively easy to measure and were
434 chosen based on their ecological importance and prevalent use in network ecology (McCann
435 2011; Delmas et al. 2019).

436 We compared the performance of both heuristic MaxEnt models in predicting these measures to
437 the one of the null and neutral models (Box 2). We conducted these comparisons using two dif-
438 ferent datasets: (1) our complete dataset including most food webs archived on Mangal, as well
439 as all food webs in the New Zealand and Tuesday Lake datasets, and (2) our *abundance dataset*,
440 a subset of the complete dataset comprising all food webs having data on their species' relative

441 abundances ($N = 19$). Indeed, of the New Zealand and Tuesday Lake datasets, 19 networks
442 had data on species' relative abundances that were used in the neutral model to better assess the
443 performance of our heuristic models. We compared our models' predictions using these two
444 datasets separately to minimize biases and to better represent food webs with abundance data
445 (tbl. 1 and tbl. 2).

446 Overall, we found that the models based on the joint degree sequence (i.e. the type II null and
447 heuristic MaxEnt models) reproduced the structure of empirical food webs much better than the
448 ones based on connectance (i.e. the type I null and heuristic MaxEnt models). This suggests that
449 the predictive power of connectance might be more limited than what was previously suggested
450 (Poisot and Gravel 2014). On the other hand, the neutral model of relative abundances was sur-
451 prisingly good at predicting the maximum trophic level and the network diameter. However,
452 with the exception of the network diameter, the type II heuristic MaxEnt model was better at
453 predicting network structure than the neutral model for most measures considered. This might
454 be because, although neutral processes are important, they act in concert with niche processes in
455 determining species interactions Canard et al. (2014). The joint degree sequence encodes infor-
456 mation on both neutral and niche processes because the number of prey and predators a species
457 has is determined by its relative abundance and biological traits. These results thus show that
458 having information on the number of prey and predators for each species substantially improves
459 the prediction of food-web structure, both compared to models solely based on connectance and
460 to the ones solely based on species relative abundances.

461 Next, the predictions of the type II heuristic MaxEnt model can be compared to its null model
462 counterpart. On average, the type II heuristic MaxEnt model was better at predicting nestedness
463 (0.62 ± 0.08) than its corresponding null model (0.73 ± 0.05 ; empirical networks: 0.63 ± 0.09) for
464 networks in our complete dataset (tbl. 1). This might in part be due to the fact that nestedness was
465 calculated using the spectral radius of the adjacency matrix, which directly leverages information
466 on the network itself just like the heuristic MaxEnt model. The proportion of self-loops (cannibal
467 species) was also better predicted by the type II heuristic MaxEnt model in comparison to the
468 type II null model. However, the type II null model was better at predicting network diameter
469 and average maximum similarity between species pairs, and predictions of the maximum trophic

470 level and the proportion of omnivorous species were similar between both types of models. We
 471 believe that this is because increasing the complexity of a food web might increase its average
 472 and maximum food-chain lengths. In comparison, the null model was more stochastic and does
 473 not necessarily produce more complex food webs with longer food-chain lengths.

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

model	rho	maxtl	diam	MxSim	Cannib	Omniv	entropy
null 1	-0.167	0.980	1.428	-0.502	2.007	1.493	0.056
MaxEnt 1	-0.226	0.831	1.274	-0.524	1.982	1.863	0.106
null 2	0.160	-0.125	0.016	0.007	1.078	0.559	-0.023
MaxEnt 2	-0.015	0.178	0.565	-0.282	0.698	0.589	0.058

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

model	rho	maxtl	diam	MxSim	Cannib	Omniv	entropy
neutral	0.367	-0.090	0.027	0.266	6.870	0.576	-0.083
null 1	-0.134	0.950	1.919	-0.369	2.077	0.614	0.068
MaxEnt 1	-0.229	1.020	1.946	-0.355	2.215	0.801	0.121

model	rho	maxtl	diam	MxSim	Cannib	Omniv	entropy
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

474 Despite this increase in maximum trophic level and network diameter in MaxEnt food webs, we
 475 found that empirical food webs are close to their maximum entropy given a fixed joint degree
 476 sequence (fig. S3). Empirical food webs in the complete dataset had an SVD entropy of $0.89 \pm$
 477 0.04, compared to an SVD entropy of 0.94 ± 0.03 for networks generated using the type II
 478 heuristic MaxEnt model. The relationship between the SVD entropy of empirical food webs
 479 and their maximum entropy is plotted in the last panel of fig. 4. As expected, the SVD entropy
 480 of maximum entropy food webs was higher than that of empirical food webs for almost all
 481 networks, confirming that our method indeed generated more complex networks. Moreover, we
 482 found no to a weak relationship between the increase in SVD entropy and the number of species,
 483 the number of interactions, and connectance (fig. S4). This suggests that the slight increase in
 484 entropy between empirical food webs and their maximum entropy configuration was sufficient
 485 to modify some of their properties, regardless of their number of species and their number of
 486 interactions.

487 A direct comparison of the structure of maximum entropy food webs, constrained by the joint
 488 degree sequence, with empirical data also supports the results depicted in tbl. 1. Indeed, in fig. 4
 489 we show how well empirical measures are predicted by the type II heuristic MaxEnt model. In
 490 accordance with our previous results, we found that nestedness was very well predicted by our
 491 model. However, the model overestimated the maximum trophic level and network diameter,
 492 especially when the sampled food web had intermediate values of these measures. In fig. S5,
 493 we show that the pairwise relationships between the four measures in fig. 4 and species richness
 494 in empirical food webs are similar (in magnitude and sign) to the ones found in food webs
 495 generated using the type II heuristic MaxEnt model. This indicates that the number of species
 496 in the network does not seem to impact the ability of the model to reproduce food-web structure.

