Deciphering probabilistic species interaction networks

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

[1] Université de Montréal [2] Université de Sherbrooke [3] Quebec Centre for Biodiversity Science [4] University of Sheffield [5] University of Toronto [6] 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?) conveys uncertainties in our knowledge of interactions and information on their variability. 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 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 or ecological feasibility (regional). 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), conditional variables and uncertainty sources.

Last revision: *December 16, 2024*

Introduction

Species interactions are variable and uncertain

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 leading to uncertainty in our knowledge of interactions. For example, the spatial and temporal uncoupling of species (e.g., nocturnal and diurnal species coexisting in the same space with different daily activity timings, Jordano 1987) and the large number of rare and cryptic interactions in a community, contribute to these knowledge gaps by making it more difficult to observe interactions (Jordano 2016).

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 within a defined time frame (Poisot *et al.* 2015). Probabilities of interspecific encounters are typically low, especially for rare species with low abundances (Canard *et al.* 2012; Canard *et al.* 2014; Vázquez *et al.* 2007). 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, such as their phenotypes (Bolnick *et al.* 2011; Gravel *et al.* 2013; Stouffer *et al.* 2011) and behavior (Choh *et al.* 2012; Pulliam 1974), are locally compatible in that specific environment (Poisot *et al.* 2015). Because these conditions are not consistently met locally, there will inevitably be instances where interactions will be observed and others where they will not.

Documenting the location and timing of interactions is difficult when accounting for the spatiotemporal variability of ecological interactions (Poisot *et al.* 2012, 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. 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), habitat characteristics (e.g., presence of refuges where prey can hide from predators, Grabowski 2004), and land use change (Tylianakis *et al.* 2007), contribute to this spatiotemporal variability by impacting species abundance and traits. Interactions may also be influenced by a third species (e.g., a more profitable prey species, Golubski & Abrams 2011; Sanders & van Veen 2012). Even under favorable circumstances, 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).

We distinguish the variability of interactions from their uncertainty. Interaction variability is defined as the changes in the occurrence or strength of interactions along spatial, temporal, or environmental axes (Poisot *et al.* 2015). It is a property of interactions that should be quantified if we aim for a comprehensive understanding of ecological networks. Stochasticity is the inherent randomness or unpredictability of interactions that lead to this variability. Conversely, uncertainty is defined as a lack of knowledge about the occurrence of interactions. When using statistical models to infer interactions, uncertainty sources include input data, parameter, and model structure uncertainties (Simmonds *et al.* 2024). Input data uncertainty arises from our inability to empirically observe all interactions and from measurement errors in environmental and biological variables used for inference. Parameter uncertainty represents a plausible range of values for a parameter whose exact value is unknown. For example, we may calculate a range of plausible values for interaction variability (e.g., there could be a 50% certainty that an interaction occurs 50% of the time). Model structure uncertainty recognizes that different statistical models may adequately predict interactions. In contrast to variability, uncertainty can be reduced by sampling additional data (except for model uncertainty, which will persist regardless of sampling effort). Simmonds *et al.* (2024) underscores the importance of quantifying and reporting these diverse sources of uncertainty, alongside ensuring their appropriate propagation to model output (such as predicted interactions) and higher-level measures (such as network structure). While recognizing that these definitions may not be universally accepted, clarifying the distinction between variability and uncertainty enables us to better comprehend the sources of our knowledge gaps about ecological interactions.

Species interactions as probabilistic objects

The recognition of the intrinsic variability and uncertainty of species interactions has led ecologists to expand their representation of ecological networks to include a probabilistic view of interactions (Dallas *et al.* 2017; Fu *et al.* 2021; Poisot *et al.* 2016). This allows filling in the Eltonian shortfall (i.e., the gap between our current knowledge and a comprehensive understanding of interactions, 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 (Carlson *et al.* 2021), especially in places where access and resources for research are scarce. A probability is a measure of how likely a specific outcome is, based on both the uncertainty and variability of interactions. Interaction probabilities may be uncertain when there is a distribution of plausible probability values. The probabilistic representation of interactions has been applied 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). By accounting for the uncertainty and variability of direct interactions, networks of probabilistic interactions (which differ from *probabilistic networks* describing the uncertainty and variability of the *whole* network) may provide a more realistic portrait of species interactions.

Probabilistic interactions differ from binary interactions. Networks of probabilistic interactions, within a Bayesian perspective, express our degree of belief (or confidence) regarding the occurrence or observation of interactions. In a frequentist approach, they represent the expected relative frequencies of interactions over many repeated trials or sampling events. In contrast, interactions are simply regarded as either occurring or not in networks of deterministic binary interactions. Based on the scale at which they are estimated, interaction probabilities may reflect our level of confidence in whether interactions will be observed, realized locally, or biologically feasible. 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 duration, thereby diminishing the uncertainty of our knowledge of 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). 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). Understanding the nuances between probabilistic and binary interactions is essential for accurately modeling and interpreting ecological networks.

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), helping us identify places where sampling is most needed to reduce this uncertainty. Statistical models can also predict networks without prior knowledge of pairwise interactions. They may do so using body size (Caron *et al.* 2024; Gravel *et al.* 2013; Petchey *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. Before being used to test ecological hypotheses, predicted networks must be validated against empirical data (Brimacombe *et al.* 2024), which could be sampled strategically to optimize the validation process. 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), are examples of common probabilistic interaction models. Null models can produce underlying distributions of network measures for null hypothesis significance testing. 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. 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 networks. These models and measures support the use of probabilistic interactions for the study of a wide range of ecological questions, from making better predictions of species distributions (Cazelles *et al.* 2016) to forecasting the impact of climate change on ecological networks (Gilman *et al.* 2010).

We lack a clear understanding of probabilistic species interactions

We still lack a precise definition of probabilistic interactions, which makes the estimation and use of these data more difficult. In this manuscript, 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: local networks describing probabilities of realized interactions, and regional networks (metawebs) describing probabilities of potential interactions. We highlight the distinctions in the ecological meaning of these two representations of interactions and examine their properties and relationships (particularly with space, time, and between each other).

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 reporting standard that could guide the documentation of all types of probabilistic interactions (Salim *et al.* 2022 discuss data standards for deterministic mutualistic networks). Clear reporting standards 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. Data documentation should outline the nature (i.e., local or regional) and type (e.g., predatory or pollination) of 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 networks. These broad principles remain relevant and applicable across different types of direct interactions. In the following sections, we discuss the definitions, conditions, and estimation of probabilistic interactions as we scale up from pairwise interactions to interactions within local and regional networks.

Pairwise interactions: the building blocks of ecological networks

What are probabilistic interactions?

Consider a scenario where an avian predator has just established itself in a northern habitat home to a small rodent. Suppose their interaction has not been previously observed, either because these species have never co-occurred before or because previous sampling failed to detect an interaction despite their co-occurrence. 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 could 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 could 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 could 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 the probability of interaction between these two species, whether through predictive models or informative prior probabilities, hinges on our understanding of these probabilities and the specific ecological processes we aim to capture.

An important aspect to consider when estimating or using interaction probabilities is knowing if they describe the probability of potential or realized interactions, as these two types of interactions have distinct meanings and sources of uncertainty and variability. A potential (regional) interaction is defined as the biological or ecological capacity of two taxa to interact (i.e., the probability that they interact if they were to encounter each other, given sufficient time and appropriate environmental conditions) 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 enough time, the probability of local interaction tends toward the probability of regional (potential) interaction. A longer duration increases the probability that species will eventually encounter each other and that local environmental conditions supporting an interaction will occur, provided that species have the biological capacity to interact. Recognizing the distinction between probabilistic regional and local interactions is crucial for accurately interpreting interaction probabilities in ecological networks.

We use the terms *metaweb* (Dunne 2006) to designate regional networks of potential interactions and *local networks* (Poisot *et al.* 2012) for those of realized interactions. Metawebs are the network analogs of the species pool, where local networks originate from a subset of both species (nodes) and interactions (edges) of the regional metaweb (Saravia *et al.* 2022). Without clear documentation, it can be challenging to know if published probabilistic interaction networks describe local or regional interactions. When probabilistic local interactions are used and interpreted incorrectly as regional interactions (and conversely), this may generate misleading findings during data analysis. A better understanding of probabilistic local and regional interaction networks would facilitate a more adequate use of interaction data (e.g., when studying network-area relationships in local networks and metawebs) and prevent misinterpretations of the biological meaning of probabilistic interactions.

What is the outcome of probabilistic interactions?

