

# Deciphering probabilistic species interaction networks

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

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

## Correspondance to:

Francis Banville — francis.banville@umontreal.ca

Representing species interactions probabilistically (how likely are they to occur?) as opposed to deterministically (are they occurring?) 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: *July 22, 2024*

## **1 Introduction**

### **2 Species interactions are variable and uncertain**

3 As we try to navigate global biodiversity change, filling in knowledge gaps about biodiversity becomes  
4 instrumental to monitoring and mitigating those changes (Abrego *et al.* 2021; Gonzalez & Londoño 2022;  
5 Hortal *et al.* 2015). However, cataloging species, populations and, in particular, ecological interactions (e.g.,  
6 predation, parasitism, and pollination) is a substantial challenge (Pascual *et al.* 2006; Polis 1991). There are  
7 methodological and biological constraints that hinder our ability to document species interactions, inevitably  
8 leading to uncertainty in our knowledge of interactions. For example, the spatial and temporal uncoupling of  
9 species (e.g., nocturnal and diurnal species coexisting in the same space with different daily activity timings,  
10 Jordano 1987) and the large number of rare and cryptic interactions in a community, contribute to these  
11 knowledge gaps by making it more difficult to observe interactions (Jordano 2016).

12 Several conditions must be satisfied for an interaction to be observed locally. First, both species must have  
13 overlapping geographic ranges, i.e. they must co-occur within the region of interest (Cazelles *et al.* 2016;  
14 Morales-Castilla *et al.* 2015). Second, they must have some probability of meeting (Poisot *et al.* 2015).  
15 Probabilities of interspecific encounters are typically low, especially for rare species with low abundances.  
16 While species' absolute abundances may impact interaction frequencies (Vázquez *et al.* 2007), encounter  
17 probabilities are determined by their relative abundances (Canard *et al.* 2012; Canard *et al.* 2014). The  
18 probability that species meet also depends on their biology, such as their phenology (Olesen *et al.* 2010; Singer  
19 & McBride 2012) and discoverability (Broom & Ruxton 2005). Finally, when species do come into contact, an  
20 interaction occurs only if their traits, such as their phenotypes (Bolnick *et al.* 2011; Gravel *et al.* 2013; Stouffer  
21 *et al.* 2011) and behavior (Choh *et al.* 2012; Pulliam 1974), are locally compatible in that specific environment  
22 (Poisot *et al.* 2015). Because these conditions are not consistently met locally, there will inevitably be instances  
23 where interactions will be observed and others where they will not.

24 Documenting the location and timing of interactions is difficult when accounting for the spatiotemporal  
25 variability of ecological interactions (Poisot *et al.* 2012, 2015). Knowing the biological capacity of two species  
26 to interact directly (via e.g., trophic interactions) is necessary but not sufficient for inferring their interaction at a  
27 specific time and space. Environmental factors, such as temperature (Angilletta *et al.* 2004), drought  
28 (Woodward *et al.* 2012), climate change (Araujo *et al.* 2011; Gilman *et al.* 2010; Woodward *et al.* 2010), and  
29 habitat modifications (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 variation 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. 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

59 taking action (Carlson *et al.* 2021), especially in places where access and resources for research are scarce. A  
60 probability is a measure of how likely a specific outcome is, based on both the uncertainty and variability of  
61 interactions. Interaction probabilities may be uncertain when there is a distribution of plausible probability  
62 values. The probabilistic representation of interactions has been applied to direct interactions, which are  
63 conceptually and mathematically analogous regardless of their biological type (e.g., predation and pollination).  
64 This is in contrast with indirect interactions (e.g., interspecific competition), which arise from distinct  
65 ecological processes and are often not directly observable (Kéfi *et al.* 2015, 2016). By accounting for the  
66 uncertainty and variability of direct interactions, networks of probabilistic interactions (which differ from  
67 *probabilistic networks* describing the uncertainty and variability of the *whole* network) may provide a more  
68 realistic portrait of species interactions.

69 Probabilistic interactions differ from binary interactions. Networks of probabilistic interactions, within a  
70 Bayesian perspective, express our degree of belief (or confidence) regarding the occurrence or observation of  
71 interactions. In contrast, interactions are simply regarded as either occurring or not in networks of deterministic  
72 binary interactions. Based on the scale at which they are estimated, interaction probabilities may reflect our  
73 level of confidence in whether interactions will be observed, realized locally, or biologically feasible. Our level  
74 of confidence should be more definitive (approaching either 0 or 1) as we extend our sampling to a broader area  
75 and over a longer duration, thereby diminishing the uncertainty of our knowledge of interactions (but not  
76 necessarily the estimation of their variability). In the broadest sense, binary interactions are also a type of  
77 probabilistic interaction, in which the numerical value of an interaction is restrained to 0 (non-occurring) or 1  
78 (occurring). In networks of probabilistic interactions, only forbidden interactions (i.e., interactions prohibited  
79 by biological traits or species absence, Jordano *et al.* 2003; Olesen *et al.* 2010) have a probability value of zero,  
80 provided that intraspecific trait variability is considered (Gonzalez-Varo & Traveset 2016). Understanding the  
81 nuances between probabilistic and binary interactions is essential for accurately modeling and interpreting  
82 ecological networks.

83 The application and development of computational methods in network ecology, often based on a probabilistic  
84 representation of interactions, can alleviate (and guide) the sampling efforts required to document species  
85 interactions (Strydom *et al.* 2021). For example, statistical models can be used to estimate the uncertainty of  
86 pairwise interactions (Cirtwill *et al.* 2019) and the probability of missing (false negatives) and spurious (false  
87 positives) interactions (Guimerà & Sales-Pardo 2009), helping us identify places where sampling is most  
88 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 distribution (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 these species have never co-occurred before, and as a result, their interaction has not been previously observed. What is the probability that the rodent is part of the diet of the avian predator, or put differently, what is the probability that they interact? Answering this question requires some clarification, as there are multiple ways to interpret and calculate interaction probabilities. We 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.

146 An important aspect to consider when estimating or using interaction probabilities is knowing if they describe  
147 the probability of potential or realized interactions, as these two types of interactions have distinct meanings and  
148 sources of uncertainty and variability. A potential (regional) interaction is defined as the biological or  
149 ecological capacity of two taxa to interact (i.e., the probability that they interact if they were to encounter each  
150 other, given sufficient time and appropriate environmental conditions) whereas a realized (local) interaction is  
151 the occurrence or observation of this interaction in a well-defined space and time (i.e., the probability that they  
152 interact locally). For two co-occurring taxa and over enough time, the probability of local interaction tends  
153 toward the probability of regional (potential) interaction. A longer duration increases the probability that  
154 species will eventually encounter each other and that local environmental conditions supporting an interaction  
155 will occur, provided that species have the biological capacity to interact. Recognizing the distinction between  
156 probabilistic regional and local interactions is crucial for accurately interpreting interaction probabilities in  
157 ecological networks.

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

## 167 What is the outcome of probabilistic interactions?

### 168 The outcome of probabilistic interactions is usually binary

169 Local networks and metawebs, like any type of network, are made of nodes and edges that may be represented  
170 at different levels of organization. The basic units of ecological networks are individuals that interact with each  
171 other (e.g., by predation in food webs, Elton 2001), forming individual-based networks (Melián *et al.* 2011).  
172 The aggregation of these individuals into more or less homogeneous groups (e.g., populations, species, families,  
173 feeding guilds) allows us to represent nodes at broader taxonomic scales, which affects our interpretation of the

174 properties of these systems (Guimarães 2020; Hemprich-Bennett *et al.* 2021).

175 Ecologists have traditionally represented interactions (edges) as binary objects that were considered realized  
176 after observing at least one individual from group  $i$  interact with at least another individual from group  $j$ . In an  
177 adjacency matrix  $B$  of binary interactions, the presence or absence of an interaction  $B_{i,j}$  between two taxa can  
178 be viewed as the result of a Bernoulli trial  $B_{i,j} \sim \text{Bernoulli}(P(B_{i,j}))$ , with  $P(B_{i,j})$  being the probability of  
179 interaction. This interaction probability characterizes our limited ecological knowledge and/or the intrinsic  
180 spatiotemporal variability of the interaction. It may be estimated through predictive models (e.g., those based  
181 on biological traits and species abundances) or expert (prior) knowledge about the interaction. In networks of  
182 probabilistic interactions, the edge values  $P(B_{i,j})$  are probabilistic events whose only two possible outcomes are  
183 the presence ( $B_{i,j} = 1$ ) or absence ( $B_{i,j} = 0$ ) of an interaction between each pair of nodes. Depending on the  
184 type of probabilistic interaction network (local network or metaweb), the mathematical formulation and  
185 interpretation of stochastic parameters like  $P(B_{i,j})$  can be linked to environmental and biological factors such as  
186 species abundances, species traits, area, and time, for example using logistic regression with continuous  
187 explanatory variables. This allows us to model the probability that at least two individuals interact under these  
188 conditions.

