

# 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 or ecological feasibility (regional). Using host-parasite interactions in Europe, we illustrate how these two network representations differ in their statistical properties, specifically: how local networks and metawebs differ in their spatial and temporal scaling of probabilistic interactions, but not in their taxonomic scaling. We present two approaches to inferring binary interactions from probabilistic ones that account for these differences and show that systematic biases arise when directly inferring local networks from metawebs. Our results underscore the importance of more rigorous descriptions of probabilistic species interaction networks that specify their type of interaction (local or regional) and uncertainty sources.

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



Last revision: June 9, 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 in neutral models (Canard *et al.* 2012; Canard *et al.*  
18 2014). The probability that species meet also depends on their biology, such as their phenology (Olesen *et al.*  
19 2010; Singer & McBride 2012) and discoverability (Broom & Ruxton 2005). Finally, when species do come  
20 into contact, an interaction occurs only if their traits, such as their phenotypes (Bolnick *et al.* 2011; Gravel *et al.*  
21 2013; Stouffer *et al.* 2011) and behavior (Choh *et al.* 2012; Pulliam 1974), are locally compatible in that  
22 specific environment (Poisot *et al.* 2015).

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

30 to the intricate nature of the system or simply by chance. If it does occur, it might go undetected, particularly if  
31 it happens infrequently. In this context, it is unsurprising that our knowledge of ecological interactions remains  
32 limited (Hortal *et al.* 2015) despite extensive biodiversity data collection (Schmeller *et al.* 2015).

33 **Species interactions as probabilistic objects**

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

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

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

## 82 We lack clear definitions of probabilistic species interactions

83 Yet, a precise definition of probabilistic interactions appears to be lacking, making the estimation and use of  
84 these data more difficult. In this manuscript, we aim to take a step back by outlining different ways in which  
85 probabilistic interactions are defined and used in network ecology. We distinguish two broad categories of  
86 probabilistic interaction networks that necessitate distinct approaches: local networks describing probabilities  
87 of realized interactions, and regional networks (metawebs) describing probabilities of potential interactions. We

88 highlight the distinctions in the ecological meaning of these two representations and examine their properties  
89 and relationships (particularly with space, time, and between each other).

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

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

### 108 **Probabilistic interactions have different interpretations**

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

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

122 An important aspect to consider when using interaction probabilities is knowing if they describe potential or  
123 realized interactions, as these two types of interactions have distinct conceptual underpinnings and sources of  
124 uncertainty. A potential (regional) interaction is defined as the biological capacity of two taxa to interact (i.e.,  
125 the probability that they interact if they were to encounter each other, given sufficient time and appropriate  
126 environmental conditions) whereas a realized (local) interaction is the occurrence or observation of this  
127 interaction in a well-defined space and time (i.e., the probability that they interact locally). For two co-occurring  
128 taxa and over enough time, the probability of local interaction tends toward the probability of regional  
129 (potential) interaction. A longer duration increases the probability that species will eventually encounter each  
130 other and that local environmental conditions supporting an interaction will occur, provided that species have  
131 the biological capacity to interact.

132 We use the terms *metaweb* (Dunne 2006) to designate regional networks of potential interactions and *local*  
133 *networks* (Poisot *et al.* 2012) for those of realized interactions. Metawebs are the network analogs of the species  
134 pool, where local networks originate from a subset of both species (nodes) and interactions (edges) of the  
135 regional metaweb (Saravia *et al.* 2022). When uncertain, potential interactions in a metaweb may be  
136 probabilistic (Strydom *et al.* 2023). Without clear documentation, it can be challenging to know if published  
137 probabilistic interaction networks describe local or regional interactions, or if so-called probabilities are in  
138 reality a form of interaction score. When probabilistic regional interactions are used and interpreted incorrectly  
139 as local interactions (and conversely), this may generate misleading findings during data analysis. A better  
140 understanding of probabilistic local and regional interactions would prevent interpretation errors (e.g., when  
141 studying network-area relationships with metawebs or local networks) and facilitate a more adequate use of  
142 interaction data.

143 **The outcome of probabilistic interactions is usually binary**

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

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

162 The variability of an interaction determines the fraction of networks in which it occurs. This can be measured  
163 by using a Binomial distribution, assuming a constant interaction probability and independence between  
164 interactions in different networks (trials). When considering uncertainties around the estimation of  $P(B_{i,j})$  or  
165 the variability of interaction probabilities, a Beta distribution may be used to represent the relative likelihood of  
166 different probability values. For example, when calculating the probability of interaction between two taxa  
167 based on their local abundances, any spatiotemporal fluctuations in their abundances would introduce  
168 variability in the interaction probability at the local scale. If we take into account the stochasticity of the  
169 interaction probability, a Beta-Binomial distribution can be used to predict the number of networks in which the  
170 interaction occurs. Empirically observing an interaction between two taxa at a given location and time provides  
171 important information that can be used to update previous estimates of  $P(B_{i,j})$ , informing us on the biological

172 capacity of both taxa to interact and the environmental conditions that enabled them to interact locally. By  
173 sampling more interactions, we can estimate their local variability more precisely.

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

175 Even though binary interaction networks constitute a highly valuable source of ecological information (Pascual  
176 *et al.* 2006), they overlook interaction strengths. Represented in a quantitative adjacency matrix  $W$ , interaction  
177 strengths better describe the energy flows, demographic impacts or frequencies of interactions between nodes  
178 (Berlow *et al.* 2004; Borrett & Scharler 2019), with  $W_{i,j}$  being a natural  $\mathbb{N}$  or real  $\mathbb{R}$  number depending on the  
179 measure. For example, they may represent local interaction rates between pairs of taxa (e.g., the flower-visiting  
180 rates of pollinators in a mutualistic network, Herrera 1989). When interaction strengths characterize predation  
181 pressure on prey, they can serve as parameters in a Lotka-Volterra model (e.g., Emmerson & Raffaelli 2004).  
182 The extra amount of ecological information in quantitative networks typically comes at a cost of greater  
183 sampling effort and data volume (Strydom *et al.* 2021), especially when using predictive models that quantify  
184 the uncertainty and variability of quantitative interactions (Berlow *et al.* 2004).

185 Like binary interaction networks, the uncertainty and variability of interaction strengths can be represented  
186 probabilistically. Interaction strengths can follow many probability distributions depending on the measure. For  
187 instance, they can follow a Poisson distribution  $W_{i,j} \sim \text{Poisson}(\lambda_{i,j}t_0)$  when predicting the number of  
188 interactions between individuals, with  $\lambda_{i,j}$  being the expected rate at which individuals of taxa  $i$  and  $j$  interact.  
189 The product  $\lambda_{i,j}t_0$  is the expected number of interactions realized by individuals during the time interval  $t_0$   
190 (e.g., the expected number of prey  $j$  consumed by all predators  $i$ ). The Poisson distribution can also be  
191 0-inflated when taking into account non-interacting taxa (e.g., Boulangeat *et al.* 2012 employ a 0-inflated model  
192 to analyze species abundance following the modeling of species presence and absence), which constitute the  
193 majority of taxa pairs in most local networks (Jordano 2016).

194 Because of the methodological difficulties typically encountered when building deterministic quantitative  
195 networks, binary interaction networks, which are easier to sample (Jordano 2016) and predict (Strydom *et al.*  
196 2021), have been more frequently studied and used. Mathematical models such as Ecopath (Plagányi &  
197 Butterworth 2004) partially mitigate these difficulties, but the number of biological parameters required to make  
198 predictions hinders their application in many systems. Moreover, most published probabilistic interaction  
199 networks (e.g., Strydom *et al.* 2022) and methods (e.g., Poisot *et al.* 2016) involve probabilistic interactions

200 whose outcome is binary. This underlines the need for better guidelines on the interpretation and manipulation  
201 of such probabilistic interactions, to ensure the appropriate use of these networks and methods. For these  
202 reasons, our primary focus is on interpreting interaction probabilities that dictate the presence or absence of  
203 interactions, in both local networks and metawebs.

## 204 **Local networks: communities interacting in space and time**

### 205 **Species interactions occur in a local context**

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

### 210 **Space and time**

211 We define space as the collection of geographic coordinates  $(x, y, z)$ , with  $(x, y)$  representing longitude and  
212 latitude coordinates, and  $z$  denoting either altitudes or depths. These point coordinates delineate the spatial  
213 boundaries of the system, which may be portrayed as a polyhedron. Ecological interactions may exhibit  
214 variations along latitudinal and altitudinal gradients, as evidenced by changes in hummingbird-plant  
215 interactions (Weinstein & Graham 2017a, b) and mosquito biting rates (e.g., Kulkarni *et al.* 2006) at different  
216 elevations. In contrast, time is defined as the specific time period within which interactions were either observed  
217 or predicted. Even though space and time are continuous variables that should yield probability *densities* of  
218 interactions (i.e., relative likelihoods of interactions occurring at infinitesimal locations and instants in time),  
219 these definitions enable them to be conceptualized as distinct patches and time segments. Treating space and  
220 time as discrete dimensions aligns with the common sampling methods of ecological networks and provides  
221 probabilities of interactions, which can be obtained by integrating probability densities over space and time. We  
222 can quantify both an area  $A_0$  and a duration  $t_0$ , which can readily be used in spatiotemporal analyses of  
223 ecological networks (Box 1).

