

The biological interpretation of probabilistic food webs

Francis Banville^{1,2,3,‡} Tanya Strydom^{1,3,‡} Timothée Poisot^{1,3}

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

‡ Equal contributions

Correspondance to:

Francis Banville — francis.banville@umontreal.ca

The stochastic nature of ecological interactions has led many biologists to adopt a probabilistic view of ecological networks. Representing species interactions probabilistically (how likely are they to interact?) as opposed to deterministically (do they interact?) allows a better assessment of their spatiotemporal variability and accounts for inherent uncertainties in observations and predictions. However, despite this growing interest, general guidelines regarding the estimation and documentation of probabilistic interaction data are still lacking. This is concerning given that their biological interpretation and statistical manipulation are contingent upon the methods and variables used to estimate them, which are poorly documented in most published datasets. In this contribution, we review how probabilistic interactions are defined in the literature at different spatial scales, from local interactions to regional networks, with a strong emphasis on food webs. These definitions are based on the distinction between the realization of an interaction at a specific time and space and its biological feasibility. We show that different network representations have different statistical behaviours when it comes to common ecological applications. Specifically, unlike taxonomic scaling, we argue that local and regional probabilistic networks differ in their spatial and temporal scaling of interactions, with regional interactions being scale-independent. To support our arguments, we built a spatiotemporally explicit model of probabilistic interactions and used empirical and simulated data in our case studies. Moreover, we suggest two approaches to sampling deterministic networks from probabilistic webs that account for these differences and argue that systematic biases arise when directly inferring local networks from subsets of regional webs. Overall, our results emphasize the need for better documentation of probabilistic ecological networks, both at the local and regional scales, to inform the appropriate reuse of these data.

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



Last revision: *April 2, 2023*

1 Introduction

2 Cataloging ecological interactions is a gargantuan task. Regardless of sampling effort, there are practical and
3 biological constraints that hinder our ability to observe all interactions in nature, such as the spatial and temporal
4 uncoupling of species and the large number of potential interactions in a community, of which the vast majority
5 are rare (Jordano 2016). Documenting the location and timing of interactions becomes even more challenging
6 when accounting for the spatiotemporal variability of ecological networks (Poisot *et al.* 2012, 2015). Indeed, it
7 is now recognized that knowing the biological capacity of two species to interact is necessary but not sufficient
8 for inferring their interaction at a specific time and space. For example, (Golubski2011ModMod?) presented
9 many cases where trophic interactions in food webs depend on the presence or abundance of a third species
10 (e.g., of a more profitable prey species). More generally, a handful of conditions must be satisfied for an
11 interaction to be observed locally. First, both species must have overlapping geographic ranges, i.e. they must
12 co-occur within the region of interest (Blanchet *et al.* 2020). Then, they must encounter locally. Probabilities of
13 interspecific encounters are typically low, especially for rare species with low relative abundances (Canard *et al.*
14 2012). Finally, their traits must be locally compatible (Poisot *et al.* 2015). This includes their phenology
15 (Olesen *et al.* 2010; Singer2012GeoMos?), behavioral choices (Pulliam 1974; Choh2012PreRol?) and
16 phenotypes [Bolnick2011WhyInt; (Stouffer2011RolBod?); Gravel2013InfFooo]. Environmental factors, such
17 as temperature (Angilletta2004TemGro?), drought (Woodward2012CliCha?), climate change (Araujo *et al.*
18 2011; Gilman2010FraCom?; Woodward2010ChaEco?), and habitat modifications
19 (Tylianakis2007HabMod?) contribute to this spatiotemporal variability of interactions by impacting species
20 abundance and traits. In this context, it is unsurprising that our knowledge of ecological interactions remains
21 limited (Hortal2015SevSho?) despite extensive biodiversity data collection (Schmeller2015GloTer?).

22 The recognition of the intrinsic variability of species interactions has led ecologists to expand their
23 representation of ecological networks to include a probabilistic view of interactions (Poisot *et al.* 2016). As
24 opposed to binary deterministic networks, in which interactions are either observed or not, probabilistic
25 networks represent our degree of belief about the realization or feasibility of pairwise interactions at the local or
26 regional scale, respectively. In other words, representing interactions probabilistically considers inherent
27 uncertainties and observation errors associated with ecological data. In the broadest sense, binary networks are
28 also a type of probabilistic networks, in which the value of interactions is restrained to 0 (non-observed) or 1
29 (observed). In probabilistic networks, only forbidden interactions (Jordano *et al.* 2003; Olesen *et al.* 2010)

would have a probability value of zero (but see Gonzalez-Varo & Traveset 2016). However, *neutral* forbidden interactions (i.e., improbable interactions between rare species, Canard *et al.* 2012) could have low probability values in a local network but high probability in a regional network (metaweb) describing the biological capacity of species to interact.

