

# Deciphering ecological networks of probabilistic interactions

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

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

## Correspondance to:

Francis Banville — francis.banville@umontreal.ca

The stochastic nature of ecological interactions has led biologists to adopt a probabilistic view of ecological networks. Representing species interactions probabilistically (how likely are they to occur?) as opposed to deterministically (are they occurring?) highlights uncertainties in our knowledge of interactions. The origin of the uncertainty captured by interaction probabilities differs depending on the method used to evaluate them. It may consist of the uncertainty of predictive models, the subjective assessment of experts, or even the empirical evaluation of interaction spatiotemporal variability. Despite this growing interest in probabilistic interactions, guidelines for the estimation and documentation of probabilistic interaction data are still lacking. This is concerning because our biological understanding of interaction probabilities and the decisions made during data analysis rely heavily on the often elusive definition of the probabilistic event. We review how probabilistic interactions are defined at different spatial scales, from local interactions to regional networks (metawebs), with a strong emphasis on trophic interactions (predatory, herbivory, and parasitic interactions). These definitions are based on the distinction between the realization of an interaction at a specific time and space (local) and its biological feasibility (regional). Through case studies using open data on host-parasite interactions in Europe, we show that these two network representations have different statistical properties when it comes to common ecological applications. Specifically, we found that local webs and metawebs differ in their spatial and temporal scaling of probabilistic interactions, but not in their taxonomic scaling. We suggest two approaches to inferring binary interactions from probabilistic ones that account for these differences and show that systematic biases arise when directly inferring local webs from subsets of metawebs. Our results underscore the importance of clear metadata for networks of probabilistic interactions that describe the type of interaction (local or regional)

and uncertainty sources. Comprehensive metadata are particularly useful for the identification of optimal sampling locations where data collection may provide the greatest amount of new information on interactions and minimize their uncertainty.

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



Last revision: *March 27, 2024*

## **1 Introduction**

2 As we try to navigate global biodiversity change, filling in knowledge gaps about biodiversity becomes  
3 instrumental to monitoring and mitigating those changes (Abrego *et al.* 2021; Gonzalez & Londoño 2022).  
4 However, cataloging species, populations and, in particular, ecological interactions (e.g., predation, parasitism,  
5 and pollination) is a substantial challenge (Pascual *et al.* 2006; Polis 1991). There are methodological and  
6 biological constraints that hinder our ability to observe all ecological interactions, leading to significant  
7 uncertainties in our understanding of these interactions. For example, the spatial and temporal uncoupling of  
8 species (e.g., nocturnal and diurnal species coexisting in the same space without interacting due to differences  
9 in the timing of their daily activities, Jordano 1987) and the large number of rare and cryptic interactions in a  
10 community contribute to these uncertainties (Jordano 2016). More generally, a handful of conditions must be  
11 satisfied for an interaction to be observed locally. First, both species must have overlapping geographic ranges,  
12 i.e. they must co-occur within the region of interest (Blanchet *et al.* 2020). Second, they must have some  
13 probability of meeting. Probabilities of interspecific encounters are typically low, especially for rare species  
14 with low abundances. While species' absolute abundances may impact interaction frequencies (Vázquez *et al.*  
15 2007), encounter probabilities are determined by their relative abundances (Canard *et al.* 2012). The probability  
16 that species meet each other also depends on their biological characteristics, such as the synchronization of their  
17 phenology (Olesen *et al.* 2010; Singer & McBride 2012) and their discoverability (e.g., Broom & Ruxton 2005).  
18 Finally, when species do come into contact, an interaction occurs only if their traits are locally compatible  
19 (Poisot *et al.* 2015), including their body phenotypes (Bolnick *et al.* 2011; Gravel *et al.* 2013; Stouffer *et al.*  
20 2011) and behavioral choices (Choh *et al.* 2012; Pulliam 1974). Interactions may also be influenced by the  
21 presence or prevalence of a third species (e.g., of a more profitable prey species, Golubski & Abrams 2011;  
22 Sanders & van Veen 2012). Documenting the location and timing of interactions becomes even more difficult  
23 when accounting for the spatiotemporal variability of ecological interactions (Poisot *et al.* 2012, 2015).  
24 Environmental factors, such as temperature (Angilletta *et al.* 2004), drought (Woodward *et al.* 2012), climate  
25 change (Araujo *et al.* 2011; Gilman *et al.* 2010; Woodward *et al.* 2010), and habitat modifications (Tylianakis *et*  
26 *al.* 2007), contribute to this spatiotemporal variability by impacting species abundance and traits. Even after  
27 satisfying all these conditions, there remains a possibility that the interaction does not occur locally, either due  
28 to the intricate nature of the system or simply by chance. If it does occur, it might still go unnoticed, particularly  
29 if it happens infrequently. In this context, it is unsurprising that our knowledge of ecological interactions  
30 remains limited (Hortal *et al.* 2015) despite extensive biodiversity data collection (Schmeller *et al.* 2015).

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

58 By accounting for the uncertainty of interactions, networks of probabilistic interactions may provide a more  
59 realistic portrait of species interactions. However, how the uncertainty of pairwise interactions propagates to  
60 network structure (i.e., community-level properties driving the functioning, dynamics, and resilience of

61 ecosystems, McCann 2007; McCann 2011; Proulx *et al.* 2005; Rooney & McCann 2012) remains to be  
62 elucidated. The application and development of computational methods in network ecology, often based on a  
63 probabilistic representation of interactions, can alleviate (and guide) the sampling efforts required to document  
64 species interactions (Strydom *et al.* 2021). For example, statistical models can be used to estimate the  
65 uncertainty of pairwise interactions (Cirtwill *et al.* 2019) and the probability of missing (false negatives) and  
66 spurious (false positives) interactions (Guimerà & Sales-Pardo 2009). Considering the high rate of false  
67 negatives in species interaction data due to the difficulty of witnessing rare interactions (Catchen *et al.* 2023),  
68 these models could inform the identification of priority sampling locations of ecological webs where data  
69 collection would yield the most valuable information, thereby reducing errors. Optimization models for  
70 sampling locations have mostly found applications in biological systems that are not networks, such as  
71 identifying priority sampling sites for disease hotspots (Andrade-Pacheco *et al.* 2020), but there is substantial  
72 promise in applying them to probabilistic ecological interactions. Statistical models can also be used to generate  
73 predictions of ecological webs without prior knowledge of pairwise interactions, for instance using body size  
74 (Gravel *et al.* 2013; Petchey *et al.* 2008), phylogeny (Elmasri *et al.* 2020; Strydom *et al.* 2022), or a  
75 combination of niche and neutral processes (Bartomeus *et al.* 2016; Pomeranz *et al.* 2019) for inference.  
76 Topological null models, which generate networks of probabilistic interactions by preserving chosen  
77 characteristics of the adjacency matrix of binary interactions while intentionally omitting others (Bascompte *et*  
78 *al.* 2003; Fortuna & Bascompte 2006), serve as other examples of common probabilistic interaction models.  
79 Null models can be used to produce underlying distributions of network measures for null hypothesis  
80 significance testing. Many measures have been developed to describe the structure (Poisot *et al.* 2016) and  
81 diversity (Godsoe *et al.* 2022; Ohlmann *et al.* 2019) of probabilistic interaction webs. These models and  
82 measures support the use of this approach for the study of a wide range of ecological questions, from making  
83 better predictions of species distribution (Cazelles *et al.* 2016) to forecasting the impact of climate change on  
84 ecological webs (Gilman *et al.* 2010).  
85 Beyond methodological difficulties that may arise when assessing probabilistic interactions, a precise definition  
86 of probabilistic interactions appears to be lacking, making the estimation and use of these data more difficult.  
87 We aim to take a step back by outlining different ways in which probabilistic interactions are defined and used in  
88 network ecology. We distinguish two broad categories of probabilistic interaction networks that necessitate  
89 distinct approaches when applied to key ecological questions: local webs describing probabilities of realized  
90 interactions, and regional webs (metawebs) describing probabilities of potential interactions. We highlight the

91 distinctions in the ecological meaning of these two representations and examine some of their properties and  
92 relationships (particularly with space, time, and between each other) through empirical case studies. Moreover,  
93 the lack of clear guidelines on the use of probabilistic interaction data is worrisome, as it affects both data  
94 producers and re-users who generate and manipulate these numbers. This is concerning because sampling  
95 strategies and decisions regarding network construction can affect our understanding of network properties  
96 (Brimacombe *et al.* 2023). There is currently no metadata standard that could guide the documentation of all  
97 types of probabilistic interactions (although see e.g., Salim *et al.* 2022 who discuss data standards for  
98 deterministic mutualistic webs). Well-defined metadata for probabilistic interactions would support more  
99 adequate manipulation and integration of interaction data from different sources and guard against possible  
100 misinterpretations arising from ambiguous definitions of probabilistic interaction networks. These metadata  
101 should outline the nature (i.e., local or regional) and type (e.g., predatory or pollination) of the interactions,  
102 provide information regarding the taxonomic level, identities, and characteristics (e.g., life stages) of the  
103 individuals involved in an interaction, present the mathematical formulation of probabilities, including clearly  
104 identified conditional variables (e.g., spatial and temporal scales), and describe the methods and contexts (e.g.,  
105 location, time, environmental conditions) in which interactions were estimated. Inadequately documented  
106 probabilistic interaction data should be used with caution when analyzing ecological webs. Our observations  
107 and advice can be applied to most types of ecological networks representing direct interactions, from food webs  
108 to host-virus networks. Even though the measurement of interaction probabilities may differ between network  
109 types (e.g., estimations often based on frequencies of interactions in plant-pollinator networks), the broad  
110 principles underlying our findings remain relevant and applicable across diverse ecological contexts.