224 **Co-occurrence**

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

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

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

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

239 **Biological and environmental factors**

240 Local interactions may also be conditioned on local environmental factors such as temperature (Angilletta *et al.*  
241 2004), precipitation (Woodward *et al.* 2012), habitat structure (Klecka & Boukal 2014), and presence of other  
242 taxa in the network (Kéfi *et al.* 2012; Pilosof *et al.* 2017). We use the variable  $E_k$  to describe the local  
243 environmental context in which interaction probabilities were estimated. For example, in a mesocosm  
244 experiment estimating interaction probabilities between predators and prey with and without shelters,  $E_k$  would  
245 represent the presence or absence of these shelters. Like co-occurrence,  $E_k$  can also be modeled  
246 probabilistically when the stochasticity or uncertainty of environmental factors is considered.  $E_k$  represents all

247 environmental variables that were taken into consideration when measuring interaction probabilities; it is a  
248 subset of all environmental factors acting on ecological interactions.

249 Other important factors that can impact interaction probabilities at the local scale are taxa local abundances  $N_{i,k}$   
250 and  $N_{j,k}$ , which affect encounter probabilities (Canard *et al.* 2012), and local traits distributions  $T_{i,k}$  and  $T_{j,k}$   
251 (e.g., body mass, longevity, and habitat breadth, Caron *et al.* 2024), which determine the ability of individuals  
252 to interact after encountering each other (Poisot *et al.* 2015). Local interaction probabilities may also be  
253 conditioned on higher-level properties of the network, which we denote by  $f(L)$ . Many topological null models  
254 (i.e., statistical models that randomize interactions by retaining certain properties of the network while  
255 excluding others) provide interaction probabilities from selected measures of network structure, such as  
256 connectance (Fortuna & Bascompte 2006) and the degree distribution (Bascompte *et al.* 2003).

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

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

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

262 Low values of  $P(L_{i,j,k}|M_{i,j})$  indicate that feasible interactions rarely occur locally, intermediate values around  
263 50% suggest considerable spatiotemporal variability, while high values indicate that regional interactions are  
264 nearly always realized locally. The local probability of interaction between a given pair of taxa is always equal  
265 to or below their probability of regional interaction.

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

267 The probability that two taxa  $i$  and  $j$  interact in a local network  $L_k$  can thus be conditioned on their  
268 co-occurrence  $X_{i,j,k}$  (or more explicitly on their occurrences  $X_{i,k}$  and  $X_{j,k}$ ), local abundances  $N_{i,k}$  and  $N_{j,k}$ , local  
269 traits distributions  $T_{i,k}$  and  $T_{j,k}$ , local environmental conditions  $E_k$ , network area (or volume)  $A_0$ , time interval  
270  $t_0$ , network properties  $f(L_k)$ , and biological feasibility  $M_{i,j}$ . The local probability of interaction is described by

271 the following expression when all of these conditional variables are included:

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

272 Although these variables correspond to distinct ecological inquiries or mechanisms related to ecological  
273 interactions, they may covary with each other, such as the possible dependence of  $X_{i,j,k}$  and  $E_k$  on spatial and  
274 temporal scales. When estimating interaction probabilities using e.g. a generalized linear model with multiple  
275 explanatory variables that might not be independent, it may become important to address collinearity. In such a  
276 case, to mitigate this issue, it may be necessary to use variable selection techniques before fitting the model to  
277 data.

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

290 When accounted for, conditional variables should be clearly described in the documentation of the data  
291 (Brimacombe *et al.* 2023), preferentially in mathematical terms to avoid any confusion in their interpretation  
292 and to limit manipulation errors during their re-use. For instance, ecologists should be explicit about their  
293 consideration ( $P(L_{i,j,k}|X_{i,j,k})$ ) or not ( $P(L_{i,j,k})$ ) of co-occurrence in their estimation of local interaction  
294 probabilities, as this can change our interpretation of the data. Understanding the scale and level of aggregation  
295 of the data enables us to more accurately study the underlying ecological processes (Clark *et al.* 2011). In  
296 Tbl. 1, we present examples of studies that used different expressions of probabilistic interactions. We included  
297 in this table the probability of empirically observing an interaction that is realized locally  $P(O_{i,j,k}|L_{i,j,k})$  to

298 underscore the distinction between local observations and actual realizations of interactions.

**Table 1: Mathematical expressions of probabilistic interactions.** The probability of interaction between two taxa  $i$  and  $j$  is interpreted differently in a local network  $L$  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). The boxes in this study that discussed them are also specified. 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
$P(L_{i,j,k} X_{i,k}, X_{j,k}, \dots)$	local	realization of the interaction given taxa co-occurrence	species distribution models	Gravel <i>et al.</i> (2019), Dansereau <i>et al.</i> (2023), Boxes 1 and 5
$P(L_{i,j,k} N_{i,k}, N_{j,k}, \dots)$	local	realization of the interaction given taxa abundances	neutral models	Canard <i>et al.</i> (2014)
$P(L_{i,j,k} T_{i,k}, T_{j,k}, \dots)$	local	realization of the interaction given local traits	trait matching models	Caron <i>et al.</i> (2024), Box 4
$P(L_{i,j,k} E_k, \dots)$	local	realization of the interaction given local environmental conditions	environmental- based models	Gravel <i>et al.</i> (2019) (temperature and precipitation)
$P(L_{i,j,k} A_0, \dots)$	local	realization of the interaction in a given area or volume	spatial models	Galiana <i>et al.</i> (2018) *, Box 3
$P(L_{i,j,k} t_0, \dots)$	local	realization of the interaction during a given time period	temporal models	Weinstein & Graham (2017a), Boxes 1 and 3
$P(L_{i,j,k} f(L_k), \dots)$	local	realization of the interaction given network structure	topological models	Fortuna & Bascompte (2006) (connectance), Stock <i>et al.</i> (2017)
$P(L_{i,j,k} M_{i,j}, \dots)$	local	realization of the interaction given that the taxa can biologically interact	spatiotemporal models	Dansereau <i>et al.</i> (2023), Boxes 2, 3, and 5

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

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

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

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

306 where the ellipsis serves as a placeholder for the conditional variables incorporated in these models. Model  
 307 structure constitutes another source of uncertainty that needs to be documented and quantified when estimating  
 308 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:

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

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

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

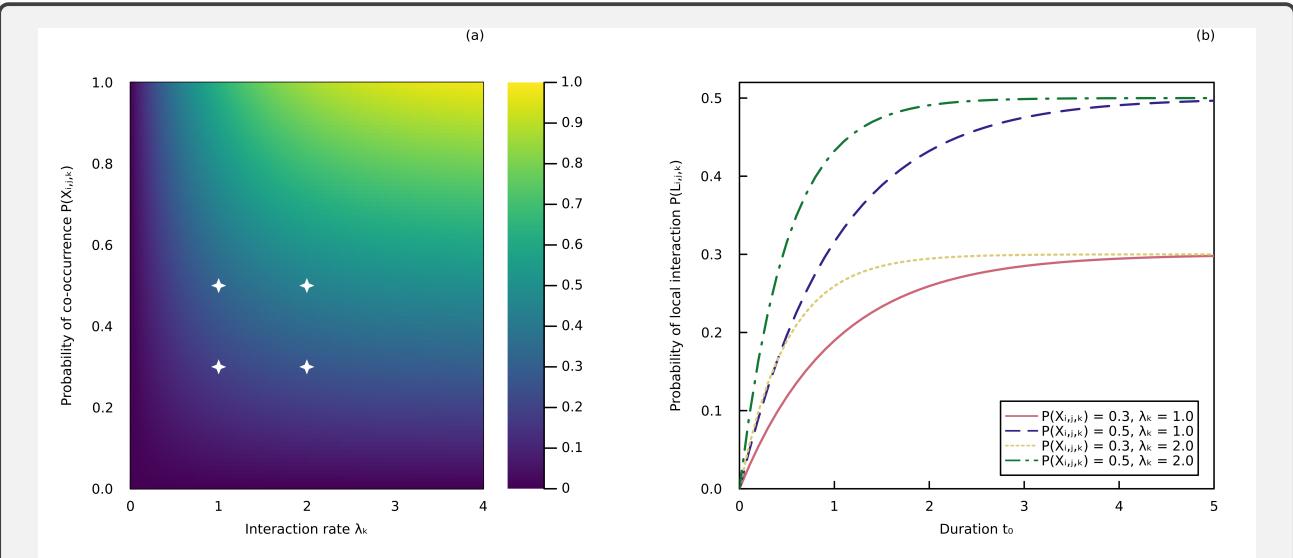
$$L_{i,j,k} \sim \text{Bernoulli}(P(L_{i,j,k})) \quad (7)$$

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

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

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

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



**Figure 1: Parameters of the spatiotemporally explicit model of interactions.** (a) Probability of local interaction given by the process model (Eq. 8) under different values of  $\lambda_k$  (interaction rate) and  $P(X_{i,j,k})$  (probability of co-occurrence), with  $t_0 = 1$  (duration). Parameters  $t_0$  and  $\lambda_k$  have complementary units (e.g.,  $t_0$  in months and  $\lambda_k$  in number of interactions per month). The parameter values used in the right panel are denoted by the white stars. (b) Scaling of the probability of interaction with the duration parameter  $t_0$ , for different values of  $\lambda_k$  and  $P(X_{i,j,k})$ .