By accounting for the uncertainty of interactions, probabilistic networks provide a more realistic portrait of species interactions and of their emerging structure. This is important given that network structure is one of the major drivers of the functioning, dynamics, and resilience of ecosystems worldwide (Proulx *et al.* 2005; **McCann2007ProBio?**; **McCann2011FooWeb?**; **Rooney2012IntFoo?**). Moreover, the application and development of computational methods in network ecology, which are often based on a probabilistic representation of interactions, can help alleviate the colossal sampling efforts required to document species interactions (Strydom *et al.* 2021). For example, statistical models can be used to estimate the uncertainty of pairwise interactions (**Cirtwill2019QuaFra?**) and the probability of missing (false negatives) and spurious (false positives) interactions (**Guimera2009MisSpu?**). Considering the high rate of false negatives in species interaction data due to the difficulty of witnessing rare interactions (Catchen *et al.* 2023), these models can inform the identification of priority sampling locations of ecological networks (e.g., Andrade-Pacheco *et al.* 2020 present an approach to identify priority sampling locations of disease hotspots). Statistical models can also be used to generate network predictions without prior knowledge about their pairwise interactions, for instance using body size (**Petchey2008SizFor?**; **Gravel2013InfFoa?**), phylogeny (Strydom *et al.* 2022; **Elmasri2020HieBay?**), or a combination of niche and neutral processes (**Bartomeus2016ComFra?**; **Pomeranz2019InfPre?**) for inference. Topological null models (e.g., **Bascompte2003NesAss?**; **Fortuna2006HabLos?**), which can be used to generate underlying distributions of network measures for null hypothesis significance testing, are other examples of common probabilistic network models. Many measures have been developed to describe the structure (Poisot *et al.* 2016) and diversity (Ohlmann *et al.* 2019; Godsoe *et al.* 2022) of probabilistic networks. These models and measures support the use of this approach for the study of a wide range of ecological questions, from making better predictions of species distribution (Cazelles *et al.* 2016) to forecasting the impact of climate change on ecological networks (**Gilman2010FraCom?**).

Despite these advances and opportunities, the lack of clear guidelines on the use of probabilistic interaction data is worrisome, especially for field and computational ecologists who manipulate and generate these numbers. Indeed, beyond methodological challenges encountered when evaluating them, there are important and perhaps more fundamental conceptual challenges when it comes to defining them. To the best of our knowledge, there is

currently no data standard that could guide the estimation and documentation of interaction probabilities (Salim *et al.* 2022 discuss data standards for deterministic mutualistic networks). General guidelines could support a more adequate integration and manipulation of interaction data from different sources and prevent ecologists from being misled by ambiguous and often diverging interpretations of probabilistic networks. In this contribution, we aim to take a step back by outlining different ways in which they were defined and used in network ecology and propose an approach to thinking about them. We distinguish two broad categories of probabilistic networks that have different statistical behaviors when applied to key ecological questions: local networks of realized interactions and regional networks (metawebs) of potential interactions. We show that these representations have different ecological and statistical implications, especially regarding the spatial and temporal scaling of interactions and the prediction of binary networks across space. Although we focus on food webs, our observations and advice can be applied to most types of ecological networks, from plant-pollinator to host-parasite networks. Indeed, with the exception of networks of indirect interactions such as competition and facilitation networks (Kefi2015NetStr?; Kefi2016HowStr?), most ecological networks describe probabilities of direct interactions, which are conceptually and mathematically analogous to each other regardless of their biological type (e.g., trophic and parasitic interactions). Overall, we argue that probabilistic networks should be better documented, clearly defined in mathematical terms, and used with caution when analyzing ecological interactions.

Stochastic representations of biological interactions

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

88 that a better understanding of these differences and concepts would alleviate interpretation errors and help
89 ecologists use these numbers more appropriately.

