

# 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 Cataloging ecological interactions is a substantial challenge. There are methodological and biological  
3 constraints that hinder our ability to observe all interactions, such as the spatial and temporal uncoupling of  
4 species (Jordano (1987)) and the large number of possible interactions in a community, of which the vast  
5 majority are rare (Jordano (2016)). Documenting the location and timing of interactions becomes even more  
6 difficult when accounting for the spatiotemporal variability of ecological networks (Poisot *et al.* (2012), Poisot  
7 *et al.* (2015)). Indeed, knowing the biological capacity of two species to interact is necessary but not sufficient  
8 for inferring their interaction at a specific time and space. For example, Golubski & Abrams (2011) presented  
9 many cases where trophic interactions in food webs depend on the presence or abundance of a third species  
10 (e.g., of a more profitable prey species). More generally, a handful of conditions must be satisfied for an  
11 interaction to be observed locally. First, both species must have overlapping geographic ranges, i.e. they must  
12 co-occur within the region of interest (Blanchet *et al.* (2020)). Then, they must encounter locally. Probabilities  
13 of interspecific encounters are typically low, especially for rare species with low relative abundances (Canard *et*  
14 *al.* (2012)). Finally, their traits must be locally compatible (Poisot *et al.* (2015)). This includes their phenology  
15 (Olesen *et al.* (2010), Singer & McBride (2012)), behavioral choices (Pulliam (1974), Choh *et al.* (2012)) and  
16 phenotypes (Bolnick *et al.* (2011), Stouffer *et al.* (2011), Gravel *et al.* (2013)). Environmental factors, such as  
17 temperature (Angilletta *et al.* (2004)), drought (Woodward *et al.* (2012)), climate change (Gilman *et al.* (2010),  
18 Woodward *et al.* (2010), Araujo *et al.* (2011)), and habitat modifications (Tylianakis *et al.* (2007)), contribute  
19 to this spatiotemporal variability of interactions by impacting species abundance and traits. In this context, it is  
20 unsurprising that our knowledge of ecological interactions remains limited (Hortal *et al.* (2015)) despite  
21 extensive biodiversity data collection (Schmeller *et al.* (2015)).

22 The recognition of the intrinsic variability of species interactions has led ecologists to expand their  
23 representation of ecological networks to include a probabilistic view of interactions (Poisot *et al.* (2016)). As  
24 opposed to binary deterministic networks, in which interactions are regarded as either occurring or not,  
25 probabilistic networks represent our degree of belief about the realization or feasibility of pairwise interactions  
26 at the local or regional scale, respectively. In other words, representing interactions probabilistically takes into  
27 account inherent uncertainties and variability associated with ecological data. In the broadest sense, binary  
28 networks are also a type of probabilistic network, in which the numerical value of an interaction is restrained to  
29 0 (non-occurring) or 1 (occurring). In probabilistic networks, only forbidden interactions (i.e., interactions

prohibited by biological traits, Jordano *et al.* (2003), Olesen *et al.* (2010)) have a probability value of zero, provided that intraspecific trait variability is considered (Gonzalez-Varo & Traveset (2016)). However, *neutrally* forbidden interactions (i.e., improbable interactions between rare species, Canard *et al.* (2012)) could have low probability values in local networks but high probability in a regional network (metaweb) describing the biological capacity of species to interact.

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

60 (2023)), both for data producers and re-users who generate and manipulate these numbers. Besides  
61 methodological difficulties that may arise when assessing probabilistic interactions, a precise definition of  
62 probabilistic interactions appears to be lacking, making the estimation and use of these data more difficult. We  
63 aim to take a step back by outlining different ways in which probabilistic interactions are defined and used in  
64 network ecology. We distinguish two broad categories of probabilistic networks that necessitate distinct  
65 approaches when applied to key ecological questions: local networks of realized interactions, and regional  
66 networks (metawebs) of potential interactions. We highlight the distinctions in the ecological meaning of these  
67 two representations and show that they yield different statistical outcomes regarding e.g. the spatial and  
68 temporal scaling of interactions and the prediction of binary networks across space. Moreover, there is currently  
69 no metadata standard that could guide the documentation of all types of probabilistic interactions (Salim *et al.*  
70 (2022) discuss data standards for deterministic mutualistic networks). Well-defined metadata for probabilistic  
71 networks would support more adequate manipulation and integration of interaction data from different sources  
72 and guard against the potential misinterpretations arising from ambiguous definitions of probabilistic networks.  
73 These metadata should outline the nature (i.e., potential or local) and type (e.g., predatory or pollination) of the  
74 interactions, provide information regarding the taxonomic level, identities, and characteristics (e.g., life stages)  
75 of the individuals involved in an interaction, present the mathematical formulation of probabilities, including  
76 clearly identified conditional variables, and describe the methods and contexts (e.g., location, time,  
77 environmental conditions) in which interactions were estimated. Inadequately documented probabilistic  
78 interaction data should be used with caution when analyzing ecological networks. Although our contribution  
79 focuses on food webs, our observations and advice can be applied to other types of ecological networks, from  
80 plant-pollinator to host-virus networks. Indeed, excluding networks of indirect interactions such as competition  
81 and facilitation networks (Kéfi *et al.* (2015), Kéfi *et al.* (2016)), most ecological networks describe probabilities  
82 of direct interactions, which are conceptually and mathematically analogous regardless of their biological type  
83 (e.g., trophic and mutualistic interactions).

## 84 **Probabilistic representations of interactions**

85 One of the first aspects to take into consideration when estimating or interpreting probabilities of interactions is  
86 knowing if they describe potential or realized interactions. A potential interaction is defined as the biological  
87 capacity of two taxa to interact (i.e., the probability that they *can* theoretically interact) whereas a realized

interaction is the materialization or observation of this interaction in a well-defined space and time (i.e., the probability that they interact locally). Here, we use the terms *metaweb* (Dunne (2006)) to designate networks of potential interactions and *local networks* (Poisot *et al.* (2012)) for those of realized interactions. Metawebs are the network analogues of the species pool, where local networks originate from a subset of both species (nodes) and interactions (edges) of the regional metaweb (Saravia *et al.* (2022)). Without clear documentation, it can be challenging to know if published probabilistic networks describe potential or realized interactions, or if so-called probabilities are in reality *interaction scores* (i.e., non-probabilistic quantitative interactions). When probabilistic potential interactions are used and interpreted as local interactions (and conversely), this may generate misleading findings during data analysis. We believe that a better understanding of the differences, similarities, and relationships between these two probabilistic representations of ecological networks would alleviate interpretation errors and facilitate a more adequate utilization of interaction data.

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

Local ecological networks and metawebs, like any type of network, are made of nodes and edges that can be represented at different levels of organization and precision. The basic unit of food webs and other ecological networks are individuals that interact with each other (e.g., by predation, Elton (2001)), forming individual-based networks (Melián *et al.* (2011)). The aggregation of these individuals into more or less homogeneous groups (e.g., populations, species, trophic species, families) allows us to represent nodes at broader taxonomic scales, which affects our interpretation of the properties of these systems (Guimarães (2020)). Moreover, edges linking these nodes can describe a variety of interaction measures. Ecologists have traditionally represented interactions as binary objects that were considered realized after observing at least one individual from group  $i$  interact with at least another individual from group  $j$ . Boolean interactions can be 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 interaction between  $i$  and  $j$  that characterizes our limited knowledge of the system and its intrinsic spatiotemporal variability. Depending on the type of networks (local or metaweb), the mathematical formulation and interpretation of stochastic parameters like  $P(i \rightarrow j)$  can be linked to environmental and biological factors such as species relative abundance, traits, area, and time (tbl. 1), for example using logistic regression with a logit link function with continuous explanatory variables. In these probabilistic network representations in which  $P(i \rightarrow j)$  are edge values, the only two possible outcomes are the presence ( $A_{i,j} = 1$ ) or absence ( $A_{i,j} = 0$ ) of an interaction between each pair of nodes. When considering uncertainties around the

117 estimation of  $P(i \rightarrow j)$ , a Beta distribution  $\text{Beta}(\alpha, \beta)$  can also be used to encompass all possible probability  
118 values. Observing an interaction between two taxa at a given location and time provides important information  
119 that can be used to update previous estimates of  $P(i \rightarrow j)$ , informing us on the biological capacity of both taxa  
120 to interact and the environmental conditions that enabled them to interact locally.