## 111 **Probabilistic representations of interactions**

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

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

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

148 **Pairwise interactions: the building blocks of ecological networks**

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

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

178 them to interact locally.

179 Even though binary interaction webs constitute a highly valuable source of ecological information (Pascual *et*  
180 *al.* 2006), they overlook important factors regarding interaction strengths. Represented in a quantitative  
181 adjacency matrix  $W$  as numbers not confined to the  $[0, 1]$  range, interaction strengths better describe the energy  
182 flows, demographic impacts or frequencies of interactions between nodes (Berlow *et al.* 2004; Borrett &  
183 Scharler 2019), with  $W_{i \rightarrow j}$  being a natural  $\mathbb{N}$  or real  $\mathbb{R}$  number depending on the measure. For example, they  
184 may represent local interaction rates between pairs of taxa (e.g., the flower-visiting rates of pollinators in a  
185 mutualistic network, Herrera 1989). When interaction strengths characterize predation pressure on prey taxa in  
186 food webs, they can serve as good estimators of the parameters describing species interactions in a  
187 Lotka-Volterra model (e.g., Emmerson & Raffaelli 2004). The extra amount of ecological information in  
188 quantitative networks typically comes at a cost of greater sampling effort and data volume in predictive models  
189 (Strydom *et al.* 2021), which can lead to relatively high levels of uncertainties when inferring quantitative webs  
190 with limited data.

191 Just like binary interaction networks, the uncertainty and spatiotemporal variability of interaction strengths can  
192 be represented probabilistically. However, the need to estimate the probability distribution of all possible values  
193 of interaction strengths can make the inference of probabilities more challenging in quantitative webs compared  
194 to binary interaction webs, which require only one probability estimate for each interaction. Interaction  
195 strengths can follow various probability distributions depending on the measure used. For instance, they can  
196 follow a Poisson distribution  $W_{i \rightarrow j} \sim \text{Poisson}(\lambda_{i \rightarrow j})$  when predicting frequencies of interactions between pairs  
197 of nodes, with  $\lambda_{i \rightarrow j}$  being the expected rate at which individuals of taxa  $i$  and  $j$  interact (e.g., the average  
198 number of prey  $j$  consumed by all predators  $i$  in a given time period). The Poisson distribution can also be  
199 0-inflated after initially modeling non-interacting taxa (e.g., Boulangeat *et al.* 2012 employ a 0-inflated model  
200 to analyze species abundance following the modeling of species presence and absence), which constitute the  
201 majority of taxa pairs in most local webs due to their typically high sparseness (Jordano 2016). Because of the  
202 methodological difficulties typically encountered when building deterministic quantitative webs (which are only  
203 partially mitigated by models such as Ecopath, Plagányi & Butterworth 2004), binary interaction webs, which  
204 are easier to sample (Jordano 2016) and predict (Strydom *et al.* 2021), have been more frequently studied and  
205 modeled. Moreover, most published networks of probabilistic interactions and methods describe probabilistic  
206 interactions whose outcome is binary (whether interaction probabilities are regarded as constant or variable,  
207 e.g. represented by a Beta distribution), which underlines the need for better guidelines regarding the

208 interpretation and manipulation of these types of webs first. For these reasons, our primary focus in this  
209 contribution will be on addressing the challenges in interpreting and using interaction probabilities in Bernoulli  
210 distributions, in both local webs and metawebs.

## 211 Local webs: communities interacting in space and time

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

220 We define space as the collection of geographic coordinates  $(x, y, z)$ , with  $(x, y)$  representing longitude and  
221 latitude coordinates, and  $z$  denoting either altitudes or depths. These point coordinates delineate the spatial  
222 boundaries of the system, which may be portrayed as a polyhedron. Ecological interactions may exhibit  
223 variations along latitudinal and altitudinal gradients, as evidenced by changes in hummingbird-plant  
224 interactions (Weinstein & Graham 2017a, b) and mosquito biting rates (e.g., Kulkarni *et al.* 2006) at different  
225 elevations. In contrast, time is defined as the specific time period within which interactions were either observed  
226 or predicted. Even though space and time are continuous variables that should yield probability *densities* of  
227 interactions (i.e., relative likelihoods of interactions occurring at infinitesimal locations and instants in time),  
228 these definitions enable them to be conceptualized as distinct patches and time segments. Treating space and  
229 time as discrete dimensions aligns with the common sampling methods of ecological networks and provides  
230 actual probabilities of interactions, which can be obtained by integrating probability densities over space and  
231 time. Furthermore, we can quantify both an area  $A$  and a duration  $t$ , which can be readily used in spatiotemporal  
232 analyses of ecological networks. For example, when studying network-area relationships (NAR, Galiana *et al.*  
233 2018), we anticipate that local probabilities of interactions scale positively with area and duration because taxa  
234 have more opportunities to interact as these dimensions expand.

235 The probability that two taxa  $i$  and  $j$  interact in a local web  $L_{x,y,z,t}$  (spatial and temporal subscripts hereafter

omitted or replaced by the shorter subscript  $l$  for clarity) can be conditioned on many environmental and biological factors. In addition to network area (or volume) and duration, they may be conditioned on taxa co-occurrence  $X_{i,j,l}$ , which is usually a Boolean describing if the geographic distributions of both taxa overlap within the study area. Co-occurrence may be modeled probabilistically, in which case it may conform to a Bernoulli distribution  $X_{i,j,l} \sim \text{Bernoulli}(P(X_{i,l}, X_{j,l}))$ , where  $X_{i,l}$  and  $X_{j,l}$  are the local occurrences (presence / absence) of both taxa. The probability of co-occurrence  $P(X_{i,l}, X_{j,l})$  can be estimated through the application of joint species distribution models (e.g., Pollock *et al.* 2014), potentially taking into account biotic interactions (Staniczenko *et al.* 2017). Given that the probability that two non-co-occurring taxa interact locally is zero (i.e.,  $P(L_{i \rightarrow j}|X_{i,j,l} = 0) = 0$ ), the probability of local interaction can be obtained by multiplying the probability of interaction given co-occurrence with the probability of co-occurrence:

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

Other important factors that can impact interaction probabilities at the local scale are taxa local abundances  $N_{i,l}$  and  $N_{j,l}$ , which affect encounter probabilities (Canard *et al.* 2012), and local traits distributions  $T_{i,l}$  and  $T_{j,l}$  (e.g., body mass, presence of given anatomical features, host resistance), which determine the ability of individuals to interact after encountering each other (Poisot *et al.* 2015). Moreover, local interactions may be conditional on local environmental factors such as temperature (Angilletta *et al.* 2004), precipitation (Woodward *et al.* 2012), habitat structure (Klecka & Boukal 2014), and presence of other taxa in the network (Kéfi *et al.* 2012; Pilosof *et al.* 2017), as described above. Here, we use the variable  $E_l$  to describe the local environmental context in which interaction probabilities were estimated. For example, if a research team conducts a mesocosm experiment to estimate interaction probabilities between predators and prey with and without shelters (a place that offers refuge and protection for prey, shielding them from predators),  $E_l$  would represent the presence or absence of these shelters. Like co-occurrence,  $E_l$  can also be modeled probabilistically when the stochasticity or uncertainty of environmental factors is considered. In sum,  $E_l$  represents all environmental variables that were taken into consideration when measuring interaction probabilities and is, therefore, a subset of all environmental factors impacting ecological interactions. Finally, local interaction probabilities may be conditioned on higher-level properties of the network, which we denote by  $f(L)$ . Many topological null models (i.e., statistical models that randomize interactions by retaining certain properties of the network while excluding others) provide interaction probabilities from selected measures of network structure, such as connectance (Fortuna & Bascompte 2006) and the degree distribution (Bascompte *et al.* 2003).

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

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

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

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

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

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

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

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

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

## Metawebs: regional catalogs of interactions

Metawebs are networks of potential interactions that have been designed for broad spatial, temporal, and taxonomic scales (e.g., species food webs at the continental scale). Potential interactions describe the biological capacity of taxa to interact, which is typically assessed at the regional scale. Metawebs of probabilistic interactions are particularly useful in situations where there is high uncertainty in the ability of taxa to interact if they were to encounter each other. This uncertainty frequently arises due to insufficient interaction data, especially for taxa that have not yet been observed to co-occur, and uncertainties in trait-matching models. As data accumulates, interactions in metawebs should tend towards binarity, approaching probability values of 0

304 (repeatedly failing to observe an interaction between two co-occurring taxa) and 1 (observing an interaction at  
305 least once). The extent of sampling effort thus influences our evaluation of probabilities of potential  
306 interactions, as sampling over a larger area or for a longer duration enables us to capture a greater number of  
307 regional interactions (McLeod *et al.* 2021). However, in contrast with local webs of probabilistic interactions,  
308 which describe interaction stochasticity in local natural settings, regional interactions are not evaluated for any  
309 particular local context. Although *neutrally* forbidden interactions (i.e., improbable interactions between rare  
310 species, Canard *et al.* 2012) tend to have low probability values in local webs, they may have higher probabilities  
311 in the metaweb (i.e., when species' biological traits can support an interaction if they were to encounter each  
312 other). Likewise, non-co-occurring taxa may have a non-zero probability of interaction in the metaweb.

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

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

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

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

324 For example, let  $i$  be a western diamondback rattlesnake (*Crotalus atrox*) and  $j$ , a wood lemming (*Myopus*  
325 *schisticolor*). These two taxa never co-occur, the rattlesnake being adapted to warm regions of North America  
326 (Castoe *et al.* 2007) and the lemming, to northern habitats of Eurasia (Fedorov *et al.* 2008). As we lack direct  
327 observations of an interaction between these two species, we have to rely on expert knowledge or trait-matching

328 models to estimate their probability of potential interaction. To accurately estimate this probability using  
329 trait-matching models, it is crucial to ensure that the set of traits considered reflects the overall traits  
330 distributions of both taxa. We could for instance consider their average body mass and the average phylogenetic  
331 proximity of lemmings to rattlesnakes' prey. Doing so, we might find a high probability of potential interaction  
332 based on these traits.