497

[Figure 4 about here.]

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

514 [Figure 5 about here.]

515 [Figure 6 about here.]

516 One of the challenges in implementing and validating a maximum entropy model is to discover
517 where its predictions break down. The results depicted in tbl. 1 and fig. 4 show that our type II
518 heuristic MaxEnt model can capture many high-level properties of food webs, but does a poor
519 job of capturing others. This suggests that, although the joint degree sequence is an important
520 driver of food-web structure, other ecological constraints might be needed to account for some
521 emerging food-web properties, especially the ones regarding food-chain lengths. Nevertheless,
522 fig. 5 and fig. 6 show that the model can reproduce surprisingly well the motifs profile, one of the
523 most ecologically informative properties of food webs. This suggests that the emerging structure
524 of food webs is mainly driven by their joint degree sequence, although higher-level properties
525 might be needed to ensure that food-chain lengths fall within realistic values.

526 **Conclusion**

527 The principle of maximum entropy is a mathematical method of finding least-biased probability
528 distributions that have some specified properties given by prior knowledge about a system. We
529 first applied this conventional MaxEnt approach on food webs to predict species level proper-
530 ties, namely the joint degree distribution and the degree distribution of maximum entropy given
531 known numbers of species and interactions. We found that the joint degree distributions of
532 maximum entropy had a similar shape to the ones of empirical food webs in high-connectance
533 systems. However, these MaxEnt distributions were more symmetric than the ones of empirical
534 food webs when connectance was low, which suggests that other constraints might be needed
535 to improve these predictions in low-connectance systems. Then, we used a slightly different
536 approach that aimed at finding heuristically the network configuration with the highest SVD
537 entropy, i.e. whose vector of relative singular values has maximum entropy. This network of
538 maximum entropy is the most complex, or random, given the specified structure. We found
539 that the heuristic maximum entropy model based on connectance did not predict the structure
540 of sampled food webs very well. However, the heuristic maximum entropy model based on the
541 entire joint degree sequence, i.e. on the number of prey and predators for each species, gave
542 more convincing results. Indeed, this model reproduced food-web structure surprisingly well,
543 including the highly informative motifs profile. Nevertheless, it was not able to predict realistic
544 food-chain lengths.

545 Our results bring to the forefront the role of the joint degree distribution in shaping food-web
546 structure. This echoes the work of Fortuna et al. (2010), who found that the degree distribution
547 of ecological networks drive their emerging structure such as their nestedness and modularity.
548 Network ecologists tend to focus on several measures of food webs when studying the ecological
549 consequences of their structure (McCann 2011; Delmas et al. 2019). In fact, following Williams
550 (2011), we believe there is a lot more ecological information in the deviation between these
551 properties in empirical systems and in their maximum entropy configuration given a fixed joint
552 degree sequence.

553 **Alternative MaxEnt models**

554 In this contribution, we used a method based on simulated annealing to find the network config-
555 uration with the highest SVD entropy while fixing some aspects of its structure. However, there
556 are different ways to generate adjacency matrices using MaxEnt. Another technique, also based
557 on simulated annealing, could begin by generating a food web randomly with fixed numbers of
558 species and interactions and calculating its joint degree distribution. Pairs of interactions could
559 then be swapped sequentially until we minimize the divergence between the calculated joint de-
560 gree distribution and the one of maximum entropy obtained analytically. In that case, this is the
561 entropy of the joint degree distribution that would be maximized, not the one of the network's
562 topology. To a certain extent, this method would thus bridge the gap between the analytical and
563 heuristic approaches presented in this article. More research is needed to compare the quality
564 of different methods generating adjacency matrices of food webs using MaxEnt.

565 Maximum entropy graph models are another type of methods that predict a distribution of ad-
566 jacency matrices under soft or hard constraints (e.g., Park and Newman 2004; Cimini et al.
567 2019). Under hard constraints, every network with a non-zero probability exactly satisfies the
568 constraints on its structure. This is in contrast with soft constraints, which require that networks
569 satisfy them on average (i.e. many networks with a non-zero probability do not have the ex-
570 act structure set by the constraints). Maximum entropy graph models are helpful because they
571 can provide probability distributions for many network properties by measuring the structure
572 of all adjacency matrices with a non-zero probability. However, we consider that our approach
573 based on simulated annealing is more flexible and more computationally efficient. Indeed, many
574 measures of food-web structure are hard to translate into mathematical constraints. Moreover,
575 because food webs are directed networks that can have self-loops, it makes the mathematical
576 derivation of maximum entropy graph models difficult. We believe that identifying heuristi-
577 cally what really constrains the topology of food webs is a useful first step before attempting to
578 derive the mathematical formulation of a maximum entropy graph model for food webs.

579 **Applications**

580 Our analytical and heuristic models can be applied for different purposes. First, they could be
581 used to generate first-order approximations of a network's properties when state variables are
582 known empirically. For example, knowing the number of species in an ecological community,
583 we can predict its number of interactions using the flexible links model and then predict its joint
584 degree distribution with minimal biases using the principle of maximum entropy. This could
585 prove particularly useful when predicting network structure at large spatial scales, subdividing
586 the study area into smaller communities (e.g., grid cells). Indeed, because species richness and
587 other ecological data are increasingly abundant (e.g., Dickinson, Zuckerberg, and Bonter 2010),
588 validated MaxEnt models can be used to respond to a wider range of macroecological questions
589 regarding food webs.

