

# Revisiting the links-species scaling relationship in food webs

ARTHUR ANDREW MEAHAN MACDONALD<sup>1,2</sup>, FRANCIS BANVILLE<sup>2,3</sup>, AND TIMOTHÉE POISOT<sup>2,3</sup>

<sup>1</sup>*Université de Montréal, Département de Sciences Biologiques, H2V 0B3, Montréal, QC, Canada*

<sup>2</sup>*Québec Centre for Biodiversity Sciences, H3A 1B1, Montréal QC, Canada*

<sup>3</sup>*Université de Montréal, Département de Sciences Biologiques, H2V 0B3, Montréal QC, Canada*

## Abstract

Predicting the number of interactions that species in a food web will establish is an important task. These trophic interactions underlie many ecological and evolutionary processes, ranging from biomass fluxes, ecosystem stability, resilience to extinction, and resistance against novel species. We investigate and compare several ways to predict the number of interactions in food webs. We conclude that a simple beta-binomial model outperforms other models, with the added desirable property of respecting biological constraints. We show how this simple relationship gives rise to a predicted distribution of several quantities related to link number in food webs, including the scaling of network structure with space, and the probability that a network will be stable.

## <sup>1</sup> Introduction

- <sup>2</sup> Community ecologists are fascinated by counting things. It is therefore no surprise that early food  
<sup>3</sup> web research paid so much attention to counting species, counting trophic links, and uncovering  
<sup>4</sup> the relationship that binds them – and it is undeniable that these inquiries kickstarted what is now  
<sup>5</sup> one of the most rapidly growing fields of ecology [1]. More species ( $S$ ) always means more links ( $L$ );  
<sup>6</sup> this scaling is universal and appears both in observed food webs and under purely neutral models

7 of food web structure [2]. In fact, these numbers underlie most measures used to describe food  
8 webs [3]. The structure of a food web, in turn, is almost always required to understand how the  
9 community functions, develops, and responds to changes [4,5], to the point where some authors  
10 suggested that describing food webs was a necessity for community ecology [6,7]. To this end, a  
11 first step is to come up with an estimate for the number of existing trophic links, through sampling  
12 or otherwise. Although both  $L$  and  $S$  can be counted in nature, the measurement of links is orders  
13 of magnitude more difficult than the observation of species [8,9]. As a result, we have far more  
14 information about values of  $S$ . In fact, the distribution of species richness across the world is  
15 probably the most frequently observed and modelled ecological phenomenon. Therefore, if we  
16 can predict  $L$  from  $S$  in an ecologically realistic way, we would be in a position to make first order  
17 approximations of food web structure at large scales, even under our current data-limited regime.

18 Measures of food web structure react most strongly to a handful of important quantities. The first  
19 and most straightforward is  $L$ , the number of trophic links among species. This quantity can be  
20 large, especially in species-rich habitats, but it cannot be arbitrarily large. It is clear to any observer  
21 of nature that of all imaginable trophic links, only a fraction actually occur. If an ecological com-  
22 munity contains  $S$  species, then the maximum number of links in its food web is  $S^2$ : a community  
23 of omnivorous cannibals. This leads to the second quantity: a ratio called *connectance* and defined  
24 by ecologists as  $Co = L/S^2$ . Connectance has become a fundamental quantity for nearly all other  
25 measures of food web structure and dynamics [10]. The third important quantity is another ratio:  
26 *linkage density*,  $L_D = L/S$ . This value represents the number of links added to the network for  
27 every additional species in the ecological system. A closely related quantity is  $L_D \times 2$ , which is the  
28 *average degree*: the average number of species with which any taxa is expected to interact, either  
29 as predator or prey. These quantities capture ecologically important aspects of a network, and all  
30 can be derived from the observation or prediction of  $L$  links among  $S$  species.

31 Because  $L$  represents such a fundamental quantity, many predictive models have been considered  
32 over the years. Here we describe three popular approaches before describing our own proposed  
33 model. The *link-species scaling (LSSL)* [11] assumes that all networks have the same *average degree*;

34 that is, most species should have the same number of links. Links are modelled as the number of  
35 species times a constant:

$$L_{LSSL} = b \times S \quad (1)$$

36 with  $b \approx 2$ . This model imagines that every species added to a community increases the number  
37 of links by two – for example, an animal which consumes one resource and is consumed by one  
38 predator. This model started to show its deficiencies when data on larger food webs became avail-  
39 able: in these larger webs,  $L$  increased faster than a linear function of  $S$ . Perhaps then all networks  
40 have the same *connectance* [12]? In other words, a food web is always equally filled, regardless of  
41 whether it has 5 or 5000 species. Under the so-called “constant connectance” model, the number  
42 of links is proportional to the richness squared,

$$L_{CC} = b \times S^2, \quad (2)$$

43 where  $b$  is a constant in  $]0, 1[$  representing the expected value of connectance. The assumption of  
44 a scaling exponent of 2 can be relaxed [12], so that  $L$  is not in direct proportion to the maximum  
45 number of links:

$$L_{PL} = b \times S^a. \quad (3)$$

46 This “power law” model can be parameterized in many ways, including spatial scaling and species  
47 area relationships [13]. It is also a general case of the previous two models, encompassing both  
48 link-species scaling ( $a = 1, b \approx 2$ ) and the strict constant connectance ( $a = 2, 0 < b < 1$ ) depend-  
49 ing on which parameters are fixed. Power laws are very flexible, and indeed this function matches  
50 empirical data well – so well that it is often treated as a “true” model which captures the scaling  
51 of link number with species richness [14–16], and from which we should draw ecological infer-  
52 ences about what shapes food webs. However, this approach is limited, because the parameters

53 of a power law relationship can arise from many mechanisms, and are difficult to reason about  
54 ecologically.

55 But the question of how informative parameters of a power law can be is moot. Indeed, both the  
56 general model and its variants share an important shortcoming: they cannot be used for predic-  
57 tion while remaining within the bounds set by ecological principles. This has two causes. First,  
58 models that are variations of  $L \approx b \times S^a$  have no constraints – with the exception of the “constant  
59 connectance” model, in which  $L_{cc}$  has a maximum value of  $S^2$ . However, we know that the num-  
60 ber of links within a food web is both lower and upper bounded [12,17]: there can be no more  
61 than  $S^2$  links, and there can be no fewer than  $S - 1$  links. This minimum of  $S - 1$  holds for food  
62 webs in which all species interact – for example, a community of plants and herbivores where no  
63 plants are inedible and all herbivores must eat [12]. Numerous simple food webs could have this  
64 minimal number of links – for example, a linear food chain wherein each trophic level consists of  
65 a single species, each of which consumes only the species below it; or a grazing herbivore which  
66 feeds on every plant in a field. Thus the number of links is constrained by ecological principles  
67 to be between  $S - 1$  and  $S^2$ , something which no present model includes. Secondly, accurate pre-  
68 dictions of  $L$  from  $S$  are often difficult because of how parameters are estimated. This is usually  
69 done using a Gaussian likelihood for  $L$ , often after log transformation of both  $L$  and  $S$ . While this  
70 approach ensures that predicted values of  $L$  are always positive, it does nothing to ensure that they  
71 stay below  $S^2$  and above  $S - 1$ . Thus a good model for  $L$  should meet these two needs: a bounded  
72 expression for the average number of links, as well as a bounded distribution for its likelihood.

73 Here we suggest a new perspective for a model of  $L$  as a function of  $S$  which respects ecological  
74 bounds, and has a bounded distribution of the likelihood. We include the minimum constraint by  
75 modelling not the total number of links, but the number in excess of the minimum. We include  
76 the maximum constraint in a similar fashion to the constant connectance model described above,  
77 by modelling the proportion of flexible links which are realized in a community.

## 78 Interlude - deriving a process-based model for the number of 79 links

80 Based on the ecological constraints discussed earlier, we know that the number of links  $L$  is an  
81 integer such that  $S - 1 \leq L \leq S^2$ . Because we know that there are at least  $S - 1$  links, there can  
82 be at most  $S^2 - (S - 1)$  links *in excess* of this quantity. The  $S - 1$  minimum links do not need to be  
83 modelled, because their existence is guaranteed as a pre-condition of observing the network. The  
84 question our model should address is therefore, how many of these  $S^2 - (S - 1)$  “flexible” links are  
85 actually present? A second key piece of information is that the presence of a link can be viewed  
86 as the outcome of a discrete stochastic event, with the alternative outcome that the link is absent.  
87 We assume that all of these flexible links have the same chance of being realized, which we call  $p$ .  
88 Then, if we aggregate across all possible species pairs, the expected number of links is

$$L_{FL} = p \times [S^2 - (S - 1)] + (S - 1), \quad (4)$$

89 where  $p \in [0, 1]$ . When  $p = 1$ ,  $L$  is at its maximum ( $S^2$ ), and when  $p = 0$  it is at the minimum value  
90 ( $S - 1$ ). We use the notation  $L_{FL}$  to represent that our model considers the number of “flexible”  
91 links in a food web; that is, the number of links in excess of the minimum but below the maximum.  
92 Because we assume that every flexible link is an independent stochastic event with only two out-  
93 comes, we can follow recent literature on probabilistic ecological networks [18] and represent  
94 them as independent Bernoulli trials with a probability of success  $p$ . This approach does not cap-  
95 ture ecological mechanisms known to act on food webs [19], but rather captures that any interac-  
96 tion is the outcome of many processes which can overall be considered probabilistic events [20].  
97 The assumption that flexible links can all be represented by Bernoulli events is an appropriate  
98 trade-off between biological realism and parameterization requirements.

99 Furthermore, the observation of  $L$  links in a food web represents an aggregation of  $S^2 - (S - 1)$  such  
100 trials. If we then assume that  $p$  is a constant for all links in a particular food web, but may vary

101 between food webs (a strong assumption which we later show is actually more stringent than what  
102 data suggest), we can model the distribution of links directly as a shifted beta-binomial variable:

$$[L|S, \mu, \phi] = \binom{S^2 - (S - 1)}{L - (S - 1)} \frac{B(L - (S - 1) + \mu\phi, S^2 - L + (1 - \mu)\phi)}{B(\mu\phi, (1 - \mu)\phi)} \quad (5)$$

103 Where  $B$  is the beta function,  $\mu$  is the average probability of a flexible link being realized (*i.e.* the  
104 average value of  $p$  across networks in the dataset) and  $\phi$  is the concentration around this value. The  
105 support of this distribution is limited to only ecologically realistic values of  $L$ : it has no probability  
106 mass below  $S - 1$  or above  $S^2$ . This means that the problem of estimating values for  $\mu$  and  $\phi$  is  
107 reduced to fitting the univariate distribution described in eq. (5). For more detailed explanation of  
108 the model derivation, fitting, and comparison, see Experimental Procedures.