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

311

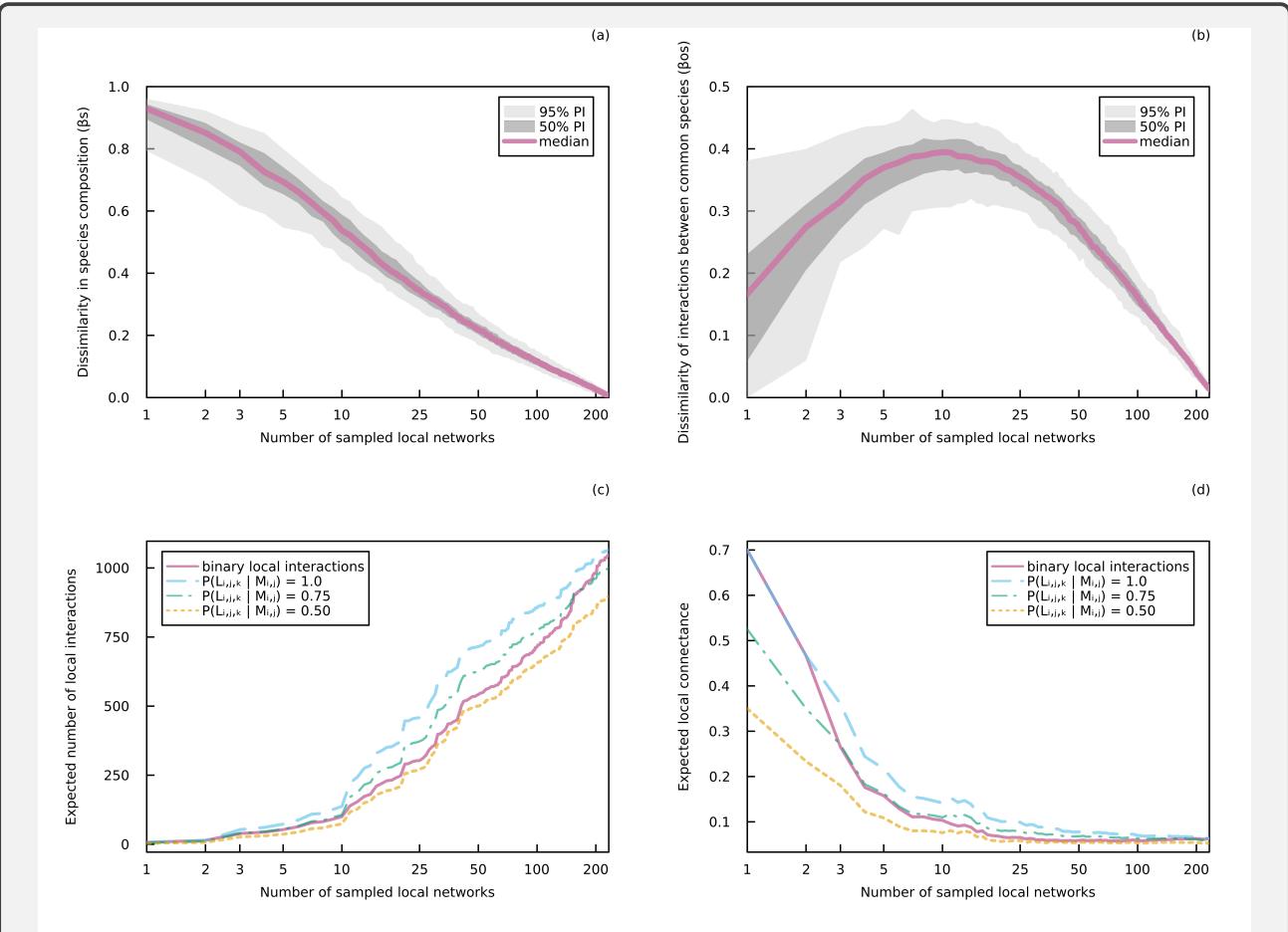
## Box 2: Dissimilarity of local host-parasite networks

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

In Fig. 2a-b, we show how the dissimilarity between the metaweb of binary interactions and aggregated local networks changes with the number of sampled local networks. We compared the metaweb and the aggregated local networks using the dissimilarity in species composition ( $\beta_S$ , Fig. 2a) and the dissimilarity of interactions between common species ( $\beta_{OS}$ , Fig. 2b) indices (Poisot *et al.* 2012). Expectingly, local

312

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 few species from the metaweb (species pool) occur locally. Moreover, we observe a peak in the dissimilarity of interactions between common species at intermediate sampling levels. This suggests that species are collected faster than their interactions. With a limited number of sampled local networks, few regional interactions are observed locally. Adding more sites brings new species, but not always their interactions. Quadratic relationships of network properties with sampling effort were also observed by McLeod *et al.* (2021).



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

Next, we investigate how the number of local interactions and connectance scale with the number of sampled (aggregated) local networks of probabilistic or binary interactions (Fig. 2c-d). By comparing the scaling relationships observed in local networks of binary and probabilistic interactions, we observe that high values of  $P(L_{i,j,k} | M_{i,j})$  lead to systematic overestimations in the number of interactions and connectance, especially when  $P(L_{i,j,k} | M_{i,j}) = 1$  (i.e., when local and regional probabilities of interactions

are equivalent). This suggests that high values of  $P(L_{i,j,k}|M_{i,j})$  do not adequately capture the variability of local interactions. However, these biases tend to diminish as the number of sampled networks increases, indicating that most interactions are eventually captured when  $P(L_{i,j,k}|M_{i,j})$  is high. In contrast, low values of  $P(L_{i,j,k}|M_{i,j})$  lead to missing interactions, resulting in an underestimation of the number of interactions and connectance. These results underscore the importance of using the appropriate level of variability when estimating local interaction probabilities.

315

## 316 Metawebs: regional catalogs of interactions

### 317 Unobserved interactions should be probabilistic

318 Metawebs (Dunne 2006) are networks of potential interactions over broad spatial, temporal, and taxonomic  
319 scales (e.g., food webs at the continental scale). Potential interactions describe the biological capacity of taxa to  
320 interact under optimal or feasible environmental conditions, which is typically assessed at the regional scale.  
321 Metawebs of probabilistic interactions are particularly useful in situations where there is uncertainty in the  
322 ability of taxa to interact. They may also be used as informative priors of local interactions.  
323 In contrast to local networks, where interaction probabilities arise from the processes of the interaction and the  
324 lack of information on the conditions, interaction probabilities in metawebs solely result from a lack of  
325 knowledge. This uncertainty arises due to insufficient interaction data, especially for taxa that have not yet been  
326 observed to co-occur, and uncertainties in trait-matching models. As data accumulates, interactions in  
327 metawebs should tend towards binarity, either taking a value of 1 (observing an interaction at least once) or  
328 approaching 0 (repeatedly failing to observe an interaction between co-occurring taxa). Confidently observing  
329 an interaction once confirms its biological feasibility, but failing to observe it (even on multiple occasions) does  
330 not ensure that it is non-feasible (e.g. due to false negatives, Catchen *et al.* 2023). While local interaction  
331 probabilities are irreducible because of local variability, uncertainty of regional interactions reduces to 0 with  
332 the addition of information. Moreover, although *neutrally* forbidden interactions (i.e., between rare species,  
333 Canard *et al.* 2012) tend to have low probability values in local networks, they may be non-random in the  
334 metaweb (i.e., a probability of 1 when species' biological traits can support an interaction if they were to  
335 encounter each other at high enough abundances). Likewise, non-co-occurring taxa may have a non-zero  
336 probability of interaction in the metaweb. Regional interaction probabilities are thus fundamentally different

337 from local interaction probabilities, both in terms of uncertainty sources and probability values.

338 The extent of sampling effort influences our evaluation of probabilities of potential interactions, as sampling

339 over a larger area or for a longer duration enables us to capture a greater number of regional interactions

340 (McLeod *et al.* 2021). However, in contrast with local networks of probabilistic interactions, which describe

341 local interaction stochasticity, regional interactions are not evaluated for any particular local context. In Box 3,

342 we discuss the differences in spatial and temporal scaling of regional interactions compared to local interactions.

