

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

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

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

## **Correspondance to:**

Francis Banville — francis.banville@umontreal.ca

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

## **1 Introduction**

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

16 The principle of maximum entropy (MaxEnt) is a statistical and topological model that can  
17 be used for both of these purposes, i.e. to make predictions of network structure and to better  
18 understand processes shaping ecological networks. Food-web properties that can be derived  
19 using MaxEnt are varied and pertain to different elements of the network (i.e. at the species  
20 (node), the interaction (edge) or the community (network) levels). For example, at the species  
21 level, MaxEnt can be used to predict the distribution of trophic levels among species, as well as  
22 the distribution of species' vulnerability (number of predators) and generality (number of preys).  
23 By contrast, at the interaction level, predictions can be made on the distribution of interaction  
24 strengths in weighted food webs. At the community level, it can give us a probability distribution  
25 of potential networks or of specific measures of their emerging structure. The applicability and  
26 performance of this method depend on the ecological information available and on our capacity  
27 to find the set of state variables that best represent natural systems and to translate them into  
28 appropriate statistical constraints. MaxEnt has been used in a wide range of disciplines, from  
29 thermodynamics to chemistry and biology (Martyushev and Seleznev 2006). It has also been

30 proven useful in ecology, e.g. in species distribution models (Phillips, Anderson, and Schapire  
31 2006) and macroecological models (Harte et al. 2008; Harte and Newman 2014). Maximizing a  
32 measure of entropy ensures that the derived probability distributions are unique and least biased  
33 under the set of constraints used. These constraints are built using state variables, i.e. variables  
34 that represent the macrostate of the system. Having a validated maximum entropy model for the  
35 system at hand allows us to make rigorous predictions using a minimal amount of data, as well  
36 as helping us describe the most important factors driving that system.

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

49 In this contribution, we used two complementary approaches to predict the structure of food  
50 webs using the principle of maximum entropy. We then compared our predictions against em-  
51 pirical data and null and neutral models commonly used in network ecology. The first approach  
52 consists in deriving constrained probability distributions of given network properties directly.  
53 We derived the joint degree distribution (a probability distribution) of maximum entropy using  
54 only the number of species  $S$  and the number of interactions  $L$  as state variables. Then, we  
55 predicted the degree distribution of maximum entropy directly from the joint degree distribu-  
56 tion since the first is the sum of the marginal distributions of the second (a species' degree is  
57 the sum of its in and out-degrees). Because of the scarcity of empirical data on the number of  
58 interactions in ecological networks, we present a method to predict  $L$  from  $S$  (Box 1), thus al-

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

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

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

84 Entropy measures the amount of information given by the outcome of a random variable. Many  
85 measures of entropy have been developed in physics (Beck 2009), but only a fraction of them  
86 could be used as an optimization measure with the principle of maximum entropy. According

87 to Beck (2009) and Khinchin (2013), a measure of entropy  $H$  should satisfy four properties  
88 in the discrete case: (1) it should be a function of a probability distribution  $p(n)$  only; (2) it  
89 should be maximized when  $p(n)$  is uniform; (3) it should not be influenced by outcomes with a  
90 null probability; and (4) it should be independent of the order of information acquisition. The  
91 Shannon's entropy (Shannon 1948)

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

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

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

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

## 103 Data and code

104 We tested our MaxEnt models (both approaches) against open food-web data queried from three  
105 different sources and integrated into what we call our *complete dataset*. First, all food webs  
106 archived on [mangal.io](https://mangal.io) (Poisot et al. 2016; Banville, Vissault, and Poisot 2021) were directly  
107 queried from the database ( $n = 235$ ). Most ecological networks archived on Mangal are multi-  
108 layer networks, i.e. networks that describe different types of interactions. We kept all networks

whose interactions were mainly of predation and herbivory types, and removed the largest network ( $S = 714$ ) for computational efficiency reasons. Then, to this set we added food webs from two different sources: the New-Zealand dataset ( $n = 21$ ; Pomeranz et al. 2018) and the Tuesday lake dataset ( $n = 2$ ; Cohen, Jonsson, and Carpenter 2003). Of these two datasets, 19 networks had data on species' relative abundances that were used in the neutral model presented in a following subsection. These networks are part of what we call our *abundance dataset*, which is a subset of our complete dataset.

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

## Analytical models: Measures of maximum entropy

### Joint degree distribution

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

We want to maximize Shannon's entropy

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

subject to the following constraints:

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

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

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

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

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

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

137 where  $\lambda_1$ ,  $\lambda_2$ , and  $\lambda_3$  are the Lagrange multipliers. The probability distribution that maximizes  
 138 entropy is obtained by finding these values. As pointed out in the introduction,  $F$  is just Shan-  
 139 non's entropy to which we added terms that each sums to zero (our constraints).  $F$  is maximized  
 140 by setting to 0 its partial derivative with respect to  $p(k_{in}, k_{out})$ . Because the derivative of a con-  
 141 stant is zero, this gives us:

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

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

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

<sup>143</sup> Then, solving eq. 9 for  $p(k_{in}, k_{out})$ , we obtain:

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

<sup>144</sup> where  $Z = e^{1+\lambda_1}$  is called the partition function. The partition function ensures that probabilities  
<sup>145</sup> sum to 1 (our normalization constraint). It can be expressed in terms of  $\lambda_2$  and  $\lambda_3$  as follows:

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

<sup>146</sup> After substituting  $p(k_{in}, k_{out})$  in eq. 5 and eq. 6, we get a nonlinear system of two equations and  
<sup>147</sup> two unknowns:

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

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

<sup>148</sup> We solved eq. 12 and eq. 13 numerically using the Julia library JuMP.jl v0.21.8 (Dunning,  
<sup>149</sup> Huchette, and Lubin 2017). JuMP.jl supports nonlinear optimization problems by providing  
<sup>150</sup> exact second derivatives that increase the accuracy and performance of its solvers. The estimated  
<sup>151</sup> values of  $\lambda_2$  and  $\lambda_3$  can be substituted in eq. 10 to have a more workable expression for the joint  
<sup>152</sup> degree distribution.