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

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

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

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

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

351 Starting from a selected set of taxa, which are usually distributed within a broad region of interest, metawebs

352 can be built using different data sources, including literature review (e.g., Maiorano *et al.* 2020), aggregated  
353 local web data (e.g., Gravel *et al.* 2019; Saravia *et al.* 2022), trait-matching models Strydom *et al.* (2022), and  
354 expert knowledge. Every pair of taxa that have confidently been observed to interact at least once can be given a  
355 probability of 1 (i.e.,  $P(M_{i \rightarrow j}) = 1$ ) since we know that they *can* interact. This differs from local webs of  
356 probabilistic interactions, where interaction events may remain stochastic (i.e.,  $P(L_{i \rightarrow j}) < 1$ ) even after  
357 empirically observing interactions due to their spatiotemporal variability. Interactions that were never observed  
358 typically have low probability values in local webs and vary from low to high values in metawebs, contingent  
359 upon taxa traits distributions (reaching 0 for forbidden links).

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

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

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

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

375 **Properties of probabilistic interaction webs**

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

387 **Application 1: Dissimilarity of binary and probabilistic interaction webs**

388 We use the collection of tripartite host-parasite networks sampled across Europe, created by Kopelke *et al.*  
389 (2017), in the following case studies. This dataset contains well-resolved binary local interactions between  
390 willows (52 species), willow-galling sawflies (96 species), and their parasitoids (126 species). Given its  
391 replicated webs spanning large spatiotemporal scales, this dataset is well-suited for analyzing a variety of  
392 ecological hypotheses and processes. Out of a total of 374 local webs, we retained those containing at least 5  
393 species, resulting in a set of 233 georeferenced local webs (networks sampled within areas of 0.1 to 0.3 km<sup>2</sup>  
394 during June and/or July spanning 29 years). We built a metaweb of binary interactions by aggregating all local  
395 interactions, which gave us a regional web composed of 274 species and 1080 interactions. In Fig. 1, we show  
396 how the dissimilarity between the metaweb and the aggregated local webs varies with the number of sampled  
397 local webs. To do so, we randomly selected one local web of binary interactions and sequentially sampled  
398 additional webs while aggregating both their species and interactions. Next, we compared the metaweb and the  
399 aggregated local webs using the dissimilarity of interactions between common species ( $\beta_{OS}$ , Fig. 1a) and the  
400 dissimilarity in species composition ( $\beta_S$ , Fig. 1b) (Poisot *et al.* 2012). We repeated this sampling process one  
401 hundred times and highlighted the median dissimilarity values across simulations, as well as the 50% and 95%  
402 percentile intervals. This shows that networks of local interactions are highly dissimilar from the metaweb, both

403 in terms of species and interactions, especially when only a limited number of sites has been sampled. Both  
404 dissimilarity indices were calculated based on the number of items shared by the two webs ( $c_{LM}$ ) and the  
405 number of items unique to the metaweb ( $u_M$ ) and to the aggregated local web ( $u_L$ ). The  $\beta_S$  dissimilarity index  
406 uses species (nodes) as items being compared, while the  $\beta_{OS}$  index assesses dissimilarity based on interactions  
407 between shared species (Poisot *et al.* 2012). Both indices were calculated following the  $\beta_W$  index of Whittaker  
408 (1960):

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

409 [Figure 1 about here.]

410 Next, we investigate the differences in scaling between networks of probabilistic or binary interactions  
411 (Fig. 1c-d). We converted these binary interaction networks into probabilistic ones using models based on  
412 simple assumptions. Our models are not designed to estimate the exact values of probabilistic interactions.  
413 Instead, their purpose is to create plausible networks that serve as illustrative examples to highlight distinctions  
414 between local webs and metawebs of probabilistic interactions. We created two metawebs of probabilistic  
415 interactions by employing constant false positive and false negative rates for all regional interactions. In the first  
416 metaweb, we set both false positive and false negative rates to zero to prevent artificially inflating the total  
417 number of links, enabling a more accurate comparison with binary interaction webs. This gave us a probability  
418 of regional interaction of 1 when at least one interaction has been observed locally and of 0 in the absence of  
419 any observed interaction between a given pair of species (Fig. 1c-d). In the second metaweb (used in the next  
420 applications), we introduced a 5% false positive rate to account for spurious interactions and a 10% false  
421 negative rate to address the elevated occurrence of missing interactions in ecological networks (Catchen *et al.*  
422 2023). We believe these rates represent reasonable estimates of missing and spurious potential interactions, but  
423 confirming their accuracy is challenging due to the unavailability of data on the actual feasibility of interaction.  
424 Observed interactions were thus given a probability of regional interaction of 95%, whereas unobserved ones  
425 were assigned a probability of 10%.

426 To build local webs of probabilistic interactions, we first recognize that local interactions must initially be  
427 biologically feasible before occurring at a specific time and space. A local probability of interaction  $P(L_{i \rightarrow j})$   
428 can be expressed as the product of the probability of local interaction given that the two taxa can potentially

429 interact  $P(L_{i \rightarrow j}|M_{i \rightarrow j} = 1)$ , which we sometimes denote as  $P(L_{i \rightarrow j}|M_{i \rightarrow j})$  for the sake of simplicity, with their  
 430 probability of regional interaction  $P(M_{i \rightarrow j})$ :

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

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

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

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

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

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

453 as the number of sampled webs increases.

## 454 Application 2: Spatial and temporal scaling of interactions

455 The investigation of network-area relationships and interaction accumulation curves is an important area of  
456 research in network ecology. First, network-area relationships elucidate the scaling of network properties (such  
457 as modularity and trophic chain lengths) with spatial scale (e.g., Galiana *et al.* 2018; Wood *et al.* 2015). The  
458 variations in network structure across spatial scales may stem from the scaling of species richness (species-area  
459 relationships, SARs), the number of interactions (Brose *et al.* 2004), and many other higher-level properties of  
460 the system (e.g., environmental heterogeneity, Thompson & Townsend 2005) with the sampled area.

461 Additionally, ecological processes occurring at distinct spatial scales, such as the spatial variability in local  
462 community composition resulting from different sequences of extinction and colonization events, can also  
463 contribute to this variation (Galiana *et al.* 2018). Next, interaction accumulation curves describe the scaling of  
464 the number of observed interactions with sampling effort (Jordano 2016). Sampling effort, which may  
465 correspond to the duration of the sampling period used to construct the network, can impact connectance  
466 (Bersier *et al.* 1999) and various measures of network structure (Banašek-Richter *et al.* 2004; McLeod *et al.*  
467 2021). Apart from sampling effort, the temporal scaling of interactions also elucidates how network structure  
468 changes with the temporal resolution of the network (Poisot *et al.* 2012), acknowledging that distinct  
469 interactions take place over time, ranging from short-term fluctuations of interactions to long-term trends. As  
470 local webs of probabilistic interactions may explicitly account for the spatiotemporal variability of interactions  
471 (as shown in ), they offer a distinct approach to investigating the scaling of network structure with space and  
472 time, in contrast to webs of binary and quantitative interactions, by making the stochasticity of interactions the  
473 focal point of the modeling process.

474 Local webs and metawebs exhibit distinct relationships with spatial and temporal scales. On one hand,  
475 metawebs of probabilistic interactions, representing biological feasibility, feature regional interactions that do  
476 not scale with space and time. This is because regional interactions depend solely on the biological capacity of  
477 two taxa to interact, regardless of their co-occurrence and local environmental conditions. However,  
478 probabilities of potential interactions may change (tending to become more definitive) upon updating previous  
479 estimates with increased sampling effort, even though they do not vary in a specific direction with the spatial  
480 and temporal extent (boundaries) of the network. The probability of two taxa potentially interacting should  
481 theoretically be the same in all metawebs in which they are present, provided that the data and methods used for

482 estimation are consistent. For example, if a smaller metaweb  $M_1$  is derived from a larger metaweb  $M_0$  by  
483 selecting the subset of taxa present in the region described by  $M_1$  and retaining all their interactions, their  
484 probabilities of interaction should be identical regardless of scale, i.e.  $P(M_{1,i \rightarrow j}) = P(M_{0,i \rightarrow j})$ . With a larger or  
485 equal number of taxa in  $M_0$  as compared to  $M_1$ , the total number of interactions is expected to be higher or at  
486 least equal in the larger web, even though pairwise probabilities remain identical.

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

501 [Figure 2 about here.]

## 502 **Box 1: A spatiotemporally explicit model of interactions**

503 Predicting local webs across time and space is a pivotal goal of network ecology, especially given the scarcity of  
504 interaction data (Strydom *et al.* 2021). Ecologists may resort to predictive models (e.g., generative Bayesian  
505 and machine learning models) to reconstruct networks at fine spatial and temporal scales with limited  
506 interaction data. For instance, real-time biomonitoring data coupled with appropriate numerical models (Bohan  
507 *et al.* 2017) can be employed to reconstruct local ecological webs, opening avenues for in-depth studies on local  
508 ecosystem functioning and dynamics. The probabilistic representation of interactions describes the inherent  
509 uncertainty in these models, typically expressed through probability distributions. We introduce and develop a

510 simple generative mechanistic model for probabilistic local interactions that takes into consideration their  
511 spatiotemporal variability (i.e., a spatiotemporally explicit model of local interactions). It is essential to note  
512 that our model is not designed for regional interactions, which are scale-independent. Rather, it could prove  
513 valuable for predicting local interactions across time and space by generating new interaction data following  
514 parameter inference.