The outcome of probabilistic interactions is usually binary

Local networks and metawebs, like any type of network, are made of nodes and edges that may be represented at different levels of organization. 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).

Ecologists have traditionally represented interactions (edges) 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 $B_{i,j}$ between two taxa can be viewed as the result of a Bernoulli trial $B_{i,j} \sim \text{Bernoulli}(\phi)$, with $\phi = P(B_{i,j} = 1)$ being the probability of interaction. This interaction probability characterizes our limited ecological knowledge and/or the intrinsic spatiotemporal variability of the interaction. 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 $P(B_{i,j} = 1)$ (which we denote as $P(B_{i,j})$ for simplicity and better readability) are probabilistic events whose only two possible outcomes are the presence ($B_{i,j} = 1$) or absence ($B_{i,j} = 0$) of an interaction between each pair of nodes. Depending on the type of probabilistic interaction network (local network or metaweb), the mathematical formulation and interpretation of stochastic parameters like $P(B_{i,j})$ can be linked to environmental and biological factors such as species abundances, species traits, area, and time, for example using logistic regression with continuous explanatory variables. This allows us to model the probability that at least two individuals interact under these conditions.

The variability of an interaction determines the number of networks in which it occurs. This number can be predicted by using a Binomial distribution, assuming a constant interaction probability and independence between interactions in different networks (trials). When considering uncertainties around the estimation of $P(B_{i,j})$, 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 uncertainty in their abundances would introduce uncertainty in the interaction probability at the local scale. If we take into account the uncertainty 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 $P(B_{i,j})$, informing us of the conditions that enabled them to interact locally. By sampling binary interactions in different contexts, we can thus estimate their local variability more precisely.

The outcome of probabilistic interactions may also be quantitative

Even though binary interaction networks constitute a highly valuable source of ecological information (Pascual *et al.* 2006), they overlook interaction strengths. Represented in a quantitative adjacency matrix W, interaction strengths better describe the energy flows, demographic impacts or frequencies of interactions between nodes (Berlow *et al.* 2004; Borrett & Scharler 2019), with $W_{i,j}$ being a natural $\mathbb{N}$ or real $\mathbb{R}$ number depending on the measure. For example, they may represent local interaction rates (e.g., the flower-visiting rates of pollinators in a mutualistic network, Herrera 1989). Relative frequencies of interactions may be used as a measure of both the strength and probability of local interactions that are biologically feasible. When interaction strengths characterize predation pressure on prey, they can serve as parameters 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 (Strydom *et al.* 2021), especially when using predictive models that quantify the uncertainty and variability of quantitative interactions (Berlow *et al.* 2004). However, if two taxa are repeatedly found together without interacting, there may be more uncertainty about their capacity to interact than their interaction strength (which would assuredly be close to 0).

Like binary interaction networks, the uncertainty and variability of interaction strengths can be represented probabilistically. Interaction strengths can follow many probability distributions depending on the measure. For instance, they can follow a Poisson distribution $W_{i,j} \sim \text{Poisson}(\lambda_{i,j} t_0)$ when predicting the number of interactions between individuals during a time interval t_0, with $\lambda_{i,j}$ being the expected rate at which individuals of taxa i and j interact (e.g., the expected number of prey j consumed by all predators i). The Poisson distribution can also be 0-inflated when taking into account 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 networks (Jordano 2016). Regardless of the measure, estimating the uncertainty of quantitative interactions enables us to consider a range of possible values of interaction strength.

Because of the methodological difficulties typically encountered when building deterministic quantitative networks, binary interaction networks, which are usually easier to sample (Jordano 2016) and predict (Strydom *et al.* 2021), have been more frequently studied and used. Mathematical models such as Ecopath (Plagányi & Butterworth 2004) partially mitigate these difficulties, but the number of biological parameters required to make predictions hinders their application in many systems. Moreover, most published probabilistic interaction networks (e.g., Strydom *et al.* 2022) and methods (e.g., Poisot *et al.* 2016) involve probabilistic interactions whose outcome is binary. This underlines the need for better guidelines on the interpretation and manipulation of probabilistic interactions with binary outcomes first, to ensure the appropriate use of these networks and methods. For these reasons, the primary focus of the remainder of this manuscript is on the interpretation of interaction probabilities that determine the presence or absence of interactions, in both local networks and metawebs.

Local networks: communities interacting in space and time

What are local probabilistic interactions?

Local networks of probabilistic interactions describe how likely taxa are to interact in a local context. Local interactions are contingent upon the environmental conditions experienced by the community and the matching of taxa's local biological traits. In local networks, edges commonly represent our degree of belief that two taxa interact in nature, but can also represent the probability of empirically *observing* this interaction (Catchen *et al.* 2023). Realized interactions occur locally without necessarily being observed (two locally interacting taxa may or may not be seen interacting during sampling), whereas observed interactions are those that have been locally recorded. Local interactions may thus arise from both the ecological (realized interactions) and sampling (observed interactions) processes taking place locally.

Local networks are delineated within a particular location and time. 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 vary 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. On the other hand, 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 probabilities of interactions, which can be obtained by integrating probability densities over space and time. We can quantify both an area A_0 and a duration t_0 with these definitions. By studying probabilistic local interaction networks, we may thus conduct spatiotemporal analyses of local interactions (Box 1), enhancing our understanding of interactions occurring in distinct environmental contexts.

What are local probabilistic interactions conditioned on?

Local interactions may be conditioned on co-occurrence

The probability that two taxa i and j interact in a local network $L_{x,y,z,t}$ (spatial and temporal subscripts hereafter replaced by the shorter subscript k for clarity) can be conditioned on many environmental and biological factors. In addition to network area (or volume) A_0 and duration t_0, they may be conditioned on taxa co-occurrence $X_{i,j,k}$, which is usually Boolean, describing if the geographic distributions of both taxa overlap within the study area. As illustrated in Box 1, co-occurrence may be modeled probabilistically, in which case it may conform to a Bernoulli distribution $X_{i,j,k} \sim \text{Bernoulli}(\phi)$, where $\phi = P(X_{i,j,k} = 1)$. The probability of co-occurrence can be calculated using the individual (marginal) occurrence probabilities $P(X_{i,k} = 1)$ and $P(X_{j,k} = 1)$ (which we denote as $P(X_{i,k})$ and $P(X_{j,k})$ for simplicity and better readability). Given that taxa occurrences are not independent of each other, the probability of co-occurrence can be calculated by multiplying the probability of occurrence of one taxon by the probability of occurrence of the other given that the first one is present:

$$P(X_{i,j,k}) = P(X_{i,k}, X_{j,k}) = P(X_{i,k}|X_{j,k})P(X_{j,k}). \tag{1}$$

Note that to keep the text concise and readable, the probability notation used in this manuscript implicitly assigns a value of 1 to binary variables (e.g., in eq. 1 the term $P(X_{i,k}|X_{j,k})$ is short for $P(X_{i,k} = 1|X_{j,k} = 1)$), unless stated otherwise. The value is only stated explicitly when it is 0 or when we wish to emphasize the value of 1.

The probability of co-occurrence $P(X_{i,j,k})$ (short for $P(X_{i,j,k} = 1)$) 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., $P(L_{i,j,k} = 1|X_{i,j,k} = 0) = 0$), the probability of local interaction can be obtained by multiplying the probability of interaction given co-occurrence with the probability of co-occurrence:

$$P(L_{i,j,k} = 1) = P(L_{i,j,k} = 1 | X_{i,j,k} = 1) \times P(X_{i,j,k} = 1). \tag{2}$$

Knowing that two taxa co-occur improves our estimation of the probability that they interact locally by mitigating a potential source of uncertainty.

Local interactions may be conditioned on different environmental and biological factors

Local interactions may also be conditioned on local environmental factors such as temperature (Angilletta *et al.* 2004), precipitation (Woodward *et al.* 2012), habitat structure (Klecka & Boukal 2014), and the presence or abundance of other taxa in the network (Kéfi *et al.* 2012; Pilosof *et al.* 2017). We use the variable E_k to describe the local environmental context in which interaction probabilities were estimated. For example, in a mesocosm experiment estimating interaction probabilities between predators and prey with and without refuges, E_k would represent the presence or absence of these refuges. Like co-occurrence, E_k can also be modeled probabilistically when the variability or uncertainty of environmental factors is considered. E_k represents all environmental variables that were taken into consideration when measuring interaction probabilities; it is therefore a subset of all environmental factors acting on ecological interactions.

