

# The biological interpretation of probabilistic food webs

Francis Banville<sup>1,2,3,‡</sup> Tanya Strydom<sup>1,3,‡</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

‡ These authors contributed equally to the work

## Correspondance to:

Francis Banville — francis.banville@umontreal.ca

This work is released by its authors under a CC-BY 4.0 license



Last revision: *February 13, 2023*

Community ecologists are increasingly thinking probabilistically when it comes to food webs and other ecological networks. Assuredly, the benefits of representing ecological interactions as probabilistic events (e.g., how likely are species to interact?) instead of binary objects (e.g., do species interact?) are numerous, from a better assessment of the spatiotemporal variation of trophic interactions to an increase capacity to reconstruct networks from sparse data. However, probabilities need to be used with caution when working with species interactions. Indeed, depending on the system at hand and the method used to build probabilistic networks, probabilities can have different interpretations that imply different ways to manipulate them. At the core of these differences lie the distinction between assessing the likelihood that two groups of individuals *can* interact and the likelihood that they *will* interact. This impacts the spatial, temporal, and taxonomic scaling of interaction probabilities, thus enlightening the need to properly define them in their ecological context. Published data on probabilistic species interactions are poorly documented in the literature, which impedes our ability to use them appropriately. With these challenges in mind, we propose a general approach to thinking about probabilities in regards to ecological interactions, with a strong focus on food webs, and call for better definitions and conceptualizations of probabilistic ecological networks, both at the local and regional scales.

# 1 Introduction

2 Cataloging species interactions across space is a gargantuan task. At the core of this challenge lies the  
3 spatiotemporal variability of ecological networks (Poisot *et al.* 2012, 2015), which makes documenting the  
4 location and timing of interactions difficult. Indeed, it is not sufficient to know that two species have the  
5 biological capacity to interact to infer the realization of their interaction at a specific time and space  
6 (Dunne 2006). Taking food webs as an example, a predator species and its potential prey must first  
7 co-occur in order for a trophic interaction to take place (Blanchet *et al.* 2020). They must then encounter,  
8 which is conditional on their relative abundances in the ecosystem and the matching of their phenology  
9 (Poisot *et al.* 2015). Finally, the interaction occurs only if the predators have a desire to consume their prey  
10 and are able to capture and ingest them (Pulliam 1974). Environmental (e.g. temperature and presence of  
11 shelters) and biological (e.g. physiological state of both species and availability of other prey species)  
12 factors contribute to this variability by impacting species co-occurrence (Araujo *et al.* 2011) and the  
13 realization of their interactions (Poisot *et al.* 2015). In this context, the development of computational  
14 methods in ecology can help alleviate the colossal sampling efforts required to document species  
15 interactions across time and space (Strydom *et al.* 2021). Having a better portrait of species interactions  
16 and the emerging structure of their food webs is important since it lays the groundwork for understanding  
17 the functioning, dynamics, and resilience of ecosystems worldwide (e.g., Proulx *et al.* 2005; Pascual *et al.*  
18 2006; Delmas *et al.* 2019).

19 The recognition of the intrinsic variability of species interactions and the emergence of numerical  
20 methods have led ecologists to rethink their representation of ecological networks, slowly moving from a  
21 binary to a probabilistic view of species interactions (Poisot *et al.* 2016). This has several benefits. For  
22 example, probabilities represent the limit of our knowledge about species interactions and can inform us  
23 about the expected number of interactions and emerging network properties despite this limited  
24 knowledge (Poisot *et al.* 2016). They are also very helpful in predicting the spatial distribution of species  
25 within networks (Cazelles *et al.* 2016) and the temporal variability of interactions (Poisot *et al.* 2015),  
26 generating new ecological data (e.g., Strydom *et al.* 2022), and identifying priority sampling locations of  
27 species interactions (see Andrade-Pacheco *et al.* 2020 for an ecological example of a sampling optimization  
28 problem). Moreover, the high rate of false negatives in ecological network data, resulting from the  
29 difficulty of witnessing interactions between rare species, makes it hard to interpret non-observations of

30 species interactions ecologically (Catchen *et al.* 2023). Using probabilities instead of yes-no interactions  
31 accounts for these observation errors; in that case, only forbidden interactions (Jordano *et al.* 2003; Olesen  
32 *et al.* 2010) would have a probability value of zero (but see Gonzalez-Varo & Traveset 2016). Many  
33 measures have been developed to describe the structure (Poisot *et al.* 2016) and diversity (Ohlmann *et al.*  
34 2019; Godsoe *et al.* 2022) of probabilistic interactions, which shows the potential of this framework in the  
35 study of a variety of ecological phenomena.

