

# The ecological interpretation of probabilistic networks

Francis Banville<sup>1,2,3</sup> Tanya Strydom<sup>1,3</sup> Penelope S. A. Blyth<sup>4</sup> Chris Brimacombe<sup>5</sup> Michael Catchen<sup>3,6</sup>  
Gabriel Dansereau<sup>1,3</sup> 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 occurring?) as opposed to deterministically (are they occurring?) allows a better assessment of their spatiotemporal variability and accounts for inherent uncertainties in both observations and predictions. However, despite this growing interest in probabilistic networks, guidelines for the estimation and documentation of probabilistic interaction data are still lacking. This is concerning given that our biological understanding of probabilistic interactions and their appropriate analyses depend on the often elusive methods and variables used for their estimation. 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. These definitions are based on the distinction between the observation of an interaction at a specific time and space and its biological feasibility. We show that different network representations have different statistical properties when it comes to common ecological applications. Specifically, we argue that local probabilistic networks and metawebs differ in their spatial and temporal scaling of interactions, with potential interactions in metawebs remaining constant across spatial and temporal scales. This is in contrast with the taxonomic scaling of interactions, which does not qualitatively differ between both types of networks. We suggest two approaches to inferring deterministic networks from probabilistic webs that account for these differences and argue that systematic biases arise when directly inferring local networks from subsets of metawebs. To support our arguments, we build a spatiotemporally explicit model of probabilistic interactions and develop different case studies using empirical data on host-parasite interactions in Europe. Overall, our results underscore the importance of clear metadata for

probabilistic ecological networks, both at the local and regional scales, to inform the appropriate reuse of interaction data.

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



Last revision: *December 5, 2023*

## **1 Introduction**

2 As we try to navigate global biodiversity change (decline?), filling in knowledge gaps about biodiversity  
3 becomes instrumental to monitoring and mitigating those changes (Gonzalez & Londoño (2022), Abrego *et al.*  
4 (2021)). However, cataloging species, populations and, in particular, ecological interactions is a substantial  
5 challenge. There are methodological and biological constraints that hinder our ability to observe all  
6 interactions, such as the spatial and temporal uncoupling of species (Jordano (1987)) and the large number of  
7 possible interactions in a community, of which the vast majority are rare (Jordano (2016)). More generally, a  
8 handful of conditions must be satisfied for an interaction to be observed locally. First, both species must have  
9 overlapping geographic ranges, i.e. they must co-occur within the region of interest (Blanchet *et al.* (2020)).  
10 Then, they must encounter locally. Probabilities of interspecific encounters are typically low, especially for rare  
11 species with low relative abundances (Canard *et al.* (2012)). Finally, their traits must be locally compatible  
12 (Poisot *et al.* (2015)). This includes their phenology (Olesen *et al.* (2010), Singer & McBride (2012)),  
13 behavioral choices (Pulliam (1974), Choh *et al.* (2012)), phenotypes (Bolnick *et al.* (2011), Stouffer *et al.*  
14 (2011), Gravel *et al.* (2013)) and even the presence or abundance of a third species (e.g., of a more profitable  
15 prey species) (Golubski & Abrams (2011), Sanders & van Veen (2012)). Documenting the location and timing  
16 of interactions becomes even more difficult when accounting for the spatiotemporal variability of ecological  
17 networks (Poisot *et al.* (2012), Poisot *et al.* (2015)). Environmental factors, such as temperature (Angilletta *et*  
18 *al.* (2004)), drought (Woodward *et al.* (2012)), climate change (Gilman *et al.* (2010), Woodward *et al.* (2010),  
19 Araujo *et al.* (2011)), and habitat modifications (Tylianakis *et al.* (2007)), contribute to this spatiotemporal  
20 variability of interactions by impacting species abundance and traits. In this context, it is unsurprising that our  
21 knowledge of ecological interactions remains limited (Hortal *et al.* (2015)) despite extensive biodiversity data  
22 collection (Schmeller *et al.* (2015)).

23 Knowing the biological capacity of two species to interact is necessary but not sufficient for inferring their  
24 interaction at a specific time and space. The recognition of the intrinsic variability of species interactions has  
25 led ecologists to expand their representation of ecological networks to include a probabilistic view of  
26 interactions (Poisot *et al.* (2016), Dallas *et al.* (2017), Fu *et al.* (2021)). This different perspective allows us to  
27 fill in the Eltonian shortfall by modeling the probability of detecting interactions, which can be an important  
28 tool for directing efforts and take action, especially in places where access and resources for research are scarce.  
29 As opposed to binary deterministic networks, in which interactions are regarded as either occurring or not,

30 probabilistic networks represent our degree of belief about the realization or feasibility of pairwise interactions  
31 at the local or regional scale, respectively. In other words, representing interactions probabilistically takes into  
32 account inherent uncertainties and variability associated with ecological data. In the broadest sense, binary  
33 networks are also a type of probabilistic network, in which the numerical value of an interaction is restrained to  
34 0 (non-occurring) or 1 (occurring). In probabilistic networks, only forbidden interactions (i.e., interactions  
35 prohibited by biological traits, Jordano *et al.* (2003), Olesen *et al.* (2010)) have a probability value of zero,  
36 provided that intraspecific trait variability is considered (Gonzalez-Varo & Traveset (2016)). However, *neutrally*  
37 forbidden interactions (i.e., improbable interactions between rare species, Canard *et al.* (2012)) could have low  
38 probability values in local networks but high probability in a regional network (metaweb) describing the  
39 biological capacity of species to interact.

40 By accounting for the uncertainty of interactions, probabilistic networks provide a more realistic portrait of  
41 species interactions and network structure (i.e. community-level properties), which is a major driver of the  
42 functioning, dynamics, and resilience of ecosystems worldwide (Proulx *et al.* (2005), McCann (2007), McCann  
43 (2011), Rooney & McCann (2012)). Moreover, the application and development of computational methods in  
44 network ecology, often based on a probabilistic representation of interactions, can alleviate (and guide) the  
45 sampling efforts required to document species interactions (Strydom *et al.* (2021)). For example, statistical  
46 models can be used to estimate the uncertainty of pairwise interactions (Cirtwill *et al.* (2019)) and the  
47 probability of missing (false negatives) and spurious (false positives) interactions (Guimerà & Sales-Pardo  
48 (2009)). Considering the high rate of false negatives in species interaction data due to the difficulty of  
49 witnessing rare interactions (Catchen *et al.* (2023)), these models could inform the identification of priority  
50 sampling locations of ecological networks where data collection would yield the most valuable information,  
51 thereby reducing errors. Optimization models for sampling locations have mostly found applications in  
52 biological systems that are not networked, such as identifying priority sampling sites for disease hotspots  
53 (Andrade-Pacheco *et al.* (2020)), but there is substantial promise in applying them to probabilistic ecological  
54 interactions. Statistical models can also be used to generate network predictions without prior knowledge of  
55 pairwise interactions, for instance using body size (Petchey *et al.* (2008), Gravel *et al.* (2013)), phylogeny  
56 (Elmasri *et al.* (2020), Strydom *et al.* (2022)), or a combination of niche and neutral processes (Bartomeus *et*  
57 *al.* (2016), Pomeranz *et al.* (2019)) for inference. Topological null models (e.g., Bascompte *et al.* (2003),  
58 Fortuna & Bascompte (2006)), which can be used to generate underlying distributions of network measures for  
59 null hypothesis significance testing, are other examples of common probabilistic network models. Many

measures have been developed to describe the structure (Poisot *et al.* (2016)) and diversity (Ohlmann *et al.* (2019), Godsoe *et al.* (2022)) of probabilistic networks. These models and measures support the use of this approach for the study of a wide range of ecological questions, from making better predictions of species distribution (Cazelles *et al.* (2016)) to forecasting the impact of climate change on ecological networks (Gilman *et al.* (2010)).

The lack of clear guidelines on the use of probabilistic interaction data is worrisome (Brimacombe *et al.* (2023)), both for data producers and re-users who generate and manipulate these numbers. Besides methodological difficulties that may arise when assessing probabilistic interactions, a precise definition of probabilistic interactions appears to be lacking, making the estimation and use of these data more difficult. We aim to take a step back by outlining different ways in which probabilistic interactions are defined and used in network ecology. We distinguish two broad categories of probabilistic networks that necessitate distinct approaches when applied to key ecological questions: local networks of realized interactions, and regional networks (metawebs) of potential interactions. We highlight the distinctions in the ecological meaning of these two representations and show that they yield different statistical outcomes regarding e.g. the spatial and temporal scaling of interactions and the prediction of binary networks across space. Moreover, there is currently no metadata standard that could guide the documentation of all types of probabilistic interactions (Salim *et al.* (2022) discuss data standards for deterministic mutualistic networks). Well-defined metadata for probabilistic networks would support more adequate manipulation and integration of interaction data from different sources and guard against the potential misinterpretations arising from ambiguous definitions of probabilistic networks. These metadata should outline the nature (i.e., potential or local) and type (e.g., predatory or pollination) of the interactions, provide information regarding the taxonomic level, identities, and characteristics (e.g., life stages) of the individuals involved in an interaction, present the mathematical formulation of probabilities, including clearly identified conditional variables, and describe the methods and contexts (e.g., location, time, environmental conditions) in which interactions were estimated. Inadequately documented probabilistic interaction data should be used with caution when analyzing ecological networks. Although our contribution focuses on food webs, our observations and advice can be applied to other types of ecological networks, from plant-pollinator to host-virus networks. Indeed, excluding networks of indirect interactions such as competition and facilitation networks (Kéfi *et al.* (2015), Kéfi *et al.* (2016)), most ecological networks describe probabilities of direct interactions, which are conceptually and mathematically analogous regardless of their biological type (e.g., trophic and mutualistic interactions).

## 90 Probabilistic representations of interactions

91 One of the first aspects to take into consideration when estimating or interpreting probabilities of interactions is  
92 knowing if they describe potential or realized interactions. A potential interaction is defined as the biological  
93 capacity of two taxa to interact (i.e., the probability that they *can* theoretically interact) whereas a realized  
94 interaction is the materialization or observation of this interaction in a well-defined space and time (i.e., the  
95 probability that they interact locally). Here, we use the terms *metaweb* (Dunne (2006)) to designate networks of  
96 potential interactions and *local networks* (Poisot *et al.* (2012)) for those of realized interactions. Metawebs are  
97 the network analogues of the species pool, where local networks originate from a subset of both species (nodes)  
98 and interactions (edges) of the regional metaweb (Saravia *et al.* (2022)). Without clear documentation, it can be  
99 challenging to know if published probabilistic networks describe potential or realized interactions, or if  
100 so-called probabilities are in reality *interaction scores* (i.e., non-probabilistic quantitative interactions). When  
101 probabilistic potential interactions are used and interpreted as local interactions (and conversely), this may  
102 generate misleading findings during data analysis. We believe that a better understanding of the differences,  
103 similarities, and relationships between these two probabilistic representations of ecological networks would  
104 alleviate interpretation errors and facilitate a more adequate utilization of interaction data.