<sup>153</sup> We predicted the joint degree distribution of maximum entropy for each food web in our com-  
<sup>154</sup> plete dataset, i.e. using their numbers of species and numbers of interactions as state variables.

<sup>155</sup> We then sampled one realization of the degree sequence for each network using the probabilities  
<sup>156</sup> given by the joint degree distribution. In fig. 1 (left panels), we show the relationship between  
<sup>157</sup>  $k_{out}$  and  $k_{in}$  standardized by the number of species in their networks, for empirical and maximum  
<sup>158</sup> entropy joint degree distributions. We see that our model predicts higher values of generality  
<sup>159</sup> and vulnerability compared to empirical food webs (i.e. relative values of  $k_{out}$  and  $k_{in}$  closer to  
<sup>160</sup> 1). However, plotting the difference between predicted and empirical values for each species

<sup>161</sup> gives a different perspective. The right panel of fig. 1 presents these differences when species  
<sup>162</sup> are ordered by their total degree in their network (i.e. by the sum of their in and out-degrees).  
<sup>163</sup> Indeed, our predicted joint degree sequences have the same number of species as their empiri-  
<sup>164</sup> cal counterparts, but they are species agnostic; in other words, instead of predicting a value for  
<sup>165</sup> each species directly, we predicted the entire joint degree sequence without taking into account  
<sup>166</sup> species' identity. When we associate predictions and empirical data according to their rank in  
<sup>167</sup> total degrees, we see that species predicted to be have a higher generality (number of preys)  
<sup>168</sup> generally have a lower vulnerability (number of preys) than what is observed (and conversely).  
<sup>169</sup> In fig. S1, we show how these differences change when species are instead ordered by their  
<sup>170</sup> out-degrees (left panel) and in-degrees (right panel), respectively.

<sup>171</sup> [Figure 1 about here.]

<sup>172</sup> We plotted the Kullback–Leibler divergence between in and out-degree distributions to compare  
<sup>173</sup> the symmetry of empirical and maximum entropy joint degree sequences fig. 2. As we expected,  
<sup>174</sup> our model predicted more similar in-degree and out-degree distributions than empirical data.  
<sup>175</sup> However, this difference decreased with connectance right panel of 2. Overall, this suggests that  
<sup>176</sup> other ecological constraints might be needed to account for the asymmetry of the joint degree  
<sup>177</sup> distribution, especially for networks with a lower connectance.

<sup>178</sup> [Figure 2 about here.]

## <sup>179</sup> Degree distribution

<sup>180</sup> The degree distribution  $p(k)$  represents the probability that a species has  $k$  interactions in a food  
<sup>181</sup> web, with  $k = k_{in} + k_{out}$ . It can thus be directly obtained from the joint degree distribution:

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

<sup>182</sup> In fig. S2, we show that the degree distribution of maximum entropy, given  $S$  and  $L$ , predicts  
<sup>183</sup> very low probabilities that a species will be isolated in its food web (i.e. having  $k = 0$ ). As

<sup>184</sup> MacDonald, Banville, and Poisot (2020) pointed out, the size of food webs should at least be  
<sup>185</sup> of  $S - 1$  interactions, since a lower number would yield isolated species, i.e. species without  
<sup>186</sup> any predators or preys. Our results show that, under our purely information-theoretic model,  
<sup>187</sup> the probability that a species is isolated is quite high when the total number of interactions is  
<sup>188</sup> below this threshold. The expected proportion of isolated species rapidly declines by orders of  
<sup>189</sup> magnitude with increasing numbers of species and interactions.

<sup>190</sup> The degree distribution could also have been obtained directly using the principle of maximum  
<sup>191</sup> entropy, as discussed in Williams (2011). This gives the following distribution:

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

<sup>192</sup> with  $Z = \sum_{k=0}^S e^{-\lambda_2 k}$ .

<sup>193</sup> This can be solved numerically using the constraint of the average degree  $\langle k \rangle = \frac{2L}{S}$  of a species.  
<sup>194</sup> Note that the mean degree is twice the value of the linkage density, because every link must be  
<sup>195</sup> counted twice when we add in and out-degrees together.

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

<sup>196</sup> The numerical solution is identical to the one we obtained using the joint degree distribution  
<sup>197</sup> as an intermediate. Ecologists wanting to model a system without considering isolated species  
<sup>198</sup> could simply change the lower limit of  $k$  to 1 and solve the resulting equation numerically.

<sup>199</sup> In this section, we showed how important measures of food-web structure, namely the degree  
<sup>200</sup> distribution and the joint degree distribution, could be derived with the principle of maximum  
<sup>201</sup> entropy using minimal knowledge on a biological community. This type of models, although  
<sup>202</sup> useful to make least-biased predictions on many network properties, can be hard to apply for  
<sup>203</sup> other measures. Indeed, there are dozens of measures of network structure (Delmas et al. 2019)  
<sup>204</sup> and many are not calculated with mathematical equations, but with algorithms. Moreover, the  
<sup>205</sup> applicability of this method to empirical systems is limited by the state variables we can actu-  
<sup>206</sup> ally measure and use. In the next section, we propose a more flexible method to predict many

207 measures of network structure simultaneously, i.e. by finding heuristically the configuration of  
 208 the network having maximum entropy given partial knowledge of its emerging structure.