189 The variability of an interaction determines the fraction of networks in which it occurs. This fraction can be  
190 predicted by using a Binomial distribution, assuming a constant interaction probability and independence  
191 between interactions in different networks (trials). When considering uncertainties around the estimation of  
192  $P(B_{i,j})$ , a Beta distribution may be used to represent the relative likelihood of different probability values. For  
193 example, when calculating the probability of interaction between two taxa based on their local abundances, any  
194 uncertainty in their abundances would introduce uncertainty in the interaction probability at the local scale. If  
195 we take into account the uncertainty of the interaction probability, a Beta-Binomial distribution can be used to  
196 predict the number of networks in which the interaction occurs. Empirically observing an interaction between  
197 two taxa at a given location and time provides important information that can be used to update previous  
198 estimates of  $P(B_{i,j})$ , informing us of the conditions that enabled them to interact locally. By sampling binary  
199 interactions in different contexts, we can thus estimate their local variability more precisely.

200 **The outcome of probabilistic interactions may also be quantitative**

201 Even though binary interaction networks constitute a highly valuable source of ecological information (Pascual  
202 *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 between pairs of taxa (e.g., the flower-visiting rates of pollinators in a mutualistic network, Herrera 1989). When interaction strengths characterize predation pressure on prey, 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).

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, representing interaction strengths probabilistically enables the propagation of uncertainty from pairwise interactions to the broader dynamics of ecological networks.

Because of the methodological difficulties typically encountered when building deterministic quantitative networks, binary interaction networks, which are 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.

232 **Local networks: communities interacting in space and time**

233 **What are local probabilistic interactions?**

234 Local networks of probabilistic interactions describe how likely taxa are to interact in a local context. Local  
235 interactions are contingent upon the environmental conditions experienced by the community and the matching  
236 of taxa's local biological traits. In local networks, edges commonly represent our degree of belief that two taxa  
237 interact in nature, but can also represent the probability of empirically *observing* this interaction (Catchen *et al.*  
238 2023). Local interactions may thus arise from both the ecological (realized interactions) and sampling  
239 (observed interactions) processes taking place locally.

240 Local networks are delineated within a particular location and time. We define space as the collection of  
241 geographic coordinates  $(x, y, z)$ , with  $(x, y)$  representing longitude and latitude coordinates, and  $z$  denoting  
242 either altitudes or depths. These point coordinates delineate the spatial boundaries of the system, which may be  
243 portrayed as a polyhedron. Ecological interactions may vary along latitudinal and altitudinal gradients, as  
244 evidenced by changes in hummingbird-plant interactions (Weinstein & Graham 2017a, b) and mosquito biting  
245 rates (e.g., Kulkarni *et al.* 2006) at different elevations. On the other hand, time is defined as the specific time  
246 period within which interactions were either observed or predicted. Even though space and time are continuous  
247 variables that should yield probability *densities* of interactions (i.e., relative likelihoods of interactions  
248 occurring at infinitesimal locations and instants in time), these definitions enable them to be conceptualized as  
249 distinct patches and time segments. Treating space and time as discrete dimensions aligns with the common  
250 sampling methods of ecological networks and provides probabilities of interactions, which can be obtained by  
251 integrating probability densities over space and time. We can quantify both an area  $A_0$  and a duration  $t_0$  with  
252 these definitions. By studying probabilistic local interaction networks, we may thus conduct spatiotemporal  
253 analyses of local interactions (Box 1), enhancing our understanding of interactions occurring in distinct  
254 environmental contexts.

255 **What are local probabilistic interactions conditioned on?**

256 **Local interactions may be conditioned on co-occurrence**

257 The probability that two taxa  $i$  and  $j$  interact in a local network  $L_{x,y,z,t}$  (spatial and temporal subscripts hereafter  
258 replaced by the shorter subscript  $k$  for clarity) can be conditioned on many environmental and biological factors.

259 In addition to network area (or volume)  $A_0$  and duration  $t_0$ , they may be conditioned on taxa co-occurrence  
260  $X_{i,j,k}$ , which is usually Boolean, describing if the geographic distributions of both taxa overlap within the study  
261 area. As illustrated in Box 1, co-occurrence may be modeled probabilistically, in which case it may conform to  
262 a Bernoulli distribution  $X_{i,j,k} \sim \text{Bernoulli}(P(X_{i,j,k}))$ . The probability of co-occurrence can be calculated using  
263 the individual (marginal) occurrence probabilities  $P(X_{i,k})$  and  $P(X_{j,k})$ . Given that taxa occurrences are not  
264 independent of each other, the probability of co-occurrence can be calculated by multiplying the probability of  
265 occurrence of one taxon by the probability of occurrence of the other given that the first one is present:

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

266 The probability of co-occurrence  $P(X_{i,j,k})$  can be estimated through the application of joint species distribution  
267 models (e.g., Pollock *et al.* 2014), potentially taking into account biotic interactions (Staniczenko *et al.* 2017).  
268 Given that the probability that two non-co-occurring taxa interact locally is zero (i.e.,  $P(L_{i,j,k}|X_{i,j,k} = 0) = 0$ ),  
269 the probability of local interaction can be obtained by multiplying the probability of interaction given  
270 co-occurrence with the probability of co-occurrence:

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

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

### 273 Local interactions may be conditioned on different environmental and biological factors

274 Local interactions may also be conditioned on local environmental factors such as temperature (Angilletta *et al.*  
275 2004), precipitation (Woodward *et al.* 2012), habitat structure (Klecka & Boukal 2014), and presence of other  
276 taxa in the network (Kéfi *et al.* 2012; Pilosof *et al.* 2017). We use the variable  $E_k$  to describe the local  
277 environmental context in which interaction probabilities were estimated. For example, in a mesocosm  
278 experiment estimating interaction probabilities between predators and prey with and without shelters,  $E_k$  would  
279 represent the presence or absence of these shelters. Like co-occurrence,  $E_k$  can also be modeled  
280 probabilistically when the variability or uncertainty of environmental factors is considered.  $E_k$  represents all  
281 environmental variables that were taken into consideration when measuring interaction probabilities; it is

282 therefore a subset of all environmental factors acting on ecological interactions.

283 Other important factors that can impact interaction probabilities at the local scale are taxa local abundances  $N_{i,k}$   
284 and  $N_{j,k}$ , which affect encounter probabilities (Canard *et al.* 2012), and local traits  $T_{i,k}$  and  $T_{j,k}$  (e.g., body  
285 mass, longevity, and habitat breadth, Caron *et al.* 2024), which determine the ability of individuals to interact  
286 after encountering each other (Poisot *et al.* 2015). Local interaction probabilities may also be conditioned on  
287 higher-level properties of the community, which we denote by  $f(L_k)$ . Many topological null models (i.e.,  
288 statistical models that randomize interactions by retaining certain properties of the network while excluding  
289 others) provide interaction probabilities from selected measures of network structure, such as connectance  
290 (Fortuna & Bascompte 2006) and the degree distribution (Bascompte *et al.* 2003). Biological factors, whether  
291 at the scale of individual taxa pairs or the community, may thus impact how we estimate and define interaction  
292 probabilities.

### 293 **Local interactions may be conditioned on biological feasibility**

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

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

298 Low values of  $P(L_{i,j,k}|M_{i,j})$  indicate that feasible interactions rarely occur locally, intermediate values around  
299 50% suggest considerable spatiotemporal variability, while high values indicate that regional interactions are  
300 nearly always realized locally. The local probability of interaction between a given pair of taxa is thus always  
301 equal to or below their probability of regional interaction. Taking into account biological feasibility in our  
302 estimation of local interaction probabilities leverages information from the metaweb to better predict the local  
303 occurrence of interactions (Dansereau *et al.* 2023; Strydom *et al.* 2021).