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

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

523 When  $\gamma > 1$ , it signifies a positive association in the geographic distributions of both taxa, indicating that the  
524 presence of one taxon enhances the probability of occurrence of the other. Attractions may be the result of  
525 positive interactions (e.g., mutualism) or positive/negative interactions (e.g., antagonism between predators and  
526 prey, Cazelles *et al.* 2016). In empirical webs,  $\gamma > 1$  holds true for the majority of species pairs (Catchen *et al.*  
527 2023). In contrast, repulsions ( $\gamma < 1$ ) may be caused e.g. by strong interspecific competition (Cazelles *et al.*  
528 2016). We model the co-occurrence  $X_{i,j,l}$  of both taxa as the outcome of a Bernoulli trial

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

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

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

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

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

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

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

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

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

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

547 **Application 3: Taxonomic scaling of interactions**

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

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

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

567 where  $g_{1,i}$  and  $g_{2,j}$  are the species of the corresponding genus and assuming independence between  
568 species-level interactions. If it is known that at least two of these species interact (i.e.,  $P(S_{g_{1,i} \rightarrow g_{2,j}}) = 1$  for at  
569 least one pair of  $(g_{1,i}, g_{2,j})$ ), it implies a probability of genus interaction equal to 1. Canard *et al.* (2012) built a  
570 species-based network following a similar approach, by using simulated interactions between individuals  
571 derived from a neutral model (i.e., a model that assumed ecological equivalence among individuals). In  
572 contrast, a more sophisticated approach is necessary when transitioning from a broader to a finer level of  
573 organization. This is because knowledge of an interaction between two genera does not guarantee that all

574 possible pairwise combinations of their species will also interact. One possible method is to build a finer-scale  
575 network by generating probabilities of interactions through random sampling from a beta distribution,  
576 parameterized by the broader-scale network.

577 Ideally, our biological interpretation of probabilistic interactions should remain consistent across a network  
578 even if it incorporates heterogeneous levels of organization, such as a network whose nodes represent both  
579 species and trophic species (groups of species sharing similar predators and prey). This is common in  
580 ecological webs where taxonomic resolution is typically low (Hemprich-Bennett *et al.* 2021; Vázquez *et al.*  
581 2022). Interaction probabilities at broader taxonomic scales should be based on probabilities of interactions  
582 between individuals, for both local webs and metawebs. For instance, in local individual-based food webs, the  
583 probability that two individuals interact reflects our degree of belief that one individual will consume the other.  
584 Similarly, in local species-based food webs, the probability that two species interact represents our degree of  
585 belief that at least one individual from the predator species consumes at least another individual from the prey  
586 species. Furthermore, in local clade-based food webs, the probability that two clades interact represents our  
587 degree of belief that at least two species from these clades interact with each other or, equivalently, that at least  
588 two individuals from these clades interact with each other. Fundamentally, the taxonomic scaling of interactions  
589 involves aggregating interactions between individuals into larger groups, which may exhibit varying degrees of  
590 homogeneity based on the organisms and the study system. In that regard, taxonomic scaling is analogous to the  
591 spatial and temporal scaling of interactions, as they all represent different ways to aggregate individuals into  
592 broader groups (either spatially, temporally, or taxonomically).

#### 593 Application 4: Sampling for binary interaction webs

594 The prediction of binary interactions through Bernoulli trials is an important application of probabilistic  
595 interaction webs. This approach proves beneficial for analyzing the structural characteristics of probabilistic  
596 interaction webs, particularly in the absence of specific analytical formulas for measures describing the structure  
597 of the complete web (Poisot *et al.* 2016). By performing independent Bernoulli trials for each interaction, a  
598 network of binary interactions may be generated. A probability distribution of network properties can be  
599 obtained by measuring network structure across multiple randomly generated networks (Poisot *et al.* 2016).  
600 This method enables the representation of the variability of network structure, albeit with possible biases when  
601 connectance is low (Chagnon 2015; Poisot & Gravel 2014). Employing this strategy to generate binary  
602 interaction networks under a null model facilitates null hypothesis significance testing, wherein the observed

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

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

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

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

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

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

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

In Fig. 4, we compare the average connectance of binary interaction networks resulting from these two sampling techniques. We drew regional and local interactions from our host-parasite networks of probabilistic interactions, generating a number of binary interaction web realizations for each site in the dataset. These two

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

636 [Figure 4 about here.]

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

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

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

#### 647 Application 5: Prediction of local webs from metawebs

648 Metawebs serve as a valuable source of ecological information for predicting local webs across time and space.  
649 Local webs of binary interactions can be reconstructed by selecting a subset of taxa and interactions from the  
650 metaweb (Dunne 2006). This implies that metawebs consistently contain more interactions than their  
651 corresponding local webs, even though their connectance is usually much smaller than the ones of local webs

652 (Gravel *et al.* 2011). Determining the list of taxa to select can be achieved empirically or through numerical  
 653 methods like range maps or species distribution models. As species composition is arguably less difficult to  
 654 sample or predict than pairwise interactions, the primary challenge lies in deciding which interactions to select  
 655 from the metaweb, a task that may necessitate advanced statistical models and ecological expertise. Inferring  
 656 the structure of local webs from the metaweb before predicting specific local pairwise interactions could hold  
 657 promise (Strydom *et al.* 2021), considering that the structure of local webs is constrained by the metaweb  
 658 (Saravia *et al.* 2022).

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

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

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

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

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

676 pairwise interactions within a network. Using simple models of  $P(L_{i \rightarrow j}|M_{i \rightarrow j})$ , as demonstrated in our case  
677 studies, represents an initial step toward the overarching objective of reconstructing local webs from metawebs.

## 678 Conclusion

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

691 Estimating local interaction probabilities independently for each taxa pair and assembling them into a network  
692 of probabilistic interactions comes with limitations. Predicting local networks of binary interactions based on  
693 these interaction probabilities assumes independence among interactions, a condition seldom respected in  
694 practice (Golubski & Abrams 2011). Relaxing this assumption is the next logical step in the stochastic  
695 representation of interactions. A more accurate representation of the stochasticity of ecological networks  
696 involves creating *probabilistic networks* ( $P(L|...)$  and  $P(M|...)$ ), rather than networks of *probabilistic*  
697 *interactions* ( $P(L_{i \rightarrow j}|...)$  and  $P(M_{i \rightarrow j}|...)$ ). Probabilistic networks describe the probability that a particular  
698 network of binary (or quantitative) interactions (its whole adjacency matrix) is realized. For example, Young *et*  
699 *al.* (2021) used a Bayesian approach to estimate the probability of different plant-pollinator network structures  
700 derived from imperfect observational data. A probability distribution of ecological networks may also be  
701 derived using the principle of maximum entropy given some structural constrained (e.g., Cimini *et al.* 2019;  
702 Park & Newman 2004). Regardless of the method employed, generating probabilistic local webs, bypassing the  
703 need to independently estimate local interaction probabilities, could lead to more accurate predictions of local

704 webs. Furthermore, probabilistic networks could serve as an alternative to null hypothesis significance testing  
705 when comparing the structure of a local web to some random expectations or, as done in Pellissier *et al.* (2018),  
706 to the metaweb. These random expectations are typically derived by performing a series of Bernoulli trials on  
707 probabilistic interactions, assuming independence, to generate a distribution of networks of binary interactions  
708 and then calculate their structure (Poisot *et al.* 2016). One could for instance compare the likelihood of an  
709 observed network to the one of the most likely network structure according to the probabilistic network  
710 distribution, thereby directly obtaining a measure of discrepancy of the empirical network. Generating  
711 probabilistic ecological networks represents a tangible challenge, one that, in the coming years, promises to  
712 unlock doors to more advanced and adequate analyses of ecological networks.

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

731 **Acknowledgment**

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

739 **References**

- 740 Abrego, N., Roslin, T., Huotari, T., Ji, Y., Schmidt, N.M., Wang, J., *et al.* (2021). [Accounting for species](#)  
741 [interactions is necessary for predicting how arctic arthropod communities respond to climate change.](#)  
742 *Ecography*, 44, 885–896.
- 743 Andrade-Pacheco, R., Rerolle, F., Lemoine, J., Hernandez, L., Meité, A., Juziwelo, L., *et al.* (2020). [Finding](#)  
744 [hotspots: Development of an adaptive spatial sampling approach.](#) *Scientific Reports*, 10, 10939.
- 745 Angilletta, M.J., Jr., Steury, T.D. & Sears, M.W. (2004). [Temperature, Growth Rate, and Body Size in](#)  
746 [Ectotherms: Fitting Pieces of a Life-History Puzzle](#)<sup>1</sup>. *Integrative and Comparative Biology*, 44, 498–509.
- 747 Araujo, M.B., Rozenfeld, A., Rahbek, C. & Marquet, P.A. (2011). [Using species co-occurrence networks to](#)  
748 [assess the impacts of climate change.](#) *Ecography*, 34, 897–908.
- 749 Banašek-Richter, C., Cattin, M.-F. & Bersier, L.-F. (2004). [Sampling effects and the robustness of quantitative](#)  
750 [and qualitative food-web descriptors.](#) *Journal of Theoretical Biology*, 226, 23–32.
- 751 Bartomeus, I., Gravel, D., Tylianakis, J.M., Aizen, M.A., Dickie, I.A. & Bernard-Verdier, M. (2016). [A](#)  
752 [common framework for identifying linkage rules across different types of interactions.](#) *Functional Ecology*,  
753 30, 1894–1903.
- 754 Bascompte, J., Jordano, P., Melián, C.J. & Olesen, J.M. (2003). [The nested assembly of plant–animal](#)  
755 [mutualistic networks.](#) *Proceedings of the National Academy of Sciences*, 100, 9383–9387.

