

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

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



Last revision: June 6, 2024

## **1 Introduction**

### **2 Species interactions are 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 (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). The probability that species  
18 meet also depends on their biology, such as their phenology (Olesen *et al.* 2010; Singer & McBride 2012) and  
19 discoverability (Broom & Ruxton 2005). Finally, when species do come into contact, an interaction occurs only  
20 if their traits are locally compatible (Poisot *et al.* 2015), including their phenotypes (Bolnick *et al.* 2011; Gravel  
21 *et al.* 2013; Stouffer *et al.* 2011) and behavior (Choh *et al.* 2012; Pulliam 1974).

22 Documenting the location and timing of interactions becomes more difficult when accounting for the  
23 spatiotemporal variability of ecological interactions (Poisot *et al.* 2012, 2015). Environmental factors, such as  
24 temperature (Angilletta *et al.* 2004), drought (Woodward *et al.* 2012), climate change (Araujo *et al.* 2011;  
25 Gilman *et al.* 2010; Woodward *et al.* 2010), and habitat modifications (Tylianakis *et al.* 2007), contribute to this  
26 spatiotemporal variability by impacting species abundance and traits. Interactions may also be influenced by a  
27 third species (e.g., a more profitable prey species, Golubski & Abrams 2011; Sanders & van Veen 2012). Even  
28 under favorable circumstances, there remains a possibility that the interaction does not occur locally, either due  
29 to the intricate nature of the system or simply by chance. If it does occur, it might go undetected, particularly if

30 it happens infrequently. In this context, it is unsurprising that our knowledge of ecological interactions remains  
31 limited (Hortal *et al.* 2015) despite extensive biodiversity data collection (Schmeller *et al.* 2015).

## 32 Species interactions as probabilistic objects

33 Knowing the biological capacity of two species to interact directly (via e.g., trophic interactions) is necessary  
34 but not sufficient for inferring their interaction at a specific time and space. The recognition of the intrinsic  
35 variability of species interactions has led ecologists to expand their representation of ecological networks to  
36 include a probabilistic view of interactions (Dallas *et al.* 2017; Fu *et al.* 2021; Poisot *et al.* 2016). This allows  
37 filling in the Eltonian shortfall (Hortal *et al.* 2015) by modeling the probability of occurrence of interactions  
38 (e.g., Gravel *et al.* 2019), which can be an important tool for directing efforts and taking action (Carlson *et al.*  
39 2021), especially in places where access and resources for research are scarce. The probabilistic representation  
40 of interactions has been applied to direct interactions, which are conceptually and mathematically analogous  
41 regardless of their biological type (e.g., predation and pollination). This is in contrast with indirect interactions  
42 (e.g., interspecific competition), which arise from distinct ecological processes and are often not directly  
43 observable (Kéfi *et al.* 2015, 2016). Representing direct interactions probabilistically can capture the  
44 spatiotemporal variability of the aforementioned ecological processes and the uncertainty in our knowledge of  
45 interactions. By accounting for the uncertainty of interactions, networks of probabilistic interactions may  
46 provide a more realistic portrait of species interactions.

47 Networks of probabilistic interactions, within a Bayesian perspective, express our degree of belief (or  
48 confidence) regarding the occurrence or observation of interactions. In contrast, interactions are simply  
49 regarded as either occurring or not in networks of deterministic binary interactions. Based on the scale at which  
50 they are estimated, interaction probabilities may reflect our level of confidence in whether interactions will be  
51 observed, realized locally, or biologically feasible. As an illustration, we could outline a situation in which there  
52 is a 50% certainty that an interaction occurs 50% of the time. Our level of confidence should be more definitive  
53 (approaching either 0 or 1) as we extend our sampling to a broader area and over a longer time period, thereby  
54 diminishing the uncertainty of our knowledge of interactions (but not necessarily the estimation of their  
55 variability). In the broadest sense, binary interactions are also a type of probabilistic interaction, in which the  
56 numerical value of an interaction is restrained to 0 (non-occurring) or 1 (occurring). In networks of  
57 probabilistic interactions, only forbidden interactions (i.e., interactions prohibited by biological traits or species  
58 absence, Jordano *et al.* 2003; Olesen *et al.* 2010) have a probability value of zero, provided that intraspecific

59 trait variability is considered (Gonzalez-Varo & Traveset 2016).

60 The application and development of computational methods in network ecology, often based on a probabilistic  
61 representation of interactions, can alleviate (and guide) the sampling efforts required to document species  
62 interactions (Strydom *et al.* 2021). For example, statistical models can be used to estimate the uncertainty of  
63 pairwise interactions (Cirtwill *et al.* 2019) and the probability of missing (false negatives) and spurious (false  
64 positives) interactions (Guimerà & Sales-Pardo 2009). Statistical models can generate predictions of ecological  
65 networks without prior knowledge of pairwise interactions. They may do so using body size (Caron *et al.* 2024;  
66 Gravel *et al.* 2013; Petchey *et al.* 2008), phylogeny (Elmasri *et al.* 2020; Strydom *et al.* 2022), or a combination  
67 of niche and neutral processes (Bartomeus *et al.* 2016; Pomeranz *et al.* 2019) for inference. Topological null  
68 models, which generate networks of probabilistic interactions by preserving chosen characteristics of the  
69 adjacency matrix of binary interactions while intentionally omitting others (Bascompte *et al.* 2003; Fortuna &  
70 Bascompte 2006), are examples of common probabilistic interaction models. Null models can be used to  
71 produce underlying distributions of network measures for null hypothesis significance testing. However, how  
72 the uncertainty of pairwise interactions propagates to network structure (i.e., community-level properties  
73 driving the functioning, dynamics, and resilience of ecosystems, McCann 2007; McCann 2011; Proulx *et al.*  
74 2005; Rooney & McCann 2012) remains to be elucidated. Many measures have been developed to describe the  
75 structure (Poisot *et al.* 2016) and diversity (Godsoe *et al.* 2022; Ohlmann *et al.* 2019) of probabilistic  
76 interaction networks. These models and measures support the use of this approach for the study of a wide range  
77 of ecological questions, from making better predictions of species distribution (Cazelles *et al.* 2016) to  
78 forecasting the impact of climate change on ecological networks (Gilman *et al.* 2010).

## 79 We lack clear definitions of probabilistic species interactions

80 Yet, a precise definition of probabilistic interactions appears to be lacking, making the estimation and use of  
81 these data more difficult. In this manuscript, we aim to take a step back by outlining different ways in which  
82 probabilistic interactions are defined and used in network ecology. We distinguish two broad categories of  
83 probabilistic interaction networks that necessitate distinct approaches: local networks describing probabilities  
84 of realized interactions, and regional networks (metawebs) describing probabilities of potential interactions. We  
85 highlight the distinctions in the ecological meaning of these two representations and examine their properties  
86 and relationships (particularly with space, time, and between each other).

87 The lack of clear guidelines on the use of probabilistic interaction data is worrisome, as it affects both data  
88 producers and re-users who generate and manipulate these numbers. This is concerning because sampling  
89 strategies and decisions regarding network construction can affect our understanding of network properties  
90 (Brimacombe *et al.* 2023). There is currently no reporting standard that could guide the documentation of all  
91 types of probabilistic interactions (although Salim *et al.* 2022 discuss data standards for deterministic  
92 mutualistic networks). Clear reporting standards for probabilistic interactions would support more adequate  
93 manipulation and integration of interaction data from different sources and guard against possible  
94 misinterpretations arising from ambiguous definitions of probabilistic interaction networks. This  
95 documentation should outline the nature (i.e., local or regional) and type (e.g., predatory or pollination) of the  
96 interactions, provide information regarding the taxonomic level, identities, and characteristics (e.g., life stages)  
97 of the individuals involved in an interaction, present the mathematical formulation of probabilities, including  
98 clearly identified conditional variables (e.g., spatial and temporal scales), and describe the methods and contexts  
99 (e.g., location, time, environmental conditions) in which interactions were estimated. Inadequately documented  
100 probabilistic interaction data should be used with caution when analyzing ecological networks. The broad  
101 principles underlying our findings remain relevant and applicable across diverse ecological contexts involving  
102 direct interactions. In the following sections, we delve into the definitions of probabilistic interactions as we  
103 scale up from pairwise interactions to higher-level representations of ecological networks, i.e. local networks  
104 and metawebs.

## 105 **Pairwise interactions: the building blocks of ecological networks**

### 106 **Probabilistic interactions have different interpretations**

107 Consider a scenario where an avian predator has just established itself in a northern habitat home to a small  
108 rodent. Suppose these species have never co-occurred before, and as a result, their interaction has not been  
109 previously observed. What is the probability that the rodent is part of the diet of the avian predator, or put  
110 differently, what is the probability that they interact? Answering this question requires some clarification, as  
111 there are multiple ways to interpret and calculate interaction probabilities. We could calculate the probability  
112 that the traits of these species match, i.e. that the avian predator possesses the biological attributes to capture  
113 and consume the rodent. We could also calculate the probability that their traits support an interaction under the  
114 typical environmental conditions of the new habitat. For example, because avian predators hunt by sight,

115 predation could be possible in the absence of snow but highly improbable when snow is present, as rodents may  
116 use it as a shelter to hide from predators. Finally, we could calculate the probability that the avian predator will  
117 consume the rodent at *that* particular location, for which the spatial and temporal boundaries need to be  
118 specified. The estimation of interaction probabilities, whether through predictive models or prior distributions,  
119 hinges on our understanding of these probabilities and the specific ecological processes we aim to capture.

120 An important aspect to consider when using interaction probabilities is knowing if they describe potential or  
121 realized interactions, as these two types of interactions have distinct conceptual underpinnings and sources of  
122 uncertainty. A potential (regional) interaction is defined as the biological capacity of two taxa to interact (i.e.,  
123 the probability that they interact if they were to encounter each other and given sufficient time) whereas a  
124 realized (local) interaction is the occurrence or observation of this interaction in a well-defined space and time  
125 (i.e., the probability that they interact locally). For two co-occurring taxa and over enough time, the probability  
126 of local interaction is equivalent to the probability of regional (potential) interaction.

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

### 137 **The outcome of probabilistic interactions is usually binary**

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

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

144 Ecologists have traditionally represented interactions (edges) as binary objects that were considered realized  
145 after observing at least one individual from group  $i$  interact with at least another individual from group  $j$ . In an  
146 adjacency matrix  $B$  of binary interactions, the presence or absence of an interaction  $B_{i,j}$  between two taxa can  
147 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  
148 interaction. This interaction probability characterizes our limited ecological knowledge and/or the intrinsic  
149 spatiotemporal variability of interactions. It may be estimated through predictive models (e.g., those based on  
150 biological traits and species abundances) or expert (prior) knowledge about the interaction. In networks of  
151 probabilistic interactions, the edge values  $P(B_{i,j})$  are probabilistic events whose only two possible outcomes are  
152 the presence ( $B_{i,j} = 1$ ) or absence ( $B_{i,j} = 0$ ) of an interaction between each pair of nodes. Depending on the  
153 type of probabilistic interaction network (local or metaweb), the mathematical formulation and interpretation of  
154 stochastic parameters like  $P(B_{i,j})$  can be linked to environmental and biological factors such as species  
155 abundance, traits, area, and time, for example using logistic regression with continuous explanatory variables.