Other important factors that can impact interaction probabilities at the local scale are taxa local abundances $N_{i,k}$ and $N_{j,k}$, which affect encounter probabilities (Canard *et al.* 2012), and local traits $T_{i,k}$ and $T_{j,k}$ (e.g., movement rates, Beardsell *et al.* 2021; Cherif *et al.* 2024), which may also impact encounter probabilities as well as the ability of individuals to interact after encountering each other (Caron *et al.* 2024; Poisot *et al.* 2015). Local interaction probabilities may also be conditioned on higher-level properties of the community (i.e., the emerging structure of ecological networks), which we denote by $f(L_k)$. 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). Biological factors, whether at the scale of individual taxa pairs or the community, may thus impact how we estimate and define interaction probabilities.

Local interactions may be conditioned on biological feasibility

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

$$P(L_{i,j,k} = 1) = P(L_{i,j,k} = 1|M_{i,j} = 1) \times P(M_{i,j} = 1), \tag{3}$$

assuming that $P(L_{i,j,k} = 1|M_{i,j} = 0) = 0$.

Low values of $P(L_{i,j,k}|M_{i,j})$ 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. The local probability of interaction between a given pair of taxa is thus always equal to or below their probability of regional interaction. Taking into account biological feasibility in our estimation of local interaction probabilities leverages information from the metaweb to better predict the local occurrence of interactions (Dansereau *et al.* 2024; Strydom *et al.* 2021).

Conditional variables must be explicitly stated

The probability that two taxa i and j interact in a local network L_k can thus be conditioned on their co-occurrence $X_{i,j,k}$ (or more explicitly on their occurrences $X_{i,k}$ and $X_{j,k}$), local abundances $N_{i,k}$ and $N_{j,k}$, local traits $T_{i,k}$ and $T_{j,k}$, local environmental conditions E_k, network area (or volume) A_0, time interval t_0, network properties $f(L_k)$, and biological feasibility $M_{i,j}$. When these conditions are absent from an expression, it may be because they have been marginalized over, which would be reflected in the overall uncertainty of the interaction. Interaction probabilities may also have been implicitly conditioned on missing variables (e.g., when estimated for specific values of these variables without explicitly including them as conditions), potentially impacting our interpretation. The local probability of interaction is described by the following expression when all of these conditional variables are included:

$$P(L_{i,j,k}|X_{i,k}, X_{j,k}, N_{i,k}, N_{j,k}, T_{i,k}, T_{j,k}, E_k, A_0, t_0, f(L_k), M_{i,j}). \tag{4}$$

These conditional variables do not all need to be considered at all times. 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. For example, Gravel *et al.* (2019) analyzed local European host-parasite networks of willow-galling sawflies and their natural enemies, all referenced in space and time, to infer probabilities of local interactions between co-occurring species. This was achieved by including temperature and precipitation as conditional variables in their models. In Box 2, we reuse these data to show the extent of variation among these local networks. We do so by measuring their dissimilarity with the regional network (metaweb aggregating all local interactions), both in terms of species composition and interactions. We built local probabilistic networks following eq. 3, showing that insufficient local variation (high probability of local interaction among potentially interacting species) results in an overestimation in both the number of interactions and connectance (i.e., the proportion of all of the non-forbidden links that are realized). This analysis was conducted for illustrative purposes, and other conditional variables could have been used to make these comparisons.

When accounted for, conditional 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 ($P(L_{i,j,k}|X_{i,j,k})$) or not ($P(L_{i,j,k})$) of co-occurrence in their estimation of local interaction probabilities, as this can change our interpretation of the data and understanding of potential uncertainty sources. Reporting the scale and level of aggregation of the data enables us to more accurately study the underlying ecological processes (Clark *et al.* 2011) and manipulate or propagate uncertainty to different aggregation levels (Simmonds *et al.* 2024). In Tbl. 1, we present examples of studies that used different expressions of probabilistic interactions with different conditional variables. We included in this table the probability of empirically observing an interaction that is realized locally $P(O_{i,j,k}|L_{i,j,k})$ to underscore the distinction between local observations and actual realizations 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 network L_k of realized interactions, a local network O_k of observed interactions, a metaweb M of potential interactions (representing the *biological* feasibility of interactions), and a metaweb M^* of potential interactions (representing the *ecological* 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 models 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 networks. The boxes in our study that discuss these expressions are also specified.

Expression	Type	Outcome	Common models	Reference
$P(L_{i,j,k} \vert X_{i,k}, X_{j,k}, ...)$	local	realization of the interaction given taxa co-occurrence	species distribution models	Gravel *et al.* (2019), Dansereau *et al.* (2024), Boxes 1 and 5
$P(L_{i,j,k} \vert N_{i,k}, N_{j,k}, ...)$	local	realization of the interaction given taxa abundances	neutral models	Canard *et al.* (2014)
$P(L_{i,j,k} \vert T_{i,k}, T_{j,k}, ...)$	local	realization of the interaction given local traits	trait matching models	Caron *et al.* (2024), Box 4
$P(L_{i,j,k} \vert E_k, ...)$	local	realization of the interaction given local environmental conditions	environmental-based models	Gravel *et al.* (2019) (temperature and precipitation)
$P(L_{i,j,k} \vert A_0, ...)$	local	realization of the interaction in a given area or volume	spatial models	Galiana *et al.* (2018) *, Box 3
$P(L_{i,j,k} \vert t_0, ...)$	local	realization of the interaction during a given time period	temporal models	Weinstein & Graham (2017a), Boxes 1 and 3
$P(L_{i,j,k} \vert f(L_k), ...)$	local	realization of the interaction given network structure	topological models	Fortuna & Bascompte (2006) (connectance), Stock *et al.* (2017)
$P(L_{i,j,k} \vert M_{i,j}, ...)$	local	realization of the interaction given that it is biologically feasible	spatiotemporal models	Dansereau *et al.* (2024), Boxes 2, 3, and 5

Expression	Type	Outcome	Common models	Reference
$P(O_{i,j,k} \| L_{i,j,k}, ...)$	local	observation of the interaction given that it is realized locally	sampling models	Catchen *et al.* (2023)
$P(M_{i,j} \| T_i, T_j)$	regional	biological feasibility of the interaction given regional traits (non-forbiddenness)	trait matching models	Strydom *et al.* (2022), Box 4
$P(M^*_{i,j} \| T_i, T_j, E)$	regional	ecological feasibility of the interaction given regional traits and environmental conditions	trait matching and environmental-based models	this study

How are local probabilistic interactions estimated?

Various statistical models can be used to estimate local interaction probabilities, some of which are presented in Tbl. 1. These models can be based on multiple conditional variables. 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 $X_{i,j,k}$ and E_k on spatial and temporal scales. When estimating interaction probabilities using e.g. a generalized linear model with multiple explanatory variables that might not all be independent, it may become important to address collinearity. In such cases, it may be necessary to use variable selection techniques before fitting the model to data to mitigate this issue. Other challenges and opportunities associated with predictive models of species interactions are reviewed in Strydom *et al.* (2021).

When using multiple competing models to estimate local interaction probabilities, rather than selecting a single model that best fits the data, model averaging may enhance our estimations. 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, given two competing models mod_1 and mod_2 with respective probabilities (or weights) $P(mod_1)$ and $P(mod_2)$, the average probability of interaction $P(L_{i,j,k})$ can be calculated as follows:

$$P(L_{i,j,k}) = P(L_{i,j,k}|mod_1) \times P(mod_1) + P(L_{i,j,k}|mod_2) \times P(mod_2). \quad (5)$$

Model averaging takes into account the uncertainty of model structure in our estimation of local interaction probabilities. Regardless of the model used for prediction, it is crucial to quantify and disclose all sources of uncertainty to understand better the validity and limitations of our predictions (Simmonds *et al.* 2024).

Box 1: A spatiotemporally explicit model of interactions

Ecologists may resort to predictive models to reconstruct local networks across time and space. We introduce and develop a simple generative Bayesian model for probabilistic local interactions, which explicitly accounts for their spatiotemporal variability. Our model is not designed for regional interactions, which do not vary spatially nor temporally. Rather, it could prove valuable for generating new data on local interactions across time and space, following parameter inference.

