

# Deciphering probabilistic species interaction networks

Francis Banville<sup>1,2,3</sup>, Tanya Strydom<sup>1,3</sup>, Penelope S. A. Blyth<sup>4</sup>, Chris Brimacombe<sup>5</sup>, Michael Catchen<sup>3,6</sup>, Gabriel Dansereau<sup>1,3</sup>, Dominique Gravel<sup>2,3</sup>, Gracielle Higino<sup>2</sup>, Thomas Malpas<sup>4</sup>, Hana Mayall<sup>4</sup>, Kari Norman<sup>1</sup>, Timothée Poisot<sup>1,3</sup>

<sup>1</sup> Université de Montréal; <sup>2</sup> Université de Sherbrooke; <sup>3</sup> Quebec Centre for Biodiversity Science; <sup>4</sup> University of Sheffield; <sup>5</sup> University of Toronto; <sup>6</sup> McGill University

## Correspondance to:

Francis Banville — francis.banville@umontreal.ca

Representing species interactions probabilistically (how likely are they to occur?) as opposed to deterministically (are they occurring?) expresses uncertainties in our knowledge of interactions. The sources of uncertainty captured by interaction probabilities depend on the method used to evaluate them: uncertainty of predictive models, subjective assessment of experts, or empirical measurement of interaction spatiotemporal variability. However, guidelines for the estimation and documentation of probabilistic interaction data are still lacking. This is concerning because our understanding and analysis of interaction probabilities depend on their sometimes elusive definition and uncertainty sources. We review how probabilistic interactions are defined at different spatial scales, from local interactions to regional networks (metawebs), with a strong emphasis on host-parasite and trophic (predatory and herbivory) interactions. These definitions are based on the distinction between the realization of an interaction at a specific time and space (local) and its biological feasibility (regional). Through case studies using host-parasite interactions in Europe, we illustrate how these two network representations differ in their statistical properties, specifically: how local networks and metawebs differ in their spatial and temporal scaling of probabilistic interactions, but not in their taxonomic scaling. We present two approaches to inferring binary interactions from probabilistic ones that account for these differences and show that systematic biases arise when directly inferring local networks from metawebs. Our results underscore the importance of more rigorous descriptions of probabilistic species interaction networks that specify their type of interaction (local or regional) and uncertainty sources.

**Keywords:**  
ecological networks  
food webs  
metaweb  
probabilistic networks  
spatial scale  
species interactions  
temporal scale

1

## Introduction

As we try to navigate global biodiversity change, filling in knowledge gaps about biodiversity becomes instrumental to monitoring and mitigating those changes (Abrego *et al.* 2021; Gonzalez & Londoño 2022; Hortal *et al.* 2015). However, cataloging species, populations and, in particular, ecological interactions (e.g., predation, parasitism, and pollination) is a substantial challenge (Pascual *et al.* 2006; Polis 1991). There are methodological and biological constraints that hinder our ability to document species interactions, inevitably introducing uncertainties in our knowledge of unobserved interactions. For example, the spatial and temporal uncoupling of species (e.g., nocturnal and diurnal species coexisting in the same space without interacting because of different daily activity timings, Jordano 1987) and the large number of rare and cryptic interactions in a community, contribute to these difficulties in documenting interactions (Jordano 2016).

More generally, several 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 (Cazelles *et al.* 2016; Morales-Castilla *et al.* 2015). Second, they must have some probability of meeting (Poisot *et al.* 2015). Probabilities of interspecific encounters are typically low, especially for rare species with low abundances. While species' absolute abundances may impact interaction frequencies (Vázquez *et al.* 2007),

encounter probabilities are determined by their relative abundances (Canard *et al.* 2012). The probability that species meet also depends on their biology, such as their phenology (Olesen *et al.* 2010; Singer & McBride 2012) and discoverability (Broom & Ruxton 2005). Finally, when species do come into contact, an interaction occurs only if their traits are locally compatible (Poisot *et al.* 2015), including their phenotypes (Bolnick *et al.* 2011; Gravel *et al.* 2013; Stouffer *et al.* 2011) and behavior (Choh *et al.* 2012; Pulliam 1974). Interactions may also be influenced by a third species (e.g., a more profitable prey species, Golubski & Abrams 2011; Sanders & van Veen 2012). Documenting the location and timing of interactions becomes even more difficult when accounting for the spatiotemporal variability of ecological interactions (Poisot *et al.* 2012, 2015). Environmental factors, such as temperature (Angilletta *et al.* 2004), drought (Woodward *et al.* 2012), climate change (Araujo *et al.* 2011; Gilman *et al.* 2010; Woodward *et al.* 2010), and habitat modifications (Tylianakis *et al.* 2007), contribute to this spatiotemporal variability by impacting species abundance and traits. Even after satisfying all these conditions, there remains a possibility that the interaction does not occur locally, either due to the intricate nature of the system or simply by chance. If it does occur, it might go undetected, particularly if it happens infrequently. 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).

Knowing the biological capacity of two species to interact directly (via e.g., trophic interactions) is necessary but not sufficient for inferring their interaction at a specific time and space. The recognition of the intrinsic variability of species interactions has led ecologists to expand their representation of ecological networks (also known as ecological webs) to include a probabilistic view of interactions (Dallas *et al.* 2017; Fu *et al.* 2021; Poisot *et al.* 2016). This different perspective allows us to fill in the Eltonian shortfall (Hortal *et al.* 2015) by modeling the probability of occurrence of interactions (e.g., Gravel *et al.* 2019), which can be an important tool for directing efforts and taking action, especially in places where access and resources for research are scarce. The probabilistic representation of interactions is thus far limited to direct interactions, which are conceptually and mathematically analogous regardless of their biological type (e.g., predation and pollination). This is in contrast with indirect interactions (e.g., interspecific competition), which arise from distinct ecological processes and are often not directly observable (Kéfi *et al.* 2015, 2016). Representing direct interactions probabilistically enables us to capture the spatiotemporal variability of the aforementioned ecological processes and the uncertainties associated with their occurrence. As opposed to webs of binary deterministic interactions, in which interactions are regarded as either occurring or not, webs of probabilistic interactions, within a Bayesian framework, express our degree of belief (or confidence) regarding the occurrence or observation of interactions. Based on the scale at which they are estimated, probabilistic interactions may reflect our level of confidence in whether interactions will be observed, realized locally, or biologically feasible. As an illustration, we could outline a situation in which there is a 50% certainty that an interaction occurs 50% of the time, or that there is a 50% certainty that it simply occurs. Our level of confidence should be more definitive (approaching either 0 or 1) as we extend our sampling to a broader area and over a longer time period, thereby diminishing the uncertainty of the interactions (but not necessarily the estimation of their variability). In the broadest sense, binary interactions are also a type of probabilistic interaction, in which the numerical value of an interaction is restrained to 0 (non-occurring) or 1 (occurring). Yet, for the sake of clarity, we omit binary interactions from our discussion of probabilistic interactions in this contribution. In networks of probabilistic interactions, only forbidden interactions (i.e., interactions prohibited by biological traits or species absence, Jordano *et al.* 2003; Olesen *et al.* 2010) have a probability value of zero, provided that intraspecific trait variability is considered (Gonzalez-Varo & Traveset 2016).

By accounting for the uncertainty of interactions, networks of probabilistic interactions may provide a more realistic portrait of species interactions. However, how the uncertainty of pairwise interactions propagates to network structure (i.e., community-level properties driving the functioning, dynamics, and resilience of ecosystems, McCann 2007; McCann 2011; Proulx *et al.* 2005; Rooney & McCann 2012) remains to be elucidated. The application and development of computational methods in network ecology, often based on a probabilistic representation of interactions, can alleviate (and guide) the 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 could inform the identification of priority sampling locations of ecological webs where data collection would yield the most valuable information, thereby reducing errors. Optimization models for sampling locations have mostly found applications in biological systems that are not networks, such as identifying priority sampling sites for disease hotspots (Andrade-Pacheco *et al.* 2020), but there is substantial promise in applying them to probabilistic ecological interactions. Statistical models can also be used to generate predictions of ecological webs without prior knowledge of pairwise interactions, for instance using body size (Gravel *et al.* 2013; Petley *et al.* 2008), 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, which generate networks of probabilistic interactions by preserving chosen characteristics of the adjacency matrix of binary interactions while intentionally omitting others (Bascompte *et al.* 2003; Fortuna & Bascompte 2006), serve as other examples of common probabilistic interaction models. Null models can be used to produce underlying distributions of network measures for null hypothesis significance testing. Many measures have been developed to describe the structure (Poisot *et al.* 2016) and diversity (Godsoe *et al.* 2022; Ohlmann *et al.* 2019) of probabilistic interaction webs. 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 webs (Gilman *et al.* 2010).

Beyond methodological difficulties that may arise when assessing probabilistic interactions, a precise definition of probabilistic interactions appears to be lacking, making the estimation and use of these data more difficult. We aim to take a step back by outlining different ways in which probabilistic interactions are defined and used in network ecology. We distinguish two broad categories of probabilistic interaction networks that necessitate distinct approaches when applied to key ecological questions: local webs describing probabilities of realized interactions, and regional webs (metawebs) describing probabilities of potential interactions. We highlight the distinctions in the ecological meaning of these two representations and examine some of their properties and relationships (particularly with space, time, and between each other) through empirical case studies. Moreover, the lack of clear guidelines on the use of probabilistic interaction data is worrisome, as it affects both data producers and re-users who generate and manipulate these numbers. This is concerning because sampling strategies and decisions regarding network construction can affect our understanding of network properties (Brimacombe *et al.* 2023). There is currently no metadata standard that could guide the documentation of all types of probabilistic interactions (although see e.g., Salim *et al.* 2022 who discuss data standards for deterministic mutualistic webs). Well-defined metadata for probabilistic interactions would support more adequate manipulation and integration of interaction data from different sources and guard against possible misinterpretations arising from ambiguous definitions of probabilistic interaction networks. These metadata should outline the nature (i.e., local or regional) and type (e.g., predatory or pollination) of the interactions, provide information regarding the taxonomic level, identities, and characteristics (e.g., life stages) of the individuals involved in an interaction, present the mathematical formulation of probabilities, including clearly identified conditional variables (e.g., spatial and temporal scales), and describe the methods and contexts (e.g., location, time, environmental conditions) in which interactions were estimated. Inadequately documented probabilistic interaction data should be used with caution when analyzing ecological webs. Our observations and advice can be applied to most types of ecological networks representing direct interactions, from food webs to host-virus networks. Even though the measurement of interaction probabilities may differ between network types (e.g., estimations often based on frequencies of interactions in plant-pollinator networks), the broad principles underlying our findings remain relevant and applicable across diverse ecological contexts.

---

## Probabilistic representations of interactions

Consider a scenario where an avian predator species has just established itself in a northern habitat home to a small rodent species. Suppose these species have never co-occurred before, and as a result, their interaction has not been previously observed. What is the probability that the rodent is part of the diet of the avian predator, or put differently, what is the probability that they interact? Answering this question requires some clarification, as there are multiple ways to interpret and calculate interaction probabilities. We may calculate the probability that the traits of these species match, i.e. that the avian predator possesses the biological attributes to capture and consume the rodent. We may also calculate the probability that their traits support an interaction under the typical environmental conditions of the new habitat. For example, because avian predators hunt by sight, predation could be possible in the absence of snow but highly improbable when snow is present, as rodents may use it as a shelter to hide from predators. Finally, we may calculate the probability that the avian predator will consume the rodent at *that* particular location, for which the spatial and temporal boundaries need to be specified. The estimation of interaction probabilities, whether through predictive models or prior distributions, hinges on our comprehension of these probabilities and the specific ecological processes we aim to capture.