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

106 Local ecological networks and metawebs, like any type of network, are made of nodes and edges that can be  
107 represented at different levels of organization and precision. The basic unit of food webs and other ecological  
108 networks are individuals that interact with each other (e.g., by predation, Elton (2001)), forming  
109 individual-based networks (Melián *et al.* (2011)). The aggregation of these individuals into more or less  
110 homogeneous groups (e.g., populations, species, trophic species, families) allows us to represent nodes at  
111 broader taxonomic scales, which affects our interpretation of the properties of these systems (Guimarães  
112 (2020)). Moreover, edges linking these nodes can describe a variety of interaction measures. Ecologists have  
113 traditionally represented interactions as binary objects that were considered realized after observing at least one  
114 individual from group  $i$  interact with at least another individual from group  $j$ . Boolean interactions can be  
115 viewed as the result of a Bernoulli process  $A_{i,j} \sim \text{Bernoulli}(P(i \rightarrow j))$ , with  $P(i \rightarrow j)$  being the probability of  
116 interaction between  $i$  and  $j$  that characterizes our limited knowledge of the system and its intrinsic  
117 spatiotemporal variability. Depending on the type of networks (local or metaweb), the mathematical

118 formulation and interpretation of stochastic parameters like  $P(i \rightarrow j)$  can be linked to environmental and  
119 biological factors such as species relative abundance, traits, area, and time (tbl. 1), for example using logistic  
120 regression with a logit link function with continuous explanatory variables. In these probabilistic network  
121 representations in which  $P(i \rightarrow j)$  are edge values, the only two possible outcomes are the presence ( $A_{i,j} = 1$ ) or  
122 absence ( $A_{i,j} = 0$ ) of an interaction between each pair of nodes. When considering uncertainties around the  
123 estimation of  $P(i \rightarrow j)$ , a Beta distribution  $\text{Beta}(\alpha, \beta)$  can also be used to encompass all possible probability  
124 values. Observing an interaction between two taxa at a given location and time provides important information  
125 that can be used to update previous estimates of  $P(i \rightarrow j)$ , informing us on the biological capacity of both taxa  
126 to interact and the environmental conditions that enabled them to interact locally.

127 Even though binary networks constitute a highly valuable source of ecological information (Pascual *et al.*  
128 (2006)), they overlook important factors regarding interaction strengths. These are represented using  
129 quantitative interactions, which better describe the energy flows, demographic impacts or frequencies of  
130 interactions between nodes (Berlow *et al.* (2004), Borrett & Scharler (2019)), with  $A_{i,j}$  being a natural number  
131  $\mathbb{N}$  or a real number  $\mathbb{R}$  depending on the measure. For example, they may represent local interaction rates  
132 between pairs of taxa (e.g., the flower visiting rates of pollinators in a mutualistic network, Herrera (1989)).  
133 Because quantitative interactions can also describe predation pressure on prey taxa in food webs, they can be  
134 good estimators of the parameters describing species interactions in a Lotka-Volterra model (e.g., Emmerson &  
135 Raffaelli (2004)). The extra amount of ecological information in quantitative networks typically comes at a cost  
136 of greater sampling effort and data volume in predictive models (Strydom *et al.* (2021)), which can lead to  
137 relatively high levels of uncertainties when inferring quantitative networks with limited data.

138 Just like binary networks, the uncertainty and spatiotemporal variability of quantitative interactions can be  
139 represented probabilistically. However, the need to estimate the probability distribution of all possible values of  
140 interaction strengths can make the inference of probabilities more challenging in quantitative networks  
141 compared to binary networks, which require only one probability estimate for each interaction. Quantitative  
142 interactions can follow various probability distributions depending on the measure used, the event's outcome  
143 being the value of interaction strength. For instance, quantitative interactions can follow a Poisson distribution  
144  $A_{i,j} \sim \text{Poisson}(\lambda_{i \rightarrow j})$  when predicting frequencies of interactions between pairs of nodes, with  $\lambda_{i \rightarrow j}$  being the  
145 expected rate at which individuals of taxa  $i$  and  $j$  interact (e.g., the average number of prey  $j$  consumed by all  
146 predators  $i$  in a given time period). The Poisson distribution can also be 0-inflated after initially modeling  
147 non-interacting taxa (e.g., Boulangeat *et al.* (2012) employ a 0-inflated model to analyze species abundance

148 following the modeling of species presence and absence), which constitute the majority of taxa pairs in most  
149 local networks due to their typically high sparseness (Jordano (2016)). Because of the methodological  
150 difficulties typically encountered when building deterministic quantitative networks, binary networks, which are  
151 easier to sample (Jordano (2016)) and predict (Strydom *et al.* (2021)), are much more documented and modeled  
152 in the literature. Moreover, most published probabilistic networks and methods describe Bernoulli interactions  
153 (whether interaction probabilities are regarded as constant or variable, e.g. represented by Beta distributions),  
154 which underlines the need for better guidelines regarding the interpretation and manipulation of these types of  
155 networks. For these reasons, our primary focus in this contribution will be on addressing the challenges in  
156 interpreting and using Bernoulli interactions, in both probabilistic local networks and metawebs.

## 157 Local networks: communities interacting in space and time

158 Probabilistic local networks describe how likely taxa are to interact at a given location and time period (i.e.,  
159 interactions are contingent upon the environmental and biological conditions of the community). In local  
160 networks, edges commonly represent our degree of belief that two taxa interact in nature, but can also represent  
161 the probability of *observing* this interaction (Catchen *et al.* (2023)). For example, Kopelke *et al.* (2017)  
162 assembled a dataset of binary local European food webs of willow-galling sawflies and their natural enemies,  
163 clearly referencing each food web in space and time. Gravel *et al.* (2019) used this dataset to infer the  
164 probabilities of locally observing interactions between co-occurring species, with the added aspect of situating  
165 local networks within the context of environmental variables. This was achieved by including temperature and  
166 precipitation as conditional variables in some of their models.

167 We define space as the set of geographic coordinates ( $x, y, z$ ) of the spatial boundaries delineating the system,  
168 whether sampled or targeted. Ecological interactions may exhibit variations spatially and along altitudinal  
169 gradients, as evidenced by changes in hummingbird-plant interactions (Weinstein & Graham (2017b) and  
170 Weinstein & Graham (2017a)) and mosquitoes biting rates (e.g., Kulkarni *et al.* (2006)) at different elevations.  
171 In contrast, time is defined as the specific time period within which interactions were either observed or  
172 predicted. With these definitions in mind, space and time can be conceptualized as distinct patches or time  
173 segments. Treating them as discrete dimensions aligns with the common sampling methods of ecological  
174 networks and provides actual probabilities of interactions across space and time. This differs from the approach  
175 of treating them as continuous variables, where edge values represent probability densities (i.e., relative  
176 likelihoods of interactions occurring at infinitesimal locations and instants in time) rather than discrete

177 probability values (which can be obtained by integrating probability densities). By employing discrete locations  
178 and time periods, we can quantify both an area  $A$  and a duration  $t$ , which can be readily used in spatiotemporal  
179 analyses of ecological networks. For example, when studying network-area relationships (NAR, Galiana *et al.*  
180 (2018)), we anticipate that local probabilities of interactions scale positively with area and duration because  
181 taxa have more opportunities to interact as these dimensions expand.

182 The probability that two taxa  $i$  and  $j$  interact locally can be conditional on many environmental and biological  
183 factors. One of these is their co-occurrence  $C_{i,j}$ , which is usually a Boolean describing if the geographic  
184 distributions of both taxa overlap within the study area. Co-occurrence can be modeled probabilistically, in  
185 which case it may conform to a Bernoulli distribution  $C_{i,j} \sim \text{Bernoulli}(P_{i,j}(x, y, z))$ , where the probability of  
186 co-occurrence  $P_{i,j}(x, y, z)$  can be estimated through the application of joint species distribution models (e.g.,  
187 Pollock *et al.* (2014)). Considering that the probability that two non co-occurring taxa interact locally is zero  
188 (i.e.,  $P_N(i \rightarrow j|C = 0) = 0$ ), the probability of local interaction can be obtained by multiplying the probability of  
189 interaction given co-occurrence with the probability of co-occurrence:

$$P_N(i \rightarrow j) = P_N(i \rightarrow j|C = 1) \times P_{i,j}(x, y, z). \quad (1)$$

190 Other important factors that can impact our estimation of interaction probabilities at the local scale are taxa  
191 relative abundance (Canard *et al.* (2012)) and traits (Poisot *et al.* (2015)), as well as environmental factors such  
192 as temperature (Angilletta *et al.* (2004)), precipitation (Woodward *et al.* (2012)), habitat structure (Klecka &  
193 Boukal (2014)), and presence of other interacting taxa in the network (Pilosof *et al.* (2017), Kéfi *et al.* (2012)).  
194 Here, we use the variable  $\Omega$  to describe the biological and ecological context in which interaction probabilities  
195 were estimated. For example, if a research team conducts a mesocosm experiment to estimate interaction  
196 probabilities between predators and prey with and without shelters,  $\Omega$  would represent the presence or absence  
197 of these shelters. Like co-occurrence,  $\Omega$  can also be modeled probabilistically when the stochasticity or  
198 uncertainty of environmental and biological factors is considered. In sum,  $\Omega$  represents all ecological and  
199 biological variables that were taken into consideration when measuring interaction probabilities and is,  
200 therefore, a subset of all factors actually impacting ecological interactions.

201 The probability that two taxa  $i$  and  $j$  interact in a local network  $N$  can thus be conditional on the area  $A$ , the time  
202 interval  $t$ , their co-occurrence  $C$  and chosen environmental and biological conditions  $\Omega$ . This gives us the  
203 following equation when all of these conditions are included in the estimation of interaction probabilities:

$$P_N(i \rightarrow j|A, t, C, \Omega). \quad (2)$$

204 The representation of the local context in which probabilities are estimated and the variables that should be  
 205 taken into consideration depend on the study system, the objectives of the study, and the resources available to  
 206 the researchers. In other words, these variables do not systematically need to be accounted for. For example, in  
 207 Gravel *et al.* (2019), for the purpose of model comparison, some models incorporated environmental variables  
 208 as conditional factors to estimate interaction probabilities, while others did not. When accounted for, these  
 209 variables should be clearly described in the documentation of the data (Brimacombe *et al.* (2023)),  
 210 preferentially in mathematical terms to avoid any confusion in their interpretation and to limit manipulation  
 211 errors during their re-use. For instance, ecologists should be explicit about their consideration of co-occurrence  
 212 in their estimation of local interaction probabilities. Indeed, it is important to specify if probability values are  
 213 conditional  $P_N(i \rightarrow j|C = 1)$  or not  $P_N(i \rightarrow j)$  on co-occurrence since this can significantly impact the  
 214 interpretation and analysis of the data. In [tbl. 1](#), we present examples of studies that used these diverse  
 215 formulations of probabilistic interactions and conditional variables.