156 Predicting the number of local networks in which the interaction between two given taxa occurs can be achieved  
157 by using a Binomial distribution, assuming a constant interaction probability and independence between  
158 networks (trials). When considering uncertainties around the estimation of  $P(B_{i,j})$  or its spatiotemporal  
159 variability, a Beta distribution may be used to represent the relative likelihood of different probability values.  
160 For example, when calculating the probability of interaction between two taxa based on their local abundances,  
161 any spatiotemporal fluctuations in their abundances would introduce variability in the interaction probability at  
162 the local scale. If we take into account the stochasticity of the interaction probability, a Beta-Binomial  
163 distribution can be used to predict the number of networks in which the interaction occurs. Empirically  
164 observing an interaction between two taxa at a given location and time provides important information that can  
165 be used to update previous estimates of  $P(B_{i,j})$ , informing us on the biological capacity of both taxa to interact  
166 and the environmental conditions that enabled them to interact locally.

## 167 **Probabilistic interactions may also describe quantitative networks**

168 Even though binary interaction networks constitute a highly valuable source of ecological information (Pascual  
169 *et al.* 2006), they overlook interaction strengths. Represented in a quantitative adjacency matrix  $W$ , interaction  
170 strengths better describe the energy flows, demographic impacts or frequencies of interactions between nodes

171 (Berlow *et al.* 2004; Borrett & Scharler 2019), with  $W_{i,j}$  being a natural  $\mathbb{N}$  or real  $\mathbb{R}$  number depending on the  
172 measure. For example, they may represent local interaction rates between pairs of taxa (e.g., the flower-visiting  
173 rates of pollinators in a mutualistic network, Herrera 1989). When interaction strengths characterize predation  
174 pressure on prey, they can serve as parameters in a Lotka-Volterra model (e.g., Emmerson & Raffaelli 2004).  
175 The extra amount of ecological information in quantitative networks typically comes at a cost of greater  
176 sampling effort and data volume in predictive models (Strydom *et al.* 2021), which can lead to relatively high  
177 levels of uncertainties when inferring quantitative networks with limited data.

178 Like binary interaction networks, the uncertainty and variability of interaction strengths can be represented  
179 probabilistically. Interaction strengths can follow many probability distributions depending on the measure. For  
180 instance, they can follow a Poisson distribution  $W_{i,j} \sim \text{Poisson}(\lambda_{i,j}t_0)$  when predicting the number of  
181 interactions between individuals, with  $\lambda_{i,j}$  being the expected rate at which individuals of taxa  $i$  and  $j$  interact.  
182 The product  $\lambda_{i,j}t_0$  is the expected number of interactions realized by individuals during the time interval  $t_0$   
183 (e.g., the expected number of prey  $j$  consumed by all predators  $i$ ). The Poisson distribution can also be  
184 0-inflated when taking into account non-interacting taxa (e.g., Boulangeat *et al.* 2012 employ a 0-inflated model  
185 to analyze species abundance following the modeling of species presence and absence), which constitute the  
186 majority of taxa pairs in most local networks (Jordano 2016).

187 Because of the methodological difficulties typically encountered when building deterministic quantitative  
188 networks (which are only partially mitigated by models such as Ecopath, Plagányi & Butterworth 2004), binary  
189 interaction networks, which are easier to sample (Jordano 2016) and predict (Strydom *et al.* 2021), have been  
190 more frequently studied and modeled. Moreover, most published probabilistic interaction networks and methods  
191 describe interactions whose outcome is binary (whether interaction probabilities are regarded as constant or  
192 variable, e.g. represented by a Beta distribution), which underlines the need for better guidelines regarding the  
193 interpretation and manipulation of these types of networks first. For these reasons, our primary focus is  
194 interpreting and using interaction probabilities in Bernoulli distributions, in both local networks and metawebs.

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

196 **Species interactions occur in a local context**

197 Local networks of probabilistic interactions describe how likely taxa are to interact at a given location and time  
198 period. Local interactions are contingent upon the environmental conditions of the community and the matching  
199 of taxa biological traits. In local networks, edges commonly represent our degree of belief that two taxa interact  
200 in nature, but can also represent the probability of empirically *observing* this interaction (Catchen *et al.* 2023).

201 **Space and time**

202 We define space as the collection of geographic coordinates  $(x, y, z)$ , with  $(x, y)$  representing longitude and  
203 latitude coordinates, and  $z$  denoting either altitudes or depths. These point coordinates delineate the spatial  
204 boundaries of the system, which may be portrayed as a polyhedron. Ecological interactions may exhibit  
205 variations along latitudinal and altitudinal gradients, as evidenced by changes in hummingbird-plant  
206 interactions (Weinstein & Graham 2017a, b) and mosquito biting rates (e.g., Kulkarni *et al.* 2006) at different  
207 elevations. In contrast, time is defined as the specific time period within which interactions were either observed  
208 or predicted. Even though space and time are continuous variables that should yield probability *densities* of  
209 interactions (i.e., relative likelihoods of interactions occurring at infinitesimal locations and instants in time),  
210 these definitions enable them to be conceptualized as distinct patches and time segments. Treating space and  
211 time as discrete dimensions aligns with the common sampling methods of ecological networks and provides  
212 probabilities of interactions, which can be obtained by integrating probability densities over space and time. We  
213 can quantify both an area  $A$  and a duration  $t$ , which can readily be used in spatiotemporal analyses of ecological  
214 networks (Box 1).

215 **Co-occurrence**

216 The probability that two taxa  $i$  and  $j$  interact in a local network  $L_{x,y,z,t}$  (spatial and temporal subscripts hereafter  
217 omitted or replaced by the shorter subscript  $k$  for clarity) can be conditioned on many environmental and  
218 biological factors. In addition to network area (or volume) and duration, they may be conditioned on taxa  
219 co-occurrence  $X_{i,j,k}$ , which is usually a Boolean describing if the geographic distributions of both taxa overlap  
220 within the study area. As illustrated in Box 1, co-occurrence may be modeled probabilistically, in which case it

may conform to a Bernoulli distribution  $X_{i,j,k} \sim \text{Bernoulli}(P(X_{i,j,k}))$ . The probability of co-occurrence can be calculated using the individual (marginal) occurrence probabilities  $P(X_{i,k})$  and  $P(X_{j,k})$ . Given that taxa occurrences are not independent of each other, the probability of co-occurrence can be calculated by multiplying the probability of occurrence of one taxon by the probability of occurrence of the other given that the first one occurs:

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

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

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

## 231 Biological and environmental factors

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

Other important factors that can impact interaction probabilities at the local scale are taxa local abundances  $N_{i,k}$  and  $N_{j,k}$ , which affect encounter probabilities (Canard *et al.* 2012), and local traits distributions  $T_{i,k}$  and  $T_{j,k}$  (e.g., body mass, longevity, and habitat breadth, Caron *et al.* 2024), which determine the ability of individuals to interact after encountering each other (Poisot *et al.* 2015). Local interaction probabilities may also be

245 conditioned on higher-level properties of the network, which we denote by  $f(L)$ . Many topological null models  
246 (i.e., statistical models that randomize interactions by retaining certain properties of the network while  
247 excluding others) provide interaction probabilities from selected measures of network structure, such as  
248 connectance (Fortuna & Bascompte 2006) and the degree distribution (Bascompte *et al.* 2003).

## 249 **Local interactions must be biologically feasible**

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

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

254 Low values of  $P(L_{i,j}|M_{i,j})$  indicate that feasible interactions rarely occur locally, intermediate values around  
255 50% suggest considerable spatiotemporal variability, while high values indicate that regional interactions are  
256 nearly always realized locally. The local probability of interaction between a given pair of taxa is always equal  
257 to or below their probability of regional interaction.

## 258 **Conditional variables must be explicitly stated**

259 The probability that two taxa  $i$  and  $j$  interact in a local network  $L$  can thus be conditioned on their co-occurrence  
260  $X_{i,j,k}$  (or more explicitly on their occurrences  $X_{i,k}$  and  $X_{j,k}$ ), local abundances  $N_{i,k}$  and  $N_{j,k}$ , local traits  
261 distributions  $T_{i,k}$  and  $T_{j,k}$ , local environmental conditions  $E_k$ , network area (or volume)  $A$ , time interval  $t$ ,  
262 network properties  $f(L)$ , and biological feasibility  $M_{i,j}$ .

263 Although these variables correspond to distinct ecological inquiries or mechanisms related to ecological  
264 interactions, they may covary with each other, such as the possible dependence of  $X_{i,j,k}$  and  $E_k$  on spatial and  
265 temporal scales. When estimating interaction probabilities using e.g. a generalized linear model with multiple  
266 explanatory variables that might not be independent, it may become important to address collinearity. In such a  
267 case, to mitigate this issue, it may be necessary to use variable selection techniques before fitting the model to  
268 data. The probability that a local interaction is realized is described by the following expression when all these

269 conditional variables are included:

$$P(L_{i,j}|X_{i,k}, X_{j,k}, N_{i,k}, N_{j,k}, T_{i,k}, T_{j,k}, E_k, A, t, f(L), M_{i,j}) \quad (4)$$

270 The representation of the local context in which probabilities are estimated and the variables that should be  
271 taken into consideration depend on the study system, the objectives of the study, and the resources available to  
272 the researchers. In other words, these variables do not systematically need to be accounted for. For example,  
273 Gravel *et al.* (2019) analyzed local European host-parasite networks of willow-galling sawflies and their natural  
274 enemies, all referenced in space and time, to infer probabilities of local interactions between co-occurring  
275 species. This was achieved by including temperature and precipitation as conditional variables in their models.  
276 In Box 2, we reuse these data to show the extent of variation among these local networks. We do so by  
277 measuring their dissimilarity with the regional network (metaweb aggregating all local interactions), both in  
278 terms of species composition and interactions. We build local probabilistic networks following eq. 3, showing  
279 that insufficient local variation (high probability of local interaction among potentially interacting species)  
280 results in an overestimation in both the number of interactions and connectance (i.e., the proportion of all of the  
281 non-forbidden links that are realized).

282 When accounted for, conditional variables should be clearly described in the documentation of the data  
283 (Brimacombe *et al.* 2023), preferentially in mathematical terms to avoid any confusion in their interpretation  
284 and to limit manipulation errors during their re-use. For instance, ecologists should be explicit about their  
285 consideration of co-occurrence in their estimation of local interaction probabilities. Indeed, it is important to  
286 specify if probability values are conditional ( $P(L_{i,j}|X_{i,j,k} = 1)$ ) or not ( $P(L_{i,j})$ ) on co-occurrence, as this can  
287 change the interpretation of the data. In Tbl. 1, we present examples of studies that used different expressions of  
288 probabilistic interactions. We have included the probability of empirically observing an interaction that is  
289 realized locally  $P(O_{i,j}|L_{i,j})$  to underscore the distinction between local observations and actual realizations of  
290 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  $L$  of realized interactions, a local network  $O$  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). Uncertainty sources include uncertainty of model parameters, model structure, and input data. The study marked with an asterisk has been conducted on binary interaction networks.