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, as these two types of interactions have distinct conceptual underpinnings and sources of uncertainty. A potential (regional) interaction is defined as

the biological capacity of two taxa to interact (i.e., the probability that they interact if they were to encounter each other and given sufficient time) whereas a realized (local) interaction is the occurrence or observation of this interaction in a well-defined space and time (i.e., the probability that they interact locally). For two co-occurring taxa and over infinite time, the probability of local interaction is equivalent to the probability of regional (potential) interaction. Our discussion focuses on the occurrence of local interactions rather than their direct observation, as one of the primary goals of describing them probabilistically is to characterize our uncertainty regarding their actual realization. We use the terms *metaweb* (Dunne 2006) to designate regional webs of potential interactions and *local webs* (Poisot *et al.* 2012) for those of realized interactions. Metawebs are the network analogs of the species pool, where local webs originate from a subset of both species (nodes) and interactions (edges) of the regional metaweb (Saravia *et al.* 2022). Without clear documentation, it can be challenging to know if published probabilistic interaction webs describe local or regional interactions (Tbl. 1 provides examples of studies employing both types of probabilistic interaction networks), or if so-called probabilities are in reality a form of interaction score (i.e., non-probabilistic quantitative interactions analogous to interaction strengths, which we won't cover further). When probabilistic regional interactions are used and interpreted incorrectly as local interactions (and conversely), this may generate misleading findings during data analysis. We believe that a better understanding of the differences, similarities, and interconnections between these two probabilistic representations of ecological interactions would alleviate interpretation errors (e.g., when studying network-area relationships with metawebs or local webs) and facilitate a more adequate use of interaction data. In the following sections, we delve into these distinctions as we scale up from pairwise interactions to higher-level representations of ecological networks, i.e. local webs and metawebs.

**2.1. Pairwise interactions: the building blocks of ecological networks** Local webs and metawebs, like any type of network, are made of nodes and edges that may be represented at different levels of organization and precision. The basic units of ecological networks are individuals that interact with each other (e.g., by predation in food webs, Elton 2001), forming individual-based networks (Melián *et al.* 2011). The aggregation of these individuals into more or less homogeneous groups (e.g., populations, species, families, feeding guilds) allows us to represent nodes at broader taxonomic scales, which affects our interpretation of the properties of these systems (Guimarães 2020; Hemprich-Bennett *et al.* 2021).

Edges linking nodes may 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  $j$ . In an adjacency matrix  $B$  of binary interactions, the presence or absence of an interaction  $B_{i \rightarrow j}$  between two taxa can be viewed as the result of a Bernoulli trial  $B_{i \rightarrow j} \sim \text{Bernoulli}(P(B_{i \rightarrow j}))$ , with  $P(B_{i \rightarrow j})$  being the probability of interaction. This interaction probability characterizes our limited ecological knowledge and/or the intrinsic spatiotemporal variability of interactions. It may be estimated through predictive models (e.g., those based on biological traits and species abundances) or expert (prior) knowledge about the interaction. In networks of probabilistic interactions, the edge values  $P(B_{i \rightarrow j})$  are probabilistic events whose only two possible outcomes are the presence ( $B_{i \rightarrow j} = 1$ ) or absence ( $B_{i \rightarrow j} = 0$ ) of an interaction between each pair of nodes. Depending on the type of probabilistic interaction network (local or metaweb), the mathematical formulation and interpretation of stochastic parameters like  $P(B_{i \rightarrow j})$  can be linked to environmental and biological factors such as species abundance, traits, area, and time, for example using logistic regression with continuous explanatory variables. Predicting the number of local webs in which the interaction between two given taxa occurs can be achieved by using a Binomial distribution, assuming a constant interaction probability and independence between networks (trials). When considering uncertainties around the estimation of  $P(B_{i \rightarrow j})$  or its spatiotemporal variability, a Beta distribution may be used to represent the relative likelihood of different probability values. For example, when calculating the probability of interaction between two taxa based on their local abundances, any spatiotemporal fluctuations in their abundances would introduce variability in the interaction probability at the local scale. If we take into account the stochasticity of the interaction probability, a Beta-Binomial distribution can be used to predict the number of networks in which the interaction occurs. Empirically observing an interaction between two taxa at a given location and time provides important information that can be used to update previous estimates of  $P(B_{i \rightarrow j})$ , informing us on the biological capacity of both taxa to interact and the environmental conditions that enabled them to interact locally.

Even though binary interaction webs constitute a highly valuable source of ecological information (Pascual *et al.* 2006), they overlook important factors regarding interaction strengths. Represented in a quantitative adjacency matrix  $W$  as numbers not confined to the  $[0, 1]$  range, interaction strengths better describe the energy flows, demographic impacts or frequencies of interactions between nodes (Berlow *et al.* 2004; Borrett & Scharler 2019), with  $W_{i \rightarrow j}$  being a natural  $\mathbb{N}$  or real  $\mathbb{R}$  number depending on the measure. For example, they may represent local interaction rates between pairs of taxa (e.g., the flower-visiting rates of pollinators

in a mutualistic network, Herrera 1989). When interaction strengths characterize predation pressure on prey taxa in food webs, they can serve as good estimators of the parameters describing species interactions in a Lotka-Volterra model (e.g., Emmerson & Raffaelli 2004). The extra amount of ecological information in quantitative networks typically comes at a cost of greater sampling effort and data volume in predictive models (Strydom *et al.* 2021), which can lead to relatively high levels of uncertainties when inferring quantitative webs with limited data.

Just like binary interaction networks, the uncertainty and spatiotemporal variability of interaction strengths can be represented probabilistically. However, the need to estimate the probability distribution of all possible values of interaction strengths can make the inference of probabilities more challenging in quantitative webs compared to binary interaction webs, which require only one probability estimate for each interaction. Interaction strengths can follow various probability distributions depending on the measure used. For instance, they can follow a Poisson distribution  $W_{i \rightarrow j} \sim \text{Poisson}(\lambda_{i \rightarrow j})$  when predicting frequencies of interactions between pairs of nodes, with  $\lambda_{i \rightarrow j}$  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  $i$  in a given time period). The Poisson distribution can also be 0-inflated after initially modeling non-interacting taxa (e.g., Boulangeat *et al.* 2012 employ a 0-inflated model to analyze species abundance following the modeling of species presence and absence), which constitute the majority of taxa pairs in most local webs due to their typically high sparseness (Jordano 2016). Because of the methodological difficulties typically encountered when building deterministic quantitative webs (which are only partially mitigated by models such as Ecopath, Plagányi & Butterworth 2004), binary interaction webs, which are easier to sample (Jordano 2016) and predict (Strydom *et al.* 2021), have been more frequently studied and modeled. Moreover, most published networks of probabilistic interactions and methods describe probabilistic interactions whose outcome is binary (whether interaction probabilities are regarded as constant or variable, e.g. represented by a Beta distribution), which underlines the need for better guidelines regarding the interpretation and manipulation of these types of webs first. For these reasons, our primary focus in this contribution will be on addressing the challenges in interpreting and using interaction probabilities in Bernoulli distributions, in both local webs and metawebs.

**2.2. Local webs: communities interacting in space and time** Local webs of probabilistic interactions describe how likely taxa are to interact at a given location and time period. Local interactions are contingent upon the environmental and biological conditions of the community. In local webs, edges commonly represent our degree of belief that two taxa interact in nature, but can also represent the probability of *observing* this interaction (Catchen *et al.* 2023). For example, Gravel *et al.* (2019) analyzed local European trophic webs of willow-galling sawflies and their natural enemies depicting binary interactions, all referenced in space and time, to infer the probabilities of locally observing interactions between co-occurring species. This was achieved by situating local webs within the context of environmental variables, i.e. by including temperature and precipitation as conditional variables in some of their models.

We define space as the collection of geographic coordinates  $(x, y, z)$ , with  $(x, y)$  representing longitude and latitude coordinates, and  $z$  denoting either altitudes or depths. These point coordinates delineate the spatial boundaries of the system, which may be portrayed as a polyhedron. Ecological interactions may exhibit variations along latitudinal and altitudinal gradients, as evidenced by changes in hummingbird-plant interactions (Weinstein & Graham 2017a, b) and mosquito biting rates (e.g., Kulkarni *et al.* 2006) at different elevations. In contrast, time is defined as the specific time period within which interactions were either observed or predicted. Even though space and time are continuous variables that should yield probability *densities* of interactions (i.e., relative likelihoods of interactions occurring at infinitesimal locations and instants in time), these definitions enable them to be conceptualized as distinct patches and time segments. Treating space and time as discrete dimensions aligns with the common sampling methods of ecological networks and provides actual probabilities of interactions, which can be obtained by integrating probability densities over space and time. Furthermore, we can quantify both an area  $A$  and a duration  $t$ , which can be readily used in spatiotemporal analyses of ecological networks. For example, when studying network-area relationships (NAR, Galiana *et al.* 2018), we anticipate that local probabilities of interactions scale positively with area and duration because taxa have more opportunities to interact as these dimensions expand.

The probability that two taxa  $i$  and  $j$  interact in a local web  $L_{x,y,z,t}$  (spatial and temporal subscripts hereafter omitted or replaced by the shorter subscript  $l$  for clarity) can be conditioned on many environmental and biological factors. In addition to network area (or volume) and duration, they may be conditioned on taxa co-occurrence  $X_{i,j,l}$ , which is usually a Boolean describing if the geographic distributions of both taxa overlap within the study area. Co-occurrence may be modeled probabilistically, in which case it may conform to a Bernoulli distribution  $X_{i,j,l} \sim \text{Bernoulli}(P(X_{i,l}, X_{j,l}))$ , where  $X_{i,l}$  and  $X_{j,l}$  are the local occurrences (presence / absence) of both taxa. The probability of co-occurrence  $P(X_{i,l}, X_{j,l})$  can be estimated through the application of joint species distribution models (e.g., Pollock *et al.* 2014), potentially taking into account biotic

interactions (Staniczenko *et al.* 2017). Given that the probability that two non-co-occurring taxa interact locally is zero (i.e.,  $P(L_{i \rightarrow j}|X_{i,j,l} = 0) = 0$ ), the probability of local interaction can be obtained by multiplying the probability of interaction given co-occurrence with the probability of co-occurrence:

$$P(L_{i \rightarrow j}) = P(L_{i \rightarrow j}|X_{i,j,l} = 1) \times P(X_{i,l} = 1, X_{j,l} = 1). \quad (1)$$

Other important factors that can impact interaction probabilities at the local scale are taxa local abundances  $N_{i,l}$  and  $N_{j,l}$ , which affect encounter probabilities (Canard *et al.* 2012), and local traits distributions  $T_{i,l}$  and  $T_{j,l}$  (e.g., body mass, presence of given anatomical features, host resistance), which determine the ability of individuals to interact after encountering each other (Poisot *et al.* 2015). Moreover, local interactions may be conditional on local environmental factors such as temperature (Angilletta *et al.* 2004), precipitation (Woodward *et al.* 2012), habitat structure (Klecka & Boukal 2014), and presence of other taxa in the network (Kéfi *et al.* 2012; Pilosof *et al.* 2017), as described above. Here, we use the variable  $E_l$  to describe the local environmental 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 (a place that offers refuge and protection for prey, shielding them from predators),  $E_l$  would represent the presence or absence of these shelters. Like co-occurrence,  $E_l$  can also be modeled probabilistically when the stochasticity or uncertainty of environmental factors is considered. In sum,  $E_l$  represents all environmental variables that were taken into consideration when measuring interaction probabilities and is, therefore, a subset of all environmental factors impacting ecological interactions. Finally, local interaction probabilities may be conditioned on higher-level properties of the network, which we denote by  $f(L)$ . Many topological null models (i.e., statistical models that randomize interactions by retaining certain properties of the network while excluding others) provide interaction probabilities from selected measures of network structure, such as connectance (Fortuna & Bascompte 2006) and the degree distribution (Bascompte *et al.* 2003).