590 Second, our analytical model can be used to generate informative priors in Bayesian analyses of
591 the structure of ecological networks (e.g., Cirtwill et al. 2019). Indeed, the probability distri-
592 bution of maximum entropy derived using MaxEnt can be used as a prior that can be constantly
593 updated with novel data. For instance, if we know the number of species and the number of
594 interactions, we can get the degree distribution of maximum entropy, as shown in this contribu-
595 tion. The degree distribution represents the probability that a species can interact (as a predator
596 or a prey) with a number of other species. Data on species interactions can be used to update
597 the prior degree distribution to generate a more accurate posterior distribution, thus improving
598 our description and understanding of the system.

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

607 Finally, our analytical and heuristic models can be used as alternative null models of ecological
608 networks to better understand and identify the ecological processes driving food-web structure.
609 Indeed, these mechanisms can be better described when analysing the deviation of empirical
610 data from MaxEnt predictions. A strong deviation would indicate that ecological mechanisms
611 not encoded in the statistical constraints are at play for the system at hand. If deviations are sys-
612 tematic, the maximum entropy model might need to be revised to include appropriate ecological
613 constraints. This revision process helps us reflect on and identify what really constrains food-
614 web structure. However, it is important to note here that tangible ecological mechanisms cannot
615 be directly inferred from statistical distributions (Warren II, Costa, and Bradford 2022). Instead,
616 by identifying the constraints of a system and by analysing empirical deviations from maximum
617 entropy predictions, MaxEnt can only help us redirect research efforts towards understanding
618 the biological mechanisms behind these constraints.

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

624 **Final remarks**

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

635 **Acknowledgments**

636 We acknowledge that this study was conducted on land within the traditional unceded territory
637 of the Saint Lawrence Iroquoian, Anishinabewaki, Mohawk, Huron-Wendat, and Omàmiwinini-
638 wak nations. This work was supported by the Institute for Data Valorisation (IVADO) and the
639 Natural Sciences and Engineering Research Council of Canada (NSERC) Collaborative Re-
640 search and Training Experience (CREATE) program, through the Computational Biodiversity
641 Science and Services (BIOS²) program.

642 **References**

- 643 Albert, R., and A. L. Barabasi. 2002. “Statistical Mechanics of Complex Networks.” *Reviews*
644 *of Modern Physics* 74 (1): 47–97. <https://doi.org/10.1103/RevModPhys.74.47>.
- 645 Banville, Francis, Steve Vissault, and Timothée Poisot. 2021. “Mangal.jl and EcologicalNet-
646 works.jl: Two Complementary Packages for Analyzing Ecological Networks in Julia.” *Journal*
647 *of Open Source Software* 6 (61): 2721. <https://doi.org/10.21105/joss.02721>.
- 648 Bartomeus, Ignasi, Dominique Gravel, Jason M. Tylianakis, Marcelo A. Aizen, Ian A. Dickie,
649 and Maud Bernard-Verdier. 2016. “A Common Framework for Identifying Linkage Rules
650 Across Different Types of Interactions.” *Functional Ecology* 30 (12): 1894–1903. <https://doi.org/10.1111/1365-2435.12666>.
- 652 Bascompte, J., P. Jordano, C. J. Melian, and J. M. Olesen. 2003. “The Nested Assembly of
653 Plant-Animal Mutualistic Networks.” *Proceedings of the National Academy of Sciences*
654 *of the United States of America* 100 (16): 9383–87. <https://doi.org/10.1073/pnas.1633576100>.
- 656 Beck, Christian. 2009. “Generalised Information and Entropy Measures in Physics.” *Contem-
657 porary Physics* 50 (4): 495–510. <https://doi.org/10.1080/00107510902823517>.
- 658 Canard, E. F., N. Mouquet, D. Mouillot, M. Stanko, D. Miklisova, and D. Gravel. 2014. “Empir-
659 ical Evaluation of Neutral Interactions in Host-Parasite Networks.” *The American Naturalist*
660 183 (4): 468–79. <https://doi.org/10.1086/675363>.

- 661 Cimini, Giulio, Tiziano Squartini, Fabio Saracco, Diego Garlaschelli, Andrea Gabrielli, and
662 Guido Caldarelli. 2019. “The Statistical Physics of Real-World Networks.” *Nature Reviews*
663 *Physics* 1 (1): 58–71. <https://doi.org/10.1038/s42254-018-0002-6>.
- 664 Cirtwill, Alyssa R., Anna Eklof, Tomas Roslin, Kate Wootton, and Dominique Gravel. 2019.
665 “A Quantitative Framework for Investigating the Reliability of Empirical Network Construc-
666 tion.” *Methods in Ecology and Evolution* 10 (6): 902–11. <https://doi.org/10.1111/2041-210X.13180>.
- 668 Cohen, Joel E., Tomas Jonsson, and Stephen R. Carpenter. 2003. “Ecological Community
669 Description Using the Food Web, Species Abundance, and Body Size.” *Proceedings of*
670 *the National Academy of Sciences* 100 (4): 1781–86. <https://doi.org/10.1073/pnas.232715699>.
- 672 Delmas, Eva, Mathilde Besson, Marie Hélène Brice, Laura A. Burkle, Giulio V. Dalla Riva,
673 Marie Josée Fortin, Dominique Gravel, et al. 2019. “Analysing Ecological Networks of
674 Species Interactions.” *Biological Reviews*. <https://doi.org/10.1111/brv.12433>.
- 675 Dickinson, Janis L., Benjamin Zuckerberg, and David N. Bonter. 2010. “Citizen Science as an
676 Ecological Research Tool: Challenges and Benefits.” In *Annual Review of Ecology, Evo-*
677 *lution, and Systematics, Vol 41*, edited by D. J. Futuyma, H. B. Shafer, and D. Simberloff,
678 41:149–72. Palo Alto: Annual Reviews. <https://doi.org/10.1146/annurev-ecolsys-102209-144636>.
- 679 Dunne, Jennifer A., Richard J. Williams, and Neo D. Martinez. 2002. “Network Structure and
680 Biodiversity Loss in Food Webs: Robustness Increases with Connectance.” *Ecology Letters*
681 5 (4): 558–67. <https://doi.org/10.1046/j.1461-0248.2002.00354.x>.
- 682 Dunning, Iain, Joey Huchette, and Miles Lubin. 2017. “JuMP: A Modeling Language for
683 Mathematical Optimization.” *SIAM Review* 59 (2): 295–320. <https://doi.org/10.1137/15M1020575>.
- 685 Fortuna, M. A., and J. Bascompte. 2006. “Habitat Loss and the Structure of Plant-Animal
686 Mutualistic Networks.” *Ecology Letters* 9 (3): 278–83. <https://doi.org/10.1111/j.1461-0248.2005.00868.x>.