**Table 1: Notation of probabilistic interactions.** The probability of interaction between two taxa  $i$  and  $j$  is interpreted differently in a metaweb  $M$  of potential interactions and a local network  $N$  of realized interactions. Each notation includes a different conditional variable, when applicable. An example of a study employing each of these notations and conditional variables is provided, with the specific variables used indicated in parentheses. The study marked with an asterisk has been carried out on binary networks. Note that interaction probabilities can be contingent upon multiple conditional variables, or none at all. Additionally, local interaction probabilities might represent the probability of observing an interaction, not necessarily its actual occurrence.

Notation	Type	Description	Reference
$P_M(i \rightarrow j)$	potential	biological feasibility of the interaction	Strydom <i>et al.</i> (2022)
$P_N(i \rightarrow j)$	local	realization of the interaction	Fortuna & Bascompte (2006) (null model)
$P_N(i \rightarrow j A)$	local	realization of the interaction in a given area	Galiana <i>et al.</i> (2018) *
$P_N(i \rightarrow j t)$	local	realization of the interaction during a given time period	Weinstein & Graham (2017a)
$P_N(i \rightarrow j C)$	local	realization of the interaction given that the taxa co-occur	Gravel <i>et al.</i> (2019)

Notation	Type	Description	Reference
$P_N(i \rightarrow j \Omega)$	local	realization of the interaction given environmental conditions	Gravel <i>et al.</i> (2019) (temperature and precipitation)

## 216 Metawebs: regional catalogs of interactions

217 Metawebs are networks of potential interactions that have been designed for broad spatial, temporal, and  
 218 taxonomic scales (e.g., species food webs at the continental scale). They represent the probability that taxa can  
 219 biologically interact regardless of their co-occurrence and local environmental conditions. Potential interactions  
 220 are by definition context-independent, i.e. they are not measured at a specific location and time. In contrast with  
 221 probabilistic local networks, which represent the stochasticity of interactions occurring in nature, probabilistic  
 222 metawebs measure our degree of belief in the capacity of two taxa to interact (i.e., the probability that their  
 223 traits could support an interaction in the right conditions). In other words, potential interactions describe the  
 224 probability that there exists at least one combination of phenotypes of taxa  $i$  and  $j$  that can interact with each  
 225 other if they were to encounter. To reduce any biases in our calculation of this probability, it is crucial to ensure  
 226 that the set of traits sampled or considered accurately reflects the overall trait distribution in both taxa. This  
 227 enhances our confidence in concluding that the taxa cannot interact when no positive observation has been  
 228 made and increases our capacity to detect an interaction when it is biologically feasible. In addition to  
 229 one-on-one interactions, the extent of sampling effort also contributes to shaping our assessment of metaweb  
 230 properties, as sampling a larger number of local networks allows us to capture more potential interactions  
 231 (McLeod *et al.* (2021a)). A probability of potential interaction in a metaweb  $M$  can be expressed as

$$P_M(i \rightarrow j), \quad (3)$$

232 which, in contrast with eq. 2, is not conditional on any spatial, temporal, or environmental variables (tbl. 1).

233 Starting from a selected set of taxa, which are usually distributed within a broad region of interest, probabilistic  
 234 metawebs can be built using different data sources, including literature review, fieldwork, and predictive models  
 235 (e.g., the metaweb of trophic interactions between Canadian mammals inferred by Strydom *et al.* (2022)).

236 Every pair of taxa that have confidently been observed to interact at least once can be given a probability of 1  
 237 (i.e.,  $P_M(i \rightarrow j) = 1$ ) since we know that they *can* interact. This is not the case in local probabilistic networks,

238 in which interaction events usually remain stochastic (i.e.,  $P_N(i \rightarrow j) < 1$ ) after empirically observing  
239 interactions because of their intrinsic spatiotemporal variability. Similarly, although rare interactions typically  
240 have low probabilities in local networks, they can have high probabilities in metawebs when the traits of both  
241 taxa are congruent. On the other hand, interactions that were never observed can have low probability values in  
242 both metawebs and local networks, going as low as 0 for forbidden links. However, because of observation  
243 errors due to taxonomic misidentifications and ecological misinterpretations (e.g., due to cryptic species and  
244 interactions, Pringle & Hutchinson (2020)), many observations of interactions are in reality false positives.  
245 Likewise, forbidden interactions can be false negatives in metawebs, e.g. if they have been assessed for specific  
246 phenotypes, locations or time. Employing Bayesian models, whether they are mechanistic or  
247 phenomenological, has the potential to enhance the accuracy of our estimation of interaction probabilities  
248 within both local and potential networks (e.g., Bartomeus *et al.* (2016), Cirtwill *et al.* (2019)). This  
249 improvement is achieved in potential networks by leveraging prior information regarding the feasibility of  
250 interactions along with the empirical data on observed interactions.

## 251 **Properties of probabilistic networks**

252 Probabilistic local networks and metawebs differ in their type of interactions (i.e., local or potential) and in the  
253 conditional variables upon which interaction values depend. These differences are significant as they influence  
254 the characteristics of probabilistic networks. Neglecting to consider them may result in misleading results and  
255 interpretation errors when analyzing the properties of probabilistic networks, which could be particularly  
256 problematic when addressing crucial ecological questions about networks. Here we compare the characteristics  
257 of local networks and metawebs through the presentation of four common applications of probabilistic  
258 interactions. All code and data to reproduce these analyses are available at the Open Science Framework (TBD).

## 259 **Host-parasite network data**

260 We use the collection of tripartite host-parasite networks sampled across Europe, created by Kopelke *et al.*  
261 (2017), in most of our case studies. This dataset contains well-resolved binary local interactions between  
262 willows (52 species), willow-galling sawflies (96 species), and their parasitoids (126 species). Given its  
263 replicated networks spanning large spatiotemporal scales, this dataset is well-suited for analyzing a variety of  
264 ecological hypotheses and processes. Out of a total of 374 local networks, we retained those containing at least

265 5 species, resulting in a set of 233 georeferenced local networks. We built a binary metaweb by aggregating all  
266 local interactions, which gave us a regional metaweb composed of 274 species and 1080 interactions. In the first  
267 two panels of fig. 1, we show how the dissimilarity of interactions between common species ( $\beta_{OS}$ ) and the  
268 dissimilarity in species composition ( $\beta_S$ ) between the metaweb and aggregated local networks (Poisot *et al.*  
269 (2012)) vary with the number of sampled local networks. This shows that networks of local interactions are  
270 highly dissimilar from the metaweb, both in terms of species and interactions, especially when only a limited  
271 number of sites has been sampled.

272 [Figure 1 about here.]

273 We converted these binary networks into probabilistic ones using models based on simple assumptions. Our  
274 models do not aim to precisely estimate the actual values of probabilistic interactions but rather to offer  
275 simplified networks as illustrative examples to underscore the differences between probabilistic local networks  
276 and metawebs. We created two probabilistic metawebs by employing constant false positive and false negative  
277 rates for all potential interactions. In the first metaweb, we set both false positive and false negative rates to zero  
278 to prevent artificially inflating the total number of links, enabling a more accurate comparison with binary  
279 networks. This gave us a probability of potential interaction of 1 when at least one interaction has been observed  
280 and of 0 in the absence of any observed interaction between a given pair of species. In the second metaweb, we  
281 introduced a 5% false positive rate to account for spurious interactions and a 10% false negative rate to capture  
282 the elevated occurrence of false negatives in ecological networks (Catchen *et al.* (2023)). Observed potential  
283 interactions were thus given a probability of 95%, whereas unobserved ones were assigned a probability of 10%.

284 To build probabilistic local networks, we first recognize that local interactions must initially be biologically  
285 feasible before occurring at a specific time and space. A local probability of interaction  $P_N(i \rightarrow j)$  can be  
286 expressed as the product of the probability of local interaction given that the two taxa can potentially interact  
287  $P_N(i \rightarrow j|M_{i \rightarrow j} = 1)$ , which we denote as  $p$  for the sake of simplicity, with their probability of potential  
288 interaction  $P_M(i \rightarrow j)$ :

$$P_N(i \rightarrow j) = P_N(i \rightarrow j|M_{i \rightarrow j} = 1) \times P_M(i \rightarrow j). \quad (4)$$

289 We built the probabilistic local networks from the binary ones by using the probabilistic metawebs and a  
290 constant value of  $p$  across interactions. Lower values of  $p$  indicate that feasible interactions rarely occur locally,

291 intermediate values around 50% suggest considerable spatiotemporal variability, while higher values indicate  
292 that potential interactions are nearly always realized. Following eq. 4, the local probabilities of interaction  
293 between a given pair of taxa consistently remained equal to or below their probability of potential interaction.

294 In the last two panels of fig. 1, we show how the aggregated number of links and connectance (i.e., the  
295 proportion of all of the  $S^2$  possible links that are realized, where  $S$  represents species richness) scale with the  
296 number of sampled local probabilistic networks, according to different values of  $p$ . When aggregating local  
297 probabilistic networks, the constancy of the probability of potential interaction across the entire study area  
298 means that any rise in the probability of local interaction is solely attributable to an increase in  $p$ . The  
299 probability  $p_{1,2}$  of local interaction among potentially interacting species in an aggregated network  $N_{1,2}$  is  
300 obtained by:

$$p_{1,2} = 1 - (1 - p_1) \times (1 - p_2), \quad (5)$$

301 where  $p_1$  and  $p_2$  are the probabilities of local interaction among two potentially interacting species in the  
302 subnetworks  $N_1$  and  $N_2$ , respectively.

303 By comparing the scaling relationships observed in binary and probabilistic networks, fig. 1 illustrates that high  
304 values of  $p$  lead to systematic overestimations in the number of links and connectance, especially when  $p = 1$   
305 (corresponding to the scenario where local probabilities of interactions are equivalent to the probabilities of  
306 potential interactions). However, these biases tend to diminish as the number of sampled networks increases.

### 307 Taxonomic scaling of interactions

308 Probabilistic networks offer a versatile approach to tackle a broad array of ecological questions, depending on  
309 their level of organization. For instance, the assemblage of interactions across ecological scales can be explored  
310 through species-based networks, while clade-based networks provide insights into macroevolutionary processes  
311 (e.g., Gomez *et al.* (2010)). Given that our interpretation of the properties and dynamics of ecological networks  
312 depends on their taxonomic scale (Guimarães (2020)), investigating the phylogenetic scaling of network  
313 structure emerges as a promising research avenue. Examining the same system at various taxonomic scales can  
314 yield meaningful and complementary ecological information, and, in our perspective, employing probabilistic  
315 networks is an effective approach for such analyses.