As indicated by eq. 2, the probability that two taxa i and j interact locally can be obtained by multiplying their probability of interaction given co-occurrence with their probability of co-occurrence. The probability of interaction given co-occurrence can be made temporally explicit by modeling it as a Poisson process with rate parameter λ_k. This parameter represents the local expected frequency of interaction between co-occurring taxa. The probability that two co-occurring taxa interact at least once during a time interval t_0 can be given by:

$$P(L_{i,j,k} = 1 | X_{i,j,k} = 1) = 1 - e^{-\lambda_k t_0}, \tag{6}$$

which tends toward 1 as $t_0 \to \infty$ if $\lambda_k > 0$. In other words, two co-occurring taxa with a nonzero rate of interaction will inevitably interact at least once in a sufficiently long time interval.

The occurrence of an interaction between i and j may be the result of a Bernoulli trial with parameter ϕ representing the probability of interaction $P(L_{i,j,k} = 1)$. A Bayesian model can be built using the preceding equations to generate new interaction data, following the inference of the λ_k and ϕ parameters.

$$L_{i,j,k} \sim \text{Bernoulli}(\phi) \tag{7}$$

$$\phi = P(X_{i,j,k} = 1)(1 - e^{-\lambda_k t_0}) \tag{8}$$

$$P(X_{i,j,k}) \sim \text{Beta}(2, 2) \tag{9}$$

$$\lambda_k \sim \text{Exponential}(2) \tag{10}$$

In Fig. 1, we show the variation in the probability of interaction under different parameter values. In the right panel, we notice that the probability of interaction always converges toward the probability of co-occurrence $P(X_{i,j,k} = 1)$, for all positive values of the interaction rate.

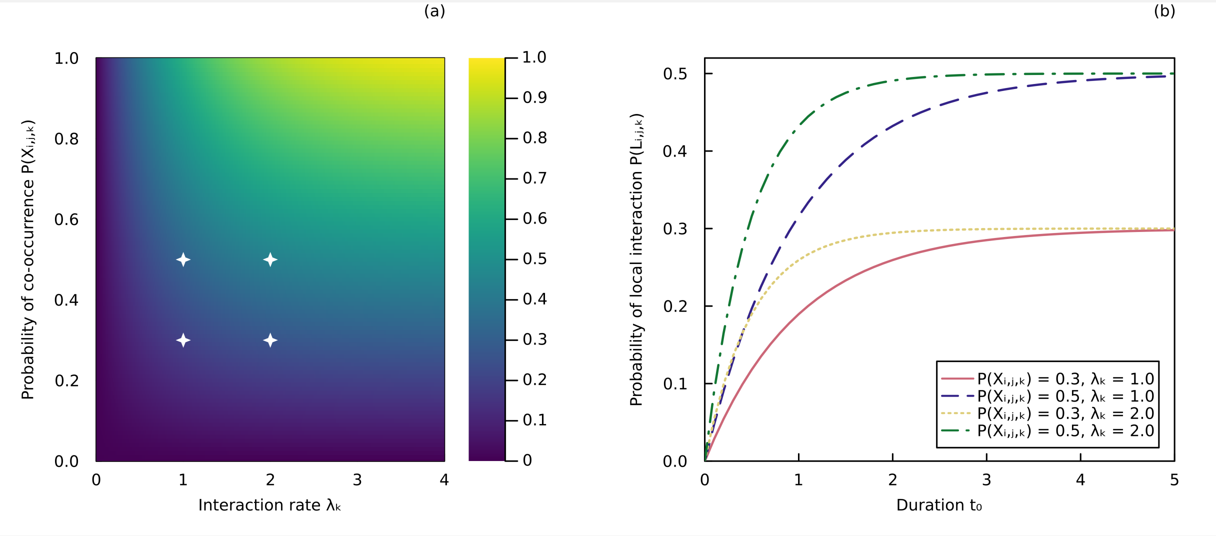

Figure 1: **Parameters of the spatiotemporally explicit model of interactions.** (a) Probability of local interaction $\phi = P(L_{i,j,k})$ (short for $P(L_{i,j,k} = 1)$) given by the process model (eq. 8) under different values of λ_k (interaction rate) and $P(X_{i,j,k})$ (probability of co-occurrence, short for $P(X_{i,j,k} = 1)$), with $t_0 = 1$ (duration). The probability of local interaction represents the probability that the two taxa will interact at least once within the given time interval. Parameters t_0 and λ_k have complementary units (e.g., t_0 in months and λ_k in number of interactions per month). 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, for different values of λ_k and $P(X_{i,j,k})$.

This model can be customized in different ways, such as linking both parameters to specific environmental or biological variables. For instance, the probability of co-occurrence could be modeled as a function of climatic variables, while the interaction rate parameter could be modeled based on taxa abundances.

Box 2: Dissimilarity of local host-parasite networks

We present a way to assess local network variability and dissimilarity regarding species composition and interactions. We do so by comparing local tripartite host-parasite networks to the metaweb using data from Kopelke *et al.* (2017). This collection of networks consists of interactions between willows, willow-galling sawflies, and their natural enemies sampled across Europe. All data manipulation and methods are described in Appendix 1. All code and data to reproduce these analyses are available on Zenodo (https://doi.org/10.5281/zenodo.12802326).

In Fig. 2a-b, we show how the dissimilarity between the metaweb of binary interactions and aggregated local networks changes with the number of sampled local networks. We compared the metaweb and the aggregated local networks using the dissimilarity in species composition (β_S, Fig. 2a) and the dissimilarity of interactions between common species (β_{OS}, Fig. 2b) indices (Poisot *et al.* 2012). Expectedly, local networks are highly dissimilar from the metaweb in terms of species composition, especially when only a limited number of sites have been sampled. This is because few species from the metaweb (species pool) occur locally. Moreover, we observe a peak in the dissimilarity of interactions between common species at intermediate sampling levels. This suggests that species are collected faster than their interactions. With a limited number of sampled local networks, few regional interactions are observed locally. Adding more sites brings new species, but not always their interactions. Quadratic relationships of network properties with sampling effort were also observed by McLeod *et al.* (2021).