343 We do so by using the host-parasite networks of Kopelke *et al.* (2017) as an illustration of spatial scaling.

### 344 **Potential interactions are sometimes linked to environmental conditions**

#### 345 **Biological feasibility**

346 Potential interactions describe what we refer to as the *biological* feasibility of interactions, which is based solely

347 on the regional traits distributions  $T_i$  and  $T_j$  of taxa  $i$  and  $j$ , respectively. We define regional trait distributions as

348 the range of phenotypes that a taxon can express across various environments. Regional traits may differ from

349 local traits  $T_{i,k}$  and  $T_{j,k}$ , which vary spatially and temporally because of phenotypic plasticity and local

350 environmental variability (Berg & Ellers 2010). A probability of potential interaction in a metaweb  $M$

351 describing the biological feasibility of interactions may be expressed as:

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

352 which, in contrast with local networks, is not conditioned on any spatial, temporal, co-occurrence or

353 environmental variables (Tbl. 1). Because phylogenetically close species share similar traits, we should expect

354 that closely related species will have similar interacting partners. We can thus use phylogeny to predict species

355 traits and infer regional interactions (Eklöf & Stouffer 2016; Stouffer *et al.* 2012; Strydom *et al.* 2022). The

356 taxonomic level at which interactions are evaluated also influences the distribution of regional traits. However,

357 as explained in Box 4, there is no fundamental difference in the taxonomic scaling of regional and local

358 interactions (i.e., how interaction probabilities change with taxonomic level) because they both depend on trait

359 aggregation.

360 The biological feasibility of interactions expresses our degree of belief that there exists at least one combination

361 of phenotypes that could support an interaction if they were to encounter each other, assuming they had enough

362 time to interact. Evaluating this probability is conducted without incorporating the environmental conditions  
363 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  
364 forbidden interactions (i.e., the probability that their traits do not support an interaction), which is based  
365 uniquely on biological traits:

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

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

### 375 Ecological feasibility

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

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

379 where  $E$  is the environmental conditions under which potential interactions are evaluated (Tbl. 1). Unlike  $E_k$ ,  
380 these environmental conditions do not represent conditions occurring at specific locations. Ecological  
381 feasibility represents the probability that two taxa interact if they were to encounter each other under given  
382 environmental conditions, assuming they had enough time to interact. Incorporating environmental conditions  
383 into a trait-matching model may be important when there is high covariance between the environment and traits.  
384 For instance, in our example involving rattlesnakes and lemmings, the probability of potential interaction

385 between these two species may be low in most environmental conditions. Western diamondback rattlesnakes  
386 may be unactive under low temperatures (Kissner *et al.* 1997), whereas wood lemmings may have low tolerance  
387 to high temperatures (Kausrud *et al.* 2008). The probability that an interaction is ecologically feasible is always  
388 lower than the probability that it is biologically feasible, even across all environmental conditions:

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

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

### 394 Metawebs are built with interaction data and trait-matching models

395 Starting from a selected set of taxa, which are usually distributed within a broad region of interest, metawebs  
396 can be built using different data sources, including literature review (e.g., Maiorano *et al.* 2020), aggregated  
397 interaction data (e.g., Gravel *et al.* 2019; Saravia *et al.* 2022), trait-matching models (e.g., Shaw *et al.* 2024;  
398 Strydom *et al.* 2022), and expert knowledge. Every pair of taxa that have confidently been observed to interact  
399 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  
400 from local networks of probabilistic interactions, where interaction events may remain stochastic (i.e.,  
401  $P(L_{i,j,k}) < 1$ ) even after empirically observing interactions due to their spatiotemporal variability. Interactions  
402 that were never observed typically have low probability values in local networks and vary from low to high  
403 values in metawebs, contingent upon taxa traits distributions (reaching 0 for forbidden links).

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

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

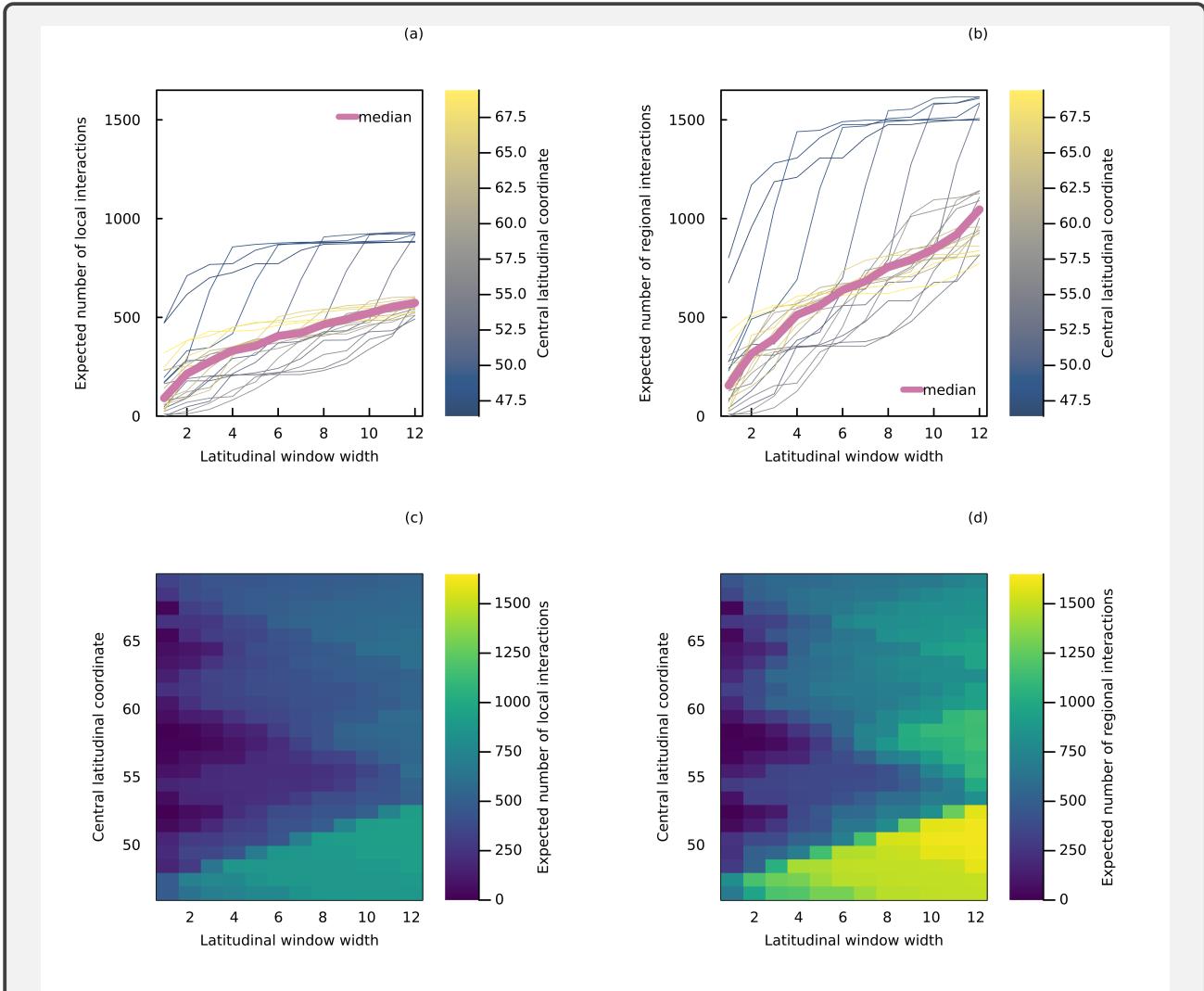
408 The reduction in the probability of regional interaction after considering that it was never observed locally (i.e.,  
409  $P(M_{i,j} = 1|O_{i,j,k} = 0, \dots) < P(M_{i,j} = 1| \dots)$ ) occurs because  $P(O_{i,j,k} = 0|M_{i,j} = 1, \dots)$  must be lower than  
410  $P(O_{i,j,k} = 0, \dots)$ , i.e. there is a higher chance of observing an interaction when it is biologically feasible.  
411 Observations of interactions may be false positives because of observation errors due to taxonomic  
412 misidentifications and ecological misinterpretations, such as those involving phylogenetically close species or  
413 cryptic species and interactions (Pringle & Hutchinson 2020). Likewise, forbidden interactions may be false  
414 negatives, e.g. if they have been evaluated based on unrepresentative or incomplete traits distributions.  
415 Employing Bayesian models proves valuable when estimating interaction probabilities in metawebs (e.g.,  
416 Bartomeus *et al.* 2016; Cirtwill *et al.* 2019). This improvement is achieved by updating prior information  
417 regarding the feasibility of interactions (e.g., experts' prior assessments of interaction probabilities) with  
418 empirical data on interactions and traits. By improving our estimation of potential interaction probabilities, we  
419 may build more reliable metawebs that adequately reflect our uncertainty on the biological feasibility of  
420 interactions.

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

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

In Fig. 3, we show how the expected *number* of local host-parasite interactions scales with the spatial

boundary of the network (represented by an expanding latitudinal window) in comparison with regional interactions. We do so using the host-parasite networks of Kopelke *et al.* (2017). The increase in the number of regional interactions is due to the inclusion of more species in a larger area. To ensure a conservative comparison between aggregated local and regional networks, we employed equal interaction probabilities (i.e., using  $P(L_{i,j,k}|M_{i,j}) = 1$ ) in both types of network. This means that local interaction probabilities could not increase further when aggregating them. Despite this, we notice that the total number of regional interactions scales more rapidly than local interactions. This is because numerous regional interactions involve species that never co-occur, and as a result, these interactions are not captured in local networks. All data manipulation and methods are described in Appendix 1.