90 **Pairwise interactions: the building blocks of ecological networks**

91 The basic unit of food webs and other ecological networks are individuals that interact with each others [e.g., by
92 predation; Elton (2001)], forming individual-based networks. The aggregation of these individuals into more or
93 less homogeneous groups (e.g., populations, species, trophic species, families) allows us to represent networks
94 at broader taxonomic scales, which impacts our interpretation of the properties and behaviour of these systems
95 (Guimarães 2020). Nodes can thus designate distinct levels of organization, whereas edges linking these nodes
96 can describe a variety of interaction measures. When using a Boolean representation of biotic interactions, the
97 observation that one individual from group (or node) i interacts with another individual from group j is enough
98 to set the interaction $A_{i,j}$ to 1. This simplified representation of food webs is a highly valuable source of
99 ecological information (Pascual *et al.* 2006) even though it overlooks important factors regarding interaction
100 strengths. These, in turn, can be represented using quantitative interactions, which better describe the energy
101 flows, demographic impacts or frequencies of interactions between nodes (Berlow *et al.* 2004; Borrett &
102 Scharler 2019), with $A_{i,j}$ being a natural number \mathbb{N} or a real number \mathbb{R} depending on the measure. For
103 example, they can be used to estimate the average number of prey individuals consumed by the predators in a
104 given time period (e.g., the average number of fish in the stomach of a piscivorous species). Interaction
105 strengths can also be used as good estimators of the parameters describing species interactions in a
106 Lotka-Volterra model (e.g., Emmerson & Raffaelli 2004). This extra amount of ecological information typically
107 comes at a cost of greater sampling effort or data requirement in predictive models (Strydom *et al.* 2021), which
108 can lead to high uncertainties when building these networks.

109 The uncertainty and spatiotemporal variability of both types of trophic interactions (Boolean and quantitative)
110 can be represented probabilistically. On one hand, Boolean interactions follow a Bernoulli distribution
111 $A_{i,j} \sim \text{Bernoulli}(p)$, with p being the probability of interactions. The only two possible outcomes are the
112 presence ($A_{i,j} = 1$) or absence ($A_{i,j} = 0$) of an interaction between the two nodes. Quantitative interactions, on
113 the other hand, can follow various probability distributions depending on the measure used. In this case, the
114 event's outcome is the value of interaction strength. For instance, these interaction strengths can follow a
115 Poisson distribution $A_{i,j} \sim \text{Poisson}(\lambda)$ when predicting frequencies of interactions between pairs of nodes, with
116 λ being the expected rate of interaction. Note that quantitative interactions can be converted to probabilistic

interactions by normalizing. The definition and interpretation of parameters like p and λ are inextricably linked to environmental and biological factors such as species relative abundance, traits, area, and time, depending on the type of interaction. Because Boolean species interactions are much more documented in the literature, our primary focus in this contribution will be on addressing the challenges in defining and interpreting p for pairwise species interactions.

Local networks: communities interacting in space and time

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 Ω 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 :

$$P_N(i \rightarrow j|A, t, C, \Omega), \quad (1)$$

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

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

157 **Metawebs: regional catalogs of interactions**

158 Metawebs are networks of potential interactions, representing the probability that two taxa can interact
159 regardless of biological plasticity, environmental variability or co-occurrence. Instead of describing stochastic
160 biological processes occurring in nature, probabilistic potential interactions can be thought of as a measure of
161 imperfect knowledge about the capacity of two taxa to interact. They are the network analogue to the species
162 pool, where local networks originate from a subset of both species (nodes) and interactions (edges). For this
163 reason, they have been initially designed for broad spatial, temporal, and taxonomic scales (e.g, species food
164 webs at the continental scale).

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

$$P_M(i \rightarrow j), \quad (2)$$

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

169 Starting from a selected set of species, metawebs can be built using different data sources, including literature
170 review, field work, and predictive models (e.g., the metaweb of Canadian mammals inferred by Strydom *et al.*
171 2022). Every pair of species that has been observed to interact at least once can be given a probability of
172 interaction of 1; we know that they *can* interact. This means that rare interactions can technically be given high
173 probabilities in the metaweb. Unobserved interactions, on the other hand, are given lower probabilities, going as
174 low as 0 for forbidden links. Two important nuances must however be made here. Because of observation errors
175 due to taxonomic misidentifications and ecological misinterpretations [e.g., due to cryptic species and
176 interactions; Pringle & Hutchinson (2020)], many observations of interactions are actually false positives.
177 Similarly, forbidden interactions can be false negatives if e.g. they have been assessed for specific phenotypes,
178 locations or time. Implementing a Bayesian framework, which updates prior probabilities of interactions with
179 empirical data, could lessen these errors.