Expression	Type	Outcome	Common models	Reference
$P(L_{i,j} X_{i,k}, X_{j,k}, \dots)$	local	realization of the interaction given taxa co-occurrence	species distribution models	Gravel <i>et al.</i> (2019), Dansereau <i>et al.</i> (2023)
$P(L_{i,j} N_{i,k}, N_{j,k}, \dots)$	local	realization of the interaction given taxa abundances	neutral models	Canard <i>et al.</i> (2014)
$P(L_{i,j} T_{i,k}, T_{j,k}, \dots)$	local	realization of the interaction given local traits	trait matching models	Caron <i>et al.</i> (2024)
$P(L_{i,j} E_k, \dots)$	local	realization of the interaction given local environmental conditions	environmental-based models	Gravel <i>et al.</i> (2019) (temperature and precipitation)
$P(L_{i,j} A, \dots)$	local	realization of the interaction in a given area or volume	spatial models	Galiana <i>et al.</i> (2018)*
$P(L_{i,j} t, \dots)$	local	realization of the interaction during a given time period	temporal models	Weinstein & Graham (2017a)
$P(L_{i,j} f(L), \dots)$	local	realization of the interaction given network structure	topological models	Fortuna & Bascompte (2006) (connectance), Stock <i>et al.</i> (2017)
$P(L_{i,j} M_{i,j}, \dots)$	local	realization of the interaction given that the taxa can biologically interact	spatiotemporal models	Dansereau <i>et al.</i> (2023)
$P(O_{i,j} L_{i,j}, \dots)$	local	observation of the interaction given that it is realized locally	sampling models	Catchen <i>et al.</i> (2023)

Expression	Type	Outcome	Common models	Reference
$P(M_{i,j} T_i, T_j)$	regional	biological feasibility of the interaction given regional traits (non-forbiddenness)	trait matching models	Strydom <i>et al.</i> (2022)
$P(M_{i,j}^* T_i, T_j, E)$	regional	ecological feasibility of the interaction given regional traits and environmental conditions	trait matching and environmental-based models	this study

291 **Interaction probabilities may be estimated with multiple models**

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

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

298 where the ellipsis serves as a placeholder for the conditional variables incorporated in these models. Model  
 299 structure constitutes another source of uncertainty that needs to be documented and quantified when estimating  
 300 interaction probabilities.

**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:

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

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  $P(L_{i,j})$ . 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.

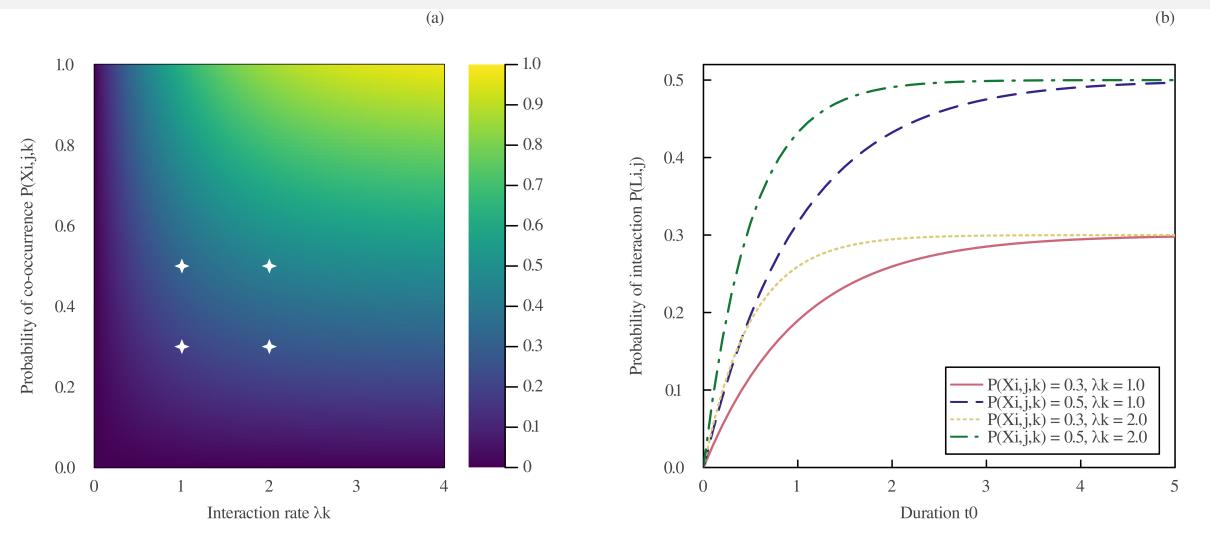
$$L_{i,j} \sim \text{Bernoulli}(P(L_{i,j})) \quad (7)$$

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

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

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

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.

303

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

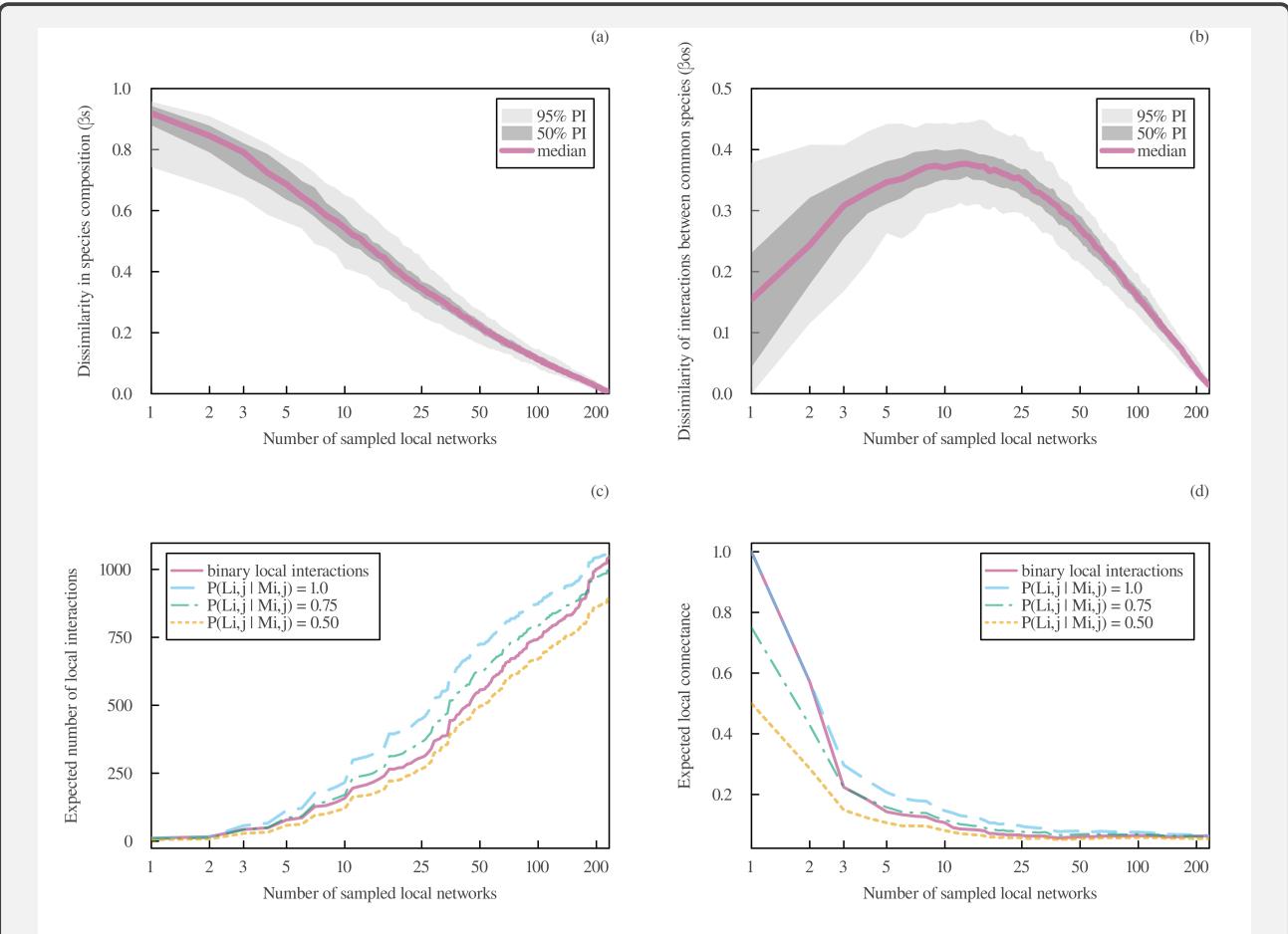
We use the collection of tripartite host-parasite networks of Kopelke *et al.* (2017), consisting of willows, willow-galling sawflies, and their natural enemies, sampled across Europe (233 local networks used). Given its replicated networks spanning large spatiotemporal scales, this dataset is well-suited for analyzing network variability, both in terms of species composition and interactions. Further details regarding data manipulation and network construction can be found in the supplementary material. 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 varies with the number of sampled local networks. To do so, we randomly selected one local network of binary interactions and sequentially sampled additional networks while aggregating both their species and interactions. We compared the metaweb and the aggregated local networks using the

304

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). We repeated this sampling process one hundred times and highlighted the median dissimilarity values across simulations, as well as the 50% and 95% percentile intervals. Both dissimilarity indices were calculated based on the number of items shared by the two networks ( $c_{LM}$ ) and the number of items unique to the metaweb ( $u_M$ ) and the aggregated local network ( $u_L$ ). The  $\beta_S$  dissimilarity index uses species (nodes) as items being compared, while the  $\beta_{OS}$  index assesses dissimilarity based on interactions between shared species. Both indices were calculated following the  $\beta_W$  index of Whittaker (1960):

$$\beta_W = \frac{c_{LM} + u_L + u_M}{(2c_{LM} + u_L + u_M)/2} - 1. \quad (11)$$



**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. Aggregated local networks were obtained by sequentially and randomly selecting a number of local networks of binary interactions and aggregating both their species and interactions. In both panels, the colored line represents the median dissimilarity across simulations and the grey areas cover the 50% and 95% percentile intervals. (c) Scaling of the number of 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  $P(L_{i,j} | M_{i,j})$  (the local probability of interaction among potentially interacting species, indices omitted in the figure for simplicity) was used for all local networks within a particular curve. Aggregated local networks of probabilistic interactions were obtained by sequentially and randomly selecting a number of local networks and aggregating both their species and interactions (with the value of  $P(L_{i,j} | M_{i,j})$  adjusting according to Eq. 12).

Expectingly, local networks are highly dissimilar from the metaweb in terms of species composition, especially when only a limited number of sites has been sampled. This is because species in the metaweb form the regional species pool, and few species 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. Interactions are more comprehensively captured when the number of sampled sites is high. Quadratic relationships of network properties with sampling effort were also observed by McLeod *et al.* (2021).

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). We built probabilistic local networks following Eq. 3, and set  $P(M_{i,j})$  to 1 when the interaction was observed at least once, and to 0 otherwise (i.e., no false positive or false negative in the metaweb). We used a constant value of  $P(L_{i,j}|M_{i,j})$  for all interactions.