The probability that two taxa  $i$  and  $j$  interact in a local web  $L$  can thus be conditioned on their co-occurrence  $X_{i,j,l}$  (or more explicitly on their occurrences  $X_{i,l}$  and  $X_{j,l}$ ), local abundances  $N_{i,l}$  and  $N_{j,l}$ , local traits distributions  $T_{i,l}$  and  $T_{j,l}$ , local environmental conditions  $E_l$ , network area (or volume)  $A$ , time interval  $t$ , and network properties  $f(L)$ . Although these variables correspond to distinct ecological inquiries or mechanisms related to ecological interactions, they may covary with each other, such as the possible dependence of  $X_{i,j,l}$  and  $E_l$  on spatial and temporal scales. When estimating interaction probabilities using e.g. a generalized linear model with multiple explanatory variables that might not be independent, it may become important to address collinearity. In such a case, to mitigate this issue, it may be necessary to use variable selection techniques before fitting the model to data. The probability that a local interaction is realized is described by the following expression when all these conditional variables are included:

$$P(L_{i \rightarrow j}|X_{i,l}, X_{j,l}, N_{i,l}, N_{j,l}, T_{i,l}, T_{j,l}, E_l, A, t, f(L)) \quad (2)$$

The representation of the local context in which probabilities are estimated and the variables that should be taken into consideration depend on the study system, the objectives of the study, and the resources available to the researchers. In other words, these variables do not systematically need to be accounted for. For example, in Gravel *et al.* (2019), for the purpose of model comparison, some models incorporated environmental variables as conditions of interaction probabilities, while others did not. When accounted for, these variables should be clearly described in the documentation of the data (Brimacombe *et al.* 2023), preferentially in mathematical terms to avoid any confusion in their interpretation and to limit manipulation errors during their re-use. For instance, 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  $P(L_{i \rightarrow j}|X_{i,j,l} = 1)$  or not  $P(L_{i \rightarrow j})$  on co-occurrence since this can significantly impact the interpretation and analysis of the data. In Tbl. 1, we present examples of studies that used these different formulations of probabilistic interactions and conditional variables. We have included the probability of empirically observing an interaction that is realized locally  $P(O_{i \rightarrow j}|L_{i \rightarrow j})$  to underscore the distinction between local observations and actual realizations of interactions, even though the focus of this manuscript is not on the observation of interactions.

**Table 1** Mathematical expressions of probabilistic interactions. The probability of interaction between two taxa  $i$  and  $j$  is interpreted differently in a local web  $L$  of realized interactions and a metaweb  $M$  of potential interactions (representing the *biological* feasibility of interactions). Each expression emphasizes a different conditional variable, the ellipsis serving as a placeholder for other variables not explicitly stated in the expression. The outcome of each of these probabilistic events, along with common uncertainty sources (often stemming from the model used for estimation), is presented alongside examples of studies that employed them (with specific variables indicated in parentheses, when applicable). The study marked with an asterisk has been conducted on binary interaction webs. Note that we also included the expression for a local web  $O$  of observed interactions and a metaweb  $M^*$  of potential interactions (representing the *ecological* feasibility of interactions) as they were mentioned in the main text, even though they are not the focus of this manuscript.

Expression	Type	Outcome	Uncertainty sources	Reference
$P(L_{i \rightarrow j} X_{i,l}, X_{j,l}, \dots)$	local	realization of the interaction given taxa co-occurrence	spatiotemporal variability	Gravel <i>et al.</i> (2019)
$P(L_{i \rightarrow j} N_{i,l}, N_{j,l}, \dots)$	local	realization of the interaction given taxa abundances	neutral models	Canard <i>et al.</i> (2014)
$P(L_{i \rightarrow j} T_{i,l}, T_{j,l}, \dots)$	local	realization of the interaction given local traits	trait matching models	Gravel <i>et al.</i> (2016)
$P(L_{i \rightarrow j} E_l, \dots)$	local	realization of the interaction given local environmental conditions	environmental-based models	Gravel <i>et al.</i> (2019) (temperature and precipitation)
$P(L_{i \rightarrow j} A, \dots)$	local	realization of the interaction in a given area or volume	spatial models	Galiana <i>et al.</i> (2018)*
$P(L_{i \rightarrow j} t, \dots)$	local	realization of the interaction during a given time period	temporal models	Weinstein & Graham (2017a)
$P(L_{i \rightarrow j} f(L), \dots)$	local	realization of the interaction given network structure	topological null models	Fortuna & Bascompte (2006) (connectance)
$P(L_{i \rightarrow j} M_{i \rightarrow j}, \dots)$	local	realization of the interaction given that the taxa can biologically interact	spatiotemporal variability	this study
$P(O_{i \rightarrow j} L_{i \rightarrow j}, \dots)$	local	observation of the interaction given that it is realized locally	sampling model	Catchen <i>et al.</i> (2023)
$P(M_{i \rightarrow j} T_i, T_j)$	regional	biological feasibility of the interaction given regional traits (non-forbiddenness)	trait matching models	Strydom <i>et al.</i> (2022)
$P(M_{i \rightarrow j}^* T_i, T_j, E)$	regional	ecological feasibility of the interaction given regional traits and environmental conditions	trait matching and environmental-based models	this study

When testing or employing multiple models to estimate local interaction probabilities, rather than selecting a single model that best fits the data, model averaging may be used to enhance our estimations. In this approach, the weighting is based on the likelihood of each model. Model weights represent the probability that each model is the most suitable for explaining the data, and may be measured using Akaike weights (Burnham & Anderson 2004; Wagenmakers & Farrell 2004). For instance, considering two models  $mod_1$  and  $mod_2$  with respective probabilities (or weights)  $P(mod_1)$  and  $P(mod_2)$ , the average probability of interaction  $P(L_{i \rightarrow j}| \dots)$  can be calculated as follows:

$$P(L_{i \rightarrow j}| \dots) = P(L_{i \rightarrow j}|mod_1, \dots) \times P(mod_1) + P(L_{i \rightarrow j}|mod_2, \dots) \times P(mod_2), \quad (3)$$

where the ellipsis serves as a placeholder for the conditional variables incorporated in these models.

**2.3. 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). Potential interactions describe the biological capacity of taxa to interact, which is typically assessed at the regional scale. Metawebs of probabilistic interactions are particularly useful in situations where there

is high uncertainty in the ability of taxa to interact if they were to encounter each other. This uncertainty frequently arises due to insufficient interaction data, especially for taxa that have not yet been observed to co-occur, and uncertainties in trait-matching models. As data accumulates, interactions in metawebs should tend towards binarity, approaching probability values of 0 (repeatedly failing to observe an interaction between two co-occurring taxa) and 1 (observing an interaction at least once). The extent of sampling effort thus influences our evaluation of probabilities of potential interactions, as sampling over a larger area or for a longer duration enables us to capture a greater number of regional interactions (McLeod *et al.* 2021). However, in contrast with local webs of probabilistic interactions, which describe interaction stochasticity in local natural settings, regional interactions are not evaluated for any particular local context. Although *neutrally* forbidden interactions (i.e., improbable interactions between rare species, Canard *et al.* 2012) tend to have low probability values in local webs, they may have higher probabilities in the metaweb (i.e., when species' biological traits can support an interaction if they were to encounter each other). Likewise, non-co-occurring taxa may have a non-zero probability of interaction in the metaweb.

Potential interactions describe what we refer to as the *biological* feasibility of interactions, which is based solely on the regional traits distributions  $T_i$  and  $T_j$  of taxa  $i$  and  $j$ , respectively. A probability of potential interaction in a metaweb  $M$  describing the biological feasibility of interactions may be expressed as:

$$P(M_{i \rightarrow j} | T_i, T_j), \quad (4)$$

which, in contrast with local webs, is not conditioned on any spatial, temporal, co-occurrence or environmental variables (Tbl. 1). Regional traits may differ from local traits  $T_{i,l}$  and  $T_{j,l}$ , which may vary spatially and temporally due to phenotypic plasticity (Berg & Ellers 2010). The biological feasibility of interactions expresses our degree of belief that there exists at least one combination of phenotypes that could support an interaction if they were to encounter each other, assuming they had infinite time to interact. Evaluating this probability is conducted without incorporating the environmental conditions under which they encounter each other into the model. It is the complement of the probability  $P(F_{i \rightarrow j} | T_i, T_j)$  of forbidden interactions based uniquely on biological traits (i.e., the probability that their traits do not support an interaction):

$$P(M_{i \rightarrow j} | T_i, T_j) = 1 - P(F_{i \rightarrow j} | T_i, T_j). \quad (5)$$

For example, let  $i$  be a western diamondback rattlesnake (*Crotalus atrox*) and  $j$ , a wood lemming (*Myopus schisticolor*). These two taxa never co-occur, the rattlesnake being adapted to warm regions of North America (Castoe *et al.* 2007) and the lemming, to northern habitats of Eurasia (Fedorov *et al.* 2008). As we lack direct observations of an interaction between these two species, we have to rely on expert knowledge or trait-matching models to estimate their probability of potential interaction. To accurately estimate this probability using trait-matching models, it is crucial to ensure that the set of traits considered reflects the overall traits distributions of both taxa. We could for instance consider their average body mass and the average phylogenetic proximity of lemmings to rattlesnakes' prey. Doing so, we might find a high probability of potential interaction based on these traits.

The biological feasibility of interactions should not be confused with what we refer to as the *ecological* feasibility of interactions. A probability of potential interaction in a metaweb  $M^*$  describing ecological feasibility of interactions may be expressed as:

$$P(M_{i \rightarrow j}^* | T_i, T_j, E), \quad (6)$$

where  $E$  is the environmental conditions under which potential interactions are evaluated (Tbl. 1). Unlike  $E_l$ , these environmental conditions do not represent conditions occurring at specific locations. Ecological feasibility represents the probability that two taxa interact if they were to encounter each other under given environmental conditions, assuming they had infinite time to interact. Incorporating environmental conditions into a trait-matching model may be important when there is high covariation between the environment and biological traits. For instance, in our example involving rattlesnakes and lemmings, the probability of potential interaction between these two species may be low in most environmental conditions. Western diamondback rattlesnakes may be unactive under low temperatures (Kissner *et al.* 1997), whereas wood lemmings may have low tolerance to high temperatures (Kausrud *et al.* 2008). The probability that an interaction is ecologically feasible is always lower than the probability that it is biologically feasible, even across all environmental conditions:

$$\int_E P(M_{i \rightarrow j}^* | T_i, T_j, E) dE \leq P(M_{i \rightarrow j} | T_i, T_j). \quad (7)$$

This is because the biological feasibility of an interaction is a prerequisite for its ecological feasibility. Biological feasibility is necessary but not sufficient for an interaction to be ecologically feasible. Our discussion of metawebs will focus on the biological feasibility of interactions since most methods developed for inferring probabilities of regional interactions do not explicitly take into account environmental conditions (e.g., Strydom *et al.* 2022).

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 (e.g., Maiorano *et al.* 2020), aggregated local web data (e.g., Gravel *et al.* 2019; Saravia *et al.* 2022), trait-matching models Strydom *et al.* (2022), and expert knowledge. 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 differs from local webs of probabilistic interactions, where interaction events may remain stochastic (i.e.,  $P(L_{i \rightarrow j}) < 1$ ) even after empirically observing interactions due to their spatiotemporal variability. Interactions that were never observed typically have low probability values in local webs and vary from low to high values in metawebs, contingent upon taxa traits distributions (reaching 0 for forbidden links).