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

181 **Statistical behaviour of networks in key ecological applications**

182 **Taxonomic agglomeration and division of nodes**

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

196 There are a lot of similarities between taxonomic and spatiotemporal scaling of probabilistic interactions.
197 Fundamentally, these types of scaling are just different ways to aggregate individuals into broader nodes, either
198 spatially, temporally, or taxonomically. However, there are also important differences between them. First, in
199 metawebs, if we know that two species have the capacity to interact, we can infer that their respective genus
200 should also be able to interact (i.e., there should be at least two individuals within these genus that can interact).
201 On the contrary, knowing that two genus can interact does not mean that all pairwise combinations of species
202 within these genus can also interact among themselves. This observation also applies to local networks. When it
203 comes to probabilistic networks, interaction probabilities at broader taxonomic scales can be directly obtained
204 from probabilities at finer scales when aggregating nodes. For example, if we have in a network n_A species from
205 genus A and n_B species from genus B , we can calculate the probability that the two genus interact as
206 $P_N(A \rightarrow B) = 1 - \prod_{i=1}^{n_A} \prod_{j=1}^{n_B} (1 - P_N(A_i \rightarrow B_j))$, where A_i and B_j are the species of the corresponding genus.
207 However, more sophisticated models need to be used when building probabilistic networks at smaller taxonomic
208 resolutions (e.g., when building a species-level network from a genus-level network). One could, for example,
209 estimate the probabilities of all pairwise species interactions by using a Beta distribution parametrised by the
210 broader-scale network.

211 [Figure 3 about here]. Conceptual figure of how a scale up of the nodes from an individual to a population to
212 any higher taxonomic group change our interpretation of the probability of interaction.

213 **Spatial and temporal scaling of probabilistic interactions**

214 Metawebs and local networks intrinsically differ in their relation to scale. On one hand, as mentioned above,
215 probabilistic metawebs are context independent, i.e., probabilistic pairwise interactions do not scale with space
216 and time because they depend solely on the biological capacity of the two taxa to interact. This implies that the
217 estimated likelihood that two species can potentially interact should be the same among all metawebs in which
218 they are present. In practice, this is rarely the case because ecologists use different methods and data to estimate
219 these probabilities of interactions (e.g., different sampling area and time period). However, in the case where
220 local metawebs $M_{x,y}$ are subsampled from their regional counterpart M_0 , we should expect edge values to be
221 identical among all networks, regardless of their spatial scale, i.e. $P_{M_{x,y}}(i \rightarrow j) = P_{M_0}(i \rightarrow j)$. On the other
222 hand, local probabilistic networks are indissociable from their spatial and temporal contexts because there are
223 more opportunities of interactions (e.g., more individuals, more trait variations, more chance of encounter) in a
224 larger area and longer time period (McLeod *et al.* 2020). Let N_0 be a local probabilistic food web delineated in

an area A_0 and N_1 a network of area $A_1 < A_0$ within A_0 . We should expect the probability that i and j interacts to scale spatially, 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 webs could be manifested through the effect of sampling effort on the observation of interactions (Jordano 2016; McLeod *et al.* 2021) or of time itself on their realization (Poisot *et al.* 2012). There are many network-area relationships (e.g., Wood *et al.* 2015; Galiana *et al.* 2018) and interaction accumulation curves (e.g, Jordano 2016) explored in the literature. These could inspire the development and testing of different equations describing the spatiotemporal scaling of probabilistic pairwise interactions in local networks, which are over the scope of this manuscript.

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

Spatiotemporally explicit model of probabilistic interactions

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

One way that probabilistic food webs can be made spatiotemporally explicit is by modelling interactions between co-occurring species as a Poisson process with rate λ . Specifically, if the total observation time for a location is t_0 , the probability that two co-occurring species i and j will interact during this time period is $P_N(i \rightarrow j | C_{i,j} = 1, t = t_0) = 1 - e^{-\lambda t_0}$, which approaches 1 when $t \rightarrow \infty$. The value of the parameter λ could be estimated using prior data on interaction strengths, if available. Additionally, we can estimate the 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 $P_{x,y}(j)$ are respectively the probabilities of occurrence of species i and j and γ is the strength of association between occurrence and co-occurrence, as defined in Cazelles *et al.* (2016). Note that in empirical networks, γ 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 λ with given environmental variables or by adding in observation errors (i.e., probability

253 of false negatives and false positives; Catchen *et al.* (2023)).