109 In this paper we will compare our flexible links model to three previous models for  $L$ . We esti-  
110 mate parameters and compare the performance of all models using open data from the `mangal.io`  
111 networks database [21]. This online, open-access database collects published information on all  
112 kinds of ecological networks, including 255 food webs detailing interactions between consumers  
113 and resources [22]. We use these data to show how our flexible links model not only outperforms  
114 existing efforts at predicting the number of links, but also has numerous desirable properties from  
115 which novel insights about the structure of food webs can be derived.

## 116 Results and Discussion

### 117 Flexible links model fits better and makes a plausible range of predictions

Table 1: Comparison of the four different models. We show Pareto-smoothed important sampling values (PSIS-LOO) and their standard deviation. PSIS-LOO is similar to information criteria in that smaller values indicate better predictive performance. We also show expected log predictive density (ELPD) differences to the maximum for all models, along with the standard error (SE) of these differences.

Model	eq.	PSIS-LOO	$\Delta \text{ELPD}$	$SE_{\Delta \text{ELPD}}$
Flexible links	4	$2520.5 \pm 44.4$	0	0
Power law [13]	3	$2564.3 \pm 46.6$	-21.9	6.5
Constant [12]	2	$2811.0 \pm 68.3$	-145.3	21.1
Link-species scaling [11]	1	$39840.1 \pm 2795.1$	-18659.8	1381.7

118 When fit to the datasets archived on `mangal.io`, all four models fit without any problematic warn-  
 119 ings (see Experimental Procedures), while our model for flexible links outperformed previous so-  
 120 lutions to the problem of modelling  $L$ . The flexible links model, which we fit via a beta-binomial  
 121 observation model, had the most favourable values of PSIS-LOO information criterion (table 1)  
 122 and of expected log predictive density (ELPD), relative to the three competing models which used  
 123 a negative binomial observation model. Pareto-smoothed important sampling serves as a guide to  
 124 model selection [23]; like other information criteria it approximates the error in cross-validation  
 125 predictions. Smaller values indicate a model which makes better predictions. The calculation of  
 126 PSIS-LOO can also provide some clues about potential model fits; in our case the algorithm sug-  
 127 gested that the constant connectance model was sensitive to extreme observations. The expected  
 128 log predictive density (ELPD), on the other hand, measures the predictive performance of the  
 129 model; here, higher values indicate more reliable predictions [23]. This suggests that the flexible  
 130 links model will make the best predictions of  $L$ .

131 To be useful to ecologists, predictions of  $L$  must stay within realistic boundaries determined by eco-  
 132 logical principles. We generated posterior predictions for all models and visualized them against  
 133 these constraints (fig. 1). The LSSL model underestimates the number of links, especially in large  
 134 networks: its predictions were frequently lower than the minimum  $S - 1$ . The constant con-

nectance and power law models also made predictions below this value, especially for small values of  $S$ . The flexible links model made roughly the same predictions, but within ecologically realistic values.

## The flexible links model makes realistic predictions for small communities

Constraints on food web structure are especially important for small communities. This is emphasized in fig. 2, which shows that all models other than the flexible links model fail to stay within realistic ecological constraints when  $S$  is small. The link-species scaling model made around 29% of unrealistic predictions of link numbers for every value of  $S$  ( $3 \leq S \leq 750$ ). The constant connectance and power law models, on the other hand, also produced unrealistic results but for small networks only: more than 20% were unrealistic for networks comprising less than 12 and 7 species, respectively. Only the flexible links model, by design, never failed to predict numbers of links between  $S - 1$  and  $S^2$ . It must be noted that unrealistic predictions are most common in the shaded area of fig. 2, which represents 90% of the empirical data we used to fit the model; therefore it matters little that models agree for large  $S$ , since there are virtually no such networks observed.

## Parameter estimates for all models

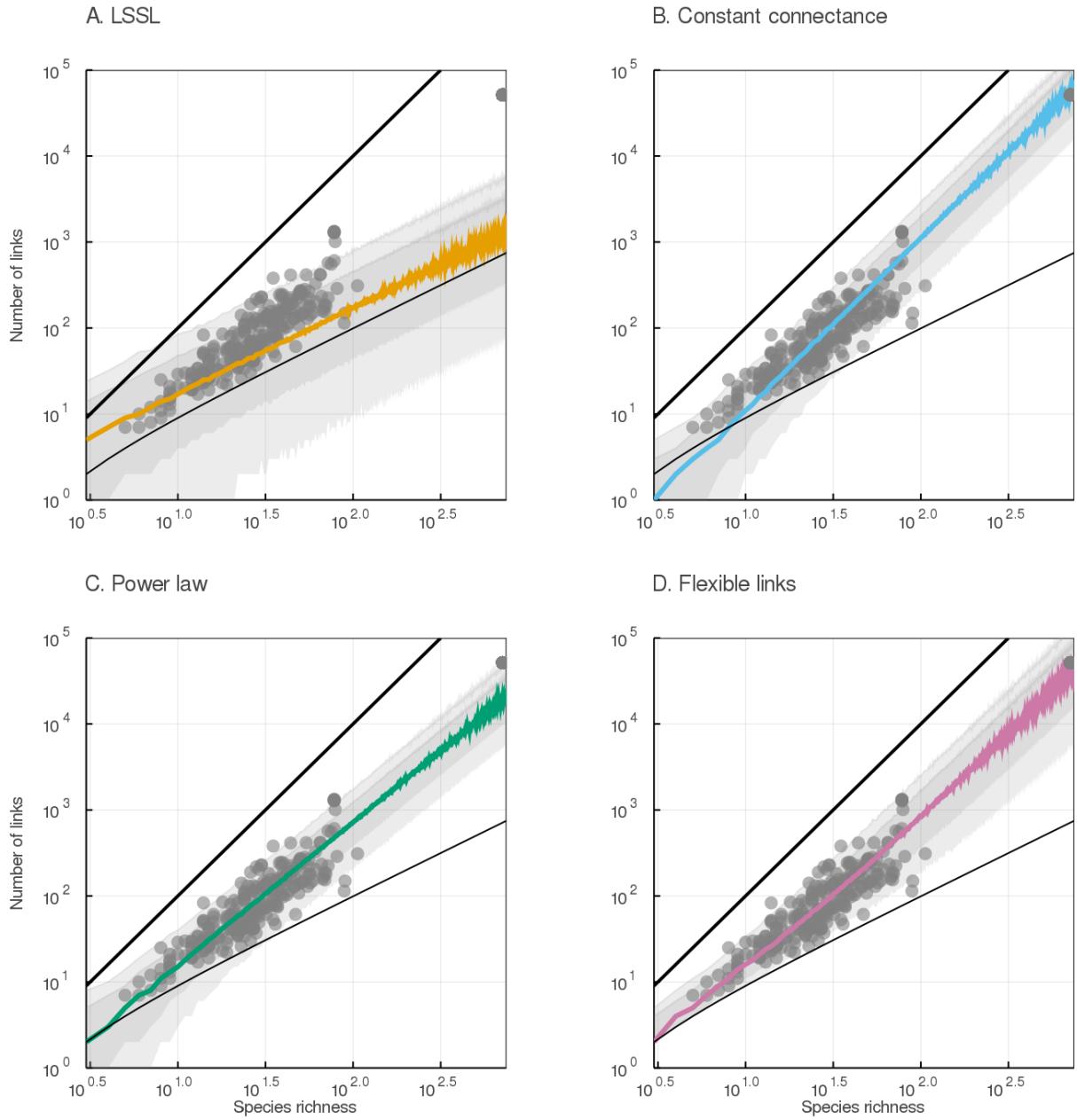
Table 2: Parameter estimates for all models. Mean and standard deviation (SD) are given for each parameter.

Model	parameter	interpretation	value	SD
$bS$ [11]	$b$	links per species	2.2	0.047
	$\kappa$	concentration of $L$ around mean	1.4	0.12
$bS^2$ [12]	$b$	proportion of links realized	0.12	0.0041
	$\kappa$	concentration of $L$ around mean	4.0	0.37
$bS^a$ [13]	$b$	proportion of relationship	0.37	0.054

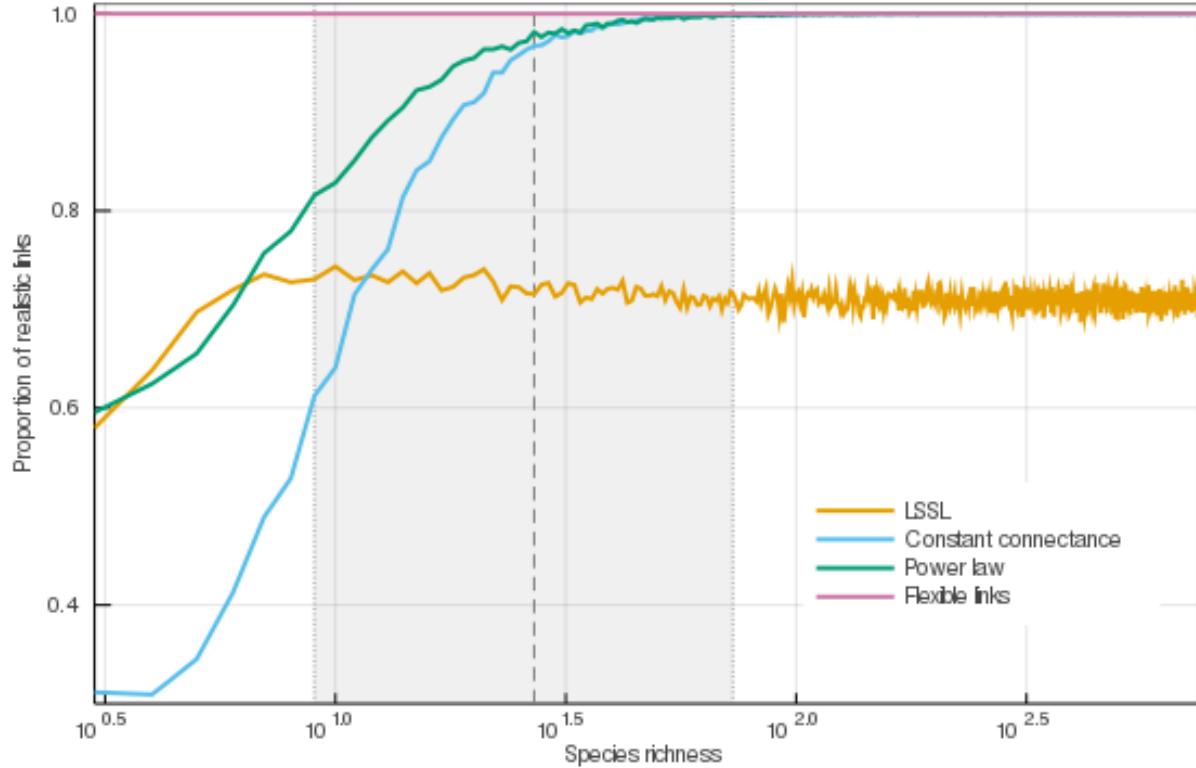
Model	parameter	interpretation	value	SD
	$a$	scaling of relationship	1.7	0.043
	$\kappa$	concentration of $L$ around mean	4.8	0.41
$(S^2 - (S - 1))p + S - 1$	$\mu$	average value of $p$	0.086	0.0037
	$\phi$	concentration around value of $\mu$	24.3	2.4