When aggregating local networks of probabilistic interactions, the constancy of the probability of regional interaction across the entire study area means that any rise in the probability of local interaction is solely attributable to an increase in  $P(L_{i,j}|M_{i,j})$ . For example, let  $L_1$  and  $L_2$  be two local networks and  $L_{1,2}$  the aggregated network. If  $P(L_{1,i,j}|M_{i,j})$  and  $P(L_{2,i,j}|M_{i,j})$  are the probabilities that two potentially interacting taxa interact respectively in  $L_1$  and  $L_2$ , the probability  $P(L_{1,2,i,j}|M_{i,j})$  that these taxa interact in the aggregated network  $L_{1,2}$  is obtained by:

$$P(L_{1,2,i,j}|M_{i,j}) = 1 - [1 - P(L_{1,i,j}|M_{i,j})] \times [1 - P(L_{2,i,j}|M_{i,j})], \quad (12)$$

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

By comparing the scaling relationships observed in local networks of binary (empirical) and probabilistic (modeled) interactions, Fig. 2c-d shows that high values of  $P(L_{i,j}|M_{i,j})$  lead to systematic overestimations in the number of interactions and connectance, especially when  $P(L_{i,j}|M_{i,j}) = 1$  (corresponding to the scenario where local probabilities of interactions are equivalent to the probabilities of regional interactions). This suggests that high values of  $P(L_{i,j}|M_{i,j})$  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  $P(L_{i,j}|M_{i,j})$  is high. In contrast, low values

of  $P(L_{i,j}|M_{i,j})$  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.

308

## 309 Metawebs: regional catalogs of interactions

### 310 Unobserved interactions should be probabilistic

311 Metawebs (Dunne 2006) are networks of potential interactions over broad spatial, temporal, and taxonomic  
312 scales (e.g., food webs at the continental scale). Potential interactions describe the biological capacity of taxa to  
313 interact, which is typically assessed at the regional scale. Metawebs of probabilistic interactions are particularly  
314 useful in situations where there is uncertainty in the ability of taxa to interact. This uncertainty frequently arises  
315 due to insufficient interaction data, especially for taxa that have not yet been observed to co-occur, and  
316 uncertainties in trait-matching models. As data accumulates, interactions in metawebs should tend towards  
317 binarity, approaching probability values of 0 (repeatedly failing to observe an interaction between two  
318 co-occurring taxa) and 1 (observing an interaction at least once). Although *neutrally* forbidden interactions  
319 (i.e., between rare species, Canard *et al.* 2012) tend to have low probability values in local networks, they may  
320 have higher probabilities in the metaweb (i.e., when species' biological traits can support an interaction if they  
321 were to encounter each other). Likewise, non-co-occurring taxa may have a non-zero probability of interaction  
322 in the metaweb.

323 The extent of sampling effort influences our evaluation of probabilities of potential interactions, as sampling  
324 over a larger area or for a longer duration enables us to capture a greater number of regional interactions  
325 (McLeod *et al.* 2021). However, in contrast with local networks of probabilistic interactions, which describe  
326 local interaction stochasticity, regional interactions are not evaluated for any particular local context. In Box 3,  
327 we discuss the differences in spatial and temporal scaling of regional interactions compared to local interactions.  
328 We do so by using the host-parasite networks of Kopelke *et al.* (2017) as an illustration of spatial scaling.

329 **Potential interactions are sometimes linked to environmental conditions**

330 **Biological feasibility**

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

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

334 which, in contrast with local networks, is not conditioned on any spatial, temporal, co-occurrence or  
335 environmental variables (Tbl. 1). Regional traits may differ from local traits  $T_{i,k}$  and  $T_{j,k}$ , which may vary  
336 spatially and temporally due to phenotypic plasticity (Berg & Ellers 2010). The taxonomic level at which  
337 interactions are evaluated influences the distribution of regional and local traits. However, as explained in Box  
338 4, there is no fundamental difference in the taxonomic scaling of regional and local interactions (i.e., how  
339 interaction probabilities change with taxonomic level), unlike their spatial and temporal scaling.

340 The biological feasibility of interactions expresses our degree of belief that there exists at least one combination  
341 of phenotypes that could support an interaction if they were to encounter each other, assuming they had enough  
342 time to interact. Evaluating this probability is conducted without incorporating the environmental conditions  
343 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  
344 forbidden interactions (i.e., the probability that their traits do not support an interaction), which is based  
345 uniquely on biological traits:

$$P(M_{i,j}|T_i, T_j) = 1 - P(F_{i,j}|T_i, T_j). \quad (14)$$

346 For example, let  $i$  be a western diamondback rattlesnake (*Crotalus atrox*) and  $j$ , a wood lemming (*Myopus*  
347 *schisticolor*). These two taxa never co-occur, the rattlesnake being adapted to warm regions of North America  
348 (Castoe *et al.* 2007) and the lemming, to northern habitats of Eurasia (Fedorov *et al.* 2008). As we lack direct  
349 observations of an interaction between these two species, we have to rely on expert knowledge or trait-matching  
350 models to estimate their probability of potential interaction. To accurately estimate this probability using  
351 trait-matching models, it is crucial to ensure that the set of traits considered reflects the overall traits

352 distributions of both taxa. We could for instance consider their average body mass and the average phylogenetic  
353 proximity of lemmings to rattlesnakes' prey. Doing so, we might find a high probability of potential interaction  
354 based on these traits.

355 **Ecological feasibility**

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

$$P(M_{i,j}^*|T_i, T_j, E), \quad (15)$$

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

$$\int_E P(M_{i,j}^*|T_i, T_j, E) dE \leq P(M_{i,j}|T_i, T_j). \quad (16)$$

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

374 **Metawebs are built with interaction data and trait-matching models**

375 Starting from a selected set of taxa, which are usually distributed within a broad region of interest, metawebs  
376 can be built using different data sources, including literature review (e.g., Maiorano *et al.* 2020), aggregated  
377 interaction data (e.g., Gravel *et al.* 2019; Saravia *et al.* 2022), trait-matching models (e.g., Shaw *et al.* 2024;  
378 Strydom *et al.* 2022), and expert knowledge. Every pair of taxa that have confidently been observed to interact  
379 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  
380 from local networks of probabilistic interactions, where interaction events may remain stochastic (i.e.,  
381  $P(L_{i,j}) < 1$ ) even after empirically observing interactions due to their spatiotemporal variability. Interactions  
382 that were never observed typically have low probability values in local networks and vary from low to high  
383 values in metawebs, contingent upon taxa traits distributions (reaching 0 for forbidden links).  
384 When incorporating local network data to estimate probabilities of regional interactions, repeatedly failing to  
385 observe an interaction between two co-occurring taxa (i.e.,  $P(O_{i,j} = 0)$ ) should decrease the probability that the  
386 interaction is biologically feasible. Using Bayes' theorem, the probability that the interaction is biologically  
387 feasible given that it was never observed locally,  $P(M_{i,j} = 1|O_{i,j} = 0, \dots)$ , may be calculated as follows:

$$P(M_{i,j} = 1|O_{i,j} = 0, \dots) = \frac{P(O_{i,j} = 0|M_{i,j} = 1, \dots) \times P(M_{i,j} = 1|\dots)}{P(O_{i,j} = 0|\dots)}. \quad (17)$$

388 The reduction in the probability of regional interaction after considering that it was never observed locally (i.e.,  
389  $P(M_{i,j} = 1|O_{i,j} = 0, \dots) < P(M_{i,j} = 1|\dots)$ ) occurs because  $P(O_{i,j} = 0|M_{i,j} = 1, \dots)$  must be lower than  
390  $P(O_{i,j} = 0, \dots)$ , i.e. there is a higher chance of observing an interaction when it is biologically feasible.  
391 Many observations of interactions are false positives because of observation errors due to taxonomic  
392 misidentifications and ecological misinterpretations (e.g., due to cryptic species and interactions, Pringle &  
393 Hutchinson 2020). Likewise, forbidden interactions may be false negatives, e.g. if they have been evaluated  
394 based on unrepresentative or incomplete traits distributions. Employing Bayesian models proves valuable when  
395 estimating interaction probabilities in metawebs (e.g., Bartomeus *et al.* 2016; Cirtwill *et al.* 2019). This  
396 improvement is achieved by updating prior information regarding the feasibility of interactions (e.g., experts'  
397 prior assessments of interaction probabilities) with empirical data on interactions and traits.

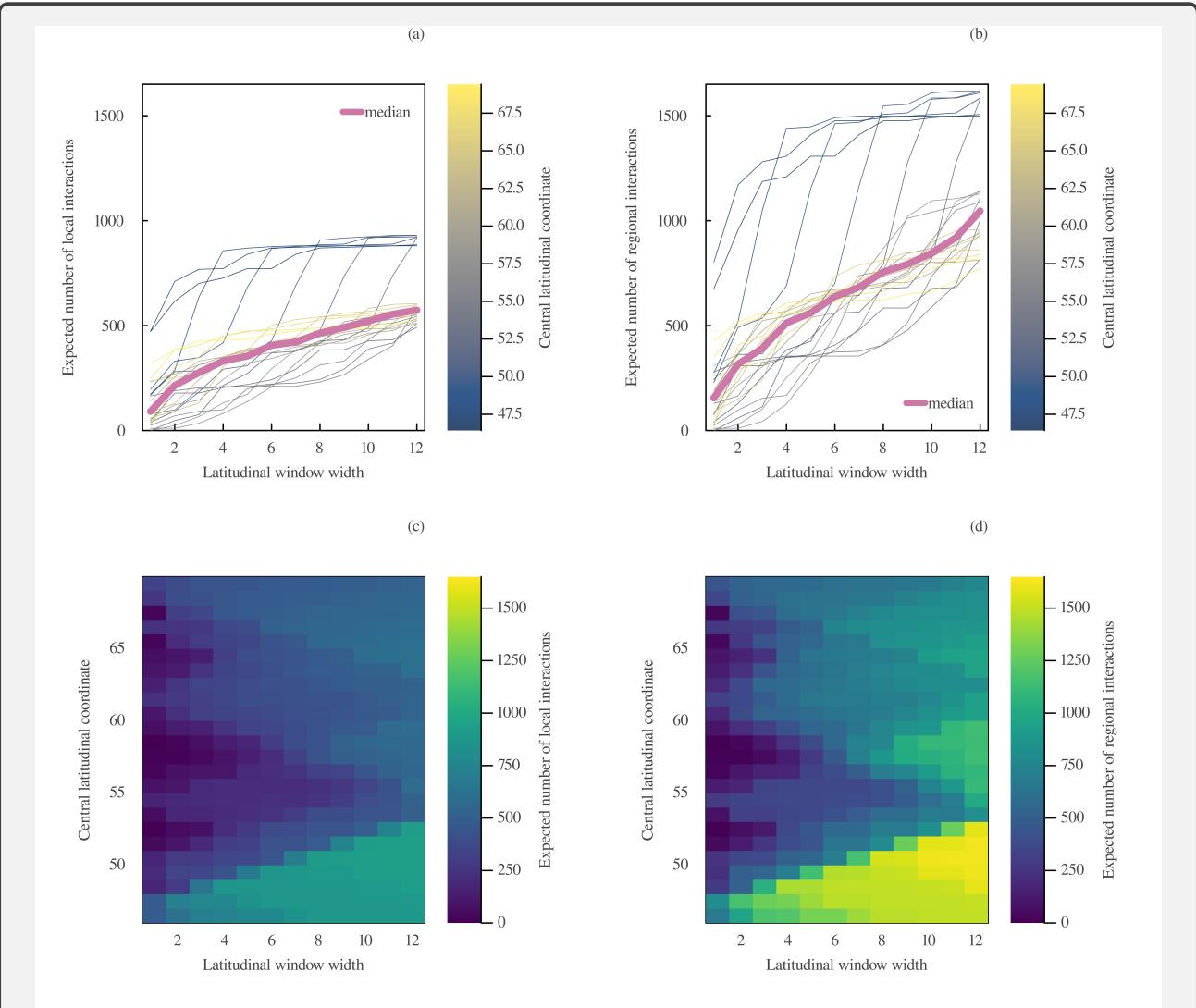
### Box 3: Spatial and temporal scaling of interactions

Network-Area Relationships document the scaling of network properties (such as modularity and connectance) with spatial scale (e.g., Galiana *et al.* 2018; Wood *et al.* 2015). The variation in network structure across spatial scales may stem from the scaling of species richness (species-area relationships, SARs), the number of interactions (Brose *et al.* 2004), and many other higher-level properties of the system (e.g., environmental heterogeneity, Thompson & Townsend 2005) with the sampled area. Likewise, interaction accumulation curves describe the scaling of the number of observed interactions with sampling effort (Jordano 2016). Sampling effort, which may correspond to the duration of the sampling period used to construct the network, can impact connectance (Bersier *et al.* 1999) and various measures of network structure (Banašek-Richter *et al.* 2004; McLeod *et al.* 2021). Apart from sampling effort, the temporal scaling of interactions also describes how network structure changes with the temporal resolution of the network, acknowledging that distinct interactions take place over time, ranging from short-term fluctuations of interactions to long-term trends. As local networks of probabilistic interactions may explicitly account for the spatiotemporal variability of interactions, they offer a distinct approach to investigating the scaling of network structure with space and time.