- 756 Berg, M.P. & Ellers, J. (2010). Trait plasticity in species interactions: A driving force of community dynamics.  
757 *Evolutionary Ecology*, 24, 617–629.
- 758 Berlow, E.L., Neutel, A.-M., Cohen, J.E., De Ruiter, P.C., Ebenman, B., Emmerson, M., *et al.* (2004).  
759 Interaction strengths in food webs: Issues and opportunities. *Journal of Animal Ecology*, 73, 585–598.
- 760 Bersier, L.-F., Dixon, P. & Sugihara, G. (1999). Scale-Invariant or Scale-Dependent Behavior of the Link  
761 Density Property in Food Webs: A Matter of Sampling Effort? *The American Naturalist*, 153, 676–682.
- 762 Blanchet, F.G., Cazelles, K. & Gravel, D. (2020). Co-occurrence is not evidence of ecological interactions.  
763 *Ecology Letters*, 23, 1050–1063.
- 764 Bohan, D.A., Vacher, C., Tamaddoni-Nezhad, A., Raybould, A., Dumbrell, A.J. & Woodward, G. (2017).  
765 Next-Generation Global Biomonitoring: Large-scale, Automated Reconstruction of Ecological Networks.  
766 *Trends in Ecology & Evolution*, 32, 477–487.
- 767 Bolnick, D.I., Amarasekare, P., Araújo, M.S., Bürger, R., Levine, J.M., Novak, M., *et al.* (2011). Why  
768 intraspecific trait variation matters in community ecology. *Trends in Ecology & Evolution*, 26, 183–192.
- 769 Borrett, S.R. & Scharler, U.M. (2019). Walk partitions of flow in Ecological Network Analysis: Review and  
770 synthesis of methods and indicators. *Ecological Indicators*, 106, 105451.
- 771 Boulangeat, I., Gravel, D. & Thuiller, W. (2012). Accounting for dispersal and biotic interactions to disentangle  
772 the drivers of species distributions and their abundances. *Ecology Letters*, 15, 584–593.
- 773 Brimacombe, C., Bodner, K., Michalska-Smith, M., Poisot, T. & Fortin, M.-J. (2023). Shortcomings of reusing  
774 species interaction networks created by different sets of researchers. *PLOS Biology*, 21, e3002068.
- 775 Broom, M. & Ruxton, G.D. (2005). You can run—or you can hide: Optimal strategies for cryptic prey against  
776 pursuit predators. *Behavioral Ecology*, 16, 534–540.
- 777 Brose, U., Ostling, A., Harrison, K. & Martinez, N.D. (2004). Unified spatial scaling of species and their  
778 trophic interactions. *Nature*, 428, 167–171.
- 779 Burnham, K.P. & Anderson, D.R. (2004). Multimodel Inference: Understanding AIC and BIC in Model  
780 Selection. *Sociological Methods & Research*, 33, 261–304.
- 781 Canard, E.F., Mouquet, N., Mouillot, D., Stanko, M., Miklisova, D. & Gravel, D. (2014). Empirical Evaluation  
782 of Neutral Interactions in Host-Parasite Networks. *The American Naturalist*, 183, 468–479.

- 783 Canard, E., Mouquet, N., Marescot, L., Gaston, K.J., Gravel, D. & Mouillot, D. (2012). [Emergence of](#)  
784 [Structural Patterns in Neutral Trophic Networks](#). *PLOS ONE*, 7, e38295.
- 785 Castoe, T.A., Spencer, C.L. & Parkinson, C.L. (2007). [Phylogeographic structure and historical demography of](#)  
786 [the western diamondback rattlesnake \(\*Crotalus Atrox\*\): A perspective on North American desert](#)  
787 [biogeography](#). *Molecular Phylogenetics and Evolution*, 42, 193–212.
- 788 Catchen, M.D., Poisot, T., Pollock, L.J. & Gonzalez, A. (2023). [The missing link: Discerning true from false](#)  
789 [negatives when sampling species interaction networks](#).
- 790 Cazelles, K., Araujo, M.B., Mouquet, N. & Gravel, D. (2016). [A theory for species co-occurrence in interaction](#)  
791 [networks](#). *Theoretical Ecology*, 9, 39–48.
- 792 Chagnon, P.-L. (2015). [Characterizing topology of ecological networks along gradients: The limits of metrics'](#)  
793 [standardization](#). *Ecological Complexity*, 22, 36–39.
- 794 Choh, Y., Ignacio, M., Sabelis, M.W. & Janssen, A. (2012). [Predator-prey role reversals, juvenile experience](#)  
795 [and adult antipredator behaviour](#). *Scientific Reports*, 2, 728.
- 796 Cimini, G., Squartini, T., Saracco, F., Garlaschelli, D., Gabrielli, A. & Caldarelli, G. (2019). [The statistical](#)  
797 [physics of real-world networks](#). *Nature Reviews Physics*, 1, 58–71.
- 798 Cirtwill, A.R., Eklöf, A., Roslin, T., Wootton, K. & Gravel, D. (2019). [A quantitative framework for](#)  
799 [investigating the reliability of empirical network construction](#). *Methods in Ecology and Evolution*, 10,  
800 902–911.
- 801 Dallas, T., Park, A.W. & Drake, J.M. (2017). [Predicting cryptic links in host-parasite networks](#). *PLOS*  
802 [Computational Biology, 13, e1005557.](#)
- 803 Dunne, J.A. (2006). The Network Structure of Food Webs. In: *Ecological networks: Linking structure and*  
804 *dynamics* (eds. Dunne, J.A. & Pascual, M.). Oxford University Press, pp. 27–86.
- 805 Elmasri, M., Farrell, M.J., Davies, T.J. & Stephens, D.A. (2020). [A hierarchical Bayesian model for predicting](#)  
806 [ecological interactions using scaled evolutionary relationships](#). *The Annals of Applied Statistics*, 14,  
807 221–240.
- 808 Elton, C.S. (2001). *Animal Ecology*. University of Chicago Press, Chicago, IL.
- 809 Emmerson, M.C. & Raffaelli, D. (2004). [Predator–prey body size, interaction strength and the stability of a real](#)  
810 [food web](#). *Journal of Animal Ecology*, 73, 399–409.

- 811 Fedorov, V.B., Goropashnaya, A.V., Boeskorov, G.G. & Cook, J.A. (2008). Comparative phylogeography and  
812 demographic history of the wood lemming (*Myopus schisticolor*): Implications for late Quaternary history  
813 of the taiga species in Eurasia. *Molecular Ecology*, 17, 598–610.
- 814 Fortuna, M.A. & Bascompte, J. (2006). Habitat loss and the structure of plant–animal mutualistic networks.  
815 *Ecology Letters*, 9, 281–286.
- 816 Fu, X., Seo, E., Clarke, J. & Hutchinson, R.A. (2021). Link Prediction Under Imperfect Detection:  
817 Collaborative Filtering for Ecological Networks. *IEEE Transactions on Knowledge and Data Engineering*,  
818 33, 3117–3128.
- 819 Galiana, N., Lurgi, M., Claramunt-López, B., Fortin, M.-J., Leroux, S., Cazelles, K., et al. (2018). The spatial  
820 scaling of species interaction networks. *Nature Ecology & Evolution*, 2, 782–790.
- 821 Gilman, S.E., Urban, M.C., Tewksbury, J., Gilchrist, G.W. & Holt, R.D. (2010). A framework for community  
822 interactions under climate change. *Trends in Ecology & Evolution*, 25, 325–331.
- 823 Godsoe, W., Murray, R. & Iritani, R. (2022). Species interactions and diversity: A unified framework using Hill  
824 numbers. *Oikos*, n/a, e09282.
- 825 Golubski, A.J. & Abrams, P.A. (2011). Modifying modifiers: What happens when interspecific interactions  
826 interact? *Journal of Animal Ecology*, 80, 1097–1108.
- 827 Gomez, J.M., Verdu, M. & Perfectti, F. (2010). Ecological interactions are evolutionarily conserved across the  
828 entire tree of life. *Nature*, 465, 918–U6.
- 829 Gonzalez, A. & Londoño, M.C. (2022). Monitor biodiversity for action. *Science*, 378, 1147–1147.
- 830 Gonzalez-Varo, J.P. & Traveset, A. (2016). The Labile Limits of Forbidden Interactions. *Trends in Ecology &*  
831 *Evolution*, 31, 700–710.
- 832 Gravel, D., Albouy, C. & Thuiller, W. (2016). The meaning of functional trait composition of food webs for  
833 ecosystem functioning. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 371,  
834 20150268.
- 835 Gravel, D., Baiser, B., Dunne, J.A., Kopelke, J.-P., Martinez, N.D., Nyman, T., et al. (2019). Bringing Elton  
836 and Grinnell together: A quantitative framework to represent the biogeography of ecological interaction  
837 networks. *Ecography*, 42, 401–415.