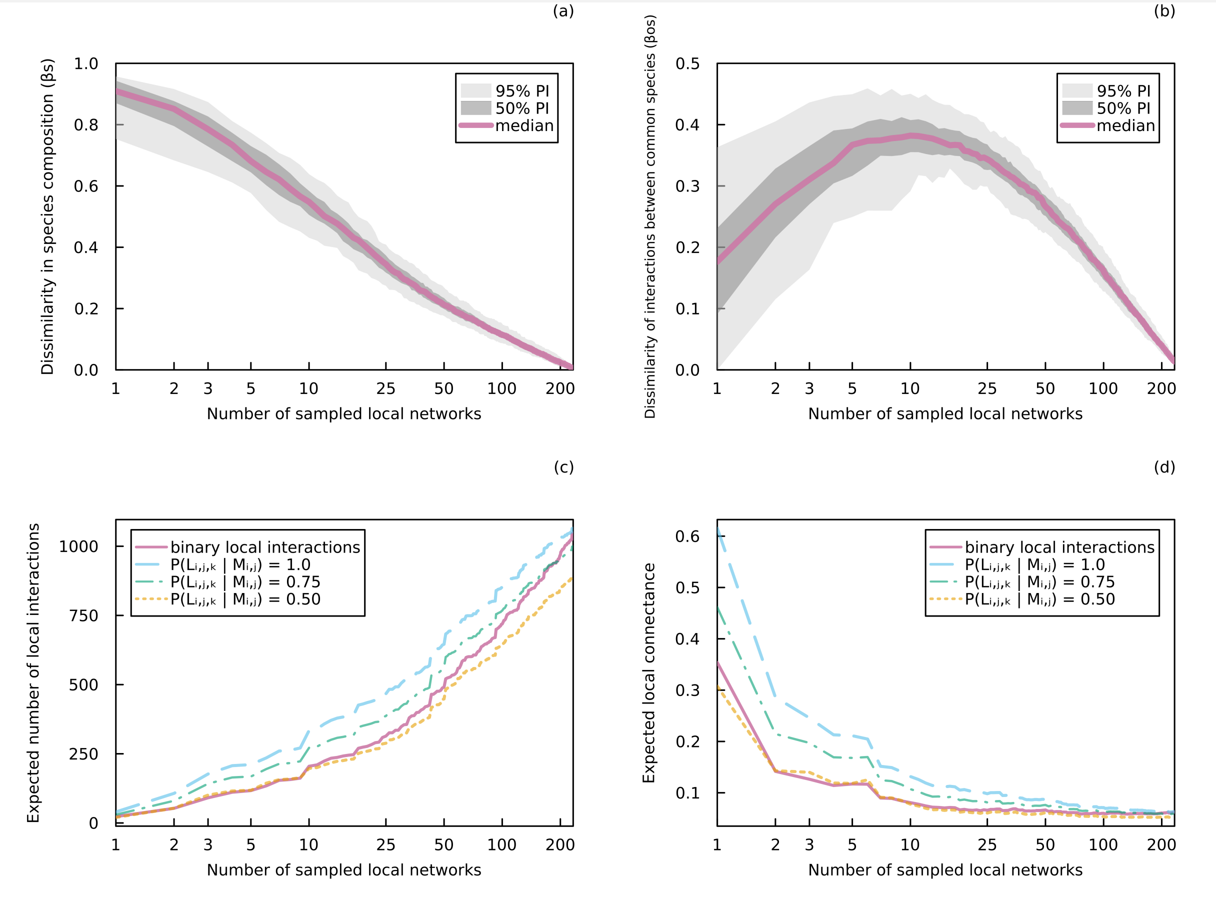

Figure 2: **Network accumulation curves.** (a) Dissimilarity in species composition and (b) dissimilarity of interactions between common species between aggregated local networks and the metaweb of binary host-parasite interactions. In both panels, the colored line represents the median dissimilarity across simulations and the grey areas cover the 50% and 95% percentile intervals. (c) Scaling of the number of interactions and (d) scaling of connectance with the number of sampled (aggregated) binary and probabilistic local interaction networks. For a better comparison with binary interactions, local networks of probabilistic interactions were derived from a metaweb of probabilistic interactions with a false positive and false negative rate of zero. A specific value of $P(L_{i,j,k}|M_{i,j})$ (the local probability of interaction among potentially interacting species) was used for all non-aggregated local networks within a particular curve. Aggregated local networks were obtained by sequentially and randomly selecting a number of local networks and aggregating both their species and interactions (with the value of $P(L_{i,j,k}|M_{i,j})$ increasing in aggregated local networks of probabilistic interactions).

Next, we investigate how the number of local interactions and connectance scale with the number of sampled (aggregated) local networks of probabilistic or binary interactions (Fig. 2c-d). By comparing the scaling relationships observed in local networks of binary and probabilistic interactions, we observe that high values of $P(L_{i,j,k}|M_{i,j})$ (short for $P(L_{i,j,k} = 1|M_{i,j} = 1)$) lead to systematic overestimations in the number of interactions and connectance, especially when $P(L_{i,j,k}|M_{i,j}) = 1$ (i.e., when local and regional probabilities of interactions are equivalent). This suggests that high values of $P(L_{i,j,k}|M_{i,j})$ do not adequately capture the variability of local interactions. However, these biases tend to diminish as the number of sampled networks increases, indicating that most interactions are eventually captured when $P(L_{i,j,k}|M_{i,j})$ is high. In contrast, low values of $P(L_{i,j,k}|M_{i,j})$ lead to missing interactions, resulting in an underestimation of the number of interactions and connectance. These results underscore the importance of using the appropriate level of variability when estimating local interaction probabilities.

Metawebs: regional catalogs of interactions

What are regional probabilistic interactions?

Metawebs (Dunne 2006) are networks of potential interactions over broad spatial, temporal, and taxonomic scales (e.g., food webs at the continental scale). They correspond to the temporal and spatial asymptotes of local interactions (Box 1). Potential interactions describe the biological capacity of taxa to interact under optimal or feasible environmental conditions given enough time, which is typically assessed at the regional scale. Metawebs of probabilistic interactions are particularly useful in situations where there is uncertainty in the ability of taxa to interact (Strydom *et al.* 2023). They may also be used as informative priors of local interactions. Therefore, building a metaweb of probabilistic interactions may be an important first step before predicting networks at finer scales.

In contrast to local networks, where interaction probabilities arise from the variability of interactions and the lack of information on the conditions, interaction probabilities in metawebs solely result from a lack of knowledge. This uncertainty 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, either taking a value of 1 (observing an interaction at least once) or approaching 0 (repeatedly failing to observe an interaction between co-occurring taxa). Confidently observing an interaction once confirms its biological feasibility, but failing to observe it (even on multiple occasions) does not ensure that it is non-feasible (e.g., due to false negatives, Catchen *et al.* 2023). While local interaction probabilities are irreducible because of local variability, the uncertainty of regional interactions reduces to 0 with the addition of information. Moreover, although *neutrally* forbidden interactions (i.e., forbidden interactions between rare species, Canard *et al.* 2012) have low probability values in local networks, they would have a probability of 1 in the metaweb (this is because the species' traits could support an interaction if they were to encounter each other at high enough abundances). Likewise, non-co-occurring taxa may have a non-zero probability of interaction in the metaweb. Regional interaction probabilities are thus fundamentally different from local interaction probabilities, both in terms of uncertainty sources and probability values.

The extent of sampling effort influences our evaluation of probabilities of regional interactions, as sampling over a larger area or for a longer duration enables us to capture a greater number of interactions (Box 1, McLeod *et al.* 2021). However, in contrast with local networks of probabilistic interactions, regional interactions are not evaluated for any particular local context (they are rather a collection of local contexts), which impacts how they scale with space and time (notably through the extent of the region covered and sampling duration). In Box 3, we discuss the differences in spatial and temporal scaling of regional interactions compared to local interactions. We do so using the host-parasite networks of Kopelke *et al.* (2017) as an illustration of spatial scaling (Box 3). Understanding the effect of spatial and temporal scales (including sampling effort) on local and regional interaction probabilities is important for effectively propagating uncertainty across scales and highlighting the fundamental differences between these two types of networks.

What are regional probabilistic interactions conditioned on?

Regional interactions describing biological feasibility are conditioned on traits

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. We define regional traits distributions as the range of phenotypes that a taxon can express across various environments. Local traits $T_{i,k}$ and $T_{j,k}$, which vary spatially and temporally because of phenotypic plasticity and local environmental variability (Berg & Ellers 2010), are a subset of regional traits. A probability of potential interaction in a metaweb M describing the biological feasibility of interactions may be expressed as:

$$P(M_{i,j}|T_i, T_j), \tag{11}$$

which, in contrast with local networks, is not conditioned on any spatial, temporal, co-occurrence or environmental variables (Tbl. 1). Because phylogenetically close species often share similar traits, we should expect that closely related species will have similar interacting partners. We can thus use phylogeny to predict species traits and infer regional interactions (Eklöf & Stouffer 2016; Stouffer *et al.* 2012; Strydom *et al.* 2022). The taxonomic level at which interactions are evaluated also influences the distribution of regional traits. However, as explained in Box 4, there is no fundamental difference in the taxonomic scaling of regional and local interactions (i.e., how interaction probabilities change with taxonomic level) because they both depend on trait aggregation.

The biological feasibility of interactions expresses our degree of belief that there exists at least one combination of phenotypes that could support a specific type of interaction if they were to encounter each other, assuming they had enough 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 $P(F_{i,j}|T_i, T_j)$ of forbidden interactions (i.e., the probability that their traits do not support an interaction), which is based uniquely on biological traits:

$$P(M_{i,j}|T_i, T_j) = 1 - P(F_{i,j}|T_i, T_j). \tag{12}$$

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 distance of lemmings to rattlesnakes' prey. Doing so, we might find a high probability of potential interaction based on these traits. This example illustrates how regional interactions describing biological feasibility may be estimated solely based on traits, without taking into account environmental conditions (which could be important to consider when e.g. an interaction is forbidden at all temperature values).

Regional interactions describing ecological feasibility are conditioned on traits and environmental conditions

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 the ecological feasibility of interactions may be expressed as:

$$P(M^{*}_{i,j}|T_i, T_j, E), \tag{13}$$

where E is the environmental conditions under which potential interactions are evaluated (Tbl. 1). Unlike E_k, 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 enough time to interact. Incorporating environmental conditions into a trait-matching model may be important when there is high covariance between the environment and 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:

$$P(M^{*}_{i,j}|T_i, T_j) = \int_E P(M^{*}_{i,j}|T_i, T_j, E) g(E|T_i, T_j) dE \leq P(M_{i,j}|T_i, T_j), \tag{14}$$

where $g(E|T_i, T_j)$ is the conditional probability density function of E given T_i and T_j.

The difference between these two regional probabilities of interaction (across all environmental conditions) arises because the biological feasibility of an interaction is a prerequisite for its ecological feasibility. In other words, biological feasibility is necessary but not sufficient for an interaction to be ecologically feasible. Our discussion of metawebs focuses 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).

How are regional probabilistic interactions estimated?