316 There are no inherent differences in the taxonomic scaling between local networks and metawebs, as only the  
317 nodes are defined taxonomically. In other words, the probability values of edges in both local networks (eq. 2)  
318 and metawebs (eq. 3) are not conditional on any taxonomic scale. The phylogenetic scale is tied to the definition  
319 of the event itself (i.e., the interaction between two taxa), not to the conditional variables. In both types of  
320 networks, transitioning to a broader level of organization can be accomplished directly by using probabilities  
321 from finer scales. For example, in a network with  $n_A$  species from genus A and  $n_B$  species from genus B, one  
322 can compute the probability that at least one species from genus A interacts with at least one species from genus  
323 B as follows:

$$P(A \rightarrow B) = 1 - \prod_{i=1}^{n_A} \prod_{j=1}^{n_B} (1 - P(A_i \rightarrow B_j)), \quad (6)$$

324 where  $A_i$  and  $B_j$  are the species of the corresponding genus. If it is known that at least two of these species  
325 interact (i.e.,  $P(A_i \rightarrow B_j) = 1$  for at least one pair of  $i, j$ ), it implies a probability of genus interaction equal to 1.  
326 Canard *et al.* (2012) built a species-based network following a similar approach, by using simulated interactions  
327 between individuals derived from a neutral model (i.e., a model that assumed ecological equivalence among  
328 individuals). In contrast, a more sophisticated approach is necessary when transitioning from a broader to a  
329 finer level of organization. This is because knowledge of an interaction between two genera does not guarantee  
330 that all possible pairwise combinations of their species will also interact. One possible method is to build a  
331 finer-scale network by generating probabilities of interactions through random sampling from a beta  
332 distribution, parameterized by the broader-scale network.

333 Ideally, our biological interpretation of probabilistic interactions should remain consistent across a network  
334 even if it incorporates heterogenous levels of organization (e.g., a network whose nodes are composed of both  
335 species and trophic species). This is common in ecological networks where taxonomic resolution is typically  
336 low (Hemprich-Bennett *et al.* (2021), Vázquez *et al.* (2022)). Interaction probabilities at broader taxonomic  
337 scales should be based on probabilities of interactions between individuals, for both local networks and  
338 metawebs. For instance, in local individual-based food webs, the probability that two individuals interact  
339 reflects our degree of belief that one individual will consume the other. Similarly, in local species-based food  
340 webs, the probability that two species interact represents our degree of belief that at least one individual from  
341 the predator species consumes at least another individual from the prey species. Furthermore, in local  
342 clade-based food webs, the probability that two clades interact represents our degree of belief that at least two

343 species from these clades interact with each other or, equivalently, that at least two individuals from these clades  
344 interact with each other. Fundamentally, the taxonomic scaling of interactions involves aggregating interactions  
345 between individuals into larger groups, which may exhibit varying degrees of homogeneity based on the  
346 organisms and the study system. In that regard, taxonomic scaling is analogous to the spatial and temporal  
347 scaling of interactions, as they all represent different ways to aggregate individuals into broader groups (either  
348 spatially, temporally, or taxonomically).

### 349 **Spatial and temporal scaling of interactions**

350 The investigation of network-area relationships (NARs) and interaction accumulation curves (IACs) is an  
351 important area of research in network ecology. First, NARs elucidate the scaling of network properties (such as  
352 modularity and trophic chain lengths) with spatial scale (e.g., Wood *et al.* (2015), Galiana *et al.* (2018)). The  
353 variations in network structure across spatial scales may stem from the scaling of species richness (species-area  
354 relationships, SARs) and the number of interactions (Brose *et al.* (2004)) with the sampled area. Additionally,  
355 ecological processes unfolding at distinct spatial scales, such as the spatial variability in local community  
356 composition resulting from different sequences of extinction and colonization events, can also contribute to this  
357 variation (Galiana *et al.* (2018)). Next, IACs describe the scaling of the number of observed interactions with  
358 sampling effort (Jordano (2016)), which can also impact various measures of network structure such as  
359 connectance (McLeod *et al.* (2021b)). Apart from sampling effort, the temporal scaling of interactions  
360 elucidates how network structure changes with the duration of the network (Poisot *et al.* (2012)),  
361 acknowledging that distinct interactions take place over time. Given that probabilistic local networks can  
362 quantify the spatiotemporal variability of interactions (eq. 2), they serve as valuable tools for developing and  
363 testing equations describing the scaling of network structure with space and time.

364 Metawebs and local networks exhibit distinct relationships with spatial and temporal scales. On one hand,  
365 probabilistic metawebs, being independent of any local context, feature potential interactions that do not scale  
366 with space and time. This is because potential interactions depend solely on the biological capacity of two taxa  
367 to interact, regardless of their co-occurrence and specific environmental conditions. This implies that the  
368 probability of two taxa potentially interacting should theoretically be the same in all metawebs in which they are  
369 present, provided that the data and methods used for estimation are consistent. As a result, if a smaller metaweb  
370  $M_1$  is derived from subsampling a regional metaweb  $M_0$  to represent potential interactions among a subset of  
371 taxa in a smaller region, their probabilities of interaction should be identical regardless of scale,

372 i.e.  $P_{M_1}(i \rightarrow j) = P_{M_0}(i \rightarrow j)$ . However, because the number of taxa may be higher in the larger metaweb, it is  
373 expected that the overall number of interactions would be higher in the larger network, despite pairwise  
374 probabilities being identical.

375 On the other hand, local interactions scale both spatially and temporally, given that they have more opportunities  
376 to be realized and observed in larger areas and longer durations. This is attributed to factors such as a higher  
377 number of individuals, greater trait variations, and increased opportunities for encounters, as highlighted by  
378 McLeod *et al.* (2020). For example, if a local probabilistic network  $N_1$  with an area  $A_1$  is derived from a larger  
379 network  $N_0$  with an area  $A_0$ , and  $A_1$  is entirely nested within  $A_0$ , interaction probabilities should be lower in the  
380 smaller network, i.e.  $P_{N_1}(i \rightarrow j|A_1 < A_0) \leq P_{N_0}(i \rightarrow j|A_0)$ . However, if  $A_1$  and  $A_0$  are disjoint, interaction  
381 probabilities could potentially be higher in the smaller area, contingent upon their environmental and biological  
382 conditions. Likewise, interaction probabilities are expected to be lower in networks with shorter durations when  
383 time intervals are nested. In fig. 2, we show how the expected number of local host-parasite interactions scales  
384 with area, represented as an expanding latitudinal window, in comparison with potential interactions. Even  
385 though we employed local probabilities of interactions equal to potential interactions for the purpose of  
386 comparison (i.e., using  $p = 1$  here), we notice that the total number of potential interactions scales more rapidly  
387 than local interactions. This is due to the fact that numerous potential interactions involve species that never  
388 co-occur, and as a result, these interactions are not captured in local networks.

389 [Figure 2 about here.]

390 **Box 1: A spatiotemporally explicit model of interactions**

391 Predicting local networks across time and space is a pivotal goal of network ecology (Strydom *et al.* (2021)). In  
392 a context of scarcity of interaction data (Jordano (2016)), ecologists must resort to predictive models for  
393 reconstructing networks at fine spatial and temporal scales. For instance, real-time biomonitoring data coupled  
394 with appropriate numerical models (Bohan *et al.* (2017)) can be employed to reconstruct local ecological  
395 networks, opening avenues for in-depth studies on local ecosystem functioning and dynamics. Apart from their  
396 predictive applications, statistical models can also be crafted for descriptive purposes, elucidating key  
397 parameters of interest such as probabilities of interactions. In such instances, the parameter values themselves  
398 offer valuable ecological insights. Various types of models (e.g., Bayesian and machine learning models) have  
399 been used for both predictive and descriptive purposes in the realm of ecological interactions (Strydom *et al.*

400 (2021)). The probabilistic representation of interactions acknowledges the inherent uncertainty in these models,  
 401 typically expressed through probability distributions. We introduce and develop a simple generative mechanistic  
 402 model for probabilistic local interactions that takes into consideration their spatiotemporal variability (i.e. a  
 403 spatiotemporally explicit model of local interactions). It is essential to note that our model is not designed for  
 404 potential interactions, which are scale-independent. Rather, it could prove valuable for predicting local  
 405 interactions across time and space by generating new interaction data following parameter inference.

406 As indicated by eq. 1, the probability that two taxa  $i$  and  $j$  interact at a specific location  $(x, y, z)$  is determined by  
 407 the product of their probability of interaction given co-occurrence and their probability of co-occurrence. First,  
 408 their co-occurrence probability can be calculated using their individual occurrence probabilities  $P_i(x, y, z)$  and  
 409  $P_j(x, y, z)$ , along with the strength of association  $\gamma$  between their occurrences and co-occurrence (Cazelles *et al.*  
 410 (2016)):

$$P_{i,j}(x, y, z) = P_i(x, y, z)P_j(x, y, z)\gamma. \quad (7)$$

411 When  $\gamma > 1$ , it signifies a positive association in the geographic distributions of both taxa, indicating that the  
 412 presence of one taxon enhances the probability of occurrence of the other. In empirical networks,  $\gamma > 1$  holds  
 413 true for the majority of species pairs (Catchen *et al.* (2023)). The co-occurrence of both taxa is modeled as the  
 414 outcome of a Bernoulli trial

$$C \sim \text{Bernoulli}(P_{i,j}(x, y, z)). \quad (8)$$

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

$$P_N(i \rightarrow j|C = 1) = 1 - e^{-\lambda t_0}, \quad (9)$$

420 which tends toward 1 as  $t_0 \rightarrow \infty$ . It is important to note that the units of  $\lambda$  and  $t_0$  are complementary. For

421 instance, if the duration  $t_0$  is measured in months,  $\lambda$  denote the expected number of interactions per month.  
 422 The occurrence of an interaction between  $i$  and  $j$  can be modeled as a Bernoulli trial with a probability of  
 423  $P_N(i \rightarrow j)$ . Consequently, a Bayesian inference model can be built based on the preceding equations to estimate  
 424 the value of the  $\lambda$  and  $\gamma$  parameters and generate novel interaction data:

$$I \sim \text{Bernoulli}(P_N(i \rightarrow j)) \quad (10)$$

425

$$P_N(i \rightarrow j) = P_i(x, y, z)P_j(x, y, z)\gamma(1 - e^{-\lambda t_0}) \quad (11)$$

$$\gamma \sim \text{Gamma}(2, 0.5) \quad (12)$$

$$\lambda \sim \text{Exponential}(2) \quad (13)$$

426 In fig. 3, we show the variation in the probability of interaction under different parameter values. In the right  
 427 panel, we notice that, irrespective of the interaction rate  $\lambda$ , the probability of interaction converges toward an  
 428 asymptote determined by the co-occurrence  $P_{i,j}(x, y, z)$  (eq. 7). This model can be customized in different ways,  
 429 such as by linking  $\lambda$  with specific environmental variables or explicitly incorporating observation errors (i.e.,  
 430 the probabilities of false negatives and false positives).

431

[Figure 3 about here.]