254 **Binary conversion of probabilistic networks through random draws**

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

272 **Prediction of local networks from probabilistic metawebs**

273 Even though the spatiotemporal variability of interactions is not considered in metawebs, they can still be useful
274 to reconstruct local networks of realized interactions. Indeed, local networks are formed from subsets of their
275 metaweb (called subnetworks), which are obtained by selecting a subset of both species and interactions (Dunne
276 2006). Because a community's composition is arguably easier to sample (or predict) than its interactions, the
277 biggest challenge is to sample links from the metaweb. This becomes a conceptual issue when we consider how
278 potential and realized interactions differ. Despite these concerns, metawebs remain an important source of
279 ecological information that can be leveraged for realistically predicting spatially explicit networks. First,

metawebs set the upper limit of species interactions (McLeod *et al.* 2021), i.e. the probability that two species interact at a specific location is always lower or equal to the probability of their potential interaction:

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

Therefore, inferring local networks from their metaweb keeping the same values of interaction probability would generate systematic biases in the prediction. In that case, these networks would instead be called *spatially explicit* or *local* metawebs (i.e., smaller-scale networks of potential interactions). Second, the structure of local networks is constrained by the one of their metaweb (Saravia *et al.* 2022). This suggests that a metaweb not only constrains the pairwise interactions of its corresponding local networks, but also their emerging properties. Inferring the structure of local networks from the metaweb could thus help estimate more realistically the likelihood that potential interactions are realized and observed locally (Strydom *et al.* 2021). [Figure 1 about here]. Empirical example of the association between the number of interactions in realized local food webs and the number of interactions in the corresponding subnetworks of their regional metaweb. We should expect the association to be linear below the 1:1 line, illustrating eq. 3.

Conclusion

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

In this contribution, we showed that network metadata are perhaps as important as interaction data themselves when it comes to interpreting probabilistic food webs in ecological terms. First, the type of probabilistic interaction (potential or realized) impacts the importance of scale, with interactions in metawebs being scale independent, both spatially and temporally. Second, the conditions in which local networks were estimated (e.g., area, time, biological and environmental factors) and the attributes of the interacting species that were considered (e.g., species co-occurrence) are essential contextual factors that impact the mathematical

305 representation of probabilities and their resulting behaviour. Third, the biological interpretation of probabilities
306 changes with the level of organization of the network due to the aggregation of individuals into broader groups.
307 All these information should be available as clear metadata so that ecologists can use probabilistic network data
308 appropriately.

309 Moreover, many statistical models in ecology that yield accurate predictions of biotic interactions are black
310 boxes difficult to interpret. Ecologists should be careful before using the output of these models as probabilistic
311 objects, since there is often a thin line between a real probability and a non-probabilistic predictive number (or
312 score). Probabilities are numbers between 0 and 1 that sum to 1 and either represent the expected frequency of a
313 phenomenon or the degree of belief that it will be realized. Non-probabilistic scores, which are more akin to
314 interaction strengths, have different mathematical properties, which impacts how we should handle these
315 numbers in a spatially or temporally explicit context. Therefore, researchers should use their expertise to assess
316 if their interaction data are actually probabilities or scores. This should also be added to the metadata before
317 sharing them, as well as the methods used to build the networks.

318 Better metadata documentation would allow researchers to use and manipulate probabilistic ecological
319 interactions according to how they were actually defined and obtained. This would support better scientific
320 practices, in particular when these data are used for ecological prediction and forecasting. For instance, we
321 showed that building a rigorous workflow to predict local networks from a probabilistic metaweb requires a
322 good understanding of the data at hand. Similarly, explicitly stating the context in which probabilistic data were
323 estimated would help using forecasting food-web models more rigorously under specific climate change and
324 habitat use scenarios. Regardless of the method and application, fostering a better foundation for probabilistic
325 reasonings in network ecology, from the very nature of probabilities and biotic interactions, is essential.

326 **Acknowledgement**

327 We acknowledge that this study was conducted on land within the traditional unceded territory of the Saint
328 Lawrence Iroquoian, Anishinabewaki, Mohawk, Huron-Wendat, and Omàmiwininiwak nations. This work was
329 supported by the Institute for Data Valorisation (IVADO) and the Natural Sciences and Engineering Research
330 Council of Canada (NSERC) Collaborative Research and Training Experience (CREATE) program, through the
331 Computational Biodiversity Science and Services (BIOS²) program. A special thank to all members of the
332 Black Holes and Revelations working group (organized by BIOS²) for their insightful discussions and valuable

333 feedback on this manuscript.