304    **Conditional variables must be explicitly stated**

305    The probability that two taxa  $i$  and  $j$  interact in a local network  $L_k$  can thus be conditioned on their  
306    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  
307    traits  $T_{i,k}$  and  $T_{j,k}$ , local environmental conditions  $E_k$ , network area (or volume)  $A_0$ , time interval  $t_0$ , network  
308    properties  $f(L_k)$ , and biological feasibility  $M_{i,j}$ . The local probability of interaction is described by the  
309    following expression when all of these conditional variables are included:

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

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

323    When accounted for, conditional variables should be clearly described in the documentation of the data  
324    (Brimacombe *et al.* 2023), preferentially in mathematical terms to avoid any confusion in their interpretation  
325    and to limit manipulation errors during their re-use. For instance, ecologists should be explicit about their  
326    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  
327    probabilities, as this can change our interpretation of the data and understanding of potential uncertainty  
328    sources. Reporting the scale and level of aggregation of the data enables us to more accurately study the  
329    underlying ecological processes (Clark *et al.* 2011) and manipulate or propagate uncertainty to different  
330    aggregation levels (Simmonds *et al.* 2024). In Tbl. 1, we present examples of studies that used different

331 expressions of probabilistic interactions with different conditional variables. We included in this table the  
 332 probability of empirically observing an interaction that is realized locally  $P(O_{i,j,k}|L_{i,j,k})$  to underscore the  
 333 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
$P(L_{i,j,k} X_{i,k}, X_{j,k}, \dots)$	local	realization of the interaction given taxa co-occurrence	species distribution models	Gravel <i>et al.</i> (2019), Dansereau <i>et al.</i> (2023), Boxes 1 and 5
$P(L_{i,j,k} N_{i,k}, N_{j,k}, \dots)$	local	realization of the interaction given taxa abundances	neutral models	Canard <i>et al.</i> (2014)
$P(L_{i,j,k} T_{i,k}, T_{j,k}, \dots)$	local	realization of the interaction given local traits	trait matching models	Caron <i>et al.</i> (2024), Box 4
$P(L_{i,j,k} E_k, \dots)$	local	realization of the interaction given local environmental conditions	environmental- based models	Gravel <i>et al.</i> (2019) (temperature and precipitation)
$P(L_{i,j,k} A_0, \dots)$	local	realization of the interaction in a given area or volume	spatial models	Galiana <i>et al.</i> (2018) *, Box 3
$P(L_{i,j,k} t_0, \dots)$	local	realization of the interaction during a given time period	temporal models	Weinstein & Graham (2017a), Boxes 1 and 3
$P(L_{i,j,k} f(L_k), \dots)$	local	realization of the interaction given network structure	topological models	Fortuna & Bascompte (2006) (connectance), Stock <i>et al.</i> (2017)

Expression	Type	Outcome	Common models	Reference
$P(L_{i,j,k} M_{i,j}, \dots)$	local	realization of the interaction given that it is biologically feasible	spatiotemporal models	Dansereau <i>et al.</i> (2023), Boxes 2, 3, and 5
$P(O_{i,j,k} L_{i,j,k}, \dots)$	local	observation of the interaction given that it is realized locally	sampling models	Catchen <i>et al.</i> (2023)
$P(M_{i,j} T_i, T_j)$	regional	biological feasibility of the interaction given regional traits (non-forbiddenness)	trait matching models	Strydom <i>et al.</i> (2022), 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

### 334 How are local probabilistic interactions estimated?

335 Various statistical models can be used to estimate local interaction probabilities, some of which are presented in  
 336 Tbl. 1. These models can be based on multiple conditional variables. Although these variables correspond to  
 337 distinct ecological inquiries or mechanisms related to ecological interactions, they may covary with each other,  
 338 such as the possible dependence of  $X_{i,j,k}$  and  $E_k$  on spatial and temporal scales. When estimating interaction  
 339 probabilities using e.g. a generalized linear model with multiple explanatory variables that might not all be  
 340 independent, it may become important to address collinearity. In such cases, it may be necessary to use variable  
 341 selection techniques before fitting the model to data to mitigate this issue. Other challenges and opportunities  
 342 associated with predictive models of species interactions are reviewed in Strydom *et al.* (2021).

343 When using multiple competing models to estimate local interaction probabilities, rather than selecting a single  
 344 model that best fits the data, model averaging may enhance our estimations. Model weights represent the  
 345 probability that each model is the most suitable for explaining the data, and may be measured using Akaike  
 346 weights (Burnham & Anderson 2004; Wagenmakers & Farrell 2004). For instance, given two competing  
 347 models  $mod_1$  and  $mod_2$  with respective probabilities (or weights)  $P(mod_1)$  and  $P(mod_2)$ , the average  
 348 probability of interaction  $P(L_{i,j,k})$  can be calculated as follows:

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

349 Model averaging takes into account the uncertainty of model structure in our estimation of local interaction  
 350 probabilities. Regardless of the model used for prediction, it is crucial to quantify and disclose all sources of  
 351 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  $\lambda_k$ . This parameter represents the local expected frequency of interaction between co-occurring taxa. The probability that two co-occurring taxa interact during a time interval  $t_0$  can be given by:

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

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

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

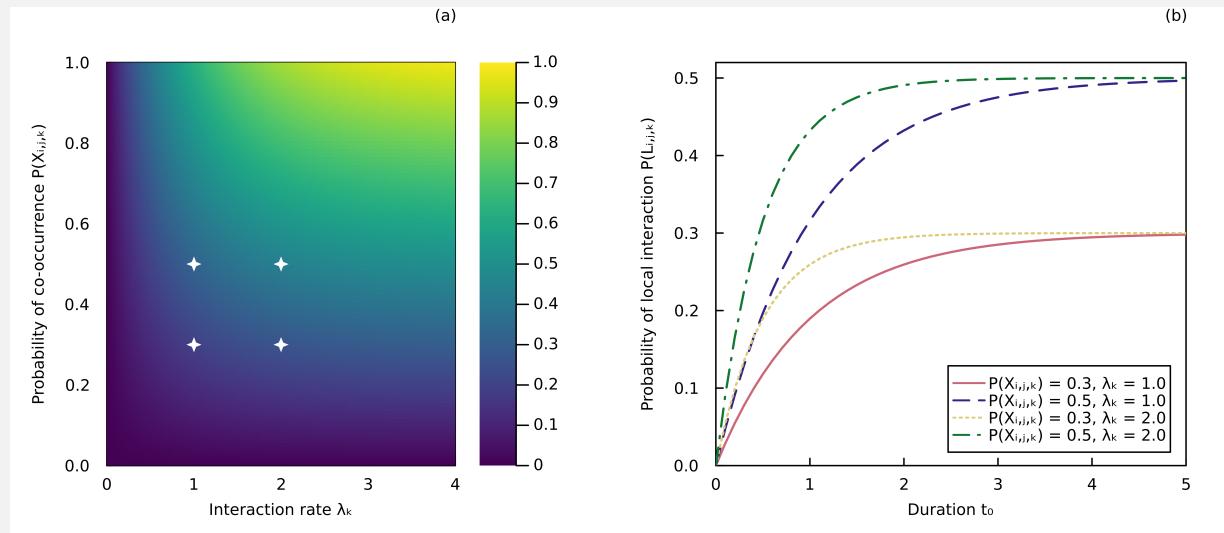
$$L_{i,j,k} \sim \text{Bernoulli}(P(L_{i,j,k})) \quad (7)$$

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

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

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

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  $P(X_{i,j,k})$ , for all positive values of the interaction rate.



**Figure 1: Parameters of the spatiotemporally explicit model of interactions.** (a) Probability of local interaction given by the process model (Eq. 8) under different values of  $\lambda_k$  (interaction rate) and  $P(X_{i,j,k})$  (probability of co-occurrence), with  $t_0 = 1$  (duration). Parameters  $t_0$  and  $\lambda_k$  have complementary units (e.g.,  $t_0$  in months and  $\lambda_k$  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  $\lambda_k$  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.

353

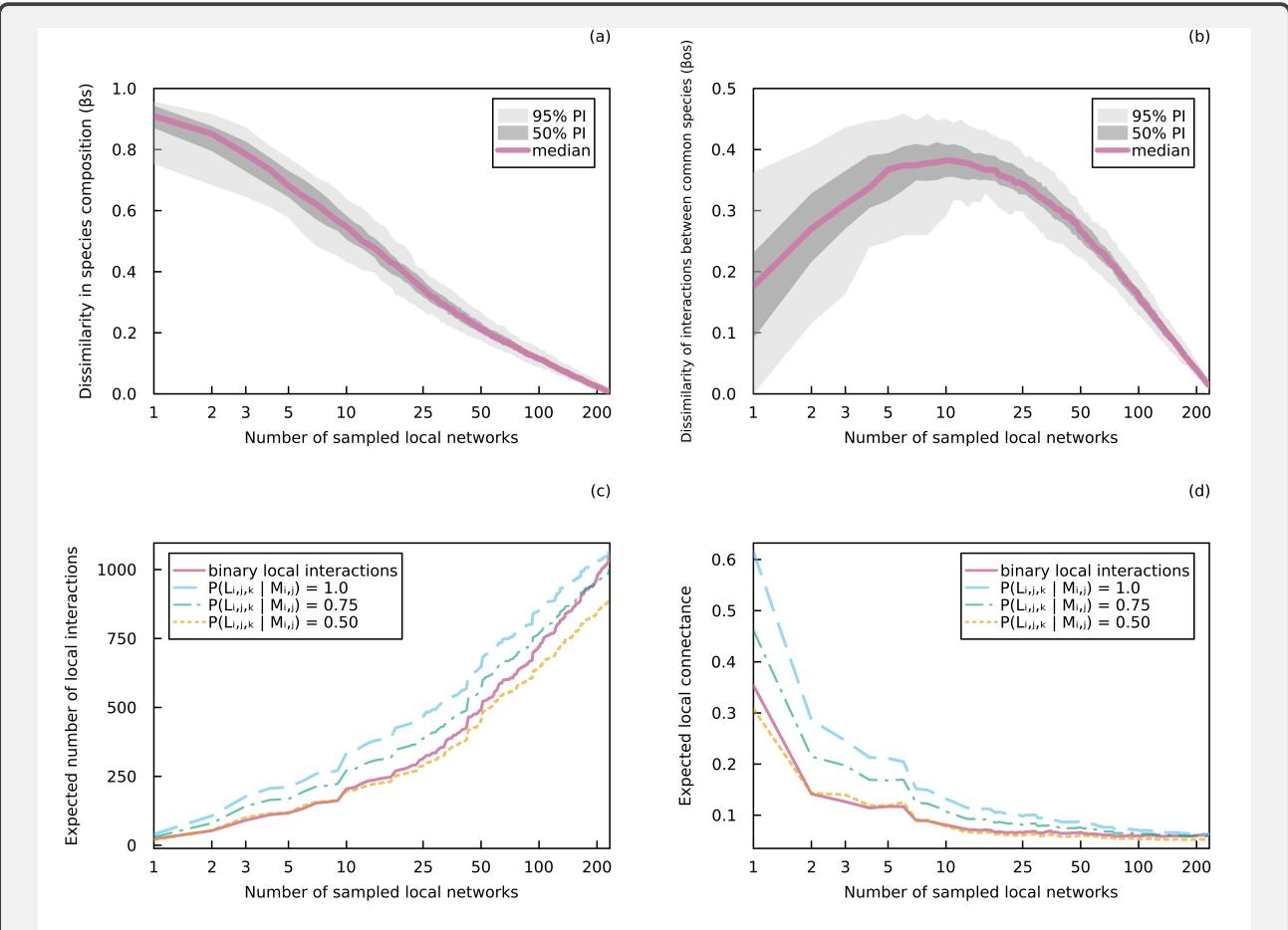
## Box 2: Dissimilarity of local host-parasite networks