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

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

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

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

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

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

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

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

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

240 [Figure 3 about here.]

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

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

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

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

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

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

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

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

## 258 Neutral model

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

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

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

## 265 Heuristical models: Networks of maximum entropy

### 266 From Shannon's to SVD entropy

267 The principle of maximum entropy can be applied on the network itself if we decompose its  
268 adjacency matrix into a non-zero vector of relative values. This is a necessary step when work-  
269 ing with food webs, which are frequently expressed as a matrix  $A = [a_{ij}]$  of Boolean values

270 representing the presence ( $a_{ij} = 1$ ) or absence ( $a_{ij} = 0$ ) of an interaction between two species  $i$   
 271 and  $j$ . Knowing one or many properties of a food web of interest (e.g., its number of species and  
 272 number of interactions), we can simulate its adjacency matrix randomly by using these known  
 273 ecological information to constrain the space of potential networks. The entropy of this hypo-  
 274 thetical matrix can then be measured after decomposing it into appropriate values. Simulating  
 275 a series of networks until we find the one having the highest entropy allows us to search for the  
 276 most complex food-web configuration given the ecological constraints used. This configuration  
 277 is the least biased one considering the information we had. In other words, the most we can  
 278 say about a network's adjacency matrix, without making more assumptions than the ones given  
 279 by our incomplete knowledge of its structure, is the one of maximum entropy. Generating the  
 280 most complex network that corresponds to this structure allows us to explore more easily other  
 281 properties of food webs under MaxEnt.

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

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

289 where  $s_i$  are the relative singular values of the adjacency matrix ( $s_i = \sigma_i / \sum_{i=1}^R \sigma_i$ , where  $\sigma_i$   
 290 are the singular values). Note that the distribution of relative singular values is analogous to  
 291 a probability distribution, with  $0 < s_i < 1$  and  $\sum s_i = 1$ . Thus, this measure also satisfies  
 292 all four properties of an appropriate entropy measure (see introduction), while being a proper  
 293 measure of the internal complexity of food webs (Strydom, Dalla Riva, and Poisot 2021). Fol-  
 294 lowing Strydom, Dalla Riva, and Poisot (2021), we standardized this measure with the rank  $R$  of  
 295 the matrix (i.e.  $J / \ln(R)$ ) to account for the difference in dimensions between networks (*sensu*

296 Pielou's evenness; Pielou 1975). We used SVD entropy to predict the network configuration  
297 of maximum entropy (i.e. of maximum complexity) heuristically given a predetermined set of  
298 constraints.

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

300 We built two types of MaxEnt network models: one based on connectance (type I MaxEnt net-  
301 work model) and the other based on the joint degree sequence (type II MaxEnt network model).  
302 They are based on the same constraints as the types I and II null models presented above. For  
303 both models, we used a simulated annealing algorithm with 4 chains, 2000 steps and an initial  
304 temperature of 0.2. For each chain, we first generated one random Boolean matrix with the same  
305 order (number of species) as empirical webs, while maintaining the total number of interactions  
306 (type I MaxEnt network model) or rows and columns sums (type II MaxEnt network model).  
307 These are our initial configurations. Then, we swapped interactions sequentially while main-  
308 taining the original connectance or the joint degree sequence for types I and II MaxEnt network  
309 model, respectively. Configurations with a higher SVD entropy than the previous one in the  
310 chain were always accepted, whereas they were accepted with a probability conditional to a de-  
311 creasing temperature when lower. The final configuration with the highest SVD entropy among  
312 the four chains constitute our estimated MaxEnt network. Even though we decided to work with  
313 point estimates, it is possible to have a (non MaxEnt) probability distribution of networks when  
314 working with the entire chains after burn-in. For each network in our complete and abundance  
315 datasets, we estimated their configuration with maximum entropy using both types of MaxEnt  
316 network models.

317 Our method is in contrast with maximum entropy graph models that predict a probability distri-  
318 bution on networks under soft or hard constraints (e.g., Park and Newman 2004; Cimini et al.  
319 2019). We believe our approach to be more flexible, easier to compute, while allowing direct  
320 comparisons of empirical food webs with more complex networks with similar structure.

321 **Structure of MaxEnt networks**

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

333 Overall, we found that models based on the joint degree sequence (type II null and MaxEnt  
334 network models) reproduced the structure of empirical networks much better than the ones based  
335 on connectance and the neutral model of relative abundances. On average, the type II MaxEnt  
336 network model was better at predicting nestedness ( $0.62 \pm 0.08$ ) than its corresponding null model  
337 ( $0.73 \pm 0.05$ ; empirical networks:  $0.63 \pm 0.09$ ), as well as the proportion of cannibal species.  
338 However, the type II null model was better at predicting network diameter and average maximum  
339 similarity between species pairs. Predictions were similar between both type II models for the  
340 maximum trophic level and the proportion of omnivorous species.

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

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

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

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

<sup>341</sup> 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

343 be considered as simple building blocks of ecological networks (Milo et al. 2002; Stouffer et al.  
344 2007). Motifs are the backbone of complex ecological networks from which network structure  
345 is built upon and play a crucial role in community dynamics and assembly (Bastolla et al. 2009;  
346 Stouffer and Bascompte 2011). In fig. 4, we show that the motifs profile of networks generated  
347 using the type II MaxEnt network model was very close to the one of empirical data, and that  
348 they made better predictions than the type II null model. This is also shown in fig. 5, where we  
349 see that relationships between motifs pairs of empirical food webs are very similar to the ones  
350 of the type II MaxEnt networks. Overall, these results suggest that the type II MaxEnt network  
351 model can reproduce many aspects of network structure. This highlights the importance of the  
352 joint degree sequence in shaping ecological networks.

