

The ecological interpretation of probabilistic networks

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 in probabilistic networks, 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. In this contribution, we review how probabilistic interactions are defined in the literature at different spatial scales, from local interactions to regional networks (metawebs), 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, we argue that local probabilistic networks and metawebs differ in their spatial and temporal scaling of interactions, with potential interactions being scale-independent. This is in contrast with the taxonomic scaling of interactions, which does not qualitatively differ between both types of networks. 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 metawebs. To support our arguments, we built a spatiotemporally explicit model of probabilistic interactions and developed different case studies using empirical and simulated data. 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.

Last revision: *May 30, 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
4 temporal uncoupling of species and the large number of potential interactions in a community, of which the vast
5 majority are rare (Jordano 2016). Documenting the location and timing of interactions becomes even more
6 challenging when accounting for the spatiotemporal variability of ecological networks (Poisot *et al.* 2012,
7 2015). Indeed, it is now recognized that knowing the biological capacity of two species to interact is necessary
8 but not sufficient for inferring their interaction at a specific time and space. For example, Golubski & Abrams
9 (2011) presented many cases where trophic interactions in food webs depend on the presence or abundance of a
10 third species (e.g., of a more profitable prey species). More generally, a handful of conditions must be satisfied
11 for an interaction to be observed locally. First, both species must have overlapping geographic ranges, i.e. they
12 must co-occur within the region of interest (Blanchet *et al.* 2020). Then, they must encounter locally.
13 Probabilities of interspecific encounters are typically low, especially for rare species with low relative
14 abundances (Canard *et al.* 2012). Finally, their traits must be locally compatible (Poisot *et al.* 2015). This
15 includes their phenology (Olesen *et al.* 2010; Singer & McBride 2012), behavioral choices (Pulliam 1974;
16 Choh *et al.* 2012) and phenotypes [Bolnick 2011 WhyInt; Stouffer *et al.* (2011); Gravel 2013 InfFoa].
17 Environmental factors, such as temperature (Angilletta *et al.* 2004), drought (Woodward *et al.* 2012), climate
18 change (Gilman *et al.* 2010; Woodward *et al.* 2010; Araujo *et al.* 2011), and habitat modifications (Tylianakis *et al.*
19 2007), contribute to this spatiotemporal variability of interactions by impacting species abundance and
20 traits. In this context, it is unsurprising that our knowledge of ecological interactions remains limited (Hortal *et al.*
21 2015) despite extensive biodiversity data collection (Schmeller *et al.* 2015).

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 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; McCann 2007; McCann 2011; Rooney & McCann 2012). 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 (Cirtwill *et al.* 2019) and the probability of missing (false negatives) and spurious (false positives) interactions (Guimerà & Sales-Pardo 2009). 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 (Petchey *et al.* 2008; Gravel *et al.* 2013), phylogeny (Elmasri *et al.* 2020; Strydom *et al.* 2022), or a combination of niche and neutral processes (Bartomeus *et al.* 2016; Pomeranz *et al.* 2019) for inference. Topological null models (e.g., Bascompte *et al.* 2003; Fortuna & Bascompte 2006), 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 (Gilman *et al.* 2010).

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, excepting networks of indirect interactions such as competition and facilitation networks (Kéfi *et al.* 2015, 2016), 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.

Probabilistic representations of ecological interactions

One of the first aspects to take into consideration when estimating or interpreting probabilities of interactions is knowing if they describe 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* theoretically 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 locally). Here, we use the terms *metaweb* to designate networks of potential interactions and *local networks* for those of realized interactions. Metawebs are the network analogue of the species pool, where local networks originate from a subset of both species (nodes) and interactions (edges) of the regional metaweb (Saravia *et al.* 2022). Frequent confusion arises among ecologists over the use of these two terms, especially in a probabilistic context. Indeed, it is often unclear when authors describe potential or realized interactions in their studies, or when so-called probabilities are actually *interaction scores*. Likewise, probabilistic potential interactions are often used and interpreted as realized interactions (and

conversely), even when the type of interaction is specified, which may generate misleading findings when analyzing these data. We believe that a better understanding of the differences and relationships between these probabilistic representations of ecological networks would alleviate interpretation errors and help ecologists use these numbers more appropriately.