- 688 Fortuna, M. A., Daniel B. Stouffer, Jens M. Olesen, Pedro Jordano, David Mouillot, Boris R.
689 Krasnov, Robert Poulin, and Jordi Bascompte. 2010. “Nestedness Versus Modularity in
690 Ecological Networks: Two Sides of the Same Coin?” *Journal of Animal Ecology* 79 (4):
691 811–17. <https://doi.org/10.1111/j.1365-2656.2010.01688.x>.
- 692 Frank, S. A., and E. Smith. 2011. “A Simple Derivation and Classification of Common Prob-
693 ability Distributions Based on Information Symmetry and Measurement Scale.” *Journal of*
694 *Evolutionary Biology* 24 (3): 469–84. <https://doi.org/10.1111/j.1420-9101.2010.02204.x>.
- 696 Gómez, José M., Francisco Perfectti, and Pedro Jordano. 2011. “The Functional Consequences
697 of Mutualistic Network Architecture.” *PLOS ONE* 6 (1): e16143. <https://doi.org/10.1371/journal.pone.0016143>.
- 699 Harremoës, Peter, and Flemming Topsøe. 2001. “Maximum Entropy Fundamentals.” *Entropy*
700 3 (3): 191–226. <https://doi.org/10.3390/e3030191>.
- 701 Harte, J., and Erica A. Newman. 2014. “Maximum Information Entropy: A Foundation for
702 Ecological Theory.” *Trends in Ecology & Evolution* 29 (7): 384–89. <https://doi.org/10.1016/j.tree.2014.04.009>.
- 704 Harte, J., T. Zillio, E. Conlisk, and A. B. Smith. 2008. “Maximum Entropy and the State-
705 Variable Approach to Macroecology.” *Ecology* 89 (10): 2700–2711. <https://doi.org/10.1890/07-1369.1>.
- 707 Hoorn, Pim van der, Gabor Lippner, and Dmitri Krioukov. 2018. “Sparse Maximum-Entropy
708 Random Graphs with a Given Power-Law Degree Distribution.” *Journal of Statistical Physics*
709 173 (3): 806–44. <https://doi.org/10.1007/s10955-017-1887-7>.
- 710 Hortal, Joaquín, Francesco de Bello, José Alexandre F. Diniz-Filho, Thomas M. Lewinsohn,
711 Jorge M. Lobo, and Richard J. Ladle. 2015. “Seven Shortfalls That Beset Large-Scale
712 Knowledge of Biodiversity.” *Annual Review of Ecology, Evolution, and Systematics* 46 (1):
713 523–49. <https://doi.org/10.1146/annurev-ecolsys-112414-054400>.
- 714 Ings, Thomas C., José M. Montoya, Jordi Bascompte, Nico Blüthgen, Lee Brown, Carsten F.
715 Dormann, François Edwards, et al. 2009. “Review: Ecological Networks Beyond Food

- 716 Webs.” *Journal of Animal Ecology* 78 (1): 253–69. <https://doi.org/10.1111/j.1365-2656.2008.01460.x>.
- 718 Jaynes, E. T. 1957a. “Information Theory and Statistical Mechanics.” *Physical Review* 106 (4):
719 620–30. <https://doi.org/10.1103/PhysRev.106.620>.
- 720 ———. 1957b. “Information Theory and Statistical Mechanics. II.” *Physical Review* 108 (2):
721 171–90. <https://doi.org/10.1103/PhysRev.108.171>.
- 722 Jordano, Pedro. 2016. “Sampling Networks of Ecological Interactions.” *Functional Ecology* 30
723 (12): 1883–93. <https://doi.org/10.1111/1365-2435.12763>.
- 724 Khinchin, A. Ya. 2013. *Mathematical Foundations of Information Theory*. Courier Corpora-
725 tion.
- 726 MacDonald, Arthur Andrew Meahan, Francis Banville, and Timothée Poisot. 2020. “Revisiting
727 the Links-Species Scaling Relationship in Food Webs.” *Patterns* 0 (0). <https://doi.org/10.1016/j.patter.2020.100079>.
- 729 Martyushev, L. M., and V. D. Seleznev. 2006. “Maximum Entropy Production Principle in
730 Physics, Chemistry and Biology.” *Physics Reports-Review Section of Physics Letters* 426
731 (1): 1–45. <https://doi.org/10.1016/j.physrep.2005.12.001>.
- 732 McCann, Kevin S. 2011. *Food Webs (MPB-50)*. *Food Webs (MPB-50)*. Princeton University
733 Press. <https://doi.org/10.1515/9781400840687>.
- 734 Milo, R., S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon. 2002. “Net-
735 work Motifs: Simple Building Blocks of Complex Networks.” *Science* 298 (5594): 824–27.
736 <https://doi.org/10.1126/science.298.5594.824>.
- 737 Park, Juyong, and M. E. J. Newman. 2004. “Statistical Mechanics of Networks.” *Physical
738 Review E* 70 (6): 066117. <https://doi.org/10.1103/PhysRevE.70.066117>.
- 739 Pascual, Department of Ecology and Evolutionary Biology Mercedes, and Visiting Professor
740 Jennifer A. Dunne. 2006. *Ecological Networks: Linking Structure to Dynamics in Food
741 Webs*. Oxford University Press, USA.