353 [Figure 4 about here.]

354 [Figure 5 about here.]

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

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

371 **Conclusion**

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

389 Our method and results could be used for different purposes. First, they could be used as first-  
390 order approximations of network structure when only state variables are known. This could  
391 prove useful when predicting network structure at large spatial scales, where few ecological in-  
392 formation is known at that scale. Second, they could be used as informative priors in Bayesian  
393 analyses of the structure of ecological networks (e.g., Cirtwill et al. 2019). Third, they could  
394 be used to make better predictions of pairwise species interactions by constraining the space of  
395 feasible networks, as discussed in Strydom et al. (2021). Finally, they could be used as alter-  
396 native null models of ecological networks to better understand ecological mechanisms driving

397 food-web structure. In that case, our model might need to be slightly adapted to give a proba-  
398 bility distribution of Boolean networks (in contrast with point estimates of maximum entropy  
399 networks).

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

## 407 Acknowledgments

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

## 414 References

- 415 Albert, R., and A. L. Barabasi. 2002. “Statistical Mechanics of Complex Networks.” *Reviews*  
416 *of Modern Physics* 74 (1): 47–97. <https://doi.org/10.1103/RevModPhys.74.47>.
- 417 Banville, Francis, Steve Vissault, and Timothée Poisot. 2021. “Mangal.jl and EcologicalNet-  
418 works.jl: Two Complementary Packages for Analyzing Ecological Networks in Julia.” *Jour-*  
419 *nal of Open Source Software* 6 (61): 2721. <https://doi.org/10.21105/joss.02721>.
- 420 Bascompte, J., P. Jordano, C. J. Melian, and J. M. Olesen. 2003. “The Nested Assembly of  
421 Plant-Animal Mutualistic Networks.” *Proceedings of the National Academy of Sciences*

- 422      *of the United States of America* 100 (16): 9383–87. <https://doi.org/10.1073/pnas.1633576100>.
- 424      Bastolla, Ugo, Miguel A. Fortuna, Alberto Pascual-Garcia, Antonio Ferrera, Bartolo Luque, and  
425      Jordi Bascompte. 2009. “The Architecture of Mutualistic Networks Minimizes Competition  
426      and Increases Biodiversity.” *Nature* 458 (7241): 1018–U91. <https://doi.org/10.1038/nature07950>.
- 428      Beck, Christian. 2009. “Generalised Information and Entropy Measures in Physics.” *Contem-  
429      porary Physics* 50 (4): 495–510. <https://doi.org/10.1080/00107510902823517>.
- 430      Cimini, Giulio, Tiziano Squartini, Fabio Saracco, Diego Garlaschelli, Andrea Gabrielli, and  
431      Guido Caldarelli. 2019. “The Statistical Physics of Real-World Networks.” *Nature Reviews  
432      Physics* 1 (1): 58–71. <https://doi.org/10.1038/s42254-018-0002-6>.
- 433      Cirtwill, Alyssa R., Anna Eklof, Tomas Roslin, Kate Wootton, and Dominique Gravel. 2019.  
434      “A Quantitative Framework for Investigating the Reliability of Empirical Network Construc-  
435      tion.” *Methods in Ecology and Evolution* 10 (6): 902–11. <https://doi.org/10.1111/2041-210X.13180>.
- 437      Cohen, Joel E., Tomas Jonsson, and Stephen R. Carpenter. 2003. “Ecological Community  
438      Description Using the Food Web, Species Abundance, and Body Size.” *Proceedings of  
439      the National Academy of Sciences* 100 (4): 1781–86. <https://doi.org/10.1073/pnas.232715699>.
- 441      Delmas, Eva, Mathilde Besson, Marie Hélène Brice, Laura A. Burkle, Giulio V. Dalla Riva,  
442      Marie Josée Fortin, Dominique Gravel, et al. 2019. “Analysing Ecological Networks of  
443      Species Interactions.” *Biological Reviews*. <https://doi.org/10.1111/brv.12433>.
- 444      Dunning, Iain, Joey Huchette, and Miles Lubin. 2017. “JuMP: A Modeling Language for  
445      Mathematical Optimization.” *SIAM Review* 59 (2): 295–320. <https://doi.org/10.1137/15M1020575>.
- 447      Fortuna, M. A., and J. Bascompte. 2006. “Habitat Loss and the Structure of Plant-Animal  
448      Mutualistic Networks.” *Ecology Letters* 9 (3): 278–83. <https://doi.org/10.1111/j.1461-0248.2005.00868.x>.