**Figure 3: Spatial scaling of interactions.** Expected number of host-parasite interactions in a network aggregating all (a) local and (b) regional probabilistic interactions within a latitudinal window of a given width. Every dashed curve corresponds to a different window centered at a given latitude (color bar), with the pink solid line representing the median number of interactions across windows. Heatmaps of the expected number of (c) local and (d) regional interactions found in windows of specified width and position (central latitude). Probabilities of regional interactions were obtained with a false positive rate of 5% and a false negative rate of 10%. Local probabilistic interactions were derived from regional probabilistic interactions by setting the value of  $P(L_{i,j,k}|M_{i,j})$  (the local probability of interaction among potentially interacting species) to 1. Aggregated local networks were obtained by aggregating both the species and interactions found within a particular latitudinal window, with the values of  $P(L_{i,j,k}|M_{i,j})$  remaining at their maximum value of 1.

423

#### Box 4: Taxonomic scaling of interactions

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

424

change with taxonomic level) is important. There are no inherent differences between the taxonomic scaling of local and regional interactions. The taxonomic level of interactions impacts the definition of nodes. Local and regional interaction probabilities are not directly conditioned on taxonomic scale. However, some conditional variables (e.g., trait distribution) may covary with taxonomic scale. In such cases, local and regional interaction probabilities would change taxonomically following the scaling of these variables.

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

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

where  $g_{1,i}$  and  $g_{2,j}$  are the species of the corresponding genus and assuming independence between species-level interactions. In contrast, a different approach is necessary when transitioning from a broader to a finer level of organization. This is because the knowledge of an interaction between two genera does not guarantee that all possible pairwise species combinations will also interact. One possible method is to build a finer-scale network by generating probabilities of interaction through random sampling from a beta distribution, parameterized by the broader-scale network.

Fundamentally, the taxonomic scaling of interactions involves aggregating interactions between individuals into larger groups. Interaction probabilities at broader taxonomic scales should thus be conform to probabilities of interactions between individuals. For example, Canard *et al.* (2012) built a species-based network using simulated individual-based networks. In local individual-based food webs, the probability that two individuals interact reflects our degree of belief that one individual will consume the other. Likewise, in local species-based food webs, the probability that two species interact represents our degree of belief that *at least* one individual from the predator species will consume at least another individual from the prey species. In that regard, taxonomic scaling is analogous to the spatial and temporal scaling of interactions, as they all represent different ways to aggregate individuals into broader groups (either spatially, temporally, or taxonomically).

### Box 5: Sampling for binary interaction networks

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

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

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

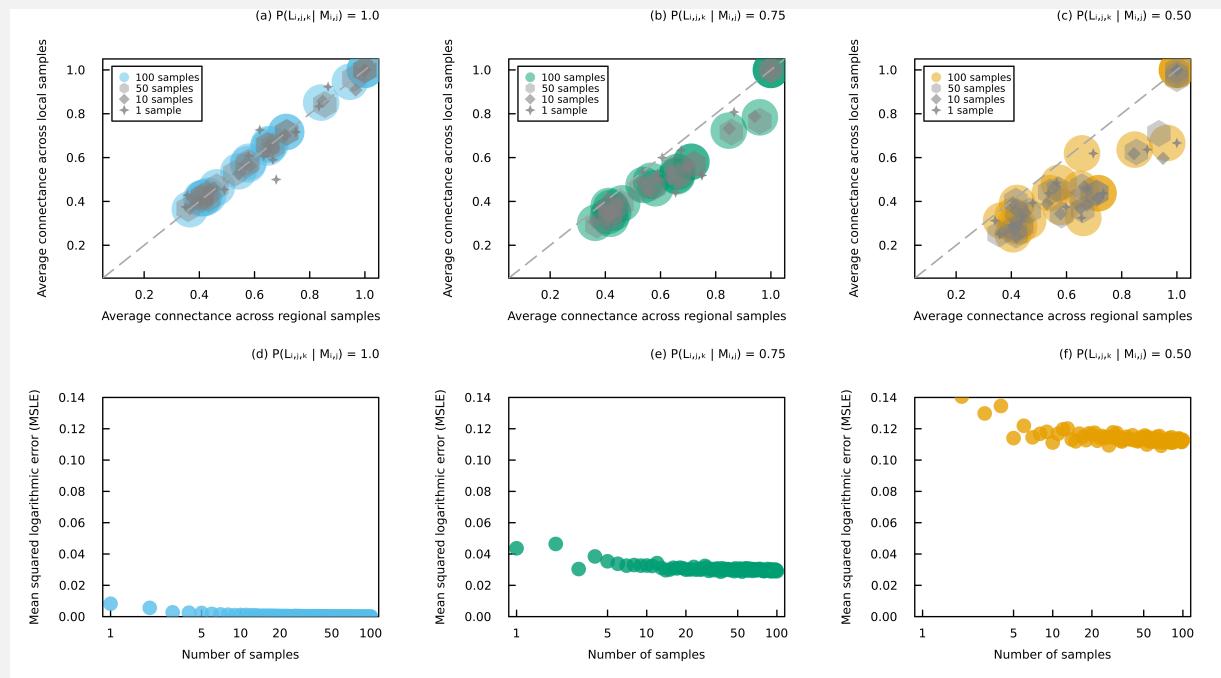
The second technique is to independently sample each local network of probabilistic interactions:

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

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

In Fig. 4, we compare the average connectance of binary interaction networks resulting from these two sampling techniques. We sampled regional and local interactions from our host-parasite networks of probabilistic interactions (Kopelke *et al.* 2017), generating a number of binary interaction network realizations for each site in the dataset. These two sampling techniques yield different outcomes, particularly for intermediate values of  $P(L_{i,j,k}|M_{i,j})$  of 0.50, which represent instances where regional

interactions do not consistently manifest locally (i.e., with the largest local variability). As anticipated, we observe that sampling binary interactions from the metaweb tends to overestimate connectance on average compared to sampling them from local networks (Fig. 4). We also observe an increase in the variability of connectance when employing a single simulation (Fig. 4a-c, cross markers), which is a more tangible representation of the process leading to the realization of local interactions in nature. All data manipulation and methods are described in Appendix 1.



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

Both sampling techniques assume independence between interactions, which might not hold true in reality.

Covariation among interactions could exist even if we do not explicitly condition interactions on others. For example, an interaction between two taxa could be more probable when another interaction occurs. The consequences of this assumption of independence on the prediction of network structure have yet to be empirically examined.

428

## 429 Future perspectives

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

### 443 Predicting local networks from metawebs

444 Metawebs are a valuable source of ecological information for predicting local networks across time and space.  
445 Local networks of binary interactions can be reconstructed by selecting a subset of taxa and interactions from  
446 the metaweb (Dunne 2006). Metawebs thus contain more interactions than local networks, even though their  
447 connectance is usually much smaller than the one of local networks (Gravel *et al.* 2011). Determining the list of  
448 taxa to select can be achieved empirically (e.g., observed occurrence data for a site) or numerically (e.g., species  
449 distribution models). As species composition is arguably easier to sample or predict than pairwise interactions,  
450 the primary challenge lies in deciding which interactions to select from the metaweb. Inferring the structure of

451 local networks from the metaweb before predicting local pairwise interactions could hold promise (Strydom *et*  
452 *al.* 2021), considering that the structure of local networks is constrained by the metaweb (Saravia *et al.* 2022).  
  
453 Inferring local networks of probabilistic interactions from a metaweb comes with its own set of challenges. For  
454 example, Dansereau *et al.* (2023) inferred spatially-explicit food webs from a metaweb of probabilistic trophic  
455 interactions between Canadian mammals. Their predicted localized food webs are downscaled versions of the  
456 metaweb (i.e., localized metawebs with the same interaction probabilities as those in the regional metaweb). To  
457 infer local networks as defined in this manuscript (i.e., describing local realizations of interactions), local  
458 interaction probabilities must be smaller than regional interaction probabilities. This decrease is due to the  
459 prerequisite that two taxa must initially possess the capacity to interact before engaging in local interactions  
460 (Eq. 3). Inferring local networks from a metaweb by maintaining identical interaction probability values  
461 introduces systematic biases into the predictions, as discussed in Box 2 (unless networks are seen as downscaled  
462 metawebs).

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

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

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

$$\int_{E_k} \int_A \int_t P(L_{i,j,k}|E_k, A, t) dt dA dE_k \leq P(M_{i,j}|T_i, T_j). \quad (18)$$

472 Estimating more precisely the probability  $P(L_{i,j,k}|M_{i,j})$  that two taxa interact locally if they can potentially  
473 interact allows for improved predictions of local networks from the metaweb of probabilistic interactions. This  
474 task is challenging due to the variability of this probability across space and time, as well as its variability

475 across pairwise interactions within a network. Using simple models of  $P(L_{i,j,k}|M_{i,j})$ , as shown in Box 2,  
476 represents an initial step toward the overarching objective of reconstructing local networks from metawebs.

