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.

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

Last revision: *June 14, 2023*

Introduction

Cataloging ecological interactions is a gargantuan task. Regardless of sampling effort, there are practical and biological constraints that hinder our ability to observe all interactions in nature, such as the spatial and temporal uncoupling of species and the large number of potential interactions in a community, of which the vast majority are rare (Jordano 2016). Documenting the location and timing of interactions becomes even more challenging when accounting for the spatiotemporal variability of ecological networks (Poisot *et al.* 2012, 2015). Indeed, it is now recognized that knowing the biological capacity of two species to interact is necessary but not sufficient for inferring their interaction at a specific time and space. For example, Golubski & Abrams (2011) presented many cases where trophic interactions in food webs depend on the presence or abundance of a third species (e.g., of a more profitable prey species). More generally, a handful of conditions must be satisfied for an interaction to be observed locally. First, both species must have overlapping geographic ranges, i.e. they must co-occur within the region of interest (Blanchet *et al.* 2020). Then, they must encounter locally. Probabilities of interspecific encounters are typically low, especially for rare species with low relative abundances (Canard *et al.* 2012). Finally, their traits must be locally compatible (Poisot *et al.* 2015). This includes their phenology (Olesen *et al.* 2010; Singer & McBride 2012), behavioral choices (Pulliam 1974; Choh *et al.* 2012) and phenotypes (Bolnick2011WhyInt; Stouffer *et al.* (2011); Gravel2013InfFooa). Environmental factors, such as temperature (Angilletta *et al.* 2004), drought (Woodward *et al.* 2012), climate change (Gilman *et al.* 2010; Woodward *et al.* 2010; Araujo *et al.* 2011), and habitat modifications (Tylianakis *et al.* 2007), contribute to this spatiotemporal variability of interactions by impacting species abundance and traits. In this context, it is unsurprising that our knowledge of ecological interactions remains limited (Hortal *et al.* 2015) despite extensive biodiversity data collection (Schmeller *et al.* 2015).

The recognition of the intrinsic variability of species interactions has led ecologists to expand their representation of ecological networks to include a probabilistic view of interactions (Poisot *et al.* 2016). As opposed to binary deterministic networks, in which interactions are either observed or not, probabilistic networks represent our degree of belief about the realization or feasibility of pairwise interactions at the local or regional scale, respectively. In other words, representing interactions probabilistically considers inherent uncertainties and observation errors associated with ecological data. In the broadest sense, binary networks are also a type of probabilistic network, in which the value of interactions is restrained to 0 (non-observed) or 1 (observed). In probabilistic networks, only forbidden interactions (Jordano *et al.* 2003; Olesen *et al.* 2010) 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 of 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 more adequate manipulation and integration 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 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 taxa 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 (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 analog 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 can be difficult to know when published probabilistic networks describe potential or realized interactions, or when so-called probabilities are in reality *interaction scores* (i.e., a type of non-probabilistic quantitative interactions). Likewise, probabilistic potential interactions are often used and interpreted as realized interactions (and conversely), which may generate misleading findings when analyzing these data. We believe that a better understanding of the differences, similarities, and relationships between these two 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 $A_{i,j} \sim \text{Bernoulli}(P(i \to j))$, with $P(i \to j)$ being the probability of interaction between i and j that characterizes our limited knowledge of the system and its intrinsic spatiotemporal variability. Depending on the type of networks (local or metaweb), the mathematical formulation and interpretation of stochastic parameters like $P(i \to j)$ 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 $P(i \to j)$ 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 $P(i \to j)$, informing us on the biological capacity of both taxa to interact and the environmental conditions that enabled them to interact locally.

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 follow a Poisson distribution $A_{i,j} \sim \text{Poisson}(\lambda_{i\to j})$ when predicting frequencies of interactions between pairs of nodes, with $\lambda_{i\to j}$ being the expected rate at which individuals of taxa i and j interact (e.g., the average number of prey j consumed 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 underlines 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 probabilistic local networks and 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-dependent). In local networks, edges commonly represent our degree of belief that two taxa interact in nature, 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 the probability of locally observing an interaction between a pair of taxa (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 during which interactions were sampled or for which they were predicted. Given that space and time are in reality continuous variables, the probability that an interaction occurs 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, the edge value could represent a probability density or a probability mass 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 allows us to measure an area A and duration t, which can be directly used in spatiotemporal analyses of ecological networks. For example, when studying network-area relationships (NAR, Galiana *et al.* 2018), we should expect local 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 $C_{i,j}$, which is usually a Boolean describing if the geographic distribution of both taxa overlaps within the study area. In local networks, the probability that the interaction is realized must be 0 when taxa do not co-occur, i.e. $P_N(i \rightarrow j | C = 0) = 0$. Co-occurrence can also be modeled probabilistically. In that case, it follows a Bernoulli distribution $C_{i,j} \sim \text{Bernoulli}(P_{i,j}(x, y))$, where the probability of co-occurrence $P_{i,j}(x, y)$ can be estimated using species distribution models (e.g., Pollock *et al.* 2014). More generally, the probability that two taxa interact locally can be obtained by the product of their probability of interaction given co-occurrence with their probability of co-occurrence:

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

Other important factors that can impact our estimation of interaction probabilities at the local scale 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, if a research team conducts a mesocosm experiment to estimate interaction probabilities between predators and prey with and without shelters, Ω would represent the presence or absence of these shelters. Like co-occurrence, Ω can also be modeled probabilistically when the stochasticity or uncertainty of environmental and biological factors is considered. 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 factors actually impacting ecological interactions.

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

$$P_N(i \to j|A, t, C, \Omega). \tag{2}$$

The local context in which probabilities are estimated and the variables that should be taken into consideration depend on the study system, the objective of the study, and the resources available to the researchers. In other words, these variables do not systematically need to be accounted for. However, when they are, they should be specified in the documentation of the data, preferentially in mathematical terms to avoid any confusion in their interpretation and to limit manipulation errors during their re-use. For example, ecologists should be explicit about their consideration of co-occurrence in their estimation of local interaction probabilities. Indeed, it is important to specify if probability values are conditional $P_N(i \to j|C = 1)$ or not $P_N(i \to j)$ on co-occurrence since this can significantly impact the interpretation and analysis of the data. In tbl. 1, we present a handful of studies of probabilistic ecological networks and their formulation of probabilistic interactions. This table illustrates the variety of definitions of probabilistic interactions found in the literature and emphasizes the need to carefully describe 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 formulas is included, with the variables used specified in parentheses.

Formula	Description	Studies
$P_M(i \to j)$	probability that the interaction is biologically feasible	
$P_N(i \to j)$	probability that the interaction is realized locally	
$P_N(i \to j\|A)$	probability that the interaction is realized locally given network area	
$P_N(i \to j\|t)$	probability that the interaction is realized locally given duration	

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

Metawebs: regional catalogs of interactions

Metawebs are networks of potential interactions that have been designed for broad spatial, temporal, and taxonomic scales (e.g, species food webs at the continental scale). They represent the probability that two taxa can biologically interact regardless of their co-occurrence and local environmental conditions. Indeed, potential interactions are by definition context-independent, i.e. they are not measured at a specific location and time. In contrast with probabilistic local networks, which represent the stochasticity of interactions occurring in nature, probabilistic metawebs measure our degree of belief in the capacity of two taxa to interact (i.e., the probability that their traits could support an interaction in the right conditions). In other words, potential interactions describe the probability that there exists at least one combination of phenotypes of taxa i and j that can interact with each other if they were to encounter. This probability of interaction, in a metaweb M, can be expressed as

$$P_M(i \rightarrow j), \tag{3}$$

which, in contrast with eq. 2, is not conditional on any spatial, temporal, or environmental variables (tbl. 1).

Starting from a selected set of taxa, which are usually distributed within a broad region of interest, metawebs can be built using different data sources, including literature review, fieldwork, and predictive models (e.g., the metaweb of Canadian mammals inferred by Strydom *et al.* 2022). Every pair of taxa that have confidently been observed to interact at least once can be given a probability of 1 (i.e., $P_M(i \rightarrow j) = 1$) since we know that they *can* interact. This is usually not the case in local probabilistic networks, in which probabilities usually remain stochastic (i.e., $P_N(i \rightarrow j) < 1$) after empirically observing interactions because of their intrinsic spatiotemporal variability. Similarly, although rare interactions typically have low probabilities in local networks, they can have high probabilities in metawebs if the traits of both taxa match. On the other hand, interactions that were never observed can have low probability values in both metawebs and local networks, going as low as 0 for forbidden links. However, 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 in metawebs, e.g. if they have been assessed for specific phenotypes, locations or time. Implementing a Bayesian framework, which updates prior probabilities with empirical data (e.g., Bartomeus *et al.* (2016), Cirtwill *et al.* (2019)), could improve our estimation of interaction probabilities in both systems.

Statistical behaviors of probabilistic networks

The differences in the mathematical formulations of probabilistic local and potential interactions can affect their statistical behaviors when applied to key ecological questions. These disparities must therefore be taken into account when analyzing probabilistic interaction data to prevent misleading results and minimize interpretation errors. Here we show four common applications of probabilistic interactions and compare the characteristics of local networks and metawebs using simulated and empirical data.