- 450 Fortuna, M. A., Daniel B. Stouffer, Jens M. Olesen, Pedro Jordano, David Mouillot, Boris R.  
451 Krasnov, Robert Poulin, and Jordi Bascompte. 2010. “Nestedness Versus Modularity in  
452 Ecological Networks: Two Sides of the Same Coin?” *Journal of Animal Ecology* 79 (4):  
453 811–17. <https://doi.org/10.1111/j.1365-2656.2010.01688.x>.
- 454 Frank, S. A., and E. Smith. 2011. “A Simple Derivation and Classification of Common Prob-  
455 ability Distributions Based on Information Symmetry and Measurement Scale.” *Journal of*  
456 *Evolutionary Biology* 24 (3): 469–84. <https://doi.org/10.1111/j.1420-9101.2010.02204.x>.
- 458 Harremoës, Peter, and Flemming Topsøe. 2001. “Maximum Entropy Fundamentals.” *Entropy*  
459 3 (3): 191–226. <https://doi.org/10.3390/e3030191>.
- 460 Harte, J., and Erica A. Newman. 2014. “Maximum Information Entropy: A Foundation for  
461 Ecological Theory.” *Trends in Ecology & Evolution* 29 (7): 384–89. <https://doi.org/10.1016/j.tree.2014.04.009>.
- 463 Harte, J., T. Zillio, E. Conlisk, and A. B. Smith. 2008. “Maximum Entropy and the State-  
464 Variable Approach to Macroecology.” *Ecology* 89 (10): 2700–2711. <https://doi.org/10.1890/07-1369.1>.
- 466 Hoorn, Pim van der, Gabor Lippner, and Dmitri Krioukov. 2018. “Sparse Maximum-Entropy  
467 Random Graphs with a Given Power-Law Degree Distribution.” *Journal of Statistical Physics*  
468 173 (3): 806–44. <https://doi.org/10.1007/s10955-017-1887-7>.
- 469 Hortal, Joaquín, Francesco de Bello, José Alexandre F. Diniz-Filho, Thomas M. Lewinsohn,  
470 Jorge M. Lobo, and Richard J. Ladle. 2015. “Seven Shortfalls That Beset Large-Scale  
471 Knowledge of Biodiversity.” *Annual Review of Ecology, Evolution, and Systematics* 46 (1):  
472 523–49. <https://doi.org/10.1146/annurev-ecolsys-112414-054400>.
- 473 Jaynes, E. T. 1957a. “Information Theory and Statistical Mechanics.” *Physical Review* 106 (4):  
474 620–30. <https://doi.org/10.1103/PhysRev.106.620>.
- 475 ———. 1957b. “Information Theory and Statistical Mechanics. II.” *Physical Review* 108 (2):  
476 171–90. <https://doi.org/10.1103/PhysRev.108.171>.

- 477 Khinchin, A. Ya. 2013. *Mathematical Foundations of Information Theory*. Courier Corpora-  
478 tion.
- 479 MacDonald, Arthur Andrew Meahan, Francis Banville, and Timothée Poisot. 2020. “Revisiting  
480 the Links-Species Scaling Relationship in Food Webs.” *Patterns* 0 (0). <https://doi.org/10.1016/j.patter.2020.100079>
- 481
- 482 Martyushev, L. M., and V. D. Seleznev. 2006. “Maximum Entropy Production Principle in  
483 Physics, Chemistry and Biology.” *Physics Reports-Review Section of Physics Letters* 426  
484 (1): 1–45. <https://doi.org/10.1016/j.physrep.2005.12.001>.
- 485 Milo, R., S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon. 2002. “Net-  
486 work Motifs: Simple Building Blocks of Complex Networks.” *Science* 298 (5594): 824–27.  
487 <https://doi.org/10.1126/science.298.5594.824>.
- 488 Park, Juyong, and M. E. J. Newman. 2004. “Statistical Mechanics of Networks.” *Physical  
489 Review E* 70 (6): 066117. <https://doi.org/10.1103/PhysRevE.70.066117>.
- 490 Phillips, Steven J., Robert P. Anderson, and Robert E. Schapire. 2006. “Maximum Entropy  
491 Modeling of Species Geographic Distributions.” *Ecological Modelling* 190 (3): 231–59.  
492 <https://doi.org/10.1016/j.ecolmodel.2005.03.026>.
- 493 Pielou, Evelyn C. 1975. “Ecological Diversity.” In. 574.524018 P5.
- 494 Poisot, Timothee, Benjamin Baiser, Jennifer A. Dunne, Sonia Kefi, Francois Massol, Nico-  
495 las Mouquet, Tamara N. Romanuk, Daniel B. Stouffer, Spencer A. Wood, and Dominique  
496 Gravel. 2016. “Mangal - Making Ecological Network Analysis Simple.” *Ecography* 39 (4):  
497 384–90. <https://doi.org/10.1111/ecog.00976>.
- 498 Pomeranz, Justin Page, Ross M. Thompson, Timothée Poisot, Jon S. Harding, and Justin P. F.  
499 Pomeranz. 2018. “Data from: Inferring Predator-Prey Interactions in Food Webs.” Dryad.  
500 <https://doi.org/10.5061/DRYAD.K59M37F>.
- 501 Shannon, C. E. 1948. “A Mathematical Theory of Communication.” *The Bell System Technical  
502 Journal* 27 (3): 379–423. <https://doi.org/10.1002/j.1538-7305.1948.tb01338.x>.

- 503 Staniczenko, Phillip P. A., Jason C. Kopp, and Stefano Allesina. 2013. "The Ghost of Nested-  
504 ness in Ecological Networks." *Nature Communications* 4 (1): 1391. <https://doi.org/10.1038/ncomms2422>.
- 505
- 506 Stock, Michiel, Timothée Poisot, and Bernard De Baets. 2021. "Optimal Transportation Theory  
507 for Species Interaction Networks." *Ecology and Evolution* 11 (9): 3841–55. <https://doi.org/10.1002/ece3.7254>.
- 508
- 509 Stouffer, Daniel B., and Jordi Bascompte. 2011. "Compartmentalization Increases Food-Web  
510 Persistence." *Proceedings of the National Academy of Sciences* 108 (9): 3648–52. <https://doi.org/10.1073/pnas.1014353108>.
- 511
- 512 Stouffer, Daniel B., Juan Camacho, Wenxin Jiang, and Luis A. Nunes Amaral. 2007. "Evidence  
513 for the Existence of a Robust Pattern of Prey Selection in Food Webs." *Proceedings of the  
514 Royal Society B-Biological Sciences* 274 (1621): 1931–40. <https://doi.org/10.1098/rspb.2007.0571>.
- 515
- 516 Strydom, Tanya, Michael D. Catchen, Francis Banville, Dominique Caron, Gabriel Dansereau,  
517 Philippe Desjardins-Proulx, Norma R. Forero-Muñoz, et al. 2021. "A Roadmap Towards  
518 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>.
- 519
- 520
- 521 Strydom, Tanya, Giulio V. Dalla Riva, and Timothee Poisot. 2021. "SVD Entropy Reveals the  
522 High Complexity of Ecological Networks." *Frontiers in Ecology and Evolution* 9: 623141.  
523 <https://doi.org/10.3389/fevo.2021.623141>.
- 524
- 525 Williams, Richard J. 2011. "Biology, Methodology or Chance? The Degree Distributions of  
526 Bipartite Ecological Networks." *PLOS ONE* 6 (3): e17645. <https://doi.org/10.1371/journal.pone.0017645>.
- 527
- 528 Williams, Richard J., and Neo D. Martinez. 2000. "Simple Rules Yield Complex Food Webs."  
*Nature* 404 (6774): 180–83. <https://doi.org/10.1038/35004572>.