121 Even though binary networks constitute a highly valuable source of ecological information (Pascual *et al.*  
122 (2006)), they overlook important factors regarding interaction strengths. These are represented using  
123 quantitative interactions, which better describe the energy flows, demographic impacts or frequencies of  
124 interactions between nodes (Berlow *et al.* (2004), Borrett & Scharler (2019)), with  $A_{i,j}$  being a natural number  
125  $\mathbb{N}$  or a real number  $\mathbb{R}$  depending on the measure. For example, they may represent local interaction rates  
126 between pairs of taxa (e.g., the flower visiting rates of pollinators in a mutualistic network, Herrera (1989)).  
127 Because quantitative interactions can also describe predation pressure on prey taxa in food webs, they can be  
128 good estimators of the parameters describing species interactions in a Lotka-Volterra model (e.g., Emmerson &  
129 Raffaelli (2004)). The extra amount of ecological information in quantitative networks typically comes at a cost  
130 of greater sampling effort and data volume in predictive models (Strydom *et al.* (2021)), which can lead to  
131 relatively high levels of uncertainties when inferring quantitative networks with limited data. Just like binary  
132 networks, the uncertainty and spatiotemporal variability of quantitative interactions can be represented  
133 probabilistically. However, the need to estimate the probability distribution of all possible value of interaction  
134 strengths can make the inference of probabilities more challenging in quantitative networks compared to binary  
135 networks, which require only one probability estimate for each interaction. Quantitative interactions can follow  
136 various probability distributions depending on the measure used, the event's outcome being the value of  
137 interaction strength. For instance, quantitative interactions can follow a Poisson distribution  
138  $A_{i,j} \sim \text{Poisson}(\lambda_{i \rightarrow j})$  when predicting frequencies of interactions between pairs of nodes, with  $\lambda_{i \rightarrow j}$  being the  
139 expected rate at which individuals of taxa  $i$  and  $j$  interact (e.g., the average number of prey  $j$  consumed by all  
140 predators  $i$  in a given time period). The Poisson distribution can also be 0-inflated after initially modeling  
141 non-interacting taxa (e.g., Boulangeat *et al.* (2012) employ a 0-inflated model to analyze species abundance  
142 following the modeling of species presence and absence), which constitute the majority of taxa pairs in most  
143 local networks due to their typically high sparseness (Jordano (2016)). Because of the methodological  
144 difficulties typically encountered when building deterministic quantitative networks, binary networks, which are  
145 easier to sample (Jordano (2016)) and predict (Strydom *et al.* (2021)), are much more documented and modeled  
146 in the literature. Moreover, most published probabilistic networks and methods describe Bernoulli interactions

147 (whether interaction probabilities are regarded as constant or variable, e.g. represented by Beta distributions),  
148 which underlines the need for better guidelines regarding the interpretation and manipulation of these types of  
149 networks. For these reasons, our primary focus in this contribution will be on addressing the challenges in  
150 interpreting and using Bernoulli interactions, in both probabilistic local networks and metawebs.

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

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

161 We define space as the set of geographic coordinates ( $x, y, z$ ) of the spatial boundaries delineating the system,  
162 whether sampled or targeted. Ecological interactions may exhibit variations spatially and along altitudinal  
163 gradients, as evidenced by changes in hummingbird-plant interactions (Weinstein & Graham (2017b) and  
164 Weinstein & Graham (2017a)) and mosquitoes biting rates (e.g., Kulkarni *et al.* (2006)) at different elevations.  
165 In contrast, time is defined as the specific time period within which interactions were either observed or  
166 predicted. With these definitions in mind, space and time can be conceptualized as distinct patches or time  
167 segments. Treating them as discrete dimensions aligns with the common sampling methods of ecological  
168 networks and provides actual probabilities of interactions across space and time. This differs from the approach  
169 of treating them as continuous variables, where edge values represent probability densities (i.e., relative  
170 likelihoods of interactions occurring at infinitesimal locations and instants in time) rather than discrete  
171 probability values (which can be obtained by integrating probability densities). By employing discrete locations  
172 and time periods, we can quantify both an area  $A$  and a duration  $t$ , which can be readily used in spatiotemporal  
173 analyses of ecological networks. For example, when studying network-area relationships (NAR, Galiana *et al.*  
174 (2018)), we anticipate that local probabilities of interactions scale positively with area and duration because  
175 taxa have more opportunities to interact as these dimensions expand.

176 The probability that two taxa  $i$  and  $j$  interact locally can be conditional on many environmental and biological  
 177 factors. One of these is their co-occurrence  $C_{i,j}$ , which is usually a Boolean describing if the geographic  
 178 distributions of both taxa overlap within the study area. Co-occurrence can be modeled probabilistically, in  
 179 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  
 180 co-occurrence  $P_{i,j}(x, y, z)$  can be estimated through the application of joint species distribution models (e.g.,  
 181 Pollock *et al.* (2014)). Considering that the probability that two non co-occurring taxa interact locally is zero  
 182 (i.e.,  $P_N(i \rightarrow j|C = 0) = 0$ ), the probability of local interaction can be obtained by multiplying the probability of  
 183 interaction given co-occurrence with the probability of co-occurrence:

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

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

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

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

198 The representation of the local context in which probabilities are estimated and the variables that should be  
 199 taken into consideration depend on the study system, the objectives of the study, and the resources available to

the researchers. In other words, these variables do not systematically need to be accounted for. For example, in Gravel *et al.* (2019), for the purpose of model comparison, some models incorporated environmental variables as conditional factors to estimate interaction probabilities, while others did not. When accounted for, these variables should be clearly described in the documentation of the data (Brimacombe *et al.* (2023)), preferentially in mathematical terms to avoid any confusion in their interpretation and to limit manipulation errors during their re-use. For instance, ecologists should be explicit about their consideration of co-occurrence in their estimation of local interaction probabilities. Indeed, it is important to specify if probability values are conditional  $P_N(i \rightarrow j|C = 1)$  or not  $P_N(i \rightarrow j)$  on co-occurrence since this can significantly impact the interpretation and analysis of the data. In [tbl. 1](#), we present examples of studies that used these diverse 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)
$P_N(i \rightarrow j \Omega)$	local	realization of the interaction given environmental conditions	Gravel <i>et al.</i> (2019) (temperature and precipitation)

210 **Metawebs: regional catalogs of interactions**

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

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

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

227 Starting from a selected set of taxa, which are usually distributed within a broad region of interest, probabilistic  
228 metawebs can be built using different data sources, including literature review, fieldwork, and predictive models  
229 (e.g., the metaweb of trophic interactions between Canadian mammals inferred by Strydom *et al.* (2022)).  
230 Every pair of taxa that have confidently been observed to interact at least once can be given a probability of 1  
231 (i.e.,  $P_M(i \rightarrow j) = 1$ ) since we know that they *can* interact. This is not the case in local probabilistic networks,  
232 in which interaction events usually remain stochastic (i.e.,  $P_N(i \rightarrow j) < 1$ ) after empirically observing  
233 interactions because of their intrinsic spatiotemporal variability. Similarly, although rare interactions typically  
234 have low probabilities in local networks, they can have high probabilities in metawebs when the traits of both  
235 taxa are congruent. On the other hand, interactions that were never observed can have low probability values in  
236 both metawebs and local networks, going as low as 0 for forbidden links. However, because of observation

237 errors due to taxonomic misidentifications and ecological misinterpretations (e.g., due to cryptic species and  
238 interactions, Pringle & Hutchinson (2020)), many observations of interactions are in reality false positives.  
239 Likewise, forbidden interactions can be false negatives in metawebs, e.g. if they have been assessed for specific  
240 phenotypes, locations or time. Employing Bayesian models, whether they are mechanistic or  
241 phenomenological, has the potential to enhance the accuracy of our estimation of interaction probabilities  
242 within both local and potential networks (e.g., Bartomeus *et al.* (2016), Cirtwill *et al.* (2019)). This  
243 improvement is achieved in potential networks by leveraging prior information regarding the feasibility of  
244 interactions along with the empirical data on observed interactions.