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

On the other hand, local interactions scale both spatially and temporally, given that they have more

opportunities to be realized and observed in larger areas and longer durations. This is attributed to factors such as a higher number of individuals, greater trait variations, and increased opportunities for encounters, as highlighted by McLeod *et al.* (2020). For example, if a local network of probabilistic interactions  $L_1$  with an area  $A_1$  is derived from 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.  $P(L_{1,i,j}|A_1 < A_0) \leq P(L_{0,i,j}|A_0)$ . However, if  $A_1$  and  $A_0$  are disjoint, interaction probabilities could be higher in the smaller area, contingent upon their environmental and biological conditions. Likewise, interaction probabilities should be lower in networks with shorter durations when time intervals are nested. In Fig. 3, we show how the expected number of local host-parasite interactions scales with area, represented as an expanding latitudinal window, in comparison with regional interactions (data from Kopelke *et al.* 2017). Even though we employed local probabilities of interactions equal to regional interactions for the purpose of comparison (i.e., using  $P(L_{i,j}|M_{i,j}) = 1$  here), we notice that the total number of regional interactions scales more rapidly than local interactions. This is because numerous regional interactions involve species that never co-occur, and as a result, these interactions are not captured in local networks.



**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 central latitudes. Probabilities of regional interactions were obtained with a false positive rate of 5% and a false negative rate of 10%. Local probabilistic interactions were derived from probabilistic regional interactions by setting the value of  $P(L_{i,j}|M_{i,j})$  (the local probability of interaction among potentially interacting species) to 1, ensuring a conservative comparison between aggregated local networks and metawebs. Aggregated local networks were obtained by aggregating both the species and interactions found within a particular latitudinal window, with the values of  $P(L_{i,j}|M_{i,j})$  remaining at their maximum value of 1 following Eq. 12. Further details on the data and network construction are presented in the supplementary material.

#### Box 4: Taxonomic scaling of interactions

Probabilistic interaction networks offer a versatile approach to tackle a broad array of ecological questions, depending on their level of organization. For instance, the assemblage of interactions across ecological scales can be explored through species-based networks, while clade-based networks provide insights into macroevolutionary processes (Gomez *et al.* 2010). Given that our interpretation of the properties and dynamics of ecological networks depends on their taxonomic level (Guimarães 2020), investigating the taxonomic scaling of interactions (i.e., how interaction probabilities change with taxonomic level) emerges as a promising research avenue.

There are no inherent differences between the taxonomic scaling of local networks and metawebs. Interaction probabilities in both local networks (Eq. 4) and metawebs (Eq. 13) are not directly conditioned on taxonomic level. The taxonomic scale is tied to the definition of the event itself (i.e., the interaction between two taxa, defined at the desired taxonomical scale), not to the variables on which interaction probabilities are conditioned. However, some conditional variables (e.g. trait distribution) may vary with taxonomic scale. In such cases, interaction probabilities would vary taxonomically following the scaling of these variables.

In both types of networks, transitioning to a broader level of organization (e.g., from a species-level network  $S$  to a genus-level network  $G$ ) can be accomplished directly by using probabilities from finer scales. For example, in a network with  $n_1$  species from genus  $g_1$  and  $n_2$  species from genus  $g_2$ , one can compute the probability that at least one species from genus  $g_1$  interacts with at least one species from genus  $g_2$  (i.e., that there is a non-zero number of species-level interactions or, equivalently, that the genus-level interaction occurs) as follows:

$$P(G_{g_1,g_2} = 1) = 1 - \prod_{i=1}^{n_1} \prod_{j=1}^{n_2} (1 - P(S_{g_1,i,g_2,j} = 1)), \quad (18)$$

where  $g_{1,i}$  and  $g_{2,j}$  are the species of the corresponding genus and assuming independence between species-level interactions. If it is known that at least two of these species interact (i.e.,  $P(S_{g_1,i,g_2,j}) = 1$  for at least one pair of  $(g_{1,i}, g_{2,j})$ ), it implies a probability of genus interaction equal to 1. Canard *et al.* (2012) built a species-based network following a similar approach, by using simulated interactions between individuals derived from a neutral model (i.e., a model that assumed ecological equivalence among individuals).

In contrast, a more sophisticated approach is necessary when transitioning from a broader to a finer level of organization. This is because the knowledge of an interaction between two genera does not guarantee that all possible pairwise combinations of their species will also interact. One possible method is to build a finer-scale network by generating probabilities of interactions through random sampling from a beta distribution, parameterized by the broader-scale network.

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

402

### **Box 5: Sampling for binary interaction networks**

The prediction of binary interactions through Bernoulli trials is an important application of probabilistic interaction networks. This approach proves beneficial for analyzing the structural characteristics of probabilistic interaction networks, particularly in the absence of specific analytical formulas for measures of network structure (Poisot *et al.* 2016). By performing independent Bernoulli trials for each interaction, a network of binary interactions may be generated. A probability distribution of network properties can be obtained by measuring network structure across multiple randomly generated networks (Poisot *et al.*

403

2016). This method enables the representation of the variability or uncertainty of network structure, albeit with possible biases when connectance is low (Chagnon 2015; Poisot & Gravel 2014). Employing this strategy to generate binary interaction networks under a null model facilitates null hypothesis significance testing, wherein the observed measure is compared against the simulated distribution (e.g., Bascompte *et al.* 2003). Yet, inferring network structure through sampling for binary interactions assumes independence among interactions, which might not accurately represent reality. Covariation among interactions could exist even if we do not explicitly condition interactions on others. For example, an interaction between two taxa could be more probable when another interaction occurs. The consequences of this assumption of independence on the prediction of network structure have yet to be empirically examined.

There are at least two approaches to sampling binary interaction networks across space from probabilistic interaction networks (i.e., when predicting a binary interaction network for each location  $k$  within a given region). Both approaches assume independence between interactions. The first approach involves performing a singular Bernoulli trial for each pair of taxa based on their regional probability of interaction:

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

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

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

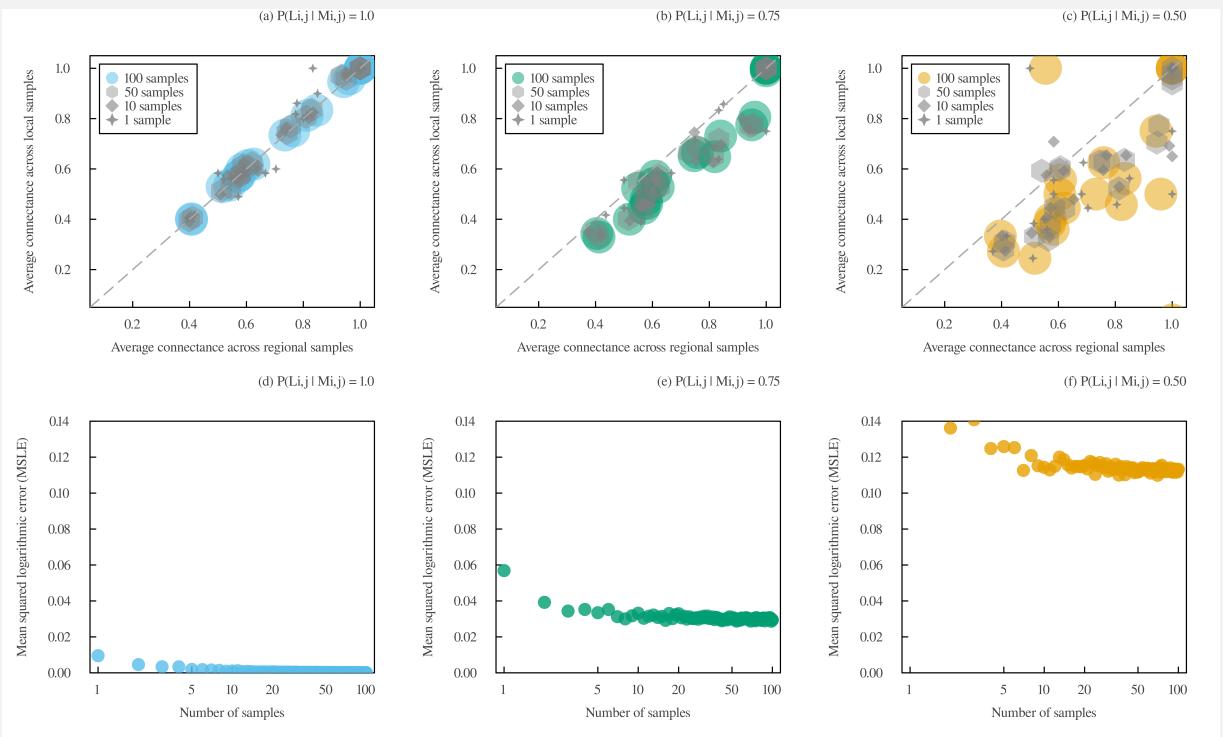
$$L_{k,i,j} \sim \text{Bernoulli}(P(L_{k,i,j})).$$

This can be achieved by first generating distinct probabilistic interaction networks for each location. These local networks of probabilistic interactions may vary in taxa composition and interaction probabilities. Because binary interaction networks are sampled independently for each location, this second approach introduces spatial variation in binary interactions. This approach may more effectively capture network structure across space and time, facilitating the investigation of ecological hypotheses about interactions at broad spatial and temporal scales.

In Fig. 4, we compare the average connectance of binary interaction networks resulting from these

two sampling techniques. We drew regional and local interactions from our host-parasite networks of probabilistic interactions (Kopelke *et al.* 2017), generating a number of binary interaction network realizations for each site in the dataset. These two sampling approaches yield different outcomes, particularly for intermediate values of  $P(L_{i,j}|M_{i,j})$ , which represent instances where regional interactions do not consistently manifest locally (i.e., with the largest local variability). Small differences between these techniques are also apparent when we equate the probability of local interaction to the probability of regional interaction (i.e., when using  $P(L_{i,j}|M_{i,j}) = 1.0$ , Fig. 4a,d), especially when the number of samples of binary interaction networks for each location is low. 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. 4a-c). Furthermore, we observe an increase in the variability of connectance when employing a single simulation (Fig. 4, cross markers), which is a more tangible representation of the process leading to the realization of local interactions in nature.