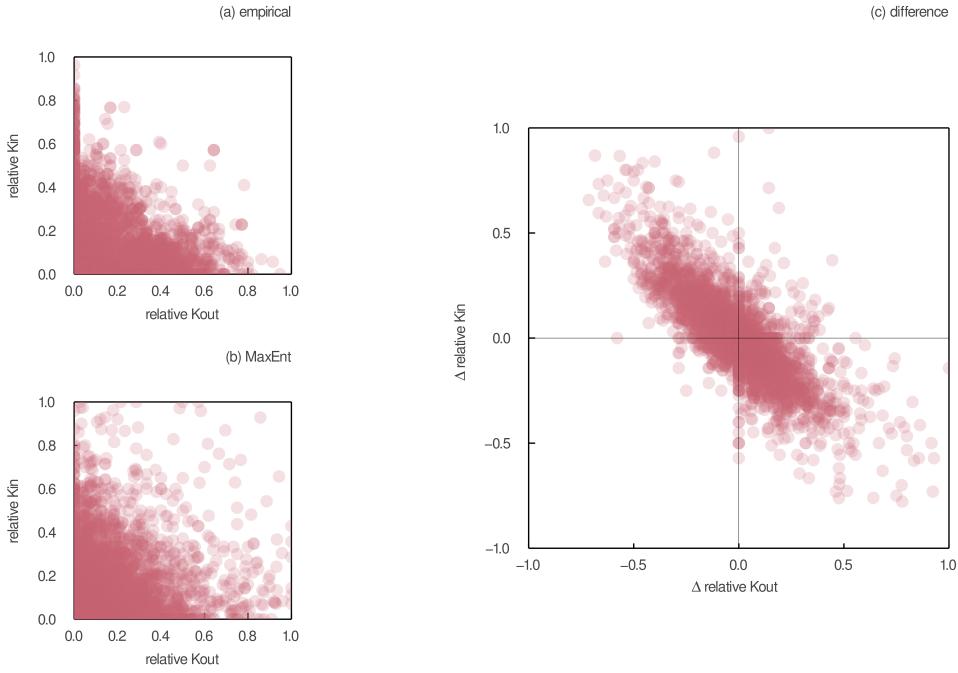


Figure 1: Relative number of predators ( $k_{in}$ ) as a function of the relative number of preys ( $k_{out}$ ) for each species in (a) empirical and (b) maximum entropy joint degree sequences. Empirical networks include all food webs archived on Mangal, as well as the New-Zealand and Tuesday lake datasets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

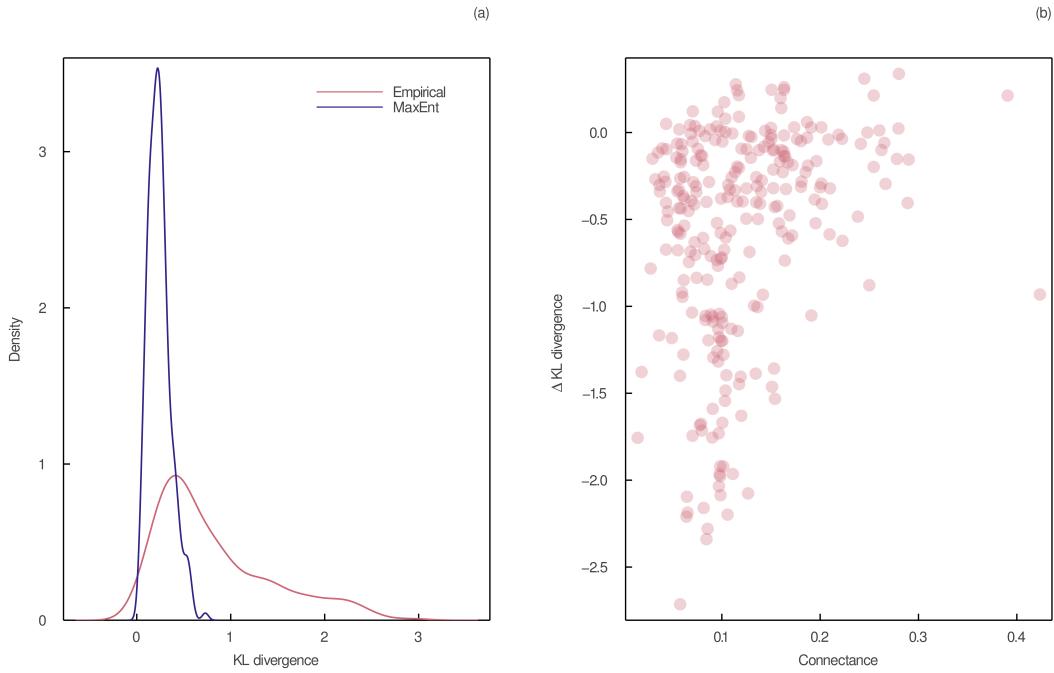


Figure 2: (a) Probability density of KL divergence between in and out-degree sequences of empirical and maximum entropy joint degree sequences. (b) Difference between the KL divergence of empirical and predicted networks as a function of connectance. In both panels, empirical networks include all food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