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 interaction data (e.g., Gravel *et al.* 2019; Saravia *et al.* 2022), trait-matching models (e.g., Shaw *et al.* 2024; Strydom *et al.* 2022), and expert knowledge, which is not a trivial challenge. Every pair of taxa that have confidently been observed to interact at least once can be given a probability of 1 (i.e., $P(M_{i,j}) = 1$) since we know that they *can* interact. This differs from local networks of probabilistic interactions, where interaction events may remain stochastic (i.e., $P(L_{i,j,k}) < 1$) even after empirically observing interactions due to their spatiotemporal variability. Interactions that were never observed typically have low probability values in local networks and vary from low to high values in metawebs, contingent upon taxa traits distributions (reaching 0 for forbidden links). The aggregation of model predictions and data from different sources thus tends to raise the number of potential interactions in metawebs.

When using local interaction data to estimate probabilities of regional interactions, repeatedly failing to observe an interaction between two co-occurring taxa 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, $P(M_{i,j} = 1|O_{i,j,k} = 0)$, may be calculated as follows:

$$P(M_{i,j} = 1|O_{i,j,k} = 0) = \frac{P(O_{i,j,k} = 0|M_{i,j} = 1) \times P(M_{i,j} = 1)}{P(O_{i,j,k} = 0)}. \tag{15}$$

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

Observations of interactions may be false positives because of observation errors due to taxonomic misidentifications and ecological misinterpretations, such as those involving phylogenetically close species or 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 proves 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. By improving our estimation of potential interaction probabilities, we may build more reliable metawebs that adequately reflect our uncertainty on the biological feasibility of interactions.

Box 3: Spatial and temporal scaling of interactions

Local networks and metawebs have distinct relationships with space (area or volume) and time (sampling effort or duration). Local probabilities of interaction scale both spatially and temporally, because local interactions have more opportunities to be realized in larger areas and longer durations. In a larger sampling area and duration, we increase the likelihood of sampling favorable conditions for interactions to occur. If a local network of probabilistic interactions L_1 with an area A_1 is compared to a larger network L_0 with an area A_0, and A_1 is entirely nested within A_0, interaction probabilities should be lower in the smaller network, i.e. $P(L_{i,j,1}|A_1 < A_0) \leq P(L_{i,j,0}|A_0)$. However, if A_1 and A_0 are disjoint, interaction probabilities could be higher in the smaller area, contingent upon local environmental and biological conditions. In contrast, regional probabilities of interaction do not scale with space and time. The probability of two taxa potentially interacting should be the same in all metawebs in which they are present regardless of scale, provided that the data and methods used for estimation are consistent. This is because they depend solely on the biological capacity of two taxa to interact, regardless of co-occurrence and local environmental conditions. However, probabilities of regional interactions may change, tending to become more definitive, with increased sampling effort.

In Fig. 3, we show how the expected *number* of local host-parasite interactions scales with the spatial boundary of the network (represented by an expanding latitudinal window) in comparison with regional interactions. We do so using the host-parasite networks of Kopelke *et al.* (2017). The increase in the number of regional interactions is due to the inclusion of more species in a larger area. To ensure a conservative comparison between aggregated local and regional networks, we employed equal interaction probabilities (i.e., using $P(L_{i,j,k}|M_{i,j}) = 1$) in both types of network. This means that local interaction probabilities could not increase further when aggregating them. Despite this, 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 local networks. All data manipulation and methods are described in Appendix 1.

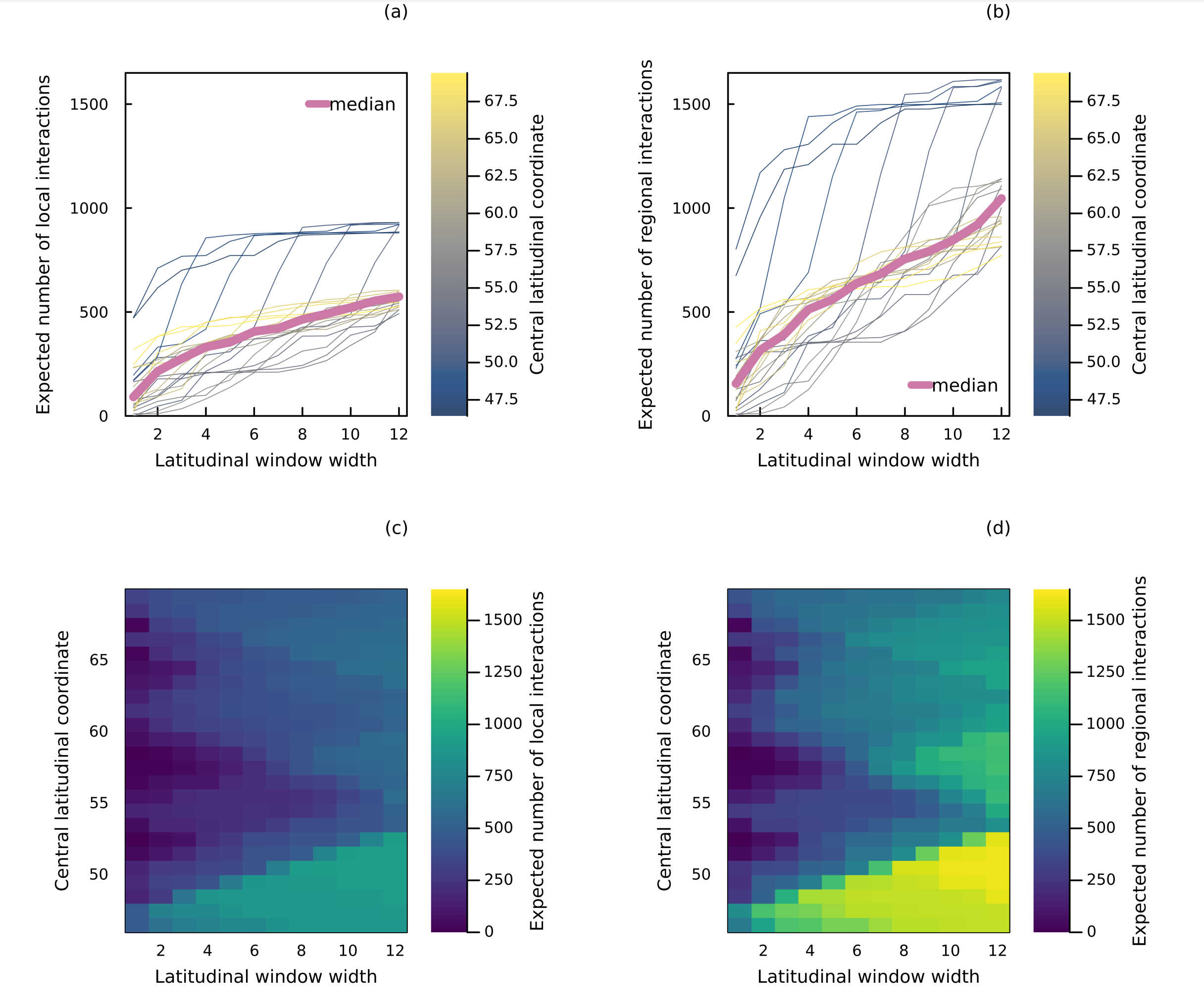

Figure 3: **Spatial scaling of interactions.** Expected number of host-parasite interactions in a network aggregating all (a) local and (b) regional probabilistic interactions within a latitudinal window of a given width. Every dashed curve corresponds to a different window centered at a given latitude (color bar), with the pink solid line representing the median number of interactions across windows. Heatmaps of the expected number of (c) local and (d) regional interactions found in windows of specified width and position (central latitude). 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 regional probabilistic interactions by setting the value of $P(L_{i,j,k}|M_{i,j})$ (the local probability of interaction among potentially interacting species) to 1. Aggregated local networks were obtained by aggregating both the species and interactions found within a particular latitudinal window, with the values of $P(L_{i,j,k}|M_{i,j})$ remaining at their maximum value of 1.

Box 4: Taxonomic scaling of interactions

Given that our interpretation of the properties of ecological networks depends on their taxonomic level (Melián *et al.* 2011), investigating the taxonomic scaling of interactions (i.e., how interaction probabilities change with taxonomic level) is important. There are no inherent differences between the taxonomic scaling of local and regional interactions. The taxonomic level of interactions impacts the definition of nodes. Local and regional interaction probabilities are not directly conditioned on taxonomic scale. However, some conditional variables (e.g., trait distribution) may covary with taxonomic scale. In such cases, local and regional interaction probabilities would change taxonomically following the scaling of these variables.