36 However, representing species interactions probabilistically can also be challenging. Beyond  
37 methodological difficulties in estimating these numbers, there are important conceptual challenges in  
38 defining what we mean by “probability of interactions.” To the best of our knowledge, because the  
39 building blocks of this mathematical representation of food webs are still being laid, there is no clear  
40 definition found in the literature or data standard when it comes to publishing data on probabilistic  
41 interactions (see Salim *et al.* 2022 for a discussion on data standardization for mutualistic networks). This  
42 is worrisome, since working with probabilistic species interactions without clear guidelines could be  
43 misleading as much for field ecologists as for computational ecologists who use and generate these data.  
44 In this contribution, we outline different ways to define and interpret interactions probabilities in network  
45 ecology and propose an approach to thinking about them. These definitions mostly depend on the study  
46 system (e.g. local network or metaweb) and on the method used to generate them. We show that different  
47 definitions can have different ecological implications, especially regarding spatial, temporal, and  
48 taxonomic scaling. Although we will focus on food webs, our observations and advice can be applied to all  
49 types of ecological networks, from plant-pollinator to host-parasite networks. Indeed, all ecological  
50 networks, whether they are unipartite or bipartite, share fundamental commonalities in their biological  
51 conceptualization and mathematical representation that support these comparisons (i.e., they all describe  
52 groups of individuals interacting with each other). Regardless of the study system, we argue that  
53 probabilities should be better documented, defined mathematically, and used with caution when  
54 describing species interactions.

## 55 Definitions and interpretations

### 56 Food-web representations

57 The basic unit of food webs and other ecological networks are individuals that interact with each others  
58 (e.g., by predation; Elton 2001), forming individual-based networks. The aggregation of these individuals  
59 into more or less homogeneous groups (e.g., populations, species, trophic species, families) allows us to  
60 represent networks at broader scales, which impacts the properties and behaviour of these systems  
61 (Guimarães 2020). A network's nodes can thus designate distinct levels of organization, whereas the edges  
62 linking these nodes can describe a variety of interaction measures. When using a Boolean (yes-no)  
63 representation of biotic interactions, the observation that one individual from group (or node)  $i$  interacts  
64 with another individual from group  $j$  is enough to set the interaction  $A_{i,j}$  to 1. This simplified  
65 representation of food webs is a highly valuable source of ecological information (Pascual *et al.* 2006) even  
66 though it overlooks important factors regarding interaction strengths. These, in turn, can be represented  
67 using weighted interactions, which better describe the energy flows, demographic impacts or frequencies  
68 of interactions between nodes (Berlow *et al.* 2004; Borrett & Scharler 2019), with  $A_{i,j} \in \mathbb{N}$  or  $\mathbb{R}$  depending  
69 on the measure. For example, they can be used to estimate the average number of prey individuals  
70 consumed by the predators in a given time period (e.g., the average number of fish in the stomach of a  
71 piscivorous species). Interaction strengths can also be used as good estimators of the parameters  
72 describing species interactions in a Lotka-Volterra model (e.g., Emmerson & Raffaelli 2004). This extra  
73 amount of ecological information typically comes at a cost of greater sampling effort or data requirement  
74 in predictive models (Strydom *et al.* 2021), which can lead to high uncertainties when building these types  
75 of networks. Therefore, important methodological and conceptual decisions must be made when  
76 sampling and building food webs.

77 The uncertainty and spatiotemporal variability of both types of trophic interactions (Boolean and  
78 weighted) can be represented probabilistically. On one hand, Boolean interactions follow a Bernoulli  
79 distribution  $A_{i,j} \sim \text{Bernoulli}(p)$ , with  $p$  being the probability of interactions. The only two possible  
80 outcomes are the presence ( $A_{i,j} = 1$ ) or absence ( $A_{i,j} = 0$ ) of an interaction between the two nodes.  
81 Weighted interactions, on the other hand, can follow various probability distributions depending on the  
82 measure used. In this case, the event's outcome is the value of interaction strength. For instance, weights  
83 can follow a Poisson distribution  $A_{i,j} \sim \text{Poisson}(\lambda)$  when predicting frequencies of interactions between

84 pairs of nodes, with  $\lambda$  being the expected rate of interaction. Note that weighted interactions can be  
85 converted to probabilistic interactions by normalizing. The definition and interpretation of parameters  
86 like  $p$  and  $\lambda$  are inextricably linked to environmental and biological factors such as species relative  
87 abundance, traits, area, and time, depending on the type of interaction. Because Boolean species  
88 interactions are much more documented in the literature, our primary focus in this contribution will be on  
89 addressing the challenges in defining and interpreting  $p$  for pairwise species interactions.