477 **Quantifying and reducing interaction uncertainty**

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

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

496 **Relaxing the independence assumption**

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

502 A more accurate representation of the stochasticity of ecological networks involves creating *probabilistic*  
503 *networks* ( $P(L|...)$  and  $P(M|...)$ ), rather than networks of *probabilistic interactions* ( $P(L_{i,j,k}|...)$  and  $P(M_{i,j}|...)$ ).  
504 Probabilistic networks describe the probability that a particular network of binary (or quantitative) interactions  
505 (its whole adjacency matrix) is realized. For example, Young *et al.* (2021) used a Bayesian approach to estimate  
506 the probability of different plant-pollinator network structures derived from imperfect observational data. A  
507 probability distribution of ecological networks may also be derived using the principle of maximum entropy  
508 given structural constrained (e.g., Cimini *et al.* 2019; Park & Newman 2004).  
  
509 Regardless of the method employed, generating probabilistic local networks could lead to more accurate  
510 predictions of local networks of binary interactions by bypassing the independence assumption. Probabilistic  
511 networks could serve as an alternative to null hypothesis significance testing when comparing the structure of a  
512 local network to some random expectations or, as done in Pellissier *et al.* (2018), to the metaweb. These random  
513 expectations are typically derived by performing a series of Bernoulli trials on probabilistic interactions,  
514 assuming independence, to generate a distribution of networks of binary interactions to calculate their structure  
515 (Poisot *et al.* 2016). One could instead compare the likelihood of an observed network to the one of the most  
516 likely network structure (according to the probabilistic network distribution), thereby directly obtaining a  
517 measure of discrepancy of the empirical network. Generating probabilistic ecological networks represents a  
518 tangible challenge, one that, in the coming years, promises to unlock doors to more advanced and adequate  
519 analyses of ecological networks.

## 520 Acknowledgment

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

528 **References**

- 529 Abrego, N., Roslin, T., Huotari, T., Ji, Y., Schmidt, N.M., Wang, J., *et al.* (2021). Accounting for species  
530 interactions is necessary for predicting how arctic arthropod communities respond to climate change.  
531 *Ecography*, 44, 885–896.
- 532 Angilletta, M.J., Jr., Steury, T.D. & Sears, M.W. (2004). Temperature, Growth Rate, and Body Size in  
533 Ectotherms: Fitting Pieces of a Life-History Puzzle<sup>1</sup>. *Integrative and Comparative Biology*, 44, 498–509.
- 534 Araujo, M.B., Rozenfeld, A., Rahbek, C. & Marquet, P.A. (2011). Using species co-occurrence networks to  
535 assess the impacts of climate change. *Ecography*, 34, 897–908.
- 536 Bartomeus, I., Gravel, D., Tylianakis, J.M., Aizen, M.A., Dickie, I.A. & Bernard-Verdier, M. (2016). A  
537 common framework for identifying linkage rules across different types of interactions. *Functional Ecology*,  
538 30, 1894–1903.
- 539 Bascompte, J., Jordano, P., Melián, C.J. & Olesen, J.M. (2003). The nested assembly of plant–animal  
540 mutualistic networks. *Proceedings of the National Academy of Sciences*, 100, 9383–9387.
- 541 Berg, M.P. & Ellers, J. (2010). Trait plasticity in species interactions: A driving force of community dynamics.  
542 *Evolutionary Ecology*, 24, 617–629.
- 543 Berlow, E.L., Neutel, A.-M., Cohen, J.E., De Ruiter, P.C., Ebenman, B., Emmerson, M., *et al.* (2004).  
544 Interaction strengths in food webs: Issues and opportunities. *Journal of Animal Ecology*, 73, 585–598.
- 545 Bolnick, D.I., Amarasekare, P., Araújo, M.S., Bürger, R., Levine, J.M., Novak, M., *et al.* (2011). Why  
546 intraspecific trait variation matters in community ecology. *Trends in Ecology & Evolution*, 26, 183–192.
- 547 Borrett, S.R. & Scharler, U.M. (2019). Walk partitions of flow in Ecological Network Analysis: Review and  
548 synthesis of methods and indicators. *Ecological Indicators*, 106, 105451.
- 549 Boulangeat, I., Gravel, D. & Thuiller, W. (2012). Accounting for dispersal and biotic interactions to disentangle  
550 the drivers of species distributions and their abundances. *Ecology Letters*, 15, 584–593.
- 551 Brimacombe, C., Bodner, K., Michalska-Smith, M., Poisot, T. & Fortin, M.-J. (2023). Shortcomings of reusing  
552 species interaction networks created by different sets of researchers. *PLOS Biology*, 21, e3002068.
- 553 Broom, M. & Ruxton, G.D. (2005). You can run—or you can hide: Optimal strategies for cryptic prey against  
554 pursuit predators. *Behavioral Ecology*, 16, 534–540.

- 555 Burnham, K.P. & Anderson, D.R. (2004). [Multimodel Inference: Understanding AIC and BIC in Model](#)  
556 [Selection](#). *Sociological Methods & Research*, 33, 261–304.
- 557 Canard, E.F., Mouquet, N., Mouillot, D., Stanko, M., Miklisova, D. & Gravel, D. (2014). [Empirical Evaluation](#)  
558 [of Neutral Interactions in Host-Parasite Networks](#). *The American Naturalist*, 183, 468–479.
- 559 Canard, E., Mouquet, N., Marescot, L., Gaston, K.J., Gravel, D. & Mouillot, D. (2012). [Emergence of](#)  
560 [Structural Patterns in Neutral Trophic Networks](#). *PLOS ONE*, 7, e38295.
- 561 Carlson, C.J., Farrell, M.J., Grange, Z., Han, B.A., Mollentze, N., Phelan, A.L., et al. (2021). [The future of](#)  
562 [zoonotic risk prediction](#). *Philosophical Transactions of the Royal Society B: Biological Sciences*, 376,  
563 20200358.
- 564 Caron, D., Brose, U., Lurgi, M., Blanchet, F.G., Gravel, D. & Pollock, L.J. (2024). [Trait-matching models](#)  
565 [predict pairwise interactions across regions, not food web properties](#). *Global Ecology and Biogeography*,  
566 33, e13807.
- 567 Castoe, T.A., Spencer, C.L. & Parkinson, C.L. (2007). [Phylogeographic structure and historical demography of](#)  
568 [the western diamondback rattlesnake \(\*Crotalus Atrox\*\): A perspective on North American desert](#)  
569 [biogeography](#). *Molecular Phylogenetics and Evolution*, 42, 193–212.
- 570 Catchen, M.D., Poisot, T., Pollock, L.J. & Gonzalez, A. (2023). [The missing link: Discerning true from false](#)  
571 [negatives when sampling species interaction networks](#).
- 572 Cazelles, K., Araujo, M.B., Mouquet, N. & Gravel, D. (2016). [A theory for species co-occurrence in interaction](#)  
573 [networks](#). *Theoretical Ecology*, 9, 39–48.
- 574 Chagnon, P.-L. (2015). [Characterizing topology of ecological networks along gradients: The limits of metrics'](#)  
575 [standardization](#). *Ecological Complexity*, 22, 36–39.
- 576 Choh, Y., Ignacio, M., Sabelis, M.W. & Janssen, A. (2012). [Predator-prey role reversals, juvenile experience](#)  
577 [and adult antipredator behaviour](#). *Scientific Reports*, 2, 728.
- 578 Cimini, G., Squartini, T., Saracco, F., Garlaschelli, D., Gabrielli, A. & Caldarelli, G. (2019). [The statistical](#)  
579 [physics of real-world networks](#). *Nature Reviews Physics*, 1, 58–71.
- 580 Cirtwill, A.R., Eklöf, A., Roslin, T., Wootton, K. & Gravel, D. (2019). [A quantitative framework for](#)  
581 [investigating the reliability of empirical network construction](#). *Methods in Ecology and Evolution*, 10,  
582 902–911.