## 245 Properties of probabilistic networks

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

## 253 Host-parasite network data

254 We use the collection of tripartite host-parasite networks sampled across Europe, created by Kopelke *et al.*  
255 (2017), in most of our case studies. This dataset contains well-resolved binary local interactions between  
256 willows (52 species), willow-galling sawflies (96 species), and their parasitoids (126 species). Given its  
257 replicated networks spanning large spatiotemporal scales, this dataset is well-suited for analyzing a variety of  
258 ecological hypotheses and processes. Out of a total of 374 local networks, we retained those containing at least  
259 5 species, resulting in a set of 233 georeferenced local networks. We built a binary metaweb by aggregating all  
260 local interactions, which gave us a regional metaweb composed of 274 species and 1080 interactions. In the first  
261 two panels of fig. 1, we show how the dissimilarity of interactions between common species ( $\beta_{OS}$ ) and the  
262 dissimilarity in species composition ( $\beta_S$ ) between the metaweb and aggregated local networks (Poisot *et al.*  
263 (2012)) vary with the number of sampled local networks. This shows that networks of local interactions are

264 highly dissimilar from the metaweb, both in terms of species and interactions, especially when only a limited  
265 number of sites has been sampled.

266 [Figure 1 about here.]

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

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

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

283 We built the probabilistic local networks from the binary ones by using the probabilistic metawebs and a  
284 constant value of  $p$  across interactions. Lower values of  $p$  indicate that feasible interactions rarely occur locally,  
285 intermediate values around 50% suggest considerable spatiotemporal variability, while higher values indicate  
286 that potential interactions are nearly always realized. Following eq. 4, the local probabilities of interaction  
287 between a given pair of taxa consistently remained equal to or below their probability of potential interaction.

288 In the last two panels of fig. 1, we show how the aggregated number of links and connectance (i.e., the  
289 proportion of all of the  $S^2$  possible links that are realized, where  $S$  represents species richness) scale with the

290 number of sampled local probabilistic networks, according to different values of  $p$ . When aggregating local  
291 probabilistic networks, the constancy of the probability of potential interaction across the entire study area  
292 means that any rise in the probability of local interaction is solely attributable to an increase in  $p$ . The  
293 probability  $p_{1,2}$  of local interaction among potentially interacting species in an aggregated network  $N_{1,2}$  is  
294 obtained by:

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

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

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

### 301 Taxonomic scaling of interactions

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

310 There are no inherent differences in the taxonomic scaling between local networks and metawebs, as only the  
311 nodes are defined taxonomically. In other words, the probability values of edges in both local networks (eq. 2)  
312 and metawebs (eq. 3) are not conditional on any taxonomic scale. The phylogenetic scale is tied to the definition  
313 of the event itself (i.e., the interaction between two taxa), not to the conditional variables. In both types of  
314 networks, transitioning to a broader level of organization can be accomplished directly by using probabilities  
315 from finer scales. For example, in a network with  $n_A$  species from genus A and  $n_B$  species from genus B, one

316 can compute the probability that at least one species from genus  $A$  interacts with at least one species from genus  
317  $B$  as follows:

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

318 where  $A_i$  and  $B_j$  are the species of the corresponding genus. If it is known that at least two of these species  
319 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.  
320 Canard *et al.* (2012) built a species-based network following a similar approach, by using simulated interactions  
321 between individuals derived from a neutral model (i.e., a model that assumed ecological equivalence among  
322 individuals). In contrast, a more sophisticated approach is necessary when transitioning from a broader to a  
323 finer level of organization. This is because knowledge of an interaction between two genera does not guarantee  
324 that all possible pairwise combinations of their species will also interact. One possible method is to build a  
325 finer-scale network by generating probabilities of interactions through random sampling from a beta  
326 distribution, parameterized by the broader-scale network.

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

343 **Spatial and temporal scaling of interactions**

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

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

369 On the other hand, local interactions scale both spatially and temporally, given that they have more opportunities  
370 to be realized and observed in larger areas and longer durations. This is attributed to factors such as a higher  
371 number of individuals, greater trait variations, and increased opportunities for encounters, as highlighted by

372 McLeod *et al.* (2020). For example, if a local probabilistic network  $N_1$  with an area  $A_1$  is derived from a larger  
373 network  $N_0$  with an area  $A_0$ , and  $A_1$  is entirely nested within  $A_0$ , interaction probabilities should be lower in the  
374 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  
375 probabilities could potentially be higher in the smaller area, contingent upon their environmental and biological  
376 conditions. Likewise, interaction probabilities are expected to be lower in networks with shorter durations when  
377 time intervals are nested. In fig. 2, we show how the expected number of local host-parasite interactions scales  
378 with area, represented as an expanding latitudinal window, in comparison with potential interactions. Even  
379 though we employed local probabilities of interactions equal to potential interactions for the purpose of  
380 comparison (i.e., using  $p = 1$  here), we notice that the total number of potential interactions scales more rapidly  
381 than local interactions. This is due to the fact that numerous potential interactions involve species that never  
382 co-occur, and as a result, these interactions are not captured in local networks.

383 [Figure 2 about here.]

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

385 Predicting local networks across time and space is a pivotal goal of network ecology (Strydom *et al.* (2021)). In  
386 a context of scarcity of interaction data (Jordano (2016)), ecologists must resort to predictive models for  
387 reconstructing networks at fine spatial and temporal scales. For instance, real-time biomonitoring data coupled  
388 with appropriate numerical models (Bohan *et al.* (2017)) can be employed to reconstruct local ecological  
389 networks, opening avenues for in-depth studies on local ecosystem functioning and dynamics. Apart from their  
390 predictive applications, statistical models can also be crafted for descriptive purposes, elucidating key  
391 parameters of interest such as probabilities of interactions. In such instances, the parameter values themselves  
392 offer valuable ecological insights. Various types of models (e.g., Bayesian and machine learning models) have  
393 been used for both predictive and descriptive purposes in the realm of ecological interactions (Strydom *et al.*  
394 (2021)). The probabilistic representation of interactions acknowledges the inherent uncertainty in these models,  
395 typically expressed through probability distributions. We introduce and develop a simple generative mechanistic  
396 model for probabilistic local interactions that takes into consideration their spatiotemporal variability (i.e. a  
397 spatiotemporally explicit model of local interactions). It is essential to note that our model is not designed for  
398 potential interactions, which are scale-independent. Rather, it could prove valuable for predicting local  
399 interactions across time and space by generating new interaction data following parameter inference.