## 432 Sampling binary networks

433 The prediction of binary interactions through random draws is an important application of probabilistic  
 434 networks. This approach proves beneficial for analyzing the structural characteristics of probabilistic networks,  
 435 particularly in the absence of specific analytical measures. By simulating independent Bernoulli trials for each  
 436 interaction, a binary network can be generated. A probability distribution of network properties can be obtained  
 437 by measuring network structure across multiple randomly generated networks (Poisot *et al.* (2016)). This  
 438 method enables the representation of the variability of network structure, albeit with potential biases when  
 439 connectance is low (Poisot & Gravel (2014), Chagnon (2015)). Employing this strategy to generate binary

440 networks under a null model facilitates null hypothesis significance testing, wherein the observed measure is  
441 compared against the simulated distribution (e.g., Bascompte *et al.* (2003)). Additionally, randomly generating  
442 binary networks across spatial and temporal dimensions aids in representing the spatiotemporal variability of  
443 network structure, allowing the testing of ecological hypotheses regarding interactions on large spatial and  
444 temporal scales.

445 There are at least two distinct approaches to sample binary networks from probabilistic webs across space, for  
446 example, when attempting to predict a binary network for each of a number of locations within a given region.  
447 The first approach involves performing a singular Bernoulli trial for each pair of taxa within the region of  
448 interest based on their regional probability of potential interaction. In employing this approach, every pair of  
449 taxa predicted to interact in the binary metaweb realization will be treated as interacting in all local networks  
450 where they co-occur. This will result in local pairwise interactions without spatial variation. The second  
451 approach is to independently sample each of the local probabilistic networks, which can be achieved by first  
452 generating distinct probabilistic networks for each location. These local probabilistic networks may vary in taxa  
453 composition and interaction probabilities. Subsequently, binary networks can be independently sampled for  
454 each location, introducing spatial variation in binary interactions.

455 In fig. 4, we compare the average connectance of binary networks resulting from these two sampling techniques,  
456 where potential and local interactions are drawn from our host-parasite probabilistic networks, generating a  
457 number of binary network realizations for each site in the dataset. These two sampling approaches yield  
458 different outcomes, particularly for lower values of  $p$ , which denote instances when potential interactions do not  
459 consistently manifest locally. Small discrepancies are also apparent between these techniques when we equate  
460 the probability of local interaction to the probability of potential interaction (i.e., when using  $p = 1.0$  in eq. 4),  
461 especially when the number of binary network samples for each location is low. As anticipated, we observe that  
462 sampling binary interactions from the metaweb tends to overestimate connectance on average compared to  
463 sampling them from local networks. Furthermore, we observe an increase in the variability of connectance  
464 when employing a single sample, representing what we consider as a more tangible process leading to the  
465 realization of local and potential interactions in nature.

466 [Figure 4 about here.]

467 The choice of a sampling approach can influence the selection of grid cell size when delineating local  
468 communities within a broader region of interest. In the first approach, pairwise interactions remain constant

469 irrespective of cell size since they are sampled only once from the regional network. However, in the second  
470 approach, local interaction probabilities are contingent on the network area. For instance, consider networks  $N_1$   
471 and  $N_2$  with an area of  $\frac{1}{2}A_0$ , both nested within  $A_0$  but disjoint from each other, forming  $N_0$ . If we treat  $N_1$  and  
472  $N_2$  as independent, the probability of interaction between taxa  $i$  and  $j$  in  $N_0$  is given by:

$$P_{N_0}(i \rightarrow j) = 1 - (1 - P_{N_1}(i \rightarrow j)) \times (1 - P_{N_2}(i \rightarrow j)). \quad (14)$$

473 Due to its larger area, the probability that the two taxa interact in  $N_0$  is equal or greater than in  $N_1$  and  $N_2$ .  
474 When sampling binary interactions from local networks, it is crucial to sample at the same spatial scale for  
475 which probabilities were estimated. Otherwise, interaction probabilities must be adjusted to align with the  
476 intended cell size, preventing potential systematic biases in predictions.

#### 477 Prediction of local networks from metawebs

478 Metawebs serve as a valuable source of ecological information for predicting local networks across time and  
479 space. Binary local networks can be reconstructed by selecting a subset of taxa and interactions from the  
480 metaweb (Dunne (2006)). This implies that metawebs consistently contain more interactions than their  
481 corresponding local networks. Determining the list of taxa to select can be achieved empirically or through  
482 methods like range maps or species distribution models. As the species composition of a community is arguably  
483 less difficult to sample or predict than its interactions, the primary challenge lies in deciding which interactions  
484 to select from the metaweb, a task that may necessitate advanced statistical models and ecological expertise.  
485 Inferring the structure of local networks from the metaweb before predicting specific local pairwise interactions  
486 could hold promise (Strydom *et al.* (2021)), considering that the structure of local networks is constrained by  
487 the metaweb (Saravia *et al.* (2022)).

488 Building probabilistic local networks from a probabilistic metaweb involves a reduction in the value of pairwise  
489 interaction probabilities. This decrease is attributed to the prerequisite that two taxa must initially possess the  
490 capacity to interact before engaging in local interactions (eq. 4). Therefore, inferring local networks from their  
491 metaweb while maintaining identical interaction probability values would introduce systematic biases into the  
492 predictions. In such cases, these networks would essentially represent smaller-scale metawebs of potential  
493 interactions, potentially leading to misinterpretations by being perceived as local interactions. As proposed by  
494 McLeod *et al.* (2021b), although metawebs lack the ability to capture the spatiotemporal variability of

495 interactions, they establish an upper limit for local interactions. In other words, the probability of two taxa  
496 interacting at a specific location and time is consistently lower or equal to the probability of their potential  
497 interaction:

$$P_N(i \rightarrow j|A, t, C, \Omega) \leq P_M(i \rightarrow j). \quad (15)$$

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

$$\int_{\Omega} \int_A \int_t P_N(i \rightarrow j|A, t, \Omega) dt dA d\Omega \leq P_M(i \rightarrow j). \quad (16)$$

502 Estimating more precisely the probability  $p$  that two taxa interact locally if they can potentially interact allows  
503 for improved predictions of local networks from a probabilistic metaweb. This task is challenging due to the  
504 variability of this probability across space and time, as well as its potential variations among pairwise  
505 interactions within a network. Using simple models of  $p$ , as demonstrated in our case studies, represents an  
506 initial step toward the overarching objective of reconstructing local networks from metawebs.

## 507 Conclusion

508 In this contribution, we underline the importance of network metadata for adequately interpreting and  
509 manipulating probabilistic interaction data. The mathematical representation of probabilities and their statistical  
510 properties depend on the type of interactions (local or potential) and the conditions in which they were  
511 estimated. We showed that probabilistic local networks and metawebs differ in their relationship to spatial and  
512 temporal scales, with potential interactions remaining consistent across scales. In contrast with metawebs, local  
513 interactions are measured in a specific context (e.g., in a given area, time, and biological and environmental  
514 conditions) and depend on taxa co-occurrence. These important conceptual differences bring to light the need  
515 to use probabilistic data with caution, for instance when generating binary network realizations across space and  
516 predicting local networks from metawebs. Clear metadata describing the type of interaction and the variables  
517 used in their estimation are required to ensure adequate data manipulation. Better data practices and

518 foundations for probabilistic thinking in network ecology could enable more reliable assessments of the  
519 spatiotemporal variability and uncertainty of biotic interactions.

520 It is essential to enhance our comprehension of both potential and local interactions, especially considering the  
521 current scarcity of interaction data. However, while sampling biological communities does decrease the  
522 uncertainty of interactions by accumulating evidence for their feasibility and local realization, there is a limit to  
523 how much we can diminish their inherent randomness. In metawebs, probabilities reflect our limited knowledge  
524 of interactions (i.e., our degree of belief that interactions are feasible), which is expected to improve with a  
525 larger volume of data. We should anticipate that potential interactions will become more definitive (with  
526 probabilities approaching 0 or 1) as we investigate various conditions, including different combinations of  
527 species traits. Conversely, in the case of local networks, which can be seen as random instances of metawebs,  
528 randomness cannot be as reduced. In local interactions, probabilities can be divided into a component  
529 representing uncertainty and another representing spatiotemporal variability. Owing to environmental  
530 heterogeneity, there will invariably be instances in which an interaction occurs and others in which it does not,  
531 across different times and locations, irrespective of the extent to which we can improve our knowledge of its  
532 biological feasibility and the local conditions that facilitate its occurrence. When local networks depict  
533 probabilities of observing interactions rather than just their actual occurrence, we must also consider the  
534 observation variability as an additional source of randomness. Every ecological process is stochastic but there is  
535 also a possibility that a phenomenon goes undetected. Quantifying and partitioning this stochasticity will  
536 enable us to make more accurate predictions about ecological interactions at various spatial and temporal  
537 scales. This will prove to be of vital importance as our time to understand nature runs out, especially at the  
538 places where the climate change impacts hit harder.

## 539 Acknowledgment

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

546 feedback on this manuscript.

547 **References**

- 548 Abrego, N., Roslin, T., Huotari, T., Ji, Y., Schmidt, N.M., Wang, J., *et al.* (2021). Accounting for species  
549 interactions is necessary for predicting how arctic arthropod communities respond to climate change.  
550 *Ecography*, 44, 885–896.
- 551 Andrade-Pacheco, R., Rerolle, F., Lemoine, J., Hernandez, L., Meité, A., Juziwelo, L., *et al.* (2020). Finding  
552 hotspots: Development of an adaptive spatial sampling approach. *Scientific Reports*, 10, 10939.
- 553 Angilletta, M.J., Jr., Steury, T.D. & Sears, M.W. (2004). Temperature, Growth Rate, and Body Size in  
554 Ectotherms: Fitting Pieces of a Life-History Puzzle<sup>1</sup>. *Integrative and Comparative Biology*, 44, 498–509.
- 555 Araujo, M.B., Rozenfeld, A., Rahbek, C. & Marquet, P.A. (2011). Using species co-occurrence networks to  
556 assess the impacts of climate change. *Ecography*, 34, 897–908.
- 557 Bartomeus, I., Gravel, D., Tylianakis, J.M., Aizen, M.A., Dickie, I.A. & Bernard-Verdier, M. (2016). A  
558 common framework for identifying linkage rules across different types of interactions. *Functional Ecology*,  
559 30, 1894–1903.
- 560 Bascompte, J., Jordano, P., Melián, C.J. & Olesen, J.M. (2003). The nested assembly of plantanimal mutualistic  
561 networks. *Proceedings of the National Academy of Sciences*, 100, 9383–9387.
- 562 Berlow, E.L., Neutel, A.-M., Cohen, J.E., De Ruiter, P.C., Ebenman, B., Emmerson, M., *et al.* (2004).  
563 Interaction strengths in food webs: Issues and opportunities. *Journal of Animal Ecology*, 73, 585–598.
- 564 Blanchet, F.G., Cazelles, K. & Gravel, D. (2020). Co-occurrence is not evidence of ecological interactions.  
565 *Ecology Letters*, 23, 1050–1063.
- 566 Bohan, D.A., Vacher, C., Tamaddoni-Nezhad, A., Raybould, A., Dumbrell, A.J. & Woodward, G. (2017).  
567 Next-Generation Global Biomonitoring: Large-scale, Automated Reconstruction of Ecological Networks.  
568 *Trends in Ecology & Evolution*, 32, 477–487.
- 569 Bolnick, D.I., Amarasekare, P., Araújo, M.S., Bürger, R., Levine, J.M., Novak, M., *et al.* (2011). Why  
570 intraspecific trait variation matters in community ecology. *Trends in Ecology & Evolution*, 26, 183–192.