- 583 Clark, J.S., Bell, D.M., Hersh, M.H., Kwit, M.C., Moran, E., Salk, C., *et al.* (2011). Individual-scale variation,  
584 species-scale differences: Inference needed to understand diversity. *Ecology Letters*, 14, 1273–1287.
- 585 Dallas, T., Park, A.W. & Drake, J.M. (2017). Predicting cryptic links in host-parasite networks. *PLOS  
586 Computational Biology*, 13, e1005557.
- 587 Dansereau, G., Barros, C. & Poisot, T. (2023). Spatially explicit predictions of food web structure from regional  
588 level data.
- 589 Dunne, J.A. (2006). The Network Structure of Food Webs. In: *Ecological networks: Linking structure and  
590 dynamics* (eds. Dunne, J.A. & Pascual, M.). Oxford University Press, pp. 27–86.
- 591 Eklöf, A. & Stouffer, D.B. (2016). The phylogenetic component of food web structure and intervality.  
592 *Theoretical Ecology*, 9, 107–115.
- 593 Elmasri, M., Farrell, M.J., Davies, T.J. & Stephens, D.A. (2020). A hierarchical Bayesian model for predicting  
594 ecological interactions using scaled evolutionary relationships. *The Annals of Applied Statistics*, 14,  
595 221–240.
- 596 Elton, C.S. (2001). *Animal Ecology*. University of Chicago Press, Chicago, IL.
- 597 Emmerson, M.C. & Raffaelli, D. (2004). Predator–prey body size, interaction strength and the stability of a real  
598 food web. *Journal of Animal Ecology*, 73, 399–409.
- 599 Fedorov, V.B., Goropashnaya, A.V., Boeskorov, G.G. & Cook, J.A. (2008). Comparative phylogeography and  
600 demographic history of the wood lemming (*Myopus schisticolor*): Implications for late Quaternary history  
601 of the taiga species in Eurasia. *Molecular Ecology*, 17, 598–610.
- 602 Fortuna, M.A. & Bascompte, J. (2006). Habitat loss and the structure of plant–animal mutualistic networks.  
603 *Ecology Letters*, 9, 281–286.
- 604 Fu, X., Seo, E., Clarke, J. & Hutchinson, R.A. (2021). Link Prediction Under Imperfect Detection:  
605 Collaborative Filtering for Ecological Networks. *IEEE Transactions on Knowledge and Data Engineering*,  
606 33, 3117–3128.
- 607 Galiana, N., Lurgi, M., Claramunt-López, B., Fortin, M.-J., Leroux, S., Cazelles, K., *et al.* (2018). The spatial  
608 scaling of species interaction networks. *Nature Ecology & Evolution*, 2, 782–790.
- 609 Gilman, S.E., Urban, M.C., Tewksbury, J., Gilchrist, G.W. & Holt, R.D. (2010). A framework for community  
610 interactions under climate change. *Trends in Ecology & Evolution*, 25, 325–331.

- 611 Godsoe, W., Murray, R. & Iritani, R. (2022). Species interactions and diversity: A unified framework using Hill  
612 numbers. *Oikos*, n/a, e09282.
- 613 Golubski, A.J. & Abrams, P.A. (2011). Modifying modifiers: What happens when interspecific interactions  
614 interact? *Journal of Animal Ecology*, 80, 1097–1108.
- 615 Gonzalez, A. & Londoño, M.C. (2022). Monitor biodiversity for action. *Science*, 378, 1147–1147.
- 616 Gonzalez-Varo, J.P. & Traveset, A. (2016). The Labile Limits of Forbidden Interactions. *Trends in Ecology &*  
617 *Evolution*, 31, 700–710.
- 618 Gravel, D., Baiser, B., Dunne, J.A., Kopelke, J.-P., Martinez, N.D., Nyman, T., et al. (2019). Bringing Elton  
619 and Grinnell together: A quantitative framework to represent the biogeography of ecological interaction  
620 networks. *Ecography*, 42, 401–415.
- 621 Gravel, D., Massol, F., Canard, E., Mouillot, D. & Mouquet, N. (2011). Trophic theory of island biogeography.  
622 *Ecology Letters*, 14, 1010–1016.
- 623 Gravel, D., Poisot, T., Albouy, C., Velez, L. & Mouillot, D. (2013). Inferring food web structure from  
624 predator–prey body size relationships. *Methods in Ecology and Evolution*, 4, 1083–1090.
- 625 Guimarães, P.R. (2020). The Structure of Ecological Networks Across Levels of Organization. *Annual Review*  
626 *of Ecology, Evolution, and Systematics*, 51, 433–460.
- 627 Guimerà, R. & Sales-Pardo, M. (2009). Missing and spurious interactions and the reconstruction of complex  
628 networks. *Proceedings of the National Academy of Sciences*, 106, 22073–22078.
- 629 Hemprich-Bennett, D.R., Oliveira, H.F.M., Le Comber, S.C., Rossiter, S.J. & Clare, E.L. (2021). Assessing the  
630 impact of taxon resolution on network structure. *Ecology*, 102, e03256.
- 631 Herrera, C.M. (1989). Pollinator abundance, morphology, and flower visitation rate: Analysis of the “quantity”  
632 component in a plant-pollinator system. *Oecologia*, 80, 241–248.
- 633 Hortal, J., de Bello, F., Diniz-Filho, J.A.F., Lewinsohn, T.M., Lobo, J.M. & Ladle, R.J. (2015). Seven Shortfalls  
634 that Beset Large-Scale Knowledge of Biodiversity. *Annual Review of Ecology, Evolution, and Systematics*,  
635 46, 523–549.
- 636 Jordano, P. (1987). Patterns of Mutualistic Interactions in Pollination and Seed Dispersal: Connectance,  
637 Dependence Asymmetries, and Coevolution. *The American Naturalist*, 129, 657–677.

- 638 Jordano, P. (2016). Sampling networks of ecological interactions. *Functional Ecology*, 30, 1883–1893.
- 639 Jordano, P., Bascompte, J. & Olesen, J.M. (2003). Invariant properties in coevolutionary networks of  
640 plant–animal interactions. *Ecology Letters*, 6, 69–81.
- 641 Kausrud, K.L., Mysterud, A., Steen, H., Vik, J.O., Østbye, E., Cazelles, B., *et al.* (2008). Linking climate  
642 change to lemming cycles. *Nature*, 456, 93–97.
- 643 Kéfi, S., Berlow, E.L., Wieters, E.A., Joppa, L.N., Wood, S.A., Brose, U., *et al.* (2015). Network structure  
644 beyond food webs: Mapping non-trophic and trophic interactions on Chilean rocky shores. *Ecology*, 96,  
645 291–303.
- 646 Kéfi, S., Berlow, E.L., Wieters, E.A., Navarrete, S.A., Petchey, O.L., Wood, S.A., *et al.* (2012). More than a  
647 meal... integrating non-feeding interactions into food webs. *Ecology Letters*, 15, 291–300.
- 648 Kéfi, S., Miele, V., Wieters, E.A., Navarrete, S.A. & Berlow, E.L. (2016). How Structured Is the Entangled  
649 Bank? The Surprisingly Simple Organization of Multiplex Ecological Networks Leads to Increased  
650 Persistence and Resilience. *PLOS Biology*, 14, e1002527.
- 651 Kissner, K.J., Forbes, M.R. & Secoy, D.M. (1997). Rattling Behavior of Prairie Rattlesnakes (*Crotalus viridis*  
652 *viridis*, Viperidae) in Relation to Sex, Reproductive Status, Body Size, and Body Temperature. *Ethology*,  
653 103, 1042–1050.
- 654 Klecka, J. & Boukal, D.S. (2014). The effect of habitat structure on prey mortality depends on predator and  
655 prey microhabitat use. *Oecologia*, 176, 183–191.
- 656 Kopelke, J.-P., Nyman, T., Cazelles, K., Gravel, D., Vissault, S. & Roslin, T. (2017). Food-web structure of  
657 willow-galling sawflies and their natural enemies across Europe. *Ecology*, 98, 1730–1730.
- 658 Kulkarni, M.A., Kweka, E., Nyale, E., Lyatuu, E., Mosha, F.W., Chandramohan, D., *et al.* (2006).  
659 Entomological Evaluation of Malaria Vectors at Different Altitudes in Hai District, Northeastern Tanzania.  
660 *Journal of Medical Entomology*, 43, 580–588.
- 661 Maiorano, L., Montemaggiori, A., Ficetola, G.F., O'Connor, L. & Thuiller, W. (2020). TETRA-EU 1.0: A  
662 species-level trophic metaweb of European tetrapods. *Global Ecology and Biogeography*, 29, 1452–1457.
- 663 McCann, K. (2007). Protecting biostructure. *Nature*, 446, 29–29.
- 664 McCann, K.S. (2011). Food webs (MPB-50). In: *Food Webs (MPB-50)*. Princeton University Press.