90 The first aspect to take into consideration when estimating or interpreting probabilities of interactions is  
91 knowing if they describe the likelihood of potential or realized interactions. A potential interaction is  
92 defined as the biological capacity of two species to interact (i.e., the probability that they *can* interact)  
93 whereas a realized interaction refers to the materialization or observation of this interaction in a delineated  
94 space and time period (i.e., the probability that they interact). Here, we will use the terms *metaweb* to  
95 designate networks of potential interactions and *local networks* for those of realized interactions. Frequent  
96 confusion arises among ecologists over the use of these two terms, especially in a probabilistic context.  
97 Indeed, in many studies of probabilistic ecological networks, it remains unclear when authors describe  
98 potential or realized interactions, or when so-called probabilities are actually *interaction scores*. Likewise,  
99 probabilistic potential interactions are often used as realized interactions (and conversely), even when the  
100 type of interaction is clearly indicated. We believe that a better understanding of these differences and  
101 concepts would alleviate interpretation errors and help ecologists use these numbers more appropriately.

## 102 **Probabilistic metawebs**

103 Metawebs are networks of potential interactions, representing the probability that two taxa can interact  
104 regardless of biological plasticity, environmental variability or co-occurrence. Instead of describing  
105 stochastic biological processes occurring in nature, probabilistic potential interactions can be thought of as  
106 a measure of imperfect knowledge about the capacity of two taxa to interact. They are the network  
107 analogue to the species pool, where local networks originate from a subset of both species (nodes) and  
108 interactions (edges). For this reason, they have been initially designed for broad spatial, temporal, and  
109 taxonomic scales (e.g, species food webs at the continental scale). However, in the next section, we argue  
110 that this concept can also be used at smaller scales, with similar ecological meaning.

111 We can express the probability that two taxa  $i$  and  $j$  can interact in a metaweb  $M$  as

$$P_M(i \rightarrow j), \quad (1)$$

which is context independent. In other words, the probability that two species can interact is not contingent on location, time, and environmental factors. Nevertheless, one aspect of a metaweb that could be conditional on these factors is the list of species (or taxa) it is built from when assembled for a specific region.

Starting from a selected set of species, metawebs can be built using different data sources, including literature review, field work, and predictive models (e.g., the metaweb of Canadian mammals inferred by Strydom *et al.* 2022). Every pair of species that has been observed to interact at least once can be given a probability of interaction of 1; we know that they *can* interact. This means that rare interactions can technically be given high probabilities in the metaweb. Unobserved interactions, on the other hand, are given lower probabilities, going as low as 0 for forbidden links. Two important nuances must however be made here. Because of observation errors due to taxonomic misidentifications and ecological misinterpretations (e.g., due to cryptic species and interactions; Pringle & Hutchinson 2020), many observations of interactions are actually false positives. Similarly, forbidden interactions can be false negatives if e.g. they have been assessed for specific phenotypes, locations or time. Implementing a Bayesian framework, which updates prior probabilities of interactions with empirical data, could lessen these errors.

## Probabilistic local networks

As opposed to metawebs, probabilistic local food webs represent the likelihood that two species will interact at a specific location and within a given time period; in other words, they are context dependant. They could also represent the likelihood of observing these interactions within a given area and time. To be specific, space is defined here as the geographic coordinates  $(x, y)$  of the spatial boundaries delineating the system, whereas time is the time interval  $t$  during which interactions were sampled or for which they were predicted. We want to point out that they are not single values, but rather continued dimensions that could be outlined differently depending on the study system. Regardless of how they were defined, they always delineate a specific area  $A$  and duration  $t$ . These could refer to the sampled area and duration or to the targeted location and time period.

Many factors could be taken into consideration when estimating the probability that a predator species  $i$  interacts with a given prey species  $j$  locally. One of the most important is species co-occurrence  $C$ , which is a Boolean describing if both species can be found at location and time  $(x, y, t)$ . Surely, the probability that the interaction is realized must be 0 when species do not co-occur ( $C = 0$ ). Interaction probabilities can also be conditional on other biological and environmental variables, such as temperature, precipitation, presence of shelters, phenotypic plasticity, phenology, and presence of other interacting species in the network. These conditions can affect species traits, which greatly impact the capacity of species to interact (Poisot *et al.* 2015). Similarly, species relative abundance is another important predictor of the probability of interaction, because it impacts the probability that species will randomly encounter (Canard *et al.* 2012; Canard *et al.* 2014; Poisot *et al.* 2015). Here, we will use the variable  $\Omega$  as a substitute for the biological and ecological context in which interaction probabilities were estimated, including the presence of higher-order interactions. This gives us the following equation for the probability of realized interaction between species (or taxa)  $i$  and  $j$  in a local network  $N$ :

$$P_N(i \rightarrow j | A, t, C, \Omega), \quad (2)$$

which can be read as the probability of local interaction between the two species in an area  $A$  and time interval  $t$ , given their co-occurrence  $C$  and specific environmental and biological conditions  $\Omega$ . These conditions do not systematically need to be specified for all studies. However, when they are, they should be made explicit in the metadata.