- 571 Borrett, S.R. & Scharler, U.M. (2019). [Walk partitions of flow in Ecological Network Analysis: Review and](#)  
572 [synthesis of methods and indicators](#). *Ecological Indicators*, 106, 105451.
- 573 Boulangeat, I., Gravel, D. & Thuiller, W. (2012). [Accounting for dispersal and biotic interactions to disentangle](#)  
574 [the drivers of species distributions and their abundances](#). *Ecology Letters*, 15, 584–593.
- 575 Brimacombe, C., Bodner, K., Michalska-Smith, M., Poisot, T. & Fortin, M.-J. (2023). [Shortcomings of reusing](#)  
576 [species interaction networks created by different sets of researchers](#). *PLOS Biology*, 21, e3002068.
- 577 Brose, U., Ostling, A., Harrison, K. & Martinez, N.D. (2004). [Unified spatial scaling of species and their](#)  
578 [trophic interactions](#). *Nature*, 428, 167–171.
- 579 Canard, E., Mouquet, N., Marescot, L., Gaston, K.J., Gravel, D. & Mouillot, D. (2012). [Emergence of](#)  
580 [Structural Patterns in Neutral Trophic Networks](#). *PLOS ONE*, 7, e38295.
- 581 Catchen, M.D., Poisot, T., Pollock, L.J. & Gonzalez, A. (2023). [The missing link: Discerning true from false](#)  
582 [negatives when sampling species interaction networks](#).
- 583 Cazelles, K., Araujo, M.B., Mouquet, N. & Gravel, D. (2016). [A theory for species co-occurrence in interaction](#)  
584 [networks](#). *Theoretical Ecology*, 9, 39–48.
- 585 Chagnon, P.-L. (2015). [Characterizing topology of ecological networks along gradients: The limits of metrics'](#)  
586 [standardization](#). *Ecological Complexity*, 22, 36–39.
- 587 Choh, Y., Ignacio, M., Sabelis, M.W. & Janssen, A. (2012). [Predator-prey role reversals, juvenile experience](#)  
588 [and adult antipredator behaviour](#). *Scientific Reports*, 2, 728.
- 589 Cirtwill, A.R., Eklöf, A., Roslin, T., Wootton, K. & Gravel, D. (2019). [A quantitative framework for](#)  
590 [investigating the reliability of empirical network construction](#). *Methods in Ecology and Evolution*, 10,  
591 902–911.
- 592 Dallas, T., Park, A.W. & Drake, J.M. (2017). [Predicting cryptic links in host-parasite networks](#). *PLOS*  
593 [Computational Biology, 13, e1005557.](#)
- 594 Dunne, J.A. (2006). The Network Structure of Food Webs. In: *Ecological networks: Linking structure and*  
595 *dynamics* (eds. Dunne, J.A. & Pascual, M.). Oxford University Press, pp. 27–86.
- 596 Elmasri, M., Farrell, M.J., Davies, T.J. & Stephens, D.A. (2020). [A hierarchical Bayesian model for predicting](#)  
597 [ecological interactions using scaled evolutionary relationships](#). *The Annals of Applied Statistics*, 14,  
598 221–240.

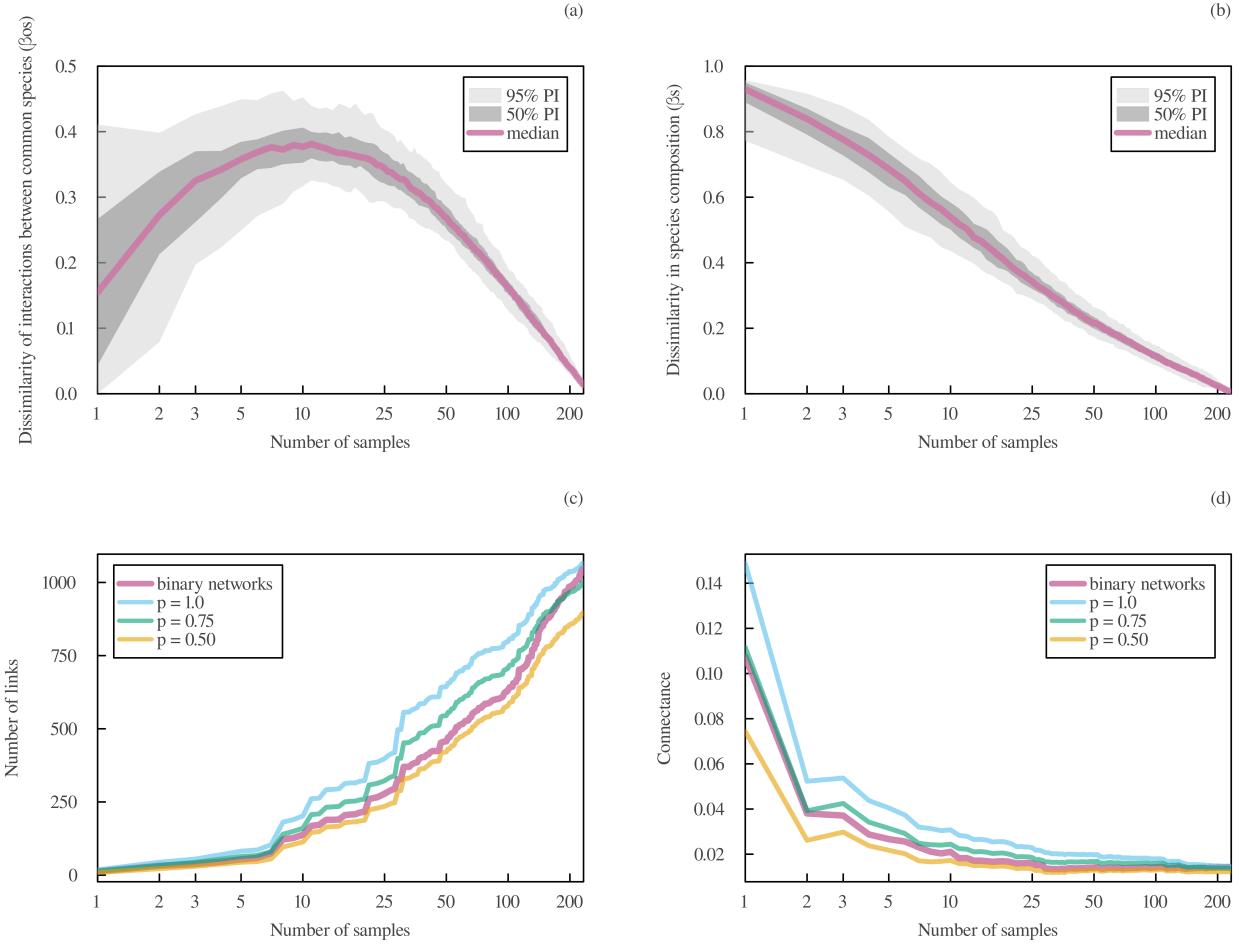
- 599 Elton, C.S. (2001). *Animal Ecology*. University of Chicago Press, Chicago, IL.
- 600 Emmerson, M.C. & Raffaelli, D. (2004). Predator-prey body size, interaction strength and the stability of a real  
601 food web. *Journal of Animal Ecology*, 73, 399–409.
- 602 Fortuna, M.A. & Bascompte, J. (2006). Habitat loss and the structure of plant-animal mutualistic networks.  
603 *Ecology Letters*, 9, 281–286.
- 604 Fu, X., Seo, E., Clarke, J. & Hutchinson, R.A. (2021). Link Prediction Under Imperfect Detection:  
605 Collaborative Filtering for Ecological Networks. *IEEE Transactions on Knowledge and Data Engineering*,  
606 33, 3117–3128.
- 607 Galiana, N., Lurgi, M., Claramunt-López, B., Fortin, M.-J., Leroux, S., Cazelles, K., et al. (2018). The spatial  
608 scaling of species interaction networks. *Nature Ecology & Evolution*, 2, 782–790.
- 609 Gilman, S.E., Urban, M.C., Tewksbury, J., Gilchrist, G.W. & Holt, R.D. (2010). A framework for community  
610 interactions under climate change. *Trends in Ecology & Evolution*, 25, 325–331.
- 611 Godsoe, W., Murray, R. & Iritani, R. (2022). Species interactions and diversity: A unified framework using Hill  
612 numbers. *Oikos*, n/a, e09282.
- 613 Golubski, A.J. & Abrams, P.A. (2011). Modifying modifiers: What happens when interspecific interactions  
614 interact? *Journal of Animal Ecology*, 80, 1097–1108.
- 615 Gomez, J.M., Verdu, M. & Perfectti, F. (2010). Ecological interactions are evolutionarily conserved across the  
616 entire tree of life. *Nature*, 465, 918–U6.
- 617 Gonzalez, A. & Londoño, M.C. (2022). Monitor biodiversity for action. *Science*, 378, 1147–1147.
- 618 Gonzalez-Varo, J.P. & Traveset, A. (2016). The Labile Limits of Forbidden Interactions. *Trends in Ecology &*  
619 *Evolution*, 31, 700–710.
- 620 Gravel, D., Baiser, B., Dunne, J.A., Kopelke, J.-P., Martinez, N.D., Nyman, T., et al. (2019). Bringing Elton  
621 and Grinnell together: A quantitative framework to represent the biogeography of ecological interaction  
622 networks. *Ecography*, 42, 401–415.
- 623 Gravel, D., Poisot, T., Albouy, C., Velez, L. & Mouillot, D. (2013). Inferring food web structure from  
624 predator-prey body size relationships. *Methods in Ecology and Evolution*, 4, 1083–1090.