- 665 McLeod, A., Leroux, S.J., Gravel, D., Chu, C., Cirtwill, A.R., Fortin, M.-J., *et al.* (2021). [Sampling and](#)  
666 [asymptotic network properties of spatial multi-trophic networks](#). *Oikos*, n/a.
- 667 Melián, C.J., Vilas, C., Baldó, F., González-Ortegón, E., Drake, P. & Williams, R.J. (2011). [Eco-evolutionary](#)  
668 [Dynamics of Individual-Based Food Webs](#). In: *Advances in Ecological Research*, The Role of Body Size in  
669 Multispecies Systems (ed. Belgrano, A.). Academic Press, pp. 225–268.
- 670 Morales-Castilla, I., Matias, M.G., Gravel, D. & Araújo, M.B. (2015). [Inferring biotic interactions from](#)  
671 [proxies](#). *Trends in Ecology & Evolution*, 30, 347–356.
- 672 Ohlmann, M., Miele, V., Dray, S., Chalmandrier, L., O'Connor, L. & Thuiller, W. (2019). [Diversity indices for](#)  
673 [ecological networks: A unifying framework using Hill numbers](#). *Ecology Letters*, 22, 737–747.
- 674 Olesen, J.M., Bascompte, J., Dupont, Y.L., Elberling, H., Rasmussen, C. & Jordano, P. (2010). [Missing and](#)  
675 [forbidden links in mutualistic networks](#). *Proceedings of the Royal Society B: Biological Sciences*, 278,  
676 725–732.
- 677 Park, J. & Newman, M.E.J. (2004). [Statistical mechanics of networks](#). *Physical Review E*, 70, 066117.
- 678 Pascual, M., Dunne, J.A. & Dunne, J.A. (2006). *Ecological Networks: Linking Structure to Dynamics in Food*  
679 *Webs*. Oxford University Press, USA.
- 680 Pellissier, L., Albouy, C., Bascompte, J., Farwig, N., Graham, C., Loreau, M., *et al.* (2018). [Comparing species](#)  
681 [interaction networks along environmental gradients](#). *Biological Reviews*, 93, 785–800.
- 682 Petchey, O.L., Beckerman, A.P., Riede, J.O. & Warren, P.H. (2008). [Size, foraging, and food web structure](#).  
683 *Proceedings of the National Academy of Sciences*, 105, 4191–4196.
- 684 Pilosof, S., Porter, M.A., Pascual, M. & Kéfi, S. (2017). [The multilayer nature of ecological networks](#). *Nature*  
685 *Ecology & Evolution*, 1, 1–9.
- 686 Plagányi, É.E. & Butterworth, D.S. (2004). [A critical look at the potential of Ecopath with ecosim to assist in](#)  
687 [practical fisheries management](#). *African Journal of Marine Science*, 26, 261–287.
- 688 Poisot, T., Canard, E., Mouillot, D., Mouquet, N. & Gravel, D. (2012). [The dissimilarity of species interaction](#)  
689 [networks](#). *Ecology Letters*, 15, 1353–1361.
- 690 Poisot, T., Cirtwill, A.R., Cazelles, K., Gravel, D., Fortin, M.-J. & Stouffer, D.B. (2016). [The structure of](#)  
691 [probabilistic networks](#). *Methods in Ecology and Evolution*, 7, 303–312.

- 692 Poisot, T. & Gravel, D. (2014). When is an ecological network complex? Connectance drives degree  
693 distribution and emerging network properties. *PeerJ*, 2, e251.
- 694 Poisot, T., Stouffer, D.B. & Gravel, D. (2015). Beyond species: Why ecological interaction networks vary  
695 through space and time. *Oikos*, 124, 243–251.
- 696 Polis, G.A. (1991). Complex Trophic Interactions in Deserts: An Empirical Critique of Food-Web Theory. *The  
697 American Naturalist*, 138, 123–155.
- 698 Pollock, L.J., Tingley, R., Morris, W.K., Golding, N., O'Hara, R.B., Parris, K.M., et al. (2014). Understanding  
699 co-occurrence by modelling species simultaneously with a Joint Species Distribution Model (JSDM).  
700 *Methods in Ecology and Evolution*, 5, 397–406.
- 701 Pomeranz, J.P.F., Thompson, R.M., Poisot, T. & Harding, J.S. (2019). Inferring predator–prey interactions in  
702 food webs. *Methods in Ecology and Evolution*, 10, 356–367.
- 703 Pringle, R.M. & Hutchinson, M.C. (2020). Resolving Food-Web Structure. *Annual Review of Ecology,  
704 Evolution, and Systematics*, 51, 55–80.
- 705 Proulx, S.R., Promislow, D.E.L. & Phillips, P.C. (2005). Network thinking in ecology and evolution. *Trends in  
706 Ecology & Evolution*, SPECIAL ISSUE: BUMPER BOOK REVIEW, 20, 345–353.
- 707 Pulliam, H.R. (1974). On the Theory of Optimal Diets. *The American Naturalist*, 108, 59–74.
- 708 Rooney, N. & McCann, K.S. (2012). Integrating food web diversity, structure and stability. *Trends in Ecology  
709 & Evolution*, 27, 40–46.
- 710 Salim, J.A., Saraiva, A.M., Zermoglio, P.F., Agostini, K., Wolowski, M., Drucker, D.P., et al. (2022). Data  
711 standardization of plant–pollinator interactions. *GigaScience*, 11, giac043.
- 712 Sanders, D. & van Veen, F.J.F. (2012). Indirect commensalism promotes persistence of secondary consumer  
713 species. *Biology Letters*, 8, 960–963.
- 714 Saravia, L.A., Marina, T.I., Kristensen, N.P., De Troch, M. & Momo, F.R. (2022). Ecological network  
715 assembly: How the regional metaweb influences local food webs. *Journal of Animal Ecology*, 91, 630–642.
- 716 Schmeller, D.S., Julliard, R., Bellingham, P.J., Böhm, M., Brummitt, N., Chiarucci, A., et al. (2015). Towards a  
717 global terrestrial species monitoring program. *Journal for Nature Conservation*, 25, 51–57.

- 718 Shaw, J.O., Dunhill, A.M., Beckerman, A.P., Dunne, J.A. & Hull, P.M. (2024). [A framework for reconstructing](#)  
719 [ancient food webs using functional trait data.](#)
- 720 Singer, M.C. & McBride, C.S. (2012). [Geographic mosaics of species' association: A definition and an](#)  
721 [example driven by plant–insect phenological synchrony.](#) *Ecology*, 93, 2658–2673.
- 722 Staniczenko, P.P.A., Sivasubramaniam, P., Suttle, K.B. & Pearson, R.G. (2017). [Linking macroecology and](#)  
723 [community ecology: Refining predictions of species distributions using biotic interaction networks.](#)
- 724 [Ecology Letters](#), 20, 693–707.
- 725 Stock, M., Poisot, T., Waegeman, W. & De Baets, B. (2017). [Linear filtering reveals false negatives in species](#)  
726 [interaction data.](#) *SCIENTIFIC REPORTS*, 7, 45908.
- 727 Stouffer, D.B., Rezende, E.L. & Amaral, L.A.N. (2011). [The role of body mass in diet contiguity and food-web](#)  
728 [structure.](#) *Journal of Animal Ecology*, 80, 632–639.
- 729 Stouffer, D.B., Sales-Pardo, M., Sirer, M.I. & Bascompte, J. (2012). [Evolutionary Conservation of Species'](#)  
730 [Roles in Food Webs.](#) *Science*, 335, 1489–1492.
- 731 Strydom, T., Bouskila, S., Banville, F., Barros, C., Caron, D., Farrell, M.J., *et al.* (2022). [Food web](#)  
732 [reconstruction through phylogenetic transfer of low-rank network representation.](#) *Methods in Ecology and*  
733 [Evolution](#), 13.
- 734 Strydom, T., Bouskila, S., Banville, F., Barros, C., Caron, D., Farrell, M.J., *et al.* (2023). [Graph embedding and](#)  
735 [transfer learning can help predict potential species interaction networks despite data limitations.](#) *Methods in*  
736 [Ecology and Evolution](#), 14, 2917–2930.
- 737 Strydom, T., Catchen, M.D., Banville, F., Caron, D., Dansereau, G., Desjardins-Proulx, P., *et al.* (2021). [A](#)  
738 [roadmap towards predicting species interaction networks \(across space and time\).](#) *Philosophical*  
739 [Transactions of the Royal Society B-Biological Sciences](#), 376, 20210063.
- 740 Tylianakis, J.M., Tscharntke, T. & Lewis, O.T. (2007). [Habitat modification alters the structure of tropical](#)  
741 [host–parasitoid food webs.](#) *Nature*, 445, 202–205.
- 742 Vázquez, D.P., Melián, C.J., Williams, N.M., Blüthgen, N., Krasnov, B.R. & Poulin, R. (2007). [Species](#)  
743 [abundance and asymmetric interaction strength in ecological networks.](#) *Oikos*, 116, 1120–1127.
- 744 Wagenmakers, E.-J. & Farrell, S. (2004). [AIC model selection using Akaike weights.](#) *Psychonomic Bulletin &*  
745 [Review](#), 11, 192–196.

- 746 Weinstein, B.G. & Graham, C.H. (2017a). [On comparing traits and abundance for predicting species](#)  
747 [interactions with imperfect detection](#). *Food Webs*, 11, 17–25.
- 748 Weinstein, B.G. & Graham, C.H. (2017b). [Persistent bill and corolla matching despite shifting temporal](#)  
749 [resources in tropical hummingbird-plant interactions](#). *Ecology Letters*, 20, 326–335.
- 750 Woodward, G., Benstead, J.P., Beveridge, O.S., Blanchard, J., Brey, T., Brown, L.E., *et al.* (2010). [Chapter 2 -](#)  
751 [Ecological Networks in a Changing Climate](#). In: *Advances in Ecological Research*, Ecological Networks  
752 (ed. Woodward, G.). Academic Press, pp. 71–138.
- 753 Woodward, G., Brown, L.E., Edwards, F.K., Hudson, L.N., Milner, A.M., Reuman, D.C., *et al.* (2012). [Climate](#)  
754 [change impacts in multispecies systems: Drought alters food web size structure in a field experiment](#).  
755 *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367, 2990–2997.
- 756 Young, J.-G., Valdovinos, F.S. & Newman, M.E.J. (2021). [Reconstruction of plant–pollinator networks from](#)  
757 [observational data](#). *Nature Communications*, 12, 3911.