400 As indicated by eq. 1, the probability that two taxa  $i$  and  $j$  interact at a specific location  $(x, y, z)$  is determined by  
 401 the product of their probability of interaction given co-occurrence and their probability of co-occurrence. First,  
 402 their co-occurrence probability can be calculated using their individual occurrence probabilities  $P_i(x, y, z)$  and  
 403  $P_j(x, y, z)$ , along with the strength of association  $\gamma$  between their occurrences and co-occurrence (Cazelles *et al.*  
 404 (2016)):

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

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

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

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

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

414 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  
 415 instance, if the duration  $t_0$  is measured in months,  $\lambda$  denote the expected number of interactions per month.  
 416 The occurrence of an interaction between  $i$  and  $j$  can be modeled as a Bernoulli trial with a probability of  
 417  $P_N(i \rightarrow j)$ . Consequently, a Bayesian inference model can be built based on the preceding equations to estimate  
 418 the value of the  $\lambda$  and  $\gamma$  parameters and generate novel interaction data:

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

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

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

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

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

425 [Figure 3 about here.]

## 426 Sampling binary networks

427 The prediction of binary interactions through random draws is an important application of probabilistic  
 428 networks. This approach proves beneficial for analyzing the structural characteristics of probabilistic networks,  
 429 particularly in the absence of specific analytical measures. By simulating independent Bernoulli trials for each  
 430 interaction, a binary network can be generated. A probability distribution of network properties can be obtained  
 431 by measuring network structure across multiple randomly generated networks (Poisot *et al.* (2016)). This  
 432 method enables the representation of the variability of network structure, albeit with potential biases when  
 433 connectance is low (Poisot & Gravel (2014), Chagnon (2015)). Employing this strategy to generate binary  
 434 networks under a null model facilitates null hypothesis significance testing, wherein the observed measure is  
 435 compared against the simulated distribution (e.g., Bascompte *et al.* (2003)). Additionally, randomly generating  
 436 binary networks across spatial and temporal dimensions aids in representing the spatiotemporal variability of  
 437 network structure, allowing the testing of ecological hypotheses regarding interactions on large spatial and  
 438 temporal scales.

439 There are at least two distinct approaches to sample binary networks from probabilistic webs across space, for

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

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

[Figure 4 about here.]

The choice of a sampling approach can influence the selection of grid cell size when delineating local communities within a broader region of interest. In the first approach, pairwise interactions remain constant irrespective of cell size since they are sampled only once from the regional network. However, in the second approach, local interaction probabilities are contingent on the network area. For instance, consider networks  $N_1$  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  $N_2$  as independent, the probability of interaction between taxa  $i$  and  $j$  in  $N_0$  is given by:

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

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

471 **Prediction of local networks from metawebs**

472 Metawebs serve as a valuable source of ecological information for predicting local networks across time and  
473 space. Binary local networks can be reconstructed by selecting a subset of taxa and interactions from the  
474 metaweb (Dunne (2006)). This implies that metawebs consistently contain more interactions than their  
475 corresponding local networks. Determining the list of taxa to select can be achieved empirically or through  
476 methods like range maps or species distribution models. As the species composition of a community is arguably  
477 less difficult to sample or predict than its interactions, the primary challenge lies in deciding which interactions  
478 to select from the metaweb, a task that may necessitate advanced statistical models and ecological expertise.  
479 Inferring the structure of local networks from the metaweb before predicting specific local pairwise interactions  
480 could hold promise (Strydom *et al.* (2021)), considering that the structure of local networks is constrained by  
481 the metaweb (Saravia *et al.* (2022)).  
482 Building probabilistic local networks from a probabilistic metaweb involves a reduction in the value of pairwise  
483 interaction probabilities. This decrease is attributed to the prerequisite that two taxa must initially possess the  
484 capacity to interact before engaging in local interactions (eq. 4). Therefore, inferring local networks from their  
485 metaweb while maintaining identical interaction probability values would introduce systematic biases into the  
486 predictions. In such cases, these networks would essentially represent smaller-scale metawebs of potential  
487 interactions, potentially leading to misinterpretations by being perceived as local interactions. As proposed by  
488 McLeod *et al.* (2021b), although metawebs lack the ability to capture the spatiotemporal variability of  
489 interactions, they establish an upper limit for local interactions. In other words, the probability of two taxa  
490 interacting at a specific location and time is consistently lower or equal to the probability of their potential  
491 interaction:

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

492 Moreover, the probability that two taxa possess the biological capacity to interact must exceed the probability of

493 them interacting at any location and time because they may never co-occur or encounter locally. Specifically,  
494 the cumulative probability of realized interactions across all spatial, temporal, and environmental conditions  
495 must be less than the probability of potential interaction, i.e.

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

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

## 501 Conclusion

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

514 It is essential to enhance our comprehension of both potential and local interactions, especially considering the  
515 current scarcity of interaction data. However, while sampling biological communities does decrease the  
516 uncertainty of interactions by accumulating evidence for their feasibility and local realization, there is a limit to  
517 how much we can diminish their inherent randomness. In metawebs, probabilities reflect our limited knowledge

518 of interactions (i.e., our degree of belief that interactions are feasible), which is expected to improve with a  
519 larger volume of data. We should anticipate that potential interactions will become more definitive (with  
520 probabilities approaching 0 or 1) as we investigate various conditions, including different combinations of  
521 species traits. Conversely, in the case of local networks, which can be seen as random instances of metawebs,  
522 randomness cannot be as reduced. In local interactions, probabilities can be divided into a component  
523 representing uncertainty and another representing spatiotemporal variability. Owing to environmental  
524 heterogeneity, there will invariably be instances in which an interaction occurs and others in which it does not,  
525 across different times and locations, irrespective of the extent to which we can improve our knowledge of its  
526 biological feasibility and the local conditions that facilitate its occurrence. When local networks depict  
527 probabilities of observing interactions rather than just their actual occurrence, we must also consider the  
528 observation variability as an additional source of randomness. Every ecological process is stochastic but there is  
529 also a possibility that a phenomenon goes undetected. Quantifying and partitioning this stochasticity will enable  
530 us to make more accurate predictions about ecological interactions at various spatial and temporal scales.

## 531 Acknowledgment

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

## 539 References

- 540 Andrade-Pacheco, R., Rerolle, F., Lemoine, J., Hernandez, L., Meïté, A., Juziwelo, L., *et al.* (2020). [Finding](#)  
541 [hotspots: Development of an adaptive spatial sampling approach](#). *Scientific Reports*, 10, 10939.  
542 Angilletta, M.J., Jr., Steury, T.D. & Sears, M.W. (2004). [Temperature, Growth Rate, and Body Size in](#)  
543 [Ectotherms: Fitting Pieces of a Life-History Puzzle1](#). *Integrative and Comparative Biology*, 44, 498–509.