151 Although we did not use the same approach to parameter estimation as previous authors, our  
 152 approach to fitting these models recovered parameter estimates that are broadly congruent with  
 153 previous works. We found a value of 2.2 for  $b$  of the LSSL model (table 2), which is close to the  
 154 original value of approximately 2 [11]. Similarly, we found a value of 0.12 for  $b$  of the constant  
 155 connectance model, which was consistent with original estimates of 0.14 [12]. Finally, the param-  
 156 eter values we found for the power law were also comparable to earlier estimates [13]. All of these  
 157 models were fit with a negative binomial observation model, which has an additional parameter,  
 158  $\kappa$ , which is sometimes called a “concentration” parameter. This value increases from the top of our  
 159 table to the bottom, in the same sequence as predictive performance improves in table 1. This in-  
 160 dicates that the model predictions are more concentrated around the mean predicted by the model  
 161 (table 2, column 1).

162 Our parameter estimates for the flexible links model are ecologically meaningful. For large com-  
 163 munities, our model should behave similarly to the constant connectance model and so it is no  
 164 surprise that  $\mu$  was about 0.09, which is close to our value of 0.12 for constant connectance. In  
 165 addition, we obtained a rather large value of 24.3 for  $\phi$ , which shrinks the variance around the  
 166 mean of  $p$  to approximately 0.003 ( $var(p) = \mu(1 - \mu)/(1 + \phi)$ ). This indicates that food webs are  
 167 largely similar in their probability of flexible links being realized (thus showing how our previous  
 168 assumption that  $p$  might vary between food webs to be more conservative than strictly required).  
 169 The flexible links model also uses fewer parameters than the power law model and makes slightly  
 170 better predictions, which accounts for its superior performance in model comparison (table 1). In  
 171 fig. S1, we compare the maximum *a posteriori* (MAP) estimates of our model parameters to their



**Figure 1: The flexible links model fits better and makes a plausible range of predictions.** The number of links is plotted as a function of species richness obtained from the posterior distributions of A) the link-species scaling, B) the constant connectance, C) the power law and D) the flexible links models. In each panel, the colored line represents the median predicted link number and the grey areas cover the 78% and 97% percentile intervals. Empirical data from the `mangal.io` database are plotted in each panel (grey dots), as well as the minimal  $S - 1$  and maximal  $S^2$  number of links (thinner and bolder black lines, respectively). Predictions from the flexible links model are always plausible: they stay within these biological boundaries.



**Figure 2: Only the flexible links model makes realistic predictions for small communities.** Here we show the proportion of posterior predictions from each of our 4 models which fall outside ecologically realistic values. The proportion of predictions in the correct range increases with species richness for the constant connectance and power law models. Shaded area shows the 5%, 50% and 95% quantiles of the distribution of  $S$ , demonstrating that many communities have potentially incorrect predictions under previous models.

172 maximum likelihood estimates (MLE).

## 173 Connectance and linkage density can be derived from a model for links

174 Of the three important quantities which describe networks ( $L$ ,  $Co$  and  $L_D$ ), we have directly mod-  
175 elled  $L$  only. However, we can use the parameter estimates from our model for  $L$  to parameterize  
176 a distribution for connectance ( $L/S^2$ ) and linkage density ( $L/S$ ). We can derive this by noticing  
177 that eq. (4) can be rearranged to show how  $Co$  and  $L_D$  are linear transformations of  $p$ :

$$Co = \frac{L}{S^2} = p \left(1 - \frac{S-1}{S^2}\right) + \frac{S-1}{S^2}, \quad (6)$$

178 and

$$L_D = \frac{L}{S} = p \left(S - \frac{S-1}{S}\right) + \frac{S-1}{S}, \quad (7)$$

179 For food webs with many species, these equations simplify: eq. (4) can be expressed as a second  
180 degree polynomial,  $L_{FL} = p \times S^2 + (1-p) \times S + (p-1)$ , whose leading term is  $p \times S^2$ . Therefore, when  
181  $S$  is large, eq. (6) and eq. (7) respectively approach  $Co = L/S^2 \approx p$  and  $L_D = L/S \approx pS$ . A study of  
182 eq. (6) and eq. (7) also provides insight into the ecological interpretation of the parameters in our  
183 equation. For example, eq. (7) implies that adding  $n$  species should increase the linkage density by  
184 approximately  $p \times n$ . The addition of 11 new species ( $p^{-1}$  according to table 2) should increase the  
185 linkage density in the food web by roughly 1, meaning that each species in the original network  
186 would be expected to develop 2 additional interactions. Similarly, eq. (6) shows that when  $S$  is  
187 large, we should expect a connectance which is a constant. Thus  $p$  has an interesting ecological  
188 interpretation: it represents the average connectance of networks large enough that the proportion  
189  $(S-1)/S^2$  is negligible.

## 190 Applications of the flexible links model to key food web questions

191 Our model is generative, and that is important and useful: we can use this model to correctly  
 192 generate predictions that look like real data. This suggests that we can adapt the model, using  
 193 either its parameters or predictions or both, to get a new perspective on many questions in network  
 194 ecology. Here we show four possible applications that we think are interesting, in that relying on  
 195 our model eliminates the need to speculate on the structure of networks, or to introduce new  
 196 hypotheses to account for it.

### 197 Probability distributions for $L_D$ and $Co$

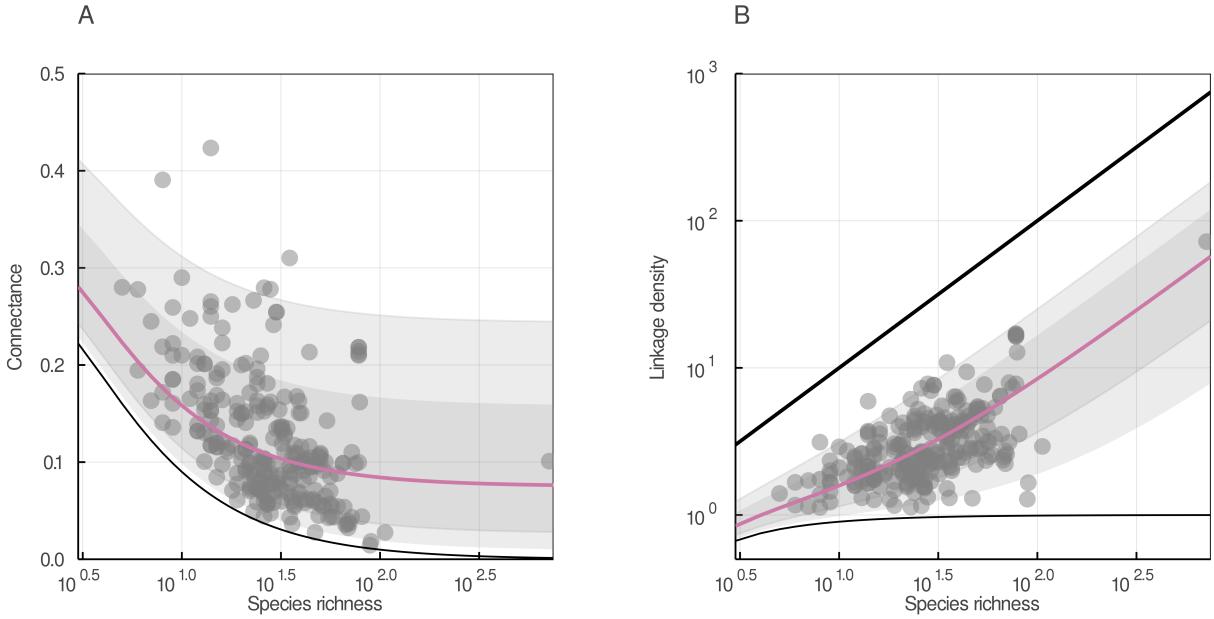
198 In a beta-binomial distribution, it is assumed that the probability of success  $p$  varies among groups  
 199 of trials according to a  $\text{Beta}(\mu\phi, (1 - \mu)\phi)$  distribution. Since  $p$  has a beta distribution, the linear  
 200 transformations described by eq. (6) and eq. (7) also describe beta distributions which have been  
 201 shifted and scaled according to the number of species  $S$  in a community. This shows that just as  $L$   
 202 must be within ecologically meaningful bounds,  $Co$  (eq. (6)) and  $L_D$  (eq. (7)) must be as well. The  
 203 connectance of a food web is bounded by  $(S - 1)/S^2$  and 1, while the linkage density is bounded  
 204 by  $(S - 1)/S$  and  $S$ .