- 625 Guimarães, P.R. (2020). The Structure of Ecological Networks Across Levels of Organization. *Annual Review*  
626 *of Ecology, Evolution, and Systematics*, 51, 433–460.
- 627 Guimerà, R. & Sales-Pardo, M. (2009). Missing and spurious interactions and the reconstruction of complex  
628 networks. *Proceedings of the National Academy of Sciences*, 106, 22073–22078.
- 629 Hemprich-Bennett, D.R., Oliveira, H.F.M., Le Comber, S.C., Rossiter, S.J. & Clare, E.L. (2021). Assessing the  
630 impact of taxon resolution on network structure. *Ecology*, 102, e03256.
- 631 Herrera, C.M. (1989). Pollinator abundance, morphology, and flower visitation rate: Analysis of the “quantity”  
632 component in a plant-pollinator system. *Oecologia*, 80, 241–248.
- 633 Hortal, J., de Bello, F., Diniz-Filho, J.A.F., Lewinsohn, T.M., Lobo, J.M. & Ladle, R.J. (2015). Seven Shortfalls  
634 that Beset Large-Scale Knowledge of Biodiversity. *Annual Review of Ecology, Evolution, and Systematics*,  
635 46, 523–549.
- 636 Jordano, P. (1987). Patterns of Mutualistic Interactions in Pollination and Seed Dispersal: Connectance,  
637 Dependence Asymmetries, and Coevolution. *The American Naturalist*, 129, 657–677.
- 638 Jordano, P. (2016). Sampling networks of ecological interactions. *Functional Ecology*, 30, 1883–1893.
- 639 Jordano, P., Bascompte, J. & Olesen, J.M. (2003). Invariant properties in coevolutionary networks of  
640 plantanimal interactions. *Ecology Letters*, 6, 69–81.
- 641 Kéfi, S., Berlow, E.L., Wieters, E.A., Joppa, L.N., Wood, S.A., Brose, U., et al. (2015). Network structure  
642 beyond food webs: Mapping non-trophic and trophic interactions on Chilean rocky shores. *Ecology*, 96,  
643 291–303.
- 644 Kéfi, S., Berlow, E.L., Wieters, E.A., Navarrete, S.A., Petchey, O.L., Wood, S.A., et al. (2012). More than a  
645 meal... integrating non-feeding interactions into food webs. *Ecology Letters*, 15, 291–300.
- 646 Kéfi, S., Miele, V., Wieters, E.A., Navarrete, S.A. & Berlow, E.L. (2016). How Structured Is the Entangled  
647 Bank? The Surprisingly Simple Organization of Multiplex Ecological Networks Leads to Increased  
648 Persistence and Resilience. *PLOS Biology*, 14, e1002527.
- 649 Klecka, J. & Boukal, D.S. (2014). The effect of habitat structure on prey mortality depends on predator and  
650 prey microhabitat use. *Oecologia*, 176, 183–191.
- 651 Kopelke, J.-P., Nyman, T., Cazelles, K., Gravel, D., Vissault, S. & Roslin, T. (2017). Food-web structure of  
652 willow-galling sawflies and their natural enemies across Europe. *Ecology*, 98, 1730–1730.

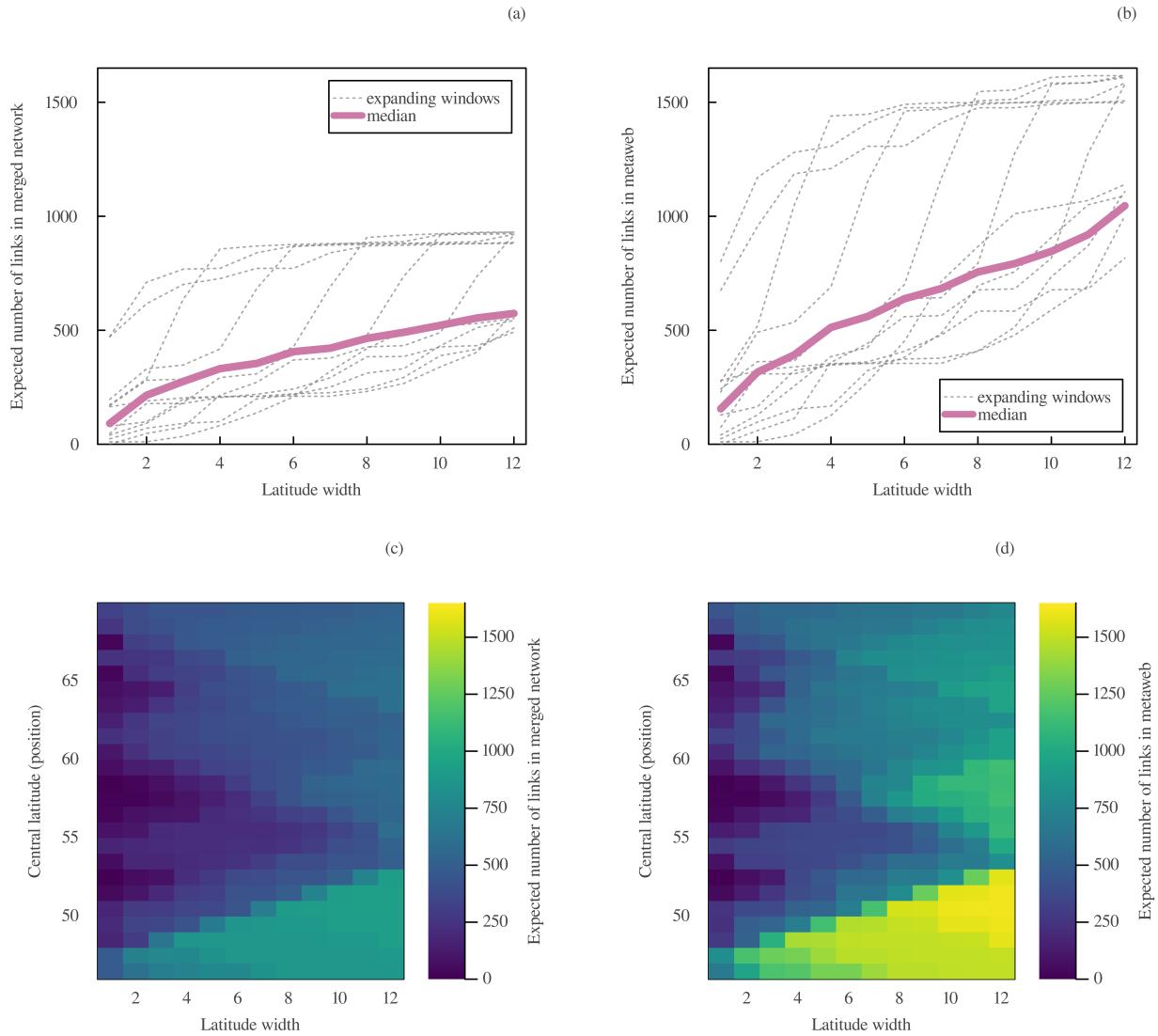
- 653 Kulkarni, M.A., Kweka, E., Nyale, E., Lyatuu, E., Mosha, F.W., Chandramohan, D., *et al.* (2006).  
654 [Entomological Evaluation of Malaria Vectors at Different Altitudes in Hai District, Northeastern Tanzania.](#)  
655 *Journal of Medical Entomology*, 43, 580–588.
- 656 McCann, K. (2007). Protecting biostructure. *Nature*, 446, 29–29.
- 657 McCann, K.S. (2011). Food webs (MPB-50). In: *Food Webs (MPB-50)*. Princeton University Press.
- 658 McLeod, A., Leroux, S.J., Gravel, D., Chu, C., Cirtwill, A.R., Fortin, M.-J., *et al.* (2021a). [Sampling and asymptotic network properties of spatial multi-trophic networks](#). *Oikos*, 130, 2250–2259.
- 660 McLeod, A., Leroux, S.J., Gravel, D., Chu, C., Cirtwill, A.R., Fortin, M.-J., *et al.* (2021b). [Sampling and asymptotic network properties of spatial multi-trophic networks](#). *Oikos*, n/a.
- 662 McLeod, A.M., Leroux, S.J. & Chu, C. (2020). [Effects of species traits, motif profiles, and environment on spatial variation in multi-trophic antagonistic networks](#). *Ecosphere*, 11, e03018.
- 664 Melián, C.J., Vilas, C., Baldó, F., González-Ortegón, E., Drake, P. & Williams, R.J. (2011). [Eco-evolutionary Dynamics of Individual-Based Food Webs](#). In: *Advances in Ecological Research*, The Role of Body Size in Multispecies Systems (ed. Belgrano, A.). Academic Press, pp. 225–268.
- 667 Ohlmann, M., Miele, V., Dray, S., Chalmandrier, L., O'Connor, L. & Thuiller, W. (2019). [Diversity indices for ecological networks: A unifying framework using Hill numbers](#). *Ecology Letters*, 22, 737–747.
- 669 Olesen, J.M., Bascompte, J., Dupont, Y.L., Elberling, H., Rasmussen, C. & Jordano, P. (2010). [Missing and forbidden links in mutualistic networks](#). *Proceedings of the Royal Society B: Biological Sciences*, 278, 725–732.
- 672 Pascual, M., Dunne, J.A. & Dunne, J.A. (2006). *Ecological Networks: Linking Structure to Dynamics in Food Webs*. Oxford University Press, USA.
- 674 Petchey, O.L., Beckerman, A.P., Riede, J.O. & Warren, P.H. (2008). [Size, foraging, and food web structure](#).  
675 *Proceedings of the National Academy of Sciences*, 105, 4191–4196.
- 676 Pilosof, S., Porter, M.A., Pascual, M. & Kéfi, S. (2017). [The multilayer nature of ecological networks](#). *Nature Ecology & Evolution*, 1, 1–9.
- 678 Poisot, T., Canard, E., Mouillot, D., Mouquet, N. & Gravel, D. (2012). [The dissimilarity of species interaction networks](#). *Ecology Letters*, 15, 1353–1361.

- 680 Poisot, T., Cirtwill, A.R., Cazelles, K., Gravel, D., Fortin, M.-J. & Stouffer, D.B. (2016). [The structure of probabilistic networks](#). *Methods in Ecology and Evolution*, 7, 303–312.
- 681
- 682 Poisot, T. & Gravel, D. (2014). [When is an ecological network complex? Connectance drives degree distribution and emerging network properties](#). *PeerJ*, 2, e251.
- 683
- 684 Poisot, T., Stouffer, D.B. & Gravel, D. (2015). [Beyond species: Why ecological interaction networks vary through space and time](#). *Oikos*, 124, 243–251.
- 685
- 686 Pollock, L.J., Tingley, R., Morris, W.K., Golding, N., O'Hara, R.B., Parris, K.M., *et al.* (2014). [Understanding co-occurrence by modelling species simultaneously with a Joint Species Distribution Model \(JSDM\)](#). *Methods in Ecology and Evolution*, 5, 397–406.
- 687
- 688
- 689 Pomeranz, J.P.F., Thompson, R.M., Poisot, T. & Harding, J.S. (2019). [Inferring predatorprey interactions in food webs](#). *Methods in Ecology and Evolution*, 10, 356–367.
- 690
- 691 Pringle, R.M. & Hutchinson, M.C. (2020). [Resolving Food-Web Structure](#). *Annual Review of Ecology, Evolution, and Systematics*, 51, 55–80.
- 692
- 693 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.
- 694
- 695 Pulliam, H.R. (1974). [On the Theory of Optimal Diets](#). *The American Naturalist*, 108, 59–74.
- 696
- 697 Rooney, N. & McCann, K.S. (2012). [Integrating food web diversity, structure and stability](#). *Trends in Ecology & Evolution*, 27, 40–46.
- 698
- 699 Salim, J.A., Saraiva, A.M., Zermoglio, P.F., Agostini, K., Wolowski, M., Drucker, D.P., *et al.* (2022). [Data standardization of plantpollinator interactions](#). *GigaScience*, 11, giac043.
- 700
- 701 Sanders, D. & van Veen, F.J.F. (2012). [Indirect commensalism promotes persistence of secondary consumer species](#). *Biology Letters*, 8, 960–963.
- 702
- 703 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.
- 704
- 705 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.