- 742 Phillips, Steven J., Robert P. Anderson, and Robert E. Schapire. 2006. "Maximum Entropy
743 Modeling of Species Geographic Distributions." *Ecological Modelling* 190 (3): 231–59.
744 <https://doi.org/10.1016/j.ecolmodel.2005.03.026>.
- 745 Pielou, Evelyn C. 1975. "Ecological Diversity." In. 574.524018 P5.
- 746 Poisot, Timothée, Benjamin Baiser, Jennifer A. Dunne, Sonia Kefi, Francois Massol, Nico-
747 las Mouquet, Tamara N. Romanuk, Daniel B. Stouffer, Spencer A. Wood, and Dominique
748 Gravel. 2016. "Mangal - Making Ecological Network Analysis Simple." *Ecography* 39 (4):
749 384–90. <https://doi.org/10.1111/ecog.00976>.
- 750 Poisot, Timothée, Gabriel Bergeron, Kevin Cazelles, Tad Dallas, Dominique Gravel, Andrew
751 MacDonald, Benjamin Mercier, Clément Violet, and Steve Vissault. 2021. "Global Knowl-
752 edge Gaps in Species Interaction Networks Data." *Journal of Biogeography* 48 (7): 1552–
753 63. <https://doi.org/10.1111/jbi.14127>.
- 754 Poisot, Timothée, and Dominique Gravel. 2014. "When Is an Ecological Network Complex?
755 Connectance Drives Degree Distribution and Emerging Network Properties." *PeerJ* 2: e251.
756 <https://doi.org/10.7717/peerj.251>.
- 757 Poisot, Timothée, Daniel B. Stouffer, and Dominique Gravel. 2015. "Beyond Species: Why
758 Ecological Interaction Networks Vary Through Space and Time." *Oikos* 124 (3): 243–51.
759 <https://doi.org/10.1111/oik.01719>.
- 760 Pomeranz, Justin P. F., Ross M. Thompson, Timothée Poisot, and Jon S. Harding. 2019. "In-
761 ferring Predatorprey Interactions in Food Webs." *Methods in Ecology and Evolution* 10 (3):
762 356–67. <https://doi.org/10.1111/2041-210X.13125>.
- 763 Pomeranz, Justin Page, Ross M. Thompson, Timothée Poisot, Jon S. Harding, and Justin P. F.
764 Pomeranz. 2018. "Data from: Inferring Predator-Prey Interactions in Food Webs." Dryad.
765 <https://doi.org/10.5061/DRYAD.K59M37F>.
- 766 Shannon, C. E. 1948. "A Mathematical Theory of Communication." *The Bell System Technical
767 Journal* 27 (3): 379–423. <https://doi.org/10.1002/j.1538-7305.1948.tb01338.x>.
- 768 Staniczenko, Phillip P. A., Jason C. Kopp, and Stefano Allesina. 2013. "The Ghost of Nested-
769 ness in Ecological Networks." *Nature Communications* 4 (1): 1391. <https://doi.org/>

- 770 [10.1038/ncomms2422](https://doi.org/10.1038/ncomms2422).
- 771 Stock, Michiel, Timothée Poisot, and Bernard De Baets. 2021. “Optimal Transportation Theory
772 for Species Interaction Networks.” *Ecology and Evolution* 11 (9): 3841–55. <https://doi.org/10.1002/ece3.7254>.
- 773
- 774 Stouffer, Daniel B., and Jordi Bascompte. 2011. “Compartmentalization Increases Food-Web
775 Persistence.” *Proceedings of the National Academy of Sciences* 108 (9): 3648–52. <https://doi.org/10.1073/pnas.1014353108>.
- 776
- 777 Stouffer, Daniel B., Juan Camacho, Wenxin Jiang, and Luis A. Nunes Amaral. 2007. “Evidence
778 for the Existence of a Robust Pattern of Prey Selection in Food Webs.” *Proceedings of the Royal Society B-Biological Sciences* 274 (1621): 1931–40. <https://doi.org/10.1098/rspb.2007.0571>.
- 779
- 780
- 781 Strydom, Tanya, Michael D. Catchesen, Francis Banville, Dominique Caron, Gabriel Dansereau,
782 Philippe Desjardins-Proulx, Norma R. Forero-Muñoz, et al. 2021. “A Roadmap Towards
783 Predicting Species Interaction Networks (across Space and Time).” *Philosophical Transactions of the Royal Society B: Biological Sciences* 376 (1837): 20210063. <https://doi.org/10.1098/rstb.2021.0063>.
- 784
- 785
- 786 Strydom, Tanya, Giulio V. Dalla Riva, and Timothee Poisot. 2021. “SVD Entropy Reveals the
787 High Complexity of Ecological Networks.” *Frontiers in Ecology and Evolution* 9: 623141.
788 <https://doi.org/10.3389/fevo.2021.623141>.
- 789
- 790
- 791
- 792 Vázquez, Diego P. 2005. “Degree Distribution in Plantanimal Mutualistic Networks: Forbidden
793 Links or Random Interactions?” *Oikos* 108 (2): 421–26. <https://doi.org/10.1111/j.0030-1299.2005.13619.x>.
- 794
- 795 Warren II, Robert J., James T. Costa, and Mark A. Bradford. 2022. “Seeing Shapes in Clouds:
796 The Fallacy of Deriving Ecological Hypotheses from Statistical Distributions.” *Oikos* 2022
797 (11): e09315. <https://doi.org/10.1111/oik.09315>.
- 798
- 799
- 800 Williams, Richard J. 2011. “Biology, Methodology or Chance? The Degree Distributions of
801 Bipartite Ecological Networks.” *PLOS ONE* 6 (3): e17645. <https://doi.org/10.1371/journal.pone.0017645>.