Taxonomic agglomeration and division of nodes

Probabilistic networks can be used to address a wide range of ecological questions based on their level of organization. For example, the assemblage of interactions across ecological scales can be studied using species-based networks, whereas clade-based networks can be used to study macroevolutionary processes (e.g., Gomez *et al.* 2010). Because our interpretation of the properties and dynamics of ecological networks depends on their taxonomic scale (Guimarães 2020), examining the phylogenetic scaling of network structure is also a promising research avenue. Analyzing the same system at different taxonomic scales can thus provide meaningful and complementary ecological information and is, in our perspective, best conducted using probabilistic networks.

Local networks are analogous to metawebs in their taxonomic scaling of interactions because only the nodes are defined taxonomically. In other words, the edge values of local networks (eq. 2) and metawebs (eq. 3) are not conditional on any taxonomic scale. It is the definition of the event itself (i.e., the interaction of the two taxa) that has a given phylogenetic scale, not their conditional variables. In both types of networks, shifting to a broader level of organization can be directly done using probabilities at finer scales. For example, if we have a network of n_A species from genus A and n_B species from genus B, we can calculate the probability that at least one species from genus A interact with at least one species from genus B as follows:

$$P(A \to B) = 1 - \prod_{i=1}^{n_A} \prod_{j=1}^{n_B} (1 - P(A_i \to B_j)), \tag{4}$$

where A_i and B_j are the species of the corresponding genus. Knowing that two species interact (i.e., $P(A_i \to B_j) = 1$) gives a probability of genus interaction of 1. Canard *et al.* (2012) built a species-based network from neutrally simulated interactions between individuals using a similar approach. In contrast, more sophisticated models need to be built when shifting to a finer level of organization. Indeed, knowing that two genera interact does not imply that all of their pairwise species combinations can also interact. One could, for example, build a finer-scale network by generating probabilities of species interactions by randomly sampling them from a beta distribution parametrized by the broader-scale network.

The biological interpretation of probabilistic interactions should not differ across a network even if it has heterogenous levels of organization, i.e. if different nodes are represented at different taxonomic scales (e.g., a network composed of species and trophic species). This is frequent in ecological networks where taxonomic resolution is typically low (Hemprich-Bennett *et al.* 2021; Vázquez *et al.* 2022). Indeed, these definitions should be based on probabilities of interactions between individuals, either locally or potentially. In local individual-based food webs, the probability that two individuals interact represents the degree of belief that one individual will actually consume the other. Similarly, in local species-based food webs, the probability that two species interact represents the degree of belief that at least one individual from the predator species consumes at least another individual from the prey species. Moreover, in local clade-based food webs, the probability that two clades interact represents the degree of belief that at least two species from these clades interact with each other or, equivalently, that at least two individuals from these clades interact with each other. Fundamentally, the taxonomic scaling of interactions is an aggregation of interactions between individuals into larger groups, which could be more or less homogeneous depending on the organisms and the study system. This type of scaling is analogous to the spatial and temporal scaling of interactions to the extent that these are different ways to aggregate individuals into broader nodes, either spatially, temporally, or taxonomically.

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.

A 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 $P_N(i \to j | C_{i,j} = 1, t = t_0) = 1 - e^{-\lambda t_0}$, which approaches 1 when $t \to \infty$. 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. Lets 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 $P_{N_0}(i \to j) = 1 - (1 - P_{N_1}(i \to j)) \times (1 - P_{N_2}(i \to j))$ 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.

Quantitative interactions can be converted to probabilistic interactions by normalizing.

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:

$$P_N(i \to j | A, t, C, \Omega) \leq P_M(i \to j). \tag{5}$$

Probabilistic metawebs give limited information on local networks. Additionally, our degree of belief that two taxa have the capacity to interact must be higher than the probability that they will actually interact (or that they will ever interact). This implies that the accumulated probability of realized interactions across all spatial, temporal, and environmental conditions must be lower than the probability of potential interaction, i.e.

$$\int_\Omega \int_A \int_t P_N(i \to j | A, t, \Omega) dt dA d\Omega \leq P_M(i \to j), \tag{6}$$

since both taxa might never co-occur or encounter locally.

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).

Conclusion

In this contribution, we underlined the importance of network metadata for adequately interpreting and manipulating probabilistic interaction data. Indeed, the mathematical representation of probabilities and their statistical behaviors depend on the type of interactions (local or potential) and the conditions in which they were estimated. We showed that probabilistic local networks and metawebs differ in their relationship to spatial and temporal scales, with potential interactions being scale-independent. In contrast, local interactions are measured in a specific context (e.g., in a given area, time, and biological and environmental conditions) and are conditional on taxa co-occurrence. These important conceptual differences bring to light the need to use probabilistic data with caution, for instance when generating binary network realizations across space and predicting local networks from metawebs. Clear metadata describing the type of interaction and the variables used in their estimation are required to ensure adequate data manipulation. Better data practices and rigorous foundations for probabilistic thinking in network ecology could enable more reliable assessments of the spatiotemporal variability and measurement uncertainty of biotic interactions.