- 838 Gravel, D., Massol, F., Canard, E., Mouillot, D. & Mouquet, N. (2011). [Trophic theory of island biogeography](#).  
839 *Ecology Letters*, 14, 1010–1016.
- 840 Gravel, D., Poisot, T., Albouy, C., Velez, L. & Mouillot, D. (2013). [Inferring food web structure from](#)  
841 [predator–prey body size relationships](#). *Methods in Ecology and Evolution*, 4, 1083–1090.
- 842 Guimarães, P.R. (2020). [The Structure of Ecological Networks Across Levels of Organization](#). *Annual Review*  
843 *of Ecology, Evolution, and Systematics*, 51, 433–460.
- 844 Guimerà, R. & Sales-Pardo, M. (2009). [Missing and spurious interactions and the reconstruction of complex](#)  
845 [networks](#). *Proceedings of the National Academy of Sciences*, 106, 22073–22078.
- 846 Hemprich-Bennett, D.R., Oliveira, H.F.M., Le Comber, S.C., Rossiter, S.J. & Clare, E.L. (2021). [Assessing the](#)  
847 [impact of taxon resolution on network structure](#). *Ecology*, 102, e03256.
- 848 Herrera, C.M. (1989). [Pollinator abundance, morphology, and flower visitation rate: Analysis of the “quantity”](#)  
849 [component in a plant-pollinator system](#). *Oecologia*, 80, 241–248.
- 850 Hortal, J., de Bello, F., Diniz-Filho, J.A.F., Lewinsohn, T.M., Lobo, J.M. & Ladle, R.J. (2015). [Seven Shortfalls](#)  
851 [that Beset Large-Scale Knowledge of Biodiversity](#). *Annual Review of Ecology, Evolution, and Systematics*,  
852 46, 523–549.
- 853 Jordano, P. (1987). [Patterns of Mutualistic Interactions in Pollination and Seed Dispersal: Connectance,](#)  
854 [Dependence Asymmetries, and Coevolution](#). *The American Naturalist*, 129, 657–677.
- 855 Jordano, P. (2016). [Sampling networks of ecological interactions](#). *Functional Ecology*, 30, 1883–1893.
- 856 Jordano, P., Bascompte, J. & Olesen, J.M. (2003). [Invariant properties in coevolutionary networks of](#)  
857 [plant–animal interactions](#). *Ecology Letters*, 6, 69–81.
- 858 Kausrud, K.L., Mysterud, A., Steen, H., Vik, J.O., Østbye, E., Cazelles, B., *et al.* (2008). [Linking climate](#)  
859 [change to lemming cycles](#). *Nature*, 456, 93–97.
- 860 Kéfi, S., Berlow, E.L., Wieters, E.A., Joppa, L.N., Wood, S.A., Brose, U., *et al.* (2015). [Network structure](#)  
861 [beyond food webs: Mapping non-trophic and trophic interactions on Chilean rocky shores](#). *Ecology*, 96,  
862 291–303.
- 863 Kéfi, S., Berlow, E.L., Wieters, E.A., Navarrete, S.A., Petchey, O.L., Wood, S.A., *et al.* (2012). [More than a](#)  
864 [meal... integrating non-feeding interactions into food webs](#). *Ecology Letters*, 15, 291–300.

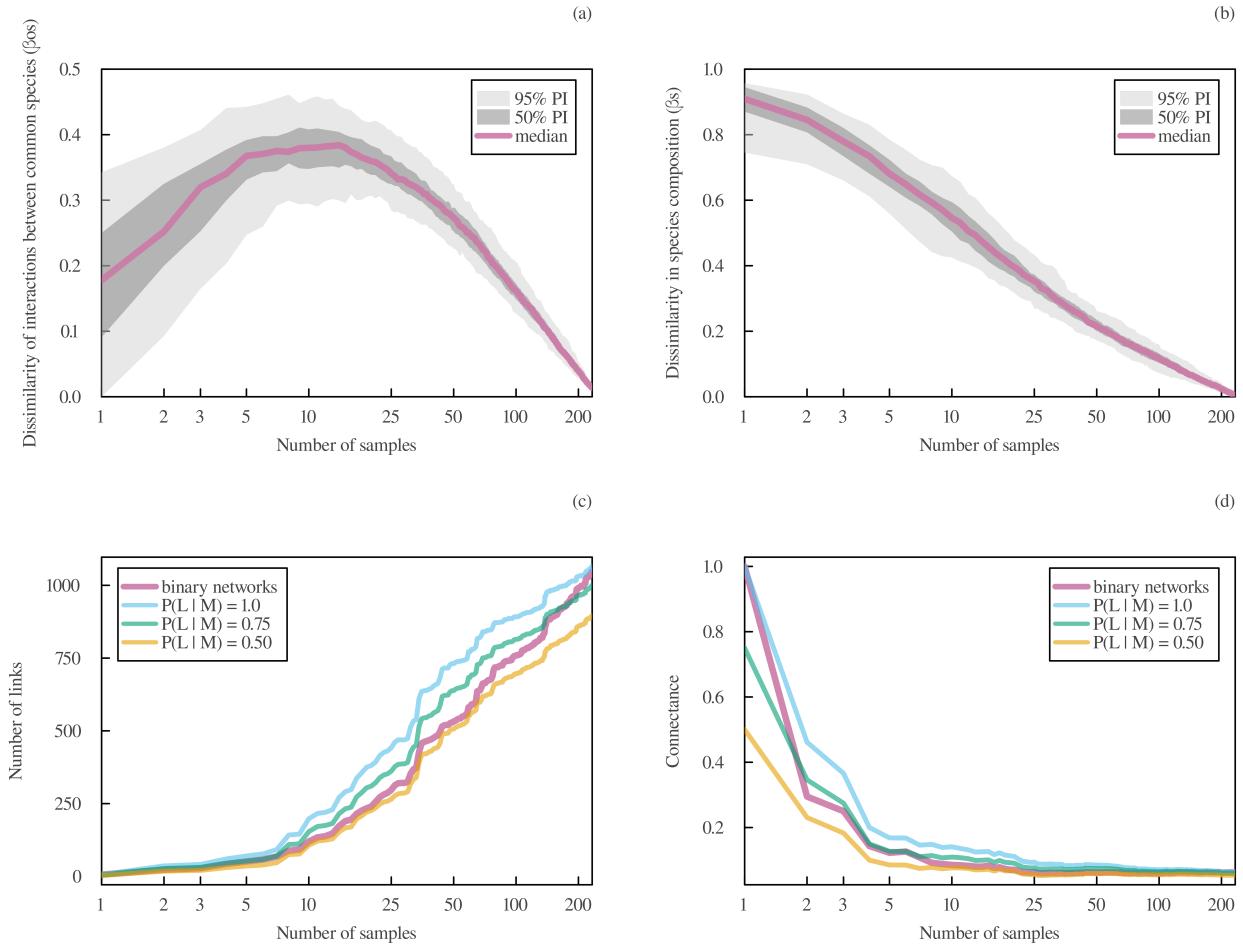
- 865 Kéfi, S., Miele, V., Wieters, E.A., Navarrete, S.A. & Berlow, E.L. (2016). How Structured Is the Entangled  
866 Bank? The Surprisingly Simple Organization of Multiplex Ecological Networks Leads to Increased  
867 Persistence and Resilience. *PLOS Biology*, 14, e1002527.
- 868 Kissner, K.J., Forbes, M.R. & Secoy, D.M. (1997). Rattling Behavior of Prairie Rattlesnakes (*Crotalus viridis*  
869 *viridis*, *Viperidae*) in Relation to Sex, Reproductive Status, Body Size, and Body Temperature. *Ethology*,  
870 103, 1042–1050.
- 871 Klecka, J. & Boukal, D.S. (2014). The effect of habitat structure on prey mortality depends on predator and  
872 prey microhabitat use. *Oecologia*, 176, 183–191.
- 873 Kopelke, J.-P., Nyman, T., Cazelles, K., Gravel, D., Vissault, S. & Roslin, T. (2017). Food-web structure of  
874 willow-galling sawflies and their natural enemies across Europe. *Ecology*, 98, 1730–1730.
- 875 Kulkarni, M.A., Kweka, E., Nyale, E., Lyatuu, E., Mosha, F.W., Chandramohan, D., *et al.* (2006).  
876 Entomological Evaluation of Malaria Vectors at Different Altitudes in Hai District, Northeastern Tanzania.  
877 *Journal of Medical Entomology*, 43, 580–588.
- 878 Maiorano, L., Montemaggioli, A., Ficetola, G.F., O'Connor, L. & Thuiller, W. (2020). TETRA-EU 1.0: A  
879 species-level trophic metaweb of European tetrapods. *Global Ecology and Biogeography*, 29, 1452–1457.
- 880 McCann, K. (2007). Protecting biostructure. *Nature*, 446, 29–29.
- 881 McCann, K.S. (2011). Food webs (MPB-50). In: *Food Webs (MPB-50)*. Princeton University Press.
- 882 McLeod, A., Leroux, S.J., Gravel, D., Chu, C., Cirtwill, A.R., Fortin, M.-J., *et al.* (2021). Sampling and  
883 asymptotic network properties of spatial multi-trophic networks. *Oikos*, n/a.
- 884 McLeod, A.M., Leroux, S.J. & Chu, C. (2020). Effects of species traits, motif profiles, and environment on  
885 spatial variation in multi-trophic antagonistic networks. *Ecosphere*, 11, e03018.
- 886 Melián, C.J., Vilas, C., Baldó, F., González-Ortegón, E., Drake, P. & Williams, R.J. (2011). Eco-evolutionary  
887 Dynamics of Individual-Based Food Webs. In: *Advances in Ecological Research*, The Role of Body Size in  
888 Multispecies Systems (ed. Belgrano, A.). Academic Press, pp. 225–268.
- 889 Ohlmann, M., Miele, V., Dray, S., Chalmandrier, L., O'Connor, L. & Thuiller, W. (2019). Diversity indices for  
890 ecological networks: A unifying framework using Hill numbers. *Ecology Letters*, 22, 737–747.
- 891 Olesen, J.M., Bascompte, J., Dupont, Y.L., Elberling, H., Rasmussen, C. & Jordano, P. (2010). Missing and  
892 forbidden links in mutualistic networks. *Proceedings of the Royal Society B: Biological Sciences*, 278,