Pairwise interactions: the building blocks of ecological networks

Local ecological networks and metawebs, like any type of networks, are made of nodes and edges that can be represented at different levels of organization and precision. The basic unit of food webs and other ecological networks are individuals that interact with each other (e.g., by predation, Elton (2001)), forming individual-based networks. The aggregation of these individuals into more or less homogeneous groups (e.g., populations, species, trophic species, families) allows us to represent nodes at broader taxonomic scales, which impacts our interpretation of the properties and behavior of these systems (Guimarães 2020). Moreover, edges linking these nodes can describe a variety of interaction measures. Ecologists have traditionally represented interactions as binary objects that were considered realized after observing at least one individual from group i interact with at least another individual from group j . Boolean interactions are actually the result of a Bernoulli process $A_{i,j} \sim \text{Bernoulli}(P_{i,j})$, with $P_{i,j}$ being the probability of interaction that characterizes our limited knowledge of interactions and their intrinsic spatiotemporal variability. Depending on the type of networks (local or metaweb), the mathematical formulation and interpretation of stochastic parameters like $P_{i,j}$ can be linked to environmental and biological factors such as species relative abundance, traits, area, and time (tbl. 1). In these probabilistic network representations in which $P_{i,j}$ are edge values, the only two possible outcomes are the presence ($A_{i,j} = 1$) or absence ($A_{i,j} = 0$) of an interaction between each pair of nodes. Observing an interaction between two taxa at a given location and time provides important information that can be used to update previous estimates of $P_{i,j}$, informing us on the biological capacity of both taxa to interact and the environmental conditions that enable the realization of their interaction in space and time.

Even though binary networks constitute a highly valuable source of ecological information (Pascual *et al.* 2006), they overlook important factors regarding interaction strengths. These are represented using quantitative interactions, which better describe the energy flows, demographic impacts or frequencies of interactions between nodes (Berlow *et al.* 2004; Borrett & Scharler 2019), with $A_{i,j}$ being a natural number \mathbb{N} or a real number \mathbb{R} depending on the measure. For example, they can represent the average number of prey individuals consumed by a predator in a given time period (e.g., the average number of fish in the stomach of a piscivorous

species). Because quantitative interactions can describe predation pressure on prey taxa, they can be good estimators of the parameters describing species interactions in a Lotka-Volterra model (e.g., Emmerson & Raffaelli 2004). However, this extra amount of ecological information typically comes at a cost of greater sampling effort or data requirement in predictive models (Strydom *et al.* 2021), which can lead to relatively high levels of uncertainties when inferring quantitative networks with limited data. Just like binary networks, the uncertainty and spatiotemporal variability of quantitative interactions can be represented probabilistically, with the difference that quantitative interactions can follow various probability distributions depending on the measure used, the event's outcome being the value of interaction strength. For instance, quantitative interactions can be distributed as a Poisson distribution $A_{i,j} \sim \text{Poisson}(\lambda_{i,j})$ when predicting frequencies of interactions between pairs of nodes, with $\lambda_{i,j}$ being the expected rate at which individuals of taxa i and j interact (e.g., the average number of prey j consumed per unit time by all predators i). The Poisson distribution can also be 0-inflated when considering non-interacting taxa, which constitute the majority of taxa pairs in most local networks due to their typically high sparseness (Jordano 2016). Because of the methodological difficulties typically encountered when building deterministic quantitative networks, binary networks, which are easier to sample (Jordano 2016) and predict (Strydom *et al.* 2021), are much more documented and modeled in the literature. Moreover, most published probabilistic networks and methods describe Bernoulli interactions, which supports the need for better guidelines regarding the interpretation and manipulation of these types of networks. For these reasons, our primary focus in this contribution will be on addressing the challenges in estimating and using Bernoulli interactions, in both local probabilistic networks and probabilistic metawebs.