- 544 Araujo, M.B., Rozenfeld, A., Rahbek, C. & Marquet, P.A. (2011). [Using species co-occurrence networks to](#)  
545 [assess the impacts of climate change](#). *Ecography*, 34, 897–908.
- 546 Bartomeus, I., Gravel, D., Tylianakis, J.M., Aizen, M.A., Dickie, I.A. & Bernard-Verdier, M. (2016). [A](#)  
547 [common framework for identifying linkage rules across different types of interactions](#). *Functional Ecology*,  
548 30, 1894–1903.
- 549 Bascompte, J., Jordano, P., Melián, C.J. & Olesen, J.M. (2003). [The nested assembly of plantanimal mutualistic](#)  
550 [networks](#). *Proceedings of the National Academy of Sciences*, 100, 9383–9387.
- 551 Berlow, E.L., Neutel, A.-M., Cohen, J.E., De Ruiter, P.C., Ebenman, B., Emmerson, M., *et al.* (2004).  
552 [Interaction strengths in food webs: Issues and opportunities](#). *Journal of Animal Ecology*, 73, 585–598.
- 553 Blanchet, F.G., Cazelles, K. & Gravel, D. (2020). [Co-occurrence is not evidence of ecological interactions](#).  
554 *Ecology Letters*, 23, 1050–1063.
- 555 Bohan, D.A., Vacher, C., Tamaddoni-Nezhad, A., Raybould, A., Dumbrell, A.J. & Woodward, G. (2017).  
556 [Next-Generation Global Biomonitoring: Large-scale, Automated Reconstruction of Ecological Networks](#).  
557 *Trends in Ecology & Evolution*, 32, 477–487.
- 558 Bolnick, D.I., Amarasekare, P., Araújo, M.S., Bürger, R., Levine, J.M., Novak, M., *et al.* (2011). [Why](#)  
559 [intraspecific trait variation matters in community ecology](#). *Trends in Ecology & Evolution*, 26, 183–192.
- 560 Borrett, S.R. & Scharler, U.M. (2019). [Walk partitions of flow in Ecological Network Analysis: Review and](#)  
561 [synthesis of methods and indicators](#). *Ecological Indicators*, 106, 105451.
- 562 Boulangeat, I., Gravel, D. & Thuiller, W. (2012). [Accounting for dispersal and biotic interactions to disentangle](#)  
563 [the drivers of species distributions and their abundances](#). *Ecology Letters*, 15, 584–593.
- 564 Brimacombe, C., Bodner, K., Michalska-Smith, M., Poisot, T. & Fortin, M.-J. (2023). [Shortcomings of reusing](#)  
565 [species interaction networks created by different sets of researchers](#). *PLOS Biology*, 21, e3002068.
- 566 Brose, U., Ostling, A., Harrison, K. & Martinez, N.D. (2004). [Unified spatial scaling of species and their](#)  
567 [trophic interactions](#). *Nature*, 428, 167–171.
- 568 Canard, E., Mouquet, N., Marescot, L., Gaston, K.J., Gravel, D. & Mouillot, D. (2012). [Emergence of](#)  
569 [Structural Patterns in Neutral Trophic Networks](#). *PLOS ONE*, 7, e38295.
- 570 Catchen, M.D., Poisot, T., Pollock, L.J. & Gonzalez, A. (2023). [The missing link: Discerning true from false](#)  
571 [negatives when sampling species interaction networks](#).

- 572 Cazelles, K., Araujo, M.B., Mouquet, N. & Gravel, D. (2016). A theory for species co-occurrence in interaction  
573 networks. *Theoretical Ecology*, 9, 39–48.
- 574 Chagnon, P.-L. (2015). Characterizing topology of ecological networks along gradients: The limits of metrics'  
575 standardization. *Ecological Complexity*, 22, 36–39.
- 576 Choh, Y., Ignacio, M., Sabelis, M.W. & Janssen, A. (2012). Predator-prey role reversals, juvenile experience  
577 and adult antipredator behaviour. *Scientific Reports*, 2, 728.
- 578 Cirtwill, A.R., Eklöf, A., Roslin, T., Wootton, K. & Gravel, D. (2019). A quantitative framework for  
579 investigating the reliability of empirical network construction. *Methods in Ecology and Evolution*, 10,  
580 902–911.
- 581 Dunne, J.A. (2006). The Network Structure of Food Webs. In: *Ecological networks: Linking structure and*  
582 *dynamics* (eds. Dunne, J.A. & Pascual, M.). Oxford University Press, pp. 27–86.
- 583 Elmasri, M., Farrell, M.J., Davies, T.J. & Stephens, D.A. (2020). A hierarchical Bayesian model for predicting  
584 ecological interactions using scaled evolutionary relationships. *The Annals of Applied Statistics*, 14,  
585 221–240.
- 586 Elton, C.S. (2001). *Animal Ecology*. University of Chicago Press, Chicago, IL.
- 587 Emmerson, M.C. & Raffaelli, D. (2004). Predator-prey body size, interaction strength and the stability of a real  
588 food web. *Journal of Animal Ecology*, 73, 399–409.
- 589 Fortuna, M.A. & Bascompte, J. (2006). Habitat loss and the structure of plant-animal mutualistic networks.  
590 *Ecology Letters*, 9, 281–286.
- 591 Galiana, N., Lurgi, M., Claramunt-López, B., Fortin, M.-J., Leroux, S., Cazelles, K., et al. (2018). The spatial  
592 scaling of species interaction networks. *Nature Ecology & Evolution*, 2, 782–790.
- 593 Gilman, S.E., Urban, M.C., Tewksbury, J., Gilchrist, G.W. & Holt, R.D. (2010). A framework for community  
594 interactions under climate change. *Trends in Ecology & Evolution*, 25, 325–331.
- 595 Godsoe, W., Murray, R. & Iritani, R. (2022). Species interactions and diversity: A unified framework using Hill  
596 numbers. *Oikos*, n/a, e09282.
- 597 Golubski, A.J. & Abrams, P.A. (2011). Modifying modifiers: What happens when interspecific interactions  
598 interact? *Journal of Animal Ecology*, 80, 1097–1108.