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

354

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 at the Open Science Framework (TBD).

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 ( $\beta_S$ , Fig. 2a) and the dissimilarity of interactions between common species ( $\beta_{OS}$ , 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  $P(L_{i,j,k}|M_{i,j})$  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.

357

## 358 Metawebs: regional catalogs of interactions

### 359 What are regional probabilistic interactions?

360 Metawebs (Dunne 2006) are networks of potential interactions over broad spatial, temporal, and taxonomic  
361 scales (e.g., food webs at the continental scale). Potential interactions describe the biological capacity of taxa to  
362 interact under optimal or feasible environmental conditions, which is typically assessed at the regional scale.

363 Metawebs of probabilistic interactions are particularly useful in situations where there is uncertainty in the  
364 ability of taxa to interact (Strydom *et al.* 2023). They may also be used as informative priors of local  
365 interactions. Therefore, building a metaweb of probabilistic interactions may be an important first step before  
366 predicting networks at finer scales.

367 In contrast to local networks, where interaction probabilities arise from the variability of interactions and the  
368 lack of information on the conditions, interaction probabilities in metawebs solely result from a lack of  
369 knowledge. This uncertainty arises due to insufficient interaction data, especially for taxa that have not yet been  
370 observed to co-occur, and uncertainties in trait-matching models. As data accumulates, interactions in  
371 metawebs should tend towards binarity, either taking a value of 1 (observing an interaction at least once) or  
372 approaching 0 (repeatedly failing to observe an interaction between co-occurring taxa). Confidently observing  
373 an interaction once confirms its biological feasibility, but failing to observe it (even on multiple occasions) does  
374 not ensure that it is non-feasible (e.g., due to false negatives, Catchen *et al.* 2023). While local interaction  
375 probabilities are irreducible because of local variability, the uncertainty of regional interactions reduces to 0  
376 with the addition of information. Moreover, although *neutrally* forbidden interactions (i.e., forbidden  
377 interactions between rare species, Canard *et al.* 2012) have low probability values in local networks, they would  
378 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 regional interactions (McLeod *et al.* 2021). However, in contrast with local networks of probabilistic interactions, regional interactions are not evaluated for any particular local context, which impacts how they scale with space and time. 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. 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:

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

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.

404 However, as explained in Box 4, there is no fundamental difference in the taxonomic scaling of regional and  
405 local interactions (i.e., how interaction probabilities change with taxonomic level) because they both depend on  
406 trait aggregation.

407 The biological feasibility of interactions expresses our degree of belief that there exists at least one combination  
408 of phenotypes that could support an interaction if they were to encounter each other, assuming they had enough  
409 time to interact. Evaluating this probability is conducted without incorporating the environmental conditions  
410 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  
411 forbidden interactions (i.e., the probability that their traits do not support an interaction), which is based  
412 uniquely on biological traits:

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

413 For example, let  $i$  be a western diamondback rattlesnake (*Crotalus atrox*) and  $j$ , a wood lemming (*Myopus*  
414 *schisticolor*). These two taxa never co-occur, the rattlesnake being adapted to warm regions of North America  
415 (Castoe *et al.* 2007) and the lemming, to northern habitats of Eurasia (Fedorov *et al.* 2008). As we lack direct  
416 observations of an interaction between these two species, we have to rely on expert knowledge or trait-matching  
417 models to estimate their probability of potential interaction. To accurately estimate this probability using  
418 trait-matching models, it is crucial to ensure that the set of traits considered reflects the overall traits  
419 distributions of both taxa. We could for instance consider their average body mass and the average phylogenetic  
420 distance of lemmings to rattlesnakes' prey. Doing so, we might find a high probability of potential interaction  
421 based on these traits. This example illustrates how regional interactions describing biological feasibility may be  
422 estimated solely based on traits, without taking into account environmental conditions.

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

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

$$P(M_{i,j}^*|T_i, T_j, E), \quad (13)$$

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

$$\int_E P(M_{i,j}^*|T_i, T_j, E) dE \leq P(M_{i,j}|T_i, T_j). \quad (14)$$

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

#### 443 **How are regional probabilistic interactions estimated?**

444 Starting from a selected set of taxa, which are usually distributed within a broad region of interest, metawebs  
 445 can be built using different data sources, including literature review (e.g., Maiorano *et al.* 2020), aggregated  
 446 interaction data (e.g., Gravel *et al.* 2019; Saravia *et al.* 2022), trait-matching models (e.g., Shaw *et al.* 2024;  
 447 Strydom *et al.* 2022), and expert knowledge. Every pair of taxa that have confidently been observed to interact  
 448 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  
 449 from local networks of probabilistic interactions, where interaction events may remain stochastic (i.e.,  
 450  $P(L_{i,j,k}) < 1$ ) even after empirically observing interactions due to their spatiotemporal variability. Interactions  
 451 that were never observed typically have low probability values in local networks and vary from low to high  
 452 values in metawebs, contingent upon taxa traits distributions (reaching 0 for forbidden links). The aggregation  
 453 of model predictions and data from different sources thus tends to raise the number of potential interactions in

454 metawebs.

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

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

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

462 Observations of interactions may be false positives because of observation errors due to taxonomic  
463 misidentifications and ecological misinterpretations, such as those involving phylogenetically close species or  
464 cryptic species and interactions (Pringle & Hutchinson 2020). Likewise, forbidden interactions may be false  
465 negatives, e.g. if they have been evaluated based on unrepresentative or incomplete traits distributions.

466 Employing Bayesian models proves valuable when estimating interaction probabilities in metawebs (e.g.,  
467 Bartomeus *et al.* 2016; Cirtwill *et al.* 2019). This improvement is achieved by updating prior information  
468 regarding the feasibility of interactions (e.g., experts' prior assessments of interaction probabilities) with  
469 empirical data on interactions and traits. By improving our estimation of potential interaction probabilities, we  
470 may build more reliable metawebs that adequately reflect our uncertainty on the biological feasibility of  
471 interactions.