Local networks: communities interacting in space and time

As opposed to metawebs, probabilistic local networks describe how likely taxa are to interact at a given location and time period (i.e., they are context-dependant). In local networks, edges commonly represent our degree of belief that two taxa interact in space and time, but can also document the probability of *observing* this interaction (Catchen *et al.* 2023). For example, Kopelke *et al.* (2017) assembled a dataset of deterministic local European food webs of willow-galling sawflies and their natural enemies, clearly referencing each food web in space and time. Because of its large number of replicated samples, this dataset can be used to infer local probabilities of observing interactions (Gravel *et al.* 2019). More generally, we define space as the geographic coordinates (x, y) of the spatial boundaries delineating the system (sampled or targeted) and time as the time interval t during which interactions were sampled or for which they were predicted. Given that space and time

are in reality continuous variables, the probability of an interaction occurring within a particular spatial and temporal setting is given by the integral of the probability density function describing the relative likelihood that this interaction is realized at any specific and infinitely small location and time. Therefore, edges could represent probability densities or masses depending on how space and time are measured. For simplicity reasons, we will consider space and time as discrete dimensions that provide actual probabilities of interactions, which is conform to how ecological interactions are usually sampled. Using space and time intervals also allows us to measure an area A and duration t , which can be used in spatiotemporal analyses of ecological networks such as the study of network-area relationships (NAR, Galiana *et al.* 2018). We should indeed expect probabilities of interactions to scale with area and duration because taxa have more opportunities to interact.

The probability that two taxa i and j interact locally can also be conditional on many environmental and biological factors. One of these is their co-occurrence C_{ij} , which is usually a Boolean describing if the geographic distribution of both taxa overlaps within the study area. The probability that the interaction is realized must be 0 when taxa do not co-occur, i.e. $P_N(i \rightarrow j|C = 0) = 0$. Probabilistically, the co-occurrence of two taxa follows a Bernoulli distribution $C_{ij} \sim \text{Bernoulli}(P_{ij}(x, y))$, where the probability of co-occurrence can be estimated using species distribution models (e.g., Pollock *et al.* 2014). More generally, the probability that both taxa interact locally can be obtained by the product of their probability of interaction given co-occurrence with their probability of co-occurrence:

$$P_N(i \rightarrow j) = P_N(i \rightarrow j|C = 1) \times P_{ij}(x, y). \quad (1)$$

Other important factors that can impact our estimation of interaction probabilities are taxa relative abundance (Canard *et al.* 2012) and traits (Poisot *et al.* 2015), as well as environmental factors such as temperature (Angilletta *et al.* 2004), precipitation (Woodward *et al.* 2012), habitat structure (Klecka & Boukal 2014), and presence of other interacting taxa in the network (Kéfi *et al.* 2012; Pilosof *et al.* 2017). Here, we will use the variable Ω to describe the biological and ecological context in which interaction probabilities were estimated. For example, a mesocosm experiment could be conducted to estimate interaction probabilities with and without shelters where prey can avoid predation. In that case, Ω would represent the presence or absence of refuges. Like co-occurrence, Ω can also be modeled probabilistically when environmental and biological factors are considered stochastic or uncertain. In sum, Ω represents all ecological and biological variables that were taken into consideration when measuring interaction probabilities and is, therefore, a subset of all elements actually

173 impacting ecological interactions.

174 The probability that two taxa i and j interacts in a local network N can thus be conditional on the area A , the
 175 time interval t , their co-occurrence C and chosen environmental and biological conditions Ω . This gives us the
 176 following equation when all these conditions are included in the estimation of interaction probabilities:

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

177 The local context in which probabilities are estimated and the variables that should be taken into consideration
 178 depend on the study system, the objective of the study, and the resources available to the researchers. In other
 179 words, these variables do not systematically need to be accounted for. However, when they are, they should be
 180 specified in the documentation of the data, preferentially in mathematical terms to avoid any confusion in their
 181 interpretation and manipulation errors during their re-use. For example, when generating interaction data,
 182 ecologists should be explicit about their consideration of co-occurrence in their estimation of local interaction
 183 probabilities since this can impact their interpretation and analysis. Indeed, it is important to specify if
 184 probability values are conditional $P_N(i \rightarrow j|C = 1)$ or not $P_N(i \rightarrow j)$ on co-occurrence. For practical reasons,
 185 probabilistic ecological networks are generally represented as adjacency matrices of probability values (i.e.,
 186 matrices of edge values). Without clear indications about C , it may be difficult to interpret if these numbers
 187 represent “raw” edge values or the product of edge and node values. In tbl. 1, we present a handful of studies
 188 that used probabilistic ecological networks and their formulation of probabilistic interactions. This table
 189 illustrates the diversity of measures used to estimate probabilistic interactions and emphasizes the need to
 190 qualify interaction data before integrating and analyzing them.