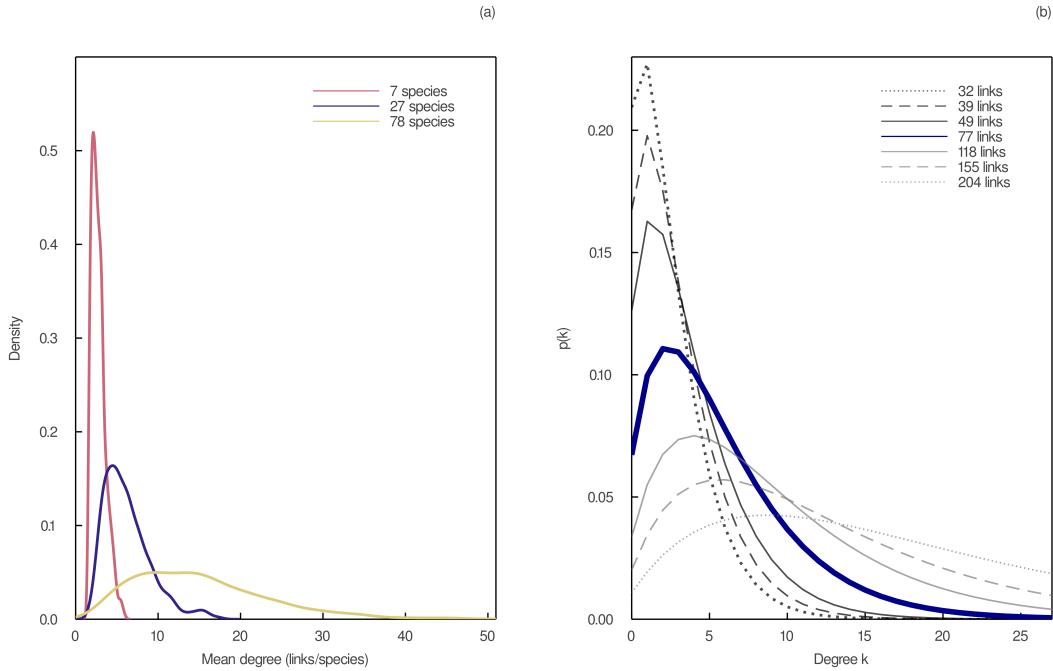


Figure 3: (a) Probability density of the predicted mean degree for different values of species richness. The number of interactions was predicted using the flexible links model fitted to all empirical networks in our complete dataset (eq. 17). (b) Degree distributions of maximum entropy for a network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

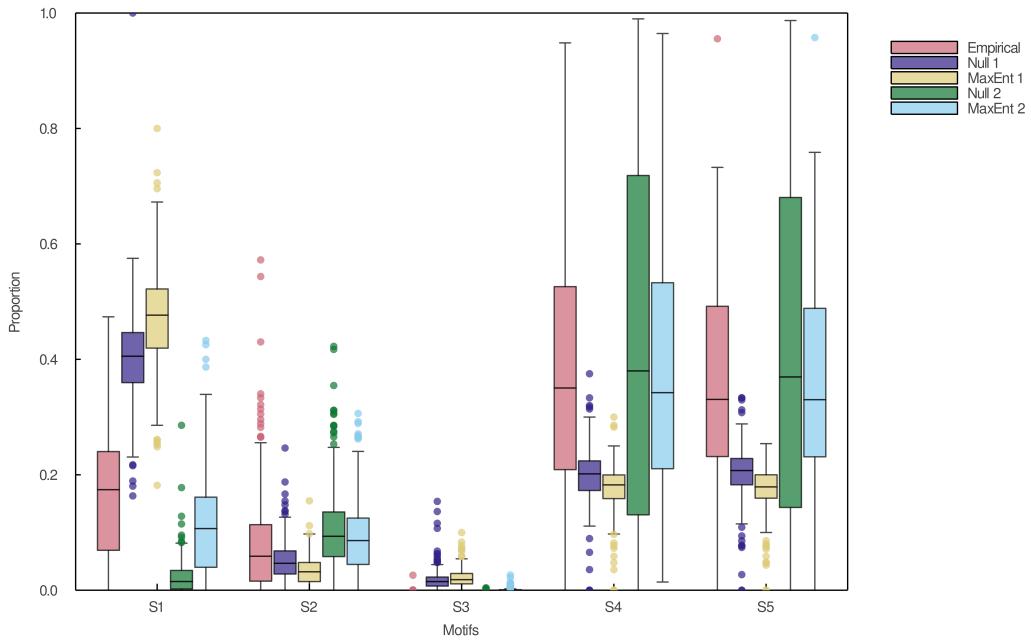


Figure 4: Motifs profile of empirical, maximum entropy, and null food webs. Empirical networks include all food webs archived on Mangal, as well as the New-Zealand and Tuesday lake datasets (our complete dataset). Maximum entropy networks were estimated using a simulated annealing algorithm to find the network of maximum SVD entropy while maintaining the connectance (type I MaxEnt network model) or the joint degree sequence (type II MaxEnt network model). The predictions of the types I and II null models are also plotted. Boxplots display the median proportions of each motif (middle horizontal lines), as well as the first (bottom horizontal lines) and third (top horizontal lines) quartiles. Vertical lines encompass all data points that fall within 1.5 times the interquartile range from both quartiles, and dots are data points that fall outside this range. Motifs names are from Stouffer et al. (2007).