⁷⁹⁸ Williams, Richard J., and Neo D. Martinez. 2000. "Simple Rules Yield Complex Food Webs."

⁷⁹⁹ *Nature* 404 (6774): 180–83. <https://doi.org/10.1038/35004572>.

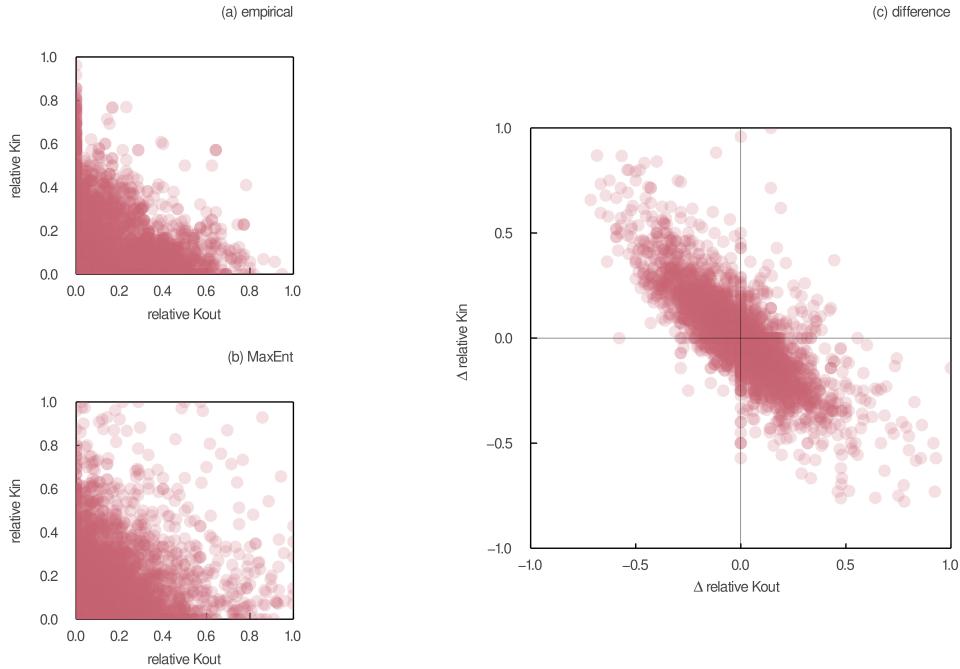


Figure 1: Relative number of predators (k_{in}) as a function of their relative number of prey (k_{out}) for each species in (a) empirical and (b) joint degree sequences obtained from the analytical MaxEnt model. Empirical networks include most food webs archived on Mangal, as well as the New Zealand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