Table 1: Interaction probabilities are interpreted differently in metawebs and local networks. Each formula includes different conditional variables and is described in plain text. A non-exhaustive list of studies using these formulations is included, with the variables they used specified in parentheses.

Formula	Description	Studies
$P_M(i \rightarrow j)$	probability that the interaction is biologically feasible	
$P_N(i \rightarrow j)$	probability that the interaction is realized locally	
$P_N(i \rightarrow j A)$	probability that the interaction is realized locally given the area	

Formula	Description	Studies
$P_N(i \rightarrow j t)$	probability that the interaction is realized locally given the time period	
$P_N(i \rightarrow j C)$	probability that the interaction is realized locally given co-occurrence	
$P_N(i \rightarrow j \Omega)$	probability that the interaction is realized locally given chosen environmental and biological factors	
$P_N(i \rightarrow j A, t, C, \Omega)$	probability that the interaction is realized locally given several factors	

Metawebs: regional catalogs of interactions

Metawebs are networks of potential interactions, representing the probability that two taxa can interact regardless of biological plasticity, environmental variability or co-occurrence. Instead of describing stochastic biological processes occurring in nature, probabilistic potential interactions can be thought of as a measure of imperfect knowledge about the capacity of two taxa to interact. For this reason, they have been initially designed for broad spatial, temporal, and taxonomic scales (e.g, species food webs at the continental scale).

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

$$P_M(i \rightarrow j), \quad (3)$$

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.

Below, we will see examples of why this matters when it comes to spatial, temporal, and taxonomic scaling of biotic interactions.

Statistical behaviour of networks in key ecological applications

Quantitative interactions can be converted to probabilistic interactions by normalizing.

Taxonomic agglomeration and division of nodes

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 calculate the probability that the two genus interact as $P_N(A \rightarrow B) = 1 - \prod_{i=1}^{n_A} \prod_{j=1}^{n_B} (1 - P_N(A_i \rightarrow B_j))$, where A_i and B_j are the species of the corresponding genus. However, more sophisticated models need to be used when building probabilistic networks at smaller taxonomic resolutions (e.g., when building a species-level network from a genus-level network). One could, for example, estimate the probabilities of all pairwise species interactions by using a Beta distribution parametrised by the broader-scale network.

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

Spatial and temporal scaling of probabilistic interactions

Metawebs and local networks intrinsically differ in their relation to scale. On one hand, as mentioned above, probabilistic metawebs are context independent, i.e., probabilistic pairwise interactions do not scale with space and time because they depend solely on the biological capacity of the two taxa to interact. This implies that the estimated likelihood that two species can potentially interact should be the same among all metawebs in which they are present. In practice, this is rarely the case because ecologists use different methods and data to estimate these probabilities of interactions (e.g., different sampling area and time period). However, in the case where local metawebs $M_{x,y}$ are subsampled from their regional counterpart M_0 , we should expect edge values to be 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 hand, local probabilistic networks are indissociable from their spatial and temporal contexts because there are more opportunities of interactions (e.g., more individuals, more trait variations, more chance of encounter) in a 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 of false negatives and false positives; Catchen *et al.* (2023)).

Binary conversion of probabilistic networks through random draws

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

Prediction of local networks from probabilistic metawebs

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

$$P_N(i \rightarrow j|A, t, C, \Omega) \leq P_M(i \rightarrow j). \quad (4)$$

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. 4.

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 representation of probabilities and their resulting behaviour. Third, the biological interpretation of probabilities changes with the level of organization of the network due to the aggregation of individuals into broader groups.

341 All these information should be available as clear metadata so that ecologists can use probabilistic network data
342 appropriately.