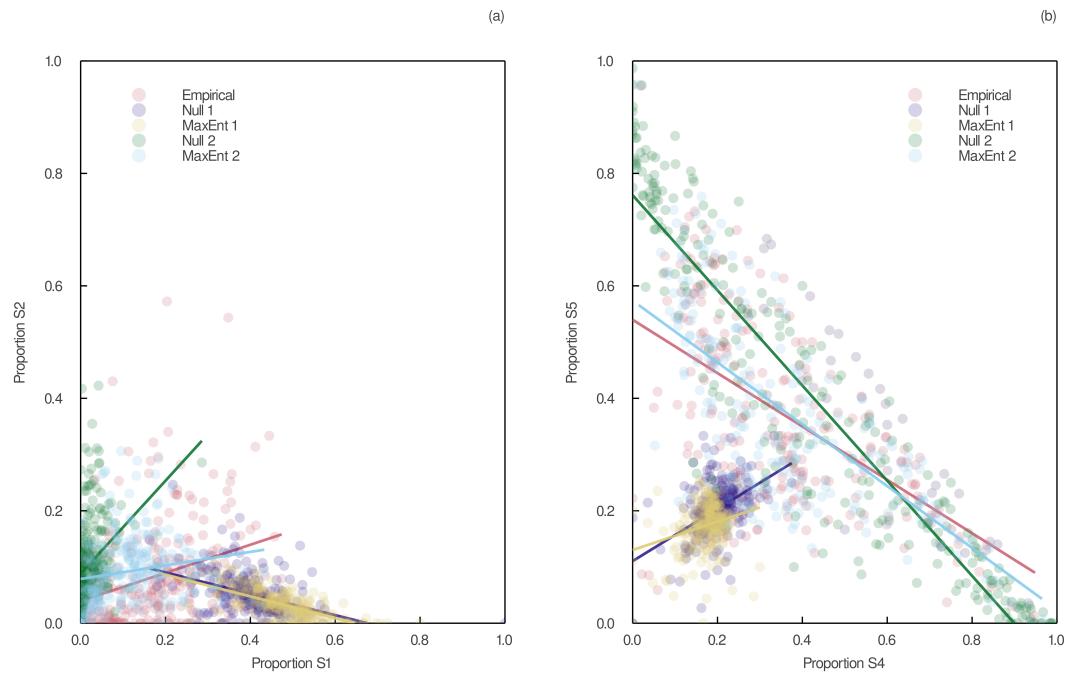


Figure 5: Relationship between motifs proportions of empirical, maximum entropy, and null food webs. Empirical networks include all food webs archived on Mangal, as well as the New-Zealand and Tuesday lake datasets (our complete dataset). Maximum entropy networks were estimated using a simulated annealing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

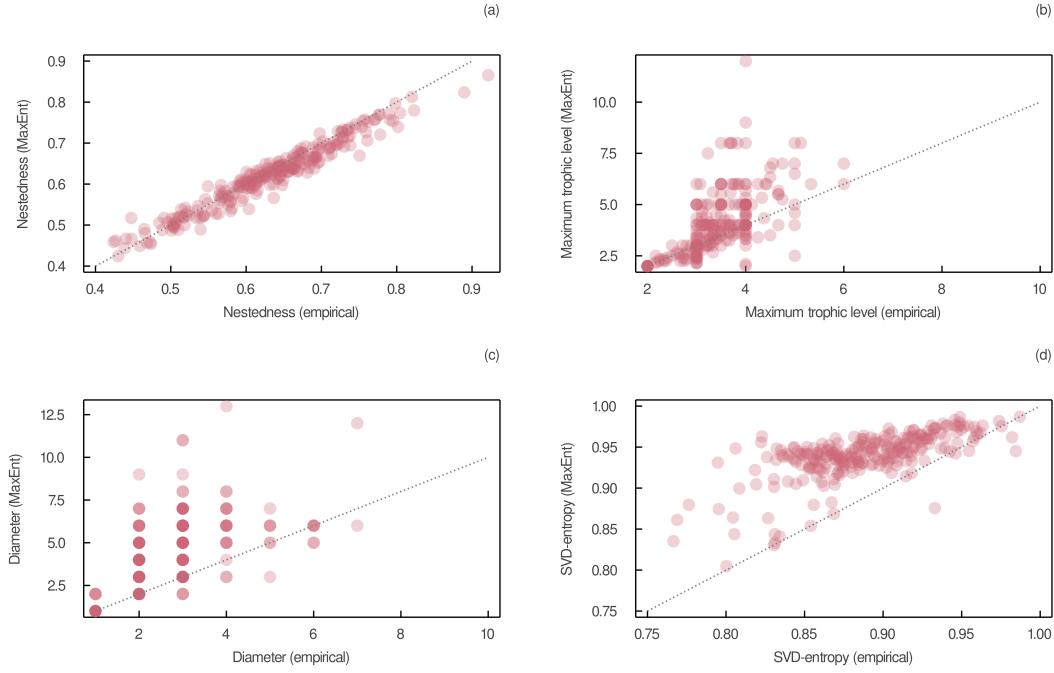


Figure 6: Relationship between the structure of empirical and maximum entropy food webs. Empirical networks include all food webs archived on Mangal, as well as the New-Zealand and Tuesday lake datasets (our complete dataset). Maximum entropy networks were estimated using a simulated annealing algorithm to find the network of maximum SVD entropy while maintaining the joint degree sequence (type II MaxEnt network model). (a) Nestedness (estimated using the spectral radius of the adjacency matrix), (b) the maximum trophic level, (c) the network diameter, and (d) the SVD entropy were measured on these empirical and maximum entropy food webs. The identity line is plotted in each panel.