405



**Figure 4: Connectance of sampled binary interaction networks.** (a-c) Average connectance of binary interaction networks obtained from the two sampling approaches 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 methods as the number of samples increases. The local probability of interaction between potentially interacting species was set to three different values: (a,d)  $P(L_{i,j}|M_{i,j}) = 1.0$ , (b,e)  $P(L_{i,j}|M_{i,j}) = 0.75$ , and (c,f)  $P(L_{i,j}|M_{i,j}) = 0.50$ . Probabilities of regional interactions were obtained with a false positive rate of 5% and a false negative rate of 10%. Further details on the data and network construction are presented in the supplementary material. 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 binary interaction networks were generated by independently sampling binary interactions for each local network of probabilistic interactions.

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

is given by:

$$P(L_{0,i,j}) = 1 - [1 - P(L_{1,i,j})] \times [1 - P(L_{2,i,j})]. \quad (19)$$

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

407

## 408 Conclusion

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

## 422 Predicting local networks from metawebs

423 Metawebs are a valuable source of ecological information for predicting local networks across time and space.  
424 Local networks of binary interactions can be reconstructed by selecting a subset of taxa and interactions from  
425 the metaweb (Dunne 2006). Metawebs thus contain more interactions than local networks, even though their  
426 connectance is usually much smaller than the one of local networks (Gravel *et al.* 2011). Determining the list of

427 taxa to select can be achieved empirically (e.g., observed occurrence data for a site) or numerically (e.g., species  
428 distribution models). As species composition is arguably easier to sample or predict than pairwise interactions,  
429 the primary challenge lies in deciding which interactions to select from the metaweb. Inferring the structure of  
430 local networks from the metaweb before predicting local pairwise interactions could hold promise (Strydom *et  
431 al.* 2021), considering that the structure of local networks is constrained by the metaweb (Saravia *et al.* 2022).

432 Inferring local networks of probabilistic interactions from a metaweb comes with its own set of challenges. For  
433 example, Dansereau *et al.* (2023) inferred spatially-explicit food webs from a metaweb of probabilistic trophic  
434 interactions between Canadian mammals. Their predicted localized food webs are downscaled versions of the  
435 metaweb (i.e., localized metawebs with the same interaction probabilities as those in the regional metaweb). To  
436 infer local networks as defined in this manuscript (i.e., describing local realizations of interactions), local  
437 interaction probabilities must be smaller than regional interaction probabilities. This decrease is due to the  
438 prerequisite that two taxa must initially possess the capacity to interact before engaging in local interactions  
439 (Eq. 3). Inferring local networks from a metaweb by maintaining identical interaction probability values  
440 introduces systematic biases into the predictions, as discussed in Box 2 (unless networks are seen as downscaled  
441 metawebs).

442 As suggested by McLeod *et al.* (2021), although metawebs do not capture the spatiotemporal variability of  
443 interactions, they establish an upper limit for local interactions (similarly for metawebs of probabilistic  
444 interactions, Strydom *et al.* 2023). In other words, the probability that two taxa interact at a specific location  
445 and time is consistently lower or equal to the probability of their regional interaction, regardless of the  
446 conditional variables considered:

$$P(L_{i,j}|...) \leq P(M_{i,j}|T_i, T_j). \quad (20)$$

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

$$\int_{E_k} \int_A \int_t P(L_{i,j}|E_k, A, t) dt dA dE_k \leq P(M_{i,j}|T_i, T_j). \quad (21)$$

451 Estimating more precisely the probability  $P(L_{i,j}|M_{i,j})$  that two taxa interact locally if they can potentially  
452 interact allows for improved predictions of local networks from the metaweb of probabilistic interactions. This  
453 task is challenging due to the variability of this probability across space and time, as well as its variability  
454 across pairwise interactions within a network. Using simple models of  $P(L_{i,j}|M_{i,j})$ , as shown in Box 2,  
455 represents an initial step toward the overarching objective of reconstructing local networks from metawebs.

## 456 Quantifying and reducing interaction uncertainty

457 While sampling biological communities decreases the uncertainty of interactions by accumulating evidence for  
458 their feasibility and local realization, there is a limit to how much we can reduce uncertainty. In metawebs,  
459 probabilities reflect our limited knowledge of interactions (i.e., our degree of belief that interactions are  
460 feasible), which is expected to improve with a larger volume of data. Regional interactions should become more  
461 definitive (with probabilities approaching 0 or 1) as we investigate various conditions, including different  
462 combinations of species traits.

463 In comparison, although local networks can be seen as random instances of metawebs, their stochasticity cannot  
464 be reduced to the same extent. Local interaction probabilities may represent both their uncertainty and  
465 spatiotemporal variability. Owing to environmental heterogeneity, there will invariably be instances in which an  
466 interaction occurs and others in which it does not, across different times and locations, irrespective of the extent  
467 to which we can improve our knowledge of its biological feasibility and the local conditions that facilitate its  
468 occurrence. When local networks describe probabilities of observing interactions rather than their actual  
469 occurrence, we must also consider observation variability (sampling error) as an additional source of  
470 stochasticity. Every ecological process is stochastic but there is also a possibility that a phenomenon goes  
471 undetected. Quantifying and partitioning this stochasticity will enable us to make more accurate predictions  
472 about ecological interactions at various spatial and temporal scales. This will prove to be of vital importance as  
473 our time to understand nature runs out, especially at locations where the impacts of climate change and habitat  
474 loss hit harder.

## 475 Relaxing the independence assumption

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

478 these interaction probabilities assumes independence among interactions, a condition seldom respected in  
479 practice (Golubski & Abrams 2011). Relaxing this assumption is the next logical step in the stochastic  
480 representation of interactions.

481 A more accurate representation of the stochasticity of ecological networks involves creating *probabilistic*  
482 *networks* ( $P(L|...)$  and  $P(M|...)$ ), rather than networks of *probabilistic interactions* ( $P(L_{i,j}|...)$  and  $P(M_{i,j}|...)$ ).  
483 Probabilistic networks describe the probability that a particular network of binary (or quantitative) interactions  
484 (its whole adjacency matrix) is realized. For example, Young *et al.* (2021) used a Bayesian approach to estimate  
485 the probability of different plant-pollinator network structures derived from imperfect observational data. A  
486 probability distribution of ecological networks may also be derived using the principle of maximum entropy  
487 given structural constrained (e.g., Cimini *et al.* 2019; Park & Newman 2004).

488 Regardless of the method employed, generating probabilistic local networks could lead to more accurate  
489 predictions of local networks of binary interactions by bypassing the independence assumption. Probabilistic  
490 networks could serve as an alternative to null hypothesis significance testing when comparing the structure of a  
491 local network to some random expectations or, as done in Pellissier *et al.* (2018), to the metaweb. These random  
492 expectations are typically derived by performing a series of Bernoulli trials on probabilistic interactions,  
493 assuming independence, to generate a distribution of networks of binary interactions to calculate their structure  
494 (Poisot *et al.* 2016). One could for instance compare the likelihood of an observed network to the one of the  
495 most likely network structure (according to the probabilistic network distribution), thereby directly obtaining a  
496 measure of discrepancy of the empirical network. Generating probabilistic ecological networks represents a  
497 tangible challenge, one that, in the coming years, promises to unlock doors to more advanced and adequate  
498 analyses of ecological networks.

## 499 Acknowledgment

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

506 feedback on this manuscript.

507 **References**

- 508 Abrego, N., Roslin, T., Huotari, T., Ji, Y., Schmidt, N.M., Wang, J., *et al.* (2021). Accounting for species  
509 interactions is necessary for predicting how arctic arthropod communities respond to climate change.  
510 *Ecography*, 44, 885–896.
- 511 Angilletta, M.J., Jr., Steury, T.D. & Sears, M.W. (2004). Temperature, Growth Rate, and Body Size in  
512 Ectotherms: Fitting Pieces of a Life-History Puzzle<sup>1</sup>. *Integrative and Comparative Biology*, 44, 498–509.
- 513 Araujo, M.B., Rozenfeld, A., Rahbek, C. & Marquet, P.A. (2011). Using species co-occurrence networks to  
514 assess the impacts of climate change. *Ecography*, 34, 897–908.
- 515 Banašek-Richter, C., Cattin, M.-F. & Bersier, L.-F. (2004). Sampling effects and the robustness of quantitative  
516 and qualitative food-web descriptors. *Journal of Theoretical Biology*, 226, 23–32.
- 517 Bartomeus, I., Gravel, D., Tylianakis, J.M., Aizen, M.A., Dickie, I.A. & Bernard-Verdier, M. (2016). A  
518 common framework for identifying linkage rules across different types of interactions. *Functional Ecology*,  
519 30, 1894–1903.
- 520 Bascompte, J., Jordano, P., Melián, C.J. & Olesen, J.M. (2003). The nested assembly of plant–animal  
521 mutualistic networks. *Proceedings of the National Academy of Sciences*, 100, 9383–9387.
- 522 Berg, M.P. & Ellers, J. (2010). Trait plasticity in species interactions: A driving force of community dynamics.  
523 *Evolutionary Ecology*, 24, 617–629.
- 524 Berlow, E.L., Neutel, A.-M., Cohen, J.E., De Ruiter, P.C., Ebenman, B., Emmerson, M., *et al.* (2004).  
525 Interaction strengths in food webs: Issues and opportunities. *Journal of Animal Ecology*, 73, 585–598.
- 526 Bersier, L.-F., Dixon, P. & Sugihara, G. (1999). Scale-Invariant or Scale-Dependent Behavior of the Link  
527 Density Property in Food Webs: A Matter of Sampling Effort? *The American Naturalist*, 153, 676–682.
- 528 Bolnick, D.I., Amarasekare, P., Araújo, M.S., Bürger, R., Levine, J.M., Novak, M., *et al.* (2011). Why  
529 intraspecific trait variation matters in community ecology. *Trends in Ecology & Evolution*, 26, 183–192.
- 530 Borrett, S.R. & Scharler, U.M. (2019). Walk partitions of flow in Ecological Network Analysis: Review and  
531 synthesis of methods and indicators. *Ecological Indicators*, 106, 105451.