Multiple difficulties of interpretation arise when the conditions are not clearly specified, which we found is often the case in the literature. For example, if  $P_N(i \rightarrow j | C = 1)$  represents the probability that two co-occurring species interact (i.e., the edge's probability value),  $P_N(i \rightarrow j)$  denotes instead the probability of interaction without knowing if they co-occur (i.e., the product of the nodes and edge's probability values). For practical reasons, probabilistic ecological networks are generally represented as matrices of probabilities (i.e., matrices of edges without node values), whose elements are thus hard to interpret without clear indications about  $C$ . Overall, when probabilities of interactions are estimated using specific values of  $A$ ,  $t$ ,  $C$ , and  $\Omega$ , ecologists should make them explicit in their metadata, preferably using mathematical equations to avoid any ambiguity. Below, we will see examples of why this matters when it comes to spatial, temporal, and taxonomic scaling of biotic interactions.



165 [Table 1 about here]. Articles using probabilistic interactions and the definitions and variables they  
166 considered.

## 167 **Applications of probabilistic interactions data**

### 168 **Inferring probabilistic local food webs from metawebs**

169 Even though the spatiotemporal variability of interactions is not considered in metawebs, they can still be  
170 useful to reconstruct local networks of realized interactions. Indeed, local networks are formed from  
171 subsets of their metaweb (called subnetworks), which are obtained by selecting a subset of both species  
172 and interactions (Dunne 2006). Because a community's composition is arguably easier to sample (or  
173 predict) than its interactions, the biggest challenge is to sample links from the metaweb. This becomes a  
174 conceptual issue when we consider how potential and realized interactions differ. Despite these concerns,  
175 metawebs remain an important source of ecological information that can be leveraged for realistically  
176 predicting spatially explicit networks. First, metawebs set the upper limit of species interactions (McLeod  
177 *et al.* 2021), i.e. the probability that two species interact at a specific location is always lower or equal to  
178 the probability of their potential interaction:

$$P_N(i \rightarrow j|A, t, C, \Omega) \leq P_M(i \rightarrow j). \quad (3)$$

179 Therefore, inferring local networks from their metaweb keeping the same values of interaction probability  
180 would generate systematic biases in the prediction. In that case, these networks would instead be called  
181 *spatially explicit* or *local* metawebs (i.e., smaller-scale networks of potential interactions). Second, the  
182 structure of local networks is constrained by the one of their metaweb (Saravia *et al.* 2022). This suggests  
183 that a metaweb not only constrains the pairwise interactions of its corresponding local networks, but also  
184 their emerging properties. Inferring the structure of local networks from the metaweb could thus help  
185 estimate more realistically the likelihood that potential interactions are realized and observed locally  
186 (Strydom *et al.* 2021).

187 [Figure 1 about here]. Empirical example of the association between the number of interactions in  
188 realized local food webs and the number of interactions in the corresponding subnetworks of their  
189 regional metaweb. We should expect the association to be linear below the 1:1 line, illustrating eq. 3.

## 190 Sampling random draws from probabilistic food webs

191 Another conceptual challenge encountered when using probabilistic food webs is the prediction of  
192 Boolean networks across space. Lets take  $n \times n$  grid cells each representing a probabilistic food web. If  
193 they contain potential interactions, a single random trial must be conducted for each pairwise interaction  
194 across the region (i.e., we should have only one random realization of the regional metaweb). On the  
195 contrary, if they represent probabilities of realized interactions, each food web must be independently  
196 sampled (i.e.,  $n^2$  independent random draws). This has direct implications on the spatial scaling of  
197 interactions. For example, let  $N_1$  and  $N_2$  be networks of area  $< A_0$  within a bigger area  $A_0$  and disjoint  
198 from each other, such as  $N_1$  and  $N_2$  form  $N_0$  (think of two contiguous cells that together delineate  $N_0$ ). All  
199 other things being equal, we should expect the probability that  $i$  and  $j$  interacts in  $A_0$  to be  
200  $P_{N_0}(i \rightarrow j) = 1 - (1 - P_{N_1}(i \rightarrow j)) \times (1 - P_{N_2}(i \rightarrow j))$  if  $N_1$  and  $N_2$  are independently sampled. This also  
201 implies that we should expect interactions to be realized in a certain number of local networks depending  
202 on the probability value, which is not the case with metawebs. Note that spatial auto-correlation and the  
203 concept of meta-network (i.e., networks of networks) could invalidate the statistical assumption of  
204 independence. Nevertheless, the fundamental difference in sampling metawebs and local networks stands  
205 even when considering these factors. This difference in sampling further sheds light on the importance of  
206 clearly defining interaction probabilities. What we consider as a *Bernoulli trial*, when randomly drawing  
207 deterministic networks from probabilistic food webs, depends on our biological interpretation of these  
208 probabilities.