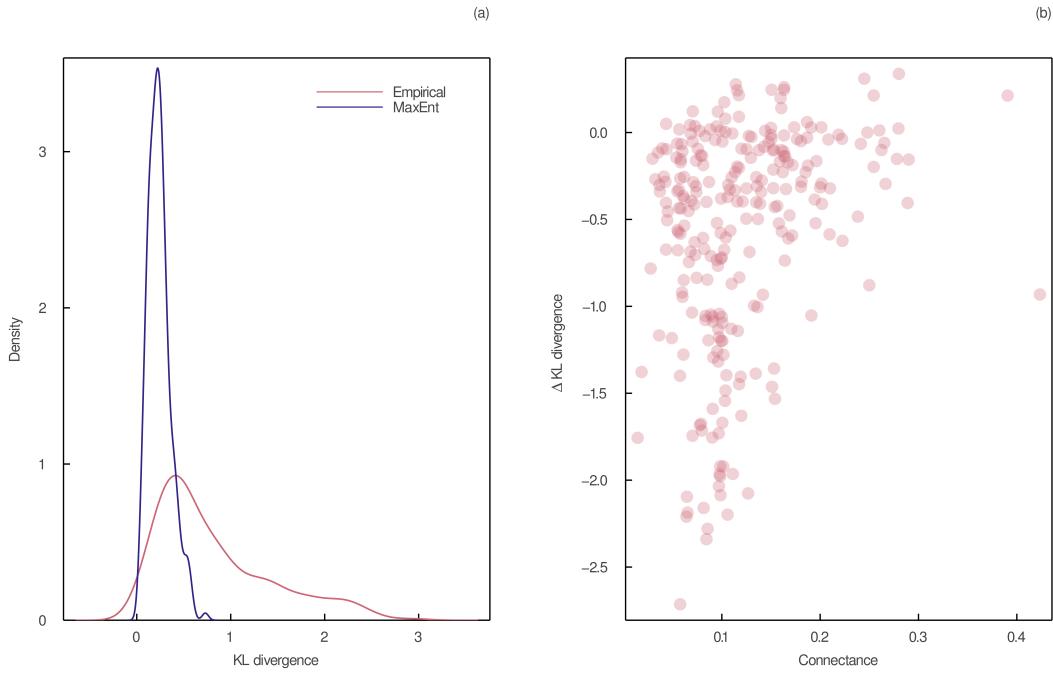


Figure 2: (a) Probability density of KL divergence between in and out-degree sequences of empirical and predicted joint degree sequences. (b) Difference between the KL divergence of empirical and predicted joint degree sequences as a function of connectance. In both panels, empirical networks include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

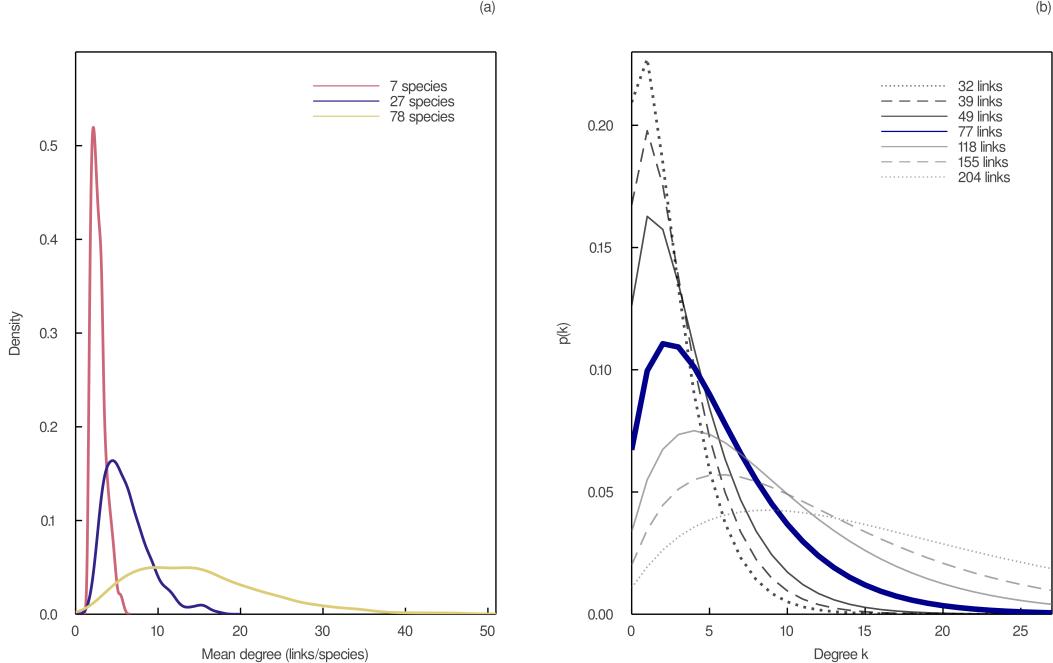


Figure 3: (a) Probability density of the mean degree of a food web obtained using different values of species richness S . The number of interactions L was simulated 1000 times using the flexible links model fitted to all empirical networks in our complete dataset (eq. 17). The mean degrees $2L/S$ were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

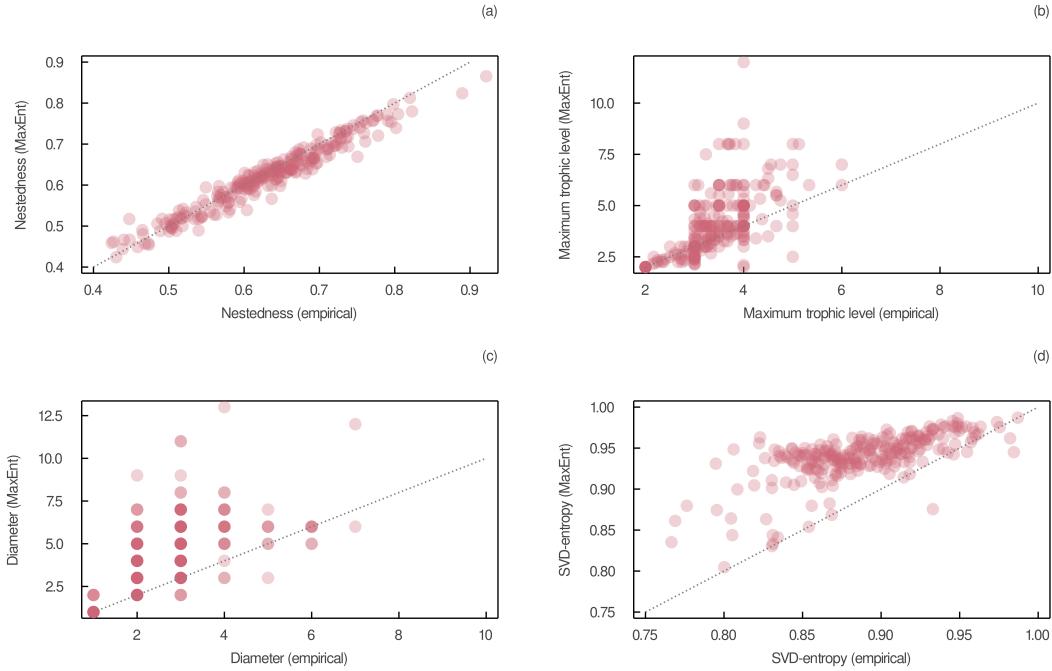


Figure 4: Relationship between the structure of empirical and maximum entropy food webs. Empirical networks include most food webs archived on Mangal, as well as the New Zealand and Tuesday Lake food webs (our complete dataset). Maximum entropy networks were obtained using the type II heuristic MaxEnt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