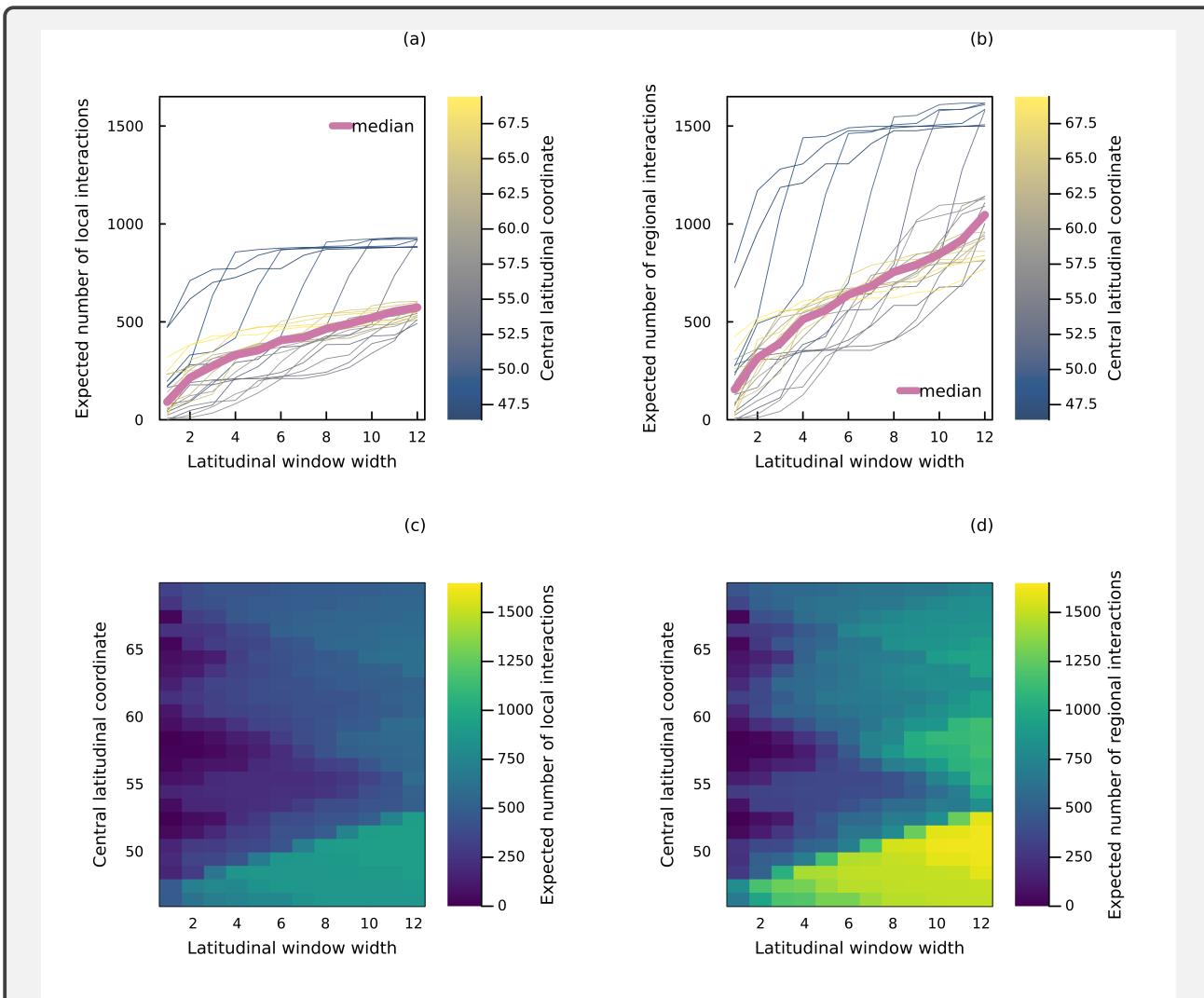
### Box 3: Spatial and temporal scaling of interactions

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,

472

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.

474

#### Box 4: Taxonomic scaling of interactions

Given that our interpretation of the properties of ecological networks depends on their taxonomic level (Guimarães 2020), investigating the taxonomic scaling of interactions (i.e., how interaction probabilities

475

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:

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

where  $g_{1,i}$  and  $g_{2,j}$  are the species of the corresponding genus and assuming independence between species-level interactions. 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

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:

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

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.

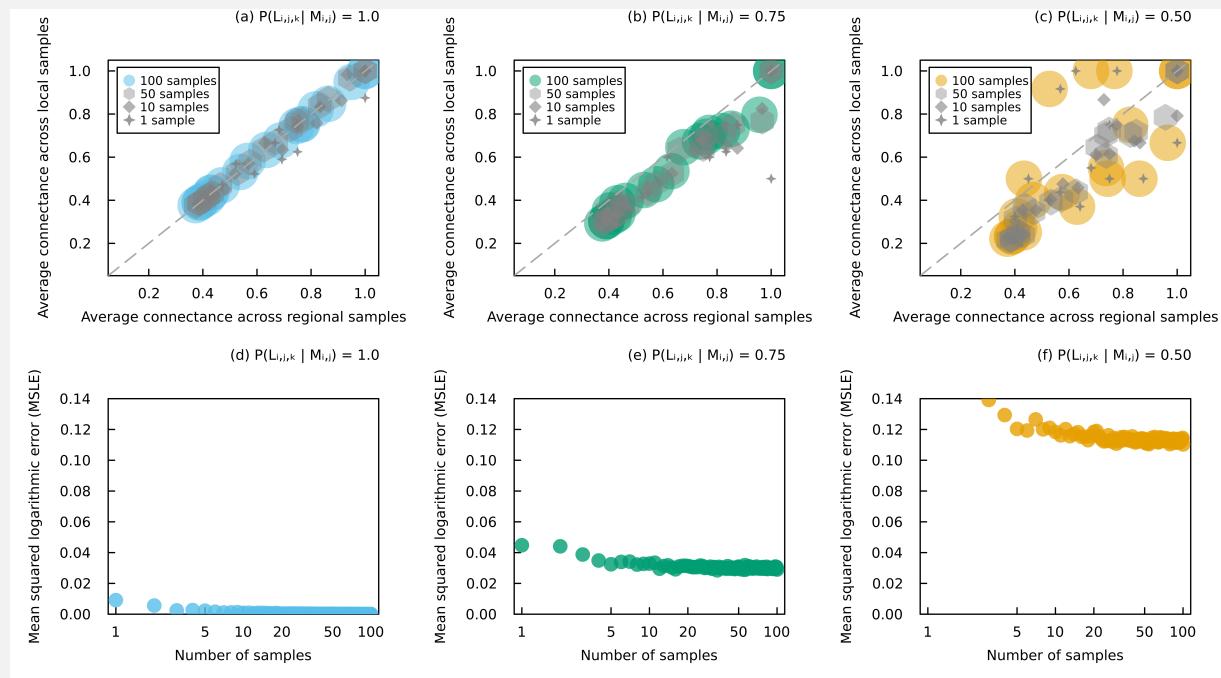
The second technique is to independently sample each local network of probabilistic interactions:

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

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, an interaction between two taxa could be more probable when another interaction occurs. The consequences of this assumption of independence on the prediction of network structure have yet to be empirically examined.

479

## 480 Future perspectives

481 In this contribution, we underline the importance of network documentation for adequately interpreting and  
482 manipulating probabilistic interaction data. The mathematical representation of probabilities and their statistical  
483 properties depend on the type of interactions (local or regional) and the conditions under which these  
484 interactions were evaluated. We show that local networks and metawebs of probabilistic interactions differ in  
485 their relationship to spatial and temporal scales (Box 3), with regional interactions remaining consistent across  
486 scales. In contrast with metawebs, local interactions are measured in a specific context (e.g., in a given area,  
487 time, and biological and environmental conditions) and depend on taxa co-occurrence. These differences bring  
488 to light the need to use probabilistic data with caution, for instance when generating network realizations of  
489 binary interactions across space (Box 5). Clear documentation describing the type of interaction and the  
490 variables used in their estimation are required to ensure adequate data manipulation. Sound data practices and  
491 foundations for probabilistic thinking in network ecology facilitate reliable assessments of the spatiotemporal  
492 variability and uncertainty of biotic interactions. Here we identify key research priorities for improving our  
493 understanding of probabilistic local and regional interactions.

### 494 Predicting local networks from metawebs

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

502 of local networks is constrained by the metaweb (Saravia *et al.* 2022).  
 503 Inferring local networks of probabilistic interactions from a metaweb comes with its own set of challenges. For  
 504 example, Dansereau *et al.* (2023) inferred spatially-explicit food webs from a metaweb of probabilistic trophic  
 505 interactions between Canadian mammals. Their predicted localized food webs are downscaled versions of the  
 506 metaweb (i.e., localized metawebs with the same interaction probabilities as those in the regional metaweb). To  
 507 infer local networks as defined in this manuscript (i.e., describing local realizations of interactions), local  
 508 interaction probabilities must be smaller than regional interaction probabilities. Inferring local networks from a  
 509 metaweb by maintaining identical interaction probability values introduces systematic biases into the  
 510 predictions, as discussed in Box 2 (unless networks are seen as downscaled metawebs).

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

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

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

$$\int_{E_k} \int_{A_0} \int_{t_0} P(L_{i,j,k}|E_k, A_0, t_0) dt_0 dA_0 dE_k \leq P(M_{i,j}|T_i, T_j). \quad (18)$$

519 Estimating more precisely the probability  $P(L_{i,j,k}|M_{i,j})$  that two taxa interact locally if they can potentially  
 520 interact allows for improved predictions of local networks from the metaweb of probabilistic interactions. This  
 521 task is challenging due to the variability of this probability across space and time, as well as its variability  
 522 across pairwise interactions within a network. Using simple models of  $P(L_{i,j,k}|M_{i,j})$ , as shown in Appendix 1,  
 523 represents an initial step toward the overarching objective of reconstructing local networks from metawebs.