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

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

360 **Acknowledgement**

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

References

- Andrade-Pacheco, R., Rerolle, F., Lemoine, J., Hernandez, L., Meité, A., Juziwelo, L., *et al.* (2020). [Finding hotspots: Development of an adaptive spatial sampling approach](#). *Scientific Reports*, 10, 10939.
- Angilletta, M.J., Jr., Steury, T.D. & Sears, M.W. (2004). [Temperature, Growth Rate, and Body Size in Ectotherms: Fitting Pieces of a Life-History Puzzle](#). *Integrative and Comparative Biology*, 44, 498–509.
- Araujo, M.B., Rozenfeld, A., Rahbek, C. & Marquet, P.A. (2011). [Using species co-occurrence networks to assess the impacts of climate change](#). *Ecography*, 34, 897–908.
- Bartomeus, I., Gravel, D., Tylianakis, J.M., Aizen, M.A., Dickie, I.A. & Bernard-Verdier, M. (2016). [A common framework for identifying linkage rules across different types of interactions](#). *Functional Ecology*, 30, 1894–1903.
- Bascompte, J., Jordano, P., Melián, C.J. & Olesen, J.M. (2003). [The nested assembly of plant-animal mutualistic networks](#). *Proceedings of the National Academy of Sciences*, 100, 9383–9387.
- Berlow, E.L., Neutel, A.-M., Cohen, J.E., De Ruiter, P.C., Ebenman, B., Emmerson, M., *et al.* (2004). [Interaction strengths in food webs: Issues and opportunities](#). *Journal of Animal Ecology*, 73, 585–598.
- Blanchet, F.G., Cazelles, K. & Gravel, D. (2020). [Co-occurrence is not evidence of ecological interactions](#). *Ecology Letters*, 23, 1050–1063.
- Bohan, D.A., Vacher, C., Tamaddon-Nezhad, A., Raybould, A., Dumbrell, A.J. & Woodward, G. (2017). [Next-Generation Global Biomonitoring: Large-scale, Automated Reconstruction of Ecological Networks](#). *Trends in Ecology & Evolution*, 32, 477–487.
- Borrett, S.R. & Scharler, U.M. (2019). [Walk partitions of flow in Ecological Network Analysis: Review and synthesis of methods and indicators](#). *Ecological Indicators*, 106, 105451.
- Canard, E., Mouquet, N., Marescot, L., Gaston, K.J., Gravel, D. & Mouillot, D. (2012). [Emergence of Structural Patterns in Neutral Trophic Networks](#). *PLOS ONE*, 7, e38295.
- Catchen, M.D., Poisot, T., Pollock, L.J. & Gonzalez, A. (2023). The missing link: Discerning true from false negatives when sampling species interaction networks.
- Cazelles, K., Araujo, M.B., Mouquet, N. & Gravel, D. (2016). [A theory for species co-occurrence in interaction networks](#). *Theoretical Ecology*, 9, 39–48.

395 Choh, Y., Ignacio, M., Sabelis, M.W. & Janssen, A. (2012). [Predator-prey role reversals, juvenile experience](#)
396 [and adult antipredator behaviour](#). *Scientific Reports*, 2, 728.

397 Cirtwill, A.R., Eklöf, A., Roslin, T., Wootton, K. & Gravel, D. (2019). [A quantitative framework for](#)
398 [investigating the reliability of empirical network construction](#). *Methods in Ecology and Evolution*, 10,
399 902–911.

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

402 Elmasri, M., Farrell, M.J., Davies, T.J. & Stephens, D.A. (2020). [A hierarchical Bayesian model for predicting](#)
403 [ecological interactions using scaled evolutionary relationships](#). *The Annals of Applied Statistics*, 14,
404 221–240.

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

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

408 Fortuna, M.A. & Bascompte, J. (2006). [Habitat loss and the structure of plantanimal mutualistic networks](#).
409 *Ecology Letters*, 9, 281–286.

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

412 Gilman, S.E., Urban, M.C., Tewksbury, J., Gilchrist, G.W. & Holt, R.D. (2010). [A framework for community](#)
413 [interactions under climate change](#). *Trends in Ecology & Evolution*, 25, 325–331.

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

416 Golubski, A.J. & Abrams, P.A. (2011). [Modifying modifiers: What happens when interspecific interactions](#)
417 [interact?](#) *Journal of Animal Ecology*, 80, 1097–1108.