205 We can convert the beta distribution for  $p$  into one for  $Co$  by replacing  $p$  with the transformation  
 206 of  $Co$  as described above (eq. (6)), and rescaling by the new range:

$$[Co|S, \mu, \phi] = \frac{\left( Co - \frac{S-1}{S^2} \right)^{\mu\phi-1} (1 - Co)^{(1-\mu)\phi-1}}{\left( 1 - \frac{S-1}{S^2} \right)^{\phi-1} \times B(\mu\phi, (1 - \mu)\phi)} \quad (8)$$

207 Similarly, we can convert the distribution for  $p$  into one for  $L_D$  by replacing  $p$  with the transfor-  
 208 mation that gives  $L_D$  (eq. (7))

$$[L_D|S, \mu, \phi] = \frac{\left( L_D - \frac{S-1}{S} \right)^{\mu\phi-1} (1 - L_D)^{(1-\mu)\phi-1}}{\left( S - \frac{S-1}{S} \right)^{\phi-1} \times B(\mu\phi, (1 - \mu)\phi)} \quad (9)$$



**Figure 3: Connectance and linkage density can be derived from a model for links.** A) Connectance and B) linkage density are plotted as a function of species richness, for the maximum *a posteriori* estimates of the flexible links model. In each panel, the colored line represents the median predicted quantity and the grey areas cover the 78% and 97% percentile intervals. Empirical data from the mangal.io database are plotted in each panel (grey dots). In A), the minimal  $(S - 1)/S^2$  connectance and in B) the minimal  $(S - 1)/S$  and maximum  $S$  linkage density are plotted (black lines).

209 In fig. 3, we show that the connectance and linkage density obtained from the equations above fit  
 210 the empirical data well.

## 211 An analytic alternative to null-model testing

212 Ecologists are often faced with the issue of comparing several networks. A common question is  
 213 whether a given network has an “unusual” number of links relative to some expectation. Tradi-  
 214 tionally these comparisons have been done by simulating a “null” distribution of random matrices  
 215 [24,25]. This is intended to allow ecologists to compare food webs to a sort of standard, hopefully  
 216 devoid of whatever biological process could alter the number of links. Importantly, this approach  
 217 assumes that (i) connectance is a fixed property of the network, ignoring any stochasticity, and (ii)  
 218 the simulated network distribution is an accurate and unbiased description of the null distribu-  
 219 tion. Yet recent advances in the study of probabilistic ecological networks show that the existence

220 of links, and connectance itself is best thought of as a probabilistic quantity [18]. Given that con-  
 221 nectance drives most of the measures of food web structure [17], it is critical to have a reliable  
 222 means of measuring differences from the expectation. We provide a way to assess whether the  
 223 number of links in a network (and therefore its connectance) is surprising. We do so using maths  
 224 rather than simulations.

225 The shifted beta-binomial can be approximated by a normal distribution with mean  $\bar{L}$  and variance  
 226  $\sigma_L^2$ :

$$L \sim Normal(\bar{L}, \sigma_L^2)$$

$$\bar{L} = (S^2 - S + 1)\mu + S - 1$$

$$\sigma_L^2 = (S^2 - S + 1)\mu(1 - \mu)\left(1 + \frac{S(S - 1)}{\phi + 1}\right) \quad (10)$$

227 This normal approximation is considered good whenever the skewness of the target distribution  
 228 is modest. In food webs, this should be true whenever communities have more than about 10  
 229 species (see Experimental Procedures). This result means that given a network with observed  
 230 species richness  $S_{obs}$  and observed links  $L_{obs}$ , we can calculate its  $z$ -score, i.e. how many standard  
 231 deviations an observation is from the population average, as

$$z = \frac{L_{obs} - \bar{L}}{\sqrt{\sigma_L^2}}. \quad (11)$$

232 A network where  $L = \bar{L}$  will have a  $z$ -score of 0, and any network with more (fewer) links will  
 233 have a positive (negative)  $z$ -score. Following this method, we computed the empirical  $z$ -scores  
 234 for the 255 food webs archived on [mangal.io](#) (fig. 4). We found that 18 webs (7.1%) had a total  
 235 number of observed links unusually higher than what was expected under the flexible links model

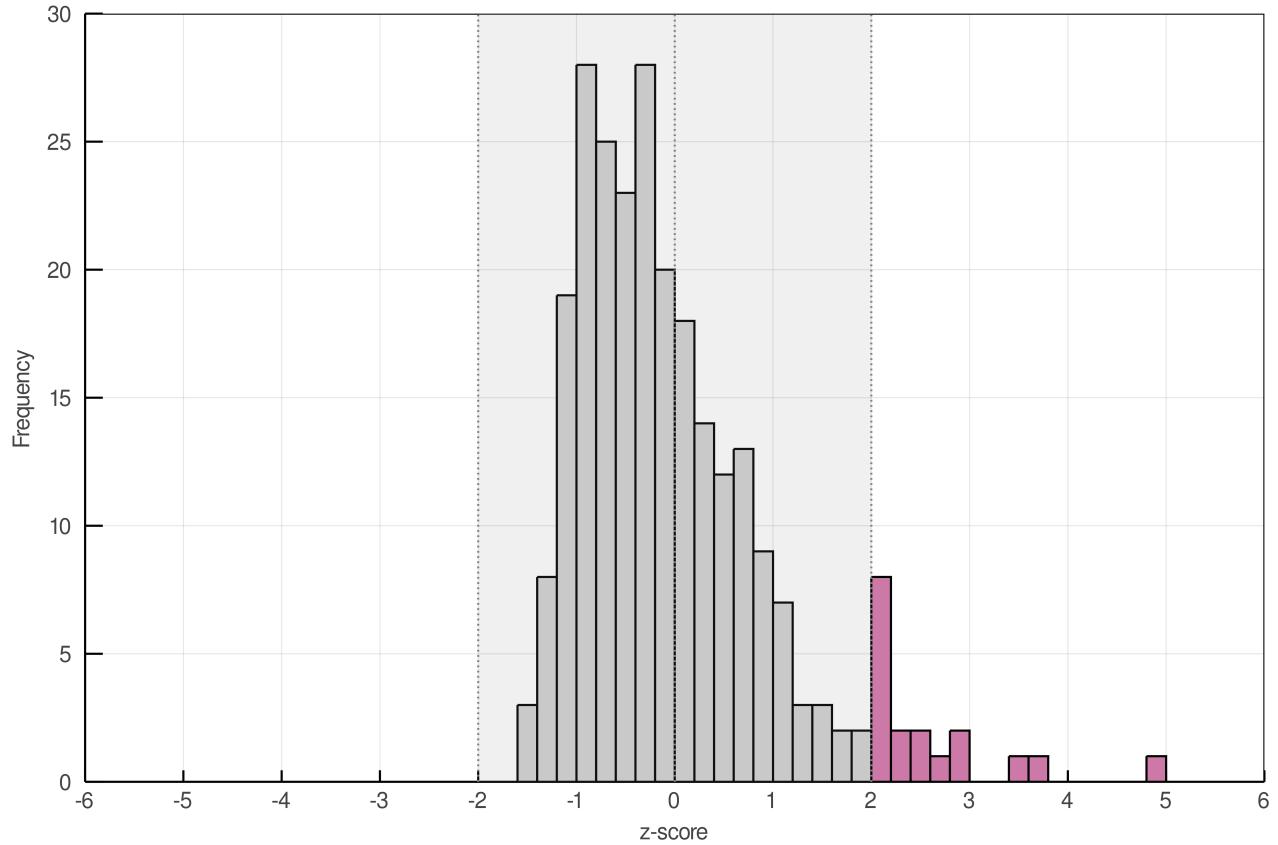
236 ( $z > 1.96$ ). These networks are interesting candidates for the study of mechanisms leading to high  
237 connectance.

238 Out of the 255 food webs, none was found to have an unusually low number of links ( $z < 1.96$ ).  
239 In fact,  $z$ -scores this low are not possible in this dataset: food webs having the minimum value of  
240  $S - 1$  links are still within two standard deviations of the mean, for this sample. However, this  
241 sample contains the full diversity of food webs found in the `mangal.io` database. Hence, this does  
242 not mean that no food web will ever have a  $z$ -score lower than -1.96. If the flexible links model  
243 is fit to data from a specific system, food webs might have a surprisingly low number of links  
244 when compared to this population average. These networks would be interesting candidates for  
245 the study of mechanisms leading to low connectance or for the identification of under-sampled  
246 webs. Ecologists can thus use our method to assess the deviation of their own food webs from  
247 their random expectations.

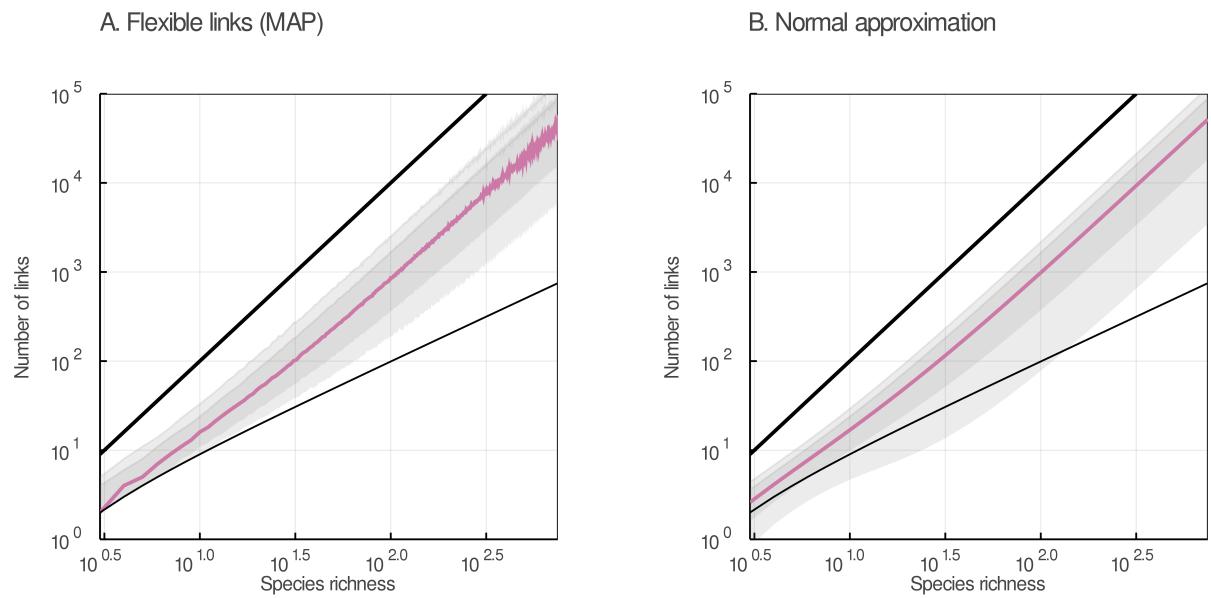
248 In fig. 5, we show that the predictions made by the normal approximation (panel B) are similar  
249 to those made by the beta distribution parameterized with the maximum *a posteriori* values of  $\mu$   
250 and  $\phi$  (panel A), although the former can undershoot the constraint on the minimum number of  
251 links. This undershooting, however, will not influence any actual  $z$ -scores, since no food webs  
252 have fewer than  $S - 1$  links and therefore no  $z$ -scores so low can ever be observed.