Acknowledgment

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

References

Andrade-Pacheco, R., Rerolle, F., Lemoine, J., Hernandez, L., Meïté, 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 Puzzle1. *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 plantanimal 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., Tamaddoni-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.

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

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

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

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

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

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

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

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

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

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

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

Gomez, J.M., Verdu, M. & Perfectti, F. (2010). Ecological interactions are evolutionarily conserved across the entire tree of life. *Nature*, 465, 918–U6.

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

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

Gravel, D., Poisot, T., Albouy, C., Velez, L. & Mouillot, D. (2013). Inferring food web structure from predatorprey body size relationships. *Methods in Ecology and Evolution*, 4, 1083–1090.

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

Guimerà, R. & Sales-Pardo, M. (2009). Missing and spurious interactions and the reconstruction of complex networks. *Proceedings of the National Academy of Sciences*, 106, 22073–22078.

Hemprich-Bennett, D.R., Oliveira, H.F.M., Le Comber, S.C., Rossiter, S.J. & Clare, E.L. (2021). Assessing the impact of taxon resolution on network structure. *Ecology*, 102, e03256.

Hortal, J., de Bello, F., Diniz-Filho, J.A.F., Lewinsohn, T.M., Lobo, J.M. & Ladle, R.J. (2015). Seven Shortfalls that Beset Large-Scale Knowledge of Biodiversity. *Annual Review of Ecology, Evolution, and Systematics*, 46, 523–549.

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

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

Kéfi, S., Berlow, E.L., Wieters, E.A., Joppa, L.N., Wood, S.A., Brose, U., *et al.* (2015). Network structure beyond food webs: Mapping non-trophic and trophic interactions on Chilean rocky shores. *Ecology*, 96, 291–303.

Kéfi, S., Berlow, E.L., Wieters, E.A., Navarrete, S.A., Petchey, O.L., Wood, S.A., *et al.* (2012). More than a meal… integrating non-feeding interactions into food webs. *Ecology Letters*, 15, 291–300.

Kéfi, S., Miele, V., Wieters, E.A., Navarrete, S.A. & Berlow, E.L. (2016). How Structured Is the Entangled Bank? The Surprisingly Simple Organization of Multiplex Ecological Networks Leads to Increased Persistence and Resilience. *PLOS Biology*, 14, e1002527.

Klecka, J. & Boukal, D.S. (2014). The effect of habitat structure on prey mortality depends on predator and prey microhabitat use. *Oecologia*, 176, 183–191.

Kopelke, J.-P., Nyman, T., Cazelles, K., Gravel, D., Vissault, S. & Roslin, T. (2017). Food-web structure of willow-galling sawflies and their natural enemies across Europe. *Ecology*, 98, 1730–1730.

McCann, K. (2007). Protecting biostructure. *Nature*, 446, 29–29.

McCann, K.S. (2011). Food webs (MPB-50). In: *Food Webs (MPB-50)*. Princeton University Press.

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

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

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

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

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

Petchey, O.L., Beckerman, A.P., Riede, J.O. & Warren, P.H. (2008). Size, foraging, and food web structure. *Proceedings of the National Academy of Sciences*, 105, 4191–4196.

Pilosof, S., Porter, M.A., Pascual, M. & Kéfi, S. (2017). The multilayer nature of ecological networks. *Nature Ecology & Evolution*, 1, 1–9.

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

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

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

Pollock, L.J., Tingley, R., Morris, W.K., Golding, N., O'Hara, R.B., Parris, K.M., *et al.* (2014). Understanding co-occurrence bymodelling species simultaneously with a Joint Species DistributionModel (JSDM). *Methods in Ecology and Evolution*, 5, 397–406.

Pomeranz, J.P.F., Thompson, R.M., Poisot, T. & Harding, J.S. (2019). Inferring predatorprey interactions in food webs. *Methods in Ecology and Evolution*, 10, 356–367.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Woodward, G., Benstead, J.P., Beveridge, O.S., Blanchard, J., Brey, T., Brown, L.E., *et al.* (2010). Chapter 2 - Ecological Networks in a Changing Climate. In: *Advances in Ecological Research*, Ecological Networks (ed. Woodward, G.). Academic Press, pp. 71–138.

Woodward, G., Brown, L.E., Edwards, F.K., Hudson, L.N., Milner, A.M., Reuman, D.C., *et al.* (2012). Climate change impacts in multispecies systems: Drought alters food web size structure in a field experiment. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367, 2990–2997.