- 893 725–732.
- 894 Park, J. & Newman, M.E.J. (2004). [Statistical mechanics of networks](#). *Physical Review E*, 70, 066117.
- 895 Pascual, M., Dunne, J.A. & Dunne, J.A. (2006). *Ecological Networks: Linking Structure to Dynamics in Food*  
896 *Webs*. Oxford University Press, USA.
- 897 Pellissier, L., Albouy, C., Bascompte, J., Farwig, N., Graham, C., Loreau, M., *et al.* (2018). [Comparing species](#)  
898 [interaction networks along environmental gradients](#). *Biological Reviews*, 93, 785–800.
- 899 Petchey, O.L., Beckerman, A.P., Riede, J.O. & Warren, P.H. (2008). [Size, foraging, and food web structure](#).  
900 *Proceedings of the National Academy of Sciences*, 105, 4191–4196.
- 901 Pilosof, S., Porter, M.A., Pascual, M. & Kéfi, S. (2017). [The multilayer nature of ecological networks](#). *Nature*  
902 *Ecology & Evolution*, 1, 1–9.
- 903 Plagányi, É.E. & Butterworth, D.S. (2004). [A critical look at the potential of Ecopath with ecosim to assist in](#)  
904 [practical fisheries management](#). *African Journal of Marine Science*, 26, 261–287.
- 905 Poisot, T., Canard, E., Mouillot, D., Mouquet, N. & Gravel, D. (2012). [The dissimilarity of species interaction](#)  
906 [networks](#). *Ecology Letters*, 15, 1353–1361.
- 907 Poisot, T., Cirtwill, A.R., Cazelles, K., Gravel, D., Fortin, M.-J. & Stouffer, D.B. (2016). [The structure of](#)  
908 [probabilistic networks](#). *Methods in Ecology and Evolution*, 7, 303–312.
- 909 Poisot, T. & Gravel, D. (2014). [When is an ecological network complex? Connectance drives degree](#)  
910 [distribution and emerging network properties](#). *PeerJ*, 2, e251.
- 911 Poisot, T., Stouffer, D.B. & Gravel, D. (2015). [Beyond species: Why ecological interaction networks vary](#)  
912 [through space and time](#). *Oikos*, 124, 243–251.
- 913 Polis, G.A. (1991). [Complex Trophic Interactions in Deserts: An Empirical Critique of Food-Web Theory](#). *The*  
914 *American Naturalist*, 138, 123–155.
- 915 Pollock, L.J., Tingley, R., Morris, W.K., Golding, N., O’Hara, R.B., Parris, K.M., *et al.* (2014). [Understanding](#)  
916 [co-occurrence by modelling species simultaneously with a Joint Species DistributionModel \(JSDM\)](#).  
917 *Methods in Ecology and Evolution*, 5, 397–406.
- 918 Pomeranz, J.P.F., Thompson, R.M., Poisot, T. & Harding, J.S. (2019). [Inferring predator–prey interactions in](#)  
919 [food webs](#). *Methods in Ecology and Evolution*, 10, 356–367.

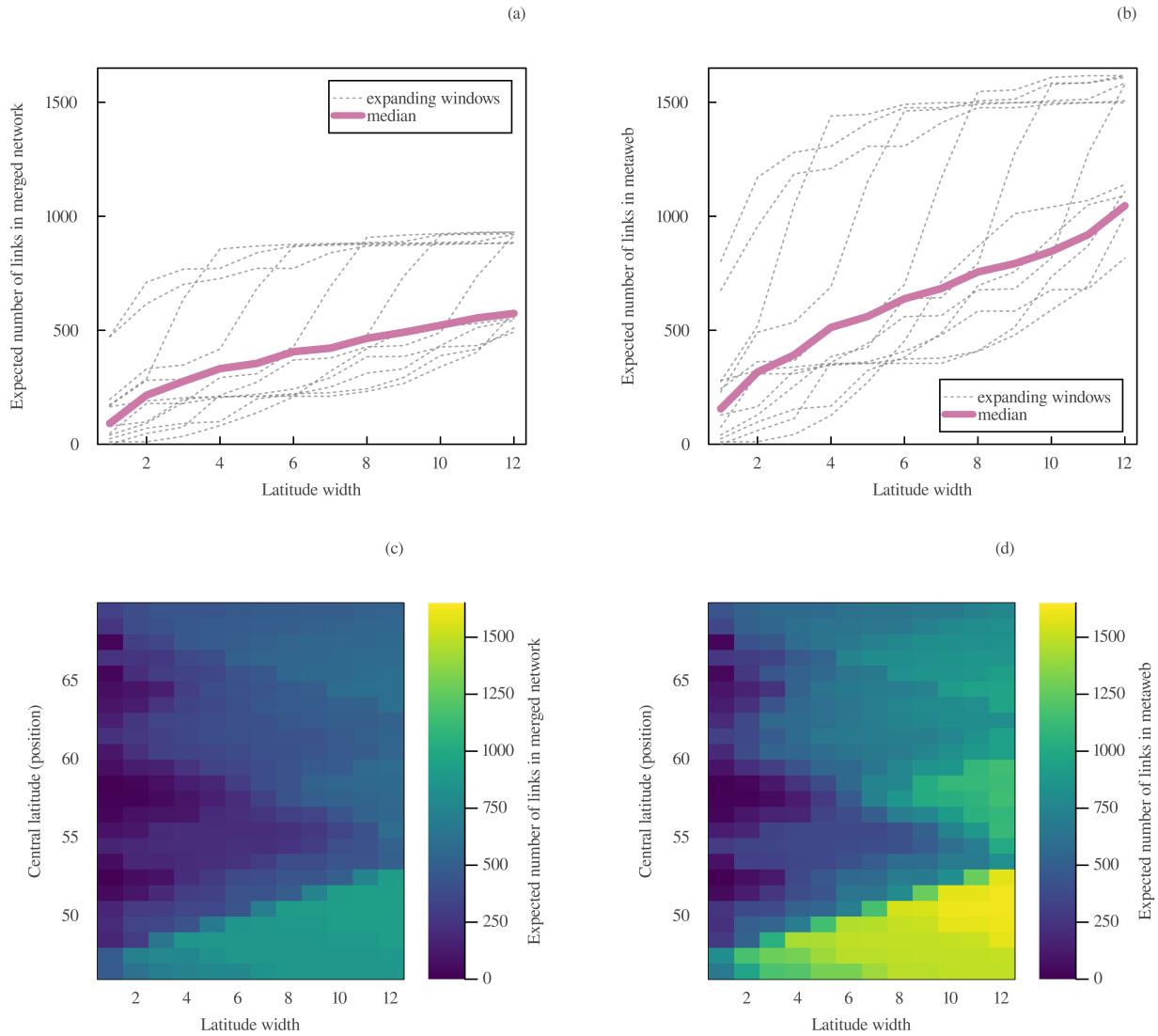
- 920 Pringle, R.M. & Hutchinson, M.C. (2020). [Resolving Food-Web Structure](#). *Annual Review of Ecology, Evolution, and Systematics*, 51, 55–80.
- 921
- 922 Proulx, S.R., Promislow, D.E.L. & Phillips, P.C. (2005). [Network thinking in ecology and evolution](#). *Trends in Ecology & Evolution*, SPECIAL ISSUE: BUMPER BOOK REVIEW, 20, 345–353.
- 923
- 924 Pulliam, H.R. (1974). [On the Theory of Optimal Diets](#). *The American Naturalist*, 108, 59–74.
- 925 Rooney, N. & McCann, K.S. (2012). [Integrating food web diversity, structure and stability](#). *Trends in Ecology & Evolution*, 27, 40–46.
- 926
- 927 Salim, J.A., Saraiva, A.M., Zermoglio, P.F., Agostini, K., Wolowski, M., Drucker, D.P., et al. (2022). [Data standardization of plant–pollinator interactions](#). *GigaScience*, 11, giac043.
- 928
- 929 Sanders, D. & van Veen, F.J.F. (2012). [Indirect commensalism promotes persistence of secondary consumer species](#). *Biology Letters*, 8, 960–963.
- 930
- 931 Saravia, L.A., Marina, T.I., Kristensen, N.P., De Troch, M. & Momo, F.R. (2022). [Ecological network assembly: How the regional metaweb influences local food webs](#). *Journal of Animal Ecology*, 91, 630–642.
- 932
- 933 Schmeller, D.S., Julliard, R., Bellingham, P.J., Böhm, M., Brummitt, N., Chiarucci, A., et al. (2015). [Towards a global terrestrial species monitoring program](#). *Journal for Nature Conservation*, 25, 51–57.
- 934
- 935 Shaw, J.O., Dunhill, A.M., Beckerman, A.P., Dunne, J.A. & Hull, P.M. (2024). [A framework for reconstructing ancient food webs using functional trait data](#).
- 936
- 937 Singer, M.C. & McBride, C.S. (2012). [Geographic mosaics of species' association: A definition and an example driven by plant–insect phenological synchrony](#). *Ecology*, 93, 2658–2673.
- 938
- 939 Staniczenko, P.P.A., Sivasubramaniam, P., Suttle, K.B. & Pearson, R.G. (2017). [Linking macroecology and community ecology: Refining predictions of species distributions using biotic interaction networks](#). *Ecology Letters*, 20, 693–707.
- 940
- 941
- 942 Stouffer, D.B., Rezende, E.L. & Amaral, L.A.N. (2011). [The role of body mass in diet contiguity and food-web structure](#). *Journal of Animal Ecology*, 80, 632–639.
- 943
- 944 Strydom, T., Bouskila, S., Banville, F., Barros, C., Caron, D., Farrell, M.J., et al. (2022). [Food web reconstruction through phylogenetic transfer of low-rank network representation](#). *Methods in Ecology and Evolution*, 13.
- 945
- 946