## 209 Describing the spatial and temporal scaling of probabilistic interactions

210 Metawebs and local networks intrinsically differ in their relation to scale. On one hand, as mentioned  
211 above, probabilistic metawebs are context independent, i.e., probabilistic pairwise interactions do not scale  
212 with space and time because they depend solely on the biological capacity of the two taxa to interact. This  
213 implies that the estimated likelihood that two species can potentially interact should be the same among  
214 all metawebs in which they are present. In practice, this is rarely the case because ecologists use different  
215 methods and data to estimate these probabilities of interactions (e.g., different sampling area and time  
216 period). However, in the case where local metawebs  $M_{x,y}$  are subsampled from their regional counterpart  
217  $M_0$ , we should expect edge values to be identical among all networks, regardless of their spatial scale,

218 i.e.  $P_{M_{x,y}}(i \rightarrow j) = P_{M_0}(i \rightarrow j)$ . On the other hand, local probabilistic networks are indissociable from their  
 219 spatial and temporal contexts because there are more opportunities of interactions (e.g., more individuals,  
 220 more trait variations, more chance of encounter) in a larger area and longer time period (McLeod *et al.*  
 221 2020). Let  $N_0$  be a local probabilistic food web delineated in an area  $A_0$  and  $N_1$  a network of area  $A_1 < A_0$   
 222 within  $A_0$ . We should expect the probability that  $i$  and  $j$  interacts to scale spatially,  
 223 i.e.  $P_{N_1}(i \rightarrow j | A < A_0) \leq P_{N_0}(i \rightarrow j | A = A_0)$ . Similarly, the temporal scaling of probabilistic local food  
 224 webs could be manifested through the effect of sampling effort on the observation of interactions (Jordano  
 225 2016; McLeod *et al.* 2021) or of time itself on their realization (Poisot *et al.* 2012). There are many  
 226 network-area relationships (e.g., Wood *et al.* 2015; Galiana *et al.* 2018) and interaction accumulation  
 227 curves (e.g. Jordano 2016) explored in the literature. These could inspire the development and testing of  
 228 different equations describing the spatiotemporal scaling of probabilistic pairwise interactions in local  
 229 networks, which are over the scope of this manuscript.

230 [Figure 2 about here]. Conceptual figure showing (1) the spatiotemporal scaling of probabilistic metawebs  
 231 and local food webs and (2) the spatial sampling of metawebs and local food webs into Boolean networks.

## 232 **Making probabilistic interactions spatiotemporally explicit**

233 The variability of species interactions spurred the development of methods aiming at predicting ecological  
 234 networks at fine spatial and temporal scales. For example, Bohan *et al.* (2017) proposed a framework to  
 235 reconstruct networks in real time using continuous biomonitoring. Here, we will build on these studies by  
 236 proposing a simple model to make probabilistic local networks spatiotemporally explicit. These types of  
 237 models could prove useful when inferring food webs across time and space from sparse data. However,  
 238 they are not suitable for metawebs, which are static objects.

239 One way that probabilistic food webs can be made spatiotemporally explicit is by modelling interactions  
 240 between co-occurring species as a Poisson process with rate  $\lambda$ . Specifically, if the total observation time for  
 241 a location is  $t_0$ , the probability that two co-occurring species  $i$  and  $j$  will interact during this time period is  
 242  $P_N(i \rightarrow j | C_{i,j} = 1, t = t_0) = 1 - e^{-\lambda t_0}$ , which approaches 1 when  $t \rightarrow \infty$ . The value of the parameter  $\lambda$   
 243 could be estimated using prior data on interaction strengths, if available. Additionally, we can estimate the  
 244 probability of co-occurrence at location  $(x, y)$  with  $P_{x,y}(C_{i,j} = 1) = P_{x,y}(i)P_{x,y}(j)\gamma$ , where  $P_{x,y}(i)$  and  
 245  $P_{x,y}(j)$  are respectively the probabilities of occurrence of species  $i$  and  $j$  and  $\gamma$  is the strength of association

between occurrence and co-occurrence, as defined in Cazelles *et al.* (2016). Note that in empirical networks,  $\gamma$  is typically  $> 1$  (Catchen *et al.* 2023). The observation of this interaction would thus follow a Bernoulli distribution with parameter  $p = p_A(x, y)p_B(x, y)\gamma(1 - e^{-\lambda t_0})$ . This simple model could be customized in many ways, e.g. by linking  $\lambda$  with given environmental variables or by adding in observation errors (i.e., probability of false negatives and false positives; Catchen *et al.* (2023)).