524 **Quantifying and reducing interaction uncertainty**

525 While sampling biological communities decreases the uncertainty of interactions by accumulating evidence for  
526 their feasibility and local realization, there is a limit to how much we can reduce uncertainty. In metawebs,  
527 probabilities reflect our limited knowledge of interactions, which is expected to improve with a larger volume of  
528 data. Regional interactions should become more definitive (with probabilities approaching 0 or 1) as we  
529 investigate various conditions, including different combinations of species traits.

530 In comparison, local interaction probabilities represent both our knowledge uncertainty and their spatiotemporal  
531 variability. Owing to environmental heterogeneity, there will invariably be instances in which an interaction  
532 occurs and others in which it does not, across different times and locations, irrespective of the extent to which  
533 we can improve our knowledge of its biological feasibility and the local conditions that facilitate its occurrence.

534 When local networks describe probabilities of observing interactions rather than their realization, we must also  
535 consider observation uncertainty (sampling error) as an additional source of uncertainty. Quantifying and  
536 partitioning this uncertainty will enable us to make more accurate predictions about ecological interactions at  
537 various spatial and temporal scales, and to identify priority sampling locations to reduce this uncertainty. This  
538 will prove to be of vital importance as our time to understand nature runs out, especially at locations where the  
539 impacts of climate change and habitat loss hit harder.

540 **Relaxing the independence assumption**

541 Estimating local interaction probabilities independently for each taxa pair and assembling them into a network  
542 of probabilistic interactions comes with limitations. Predicting local networks of binary interactions based on  
543 these interaction probabilities assumes independence among interactions, a condition seldom respected in  
544 practice (Golubski & Abrams 2011). Relaxing this assumption is the next logical step in the stochastic  
545 representation of interactions.

546 A more accurate representation of the uncertainty and variability of ecological networks involves creating  
547 *probabilistic networks* ( $P(L_k)$  and  $P(M)$ ), rather than networks of *probabilistic interactions* ( $P(L_{i,j,k})$  and  
548  $P(M_{i,j})$ ). Probabilistic networks describe the probability that a particular network of binary (or quantitative)  
549 interactions (its whole adjacency matrix) is realized. For example, Young *et al.* (2021) used a Bayesian  
550 approach to estimate the probability of different plant-pollinator network structures derived from imperfect  
551 observational data. A probability distribution of ecological networks may also be derived using the principle of

552 maximum entropy given structural constrained (e.g., Cimini *et al.* 2019; Park & Newman 2004).  
553 Regardless of the method used, generating probabilistic local networks could lead to more accurate predictions  
554 of local networks of binary interactions by bypassing the independence assumption. Probabilistic networks  
555 could serve as an alternative to null hypothesis significance testing when comparing the structure of a local  
556 network to some random expectations or, as done in Pellissier *et al.* (2018) and Box 2, to the metaweb. These  
557 random expectations are typically derived by performing a series of Bernoulli trials on probabilistic  
558 interactions, assuming independence, to generate a distribution of networks of binary interactions to calculate  
559 their structure (Poisot *et al.* 2016). One could instead compare the likelihood of an observed network to the one  
560 of the most likely network structure (according to the probabilistic network distribution), thereby directly  
561 obtaining a measure of discrepancy of the empirical network. Generating probabilistic ecological networks  
562 represents a tangible challenge, one that, in the coming years, promises to unlock doors to more advanced and  
563 adequate analyses of ecological networks.

## 564 Acknowledgment

565 We acknowledge that this study was conducted on land within the traditional unceded territory of the Saint  
566 Lawrence Iroquoian, Anishinabewaki, Mohawk, Huron-Wendat, and Omàmiwininiwak nations. This work was  
567 supported by the Institute for Data Valorisation (IVADO) and the Natural Sciences and Engineering Research  
568 Council of Canada (NSERC) Collaborative Research and Training Experience (CREATE) program, through the  
569 Computational Biodiversity Science and Services (BIOS<sup>2</sup>) program. A special thanks to all members of the  
570 Black Holes and Revelations working group (organized by BIOS<sup>2</sup>) and the Poisot Lab for their insightful  
571 discussions and valuable feedback on this manuscript.

## 572 References

- 573 Abrego, N., Roslin, T., Huotari, T., Ji, Y., Schmidt, N.M., Wang, J., *et al.* (2021). Accounting for species  
574 interactions is necessary for predicting how arctic arthropod communities respond to climate change.  
575 *Ecography*, 44, 885–896.
- 576 Angilletta, M.J., Jr., Steury, T.D. & Sears, M.W. (2004). Temperature, Growth Rate, and Body Size in  
577 Ectotherms: Fitting Pieces of a Life-History Puzzle<sup>1</sup>. *Integrative and Comparative Biology*, 44, 498–509.

- 578 Araujo, M.B., Rozenfeld, A., Rahbek, C. & Marquet, P.A. (2011). [Using species co-occurrence networks to](#)  
579 [assess the impacts of climate change](#). *Ecography*, 34, 897–908.
- 580 Bartomeus, I., Gravel, D., Tylianakis, J.M., Aizen, M.A., Dickie, I.A. & Bernard-Verdier, M. (2016). [A](#)  
581 [common framework for identifying linkage rules across different types of interactions](#). *Functional Ecology*,  
582 30, 1894–1903.
- 583 Bascompte, J., Jordano, P., Melián, C.J. & Olesen, J.M. (2003). [The nested assembly of plant–animal](#)  
584 [mutualistic networks](#). *Proceedings of the National Academy of Sciences*, 100, 9383–9387.
- 585 Berg, M.P. & Ellers, J. (2010). [Trait plasticity in species interactions: A driving force of community dynamics](#).  
586 *Evolutionary Ecology*, 24, 617–629.
- 587 Berlow, E.L., Neutel, A.-M., Cohen, J.E., De Ruiter, P.C., Ebenman, B., Emmerson, M., *et al.* (2004).  
588 [Interaction strengths in food webs: Issues and opportunities](#). *Journal of Animal Ecology*, 73, 585–598.
- 589 Bolnick, D.I., Amarasekare, P., Araújo, M.S., Bürger, R., Levine, J.M., Novak, M., *et al.* (2011). [Why](#)  
590 [intraspecific trait variation matters in community ecology](#). *Trends in Ecology & Evolution*, 26, 183–192.
- 591 Borrett, S.R. & Scharler, U.M. (2019). [Walk partitions of flow in Ecological Network Analysis: Review and](#)  
592 [synthesis of methods and indicators](#). *Ecological Indicators*, 106, 105451.
- 593 Boulangeat, I., Gravel, D. & Thuiller, W. (2012). [Accounting for dispersal and biotic interactions to disentangle](#)  
594 [the drivers of species distributions and their abundances](#). *Ecology Letters*, 15, 584–593.
- 595 Brimacombe, C., Bodner, K. & Fortin, M.-J. (2024). [Applying a method before its proof of concept: A](#)  
596 [cautionary tale using inferred food webs](#). *Global Change Biology*, 30, e17360.
- 597 Brimacombe, C., Bodner, K., Michalska-Smith, M., Poisot, T. & Fortin, M.-J. (2023). [Shortcomings of reusing](#)  
598 [species interaction networks created by different sets of researchers](#). *PLOS Biology*, 21, e3002068.
- 599 Broom, M. & Ruxton, G.D. (2005). [You can run—or you can hide: Optimal strategies for cryptic prey against](#)  
600 [pursuit predators](#). *Behavioral Ecology*, 16, 534–540.
- 601 Burnham, K.P. & Anderson, D.R. (2004). [Multimodel Inference: Understanding AIC and BIC in Model](#)  
602 [Selection](#). *Sociological Methods & Research*, 33, 261–304.
- 603 Canard, E.F., Mouquet, N., Mouillot, D., Stanko, M., Miklisova, D. & Gravel, D. (2014). [Empirical Evaluation](#)  
604 [of Neutral Interactions in Host-Parasite Networks](#). *The American Naturalist*, 183, 468–479.