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

420 Gravel, D., Baiser, B., Dunne, J.A., Kopelke, J.-P., Martinez, N.D., Nyman, T., *et al.* (2019). [Bringing Elton](#)
421 [and Grinnell together: A quantitative framework to represent the biogeography of ecological interaction](#)
422 [networks](#). *Ecography*, 42, 401–415.

- 423 Gravel, D., Poisot, T., Albouy, C., Velez, L. & Mouillot, D. (2013). [Inferring food web structure from](#)
424 [predatorprey body size relationships](#). *Methods in Ecology and Evolution*, 4, 1083–1090.
- 425 Guimarães, P.R. (2020). [The Structure of Ecological Networks Across Levels of Organization](#). *Annual Review*
426 *of Ecology, Evolution, and Systematics*, 51, 433–460.
- 427 Guimerà, R. & Sales-Pardo, M. (2009). [Missing and spurious interactions and the reconstruction of complex](#)
428 [networks](#). *Proceedings of the National Academy of Sciences*, 106, 22073–22078.
- 429 Hortal, J., de Bello, F., Diniz-Filho, J.A.F., Lewinsohn, T.M., Lobo, J.M. & Ladle, R.J. (2015). [Seven Shortfalls](#)
430 [that Beset Large-Scale Knowledge of Biodiversity](#). *Annual Review of Ecology, Evolution, and Systematics*,
431 46, 523–549.
- 432 Jordano, P. (2016). [Sampling networks of ecological interactions](#). *Functional Ecology*, 30, 1883–1893.
- 433 Jordano, P., Bascompte, J. & Olesen, J.M. (2003). [Invariant properties in coevolutionary networks of](#)
434 [plantanimal interactions](#). *Ecology Letters*, 6, 69–81.
- 435 Kéfi, S., Berlow, E.L., Wieters, E.A., Joppa, L.N., Wood, S.A., Brose, U., *et al.* (2015). [Network structure](#)
436 [beyond food webs: Mapping non-trophic and trophic interactions on Chilean rocky shores](#). *Ecology*, 96,
437 291–303.
- 438 Kéfi, S., Berlow, E.L., Wieters, E.A., Navarrete, S.A., Petchey, O.L., Wood, S.A., *et al.* (2012). [More than a](#)
439 [meal... integrating non-feeding interactions into food webs](#). *Ecology Letters*, 15, 291–300.
- 440 Kéfi, S., Miele, V., Wieters, E.A., Navarrete, S.A. & Berlow, E.L. (2016). [How Structured Is the Entangled](#)
441 [Bank? The Surprisingly Simple Organization of Multiplex Ecological Networks Leads to Increased](#)
442 [Persistence and Resilience](#). *PLOS Biology*, 14, e1002527.
- 443 Klecka, J. & Boukal, D.S. (2014). [The effect of habitat structure on prey mortality depends on predator and](#)
444 [prey microhabitat use](#). *Oecologia*, 176, 183–191.
- 445 Kopelke, J.-P., Nyman, T., Cazelles, K., Gravel, D., Vissault, S. & Roslin, T. (2017). [Food-web structure of](#)
446 [willow-galling sawflies and their natural enemies across Europe](#). *Ecology*, 98, 1730–1730.
- 447 McCann, K. (2007). Protecting biostructure. *Nature*, 446, 29–29.
- 448 McCann, K.S. (2011). Food webs (MPB-50). In: *Food Webs (MPB-50)*. Princeton University Press.