## 253 We should see many different network-area relationships

254 Our results bear important consequences for the nascent field of studying network-area relation-  
255 ships [26]. As it has long been observed that not all species in a food web diffuse equally through  
256 space [27], understanding how the shape of networks varies when the area increases is an impor-  
257 tant goal, and in fact underpins the development of a macroecological theory of food webs [28].  
258 Using a power-law as the acceptable relationship between species and area [29,30], the core idea  
259 of studying NAR is to predict network structure as a consequence of the effect of spatial scale on  
260 species richness [26]. Drawing on these results, we provide in fig. 6 a simple illustration of the fact  
261 that, due to the dispersal of values of  $L$ , the relationship between  $L/S$  and area can have a really



**Figure 4: Empirical distribution of food web  $z$ -scores** The  $z$ -scores of all food webs archived on `mangal.io` have been computed using eq. (11). Food webs with an absolute  $z$ -score above 1.96 are in pink. The shaded region comprises all food webs with an absolute  $z$ -score below 1.96.



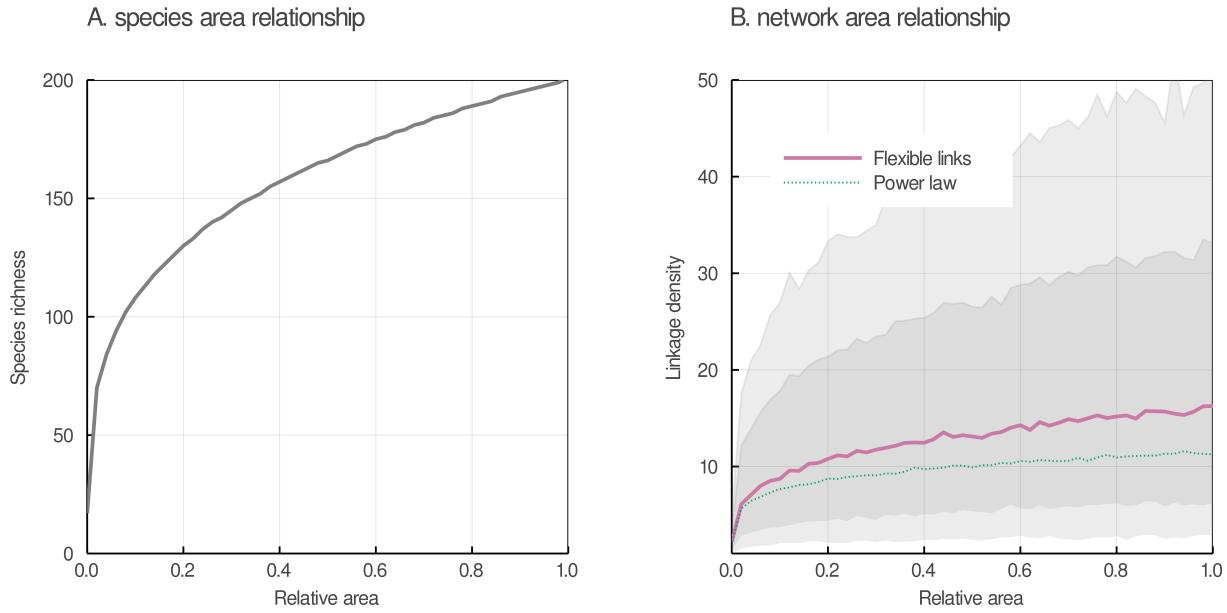
**Figure 5: The shifted beta-binomial distribution can be approximated by a normal distribution.** The number of links is plotted as a function of species richness obtained from A) the maximum *a posteriori* estimates of the flexible links model and B) its normal approximation. In each panel, the colored line represents the median predicted link number and the grey areas cover the 78% and 97% percentile intervals. The minimal  $S - 1$  and maximal  $S^2$  numbers of links are plotted in each panel (thinner and bolder black lines, respectively).

wide confidence interval. While our posterior predictions generally match the empirical results on this topic [31], they suggest that we will observe many relationships between network structure and space, and that picking out the signal of network-area relationships might be difficult.

As of now, not many NARs have been documented empirically; but after the arguments made by [26] which tie the shape of these relationships to macroecological processes, we fully expect these relationships to be more frequently described moving forward. Our results suggest that our expectation of the amount of noise in these relationships should be realistic; while it is clear that these relationships exist, because of the scaling of dispersion in the number of links with the number of species, they will necessarily be noisy. Any described relationships will exist within extremely wide confidence intervals, and it might require a large quantity of empirical data to properly characterize them. As such, our model can help in assessing the difficulty of capturing some foundational relationships of food web structure.

#### 274 Stability imposes a limit on network size

Can organisms really interact with an infinite number of partners? According to eq. (7), at large values of  $S$ , the linkage density scales according to  $p \times S$  (which is supported by empirical data), and so species are expected to have on average  $2 \times p \times S$  interactions. A useful concept in evolutionary biology is the “Darwinian demon” [32], *i.e.* an organism that would have infinite fitness in infinite environments. Our model seems to predict the emergence of what we call Eltonian demons, which can have arbitrarily large number of interactions. Yet we know that constraints on handling time of prey, for example, imposes hard limits on diet breadth [33]. This result suggests that there are other limitations to the size of food webs; indeed, the fact that  $L/S$  increases to worryingly large values only matters if ecological processes allow  $S$  to be large enough. It is known that food webs can get as high as energy transfer allows [5], and as wide as competition allows [34]. Furthermore, in more species-rich communities there is a greater diversity of functional traits among the interacting organisms; this limits interactions, because traits determine suitable interaction partners [35,36]. In short, and as fig. 2 suggests, since food webs are likely to be constrained to remain within an



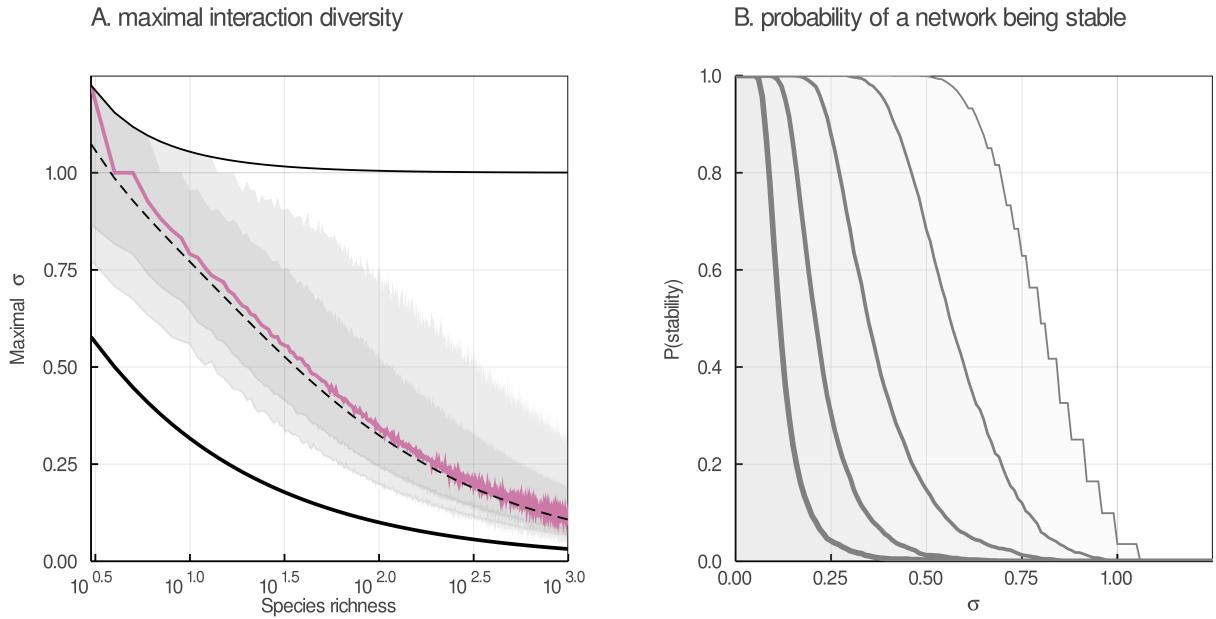
**Figure 6: Many different network-area relationships are supported by the data.** Representing the species richness as  $S = k \times A^z$  (panel A), with  $A$  being the relative area size,  $k = 200$  being the maximal species richness, and  $z = 0.27$  a scaling exponent [26]. We then use the posterior distribution of  $L$  to predict how  $L_D$  should scale with  $A$ . We compare the predictions of our model to that of the generally accepted power law (eq. (3)). While our model predicts a larger linkage density in larger areas (panel B), the confidence intervals around this prediction (grey areas covering the 78% and 97% percentile intervals) are extremely large. In particular, our model scales faster than the power law, but the confidence interval is high (due to the scaling of variance with  $S$ , eq. (10)). This suggests that we may observe either very weak, or very strong, effects of area on networks.

acceptable richness, we have no reason to anticipate that  $p \times S$  will keep growing infinitely.

Network structure may itself prevent  $S$  from becoming large. May [37] suggested that a network of richness  $S$  and connectance  $Co$  is stable as long as the criteria  $\sigma\sqrt{S \times Co} < 1$  is satisfied, with  $\sigma$  being the standard deviation of the strengths of interactions. Although this criteria is not necessarily stringent enough for the stability of food webs [38,39], it still defines an approximate maximum value  $\sigma^*$  which is the value above which the system is expected to be unstable. This threshold is  $\sigma^* = 1/\sqrt{L_D}$ , where  $L_D$  is defined as in eq. (7). We illustrate this result in fig. 7, which reveals that  $\sigma^*$  falls towards 0 for larger species richness. The result in fig. 7 is in agreement with previous simulations, placing the threshold for stability at about 1200 species in food webs. These results show how ecological limitations, for example on connectance and the resulting stability of the system, can limit the size of food webs [38,40]. In the second panel, we show that networks of increasing richness (thicker lines, varying on a log-scale from  $10^1$  to  $10^3$ ) have a lower probability of being stable, based on the proportion of stable networks in our posterior samples.