## Exploring different taxonomic resolutions

The properties of ecological networks depend on their level of organization (Guimarães 2020). Indeed, at different taxonomic scales, different behaviours and dynamics can be observed and distinct ecological questions can be answered (e.g., exploring evolutionary dynamics at broad taxonomic scales). Because of these reasons, it could be important to analyse the same network at different taxonomic scales. However, we want to emphasize here that many networks do not have an homogenous level of organisation (Vázquez *et al.* 2022). Indeed, different nodes within the same network can be represented at different taxonomic scales (e.g., a network composed of species and trophic species). This becomes important when we consider that the biological interpretation of interaction probabilities depends on the nodes' resolution. For example, in individual-based networks, the probability that two individuals interact could represent the degree of belief that one will actually consume the other. In species-based networks, the probability that two species interact could rather represent the degree of belief that *at least* one individual from the predator species will eat *at least* another individual from the prey species. This distinction in interpretation impacts the way probability values change with taxonomic scale.

There are a lot of similarities between taxonomic and spatiotemporal scaling of probabilistic interactions. Fundamentally, these types of scaling are just different ways to aggregate individuals into broader nodes, either spatially, temporally, or taxonomically. However, there are also important differences between them. First, in metawebs, if we know that two species have the capacity to interact, we can infer that their respective genus should also be able to interact (i.e., there should be at least two individuals within these genus that can interact). On the contrary, knowing that two genus can interact does not mean that all pairwise combinations of species within these genus can also interact among themselves. This observation also applies to local networks. When it comes to probabilistic networks, interaction probabilities at broader taxonomic scales can be directly obtained from probabilities at finer scales when aggregating nodes. For example, if we have in a network  $n_A$  species from genus  $A$  and  $n_B$  species from genus  $B$ , we can

275 calculate the probability that the two genus interact as  $P_N(A \rightarrow B) = 1 - \prod_{i=1}^{n_A} \prod_{j=1}^{n_B} (1 - P_N(A_i \rightarrow B_j))$ ,  
 276 where  $A_i$  and  $B_j$  are the species of the corresponding genus. However, more sophisticated models need to  
 277 be used when building probabilistic networks at smaller taxonomic resolutions (e.g., when building a  
 278 species-level network from a genus-level network). One could, for example, estimate the probabilities of  
 279 all pairwise species interactions by using a Beta distribution parametrised by the broader-scale network.  
 280 [Figure 3 about here]. Conceptual figure of how a scale up of the nodes from an individual to a population  
 281 to any higher taxonomic group change our interpretation of the probability of interaction.

## 282 Conclusion

283 The emergence of probabilistic thinking in network ecology has paved the way to a better assessment of  
 284 the spatiotemporal variability and uncertainty of biotic interactions. However, measuring probabilities  
 285 empirically can be strenuous given the difficulties of deciphering species and interactions (Pringle &  
 286 Hutchinson 2020). In this context, the development of computational methods makes it possible to  
 287 estimate interaction probabilities at large scales, which in turn can pinpoint where we should go to  
 288 optimise our sampling effort for better resolving local food webs.

289 In this contribution, we showed that network metadata are perhaps as important as interaction data  
 290 themselves when it comes to interpreting probabilistic food webs in ecological terms. First, the type of  
 291 probabilistic interaction (potential or realized) impacts the importance of scale, with interactions in  
 292 metawebs being scale independent, both spatially and temporally. Second, the conditions in which local  
 293 networks were estimated (e.g., area, time, biological and environmental factors) and the attributes of the  
 294 interacting species that were considered (e.g., species co-occurrence) are essential contextual factors that  
 295 impact the mathematical representation of probabilities and their resulting behaviour. Third, the  
 296 biological interpretation of probabilities changes with the level of organization of the network due to the  
 297 aggregation of individuals into broader groups. All these information should be available as clear  
 298 metadata so that ecologists can use probabilistic network data appropriately.

299 Moreover, many statistical models in ecology that yield accurate predictions of biotic interactions are black  
 300 boxes difficult to interpret. Ecologists should be careful before using the output of these models as  
 301 probabilistic objects, since there is often a thin line between a real probability and a non-probabilistic  
 302 predictive number (or score). Probabilities are numbers between 0 and 1 that sum to 1 and either represent

the expected frequency of a phenomenon or the degree of belief that it will be realized. Non-probabilistic scores, which are more akin to interaction weights, have different mathematical properties, which impacts how we should handle these numbers in a spatially or temporally explicit context. Therefore, researchers should use their expertise to assess if their interaction data are actually probabilities or scores. This should also be added to the metadata before sharing them, as well as the methods used to build the networks. Better metadata documentation would allow researchers to use and manipulate probabilistic ecological interactions according to how they were actually defined and obtained. This would support better scientific practices, in particular when these data are used for ecological prediction and forecasting. For instance, we showed that building a rigorous workflow to predict local networks from a probabilistic metaweb requires a good understanding of the data at hand. Similarly, explicitly stating the context in which probabilistic data were estimated would help using forecasting food-web models more rigorously under specific climate change and habitat use scenarios. Regardless of the method and application, fostering a better foundation for probabilistic reasonings in network ecology, from the very nature of probabilities and biotic interactions, is essential.