In both types of interactions, transitioning to a broader level of organization (e.g., from a species-level network S to a genus-level network G) can be done using interaction probabilities from finer scales. For example, in a network with n_1 species of genus g_1 and n_2 species of genus g_2, one can calculate the probability that at least one species from genus g_1 interacts with at least one species from genus g_2 (i.e., the probability that the genus-level interaction occurs) as follows:

$$P(G_{g_1,g_2}) = 1 - \prod_{i=1}^{n_1}\prod_{j=1}^{n_2}(1 - P(S_{g_{1,i},g_{2,j}})), \tag{16}$$

where $g_{1,i}$ and $g_{2,j}$ are the species of the corresponding genus. This equation assumes independence between species-level interactions, which may not hold true in practice due to the strong phylogenetic signal frequently encountered in species interactions (Gomez *et al.* 2010). In contrast, a different approach is necessary when transitioning from a broader to a finer level of organization. This is because the knowledge of an interaction between two genera does not guarantee that all possible pairwise species combinations will also interact. One possible method is to build a finer-scale network by generating probabilities of interaction through random sampling from a beta distribution, parameterized by the broader-scale network.

Fundamentally, the taxonomic scaling of interactions involves aggregating interactions between individuals into larger groups. Interaction probabilities at broader taxonomic scales should thus conform to probabilities of interactions between individuals. For example, Canard *et al.* (2012) built a species-based network using simulated individual-based networks. In local individual-based food webs, the probability that two individuals interact reflects our degree of belief that one individual will consume the other. Likewise, 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 will consume at least another individual from the prey species. 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).

Box 5: Sampling for binary interaction networks

Local networks of binary interactions may be predicted by performing independent Bernoulli trials for each probabilistic interaction. This is particularly useful when analyzing the structure of probabilistic interaction networks in the absence of specific analytical formulas (Poisot *et al.* 2016), even though it may introduce biases in our estimations when connectance is low (Chagnon 2015; Poisot & Gravel 2014). There are at least two techniques to sampling binary interaction networks across space, each predicting a binary interaction network for each location k within a given region. The first technique involves performing a single Bernoulli trial for each pair of taxa based on their regional probability of interaction:

$$M_{i,j} \sim \text{Bernoulli}(\phi_{i,j}),$$

where $\phi_{i,j} = P(M_{i,j} = 1)$.

In employing this technique, 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 localized networks where they co-occur, i.e. $L_{i,j,k} = M_{i,j}$ when $X_{i,j,k} = 1$. This will result in local pairwise interactions without spatial variation.

The second technique is to independently sample each local network of probabilistic interactions:

$$L_{i,j,k} \sim \text{Bernoulli}(\phi_{i,j,k}),$$

where $\phi_{i,j,k} = P(L_{i,j,k} = 1)$.

This can be achieved by first generating distinct probabilistic interaction networks for each location. Because binary interactions are sampled independently for each location, this second technique captures network structure across space and time more effectively. When sampling binary interactions from local interaction probabilities, it is crucial to sample at the same spatial scale for which probabilities were estimated to prevent systematic biases in predictions.

In Fig. 4, we compare the average connectance of binary interaction networks resulting from these two sampling techniques. We sampled regional and local interactions from our host-parasite networks of probabilistic interactions (Kopelke *et al.* 2017), generating a number of binary interaction network realizations for each site in the dataset. These two sampling techniques yield different outcomes, particularly for intermediate values of $P(L_{i,j,k}|M_{i,j})$ of 0.50, which represent instances where regional interactions do not consistently manifest locally (i.e., with the largest local variability). As anticipated, we observe that sampling binary interactions from the metaweb tends to overestimate connectance on average compared to sampling them from local networks (Fig. 4). We also observe an increase in the variability of connectance when employing a single simulation (Fig. 4a-c, cross markers), which is a more tangible representation of the process leading to the realization of local interactions in nature. All data manipulation and methods are described in Appendix 1.

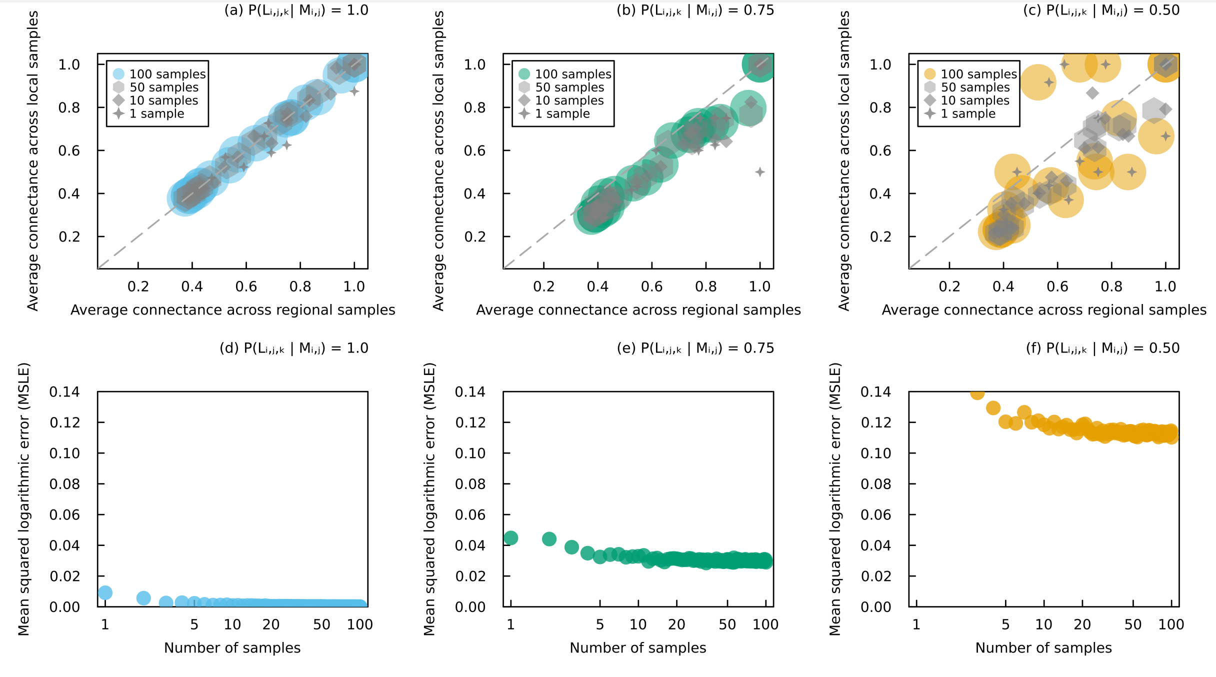

Figure 4: **Connectance of sampled binary interaction networks.** (a-c) Average connectance of binary interaction networks obtained from the two sampling techniques for 20 randomly selected host-parasite networks. Cross markers represent the connectance of a single sample for each network, diamond markers the average connectance across 10 samples, hexagon markers the average connectance across 50 samples, and the colored circles the average connectance across 100 samples (marker size proportional to the number of samples). (d-f) Reduction in the mean squared logarithmic error between the average connectance of binary interaction networks (all 233 host-parasite networks) obtained from these two sampling techniques as the number of samples increases. The local probability of interaction between potentially interacting species was set to three different values: (a,d) $P(L_{i,j,k}|M_{i,j}) = 1.0$, (b,e) $P(L_{i,j,k}|M_{i,j}) = 0.75$, and (c,f) $P(L_{i,j,k}|M_{i,j}) = 0.50$. Probabilities of regional interactions were obtained with a false positive rate of 5% and a false negative rate of 10%. Regional samples were obtained by randomly sampling binary interactions from the probabilistic interaction metaweb, and then propagating this result to all local networks that include the species potentially engaged in the interactions. Local samples were obtained by independently sampling binary interactions for each local network of probabilistic interactions.

Both sampling techniques assume independence between interactions, which might not hold true in reality. Covariation among interactions could exist even if we do not explicitly condition interactions on others. For example, the probability that two taxa interact could change with the realization of another interaction or the presence or abundance of other taxa (Kéfi *et al.* 2012; Pilosof *et al.* 2017). The consequences of this assumption of independence on the prediction of network structure have yet to be empirically examined. Sampling whole networks (or graphs) instead of pairwise interactions may eliminate the need for this assumption of independence (Battiston *et al.* 2020).