## Conclusions

Here we derived eq. (4), a model for the prediction of the number of links in ecological networks using a beta-binomial distribution for  $L$ , and show how it outperforms previous and more commonly used models describing this relationship. More importantly, we showed that our model has parameters with a clear ecological interpretation (specifically, the value of  $p$  in eq. (4) is the expected value of the connectance when  $S$  is large), and makes predictions which remain within biological boundaries. There are a variety of “structural” models for food webs, such as the niche model [41], the cascade model [42], the DBM [35] and ADBM [19], the minimum potential niche model [43], and the nested hierarchy model [44] to name a few. All of these models make predictions of food web structure: based on some parameters (usually  $S$  and  $L$ , and sometimes vectors of species-level parameters) they output an adjacency matrix  $\mathbf{A}_{S \times S}$  which contains either the presence or strength of trophic interactions. Therefore, these models require estimated values of  $L$  for a particular value of  $S$ , with the additional result that  $\sum \mathbf{A} = L$ . Our approach can serve to improve



**Figure 7: Stability imposes a limit on network size.** Using eq. (7), we can calculate the maximum standard deviation in the strength of interactions which should ensure food web stability,  $\sigma^* = 1/\sqrt{L_D}$  (panel A). The colored line represent the median value of maximum standard deviation, based on the posterior distribution of the flexible links model, and the grey areas cover the 78% and 97% percentile intervals. The fine and dark lines indicate the maximum and minimum values of maximum standard deviation, respectively. The dotted line shows the maximum for the average  $L_D$ , as given by eq. (7). The maximum standard deviation falls sharply when the number of species increases, which will limit the stability of large food webs, and therefore explain why Eltonian demons should not emerge. In panel B, we show the probability of a network with  $S$  species being stable, based on draws from the posterior distribution, for  $10 \leq S \leq 1000$  - larger networks (thicker lines) are increasingly unlikely to be stable.

314 the realism of these models, by imposing that the values of  $L$  they use are within realistic bound-  
315 aries. For example, a common use of structural models is to generate a set of “null” predictions:  
316 possible values of  $\mathbf{A}$  and  $L$  in the absence of the mechanism of interest. Empirical networks are  
317 then compared to this set of predictions, and are said to be significant if they are more extreme than  
318 95% of the observations [3]. A challenge in this approach is that structural models may generate  
319 a wide range of predictions, including ecologically impossible values, leading a high false nega-  
320 tive rate. This could be remedied by filtering this set of predictions according to our flexible links  
321 model, resulting in a narrower set of null predictions and a lower false negative rate. In general,  
322 our approach is complementary to other attempts to create ecologically-realistic food web models;  
323 for example, probabilistic models of the number of links per species which stay within ecological  
324 values [45].

325 This model also casts new light on previous results on the structure of food webs: small and large  
326 food webs behave differently [15]. Specifically, ecological networks most strongly deviate from  
327 scale free expectations when connectance is high [46]. In our model, this behaviour emerges nat-  
328 urally: connectance increases sharply as species richness decreases (fig. 3) – that is, where the ad-  
329 ditive term  $(S - 1)/S^2$  in eq. (6) becomes progressively larger. In a sense, small ecological networks  
330 are different only due to the low values of  $S$ . Small networks have only a very limited number of  
331 flexible links, and this drives connectance to be greater. Connectance in turn has implications for  
332 many ecological properties. Connectance is more than the proportion of realized interactions. It  
333 has been associated with some of the most commonly used network measures [17], and contains  
334 meaningful information on the stability [46,47] and dynamics [48] of ecological communities. A  
335 probability distribution for connectance not only accounts for the variability between networks,  
336 but can be used to describe fundamental properties of food webs and to identify ecological and  
337 evolutionary mechanisms shaping communities. A recent research direction has been to reveal  
338 its impact on resistance to invasion: denser networks with a higher connectance are comparatively  
339 more difficult to invade [49]; different levels of connectance are also associated with different com-  
340 binations of primary producers, consumers, and apex predators [41], which in turns determines  
341 which kind of species will have more success invading the network [50]. Because we can infer

342 connectance from the richness of a community, our model also ties the invasion resistance of a  
343 network to its species richness.

344 The relationship between  $L$  and  $S$  has underpinned most of the literature on food web structure  
345 since the 1980s. Additional generations of data have allowed us to progress from the link-species  
346 scaling law, to constant connectance, to more general formulations based on a power law. Our  
347 model breaks with this tradition of iterating over the same family of relationships, and instead  
348 draws from our knowledge of ecological processes, and from novel tools in probabilistic program-  
349 ming. As a result, we provide predictions of the number of links which are closer to empirical data,  
350 stimulate new ecological insights, and can be safely assumed to always fall within realistic values.

351 The results presented in fig. 6 (which reproduces results from [26]) and fig. 7 (which reproduces  
352 results from [38]) may seem largely confirmatory; in fact, the ability of our model to reach the con-  
353 clusions of previous milestone studies in food web ecology is a strong confirmation of its validity.

354 We would like to point out that these approaches would usually require ecologists to make infer-  
355 ences not only on the parameters of interests, but also on the properties of a network for a given  
356 species richness. In contrast, our model allows a real economy of parameters and offers ecologists  
357 the ability to get several key elements of network structure for free if only the species richness is  
358 known.

## 359 **Experimental Procedures**

### 360 **Availability of code and data**

361 All code and data to reproduce this article is available at the Open Science Framework (DOI:  
362 [10.17605/OSF.IO/YGPZ2](https://doi.org/10.17605/OSF.IO/YGPZ2)).

363 **Bayesian model definitions**

364 Generative models are flexible and powerful tools for understanding and predicting natural phe-  
365 nomena. These models aim to create simulated data with the same properties as observations.  
366 Creating such a model involves two key components: a mathematical expression which repre-  
367 sents the ecological process being studied, and a distribution which represents our observations  
368 of this process. Both of these components can capture our ecological understanding of a system,  
369 including any constraints on the quantities studied.

370 Bayesian models are a common set of generative models, frequently used to study ecological sys-  
371 tems. Here, we define Bayesian models for all 4 of the models described in eq. (1), eq. (2), eq. (3)  
372 and eq. (4). We use notation from [51], writing out both the likelihood and the prior as a product  
373 over all 255 food webs in the `mangal.io` database.

374 Link-species scaling (LSSL) model:

$$[b, \kappa | \mathbf{L}, \mathbf{S}] \propto \prod_{i=1}^{255} \text{negative binomial}(L_i | b \times S_i, e^\kappa) \times \text{normal}(b | 0.7, 0.02) \times \text{normal}(\kappa | 2, 1)$$

375 Constant connectance model:

$$[b, \kappa | \mathbf{L}, \mathbf{S}] \propto \prod_{i=1}^{255} \text{negative binomial}(L_i | b \times S_i^2, e^\kappa) \times \text{beta}(b | 3, 7) \times \text{normal}(\kappa | 2, 1)$$

376 Power law model:

$$[b, a, \kappa | \mathbf{L}, \mathbf{S}] \propto \prod_{i=1}^{255} \text{negative binomial}(L_i | \exp(b) \times S_i^a, e^\kappa) \times \text{normal}(b | -3, 1) \times \text{normal}(a | 2, 0.6) \times \text{normal}(\kappa | 2, 1)$$

377 Flexible links model:

$$[\mu, \phi | \mathbf{L}, \mathbf{S}] \propto \prod_{i=1}^{255} \text{beta binomial}(L_i - S_i + 1 | S_i^2 - S_i + 1, \mu \times e^\phi, (1 - \mu) \times e^\phi) \times \text{beta}(\mu | 3, 7) \times \text{normal}(\phi | 3, 0.5)$$

378 Note that while  $e^\phi$  is shown in these equations for clarity, in the text we use  $\phi$  to refer to the  
 379 parameter after exponentiation. In the above equations, bold type indicates a *vector* of values; we  
 380 use capital letters for  $\mathbf{L}$  and  $\mathbf{S}$  for consistency with the main text.

381 Because we want to compare all our models using information criteria, we were required to use  
 382 a discrete likelihood to fit all models. Our model uses a discrete likelihood by default, but the  
 383 previous three models (LSSL, constant connectance and the power law) normally do not. Instead,  
 384 these models have typically been fit with Gaussian likelihoods, sometimes after log-transforming  
 385  $L$  and  $S$ . For example, eq. (3) becomes a linear relationship between  $\log(L)$  and  $\log(S)$ . This en-  
 386 sures that predictions of  $L$  are always positive, but allows otherwise unconstrained variation on  
 387 both sides of the mean. To keep this same spirit, we chose the negative binomial distribution for  
 388 observations. This distribution is limited to positive integers, and can vary on both sides of the  
 389 mean relationship.

390 We selected priors for our Bayesian models using a combination of literature and domain expertise.  
 391 For example, we chose our prior distribution for  $p$  based on [12], who gave a value of constant  
 392 connectance equal to 0.14. While the prior we use is “informative”, it is weakly so; as [12] did  
 393 not provide an estimate of the variance for his value we chose a relatively large variation around  
 394 that mean. However, no information is available in the literature to inform a choice of prior for  
 395 concentration parameters  $\kappa$  and  $\phi$ . For these values, we followed the advice of [52] and performed  
 396 prior predictive checks. Specifically, we chose priors that generated a wide range of values for  $L_i$ ,  
 397 but which did not frequently predict webs of either maximum or minimum connectance, neither  
 398 of which are observed in nature.

399 **Explanation of shifted beta-binomial distribution**

400 Equation eq. (4) implies that  $L_{FL}$  has a binomial distribution, with  $S^2 - S + 1$  trials and a probability  
401  $p$  of any flexible link being realized:

$$[L|S, p] = \binom{S^2 - (S - 1)}{L - (S - 1)} p^{L-(S-1)} (1 - p)^{S^2-L},$$

402 This is often termed a *shifted binomial distribution*.

403 We also note that ecological communities are different in many ways besides their number of  
404 species ( $S$ ). Although we assume  $p$  to be fixed within one community, the precise value of  $p$  will  
405 change from one community to another. With this assumption, our likelihood becomes a shifted  
406 beta-binomial distribution:

$$[L|S, \mu, \phi] = \binom{S^2 - (S - 1)}{L - (S - 1)} \frac{B(L - (S - 1) + \mu\phi, S^2 - L + (1 - \mu)\phi)}{B(\mu\phi, (1 - \mu)\phi)} \quad (12)$$

407 Where  $B$  is the beta function. Thus, the problem of fitting this model becomes one of estimating  
408 the parameters of this univariate probability distribution.