When incorporating local network data to estimate probabilities of regional interactions, repeatedly failing to observe an interaction between two co-occurring taxa (i.e.,  $P(O_{i \rightarrow j} = 0)$ ) should decrease the probability that the interaction is biologically feasible. Using Bayes' theorem, the probability that the interaction is biologically feasible given that it was never observed locally,  $P(M_{i \rightarrow j} = 1|O_{i \rightarrow j} = 0, \dots)$ , may be calculated as follows:

$$P(M_{i \rightarrow j} = 1|O_{i \rightarrow j} = 0, \dots) = \frac{P(O_{i \rightarrow j} = 0|M_{i \rightarrow j} = 1, \dots) \times P(M_{i \rightarrow j} = 1|\dots)}{P(O_{i \rightarrow j} = 0|\dots)}. \quad (8)$$

The reduction in the probability of regional interaction after considering that it was never observed locally (i.e.,  $P(M_{i \rightarrow j} = 1|O_{i \rightarrow j} = 0, \dots) < P(M_{i \rightarrow j} = 1|\dots)$ ) occurs because  $P(O_{i \rightarrow j} = 0|M_{i \rightarrow j} = 1, \dots)$  must be lower than  $P(O_{i \rightarrow j} = 0, \dots)$ , i.e. there is a higher chance of observing an interaction when we know it is biologically feasible.

Many observations of interactions are false positives because of observation errors due to taxonomic misidentifications and ecological misinterpretations (e.g., due to cryptic species and interactions, Pringle & Hutchinson 2020). Likewise, forbidden interactions may be false negatives, e.g. if they have been evaluated based on unrepresentative or incomplete traits distributions. Employing Bayesian models could prove valuable when estimating interaction probabilities in metawebs (e.g., Bartomeus *et al.* 2016; Cirtwill *et al.* 2019). This improvement is achieved by updating prior information regarding the feasibility of interactions (e.g., experts' prior assessments of interaction probabilities) with empirical data on interactions and traits.

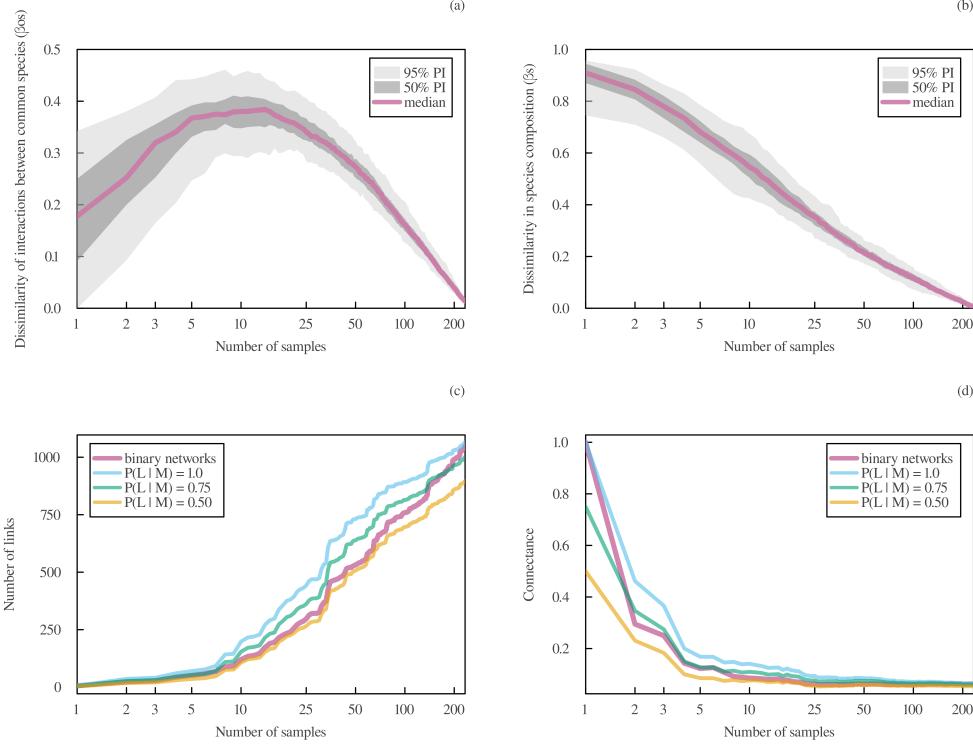
## 3

---

### Properties of probabilistic interaction webs

Local webs and metawebs of probabilistic interactions differ in their type of interactions (i.e., realized or potential) and in the conditional variables upon which interaction values depend. These differences are significant as they influence the characteristics of probabilistic interaction networks. Neglecting to consider them may result in misleading results and interpretation errors when analyzing the properties of probabilistic interaction webs, which could be particularly problematic when addressing crucial ecological questions about networks. Here we compare the characteristics of local webs and metawebs through the presentation of five applications of probabilistic interactions: (1) describing the dissimilarity between local webs and metawebs of binary and probabilistic interactions (2) describing their spatial and temporal scaling, (3) describing their taxonomic scaling, (4) sampling for binary interaction webs, and (5) reconstructing local webs of probabilistic interactions from metawebs. All code and data to reproduce these analyses are available at the Open Science Framework (TBD).

**3.1. Application 1: Dissimilarity of binary and probabilistic interaction webs** We use the collection of tripartite host-parasite networks sampled across Europe, created by Kopelke *et al.* (2017), in the following case studies. This dataset contains well-resolved binary local interactions between willows (52 species), willow-galling sawflies (96 species), and their parasitoids (126 species). Given its replicated webs spanning large spatiotemporal scales, this dataset is well-suited for analyzing a variety of ecological hypotheses and processes. Out of a total of 374 local webs, we retained those containing at least 5 species, resulting in a set of



**Figure 1 Network accumulation curves.** (a) Dissimilarity of interactions between common species and (b) dissimilarity in species composition between aggregated local webs and the metaweb of binary host-parasite interactions. Aggregated local webs were obtained by sequentially and randomly selecting a number of local webs of binary interactions and aggregating both their species and interactions. In both panels, the colored line represents the median dissimilarity across simulations and the grey areas cover the 50% and 95% percentile intervals. (c) Scaling of the number of links and (d) scaling of connectance with the number of sampled binary and probabilistic local interaction webs. For a better comparison with binary interactions, local webs of probabilistic interactions were derived from the metaweb of probabilistic interactions with a false positive and false negative rate of zero. A specific value of  $P(L_{i \rightarrow j}|M_{i \rightarrow j})$  (the local probability of interaction among potentially interacting species, indices omitted in the figure for simplicity) was used for all local webs within a particular curve. Aggregated local webs of probabilistic interactions were obtained by sequentially and randomly selecting a number of local webs and aggregating both their species and interactions (with the value of  $P(L_{i \rightarrow j}|M_{i \rightarrow j})$  adjusting according to Eq. 11).

233 georeferenced local webs (networks sampled within areas of 0.1 to 0.3 km<sup>2</sup> during June and/or July spanning 29 years). We built a metaweb of binary interactions by aggregating all local interactions, which gave us a regional web composed of 274 species and 1080 interactions. In Fig. 1, we show how the dissimilarity between the metaweb and the aggregated local webs varies with the number of sampled local webs. To do so, we randomly selected one local web of binary interactions and sequentially sampled additional webs while aggregating both their species and interactions. Next, we compared the metaweb and the aggregated local webs using the dissimilarity of interactions between common species ( $\beta_{OS}$ , Fig. 1a) and the dissimilarity in species composition ( $\beta_S$ , Fig. 1b) (Poisot *et al.* 2012). We repeated this sampling process one hundred times and highlighted the median dissimilarity values across simulations, as well as the 50% and 95% percentile intervals. This shows that networks of local interactions are highly dissimilar from the metaweb, both in terms of species and interactions, especially when only a limited number of sites has been sampled. Both dissimilarity indices were calculated based on the number of items shared by the two webs ( $c_{LM}$ ) and the number of items unique to the metaweb ( $u_M$ ) and to the aggregated local web ( $u_L$ ). The  $\beta_S$  dissimilarity index uses species (nodes) as items being compared, while the  $\beta_{OS}$  index assesses dissimilarity based on interactions between shared species (Poisot *et al.* 2012). Both indices were calculated following the  $\beta_W$  index of Whittaker (1960):

$$\beta_W = \frac{c_{LM} + u_L + u_M}{(2c_{LM} + u_L + u_M)/2} - 1. \quad (9)$$

Next, we investigate the differences in scaling between networks of probabilistic or binary interactions (Fig. 1c-d). We converted these binary interaction networks into probabilistic ones using models based on simple assumptions. Our models are not designed to estimate the exact values of probabilistic interactions. Instead, their purpose is to create plausible networks that serve as illustrative examples to highlight distinctions between local webs and metawebs of probabilistic interactions. We created two metawebs of probabilistic interactions by employing constant false positive and false negative rates for all regional interactions. In the first metaweb, we set both false positive and false negative rates to zero to prevent artificially inflating the total number of links, enabling a more accurate comparison with binary interaction webs. This gave us a probability of regional interaction of 1 when at least one interaction has been observed locally and of 0 in the absence of any observed interaction between a given pair of species (Fig. 1c-d). In the second metaweb (used in the next applications), we introduced a 5% false positive rate to account for spurious interactions and a 10% false negative rate to address the elevated occurrence of missing interactions in ecological

networks (Catchen *et al.* 2023). We believe these rates represent reasonable estimates of missing and spurious potential interactions, but confirming their accuracy is challenging due to the unavailability of data on the actual feasibility of interaction. Observed interactions were thus given a probability of regional interaction of 95%, whereas unobserved ones were assigned a probability of 10%.

To build local webs of probabilistic interactions, we first recognize that local interactions must initially be biologically feasible before occurring at a specific time and space. A local probability of interaction  $P(L_{i \rightarrow j})$  can be expressed as the product of the probability of local interaction given that the two taxa can potentially interact  $P(L_{i \rightarrow j}|M_{i \rightarrow j} = 1)$ , which we sometimes denote as  $P(L_{i \rightarrow j}|M_{i \rightarrow j})$  for the sake of simplicity, with their probability of regional interaction  $P(M_{i \rightarrow j})$ :

$$P(L_{i \rightarrow j}) = P(L_{i \rightarrow j}|M_{i \rightarrow j} = 1) \times P(M_{i \rightarrow j} = 1). \quad (10)$$

We built local webs of probabilistic interactions using the taxa found in the empirical local networks and attributing pairwise interaction probabilities based on the metawebs of probabilistic interactions and a constant value of  $P(L_{i \rightarrow j}|M_{i \rightarrow j})$  across interactions. Low values of  $P(L_{i \rightarrow j}|M_{i \rightarrow j})$  indicate that feasible interactions rarely occur locally, intermediate values around 50% suggest considerable spatiotemporal variability, while high values indicate that regional interactions are nearly always realized locally. Following Eq. 10, the local probability of interaction between a given pair of taxa consistently remained equal to or below their probability of regional interaction.

In the last two panels of Fig. 1 (c-d), we show how the aggregated number of links and connectance (i.e., the proportion of all of the non-forbidden links that are realized) scale with the number of sampled local webs, according to different values of  $P(L_{i \rightarrow j}|M_{i \rightarrow j})$ . When aggregating local webs of probabilistic interactions, the constancy of the probability of regional interaction across the entire study area means that any rise in the probability of local interaction is solely attributable to an increase in  $P(L_{i \rightarrow j}|M_{i \rightarrow j})$ . For example, let  $L_1$  and  $L_2$  be two local networks and  $L_{1,2}$  the aggregated web. If  $P(L_{1,i \rightarrow j}|M_{i \rightarrow j})$  and  $P(L_{2,i \rightarrow j}|M_{i \rightarrow j})$  are the probabilities that two potentially interacting taxa interact respectively in  $L_1$  and  $L_2$ , the probability  $P(L_{1,2,i \rightarrow j}|M_{i \rightarrow j})$  that these taxa interact in the aggregated web  $L_{1,2}$  is obtained by:

$$P(L_{1,2,i \rightarrow j}|M_{i \rightarrow j}) = 1 - (1 - P(L_{1,i \rightarrow j}|M_{i \rightarrow j})) \times (1 - P(L_{2,i \rightarrow j}|M_{i \rightarrow j})), \quad (11)$$

assuming independence between the interaction of the two taxa in different networks. This equation represents the probability that the interaction is realized in either (1) exclusively the local web  $L_1$ , (2) exclusively the local web  $L_2$  or (3) both local webs, given that the two taxa have the biological capacity to interact.