- 532 Boulangeat, I., Gravel, D. & Thuiller, W. (2012). Accounting for dispersal and biotic interactions to disentangle  
533 the drivers of species distributions and their abundances. *Ecology Letters*, 15, 584–593.
- 534 Brimacombe, C., Bodner, K., Michalska-Smith, M., Poisot, T. & Fortin, M.-J. (2023). Shortcomings of reusing  
535 species interaction networks created by different sets of researchers. *PLOS Biology*, 21, e3002068.
- 536 Broom, M. & Ruxton, G.D. (2005). You can run—or you can hide: Optimal strategies for cryptic prey against  
537 pursuit predators. *Behavioral Ecology*, 16, 534–540.
- 538 Brose, U., Ostling, A., Harrison, K. & Martinez, N.D. (2004). Unified spatial scaling of species and their  
539 trophic interactions. *Nature*, 428, 167–171.
- 540 Burnham, K.P. & Anderson, D.R. (2004). Multimodel Inference: Understanding AIC and BIC in Model  
541 Selection. *Sociological Methods & Research*, 33, 261–304.
- 542 Canard, E.F., Mouquet, N., Mouillot, D., Stanko, M., Miklisova, D. & Gravel, D. (2014). Empirical Evaluation  
543 of Neutral Interactions in Host-Parasite Networks. *The American Naturalist*, 183, 468–479.
- 544 Canard, E., Mouquet, N., Marescot, L., Gaston, K.J., Gravel, D. & Mouillot, D. (2012). Emergence of  
545 Structural Patterns in Neutral Trophic Networks. *PLOS ONE*, 7, e38295.
- 546 Carlson, C.J., Farrell, M.J., Grange, Z., Han, B.A., Mollentze, N., Phelan, A.L., et al. (2021). The future of  
547 zoonotic risk prediction. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 376,  
548 20200358.
- 549 Caron, D., Brose, U., Lurgi, M., Blanchet, F.G., Gravel, D. & Pollock, L.J. (2024). Trait-matching models  
550 predict pairwise interactions across regions, not food web properties. *Global Ecology and Biogeography*,  
551 33, e13807.
- 552 Castoe, T.A., Spencer, C.L. & Parkinson, C.L. (2007). Phylogeographic structure and historical demography of  
553 the western diamondback rattlesnake (*Crotalus Atrox*): A perspective on North American desert  
554 biogeography. *Molecular Phylogenetics and Evolution*, 42, 193–212.
- 555 Catchen, M.D., Poisot, T., Pollock, L.J. & Gonzalez, A. (2023). The missing link: Discerning true from false  
556 negatives when sampling species interaction networks.
- 557 Cazelles, K., Araujo, M.B., Mouquet, N. & Gravel, D. (2016). A theory for species co-occurrence in interaction  
558 networks. *Theoretical Ecology*, 9, 39–48.

- 559 Chagnon, P.-L. (2015). [Characterizing topology of ecological networks along gradients: The limits of metrics' standardization](#). *Ecological Complexity*, 22, 36–39.
- 560
- 561 Choh, Y., Ignacio, M., Sabelis, M.W. & Janssen, A. (2012). [Predator-prey role reversals, juvenile experience and adult antipredator behaviour](#). *Scientific Reports*, 2, 728.
- 562
- 563 Cimini, G., Squartini, T., Saracco, F., Garlaschelli, D., Gabrielli, A. & Caldarelli, G. (2019). [The statistical physics of real-world networks](#). *Nature Reviews Physics*, 1, 58–71.
- 564
- 565 Cirtwill, A.R., Eklöf, A., Roslin, T., Wootton, K. & Gravel, D. (2019). [A quantitative framework for investigating the reliability of empirical network construction](#). *Methods in Ecology and Evolution*, 10, 902–911.
- 566
- 567
- 568 Dallas, T., Park, A.W. & Drake, J.M. (2017). [Predicting cryptic links in host-parasite networks](#). *PLOS Computational Biology*, 13, e1005557.
- 569
- 570 Dansereau, G., Barros, C. & Poisot, T. (2023). Spatially explicit predictions of food web structure from regional level data.
- 571
- 572 Dunne, J.A. (2006). The Network Structure of Food Webs. In: *Ecological networks: Linking structure and dynamics* (eds. Dunne, J.A. & Pascual, M.). Oxford University Press, pp. 27–86.
- 573
- 574 Elmasri, M., Farrell, M.J., Davies, T.J. & Stephens, D.A. (2020). [A hierarchical Bayesian model for predicting ecological interactions using scaled evolutionary relationships](#). *The Annals of Applied Statistics*, 14, 221–240.
- 575
- 576
- 577 Elton, C.S. (2001). *Animal Ecology*. University of Chicago Press, Chicago, IL.
- 578
- 579 Emmerson, M.C. & Raffaelli, D. (2004). [Predator–prey body size, interaction strength and the stability of a real food web](#). *Journal of Animal Ecology*, 73, 399–409.
- 580
- 581 Fedorov, V.B., Goropashnaya, A.V., Boeskorov, G.G. & Cook, J.A. (2008). [Comparative phylogeography and demographic history of the wood lemming \(\*Myopus schisticolor\*\): Implications for late Quaternary history of the taiga species in Eurasia](#). *Molecular Ecology*, 17, 598–610.
- 582
- 583 Fortuna, M.A. & Bascompte, J. (2006). [Habitat loss and the structure of plant–animal mutualistic networks](#). *Ecology Letters*, 9, 281–286.
- 584
- 585 Fu, X., Seo, E., Clarke, J. & Hutchinson, R.A. (2021). [Link Prediction Under Imperfect Detection: Collaborative Filtering for Ecological Networks](#). *IEEE Transactions on Knowledge and Data Engineering*,
- 586

- 587 33, 3117–3128.
- 588 Galiana, N., Lurgi, M., Claramunt-López, B., Fortin, M.-J., Leroux, S., Cazelles, K., *et al.* (2018). [The spatial scaling of species interaction networks](#). *Nature Ecology & Evolution*, 2, 782–790.
- 590 Gilman, S.E., Urban, M.C., Tewksbury, J., Gilchrist, G.W. & Holt, R.D. (2010). [A framework for community interactions under climate change](#). *Trends in Ecology & Evolution*, 25, 325–331.
- 592 Godsoe, W., Murray, R. & Iritani, R. (2022). [Species interactions and diversity: A unified framework using Hill numbers](#). *Oikos*, n/a, e09282.
- 594 Golubski, A.J. & Abrams, P.A. (2011). [Modifying modifiers: What happens when interspecific interactions interact?](#) *Journal of Animal Ecology*, 80, 1097–1108.
- 596 Gomez, J.M., Verdu, M. & Perfectti, F. (2010). [Ecological interactions are evolutionarily conserved across the entire tree of life](#). *Nature*, 465, 918–U6.
- 598 Gonzalez, A. & Londoño, M.C. (2022). [Monitor biodiversity for action](#). *Science*, 378, 1147–1147.
- 599 Gonzalez-Varo, J.P. & Traveset, A. (2016). [The Labile Limits of Forbidden Interactions](#). *Trends in Ecology & Evolution*, 31, 700–710.
- 601 Gravel, D., Baiser, B., Dunne, J.A., Kopalke, J.-P., Martinez, N.D., Nyman, T., *et al.* (2019). [Bringing Elton and Grinnell together: A quantitative framework to represent the biogeography of ecological interaction networks](#). *Ecography*, 42, 401–415.
- 604 Gravel, D., Massol, F., Canard, E., Mouillot, D. & Mouquet, N. (2011). [Trophic theory of island biogeography](#). *Ecology Letters*, 14, 1010–1016.
- 606 Gravel, D., Poisot, T., Albouy, C., Velez, L. & Mouillot, D. (2013). [Inferring food web structure from predator–prey body size relationships](#). *Methods in Ecology and Evolution*, 4, 1083–1090.
- 608 Guimarães, P.R. (2020). [The Structure of Ecological Networks Across Levels of Organization](#). *Annual Review of Ecology, Evolution, and Systematics*, 51, 433–460.
- 610 Guimerà, R. & Sales-Pardo, M. (2009). [Missing and spurious interactions and the reconstruction of complex networks](#). *Proceedings of the National Academy of Sciences*, 106, 22073–22078.
- 612 Hemprich-Bennett, D.R., Oliveira, H.F.M., Le Comber, S.C., Rossiter, S.J. & Clare, E.L. (2021). [Assessing the impact of taxon resolution on network structure](#). *Ecology*, 102, e03256.

- 614 Herrera, C.M. (1989). [Pollinator abundance, morphology, and flower visitation rate: Analysis of the “quantity”](#)  
615 [component in a plant-pollinator system.](#) *Oecologia*, 80, 241–248.
- 616 Hortal, J., de Bello, F., Diniz-Filho, J.A.F., Lewinsohn, T.M., Lobo, J.M. & Ladle, R.J. (2015). [Seven Shortfalls](#)  
617 [that Beset Large-Scale Knowledge of Biodiversity.](#) *Annual Review of Ecology, Evolution, and Systematics*,  
618 46, 523–549.
- 619 Jordano, P. (1987). [Patterns of Mutualistic Interactions in Pollination and Seed Dispersal: Connectance,](#)  
620 [Dependence Asymmetries, and Coevolution.](#) *The American Naturalist*, 129, 657–677.
- 621 Jordano, P. (2016). [Sampling networks of ecological interactions.](#) *Functional Ecology*, 30, 1883–1893.
- 622 Jordano, P., Bascompte, J. & Olesen, J.M. (2003). [Invariant properties in coevolutionary networks of](#)  
623 [plant–animal interactions.](#) *Ecology Letters*, 6, 69–81.
- 624 Kausrud, K.L., Mysterud, A., Steen, H., Vik, J.O., Østbye, E., Cazelles, B., *et al.* (2008). [Linking climate](#)  
625 [change to lemming cycles.](#) *Nature*, 456, 93–97.
- 626 Kéfi, S., Berlow, E.L., Wieters, E.A., Joppa, L.N., Wood, S.A., Brose, U., *et al.* (2015). [Network structure](#)  
627 [beyond food webs: Mapping non-trophic and trophic interactions on Chilean rocky shores.](#) *Ecology*, 96,  
628 291–303.
- 629 Kéfi, S., Berlow, E.L., Wieters, E.A., Navarrete, S.A., Petchey, O.L., Wood, S.A., *et al.* (2012). [More than a](#)  
630 [meal... integrating non-feeding interactions into food webs.](#) *Ecology Letters*, 15, 291–300.
- 631 Kéfi, S., Miele, V., Wieters, E.A., Navarrete, S.A. & Berlow, E.L. (2016). [How Structured Is the Entangled](#)  
632 [Bank? The Surprisingly Simple Organization of Multiplex Ecological Networks Leads to Increased](#)  
633 [Persistence and Resilience.](#) *PLOS Biology*, 14, e1002527.
- 634 Kissner, K.J., Forbes, M.R. & Secoy, D.M. (1997). [Rattling Behavior of Prairie Rattlesnakes \(\*Crotalus viridis\*](#)  
635 [viridis, Viperidae\) in Relation to Sex, Reproductive Status, Body Size, and Body Temperature.](#) *Ethology*,  
636 103, 1042–1050.
- 637 Klecka, J. & Boukal, D.S. (2014). [The effect of habitat structure on prey mortality depends on predator and](#)  
638 [prey microhabitat use.](#) *Oecologia*, 176, 183–191.
- 639 Kopelke, J.-P., Nyman, T., Cazelles, K., Gravel, D., Vissault, S. & Roslin, T. (2017). [Food-web structure of](#)  
640 [willow-galling sawflies and their natural enemies across Europe.](#) *Ecology*, 98, 1730–1730.