- 605 Canard, E., Mouquet, N., Marescot, L., Gaston, K.J., Gravel, D. & Mouillot, D. (2012). [Emergence of](#)  
606 [Structural Patterns in Neutral Trophic Networks](#). *PLOS ONE*, 7, e38295.
- 607 Carlson, C.J., Farrell, M.J., Grange, Z., Han, B.A., Mollentze, N., Phelan, A.L., *et al.* (2021). [The future of](#)  
608 [zoonotic risk prediction](#). *Philosophical Transactions of the Royal Society B: Biological Sciences*, 376,  
609 20200358.
- 610 Caron, D., Brose, U., Lurgi, M., Blanchet, F.G., Gravel, D. & Pollock, L.J. (2024). [Trait-matching models](#)  
611 [predict pairwise interactions across regions, not food web properties](#). *Global Ecology and Biogeography*,  
612 33, e13807.
- 613 Castoe, T.A., Spencer, C.L. & Parkinson, C.L. (2007). [Phylogeographic structure and historical demography of](#)  
614 [the western diamondback rattlesnake \(\*Crotalus Atrox\*\): A perspective on North American desert](#)  
615 [biogeography](#). *Molecular Phylogenetics and Evolution*, 42, 193–212.
- 616 Catchen, M.D., Poisot, T., Pollock, L.J. & Gonzalez, A. (2023). [The missing link: Discerning true from false](#)  
617 [negatives when sampling species interaction networks](#).
- 618 Cazelles, K., Araujo, M.B., Mouquet, N. & Gravel, D. (2016). [A theory for species co-occurrence in interaction](#)  
619 [networks](#). *Theoretical Ecology*, 9, 39–48.
- 620 Chagnon, P.-L. (2015). [Characterizing topology of ecological networks along gradients: The limits of metrics'](#)  
621 [standardization](#). *Ecological Complexity*, 22, 36–39.
- 622 Choh, Y., Ignacio, M., Sabelis, M.W. & Janssen, A. (2012). [Predator-prey role reversals, juvenile experience](#)  
623 [and adult antipredator behaviour](#). *Scientific Reports*, 2, 728.
- 624 Cimini, G., Squartini, T., Saracco, F., Garlaschelli, D., Gabrielli, A. & Caldarelli, G. (2019). [The statistical](#)  
625 [physics of real-world networks](#). *Nature Reviews Physics*, 1, 58–71.
- 626 Cirtwill, A.R., Eklöf, A., Roslin, T., Wootton, K. & Gravel, D. (2019). [A quantitative framework for](#)  
627 [investigating the reliability of empirical network construction](#). *Methods in Ecology and Evolution*, 10,  
628 902–911.
- 629 Clark, J.S., Bell, D.M., Hersh, M.H., Kwit, M.C., Moran, E., Salk, C., *et al.* (2011). [Individual-scale variation,](#)  
630 [species-scale differences: Inference needed to understand diversity](#). *Ecology Letters*, 14, 1273–1287.
- 631 Dallas, T., Park, A.W. & Drake, J.M. (2017). [Predicting cryptic links in host-parasite networks](#). *PLOS*  
632 [Computational Biology](#), 13, e1005557.

- 633 Dansereau, G., Barros, C. & Poisot, T. (2023). Spatially explicit predictions of food web structure from regional  
634 level data.
- 635 Dunne, J.A. (2006). The Network Structure of Food Webs. In: *Ecological networks: Linking structure and*  
636 *dynamics* (eds. Dunne, J.A. & Pascual, M.). Oxford University Press, pp. 27–86.
- 637 Eklöf, A. & Stouffer, D.B. (2016). The phylogenetic component of food web structure and intervality.  
638 *Theoretical Ecology*, 9, 107–115.
- 639 Elmasri, M., Farrell, M.J., Davies, T.J. & Stephens, D.A. (2020). A hierarchical Bayesian model for predicting  
640 ecological interactions using scaled evolutionary relationships. *The Annals of Applied Statistics*, 14,  
641 221–240.
- 642 Elton, C.S. (2001). *Animal Ecology*. University of Chicago Press, Chicago, IL.
- 643 Emmerson, M.C. & Raffaelli, D. (2004). Predator–prey body size, interaction strength and the stability of a real  
644 food web. *Journal of Animal Ecology*, 73, 399–409.
- 645 Fedorov, V.B., Goropashnaya, A.V., Boeskorov, G.G. & Cook, J.A. (2008). Comparative phylogeography and  
646 demographic history of the wood lemming (*Myopus schisticolor*): Implications for late Quaternary history  
647 of the taiga species in Eurasia. *Molecular Ecology*, 17, 598–610.
- 648 Fortuna, M.A. & Bascompte, J. (2006). Habitat loss and the structure of plant–animal mutualistic networks.  
649 *Ecology Letters*, 9, 281–286.
- 650 Fu, X., Seo, E., Clarke, J. & Hutchinson, R.A. (2021). Link Prediction Under Imperfect Detection:  
651 Collaborative Filtering for Ecological Networks. *IEEE Transactions on Knowledge and Data Engineering*,  
652 33, 3117–3128.
- 653 Galiana, N., Lurgi, M., Claramunt-López, B., Fortin, M.-J., Leroux, S., Cazelles, K., et al. (2018). The spatial  
654 scaling of species interaction networks. *Nature Ecology & Evolution*, 2, 782–790.
- 655 Gilman, S.E., Urban, M.C., Tewksbury, J., Gilchrist, G.W. & Holt, R.D. (2010). A framework for community  
656 interactions under climate change. *Trends in Ecology & Evolution*, 25, 325–331.
- 657 Godsoe, W., Murray, R. & Iritani, R. (2022). Species interactions and diversity: A unified framework using Hill  
658 numbers. *Oikos*, n/a, e09282.
- 659 Golubski, A.J. & Abrams, P.A. (2011). Modifying modifiers: What happens when interspecific interactions  
660 interact? *Journal of Animal Ecology*, 80, 1097–1108.

- 661 Gonzalez, A. & Londoño, M.C. (2022). Monitor biodiversity for action. *Science*, 378, 1147–1147.
- 662 Gonzalez-Varo, J.P. & Traveset, A. (2016). The Labile Limits of Forbidden Interactions. *Trends in Ecology &*  
663 *Evolution*, 31, 700–710.
- 664 Gravel, D., Baiser, B., Dunne, J.A., Kopelke, J.-P., Martinez, N.D., Nyman, T., *et al.* (2019). Bringing Elton  
665 and Grinnell together: A quantitative framework to represent the biogeography of ecological interaction  
666 networks. *Ecography*, 42, 401–415.
- 667 Gravel, D., Poisot, T., Albouy, C., Velez, L. & Mouillot, D. (2013). Inferring food web structure from  
668 predator–prey body size relationships. *Methods in Ecology and Evolution*, 4, 1083–1090.
- 669 Guimarães, P.R. (2020). The Structure of Ecological Networks Across Levels of Organization. *Annual Review*  
670 *of Ecology, Evolution, and Systematics*, 51, 433–460.
- 671 Guimerà, R. & Sales-Pardo, M. (2009). Missing and spurious interactions and the reconstruction of complex  
672 networks. *Proceedings of the National Academy of Sciences*, 106, 22073–22078.
- 673 Hemprich-Bennett, D.R., Oliveira, H.F.M., Le Comber, S.C., Rossiter, S.J. & Clare, E.L. (2021). Assessing the  
674 impact of taxon resolution on network structure. *Ecology*, 102, e03256.
- 675 Herrera, C.M. (1989). Pollinator abundance, morphology, and flower visitation rate: Analysis of the “quantity”  
676 component in a plant-pollinator system. *Oecologia*, 80, 241–248.
- 677 Hortal, J., de Bello, F., Diniz-Filho, J.A.F., Lewinsohn, T.M., Lobo, J.M. & Ladle, R.J. (2015). Seven Shortfalls  
678 that Beset Large-Scale Knowledge of Biodiversity. *Annual Review of Ecology, Evolution, and Systematics*,  
679 46, 523–549.
- 680 Jordano, P. (1987). Patterns of Mutualistic Interactions in Pollination and Seed Dispersal: Connectance,  
681 Dependence Asymmetries, and Coevolution. *The American Naturalist*, 129, 657–677.
- 682 Jordano, P. (2016). Sampling networks of ecological interactions. *Functional Ecology*, 30, 1883–1893.
- 683 Jordano, P., Bascompte, J. & Olesen, J.M. (2003). Invariant properties in coevolutionary networks of  
684 plant–animal interactions. *Ecology Letters*, 6, 69–81.
- 685 Kausrud, K.L., Mysterud, A., Steen, H., Vik, J.O., Østbye, E., Cazelles, B., *et al.* (2008). Linking climate  
686 change to lemming cycles. *Nature*, 456, 93–97.