- 449 McLeod, A., Leroux, S.J., Gravel, D., Chu, C., Cirtwill, A.R., Fortin, M.-J., *et al.* (2021). [Sampling and](#)
450 [asymptotic network properties of spatial multi-trophic networks](#). *Oikos*, n/a.
- 451 McLeod, A.M., Leroux, S.J. & Chu, C. (2020). [Effects of species traits, motif profiles, and environment on](#)
452 [spatial variation in multi-trophic antagonistic networks](#). *Ecosphere*, 11, e03018.
- 453 Ohlmann, M., Miele, V., Dray, S., Chalmandrier, L., O'Connor, L. & Thuiller, W. (2019). [Diversity indices for](#)
454 [ecological networks: A unifying framework using Hill numbers](#). *Ecology Letters*, 22, 737–747.
- 455 Olesen, J.M., Bascompte, J., Dupont, Y.L., Elberling, H., Rasmussen, C. & Jordano, P. (2010). [Missing and](#)
456 [forbidden links in mutualistic networks](#). *Proceedings of the Royal Society B: Biological Sciences*, 278,
457 725–732.
- 458 Pascual, M., Dunne, J.A. & Dunne, J.A. (2006). *Ecological Networks: Linking Structure to Dynamics in Food*
459 *Webs*. Oxford University Press, USA.
- 460 Petchey, O.L., Beckerman, A.P., Riede, J.O. & Warren, P.H. (2008). [Size, foraging, and food web structure](#).
461 *Proceedings of the National Academy of Sciences*, 105, 4191–4196.
- 462 Pilosof, S., Porter, M.A., Pascual, M. & Kéfi, S. (2017). [The multilayer nature of ecological networks](#). *Nature*
463 *Ecology & Evolution*, 1, 1–9.
- 464 Poisot, T., Canard, E., Mouillot, D., Mouquet, N. & Gravel, D. (2012). [The dissimilarity of species interaction](#)
465 [networks](#). *Ecology Letters*, 15, 1353–1361.
- 466 Poisot, T., Cirtwill, A.R., Cazelles, K., Gravel, D., Fortin, M.-J. & Stouffer, D.B. (2016). [The structure of](#)
467 [probabilistic networks](#). *Methods in Ecology and Evolution*, 7, 303–312.
- 468 Poisot, T., Stouffer, D.B. & Gravel, D. (2015). [Beyond species: Why ecological interaction networks vary](#)
469 [through space and time](#). *Oikos*, 124, 243–251.
- 470 Pollock, L.J., Tingley, R., Morris, W.K., Golding, N., O'Hara, R.B., Parris, K.M., *et al.* (2014). [Understanding](#)
471 [co-occurrence by modelling species simultaneously with a Joint Species Distribution Model \(JSDM\)](#).
472 *Methods in Ecology and Evolution*, 5, 397–406.
- 473 Pomeranz, J.P.F., Thompson, R.M., Poisot, T. & Harding, J.S. (2019). [Inferring predator-prey interactions in](#)
474 [food webs](#). *Methods in Ecology and Evolution*, 10, 356–367.
- 475 Pringle, R.M. & Hutchinson, M.C. (2020). [Resolving Food-Web Structure](#). *Annual Review of Ecology,*
476 *Evolution, and Systematics*, 51, 55–80.

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

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

480 Rooney, N. & McCann, K.S. (2012). [Integrating food web diversity, structure and stability](#). *Trends in Ecology*
481 *& Evolution*, 27, 40–46.

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

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

486 Schmeller, D.S., Julliard, R., Bellingham, P.J., Böhm, M., Brummitt, N., Chiarucci, A., *et al.* (2015). [Towards a](#)
487 [global terrestrial species monitoring program](#). *Journal for Nature Conservation*, 25, 51–57.

488 Singer, M.C. & McBride, C.S. (2012). [Geographic mosaics of species' association: A definition and an](#)
489 [example driven by plantinsect phenological synchrony](#). *Ecology*, 93, 2658–2673.

490 Stouffer, D.B., Rezende, E.L. & Amaral, L.A.N. (2011). [The role of body mass in diet contiguity and food-web](#)
491 [structure](#). *Journal of Animal Ecology*, 80, 632–639.

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

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

498 Tylianakis, J.M., Tschardtke, T. & Lewis, O.T. (2007). [Habitat modification alters the structure of tropical](#)
499 [hostparasitoid food webs](#). *Nature*, 445, 202–205.

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

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

- 504 Woodward, G., Benstead, J.P., Beveridge, O.S., Blanchard, J., Brey, T., Brown, L.E., *et al.* (2010). [Chapter 2 -](#)
505 [Ecological Networks in a Changing Climate](#). In: *Advances in Ecological Research*, Ecological Networks
506 (ed. Woodward, G.). Academic Press, pp. 71–138.
- 507 Woodward, G., Brown, L.E., Edwards, F.K., Hudson, L.N., Milner, A.M., Reuman, D.C., *et al.* (2012). [Climate](#)
508 [change impacts in multispecies systems: Drought alters food web size structure in a field experiment](#).
509 *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367, 2990–2997.