- 641 Kulkarni, M.A., Kweka, E., Nyale, E., Lyatuu, E., Mosha, F.W., Chandramohan, D., *et al.* (2006).  
642 [Entomological Evaluation of Malaria Vectors at Different Altitudes in Hai District, Northeastern Tanzania.](#)  
643 *Journal of Medical Entomology*, 43, 580–588.
- 644 Maiorano, L., Montemaggiori, A., Ficetola, G.F., O'Connor, L. & Thuiller, W. (2020). [TETRA-EU 1.0: A](#)  
645 [species-level trophic metaweb of European tetrapods](#). *Global Ecology and Biogeography*, 29, 1452–1457.
- 646 McCann, K. (2007). Protecting biostructure. *Nature*, 446, 29–29.
- 647 McCann, K.S. (2011). Food webs (MPB-50). In: *Food Webs (MPB-50)*. Princeton University Press.
- 648 McLeod, A., Leroux, S.J., Gravel, D., Chu, C., Cirtwill, A.R., Fortin, M.-J., *et al.* (2021). [Sampling and](#)  
649 [asymptotic network properties of spatial multi-trophic networks](#). *Oikos*, n/a.
- 650 McLeod, A.M., Leroux, S.J. & Chu, C. (2020). [Effects of species traits, motif profiles, and environment on](#)  
651 [spatial variation in multi-trophic antagonistic networks](#). *Ecosphere*, 11, e03018.
- 652 Melián, C.J., Vilas, C., Baldó, F., González-Ortegón, E., Drake, P. & Williams, R.J. (2011). [Eco-evolutionary](#)  
653 [Dynamics of Individual-Based Food Webs](#). In: *Advances in Ecological Research*, The Role of Body Size in  
654 Multispecies Systems (ed. Belgrano, A.). Academic Press, pp. 225–268.
- 655 Morales-Castilla, I., Matias, M.G., Gravel, D. & Araújo, M.B. (2015). [Inferring biotic interactions from](#)  
656 [proxies](#). *Trends in Ecology & Evolution*, 30, 347–356.
- 657 Ohlmann, M., Miele, V., Dray, S., Chalmandrier, L., O'Connor, L. & Thuiller, W. (2019). [Diversity indices for](#)  
658 [ecological networks: A unifying framework using Hill numbers](#). *Ecology Letters*, 22, 737–747.
- 659 Olesen, J.M., Bascompte, J., Dupont, Y.L., Elberling, H., Rasmussen, C. & Jordano, P. (2010). [Missing and](#)  
660 [forbidden links in mutualistic networks](#). *Proceedings of the Royal Society B: Biological Sciences*, 278,  
661 725–732.
- 662 Park, J. & Newman, M.E.J. (2004). [Statistical mechanics of networks](#). *Physical Review E*, 70, 066117.
- 663 Pascual, M., Dunne, J.A. & Dunne, J.A. (2006). *Ecological Networks: Linking Structure to Dynamics in Food*  
664 *Webs*. Oxford University Press, USA.
- 665 Pellissier, L., Albouy, C., Bascompte, J., Farwig, N., Graham, C., Loreau, M., *et al.* (2018). [Comparing species](#)  
666 [interaction networks along environmental gradients](#). *Biological Reviews*, 93, 785–800.

- 667 Petchey, O.L., Beckerman, A.P., Riede, J.O. & Warren, P.H. (2008). [Size](#), foraging, and food web structure.  
668 *Proceedings of the National Academy of Sciences*, 105, 4191–4196.
- 669 Pilosof, S., Porter, M.A., Pascual, M. & Kéfi, S. (2017). [The multilayer nature of ecological networks](#). *Nature  
670 Ecology & Evolution*, 1, 1–9.
- 671 Plagányi, É.E. & Butterworth, D.S. (2004). [A critical look at the potential of Ecopath with ecosim to assist in  
672 practical fisheries management](#). *African Journal of Marine Science*, 26, 261–287.
- 673 Poisot, T., Canard, E., Mouillot, D., Mouquet, N. & Gravel, D. (2012). [The dissimilarity of species interaction  
674 networks](#). *Ecology Letters*, 15, 1353–1361.
- 675 Poisot, T., Cirtwill, A.R., Cazelles, K., Gravel, D., Fortin, M.-J. & Stouffer, D.B. (2016). [The structure of  
676 probabilistic networks](#). *Methods in Ecology and Evolution*, 7, 303–312.
- 677 Poisot, T. & Gravel, D. (2014). [When is an ecological network complex? Connectance drives degree  
678 distribution and emerging network properties](#). *PeerJ*, 2, e251.
- 679 Poisot, T., Stouffer, D.B. & Gravel, D. (2015). [Beyond species: Why ecological interaction networks vary  
680 through space and time](#). *Oikos*, 124, 243–251.
- 681 Polis, G.A. (1991). [Complex Trophic Interactions in Deserts: An Empirical Critique of Food-Web Theory](#). *The  
682 American Naturalist*, 138, 123–155.
- 683 Pollock, L.J., Tingley, R., Morris, W.K., Golding, N., O'Hara, R.B., Parris, K.M., *et al.* (2014). [Understanding  
684 co-occurrence bymodelling species simultaneously with a Joint Species DistributionModel \(JSDM\)](#).  
685 *Methods in Ecology and Evolution*, 5, 397–406.
- 686 Pomeranz, J.P.F., Thompson, R.M., Poisot, T. & Harding, J.S. (2019). [Inferring predator–prey interactions in  
687 food webs](#). *Methods in Ecology and Evolution*, 10, 356–367.
- 688 Pringle, R.M. & Hutchinson, M.C. (2020). [Resolving Food-Web Structure](#). *Annual Review of Ecology,  
689 Evolution, and Systematics*, 51, 55–80.
- 690 Proulx, S.R., Promislow, D.E.L. & Phillips, P.C. (2005). [Network thinking in ecology and evolution](#). *Trends in  
691 Ecology & Evolution*, SPECIAL ISSUE: BUMPER BOOK REVIEW, 20, 345–353.
- 692 Pulliam, H.R. (1974). [On the Theory of Optimal Diets](#). *The American Naturalist*, 108, 59–74.

- 693 Rooney, N. & McCann, K.S. (2012). Integrating food web diversity, structure and stability. *Trends in Ecology and Evolution*, 27, 40–46.
- 695 Salim, J.A., Saraiva, A.M., Zermoglio, P.F., Agostini, K., Wolowski, M., Drucker, D.P., et al. (2022). Data  
696 standardization of plant–pollinator interactions. *GigaScience*, 11, giac043.
- 697 Sanders, D. & van Veen, F.J.F. (2012). Indirect commensalism promotes persistence of secondary consumer  
698 species. *Biology Letters*, 8, 960–963.
- 699 Saravia, L.A., Marina, T.I., Kristensen, N.P., De Troch, M. & Momo, F.R. (2022). Ecological network  
700 assembly: How the regional metaweb influences local food webs. *Journal of Animal Ecology*, 91, 630–642.
- 701 Schmeller, D.S., Julliard, R., Bellingham, P.J., Böhm, M., Brummitt, N., Chiarucci, A., et al. (2015). Towards a  
702 global terrestrial species monitoring program. *Journal for Nature Conservation*, 25, 51–57.
- 703 Shaw, J.O., Dunhill, A.M., Beckerman, A.P., Dunne, J.A. & Hull, P.M. (2024). A framework for reconstructing  
704 ancient food webs using functional trait data.
- 705 Singer, M.C. & McBride, C.S. (2012). Geographic mosaics of species' association: A definition and an  
706 example driven by plant–insect phenological synchrony. *Ecology*, 93, 2658–2673.
- 707 Staniczenko, P.P.A., Sivasubramaniam, P., Suttle, K.B. & Pearson, R.G. (2017). Linking macroecology and  
708 community ecology: Refining predictions of species distributions using biotic interaction networks.  
709 *Ecology Letters*, 20, 693–707.
- 710 Stock, M., Poisot, T., Waegeman, W. & De Baets, B. (2017). Linear filtering reveals false negatives in species  
711 interaction data. *SCIENTIFIC REPORTS*, 7, 45908.
- 712 Stouffer, D.B., Rezende, E.L. & Amaral, L.A.N. (2011). The role of body mass in diet contiguity and food-web  
713 structure. *Journal of Animal Ecology*, 80, 632–639.
- 714 Strydom, T., Bouskila, S., Banville, F., Barros, C., Caron, D., Farrell, M.J., et al. (2022). Food web  
715 reconstruction through phylogenetic transfer of low-rank network representation. *Methods in Ecology and  
716 Evolution*, 13.
- 717 Strydom, T., Bouskila, S., Banville, F., Barros, C., Caron, D., Farrell, M.J., et al. (2023). Graph embedding and  
718 transfer learning can help predict potential species interaction networks despite data limitations. *Methods in  
719 Ecology and Evolution*, 14, 2917–2930.

- 720 Strydom, T., Catchen, M.D., Banville, F., Caron, D., Dansereau, G., Desjardins-Proulx, P., *et al.* (2021). **A**  
721 **roadmap towards predicting species interaction networks (across space and time).** *Philosophical*  
722 *Transactions of the Royal Society B-Biological Sciences*, 376, 20210063.
- 723 Thompson, R.M. & Townsend, C.R. (2005). **Food-Web Topology Varies with Spatial Scale in a Patchy**  
724 **Environment.** *Ecology*, 86, 1916–1925.
- 725 Tylianakis, J.M., Tscharntke, T. & Lewis, O.T. (2007). **Habitat modification alters the structure of tropical**  
726 **host-parasitoid food webs.** *Nature*, 445, 202–205.
- 727 Vázquez, D.P., Melián, C.J., Williams, N.M., Blüthgen, N., Krasnov, B.R. & Poulin, R. (2007). **Species**  
728 **abundance and asymmetric interaction strength in ecological networks.** *Oikos*, 116, 1120–1127.
- 729 Vázquez, D.P., Peralta, G., Cagnolo, L. & Santos, M. (2022). **Ecological interaction networks. What we know,**  
730 **what we don't, and why it matters.** *Ecología Austral*, 32, 670–697.
- 731 Wagenmakers, E.-J. & Farrell, S. (2004). **AIC model selection using Akaike weights.** *Psychonomic Bulletin &*  
732 *Review*, 11, 192–196.
- 733 Weinstein, B.G. & Graham, C.H. (2017a). **On comparing traits and abundance for predicting species**  
734 **interactions with imperfect detection.** *Food Webs*, 11, 17–25.
- 735 Weinstein, B.G. & Graham, C.H. (2017b). **Persistent bill and corolla matching despite shifting temporal**  
736 **resources in tropical hummingbird-plant interactions.** *Ecology Letters*, 20, 326–335.
- 737 Whittaker, R.H. (1960). **Vegetation of the Siskiyou Mountains, Oregon and California.** *Ecological*  
738 *Monographs*, 30, 279–338.
- 739 Wood, S.A., Russell, R., Hanson, D., Williams, R.J. & Dunne, J.A. (2015). **Effects of spatial scale of sampling**  
740 **on food web structure.** *Ecology and Evolution*, 5, 3769–3782.
- 741 Woodward, G., Benstead, J.P., Beveridge, O.S., Blanchard, J., Brey, T., Brown, L.E., *et al.* (2010). **Chapter 2 -**  
742 **Ecological Networks in a Changing Climate.** In: *Advances in Ecological Research*, Ecological Networks  
743 (ed. Woodward, G.). Academic Press, pp. 71–138.
- 744 Woodward, G., Brown, L.E., Edwards, F.K., Hudson, L.N., Milner, A.M., Reuman, D.C., *et al.* (2012). **Climate**  
745 **change impacts in multispecies systems: Drought alters food web size structure in a field experiment.**  
746 *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367, 2990–2997.

- 747 Young, J.-G., Valdovinos, F.S. & Newman, M.E.J. (2021). [Reconstruction of plant–pollinator networks from](#)  
748 [observational data](#). *Nature Communications*, 12, 3911.