By comparing the scaling relationships observed in webs of binary and probabilistic interactions; Fig. 1 illustrates that high values of  $P(L_{i \rightarrow j}|M_{i \rightarrow j})$  lead to systematic overestimations in the number of links and connectance, especially when  $P(L_{i \rightarrow j}|M_{i \rightarrow j}) = 1$  (corresponding to the scenario where local probabilities of interactions are equivalent to the probabilities of regional interactions). However, these biases tend to diminish as the number of sampled webs increases.

**3.2. Application 2: Spatial and temporal scaling of interactions** The investigation of network-area relationships and interaction accumulation curves is an important area of research in network ecology. First, network-area relationships elucidate the scaling of network properties (such as modularity and trophic chain lengths) with spatial scale (e.g., Galiana *et al.* 2018; Wood *et al.* 2015). The variations in network structure across spatial scales may stem from the scaling of species richness (species-area relationships, SARs), the number of interactions (Brose *et al.* 2004), and many other higher-level properties of the system (e.g., environmental heterogeneity, Thompson & Townsend 2005) with the sampled area. Additionally, ecological processes occurring at distinct spatial scales, such as the spatial variability in local community composition resulting from different sequences of extinction and colonization events, can also contribute to this variation (Galiana *et al.* 2018). Next, interaction accumulation curves describe the scaling of the number of observed interactions with sampling effort (Jordano 2016). Sampling effort, which may correspond to the duration of the sampling period used to construct the network, can impact connectance (Bersier *et al.* 1999) and various measures of network structure (Banašek-Richter *et al.* 2004; McLeod *et al.* 2021). Apart from sampling effort, the temporal scaling of interactions also elucidates how network structure changes with the temporal resolution of the network (Poisot *et al.* 2012), acknowledging that distinct interactions take place over time, ranging from short-term fluctuations of interactions to long-term trends. As local webs of probabilistic interactions may explicitly account for the spatiotemporal variability of interactions (as shown in Box 1), they offer a distinct approach to investigating the scaling of network structure with space and time, in contrast to

webs of binary and quantitative interactions, by making the stochasticity of interactions the focal point of the modeling process.

Local webs and metawebs exhibit distinct relationships with spatial and temporal scales. On one hand, metawebs of probabilistic interactions, representing biological feasibility, feature regional interactions that do not scale with space and time. This is because regional interactions depend solely on the biological capacity of two taxa to interact, regardless of their co-occurrence and local environmental conditions. However, probabilities of potential interactions may change (tending to become more definitive) upon updating previous estimates with increased sampling effort, even though they do not vary in a specific direction with the spatial and temporal extent (boundaries) of the network. The probability of two taxa potentially interacting should theoretically be the same in all metawebs in which they are present, provided that the data and methods used for estimation are consistent. For example, if a smaller metaweb  $M_1$  is derived from a larger metaweb  $M_0$  by selecting the subset of taxa present in the region described by  $M_1$  and retaining all their interactions, their probabilities of interaction should be identical regardless of scale, i.e.  $P(M_{1,i \rightarrow j}) = P(M_{0,i \rightarrow j})$ . With a larger or equal number of taxa in  $M_0$  as compared to  $M_1$ , the total number of interactions is expected to be higher or at least equal in the larger web, even though pairwise probabilities remain identical.

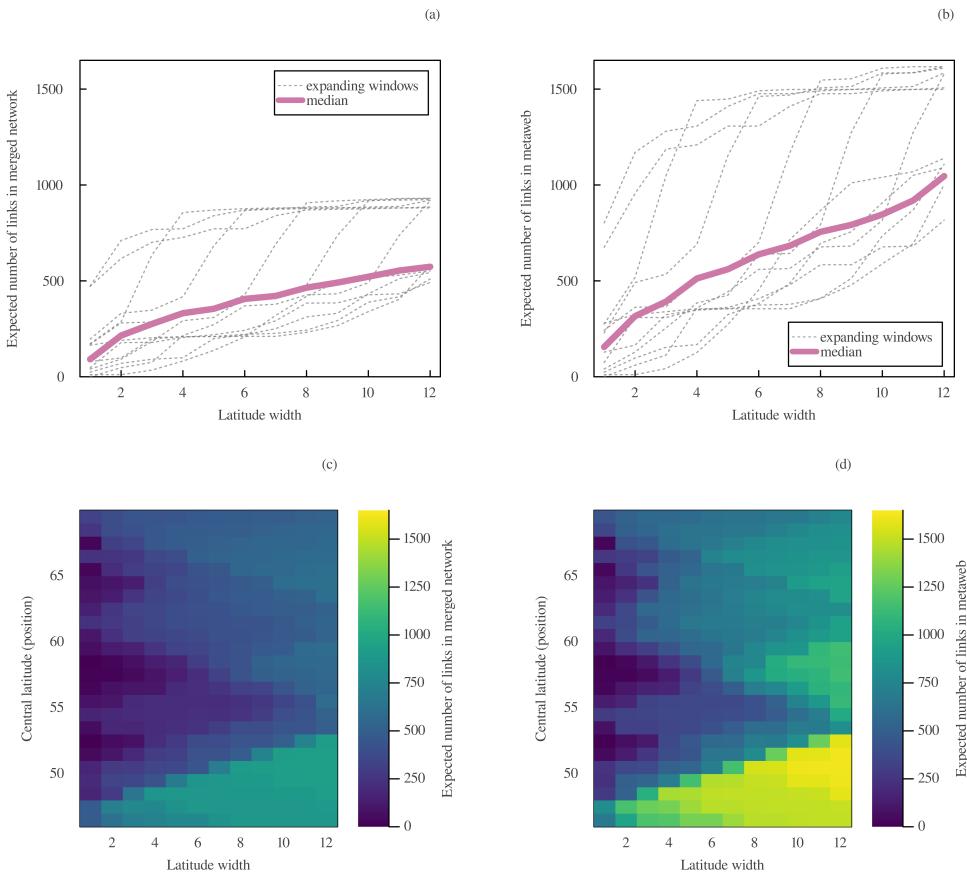
On the other hand, local interactions scale both spatially and temporally, given that they have more opportunities to be realized and observed in larger areas and longer durations. This is attributed to factors such as a higher number of individuals, greater trait variations, and increased opportunities for encounters, as highlighted by McLeod *et al.* (2020). For example, if a local web of probabilistic interactions  $L_1$  with an area  $A_1$  is derived from a larger web  $L_0$  with an area  $A_0$ , and  $A_1$  is entirely nested within  $A_0$ , interaction probabilities should be lower in the smaller web, i.e.  $P(L_{1,i \rightarrow j}|A_1 < A_0) \leq P(L_{0,i \rightarrow j}|A_0)$ . However, if  $A_1$  and  $A_0$  are disjoint, interaction probabilities could be higher in the smaller area, contingent upon their environmental and biological conditions. Likewise, interaction probabilities are expected to be lower in webs with shorter durations when time intervals are nested. In Fig. 2, we show how the expected number of local host-parasite interactions scales with area, represented as an expanding latitudinal window, in comparison with regional interactions. Even though we employed local probabilities of interactions equal to regional interactions for the purpose of comparison (i.e., using  $P(L_{i \rightarrow j}|M_{i \rightarrow j}) = 1$  here), we notice that the total number of regional interactions scales more rapidly than local interactions. This is because numerous regional interactions involve species that never co-occur, and as a result, these interactions are not captured in any local web.

**3.3. Box 1: A spatiotemporally explicit model of interactions** Predicting local webs across time and space is a pivotal goal of network ecology, especially given the scarcity of interaction data (Strydom *et al.* 2021). Ecologists may resort to predictive models (e.g., generative Bayesian and machine learning models) to reconstruct networks at fine spatial and temporal scales with limited interaction data. For instance, real-time biomonitoring data coupled with appropriate numerical models (Bohan *et al.* 2017) can be employed to reconstruct local ecological webs, opening avenues for in-depth studies on local ecosystem functioning and dynamics. The probabilistic representation of interactions describes the inherent uncertainty in these models, typically expressed through probability distributions. We introduce and develop a simple generative mechanistic model for probabilistic local interactions that takes into consideration their spatiotemporal variability (i.e., a spatiotemporally explicit model of local interactions). It is essential to note that our model is not designed for regional interactions, which are scale-independent. Rather, it could prove valuable for predicting local interactions across time and space by generating new interaction data following parameter inference.

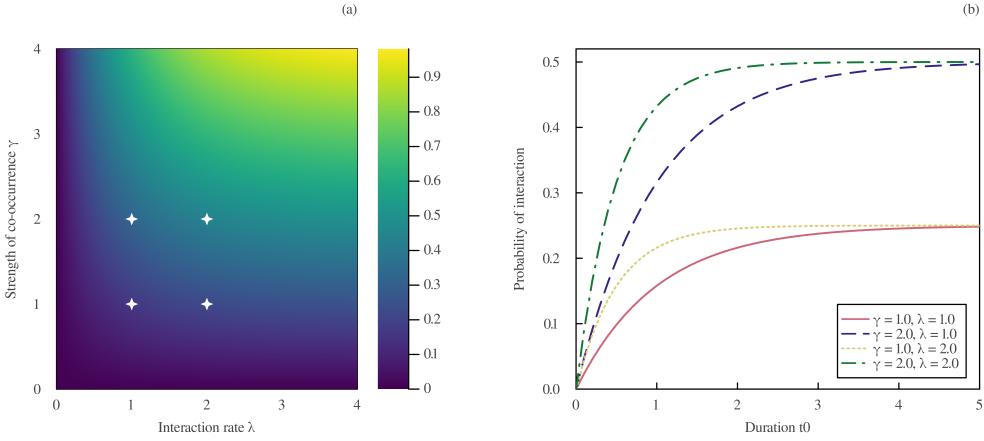
As indicated by Eq. 1, the probability that two taxa  $i$  and  $j$  interact at a specific location and time  $l$  may be determined by the product of their probability of interaction given co-occurrence and their probability of co-occurrence. First, their co-occurrence probability can be calculated using their individual (marginal) occurrence probabilities  $P(X_{i,l})$  and  $P(X_{j,l})$ . Given that taxa occurrences are not independent of each other, the joint probability of both taxa occurring together can be calculated by multiplying the probability of one taxon being present by the conditional probability of the other occurring when the first one is present. Alternatively, the joint occurrence probability can be obtained by multiplying the marginal probabilities by the strength of association  $\gamma$  between the occurrences of both taxa (Cazelles *et al.* 2016):

$$P(X_{i,l}, X_{j,l}) = P(X_{i,l}|X_{j,l})P(X_{j,l}) = P(X_{i,l})P(X_{j,l})\gamma. \quad (12)$$

When  $\gamma > 1$ , it signifies a positive association in the geographic distributions of both taxa, indicating that the presence of one taxon enhances the probability of occurrence of the other. Attractions may be the result of positive interactions (e.g., mutualism) or positive/negative interactions (e.g., antagonism between predators and prey, Cazelles *et al.* 2016). In empirical webs,  $\gamma > 1$  holds true for the majority of species pairs (Catchen



**Figure 2 Spatial scaling of interactions.** Expected number of host-parasite interactions in a network aggregating all (a) local and (b) regional probabilistic interactions within a latitudinal window of a given length. Every dashed curve corresponds to a different expanding window characterized by a different central latitude, with the colored solid line representing the median number of interactions. Heatmaps of the expected number of (c) local and (d) regional interactions found in windows of specified length and central latitudes. Probabilities of regional interactions were obtained with a false positive rate of 5% and a false negative rate of 10%. Local probabilistic interactions were derived from probabilistic regional interactions by setting the value of  $P(L_{i \rightarrow j} | M_{i \rightarrow j})$  (the local probability of interaction among potentially interacting species) to 1, ensuring a conservative comparison between aggregated local webs and metawebs. Aggregated local webs were obtained by aggregating both the species and interactions found within a particular latitudinal window, with the values of  $P(L_{i \rightarrow j} | M_{i \rightarrow j})$  remaining at their maximum value of 1 following Eq. 11.