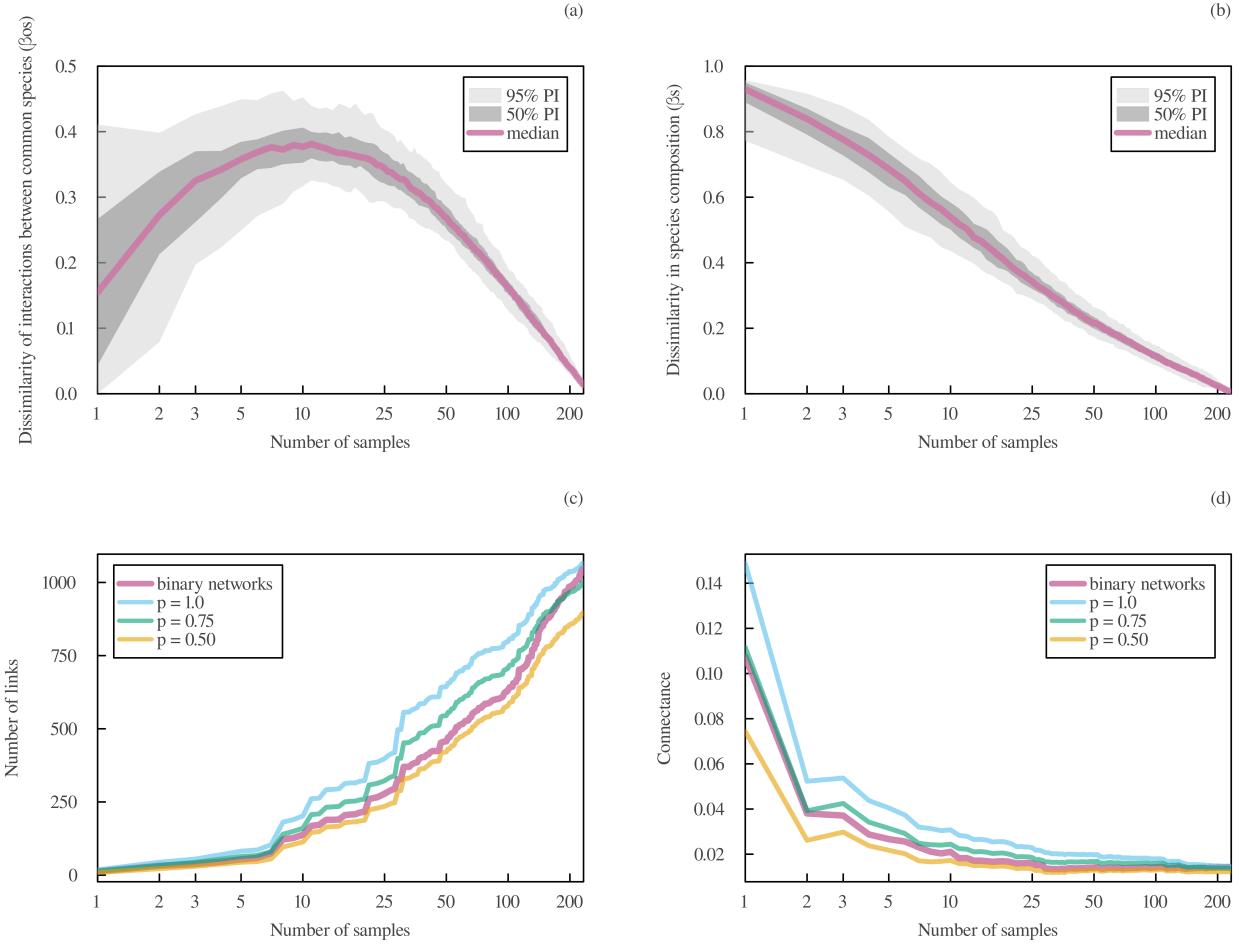
- 599 Gomez, J.M., Verdu, M. & Perfectti, F. (2010). Ecological interactions are evolutionarily conserved across the  
600 entire tree of life. *Nature*, 465, 918–U6.
- 601 Gonzalez-Varo, J.P. & Traveset, A. (2016). The Labile Limits of Forbidden Interactions. *Trends in Ecology &*  
602 *Evolution*, 31, 700–710.
- 603 Gravel, D., Baiser, B., Dunne, J.A., Kopalke, J.-P., Martinez, N.D., Nyman, T., et al. (2019). Bringing Elton  
604 and Grinnell together: A quantitative framework to represent the biogeography of ecological interaction  
605 networks. *Ecography*, 42, 401–415.
- 606 Gravel, D., Poisot, T., Albouy, C., Velez, L. & Mouillot, D. (2013). Inferring food web structure from  
607 predator-prey body size relationships. *Methods in Ecology and Evolution*, 4, 1083–1090.
- 608 Guimarães, P.R. (2020). The Structure of Ecological Networks Across Levels of Organization. *Annual Review*  
609 *of Ecology, Evolution, and Systematics*, 51, 433–460.
- 610 Guimerà, R. & Sales-Pardo, M. (2009). Missing and spurious interactions and the reconstruction of complex  
611 networks. *Proceedings of the National Academy of Sciences*, 106, 22073–22078.
- 612 Hemprich-Bennett, D.R., Oliveira, H.F.M., Le Comber, S.C., Rossiter, S.J. & Clare, E.L. (2021). Assessing the  
613 impact of taxon resolution on network structure. *Ecology*, 102, e03256.
- 614 Herrera, C.M. (1989). Pollinator abundance, morphology, and flower visitation rate: Analysis of the “quantity”  
615 component in a plant-pollinator system. *Oecologia*, 80, 241–248.
- 616 Hortal, J., de Bello, F., Diniz-Filho, J.A.F., Lewinsohn, T.M., Lobo, J.M. & Ladle, R.J. (2015). Seven Shortfalls  
617 that Beset Large-Scale Knowledge of Biodiversity. *Annual Review of Ecology, Evolution, and Systematics*,  
618 46, 523–549.
- 619 Jordano, P. (1987). Patterns of Mutualistic Interactions in Pollination and Seed Dispersal: Connectance,  
620 Dependence Asymmetries, and Coevolution. *The American Naturalist*, 129, 657–677.
- 621 Jordano, P. (2016). Sampling networks of ecological interactions. *Functional Ecology*, 30, 1883–1893.
- 622 Jordano, P., Bascompte, J. & Olesen, J.M. (2003). Invariant properties in coevolutionary networks of  
623 plant-animal interactions. *Ecology Letters*, 6, 69–81.
- 624 Kéfi, S., Berlow, E.L., Wieters, E.A., Joppa, L.N., Wood, S.A., Brose, U., et al. (2015). Network structure  
625 beyond food webs: Mapping non-trophic and trophic interactions on Chilean rocky shores. *Ecology*, 96,  
626 291–303.

- 627 Kéfi, S., Berlow, E.L., Wieters, E.A., Navarrete, S.A., Petchey, O.L., Wood, S.A., *et al.* (2012). [More than a meal... integrating non-feeding interactions into food webs](#). *Ecology Letters*, 15, 291–300.
- 628
- 629 Kéfi, S., Miele, V., Wieters, E.A., Navarrete, S.A. & Berlow, E.L. (2016). [How Structured Is the Entangled Bank? The Surprisingly Simple Organization of Multiplex Ecological Networks Leads to Increased Persistence and Resilience](#). *PLOS Biology*, 14, e1002527.
- 630
- 631
- 632 Klecka, J. & Boukal, D.S. (2014). [The effect of habitat structure on prey mortality depends on predator and prey microhabitat use](#). *Oecologia*, 176, 183–191.
- 633
- 634 Kopelke, J.-P., Nyman, T., Cazelles, K., Gravel, D., Vissault, S. & Roslin, T. (2017). [Food-web structure of willow-galling sawflies and their natural enemies across Europe](#). *Ecology*, 98, 1730–1730.
- 635
- 636 Kulkarni, M.A., Kweka, E., Nyale, E., Lyatuu, E., Mosha, F.W., Chandramohan, D., *et al.* (2006). [Entomological Evaluation of Malaria Vectors at Different Altitudes in Hai District, Northeastern Tanzania](#). *Journal of Medical Entomology*, 43, 580–588.
- 637
- 638
- 639 McCann, K. (2007). Protecting biostructure. *Nature*, 446, 29–29.
- 640 McCann, K.S. (2011). Food webs (MPB-50). In: *Food Webs (MPB-50)*. Princeton University Press.
- 641 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.
- 642
- 643 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.
- 644
- 645 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.
- 646
- 647 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.
- 648
- 649
- 650 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.
- 651
- 652 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,
- 653
- 654 725–732.