334 **References**

- 335 Andrade-Pacheco, R., Rerolle, F., Lemoine, J., Hernandez, L., Meïté, A., Juziwelo, L., *et al.* (2020). [Finding](#)
336 [hotspots: Development of an adaptive spatial sampling approach](#). *Scientific Reports*, 10, 10939.
- 337 Araujo, M.B., Rozenfeld, A., Rahbek, C. & Marquet, P.A. (2011). [Using species co-occurrence networks to](#)
338 [assess the impacts of climate change](#). *Ecography*, 34, 897–908.
- 339 Berlow, E.L., Neutel, A.-M., Cohen, J.E., De Ruiter, P.C., Ebenman, B., Emmerson, M., *et al.* (2004).
340 [Interaction strengths in food webs: Issues and opportunities](#). *Journal of Animal Ecology*, 73, 585–598.
- 341 Blanchet, F.G., Cazelles, K. & Gravel, D. (2020). [Co-occurrence is not evidence of ecological interactions](#).
342 *Ecology Letters*, 23, 1050–1063.
- 343 Bohan, D.A., Vacher, C., Tamaddoni-Nezhad, A., Raybould, A., Dumbrell, A.J. & Woodward, G. (2017).
344 [Next-Generation Global Biomonitoring: Large-scale, Automated Reconstruction of Ecological Networks](#).
345 *Trends in Ecology & Evolution*, 32, 477–487.
- 346 Borrett, S.R. & Scharler, U.M. (2019). [Walk partitions of flow in Ecological Network Analysis: Review and](#)
347 [synthesis of methods and indicators](#). *Ecological Indicators*, 106, 105451.
- 348 Canard, E.F., Mouquet, N., Mouillot, D., Stanko, M., Miklisova, D. & Gravel, D. (2014). [Empirical Evaluation](#)
349 [of Neutral Interactions in Host-Parasite Networks](#). *The American Naturalist*, 183, 468–479.
- 350 Canard, E., Mouquet, N., Mareshcot, L., Gaston, K.J., Gravel, D. & Mouillot, D. (2012). [Emergence of](#)
351 [Structural Patterns in Neutral Trophic Networks](#). *PLOS ONE*, 7, e38295.
- 352 Catchen, M.D., Poisot, T., Pollock, L.J. & Gonzalez, A. (2023). The missing link: Discerning true from false
353 negatives when sampling species interaction networks.
- 354 Cazelles, K., Araujo, M.B., Mouquet, N. & Gravel, D. (2016). [A theory for species co-occurrence in interaction](#)
355 [networks](#). *Theoretical Ecology*, 9, 39–48.
- 356 Dunne, J.A. (2006). The Network Structure of Food Webs. In: *Ecological networks: Linking structure and*
357 *dynamics* (eds. Dunne, J.A. & Pascual, M.). Oxford University Press, pp. 27–86.
- 358 Elton, C.S. (2001). *Animal Ecology*. University of Chicago Press, Chicago, IL.

359 Emmerson, M.C. & Raffaelli, D. (2004). [Predatorprey body size, interaction strength and the stability of a real](#)
360 [food web](#). *Journal of Animal Ecology*, 73, 399–409.

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

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

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

367 Guimarães, P.R. (2020). [The Structure of Ecological Networks Across Levels of Organization](#). *Annual Review*
368 *of Ecology, Evolution, and Systematics*, 51, 433–460.

369 Jordano, P. (2016). [Sampling networks of ecological interactions](#). *Functional Ecology*, 30, 1883–1893.

370 Jordano, P., Bascompte, J. & Olesen, J.M. (2003). [Invariant properties in coevolutionary networks of](#)
371 [plantanimal interactions](#). *Ecology Letters*, 6, 69–81.

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

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

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

378 Olesen, J.M., Bascompte, J., Dupont, Y.L., Elberling, H., Rasmussen, C. & Jordano, P. (2010). [Missing and](#)
379 [forbidden links in mutualistic networks](#). *Proceedings of the Royal Society B: Biological Sciences*, 278,
380 725–732.

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

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

385 PoisoT, T., Cirtwill, A.R., Cazelles, K., Gravel, D., Fortin, M.-J. & Stouffer, D.B. (2016). [The structure of](#)
386 [probabilistic networks](#). *Methods in Ecology and Evolution*, 7, 303–312.

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

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

391 Proulx, S.R., Promislow, D.E.L. & Phillips, P.C. (2005). [Network thinking in ecology and evolution](#). *Trends in*
392 *Ecology & Evolution*, SPECIAL ISSUE: BUMPER BOOK REVIEW, 20, 345–353.

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

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

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

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

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

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

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