- 947 Strydom, T., Bouskila, S., Banville, F., Barros, C., Caron, D., Farrell, M.J., *et al.* (2023). Graph embedding and  
948 transfer learning can help predict potential species interaction networks despite data limitations. *Methods in*  
949 *Ecology and Evolution*, 14, 2917–2930.
- 950 Strydom, T., Catchen, M.D., Banville, F., Caron, D., Dansereau, G., Desjardins-Proulx, P., *et al.* (2021). A  
951 roadmap towards predicting species interaction networks (across space and time). *Philosophical*  
952 *Transactions of the Royal Society B-Biological Sciences*, 376, 20210063.
- 953 Thompson, R.M. & Townsend, C.R. (2005). Food-Web Topology Varies with Spatial Scale in a Patchy  
954 Environment. *Ecology*, 86, 1916–1925.
- 955 Tylianakis, J.M., Tscharntke, T. & Lewis, O.T. (2007). Habitat modification alters the structure of tropical  
956 host-parasitoid food webs. *Nature*, 445, 202–205.
- 957 Vázquez, D.P., Melián, C.J., Williams, N.M., Blüthgen, N., Krasnov, B.R. & Poulin, R. (2007). Species  
958 abundance and asymmetric interaction strength in ecological networks. *Oikos*, 116, 1120–1127.
- 959 Vázquez, D.P., Peralta, G., Cagnolo, L. & Santos, M. (2022). Ecological interaction networks. What we know,  
960 what we don't, and why it matters. *Ecología Austral*, 32, 670–697.
- 961 Wagenmakers, E.-J. & Farrell, S. (2004). AIC model selection using Akaike weights. *Psychonomic Bulletin &*  
962 *Review*, 11, 192–196.
- 963 Weinstein, B.G. & Graham, C.H. (2017a). On comparing traits and abundance for predicting species  
964 interactions with imperfect detection. *Food Webs*, 11, 17–25.
- 965 Weinstein, B.G. & Graham, C.H. (2017b). Persistent bill and corolla matching despite shifting temporal  
966 resources in tropical hummingbird-plant interactions. *Ecology Letters*, 20, 326–335.
- 967 Whittaker, R.H. (1960). Vegetation of the Siskiyou Mountains, Oregon and California. *Ecological*  
968 *Monographs*, 30, 279–338.
- 969 Wood, S.A., Russell, R., Hanson, D., Williams, R.J. & Dunne, J.A. (2015). Effects of spatial scale of sampling  
970 on food web structure. *Ecology and Evolution*, 5, 3769–3782.
- 971 Woodward, G., Benstead, J.P., Beveridge, O.S., Blanchard, J., Brey, T., Brown, L.E., *et al.* (2010). Chapter 2 -  
972 Ecological Networks in a Changing Climate. In: *Advances in Ecological Research*, Ecological Networks  
973 (ed. Woodward, G.). Academic Press, pp. 71–138.

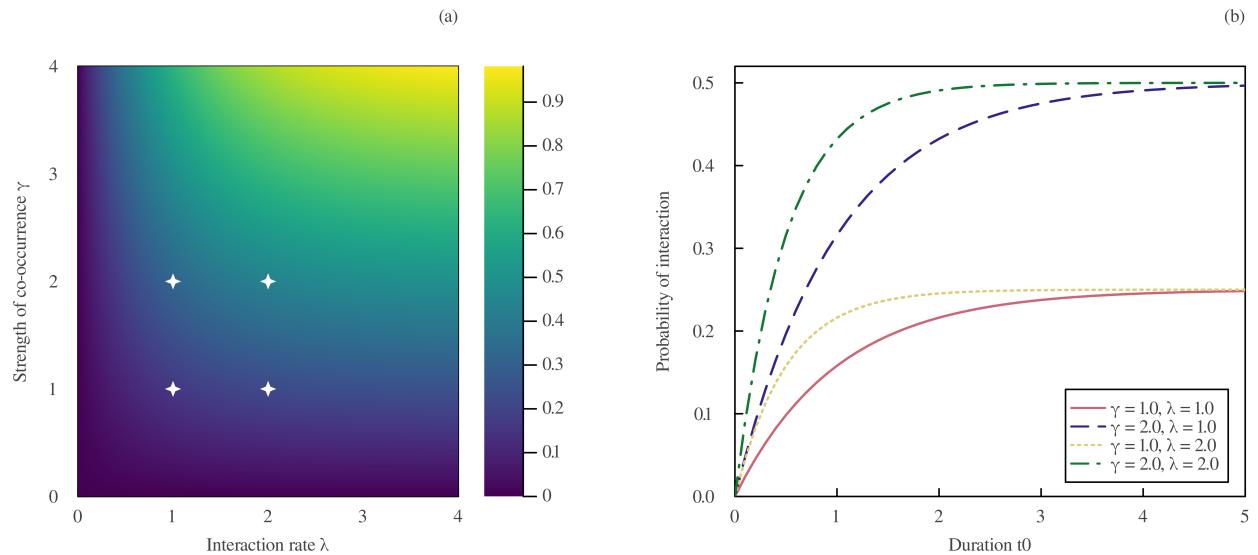
- 974 Woodward, G., Brown, L.E., Edwards, F.K., Hudson, L.N., Milner, A.M., Reuman, D.C., *et al.* (2012). [Climate](#)  
975 [change impacts in multispecies systems: Drought alters food web size structure in a field experiment.](#)
- 976 *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367, 2990–2997.
- 977 Young, J.-G., Valdovinos, F.S. & Newman, M.E.J. (2021). [Reconstruction of plant–pollinator networks from](#)  
978 [observational data](#). *Nature Communications*, 12, 3911.



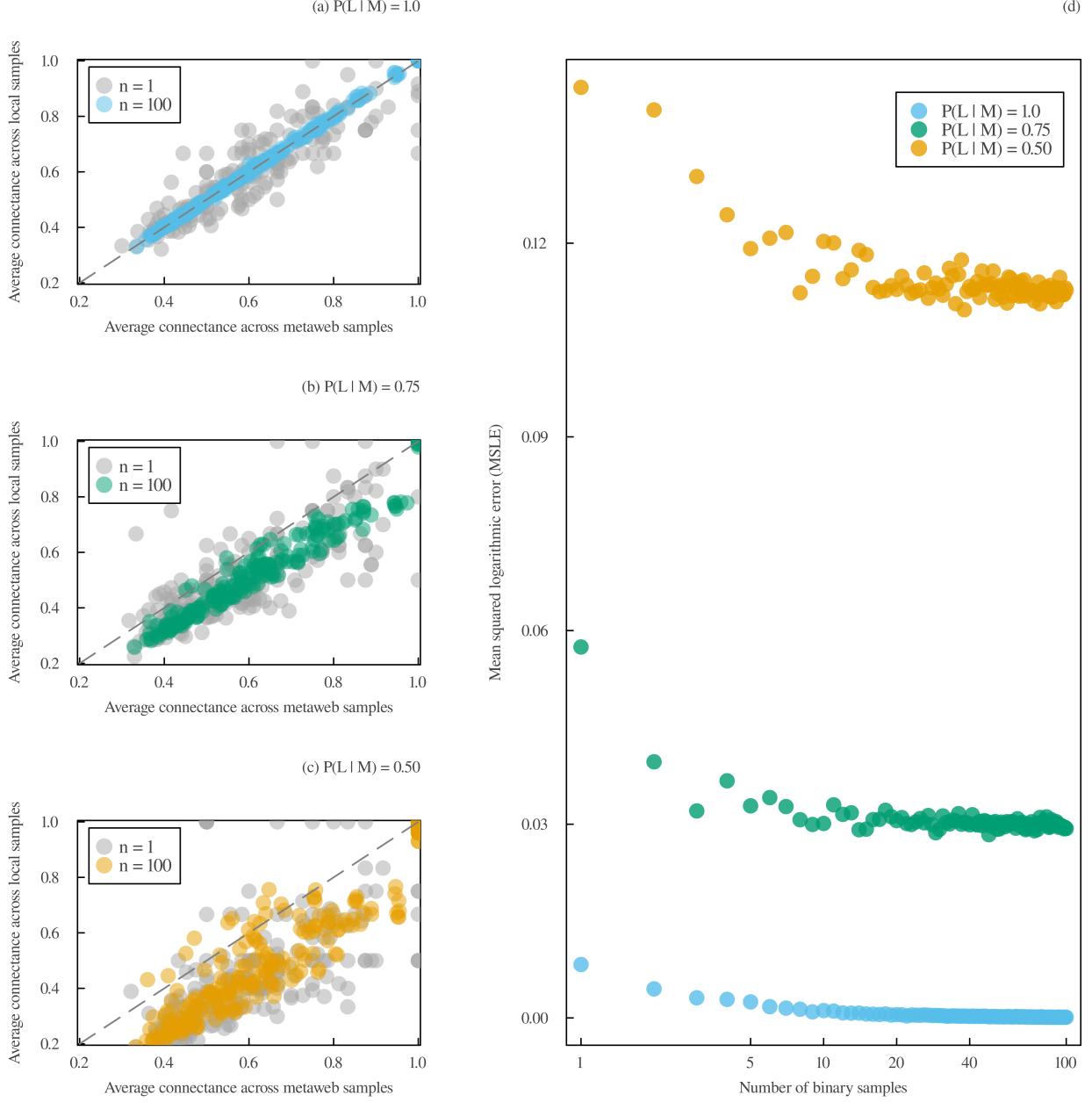
**Figure 1: Network accumulation curves.** (a) Dissimilarity of interactions between common species and (b) dissimilarity in species composition between aggregated local webs and the metaweb of binary host-parasite interactions. Aggregated local webs were obtained by sequentially and randomly selecting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



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



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



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