- 706 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.
- 707
- 708 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.
- 709
- 710 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.
- 711
- 712
- 713 Strydom, T., Catchen, M.D., Banville, F., Caron, D., Dansereau, G., Desjardins-Proulx, P., et al. (2021). [A roadmap towards predicting species interaction networks \(across space and time\)](#). *Philosophical Transactions of the Royal Society B-Biological Sciences*, 376, 20210063.
- 714
- 715
- 716 Tylianakis, J.M., Tscharntke, T. & Lewis, O.T. (2007). [Habitat modification alters the structure of tropical host-parasitoid food webs](#). *Nature*, 445, 202–205.
- 717
- 718 Vázquez, D.P., Peralta, G., Cagnolo, L., Santos, M. & Igual, E. autores contribuyeron por. (2022). [Ecological interaction networks. What we know, what we don't, and why it matters](#). *Ecología Austral*, 32, 670–697.
- 719
- 720 Weinstein, B.G. & Graham, C.H. (2017a). [On comparing traits and abundance for predicting species interactions with imperfect detection](#). *Food Webs*, 11, 17–25.
- 721
- 722 Weinstein, B.G. & Graham, C.H. (2017b). [Persistent bill and corolla matching despite shifting temporal resources in tropical hummingbird-plant interactions](#). *Ecology Letters*, 20, 326–335.
- 723
- 724 Wood, S.A., Russell, R., Hanson, D., Williams, R.J. & Dunne, J.A. (2015). [Effects of spatial scale of sampling on food web structure](#). *Ecology and Evolution*, 5, 3769–3782.
- 725
- 726 Woodward, G., Benstead, J.P., Beveridge, O.S., Blanchard, J., Brey, T., Brown, L.E., et al. (2010). [Chapter 2 - Ecological Networks in a Changing Climate](#). In: *Advances in Ecological Research*, Ecological Networks (ed. Woodward, G.). Academic Press, pp. 71–138.
- 727
- 728
- 729 Woodward, G., Brown, L.E., Edwards, F.K., Hudson, L.N., Milner, A.M., Reuman, D.C., et al. (2012). [Climate change impacts in multispecies systems: Drought alters food web size structure in a field experiment](#). *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367, 2990–2997.
- 730
- 731



**Figure 1: Network accumulation curves.** (a) Dissimilarity of interactions between common species and (b) dissimilarity in species composition between aggregated local networks and the binary metaweb of host-parasite interactions. Aggregated local networks were obtained by sequentially and randomly selecting a number of binary local networks 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 networks. For a better comparison with binary networks, local probabilistic networks were derived from the probabilistic metaweb with a false positive and false negative rates of zero. A specific value of  $p$  (the local probability of interaction among potentially interacting species) was used for all local networks within a particular curve. Aggregated probabilistic local networks were obtained by sequentially and randomly selecting a number of probabilistic local networks and aggregating both their species and interactions (with the value of  $p$  adjusting according to eq. 5).



**Figure 2: Spatial scaling of interactions.** Expected number of host-parasite interactions in a network aggregating all probabilistic (a) local and (b) potential 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) potential interactions found in windows of specified length and central latitudes. Probabilities of potential interactions were obtained with a false positive rate of 5% and a false negative rate of 10%. Local probabilistic interactions were derived from the probabilistic metaweb by setting the value of  $p$  (the local probability of interaction among potentially interacting species) to 1, ensuring a conservative comparison between aggregated local networks and metawebs. Aggregated local networks were obtained by aggregating both the species and interactions found within a particular latitudinal window, with the values of  $p$  remaining at their maximum value of 1 following eq. 5.

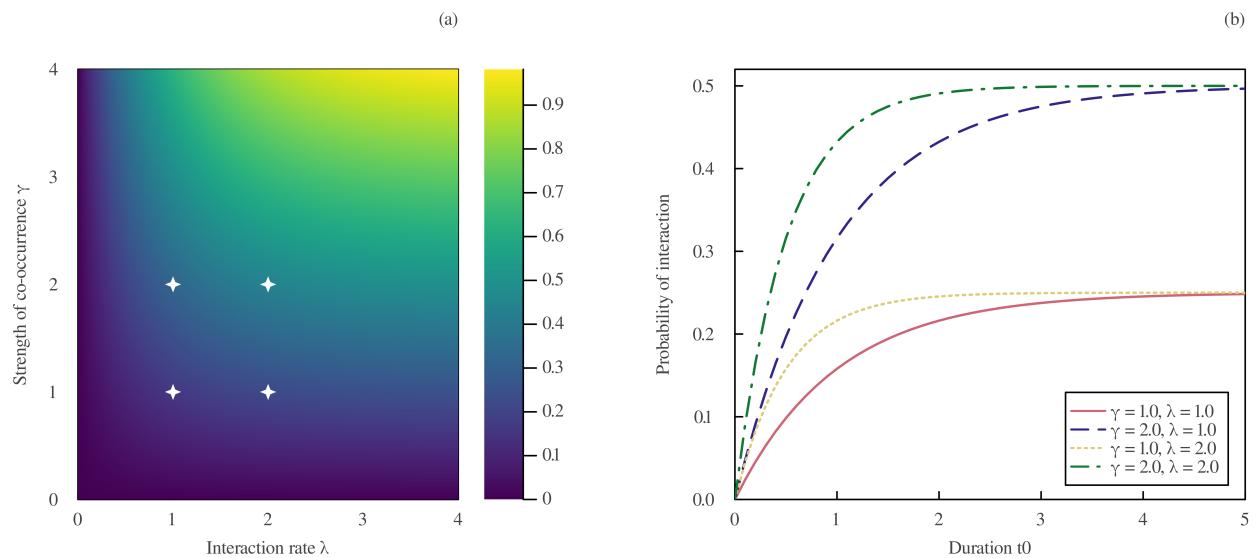
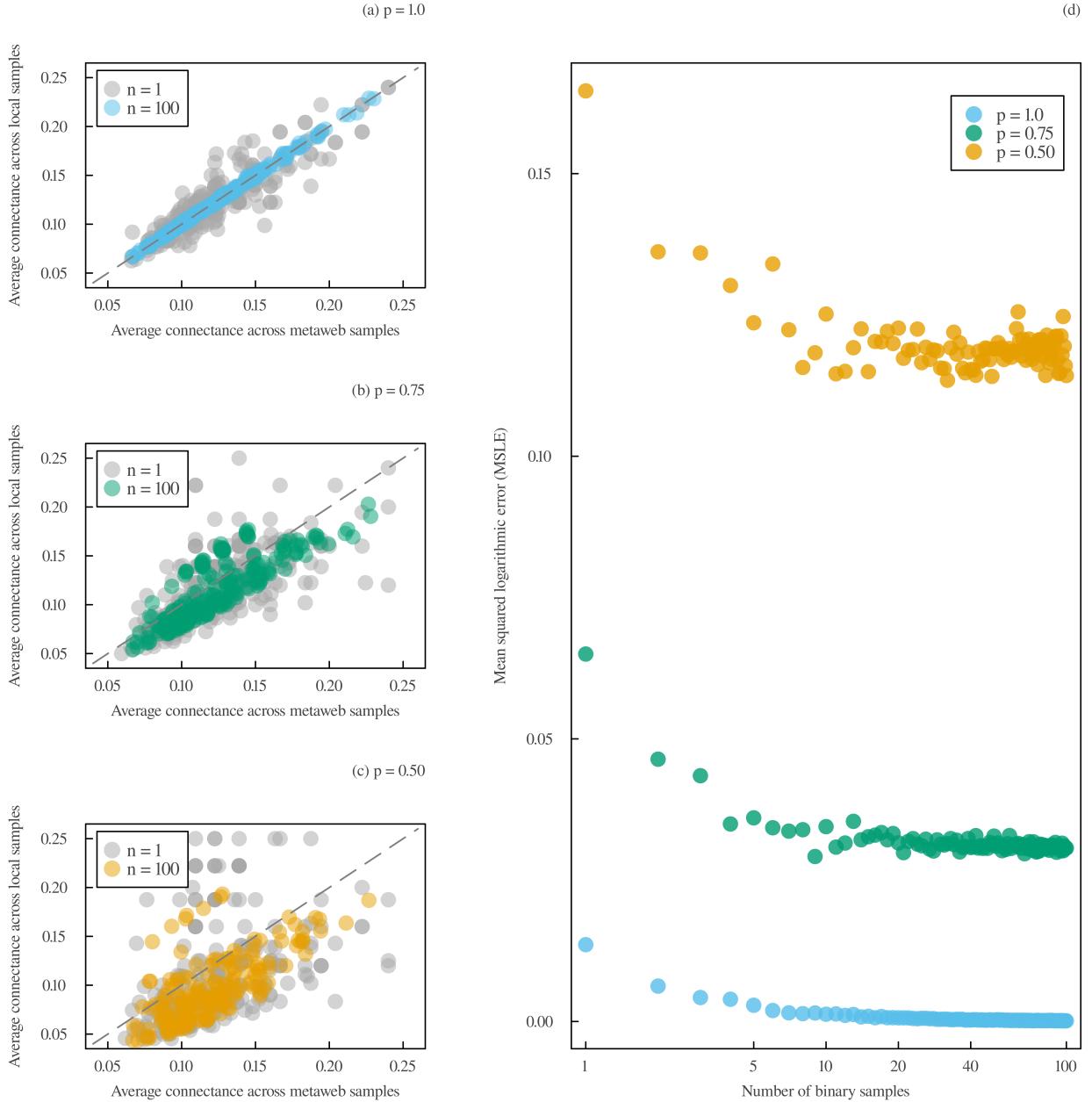


Figure 3: **Parameters of the spatiotemporally explicit model of interactions.** (a) Probability of local interaction given by the process model (eq. 11) under different values of  $\lambda$  and  $\gamma$ , with  $t_0 = 1$ . The parameter values used in the right panel are denoted by the white stars. (b) Scaling of the probability of interaction with the duration parameter  $t_0$  in eq. 11, for different values of  $\lambda$  and  $\gamma$ . In both panels, the individual probabilities of occurrence  $P_i(x, y, z)$  and  $P_j(x, y, z)$  are set to a constant value of 0.5.



**Figure 4: Connectance of sampled binary networks.** Comparison between the average connectance of binary network samples obtained from the probabilistic local networks and metaweb. Each dot corresponds to a different site. The local probability of interaction between potentially interacting species was set to three different values: (a)  $p = 1.0$ , (b)  $p = 0.75$ , and (c)  $p = 0.50$ . Grey dots represent the outcome of a single trial, while colored dots represent the average connectance of each network across 100 trials. (d) Reduction in the mean squared logarithmic error between the average connectance of binary networks obtained from these two sampling methods as the number of trials increases, for the same values of  $p$  used in panels a-c. Probabilities of potential 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 metaweb, and then propagating this result to all local networks that include the species potentially engaged in the interactions. Local binary networks were generated by independently sampling binary interactions for each local probabilistic network.