- 687 Kéfi, S., Berlow, E.L., Wieters, E.A., Joppa, L.N., Wood, S.A., Brose, U., *et al.* (2015). Network structure  
688 beyond food webs: Mapping non-trophic and trophic interactions on Chilean rocky shores. *Ecology*, 96,  
689 291–303.
- 690 Kéfi, S., Berlow, E.L., Wieters, E.A., Navarrete, S.A., Petchey, O.L., Wood, S.A., *et al.* (2012). More than a  
691 meal... integrating non-feeding interactions into food webs. *Ecology Letters*, 15, 291–300.
- 692 Kéfi, S., Miele, V., Wieters, E.A., Navarrete, S.A. & Berlow, E.L. (2016). How Structured Is the Entangled  
693 Bank? The Surprisingly Simple Organization of Multiplex Ecological Networks Leads to Increased  
694 Persistence and Resilience. *PLOS Biology*, 14, e1002527.
- 695 Kissner, K.J., Forbes, M.R. & Secoy, D.M. (1997). Rattling Behavior of Prairie Rattlesnakes (*Crotalus viridis*  
696 *viridis*, Viperidae) in Relation to Sex, Reproductive Status, Body Size, and Body Temperature. *Ethology*,  
697 103, 1042–1050.
- 698 Klecka, J. & Boukal, D.S. (2014). The effect of habitat structure on prey mortality depends on predator and  
699 prey microhabitat use. *Oecologia*, 176, 183–191.
- 700 Kopelke, J.-P., Nyman, T., Cazelles, K., Gravel, D., Vissault, S. & Roslin, T. (2017). Food-web structure of  
701 willow-galling sawflies and their natural enemies across Europe. *Ecology*, 98, 1730–1730.
- 702 Kulkarni, M.A., Kweka, E., Nyale, E., Lyatuu, E., Mosha, F.W., Chandramohan, D., *et al.* (2006).  
703 Entomological Evaluation of Malaria Vectors at Different Altitudes in Hai District, Northeastern Tanzania.  
704 *Journal of Medical Entomology*, 43, 580–588.
- 705 Maiorano, L., Montemaggioli, A., Ficetola, G.F., O'Connor, L. & Thuiller, W. (2020). TETRA-EU 1.0: A  
706 species-level trophic metaweb of European tetrapods. *Global Ecology and Biogeography*, 29, 1452–1457.
- 707 McCann, K. (2007). Protecting biostructure. *Nature*, 446, 29–29.
- 708 McCann, K.S. (2011). Food webs (MPB-50). In: *Food Webs (MPB-50)*. Princeton University Press.
- 709 McLeod, A., Leroux, S.J., Gravel, D., Chu, C., Cirtwill, A.R., Fortin, M.-J., *et al.* (2021). Sampling and  
710 asymptotic network properties of spatial multi-trophic networks. *Oikos*, n/a.
- 711 Melián, C.J., Vilas, C., Baldó, F., González-Ortegón, E., Drake, P. & Williams, R.J. (2011). Eco-evolutionary  
712 Dynamics of Individual-Based Food Webs. In: *Advances in Ecological Research*, The Role of Body Size in  
713 Multispecies Systems (ed. Belgrano, A.). Academic Press, pp. 225–268.

- 714 Morales-Castilla, I., Matias, M.G., Gravel, D. & Araújo, M.B. (2015). [Inferring biotic interactions from](#)  
715 [proxies](#). *Trends in Ecology & Evolution*, 30, 347–356.
- 716 Ohlmann, M., Miele, V., Dray, S., Chalmandrier, L., O'Connor, L. & Thuiller, W. (2019). [Diversity indices for](#)  
717 [ecological networks: A unifying framework using Hill numbers](#). *Ecology Letters*, 22, 737–747.
- 718 Olesen, J.M., Bascompte, J., Dupont, Y.L., Elberling, H., Rasmussen, C. & Jordano, P. (2010). [Missing and](#)  
719 [forbidden links in mutualistic networks](#). *Proceedings of the Royal Society B: Biological Sciences*, 278,  
720 725–732.
- 721 Park, J. & Newman, M.E.J. (2004). [Statistical mechanics of networks](#). *Physical Review E*, 70, 066117.
- 722 Pascual, M., Dunne, J.A. & Dunne, J.A. (2006). *Ecological Networks: Linking Structure to Dynamics in Food*  
723 [Webs](#). Oxford University Press, USA.
- 724 Pellissier, L., Albouy, C., Bascompte, J., Farwig, N., Graham, C., Loreau, M., *et al.* (2018). [Comparing species](#)  
725 [interaction networks along environmental gradients](#). *Biological Reviews*, 93, 785–800.
- 726 Petchey, O.L., Beckerman, A.P., Riede, J.O. & Warren, P.H. (2008). [Size, foraging, and food web structure](#).  
727 *Proceedings of the National Academy of Sciences*, 105, 4191–4196.
- 728 Pilosof, S., Porter, M.A., Pascual, M. & Kéfi, S. (2017). [The multilayer nature of ecological networks](#). *Nature*  
729 [Ecology & Evolution](#), 1, 1–9.
- 730 Plagányi, É.E. & Butterworth, D.S. (2004). [A critical look at the potential of Ecopath with ecosim to assist in](#)  
731 [practical fisheries management](#). *African Journal of Marine Science*, 26, 261–287.
- 732 Poisot, T., Canard, E., Mouillot, D., Mouquet, N. & Gravel, D. (2012). [The dissimilarity of species interaction](#)  
733 [networks](#). *Ecology Letters*, 15, 1353–1361.
- 734 Poisot, T., Cirtwill, A.R., Cazelles, K., Gravel, D., Fortin, M.-J. & Stouffer, D.B. (2016). [The structure of](#)  
735 [probabilistic networks](#). *Methods in Ecology and Evolution*, 7, 303–312.
- 736 Poisot, T. & Gravel, D. (2014). [When is an ecological network complex? Connectance drives degree](#)  
737 [distribution and emerging network properties](#). *PeerJ*, 2, e251.
- 738 Poisot, T., Stouffer, D.B. & Gravel, D. (2015). [Beyond species: Why ecological interaction networks vary](#)  
739 [through space and time](#). *Oikos*, 124, 243–251.

- 740 Polis, G.A. (1991). [Complex Trophic Interactions in Deserts: An Empirical Critique of Food-Web Theory](#). *The American Naturalist*, 138, 123–155.
- 741
- 742 Pollock, L.J., Tingley, R., Morris, W.K., Golding, N., O’Hara, R.B., Parris, K.M., *et al.* (2014). [Understanding co-occurrence by modelling species simultaneously with a Joint Species Distribution Model \(JSDM\)](#). *Methods in Ecology and Evolution*, 5, 397–406.
- 743
- 744
- 745 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.
- 746
- 747 Pringle, R.M. & Hutchinson, M.C. (2020). [Resolving Food-Web Structure](#). *Annual Review of Ecology, Evolution, and Systematics*, 51, 55–80.
- 748
- 749 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.
- 750
- 751 Pulliam, H.R. (1974). [On the Theory of Optimal Diets](#). *The American Naturalist*, 108, 59–74.
- 752 Rooney, N. & McCann, K.S. (2012). [Integrating food web diversity, structure and stability](#). *Trends in Ecology & Evolution*, 27, 40–46.
- 753
- 754 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.
- 755
- 756 Sanders, D. & van Veen, F.J.F. (2012). [Indirect commensalism promotes persistence of secondary consumer species](#). *Biology Letters*, 8, 960–963.
- 757
- 758 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.
- 759
- 760 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.
- 761
- 762 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](#).
- 763
- 764 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.
- 765
- 766

- 767 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.
- 768
- 769 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](#).
- 770
- 771 *Ecology Letters*, 20, 693–707.
- 772 Stock, M., Poisot, T., Waegeman, W. & De Baets, B. (2017). [Linear filtering reveals false negatives in species interaction data](#). *SCIENTIFIC REPORTS*, 7, 45908.
- 773
- 774 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.
- 775
- 776 Stouffer, D.B., Sales-Pardo, M., Sirer, M.I. & Bascompte, J. (2012). [Evolutionary Conservation of Species' Roles in Food Webs](#). *Science*, 335, 1489–1492.
- 777
- 778 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.
- 779
- 780
- 781 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.
- 782
- 783
- 784 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.
- 785
- 786
- 787 Tylianakis, J.M., Tscharntke, T. & Lewis, O.T. (2007). [Habitat modification alters the structure of tropical host–parasitoid food webs](#). *Nature*, 445, 202–205.
- 788
- 789 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.
- 790
- 791 Wagenmakers, E.-J. & Farrell, S. (2004). [AIC model selection using Akaike weights](#). *Psychonomic Bulletin & Review*, 11, 192–196.
- 792
- 793 Weinstein, B.G. & Graham, C.H. (2017a). [On comparing traits and abundance for predicting species interactions with imperfect detection](#). *Food Webs*, 11, 17–25.
- 794

- 795 Weinstein, B.G. & Graham, C.H. (2017b). Persistent bill and corolla matching despite shifting temporal  
796 resources in tropical hummingbird-plant interactions. *Ecology Letters*, 20, 326–335.
- 797 Woodward, G., Benstead, J.P., Beveridge, O.S., Blanchard, J., Brey, T., Brown, L.E., et al. (2010). Chapter 2 -  
798 Ecological Networks in a Changing Climate. In: *Advances in Ecological Research*, Ecological Networks  
799 (ed. Woodward, G.). Academic Press, pp. 71–138.
- 800 Woodward, G., Brown, L.E., Edwards, F.K., Hudson, L.N., Milner, A.M., Reuman, D.C., et al. (2012). Climate  
801 change impacts in multispecies systems: Drought alters food web size structure in a field experiment.  
802 *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367, 2990–2997.
- 803 Young, J.-G., Valdovinos, F.S. & Newman, M.E.J. (2021). Reconstruction of plant–pollinator networks from  
804 observational data. *Nature Communications*, 12, 3911.