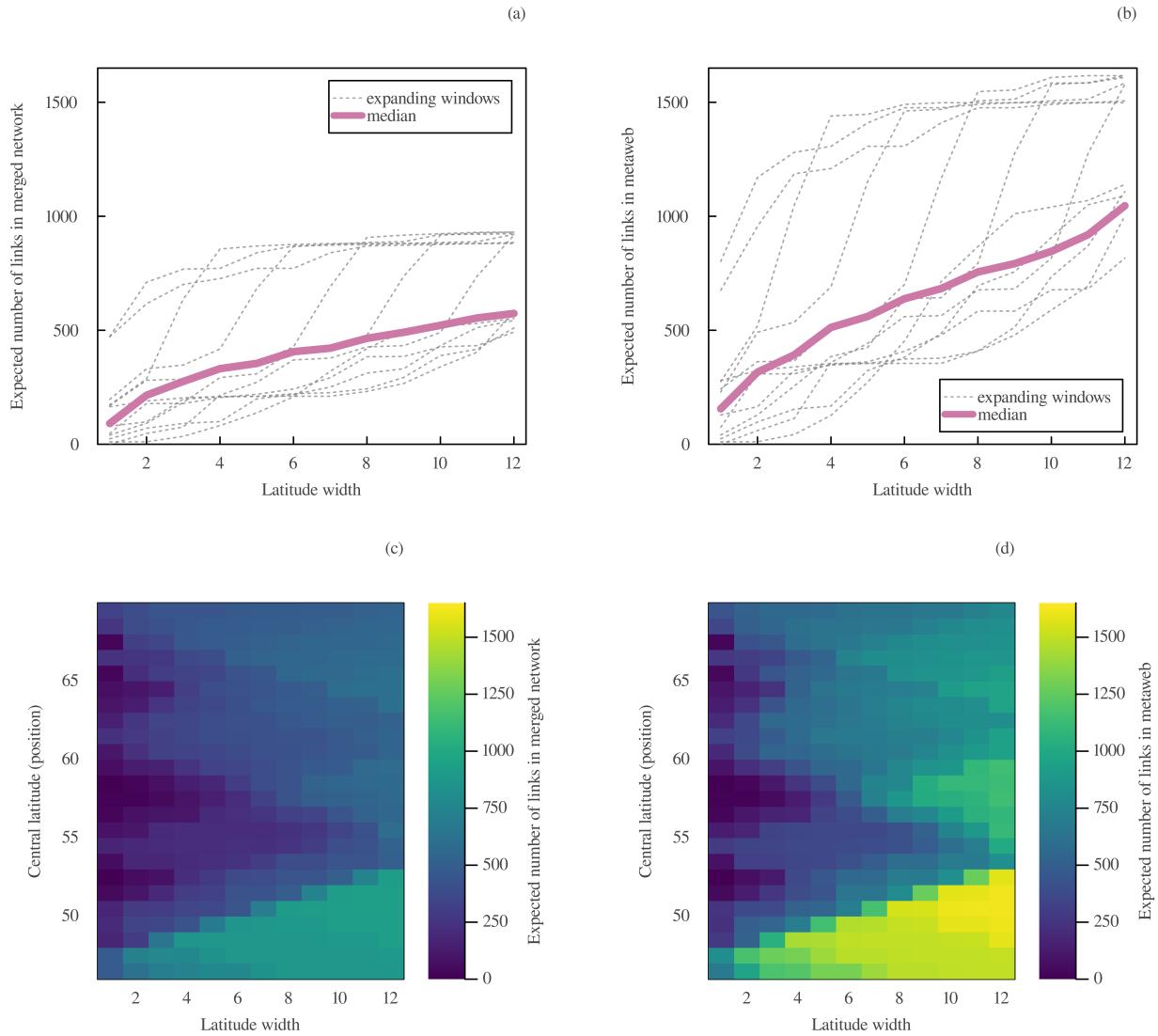
- 655 Pascual, M., Dunne, J.A. & Dunne, J.A. (2006). *Ecological Networks: Linking Structure to Dynamics in Food*  
656 *Webs*. Oxford University Press, USA.
- 657 Petchey, O.L., Beckerman, A.P., Riede, J.O. & Warren, P.H. (2008). [Size, foraging, and food web structure](#).  
658 *Proceedings of the National Academy of Sciences*, 105, 4191–4196.
- 659 Pilosof, S., Porter, M.A., Pascual, M. & Kéfi, S. (2017). [The multilayer nature of ecological networks](#). *Nature*  
660 *Ecology & Evolution*, 1, 1–9.
- 661 Poisot, T., Canard, E., Mouillot, D., Mouquet, N. & Gravel, D. (2012). [The dissimilarity of species interaction](#)  
662 [networks](#). *Ecology Letters*, 15, 1353–1361.
- 663 Poisot, T., Cirtwill, A.R., Cazelles, K., Gravel, D., Fortin, M.-J. & Stouffer, D.B. (2016). [The structure of](#)  
664 [probabilistic networks](#). *Methods in Ecology and Evolution*, 7, 303–312.
- 665 Poisot, T. & Gravel, D. (2014). [When is an ecological network complex? Connectance drives degree](#)  
666 [distribution and emerging network properties](#). *PeerJ*, 2, e251.
- 667 Poisot, T., Stouffer, D.B. & Gravel, D. (2015). [Beyond species: Why ecological interaction networks vary](#)  
668 [through space and time](#). *Oikos*, 124, 243–251.
- 669 Pollock, L.J., Tingley, R., Morris, W.K., Golding, N., O’Hara, R.B., Parris, K.M., *et al.* (2014). [Understanding](#)  
670 [co-occurrence bymodelling species simultaneously with a Joint Species DistributionModel \(JSDM\)](#).  
671 *Methods in Ecology and Evolution*, 5, 397–406.
- 672 Pomeranz, J.P.F., Thompson, R.M., Poisot, T. & Harding, J.S. (2019). [Inferring predatorprey interactions in](#)  
673 [food webs](#). *Methods in Ecology and Evolution*, 10, 356–367.
- 674 Pringle, R.M. & Hutchinson, M.C. (2020). [Resolving Food-Web Structure](#). *Annual Review of Ecology,*  
675 *Evolution, and Systematics*, 51, 55–80.
- 676 Proulx, S.R., Promislow, D.E.L. & Phillips, P.C. (2005). [Network thinking in ecology and evolution](#). *Trends in*  
677 *Ecology & Evolution*, SPECIAL ISSUE: BUMPER BOOK REVIEW, 20, 345–353.
- 678 Pulliam, H.R. (1974). [On the Theory of Optimal Diets](#). *The American Naturalist*, 108, 59–74.
- 679 Rooney, N. & McCann, K.S. (2012). [Integrating food web diversity, structure and stability](#). *Trends in Ecology*  
680 *& Evolution*, 27, 40–46.

- 681 Salim, J.A., Saraiva, A.M., Zermoglio, P.F., Agostini, K., Wolowski, M., Drucker, D.P., *et al.* (2022). [Data](#)  
682 [standardization of plantpollinator interactions](#). *GigaScience*, 11, giac043.
- 683 Saravia, L.A., Marina, T.I., Kristensen, N.P., De Troch, M. & Momo, F.R. (2022). [Ecological network](#)  
684 [assembly: How the regional metaweb influences local food webs](#). *Journal of Animal Ecology*, 91, 630–642.
- 685 Schmeller, D.S., Julliard, R., Bellingham, P.J., Böhm, M., Brummitt, N., Chiarucci, A., *et al.* (2015). [Towards a](#)  
686 [global terrestrial species monitoring program](#). *Journal for Nature Conservation*, 25, 51–57.
- 687 Singer, M.C. & McBride, C.S. (2012). [Geographic mosaics of species' association: A definition and an](#)  
688 [example driven by plantinsect phenological synchrony](#). *Ecology*, 93, 2658–2673.
- 689 Stouffer, D.B., Rezende, E.L. & Amaral, L.A.N. (2011). [The role of body mass in diet contiguity and food-web](#)  
690 [structure](#). *Journal of Animal Ecology*, 80, 632–639.
- 691 Strydom, T., Bouskila, S., Banville, F., Barros, C., Caron, D., Farrell, M.J., *et al.* (2022). [Food web](#)  
692 [reconstruction through phylogenetic transfer of low-rank network representation](#). *Methods in Ecology and*  
693 *Evolution*, 13.
- 694 Strydom, T., Catchen, M.D., Banville, F., Caron, D., Dansereau, G., Desjardins-Proulx, P., *et al.* (2021). [A](#)  
695 [roadmap towards predicting species interaction networks \(across space and time\)](#). *Philosophical*  
696 *Transactions of the Royal Society B-Biological Sciences*, 376, 20210063.
- 697 Tylianakis, J.M., Tscharntke, T. & Lewis, O.T. (2007). [Habitat modification alters the structure of tropical](#)  
698 [hostparasitoid food webs](#). *Nature*, 445, 202–205.
- 699 Vázquez, D.P., Peralta, G., Cagnolo, L., Santos, M. & Igual, E. autores contribuyeron por. (2022). [Ecological](#)  
700 [interaction networks. What we know, what we don't, and why it matters](#). *Ecología Austral*, 32, 670–697.
- 701 Weinstein, B.G. & Graham, C.H. (2017a). [On comparing traits and abundance for predicting species](#)  
702 [interactions with imperfect detection](#). *Food Webs*, 11, 17–25.
- 703 Weinstein, B.G. & Graham, C.H. (2017b). [Persistent bill and corolla matching despite shifting temporal](#)  
704 [resources in tropical hummingbird-plant interactions](#). *Ecology Letters*, 20, 326–335.
- 705 Wood, S.A., Russell, R., Hanson, D., Williams, R.J. & Dunne, J.A. (2015). [Effects of spatial scale of sampling](#)  
706 [on food web structure](#). *Ecology and Evolution*, 5, 3769–3782.
- 707 Woodward, G., Benstead, J.P., Beveridge, O.S., Blanchard, J., Brey, T., Brown, L.E., *et al.* (2010). [Chapter 2 -](#)  
708 [Ecological Networks in a Changing Climate](#). In: *Advances in Ecological Research*, Ecological Networks