**Figure 3 Parameters of the spatiotemporally explicit model of interactions.** (a) Probability of local interaction given by the process model (Eq. 16) under different values of  $\lambda_l$  and  $\gamma$ , with  $t_0 = 1$ . The parameter values used in the right panel are denoted by the white stars. (b) Scaling of the probability of interaction with the duration parameter  $t_0$  in Eq. 16, for different values of  $\lambda_l$  and  $\gamma$ . In both panels, the marginal probabilities of occurrence  $P(X_{i,l})$  and  $P(X_{j,l})$  are set to a constant value of 0.5.

et al. 2023). In contrast, repulsions ( $\gamma < 1$ ) may be caused e.g. by strong interspecific competition (Cazelles et al. 2016). We model the co-occurrence  $X_{i,j,l}$  of both taxa as the outcome of a Bernoulli trial

$$X_{i,j,l} \sim \text{Bernoulli}(P(X_{i,l}, X_{j,l})). \quad (13)$$

Next, the probability of interaction given co-occurrence can be made temporally explicit by modeling it as a Poisson process with rate parameter  $\lambda_l$ . This parameter represents the local expected frequency of interaction between the taxa within a defined time interval and can be estimated using prior data on interaction strengths, when accessible. The probability that two co-occurring taxa engage in an interaction during a time period  $t_0$  is given by:

$$P(L_{i→j}|X_{i,j,l}=1) = 1 - e^{-\lambda_l t_0}, \quad (14)$$

which tends toward 1 as  $t_0 \rightarrow \infty$  if  $\lambda_l > 0$ . In other words, two co-occurring taxa with a nonzero rate of interaction will inevitably interact in a sufficiently long time period. It is important to note that the units of  $\lambda_l$  and  $t_0$  are complementary. For instance, if the duration  $t_0$  is measured in months,  $\lambda_l$  denote the expected number of interactions per month.

The occurrence of an interaction between  $i$  and  $j$  can be modeled as a Bernoulli trial with a probability of  $P(L_{i→j})$ . Consequently, a Bayesian inference model can be built based on the preceding equations to estimate the value of the  $\lambda_l$  and  $\gamma$  parameters and generate novel interaction data:

$$L_{i→j} \sim \text{Bernoulli}(P(L_{i→j})) \quad (15)$$

$$P(L_{i→j}) = P(X_{i,l})P(X_{j,l})\gamma(1 - e^{-\lambda_l t_0}) \quad (16)$$

$$\gamma \sim \text{Gamma}(2, 0.5) \quad (17)$$

$$\lambda_l \sim \text{Exponential}(2) \quad (18)$$

In Fig. 3, we show the variation in the probability of interaction under different parameter values. In the right panel, we notice that, irrespective of the interaction rate  $\lambda_l$ , the probability of interaction converges toward an asymptote determined by the probability of co-occurrence  $P(X_{i,j,l})$  (Eq. 12). This model can be customized in different ways, such as by linking  $\lambda_l$  with specific environmental variables or explicitly incorporating observation errors (i.e., the probabilities of false negatives and false positives).

**3.4. Application 3: Taxonomic scaling of interactions** Probabilistic interaction networks offer a versatile approach to tackle a broad array of ecological questions, depending on their level of organization. For instance, the assemblage of interactions across ecological scales can be explored through species-based networks, while clade-based networks provide insights into macroevolutionary processes (e.g., Gomez *et al.* 2010). Given that our interpretation of the properties and dynamics of ecological webs depends on their taxonomic level (Guimarães 2020), investigating the taxonomic scaling of interactions (i.e., how interaction probabilities change with taxonomic level) emerges as a promising research avenue. Examining the same system at various taxonomic scales can yield meaningful and complementary ecological information, and, in our perspective, employing webs of probabilistic interactions is an effective approach for such analyses.

There are no inherent differences in the taxonomic scaling between local and metawebs, as only the nodes are defined taxonomically. In other words, the probability values of edges in both local (Eq. 2) and metawebs (Eq. 4) are not conditional on any taxonomic level. The taxonomic scale is tied to the definition of the event itself (i.e., the interaction between two taxa, defined at the desired taxonomical scale), not to the variables on which interaction probabilities are conditional. In both types of webs, transitioning to a broader level of organization (e.g., from a species-level web  $S$  to a genus-level web  $G$ ) can be accomplished directly by using probabilities from finer scales. For example, in a network with  $n_1$  species from genus  $g_1$  and  $n_2$  species from genus  $g_2$ , one can compute the probability that at least one species from genus  $g_1$  interacts with at least one species from genus  $g_2$  (i.e., that there is a non-zero number of species-level interactions or, equivalently, that the genus-level interaction occurs) as follows:

$$P(G_{g_1 \rightarrow g_2} = 1) = 1 - \prod_{i=1}^{n_1} \prod_{j=1}^{n_2} (1 - P(S_{g_{1i} \rightarrow g_{2j}} = 1)), \quad (19)$$

where  $g_{1,i}$  and  $g_{2,j}$  are the species of the corresponding genus and assuming independence between species-level interactions. If it is known that at least two of these species interact (i.e.,  $P(S_{g_{1,i} \rightarrow g_{2,j}}) = 1$  for at least one pair of  $(g_{1,i}, g_{2,j})$ ), it implies a probability of genus interaction equal to 1. Canard *et al.* (2012) built a species-based network following a similar approach, by using simulated interactions between individuals derived from a neutral model (i.e., a model that assumed ecological equivalence among individuals). In contrast, a more sophisticated approach is necessary when transitioning from a broader to a finer level of organization. This is because knowledge of an interaction between two genera does not guarantee that all possible pairwise combinations of their species will also interact. One possible method is to build a finer-scale network by generating probabilities of interactions through random sampling from a beta distribution, parameterized by the broader-scale network.

Ideally, our biological interpretation of probabilistic interactions should remain consistent across a network even if it incorporates heterogeneous levels of organization, such as a network whose nodes represent both species and trophic species (groups of species sharing similar predators and prey). This is common in ecological webs where taxonomic resolution is typically low (Hemprich-Bennett *et al.* 2021; Vázquez *et al.* 2022). Interaction probabilities at broader taxonomic scales should be based on probabilities of interactions between individuals, for both local webs and metawebs. For instance, in local individual-based food webs, the probability that two individuals interact reflects our degree of belief that one individual will consume the other. Similarly, in local species-based food webs, the probability that two species interact represents our degree of belief that at least one individual from the predator species consumes at least another individual from the prey species. Furthermore, in local clade-based food webs, the probability that two clades interact represents our 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 involves aggregating interactions between individuals into larger groups, which may exhibit varying degrees of homogeneity based on the organisms and the study system. In that regard, taxonomic scaling is analogous to the spatial and temporal scaling of interactions, as they all represent different ways to aggregate individuals into broader groups (either spatially, temporally, or taxonomically).

**3.5. Application 4: Sampling for binary interaction webs** The prediction of binary interactions through Bernoulli trials is an important application of probabilistic interaction webs. This approach proves beneficial for analyzing the structural characteristics of probabilistic interaction webs, particularly in the absence of specific analytical formulas for measures describing the structure of the complete web (Poisot *et al.* 2016). By performing independent Bernoulli trials for each interaction, a network of binary interactions may be generated. A probability distribution of network properties can be obtained by measuring network structure across multiple randomly generated networks (Poisot *et al.* 2016). This method enables the representation of the variability of network structure, albeit with possible biases when connectance is low (Chagnon 2015; Poisot &

Gravel 2014). Employing this strategy to generate binary interaction networks under a null model facilitates null hypothesis significance testing, wherein the observed measure is compared against the simulated distribution (e.g., Bascompte *et al.* 2003). Additionally, the random generation of binary interaction networks, from a probabilistic interaction web that accounts for the spatiotemporal variability of interactions, may effectively capture network structure across space and time. This facilitates the investigation of ecological hypotheses about interactions at broad spatial and temporal scales. Yet, inferring network structure through sampling for binary interactions assumes independence among interactions, which might not accurately represent reality. Covariation among interactions could exist even if we do not explicitly condition interactions on others. For example, an interaction between two taxa could be more probable when another interaction occurs. The consequences of this assumption of independence on the prediction of network structure have yet to be empirically examined.

There are at least two distinct approaches to sample binary interaction networks from probabilistic interaction webs across space, for example, when attempting to predict a binary interaction network for each location  $l$  within a given region. Both approaches assume independence between interactions. The first approach involves performing a singular Bernoulli trial for each pair of taxa based on their regional probability of interaction:

$$M_{i \rightarrow j} \sim \text{Bernoulli}(P(M_{i \rightarrow j})).$$

In employing this approach, we predict a single metaweb of binary interactions for each simulation. Every pair of taxa predicted to interact in this metaweb will be treated as interacting in all local webs where they co-occur, i.e.  $L_{l,i \rightarrow j} = M_{i \rightarrow j}$  when  $X_{i,j,l} = 1$ . This will result in local pairwise interactions without spatial variation.

The second approach is to independently sample each local web of probabilistic interactions:

$$L_{l,i \rightarrow j} \sim \text{Bernoulli}(P(L_{l,i \rightarrow j})).$$

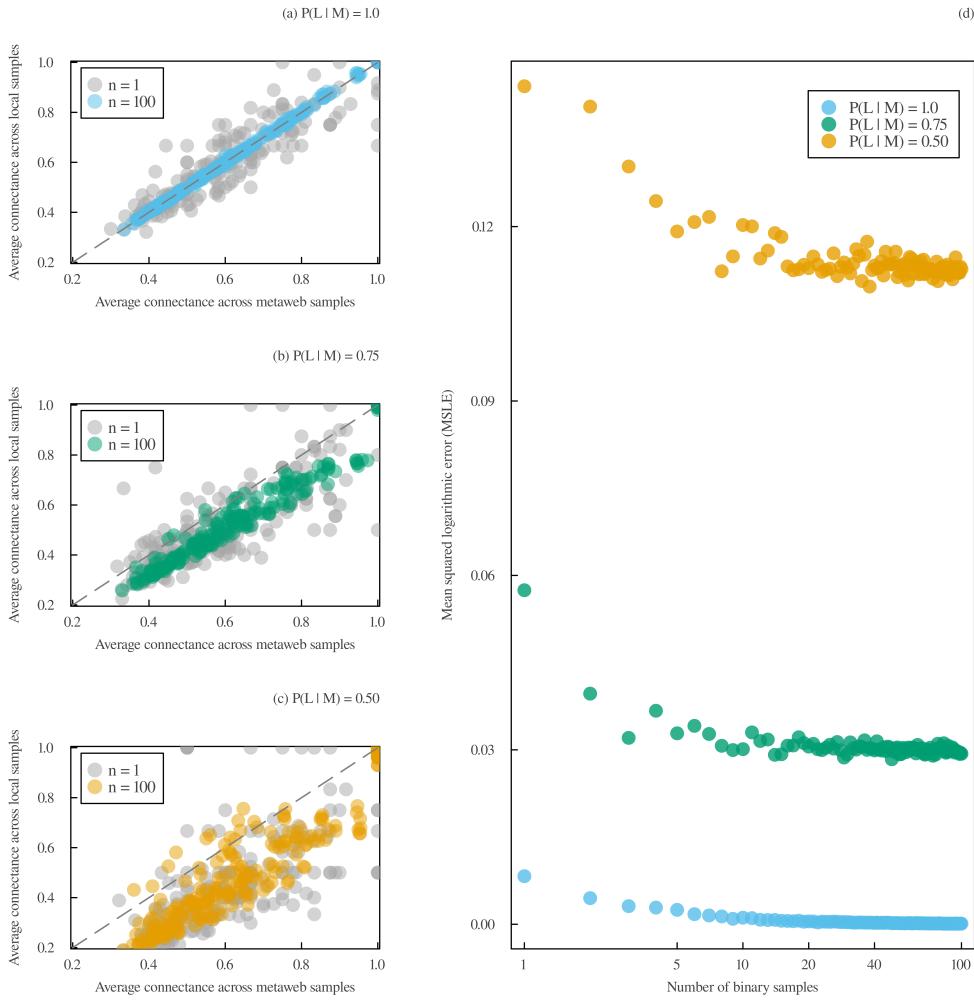
This can be achieved by first generating distinct probabilistic interaction networks for each location. These local webs of probabilistic interactions may vary in taxa composition and interaction probabilities. Because binary interaction networks are sampled independently for each location, this second approach introduces spatial variation in binary interactions.