## Acknowledgement

We acknowledge that this study was conducted on land within the traditional unceded territory of the Saint Lawrence Iroquoian, Anishinabewaki, Mohawk, Huron-Wendat, and Omàmiwininiwak nations. This work was supported by the Institute for Data Valorisation (IVADO) and the Natural Sciences and Engineering Research Council of Canada (NSERC) Collaborative Research and Training Experience (CREATE) program, through the Computational Biodiversity Science and Services (BIOS<sup>2</sup>) program. A special thank to all members of the Black Holes and Revelations working group (organized by BIOS<sup>2</sup>) for their insightful discussions and valuable feedbacks on this manuscript.

## References

Andrade-Pacheco, R., Rerolle, F., Lemoine, J., Hernandez, L., Meïté, A., Juziwele, L., *et al.* (2020). Finding hotspots: Development of an adaptive spatial sampling approach. *Scientific Reports*, 10, 10939.

328 Araujo, M.B., Rozenfeld, A., Rahbek, C. & Marquet, P.A. (2011). Using species co-occurrence networks to  
 329 assess the impacts of climate change. *Ecography*, 34, 897–908.

330 Berlow, E.L., Neutel, A.-M., Cohen, J.E., De Ruiter, P.C., Ebenman, B., Emmerson, M., *et al.* (2004).  
 331 Interaction strengths in food webs: Issues and opportunities. *Journal of Animal Ecology*, 73, 585–598.

332 Blanchet, F.G., Cazelles, K. & Gravel, D. (2020). Co-occurrence is not evidence of ecological interactions.  
 333 *Ecology Letters*, 23, 1050–1063.

334 Bohan, D.A., Vacher, C., Tamaddon-Nezhad, A., Raybould, A., Dumbrell, A.J. & Woodward, G. (2017).  
 335 Next-Generation Global Biomonitoring: Large-scale, Automated Reconstruction of Ecological  
 336 Networks. *Trends in Ecology & Evolution*, 32, 477–487.

337 Borrett, S.R. & Scharler, U.M. (2019). Walk partitions of flow in Ecological Network Analysis: Review and  
 338 synthesis of methods and indicators. *Ecological Indicators*, 106, 105451.

339 Canard, E.F., Mouquet, N., Mouillot, D., Stanko, M., Miklisova, D. & Gravel, D. (2014). Empirical  
 340 Evaluation of Neutral Interactions in Host-Parasite Networks. *The American Naturalist*, 183, 468–479.

341 Canard, E., Mouquet, N., Marescot, L., Gaston, K.J., Gravel, D. & Mouillot, D. (2012). Emergence of  
 342 Structural Patterns in Neutral Trophic Networks. *PLOS ONE*, 7, e38295.

343 Catchen, M.D., Poisot, T., Pollock, L.J. & Gonzalez, A. (2023). The missing link: Discerning true from false  
 344 negatives when sampling species interaction networks.

345 Cazelles, K., Araujo, M.B., Mouquet, N. & Gravel, D. (2016). A theory for species co-occurrence in  
 346 interaction networks. *Theoretical Ecology*, 9, 39–48.

347 Delmas, E., Besson, M., Brice, M.-H., Burkle, L.A., Dalla Riva, G.V., Fortin, M.-J., *et al.* (2019). Analysing  
 348 ecological networks of species interactions. *Biological Reviews*, 94, 16–36.

349 Dunne, J.A. (2006). The Network Structure of Food Webs. In: *Ecological networks: Linking structure and*  
 350 *dynamics* (eds. Dunne, J.A. & Pascual, M.). Oxford University Press, pp. 27–86.

351 Elton, C.S. (2001). *Animal Ecology*. University of Chicago Press, Chicago, IL.

352 Emmerson, M.C. & Raffaelli, D. (2004). Predator-prey body size, interaction strength and the stability of a  
 353 real food web. *Journal of Animal Ecology*, 73, 399–409.

354 Galiana, N., Lurgi, M., Claramunt-López, B., Fortin, M.-J., Leroux, S., Cazelles, K., *et al.* (2018). The  
 355 spatial scaling of species interaction networks. *Nature Ecology & Evolution*, 2, 782–790.

356 Godsoe, W., Murray, R. & Iritani, R. (2022). Species interactions and diversity: A unified framework using  
 357 Hill numbers. *Oikos*, n/a, e09282.