409 **Model fitting - data and software**

410 We evaluated our model against 255 empirical food webs, available in the online database `mangal.io`  
411 [21]. We queried metadata (number of nodes and number of links) for all networks, and consid-  
412 ered as food webs all networks having interactions of predation and herbivory. We use Stan [53]  
413 which implements Bayesian inference using Hamiltonian Monte Carlo. We ran all models using  
414 four chains and 2000 iterations per chain. In our figures we use the posterior predictive distri-  
415 bution, which is a distribution described by the model after conditioning on the data. There are  
416 numerous ways to display a probability distribution; here we have chosen to do so using the expec-  
417 tation (mean) and two arbitrary percentile intervals: 78% and 97%. These intervals were chosen

418 based on the recommendations of [54], and allowed us to capture most of the probability density  
 419 in the tails of the posterior distributions.  
 420 Stan provides a number of diagnostics for samples from the posterior distribution, including  $\hat{R}$ ,  
 421 effective sample size, and measures of effective tree depth and divergent iterations. None of these  
 422 indicated problems with the posterior sampling. All models converged with no warnings; this  
 423 indicates that it is safe to make inferences about the parameter estimates and to compare the mod-  
 424 els. However, the calculation of PSIS-LOO for the LSSL model warned of problematic values of the  
 425 Pareto-k diagnostic statistic. This indicates that the model is heavily influenced by large values.  
 426 Additionally, we had to drop the largest observation ( $> 50\,000$  links) from all datasets in order  
 427 to calculate PSIS-LOO for the LSSL model. Taken together, this suggests that the LSSL model is  
 428 insufficiently flexible to accurately reproduce the data.

## 429 **Normal approximation and analytic z-scores**

430 We propose using a normal approximation to the beta-binomial distribution, to calculate analytic  
 431 z-scores. This is based on a well-known similarity between the shape of a normal distribution and  
 432 a binomial distribution. This approximation is considered good whenever the absolute skewness  
 433 is less than 0.3 [55], that is whenever:

$$\frac{1}{\sqrt{S^2 - S + 1}} \left( \sqrt{\frac{1-\mu}{\mu}} - \sqrt{\frac{\mu}{1-\mu}} \right) < 0.3$$

434 The beta-binomial distribution is close to the binomial distribution. The error in approximating  
 435 the former with the latter is on the order of the inverse square of the parameter  $\phi$  [56], which for  
 436 our model is less than 0.0017.

<sup>437</sup> **Acknowledgments**

<sup>438</sup> We thank three anonymous reviewers, as well as Owen Petchey and Andrew Beckerman, for their  
<sup>439</sup> helpful comments and suggestions. Funding was provided by the Institute for data valorization  
<sup>440</sup> (IVADO) and the NSERC CREATE Computational Biodiversity Science and Services (BIOS<sup>2</sup>) train-  
<sup>441</sup> ing program. TP acknowledges funding from the John Evans Leaders' funds from the Canadian  
<sup>442</sup> Foundation for Innovation. This work was made possible thanks to the effort of volunteers devel-  
<sup>443</sup> oping high-quality free and open source software for research.

<sup>444</sup> **References**

- <sup>445</sup> 1. Borrett, S.R., Moody, J., and Edelmann, A. (2014). The rise of Network Ecology: Maps of the  
<sup>446</sup> topic diversity and scientific collaboration. Ecological Modelling 293, 111–127.
- <sup>447</sup> 2. Canard, E., Mouquet, N., Marescot, L., Gaston, K.J., Gravel, D., and Mouillot, D. (2012). Emer-  
<sup>448</sup> gence of Structural Patterns in Neutral Trophic Networks. PLOS ONE 7, e38295. Available at:  
<sup>449</sup> <http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0038295>.
- <sup>450</sup> 3. Delmas, E., Besson, M., Brice, M.-H., Burkle, L.A., Dalla Riva, G.V., Fortin, M.-J., Gravel, D.,  
<sup>451</sup> Guimarães, P.R., Hembry, D.H., and Newman, E.A. *et al.* (2018). Analysing ecological net-  
<sup>452</sup> works of species interactions. Biological Reviews, 112540. Available at: <http://doi.wiley.com/10.1111/brv.12433>.
- <sup>454</sup> 4. McCann, K.S. (2012). Food webs (Princeton: Princeton University Press).
- <sup>455</sup> 5. Thompson, R.M., Brose, U., Dunne, J.A., Hall, R.O., Hladyz, S., Kitching, R.L., Martinez, N.D.,  
<sup>456</sup> Rantala, H., Romanuk, T.N., and Stouffer, D.B. *et al.* (2012). Food webs: reconciling the struc-  
<sup>457</sup> ture and function of biodiversity. Trends in Ecology & Evolution 27, 689–697.
- <sup>458</sup> 6. McCann, K. (2007). Protecting biostructure. Nature 446, 29–29. Available at: <http://www.nature.com/nature/journal/v446/n7131/full/446029a.html>.

- 460 7. Seibold, S., Cadotte, M.W., MacIvor, J.S., Thorn, S., and Müller, J. (2018). The Necessity of Multi-  
461 trophic Approaches in Community Ecology. *Trends in Ecology & Evolution* 33, 754–764. Available  
462 at: [https://www.cell.com/trends/ecology-evolution/abstract/S0169-5347\(18\)30147-2](https://www.cell.com/trends/ecology-evolution/abstract/S0169-5347(18)30147-2).
- 463
- 464 8. Jordano, P. (2016). Sampling networks of ecological interactions. *Functional Ecology* 30, 1883–  
465 1893. Available at: <http://doi.wiley.com/10.1111/1365-2435.12763>.
- 466 9. Jordano, P. (2016). Chasing Ecological Interactions. *PLOS Biol* 14, e1002559. Available at:  
467 <http://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.1002559>.
- 468 10. Pascual, M., and Dunne, J.A. (2006). Ecological Networks: Linking Structure to Dynamics in  
469 Food Webs (Oxford University Press, USA).
- 470 11. Cohen, J.E., and Briand, F. (1984). Trophic links of community food webs. *Proc Natl Acad Sci  
471 U S A* 81, 4105–4109.
- 472 12. Martinez, N.D. (1992). Constant Connectance in Community Food Webs. *The American  
473 Naturalist* 139, 1208–1218. Available at: <http://www.jstor.org/stable/2462337>.
- 474 13. Brose, U., Ostling, A., Harrison, K., and Martinez, N.D. (2004). Unified spatial scaling of  
475 species and their trophic interactions. *Nature* 428, 167–171. Available at: [http://dx.doi.org/10.1038/nature02297](http://dx.doi.<br/>476 org/10.1038/nature02297).
- 477 14. Winemiller, K., Pianka, E., Vitt, L., and Joern, A. (2001). Food Web Laws or Niche Theory?  
478 Six Independent Empirical Tests. *The American Naturalist* 158, 193–199. Available at: [https://www.journals.uchicago.edu/doi/abs/10.1086/321315](https://<br/>479 //www.journals.uchicago.edu/doi/abs/10.1086/321315).
- 480 15. Garlaschelli, D., Caldarelli, G., and Pietronero, L. (2003). Universal scaling relations in food  
481 webs. *Nature* 423, 165–168. Available at: <https://www.nature.com/articles/nature01604>.
- 482 16. Riede, J.O., Rall, B.C., Banasek-Richter, C., Navarrete, S.A., Wieters, E.A., Emmerson, M.C.,  
483 Jacob, U., and Brose, U. (2010). Chapter 3 - Scaling of Food-Web Properties with Diversity and  
484 Complexity Across Ecosystems. In *Advances in Ecological Research Ecological Networks.*, G.  
485 Woodward, ed. (Academic Press), pp. 139–170. Available at: <http://www.sciencedirect>.

- 486       [com/science/article/pii/B9780123813633000034](https://science.org/doi/article/pii/B9780123813633000034).
- 487     17. Poisot, T., and Gravel, D. (2014). When is an ecological network complex? Connectance drives  
488       degree distribution and emerging network properties. PeerJ 2, e251. Available at: [https://  
489       //peerj.com/articles/251](https://peerj.com/articles/251).
- 490     18. Poisot, T., Cirtwill, A.R., Cazelles, K., Gravel, D., Fortin, M.-J., and Stouffer, D.B. (2016). The  
491       structure of probabilistic networks. Methods in Ecology and Evolution 7, 303–312. Available  
492       at: <http://doi.wiley.com/10.1111/2041-210X.12468>.
- 493     19. Petchey, O.L., Beckerman, A.P., Riede, J.O., and Warren, P.H. (2008). Size, foraging, and food  
494       web structure. Proceedings of the National Academy of Sciences 105, 4191–4196. Available at:  
495       <http://www.pnas.org/cgi/doi/10.1073/pnas.0710672105>.
- 496     20. Poisot, T., Stouffer, D.B., and Gravel, D. (2015). Beyond species: why ecological interaction  
497       networks vary through space and time. Oikos 124, 243–251. Available at: [http://doi.wiley.  
498       com/10.1111/oik.01719](http://doi.wiley.com/10.1111/oik.01719).
- 499     21. Poisot, T., Baiser, B., Dunne, J.A., Kéfi, S., Massol, F., Mouquet, N., Romanuk, T.N., Stouffer,  
500       D.B., Wood, S.A., and Gravel, D. (2016). mangal - making ecological network analysis simple.  
501       Ecography 39, 384–390. Available at: <http://doi.wiley.com/10.1111/ecog.00976>.
- 502     22. Poisot, T., Bergeron, G., Cazelles, K., Dallas, T., Gravel, D., Macdonald, A., Mercier, B., Vi-  
503       olet, C., and Vissault, S. (2020). Environmental biases in the study of ecological networks at  
504       the planetary scale. bioRxiv, 2020.01.27.921429. Available at: [https://www.biorxiv.org/  
content/10.1101/2020.01.27.921429v1](https://www.biorxiv.org/<br/>505       content/10.1101/2020.01.27.921429v1).
- 506     23. Vehtari, A., Gelman, A., and Gabry, J. (2017). Practical Bayesian model evaluation using leave-  
507       one-out cross-validation and WAIC. Statistics and Computing 27, 1413–1432. Available at:  
508       <https://doi.org/10.1007/s11222-016-9696-4>.
- 509     24. Fortuna, M.A., and Bascompte, J. (2006). Habitat loss and the structure of plant-animal mutu-  
510       alistic networks: Mutualistic networks and habitat loss. Ecology Letters 9, 281–286. Available  
511       at: <http://doi.wiley.com/10.1111/j.1461-0248.2005.00868.x>.