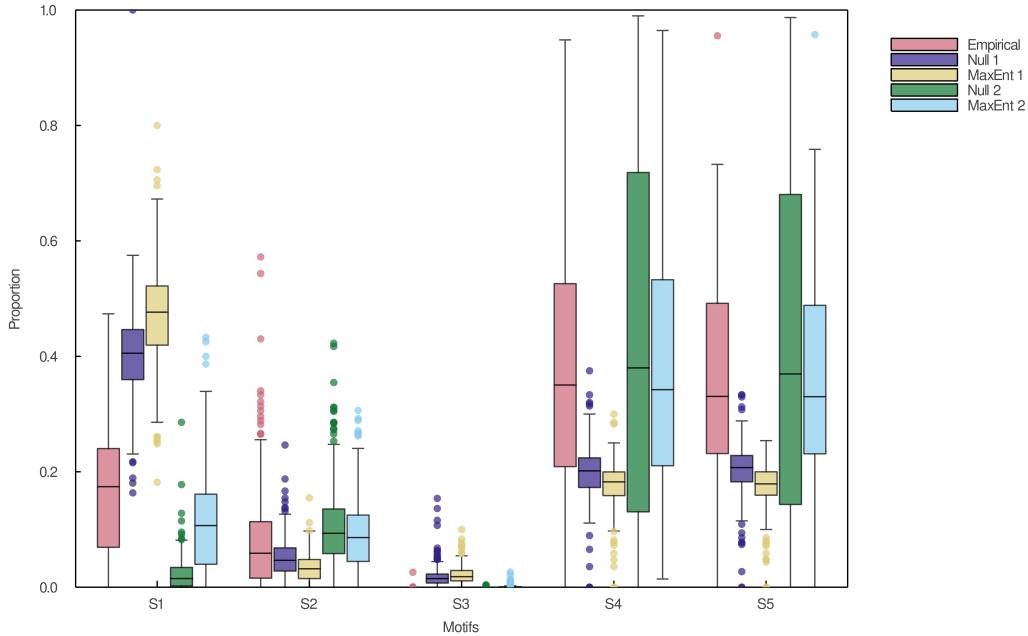


Figure 5: Motifs profile of empirical and predicted food webs. Empirical networks include most food webs archived on Mangal, as well as the New Zealand and Tuesday Lake datasets (our complete dataset). Null 1: Type I null model based on connectance. MaxEnt 1: Type I heuristic MaxEnt model based on connectance.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

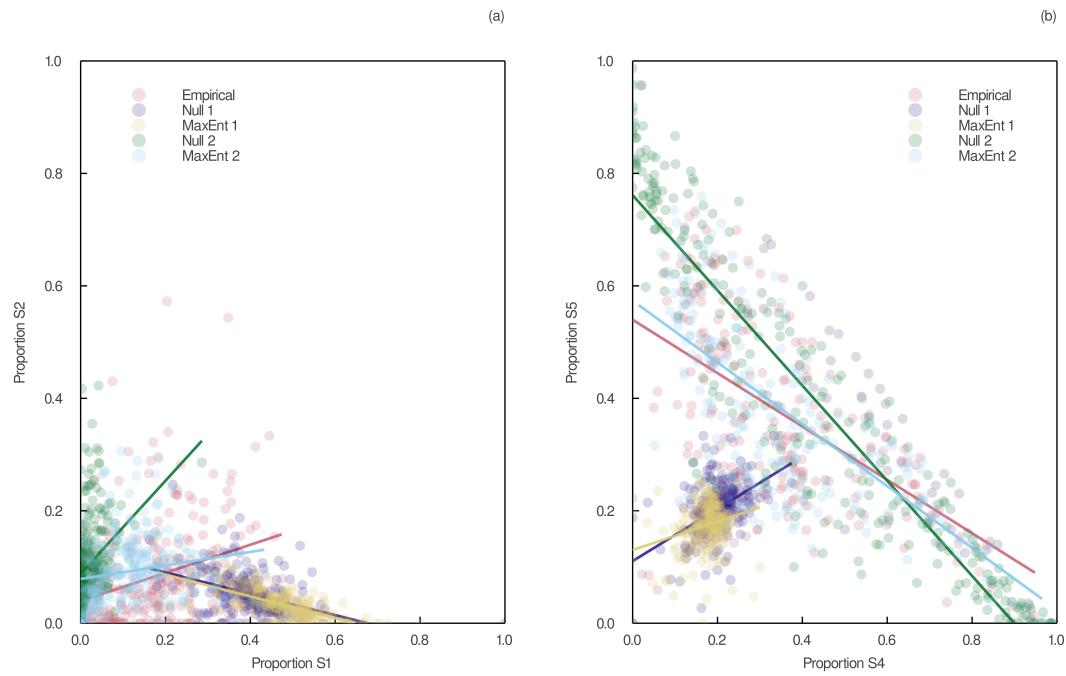


Figure 6: Pairwise relationships between motifs proportions of empirical and predicted food webs. Empirical networks include most food webs archived on Mangal, as well as the New Zealand and Tuesday Lake datasets (our complete dataset). Null 1: Type I null model based on connectance. MaxEnt 1: Type I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