Future perspectives

In this contribution, we underline the importance of network documentation 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 show that local networks and metawebs of probabilistic interactions differ in their relationship to spatial and temporal scales (Box 3), 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 (Box 5). Clear documentation 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 facilitate reliable assessments of the spatiotemporal variability and uncertainty of biotic interactions. Here we identify key research priorities for improving our understanding of probabilistic local and regional interactions.

Predicting local networks from metawebs

Metawebs are a valuable source of ecological information for predicting local networks across time and space. Local networks of binary interactions can be reconstructed by selecting a subset of taxa and interactions from the metaweb (Dunne 2006). Determining the list of taxa to select can be achieved empirically (e.g., observed occurrence data for a site) or numerically (e.g., species distribution models). As species composition is arguably easier to sample and predict than pairwise interactions, the primary challenge lies in deciding which interactions to select from the metaweb. Inferring the structure of local networks from the metaweb before predicting local pairwise interactions could hold promise (Strydom *et al.* 2021), considering that the structure of local networks is constrained by the metaweb (Saravia *et al.* 2022).

While predicting local binary interactions from a metaweb is not be a simple task, inferring local networks of probabilistic interactions from a metaweb comes with its own set of challenges. For example, Dansereau *et al.* (2024) inferred spatially-explicit food webs from a metaweb of probabilistic trophic interactions between Canadian mammals. Their predicted localized food webs are downscaled versions of the metaweb (i.e., localized metawebs with the same interaction probabilities as those in the regional metaweb). To infer local networks as defined in this manuscript (i.e., describing local realizations of interactions), local interaction probabilities must be smaller than regional interaction probabilities. Inferring local networks from a metaweb by maintaining identical interaction probability values introduces systematic biases into the predictions, as discussed in Box 2 (unless networks are seen as downscaled metawebs).

As suggested by McLeod *et al.* (2021), metawebs establish an upper limit for local interactions (similarly for metawebs of probabilistic interactions, Strydom *et al.* 2023). In other words, the probability that two taxa interact at a specific location and time is consistently lower or equal to the probability of their regional interaction, regardless of the conditional variables considered:

$$P(L_{i,j,k}|...) \leq P(M_{i,j}|T_i, T_j). \tag{17}$$

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

$$\int_{E_k} \int_{A_0} \int_{t_0} P(L_{i,j,k}|E_k, A_0, t_0) g(E_k, A_0, t_0) \, \mathrm{d}t_0 \, \mathrm{d}A_0 \, \mathrm{d}E_k \leq P(M_{i,j}|T_i, T_j), \tag{18}$$

where $g(E_k, A_0, t_0)$ is the joint density function of E_k, A_0, and t_0.

Estimating more precisely the probability $P(L_{i,j,k}|M_{i,j})$ that two taxa interact locally if they can potentially interact allows for improved predictions of local networks 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 $P(L_{i,j,k}|M_{i,j})$, as shown in Appendix 1, represents an initial step toward the overarching objective of reconstructing local networks from metawebs.

Quantifying and reducing interaction uncertainty

While sampling biological communities decreases 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, 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, local interaction probabilities represent both our knowledge uncertainty and their 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 networks describe probabilities of observing interactions rather than their realization, we must also consider observation uncertainty (sampling error) as an additional source of uncertainty. Quantifying and partitioning this uncertainty will enable us to make more accurate predictions about ecological interactions at various spatial and temporal scales, and to identify priority sampling locations to reduce this uncertainty. 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.

Relaxing the independence assumption

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). The occurrence of an interaction may depend on the realization of other interactions or the presence or abundance of other taxa in the network (Kéfi *et al.* 2012; Pilosof *et al.* 2017). Relaxing this assumption of independence is the next logical step in the stochastic representation of interactions.

A more accurate representation of the uncertainty and variability of ecological networks involves creating *probabilistic networks* ($P(L_k)$ and $P(M)$), rather than networks of *probabilistic interactions* ($P(L_{i,j,k})$ and $P(M_{i,j})$). 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 structural constrained (e.g., Cimini *et al.* 2019; Park & Newman 2004).

Regardless of the method used, generating probabilistic local networks could lead to more accurate predictions of local networks of binary interactions by bypassing the independence assumption. Probabilistic networks could serve as an alternative to null hypothesis significance testing when comparing the structure of a local network to some random expectations or, as done in Pellissier *et al.* (2018) and Box 2, 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 to calculate their structure (Poisot *et al.* 2016). One could instead 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.

Acknowledgment

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

Angilletta, M.J., Jr., Steury, T.D. & Sears, M.W. (2004). Temperature, Growth Rate, and Body Size in Ectotherms: Fitting Pieces of a Life-History Puzzle1. *Integrative and Comparative Biology*, 44, 498–509.

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

Bartomeus, I., Gravel, D., Tylianakis, J.M., Aizen, M.A., Dickie, I.A. & Bernard-Verdier, M. (2016). A common framework for identifying linkage rules across different types of interactions. *Functional Ecology*, 30, 1894–1903.

Bascompte, J., Jordano, P., Melián, C.J. & Olesen, J.M. (2003). The nested assembly of plant–animal mutualistic networks. *Proceedings of the National Academy of Sciences*, 100, 9383–9387.

Battiston, F., Cencetti, G., Iacopini, I., Latora, V., Lucas, M., Patania, A., *et al.* (2020). Networks beyond pairwise interactions: Structure and dynamics. *Physics Reports*, Networks beyond pairwise interactions: Structure and dynamics, 874, 1–92.

Beardsell, A., Gravel, D., Berteaux, D., Gauthier, G., Clermont, J., Careau, V., *et al.* (2021). Derivation of Predator Functional Responses Using a Mechanistic Approach in a Natural System. *Frontiers in Ecology and Evolution*, 9.

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.

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. & Fortin, M.-J. (2024). Applying a method before its proof of concept: A cautionary tale using inferred food webs. *Global Change Biology*, 30, e17360.

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.

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.

Carlson, C.J., Farrell, M.J., Grange, Z., Han, B.A., Mollentze, N., Phelan, A.L., *et al.* (2021). The future of zoonotic risk prediction. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 376, 20200358.

Caron, D., Brose, U., Lurgi, M., Blanchet, F.G., Gravel, D. & Pollock, L.J. (2024). Trait-matching models predict pairwise interactions across regions, not food web properties. *Global Ecology and Biogeography*, 33, e13807.

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.

Cherif, M., Brose, U., Hirt, M.R., Ryser, R., Silve, V., Albert, G., *et al.* (2024). The environment to the rescue: Can physics help predict predator-prey interactions? *BIOLOGICAL REVIEWS*.

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.

Clark, J.S., Bell, D.M., Hersh, M.H., Kwit, M.C., Moran, E., Salk, C., *et al.* (2011). Individual-scale variation, species-scale differences: Inference needed to understand diversity. *Ecology Letters*, 14, 1273–1287.

Dallas, T., Park, A.W. & Drake, J.M. (2017). Predicting cryptic links in host-parasite networks. *PLOS Computational Biology*, 13, e1005557.

Dansereau, G., Barros, C. & Poisot, T. (2024). Spatially explicit predictions of food web structure from regional-level data. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 379, 20230166.

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.

Eklöf, A. & Stouffer, D.B. (2016). The phylogenetic component of food web structure and intervality. *Theoretical Ecology*, 9, 107–115.

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.

Grabowski, J.H. (2004). Habitat Complexity Disrupts Predator–Prey Interactions but Not the Trophic Cascade on Oyster Reefs. *Ecology*, 85, 995–1004.

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

Gravel, D., Poisot, T., Albouy, C., Velez, L. & Mouillot, D. (2013). Inferring food web structure from 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., Montemaggiori, 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.

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 bymodelling species simultaneously with a Joint Species DistributionModel (JSDM). *Methods in Ecology and Evolution*, 5, 397–406.

Pomeranz, J.P.F., Thompson, R.M., Poisot, T. & Harding, J.S. (2019). Inferring 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.

Simmonds, E.G., Adjei, K.P., Cretois, B., Dickel, L., González-Gil, R., Laverick, J.H., *et al.* (2024). Recommendations for quantitative uncertainty consideration in ecology and evolution. *Trends in Ecology & Evolution*, 39, 328–337.

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.

Stock, M., Poisot, T., Waegeman, W. & De Baets, B. (2017). Linear filtering reveals false negatives in species interaction data. *SCIENTIFIC REPORTS*, 7, 45908.

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.

Stouffer, D.B., Sales-Pardo, M., Sirer, M.I. & Bascompte, J. (2012). Evolutionary Conservation of Species' Roles in Food Webs. *Science*, 335, 1489–1492.

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.

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.

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.

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.