- 709 (ed. Woodward, G.). Academic Press, pp. 71–138.
- 710 Woodward, G., Brown, L.E., Edwards, F.K., Hudson, L.N., Milner, A.M., Reuman, D.C., *et al.* (2012). [Climate](#)  
711 [change impacts in multispecies systems: Drought alters food web size structure in a field experiment.](#)
- 712 *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367, 2990–2997.



**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.

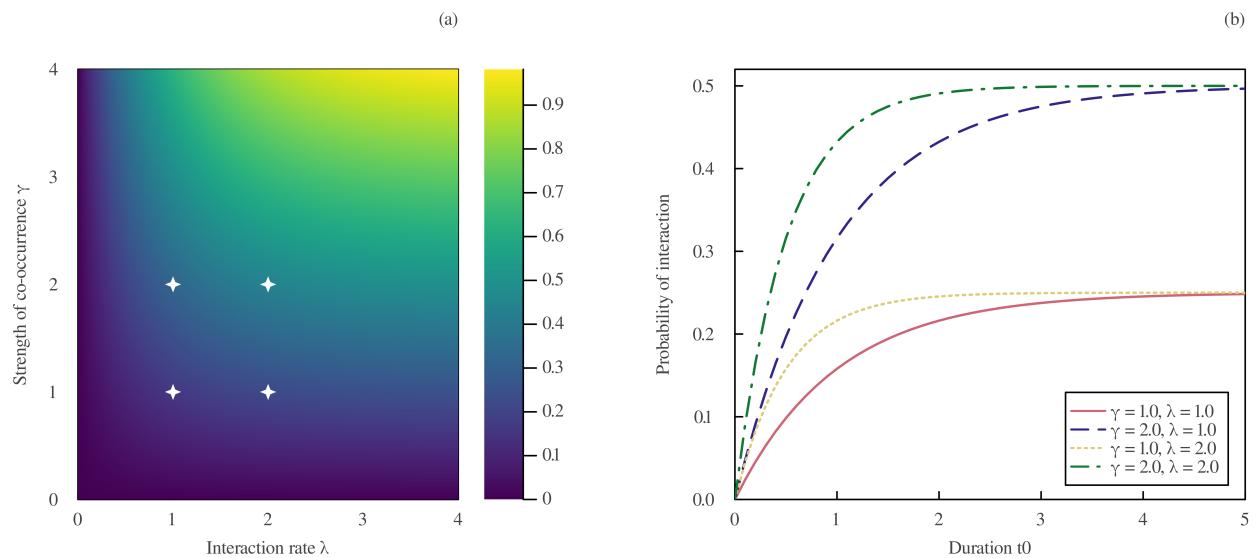
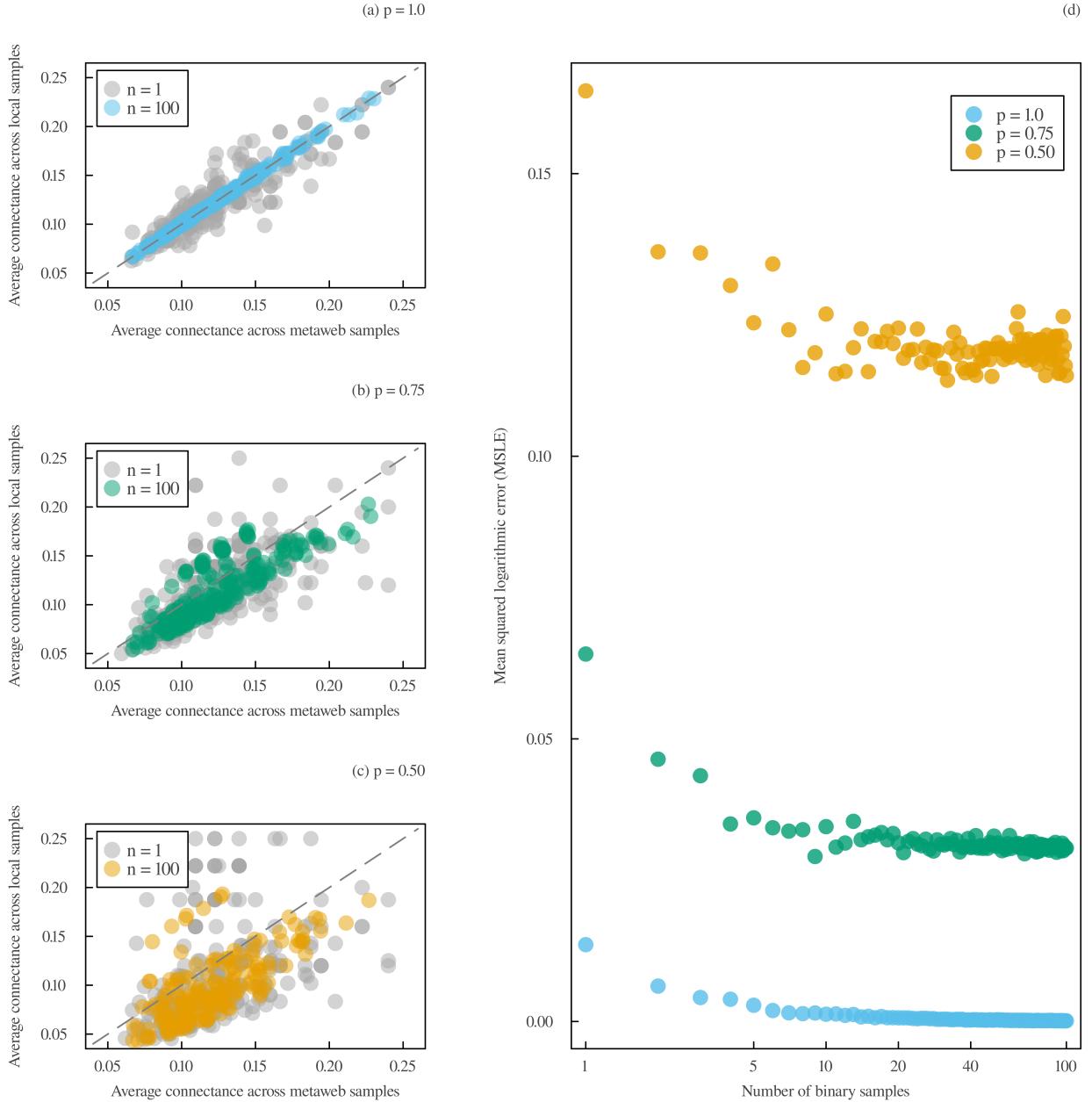


Figure 3: **Parameters of the spatiotemporally explicit model of interactions.** (a) Probability of local interaction given by the process model (eq. 11) under different values of  $\lambda$  and  $\gamma$ , with  $t_0 = 1$ . The parameter values used in the right panel are denoted by the white stars. (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.