358 Gonzalez-Varo, J.P. & Traveset, A. (2016). The Labile Limits of Forbidden Interactions. *Trends in Ecology &*  
 359 *Evolution*, 31, 700–710.

360 Guimarães, P.R. (2020). The Structure of Ecological Networks Across Levels of Organization. *Annual*  
 361 *Review of Ecology, Evolution, and Systematics*, 51, 433–460.

362 Jordano, P. (2016). Sampling networks of ecological interactions. *Functional Ecology*, 30, 1883–1893.

363 Jordano, P., Bascompte, J. & Olesen, J.M. (2003). Invariant properties in coevolutionary networks of  
 364 plant-animal interactions. *Ecology Letters*, 6, 69–81.

365 McLeod, A., Leroux, S.J., Gravel, D., Chu, C., Cirtwill, A.R., Fortin, M.-J., *et al.* (2021). Sampling and  
 366 asymptotic network properties of spatial multi-trophic networks. *Oikos*, n/a.

367 McLeod, A.M., Leroux, S.J. & Chu, C. (2020). Effects of species traits, motif profiles, and environment on  
 368 spatial variation in multi-trophic antagonistic networks. *Ecosphere*, 11, e03018.

369 Ohlmann, M., Miele, V., Dray, S., Chalmandrier, L., O'Connor, L. & Thuiller, W. (2019). Diversity indices  
 370 for ecological networks: A unifying framework using Hill numbers. *Ecology Letters*, 22, 737–747.

371 Olesen, J.M., Bascompte, J., Dupont, Y.L., Elberling, H., Rasmussen, C. & Jordano, P. (2010). Missing and  
 372 forbidden links in mutualistic networks. *Proceedings of the Royal Society B: Biological Sciences*, 278,  
 373 725–732.

374 Pascual, M., Dunne, J.A. & Dunne, J.A. (2006). *Ecological Networks: Linking Structure to Dynamics in Food*  
 375 *Webs*. Oxford University Press, USA.

376 Poisot, T., Canard, E., Mouillot, D., Mouquet, N. & Gravel, D. (2012). The dissimilarity of species  
 377 interaction networks. *Ecology Letters*, 15, 1353–1361.

378 Poisot, T., Cirtwill, A.R., Cazelles, K., Gravel, D., Fortin, M.-J. & Stouffer, D.B. (2016). The structure of  
 379 probabilistic networks. *Methods in Ecology and Evolution*, 7, 303–312.



380 PoisoT, T., Stouffer, D.B. & Gravel, D. (2015). Beyond species: Why ecological interaction networks vary  
 381 through space and time. *Oikos*, 124, 243–251.

382 Pringle, R.M. & Hutchinson, M.C. (2020). Resolving Food-Web Structure. *Annual Review of Ecology,*  
 383 *Evolution, and Systematics*, 51, 55–80.

384 Proulx, S.R., Promislow, D.E.L. & Phillips, P.C. (2005). Network thinking in ecology and evolution. *Trends*  
 385 *in Ecology & Evolution*, SPECIAL issue: BUMPER book REVIEW, 20, 345–353.

386 Pulliam, H.R. (1974). On the Theory of Optimal Diets. *The American Naturalist*, 108, 59–74.

387 Salim, J.A., Saraiva, A.M., Zermoglio, P.F., Agostini, K., Wolowski, M., Drucker, D.P., *et al.* (2022). Data  
 388 standardization of plantpollinator interactions. *GigaScience*, 11, giac043.

389 Saravia, L.A., Marina, T.I., Kristensen, N.P., De Troch, M. & Momo, F.R. (2022). Ecological network  
 390 assembly: How the regional metaweb influences local food webs. *Journal of Animal Ecology*, 91,  
 391 630–642.

392 Strydom, T., Bouskila, S., Banville, F., Barros, C., Caron, D., Farrell, M.J., *et al.* (2022). Food web  
 393 reconstruction through phylogenetic transfer of low-rank network representation. *Methods in Ecology*  
 394 *and Evolution*, 13.

395 Strydom, T., Catchen, M.D., Banville, F., Caron, D., Dansereau, G., Desjardins-Proulx, P., *et al.* (2021). A  
 396 roadmap towards predicting species interaction networks (across space and time). *Philosophical*  
 397 *Transactions of the Royal Society B-Biological Sciences*, 376, 20210063.

398 Vázquez, D.P., Peralta, G., Cagnolo, L., Santos, M. & Igual, E. autores contribuyeron por. (2022). Ecological  
 399 interaction networks. What we know, what we don't, and why it matters. *Ecología Austral*, 32, 670–697.

400 Wood, S.A., Russell, R., Hanson, D., Williams, R.J. & Dunne, J.A. (2015). Effects of spatial scale of  
 401 sampling on food web structure. *Ecology and Evolution*, 5, 3769–3782.