In Fig. 4, we compare the average connectance of binary interaction networks resulting from these two sampling techniques. We drew regional and local interactions from our host-parasite networks of probabilistic interactions, generating a number of binary interaction web realizations for each site in the dataset. These two sampling approaches yield different outcomes, particularly for lower values of  $P(L_{i \rightarrow j}|M_{i \rightarrow j})$ , which represent instances where regional interactions do not consistently manifest locally. Small discrepancies between these techniques are also apparent when we equate the probability of local interaction to the probability of regional interaction (i.e., when using  $P(L_{i \rightarrow j}|M_{i \rightarrow j}) = 1.0$  in Eq. 10, Fig. 4a), especially when the number of simulations of binary interaction networks for each location is low ( $n = 1$ ). As anticipated, we observe that sampling binary interactions from the metaweb tends to overestimate connectance on average compared to sampling them from local webs (Fig. 4a-c). Furthermore, we observe an increase in the variability of connectance when employing a single simulation (Fig. 4, gray markers), which in our opinion is a more tangible representation of the process leading to the realization of local interactions in nature.

The choice of a sampling approach can influence the selection of grid cell size when delineating local communities within a broader region of interest. In the first approach, pairwise interactions remain constant irrespective of cell size because they are sampled only once from the metaweb. However, in the second approach, local interaction probabilities are contingent on network area. For instance, consider the local webs  $L_1$  and  $L_2$  with an area of  $\frac{1}{2}A_0$ , both nested within  $A_0$  but disjoint from each other, forming  $L_0$ . If we treat  $L_1$  and  $L_2$  as independent, the probability of interaction between taxa  $i$  and  $j$  in  $L_0$  is given by:

$$P(L_{0,i \rightarrow j}) = 1 - (1 - P(L_{1,i \rightarrow j}) \times P(L_{2,i \rightarrow j})). \quad (20)$$

Due to its larger area, the probability that the two taxa interact in  $L_0$  is equal or greater than in  $L_1$  and  $L_2$ . When sampling binary interactions from local webs, it is crucial to sample at the same spatial scale for which probabilities were estimated. Otherwise, interaction probabilities must be adjusted to align with the intended cell size, preventing systematic biases in predictions.



**Figure 4 Connectance of sampled binary interaction webs.** Comparison between the average connectance of binary interaction network samples obtained from the local webs and metawebs of probabilistic interactions. Each dot corresponds to a different site. The local probability of interaction between potentially interacting species was set to three different values: (a)  $P(L_{i \rightarrow j} \mid M_{i \rightarrow j}) = 1.0$ , (b)  $P(L_{i \rightarrow j} \mid M_{i \rightarrow j}) = 0.75$ , and (c)  $P(L_{i \rightarrow j} \mid M_{i \rightarrow j}) = 0.50$ . Grey dots represent the outcome of a single simulation, while colored dots represent the average connectance of each network across 100 simulations. (d) Reduction in the mean squared logarithmic error between the average connectance of binary interaction networks obtained from these two sampling methods as the number of simulations increases, for the same values of  $P(L_{i \rightarrow j} \mid M_{i \rightarrow j})$  used in panels a-c. Probabilities of regional interactions were obtained with a false positive rate of 5% and a false negative rate of 10%. Metaweb samples were obtained by randomly sampling binary interactions from the probabilistic interaction metaweb, and then propagating this result to all local webs that include the species potentially engaged in the interactions. Local binary interaction webs were generated by independently sampling binary interactions for each local web of probabilistic interactions.

**3.6. Application 5: Prediction of local webs from metawebs** Metawebs serve as a valuable source of ecological information for predicting local webs across time and space. Local webs of binary interactions can be reconstructed by selecting a subset of taxa and interactions from the metaweb (Dunne 2006). This implies that metawebs consistently contain more interactions than their corresponding local webs, even though their connectance is usually much smaller than the ones of local webs (Gravel *et al.* 2011). Determining the list of taxa to select can be achieved empirically or through numerical methods like range maps or species distribution models. As species composition is arguably less difficult to sample or predict than pairwise interactions, the primary challenge lies in deciding which interactions to select from the metaweb, a task that may necessitate advanced statistical models and ecological expertise. Inferring the structure of local webs from the metaweb before predicting specific local pairwise interactions could hold promise (Strydom *et al.* 2021), considering that the structure of local webs is constrained by the metaweb (Saravia *et al.* 2022).

Building local webs of probabilistic interactions from a metaweb of probabilistic interactions involves a reduction in the value of pairwise interaction probabilities. This decrease is attributed to the prerequisite that two taxa must initially possess the capacity to interact before engaging in local interactions (Eq. 10). Therefore, inferring local webs from their metaweb while maintaining identical interaction probability values would introduce systematic biases into the predictions. In such cases, these networks would essentially represent smaller-scale metawebs of potential interactions, possibly leading to misinterpretations by being perceived as local interactions. As proposed by McLeod *et al.* (2021), although metawebs do not capture the spatiotemporal variability of interactions, they establish an upper limit for local interactions (similarly for metawebs of probabilistic interactions, Strydom *et al.* 2023). In other words, the probability of two taxa interacting at a specific location and time is consistently lower or equal to the probability of their regional interaction:

$$P(L_{i \rightarrow j} | X_{i,l}, X_{j,l}, N_{i,l}, N_{j,l}, T_{i,l}, T_{j,l}, E_l, A, t, f(L)) \leq P(M_{i \rightarrow j} | T_i, T_j). \quad (21)$$

Moreover, the probability that two taxa possess the biological capacity to interact must exceed the probability of them interacting at any location and time because they may never co-occur or encounter locally. Specifically, the cumulative probability of local interactions across all spatial, temporal, and environmental conditions must be less than the probability of regional interaction, i.e.

$$\int_{E_l} \int_A \int_t P(L_{i \rightarrow j} | E_l, A, t) dt dA dE_l \leq P(M_{i \rightarrow j} | T_i, T_j). \quad (22)$$

Estimating more precisely the probability  $P(L_{i \rightarrow j} | M_{i \rightarrow j})$  that two taxa interact locally if they can potentially interact allows for improved predictions of local webs from the metaweb of probabilistic interactions. This task is challenging due to the variability of this probability across space and time, as well as its variability across pairwise interactions within a network. Using simple models of  $P(L_{i \rightarrow j} | M_{i \rightarrow j})$ , as demonstrated in our case studies, represents an initial step toward the overarching objective of reconstructing local webs from metawebs.

## 4

---

### Conclusion

In this contribution, we underline the importance of network metadata for adequately interpreting and manipulating probabilistic interaction data. The mathematical representation of probabilities and their statistical properties depend on the type of interactions (local or regional) and the conditions under which these interactions were evaluated. We showed that local webs and metawebs of probabilistic interactions differ in their relationship to spatial and temporal scales, with regional interactions remaining consistent across scales. In contrast with metawebs, local interactions are measured in a specific context (e.g., in a given area, time, and biological and environmental conditions) and depend on taxa co-occurrence. These differences bring to light the need to use probabilistic data with caution, for instance when generating network realizations of binary interactions across space and predicting local webs from metawebs. Clear metadata describing the type of interaction and the variables used in their estimation are required to ensure adequate data manipulation. Sound data practices and foundations for probabilistic thinking in network ecology could facilitate reliable assessments of the spatiotemporal variability and uncertainty of biotic interactions.

Estimating local interaction probabilities independently for each taxa pair and assembling them into a network of probabilistic interactions comes with limitations. Predicting local networks of binary interactions based

on these interaction probabilities assumes independence among interactions, a condition seldom respected in practice (Golubski & Abrams 2011). Relaxing this assumption is the next logical step in the stochastic representation of interactions. A more accurate representation of the stochasticity of ecological networks involves creating *probabilistic networks* ( $P(L|...)$  and  $P(M|...)$ ), rather than networks of *probabilistic interactions* ( $P(L_{i \rightarrow j}|...)$  and  $P(M_{i \rightarrow j}|...)$ ). Probabilistic networks describe the probability that a particular network of binary (or quantitative) interactions (its whole adjacency matrix) is realized. For example, Young *et al.* (2021) used a Bayesian approach to estimate the probability of different plant-pollinator network structures derived from imperfect observational data. A probability distribution of ecological networks may also be derived using the principle of maximum entropy given some structural constrained (e.g., Cimini *et al.* 2019; Park & Newman 2004). Regardless of the method employed, generating probabilistic local webs, bypassing the need to independently estimate local interaction probabilities, could lead to more accurate predictions of local webs. Furthermore, probabilistic networks could serve as an alternative to null hypothesis significance testing when comparing the structure of a local web to some random expectations or, as done in Pellissier *et al.* (2018), to the metaweb. These random expectations are typically derived by performing a series of Bernoulli trials on probabilistic interactions, assuming independence, to generate a distribution of networks of binary interactions and then calculate their structure (Poisot *et al.* 2016). One could for instance compare the likelihood of an observed network to the one of the most likely network structure according to the probabilistic network distribution, thereby directly obtaining a measure of discrepancy of the empirical network. Generating probabilistic ecological networks represents a tangible challenge, one that, in the coming years, promises to unlock doors to more advanced and adequate analyses of ecological networks.

It is essential to enhance our comprehension of both regional and local interactions, especially considering the current scarcity of interaction data. While sampling biological communities does decrease the uncertainty of interactions by accumulating evidence for their feasibility and local realization, there is a limit to how much we can reduce uncertainty. In metawebs, probabilities reflect our limited knowledge of interactions (i.e., our degree of belief that interactions are feasible), which is expected to improve with a larger volume of data. Regional interactions should become more definitive (with probabilities approaching 0 or 1) as we investigate various conditions, including different combinations of species traits. In comparison, although local webs can be seen as random instances of metawebs, their randomness cannot be reduced to the same extent. Local interaction probabilities may represent both their uncertainty and spatiotemporal variability. Owing to environmental heterogeneity, there will invariably be instances in which an interaction occurs and others in which it does not, across different times and locations, irrespective of the extent to which we can improve our knowledge of its biological feasibility and the local conditions that facilitate its occurrence. When local webs describe probabilities of observing interactions rather than their actual occurrence, we must also consider observation variability (sampling error) as an additional source of stochasticity. Every ecological process is stochastic but there is also a possibility that a phenomenon goes undetected. Quantifying and partitioning this stochasticity will enable us to make more accurate predictions about ecological interactions at various spatial and temporal scales. This will prove to be of vital importance as our time to understand nature runs out, especially at locations where the impacts of climate change and habitat loss hit harder.

---

## 5

### 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 (BIOS<sup>2</sup>) program. A special thanks to all members of the Black Holes and Revelations working group (organized by BIOS<sup>2</sup>) for their insightful discussions and valuable feedback on this manuscript.