- 512 25. Bascompte, J., Jordano, P., Melian, C.J., and Olesen, J.M. (2003). The nested assembly of  
513 plant-animal mutualistic networks. *Proceedings of the National Academy of Sciences* *100*,  
514 9383–9387. Available at: <http://www.pnas.org/cgi/doi/10.1073/pnas.1633576100>.
- 515 26. Galiana, N., Lurgi, M., Claramunt-López, B., Fortin, M.-J., Leroux, S., Cazelles, K., Gravel, D.,  
516 and Montoya, J.M. (2018). The spatial scaling of species interaction networks. *Nature Ecology  
517 & Evolution* *2*, 782–790. Available at: <https://www.nature.com/articles/s41559-018-0517-3>.
- 518 27. Holt, R.D., Lawton, J.H., Polis, G.A., and Martinez, N.D. (1999). Trophic Rank and the Species–  
519 Area Relationship. *Ecology* *80*, 1495–1504. Available at: <https://esajournals.onlinelibrary.wiley.com/doi/abs/10.1890/0012-9658%281999%29080%5B1495%3ATRATSA%5D2.0.CO%3B2>.
- 520 28. Baiser, B., Gravel, D., Cirtwill, A.R., Dunne, J.A., Fahimipour, A.K., Gilarranz, L.J., Gro-  
521 chow, J.A., Li, D., Martinez, N.D., and McGrew, A. *et al.* (2019). Ecogeographical rules and  
522 the macroecology of food webs. *Global Ecology and Biogeography* *0*. Available at: <https://onlinelibrary.wiley.com/doi/abs/10.1111/geb.12925>.
- 523 29. Dengler, J. (2009). Which function describes the species–area relationship best? A review and  
524 empirical evaluation. *Journal of Biogeography* *36*, 728–744. Available at: <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-2699.2008.02038.x>.
- 525 30. Williamson, M., Gaston, K.J., and Lonsdale, W.M. (2001). The species–area relationship  
526 does not have an asymptote! *Journal of Biogeography* *28*, 827–830. Available at: <https://onlinelibrary.wiley.com/doi/abs/10.1046/j.1365-2699.2001.00603.x>.
- 527 31. Wood, S.A., Russell, R., Hanson, D., Williams, R.J., and Dunne, J.A. (2015). Effects of spatial  
528 scale of sampling on food web structure. *Ecology and Evolution* *5*, 3769–3782. Available at:  
529 <https://onlinelibrary.wiley.com/doi/abs/10.1002/ece3.1640>.
- 530 32. Law, R. (1979). Optimal Life Histories Under Age-Specific Predation. *The American Naturalist*  
531 *114*, 399–417. Available at: <https://www.jstor.org/stable/2460187>.
- 532 33. Forister, M.L., and Jenkins, S.H. (2017). A Neutral Model for the Evolution of Diet Breadth.  
533 The *American Naturalist* *190*, E40–E54. Available at: <http://www.journals.uchicago.edu/>

- 538        [doi/abs/10.1086/692325](https://doi.org/10.1086/692325).
- 539    34. Kéfi, S., Berlow, E.L., Wieters, E.A., Navarrete, S.A., Petchey, O.L., Wood, S.A., Boit, A., Joppa,  
540           L.N., Lafferty, K.D., and Williams, R.J. *et al.* (2012). More than a meal... integrating non-  
541           feeding interactions into food webs: More than a meal .... *Ecology Letters* 15, 291–300. Available  
542           at: <http://doi.wiley.com/10.1111/j.1461-0248.2011.01732.x>.
- 543    35. Beckerman, A.P., Petchey, O.L., and Warren, P.H. (2006). Foraging biology predicts food web  
544           complexity. *Proceedings of the National Academy of Sciences* 103, 13745–13749. Available at:  
545           <https://www.pnas.org/content/103/37/13745>.
- 546    36. Bartomeus, I., Gravel, D., Tylianakis, J.M., Aizen, M.A., Dickie, I.A., and Bernard-Verdier, M.  
547           (2016). A common framework for identifying linkage rules across different types of interac-  
548           tions. *Functional Ecology* 30, 1894–1903. Available at: <https://besjournals.onlinelibrary.wiley.com/doi/abs/10.1111/1365-2435.12666>.
- 549    37. May, R.M. (1972). Will a Large Complex System be Stable? *Nature* 238, 413–414. Available at:  
550           <http://www.nature.com/doifinder/10.1038/238413a0>.
- 551    38. Allesina, S., and Tang, S. (2012). Stability criteria for complex ecosystems. *Nature* 483, 205–  
552           208. Available at: <https://www.nature.com/articles/nature10832>.
- 553    39. Allesina, S., and Tang, S. (2015). The stability–complexity relationship at age 40: a random  
554           matrix perspective. *Population Ecology* 57, 63–75. Available at: <https://link.springer.com/article/10.1007/s10144-014-0471-0>.
- 555    40. Borrelli, J.J., Allesina, S., Amarasekare, P., Arditi, R., Chase, I., Damuth, J., Holt, R.D., Logofet,  
556           D.O., Novak, M., and Rohr, R.P. *et al.* (2015). Selection on stability across ecological scales.  
557           *Trends in Ecology & Evolution* 30, 417–425. Available at: <http://www.sciencedirect.com/science/article/pii/S0169534715001238>.
- 558    41. Williams, R.J., and Martinez, N.D. (2000). Simple rules yield complex food webs. *Nature* 404,  
559           180–183. Available at: <http://www.nature.com/articles/35004572>.
- 560

- 563 42. Cohen, J.E., and Newman, C.M. (1985). A stochastic theory of community food webs I. Models  
564 and aggregated data. *Proceedings of the Royal Society of London. Series B. Biological Sci-  
565 ences* 224, 421–448. Available at: <https://royalsocietypublishing.org/doi/10.1098/rspb.1985.0042>.
- 566
- 567 43. Allesina, S., Alonso, D., and Pascual, M. (2008). A General Model for Food Web Structure.  
568 *Science* 320, 658–661. Available at: <https://www.sciencemag.org/lookup/doi/10.1126/science.1156269>.
- 569
- 570 44. Cattin, M.-F., Bersier, L.-F., Banašek-Richter, C., Baltensperger, R., and Gabriel, J.-P. (2004).  
571 Phylogenetic constraints and adaptation explain food-web structure. *Nature* 427, 835–839.  
572 Available at: <http://www.nature.com/nature/journal/v427/n6977/full/nature02327.html>.
- 573
- 574 45. Kozubowski, T.J., Panorska, A.K., and Forister, M.L. (2015). A discrete truncated Pareto distri-  
575 bution. *Statistical Methodology* 26, 135–150. Available at: <http://www.sciencedirect.com/science/article/pii/S1572312715000313>.
- 576
- 577 46. Dunne, J.A., Williams, R.J., and Martinez, N.D. (2002). Food-web structure and network the-  
578 ory: The role of connectance and size. *Proceedings of the National Academy of Sciences* 99,  
579 12917–12922. Available at: <http://www.pnas.org/cgi/doi/10.1073/pnas.192407699>.
- 580
- 581 47. Montoya, J.M., Pimm, S.L., and Solé, R.V. (2006). Ecological networks and their fragility.  
Nature 442, 259–264. Available at: <http://www.nature.com/articles/nature04927>.
- 582
- 583 48. Vieira, M.C., and Almeida-Neto, M. (2015). A simple stochastic model for complex coextinc-  
584 tions in mutualistic networks: robustness decreases with connectance. *Ecology Letters* 18,  
144–152. Available at: <https://onlinelibrary.wiley.com/doi/abs/10.1111/ele.12394>.
- 585
- 586 49. Smith-Ramesh, L.M., Moore, A.C., and Schmitz, O.J. (2016). Global synthesis suggests that  
food web connectance correlates to invasion resistance. *Global Change Biology*, n/a–n/a. Avail-  
587 able at: <http://onlinelibrary.wiley.com/doi/10.1111/gcb.13460/abstract>.

- 588 50. Baiser, B., Russell, G.J., and Lockwood, J.L. (2010). Connectance determines invasion success  
589 via trophic interactions in model food webs. *Oikos* *119*, 1970–1976. Available at: [http://  
590 onlinelibrary.wiley.com/doi/10.1111/j.1600-0706.2010.18557.x/abstract](http://onlinelibrary.wiley.com/doi/10.1111/j.1600-0706.2010.18557.x/abstract).
- 591 51. Hobbs, N.T., and Hooten, M.B. (2015). Bayesian models: a statistical primer for ecologists  
592 (Princeton, New Jersey: Princeton University Press).
- 593 52. Gabry, J., Simpson, D., Vehtari, A., Betancourt, M., and Gelman, A. (2019). Visualization in  
594 Bayesian workflow. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* *182*,  
595 389–402. Available at: <http://arxiv.org/abs/1709.01449>.
- 596 53. Carpenter, B., Gelman, A., Hoffman, M.D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M.,  
597 Guo, J., Li, P., and Riddell, A. (2017). Stan: A Probabilistic Programming Language. *Journal*  
598 *of Statistical Software* *76*, 1–32. Available at: [https://www.jstatsoft.org/index.php/jss/  
599 article/view/v076i01](https://www.jstatsoft.org/index.php/jss/article/view/v076i01).
- 600 54. McElreath, R. (2020). *Statistical Rethinking: A Bayesian Course with Examples in R and STAN*  
601 (CRC Press).
- 602 55. Box, G.E.P., Hunter, J.S., and Hunter, W.G. (2005). *Statistics for experimenters: design, inno-*  
603 *vation, and discovery* (Wiley-Interscience).
- 604 56. Teerapabolarn, K. (2014). An improved binomial approximation for the beta binomial dis-  
605 tribution. *International Journal of Pure and Applied Mathematics* *97*. Available at: [http:  
606 //www.ijpam.eu/contents/2014-97-4/10/](http://www.ijpam.eu/contents/2014-97-4/10/).