---

### References

- Abrego, N., Roslin, T., Huotari, T., Ji, Y., Schmidt, N.M., Wang, J., *et al.* (2021). Accounting for species interactions is necessary for predicting how arctic arthropod communities respond to climate change. *Ecography*, 44, 885–896.

- 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.
- Banašek-Richter, C., Cattin, M.-F. & Bersier, L.-F. (2004). [Sampling effects and the robustness of quantitative and qualitative food-web descriptors](#). *Journal of Theoretical Biology*, 226, 23–32.
- 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.
- Berg, M.P. & Ellers, J. (2010). [Trait plasticity in species interactions: A driving force of community dynamics](#). *Evolutionary Ecology*, 24, 617–629.
- 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.
- Bersier, L.-F., Dixon, P. & Sugihara, G. (1999). [Scale-Invariant or Scale-Dependent Behavior of the Link Density Property in Food Webs: A Matter of Sampling Effort?](#) *The American Naturalist*, 153, 676–682.
- 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.
- Bolnick, D.I., Amarasekare, P., Araújo, M.S., Bürger, R., Levine, J.M., Novak, M., et al. (2011). [Why intraspecific trait variation matters in community ecology](#). *Trends in Ecology & Evolution*, 26, 183–192.
- 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.
- Boulangeat, I., Gravel, D. & Thuiller, W. (2012). [Accounting for dispersal and biotic interactions to disentangle the drivers of species distributions and their abundances](#). *Ecology Letters*, 15, 584–593.
- Brimacombe, C., Bodner, K., Michalska-Smith, M., Poisot, T. & Fortin, M.-J. (2023). [Shortcomings of reusing species interaction networks created by different sets of researchers](#). *PLOS Biology*, 21, e3002068.
- Broom, M. & Ruxton, G.D. (2005). [You can run—or you can hide: Optimal strategies for cryptic prey against pursuit predators](#). *Behavioral Ecology*, 16, 534–540.
- Brose, U., Ostling, A., Harrison, K. & Martinez, N.D. (2004). [Unified spatial scaling of species and their trophic interactions](#). *Nature*, 428, 167–171.
- Burnham, K.P. & Anderson, D.R. (2004). [Multimodel Inference: Understanding AIC and BIC in Model Selection](#). *Sociological Methods & Research*, 33, 261–304.
- Canard, E.F., Mouquet, N., Mouillot, D., Stanko, M., Miklisova, D. & Gravel, D. (2014). [Empirical Evaluation of Neutral Interactions in Host-Parasite Networks](#). *The American Naturalist*, 183, 468–479.
- 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.
- Castoe, T.A., Spencer, C.L. & Parkinson, C.L. (2007). [Phylogeographic structure and historical demography of the western diamondback rattlesnake \(\*Crotalus Atrox\*\): A perspective on North American desert biogeography](#). *Molecular Phylogenetics and Evolution*, 42, 193–212.
- 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.

- Chagnon, P.-L. (2015). Characterizing topology of ecological networks along gradients: The limits of metrics' standardization. *Ecological Complexity*, 22, 36–39.
- Choh, Y., Ignacio, M., Sabelis, M.W. & Janssen, A. (2012). Predator-prey role reversals, juvenile experience and adult antipredator behaviour. *Scientific Reports*, 2, 728.
- Cimini, G., Squartini, T., Saracco, F., Garlaschelli, D., Gabrielli, A. & Caldarelli, G. (2019). The statistical physics of real-world networks. *Nature Reviews Physics*, 1, 58–71.
- 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.
- Dallas, T., Park, A.W. & Drake, J.M. (2017). Predicting cryptic links in host-parasite networks. *PLOS Computational Biology*, 13, e1005557.
- 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). Predator-prey body size, interaction strength and the stability of a real food web. *Journal of Animal Ecology*, 73, 399–409.
- Fedorov, V.B., Goropashnaya, A.V., Boeskorov, G.G. & Cook, J.A. (2008). Comparative phylogeography and demographic history of the wood lemming (*Myopus schisticolor*): Implications for late Quaternary history of the taiga species in Eurasia. *Molecular Ecology*, 17, 598–610.
- Fortuna, M.A. & Bascompte, J. (2006). Habitat loss and the structure of plant-animal mutualistic networks. *Ecology Letters*, 9, 281–286.
- Fu, X., Seo, E., Clarke, J. & Hutchinson, R.A. (2021). Link Prediction Under Imperfect Detection: Collaborative Filtering for Ecological Networks. *IEEE Transactions on Knowledge and Data Engineering*, 33, 3117–3128.
- 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, A. & Londoño, M.C. (2022). Monitor biodiversity for action. *Science*, 378, 1147–1147.
- Gonzalez-Varo, J.P. & Traveset, A. (2016). The Labile Limits of Forbidden Interactions. *Trends in Ecology & Evolution*, 31, 700–710.
- Gravel, D., Albouy, C. & Thuiller, W. (2016). The meaning of functional trait composition of food webs for ecosystem functioning. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 371, 20150268.
- 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., Massol, F., Canard, E., Mouillot, D. & Mouquet, N. (2011). Trophic theory of island biogeography. *Ecology Letters*, 14, 1010–1016.
- Gravel, D., Poisot, T., Albouy, C., Velez, L. & Mouillot, D. (2013). Inferring food web structure from predator-prey 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.
- Herrera, C.M. (1989). **Pollinator abundance, morphology, and flower visitation rate: Analysis of the “quantity” component in a plant-pollinator system**. *Oecologia*, 80, 241–248.
- 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. (1987). **Patterns of Mutualistic Interactions in Pollination and Seed Dispersal: Connectance, Dependence Asymmetries, and Coevolution**. *The American Naturalist*, 129, 657–677.
- 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 plant-animal interactions**. *Ecology Letters*, 6, 69–81.
- Kausrud, K.L., Mysterud, A., Steen, H., Vik, J.O., Østbye, E., Cazelles, B., et al. (2008). **Linking climate change to lemming cycles**. *Nature*, 456, 93–97.
- 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.
- Kissner, K.J., Forbes, M.R. & Secoy, D.M. (1997). **Rattling Behavior of Prairie Rattlesnakes (*Crotalus viridis viridis*, *Viperidae*) in Relation to Sex, Reproductive Status, Body Size, and Body Temperature**. *Ethology*, 103, 1042–1050.
- 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.
- Kulkarni, M.A., Kweka, E., Nyale, E., Lyatuu, E., Mosha, F.W., Chandramohan, D., et al. (2006). **Entomological Evaluation of Malaria Vectors at Different Altitudes in Hai District, Northeastern Tanzania**. *Journal of Medical Entomology*, 43, 580–588.
- Maiorano, L., Montemaggioli, A., Ficetola, G.F., O'Connor, L. & Thuiller, W. (2020). **TETRA-EU 1.0: A species-level trophic metaweb of European tetrapods**. *Global Ecology and Biogeography*, 29, 1452–1457.
- 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.
- Melián, C.J., Vilas, C., Baldó, F., González-Ortegón, E., Drake, P. & Williams, R.J. (2011). **Eco-evolutionary Dynamics of Individual-Based Food Webs**. In: *Advances in Ecological Research*, The Role of Body Size in Multispecies Systems (ed. Belgrano, A.). Academic Press, pp. 225–268.

- Morales-Castilla, I., Matias, M.G., Gravel, D. & Araújo, M.B. (2015). Inferring biotic interactions from proxies. *Trends in Ecology & Evolution*, 30, 347–356.
- 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.
- Park, J. & Newman, M.E.J. (2004). Statistical mechanics of networks. *Physical Review E*, 70, 066117.
- Pascual, M., Dunne, J.A. & Dunne, J.A. (2006). *Ecological Networks: Linking Structure to Dynamics in Food Webs*. Oxford University Press, USA.
- Pellissier, L., Albouy, C., Bascompte, J., Farwig, N., Graham, C., Loreau, M., et al. (2018). Comparing species interaction networks along environmental gradients. *Biological Reviews*, 93, 785–800.
- 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.
- Plagányi, É.E. & Butterworth, D.S. (2004). A critical look at the potential of Ecopath with ecosim to assist in practical fisheries management. *African Journal of Marine Science*, 26, 261–287.
- 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. & Gravel, D. (2014). When is an ecological network complex? Connectance drives degree distribution and emerging network properties. *PeerJ*, 2, e251.
- Poisot, T., Stouffer, D.B. & Gravel, D. (2015). Beyond species: Why ecological interaction networks vary through space and time. *Oikos*, 124, 243–251.
- Polis, G.A. (1991). Complex Trophic Interactions in Deserts: An Empirical Critique of Food-Web Theory. *The American Naturalist*, 138, 123–155.
- Pollock, L.J., Tingley, R., Morris, W.K., Golding, N., O'Hara, R.B., Parris, K.M., et al. (2014). Understanding co-occurrence by modelling 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 predator–prey 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 plant–pollinator interactions. *GigaScience*, 11, giac043.
- Sanders, D. & van Veen, F.J.F. (2012). Indirect commensalism promotes persistence of secondary consumer species. *Biology Letters*, 8, 960–963.
- 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.

- Shaw, J.O., Dunhill, A.M., Beckerman, A.P., Dunne, J.A. & Hull, P.M. (2024). [A framework for reconstructing ancient food webs using functional trait data](#).
- Singer, M.C. & McBride, C.S. (2012). [Geographic mosaics of species' association: A definition and an example driven by plant–insect phenological synchrony](#). *Ecology*, 93, 2658–2673.
- Staniczenko, P.P.A., Sivasubramaniam, P., Suttle, K.B. & Pearson, R.G. (2017). [Linking macroecology and community ecology: Refining predictions of species distributions using biotic interaction networks](#). *Ecology Letters*, 20, 693–707.
- 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., Bouskila, S., Banville, F., Barros, C., Caron, D., Farrell, M.J., et al. (2023). [Graph embedding and transfer learning can help predict potential species interaction networks despite data limitations](#). *Methods in Ecology and Evolution*, 14, 2917–2930.
- 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.
- Thompson, R.M. & Townsend, C.R. (2005). [Food-Web Topology Varies with Spatial Scale in a Patchy Environment](#). *Ecology*, 86, 1916–1925.
- Tylianakis, J.M., Tscharntke, T. & Lewis, O.T. (2007). [Habitat modification alters the structure of tropical host-parasitoid food webs](#). *Nature*, 445, 202–205.
- Vázquez, D.P., Melián, C.J., Williams, N.M., Blüthgen, N., Krasnov, B.R. & Poulin, R. (2007). [Species abundance and asymmetric interaction strength in ecological networks](#). *Oikos*, 116, 1120–1127.
- Vázquez, D.P., Peralta, G., Cagnolo, L. & Santos, M. (2022). [Ecological interaction networks. What we know, what we don't, and why it matters](#). *Ecología Austral*, 32, 670–697.
- Wagenmakers, E.-J. & Farrell, S. (2004). [AIC model selection using Akaike weights](#). *Psychonomic Bulletin & Review*, 11, 192–196.
- Weinstein, B.G. & Graham, C.H. (2017a). [On comparing traits and abundance for predicting species interactions with imperfect detection](#). *Food Webs*, 11, 17–25.
- Weinstein, B.G. & Graham, C.H. (2017b). [Persistent bill and corolla matching despite shifting temporal resources in tropical hummingbird-plant interactions](#). *Ecology Letters*, 20, 326–335.
- Whittaker, R.H. (1960). [Vegetation of the Siskiyou Mountains, Oregon and California](#). *Ecological Monographs*, 30, 279–338.
- 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.
- Young, J.-G., Valdovinos, F.S. & Newman, M.E.J. (2021). [Reconstruction of plant–pollinator networks from observational data](#). *Nature Communications*, 12, 3911.