

# 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 they are to occur?) 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 (predatory, herbivory, and parasitic interactions). These definitions are based on the distinction between the realization of an interaction at a specific time and space (local web) and its biological feasibility (metaweb). We show that different network representations have different statistical properties when it comes to common ecological applications. Specifically, we argue that probabilistic local 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 webs from subsets of metawebs. To support our arguments, we develop different case studies using open empirical data on host-parasite interactions in Europe. 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. Comprehensive metadata, providing explicit details on the sources of uncertainty in probabilistic webs, not only supports more rigorous statistical analyses but also broadens the scope of applications of interaction data. This clarity may help with the identification of optimal sampling locations to effectively minimize the uncertainty of trophic interactions.

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



Last revision: *February 11, 2024*

## **1 Introduction**

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

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

54 By accounting for the uncertainty of interactions, probabilistic webs may provide a more realistic portrait of  
55 species interactions and network structure (i.e. community-level properties), which are major drivers of the  
56 functioning, dynamics, and resilience of ecosystems worldwide (Proulx *et al.* (2005), McCann (2007), McCann  
57 (2011), Rooney & McCann (2012)). Moreover, the application and development of computational methods in  
58 network ecology, often based on a probabilistic representation of interactions, can alleviate (and guide) the  
59 sampling efforts required to document species interactions (Strydom *et al.* (2021)). For example, statistical  
60 models can be used to estimate the uncertainty of pairwise interactions (Cirtwill *et al.* (2019)) and the

61 probability of missing (false negatives) and spurious (false positives) interactions (Guimerà & Sales-Pardo  
62 (2009)). Considering the high rate of false negatives in species interaction data due to the difficulty of  
63 witnessing rare interactions (Catchen *et al.* (2023)), these models could inform the identification of priority  
64 sampling locations of ecological webs where data collection would yield the most valuable information, thereby  
65 reducing errors. Optimization models for sampling locations have mostly found applications in biological  
66 systems that are not networks, such as identifying priority sampling sites for disease hotspots (Andrade-Pacheco  
67 *et al.* (2020)), but there is substantial promise in applying them to probabilistic ecological interactions.  
68 Statistical models can also be used to generate predictions of ecological webs without prior knowledge of  
69 pairwise interactions, for instance using body size (Petchey *et al.* (2008), Gravel *et al.* (2013)), phylogeny  
70 (Elmasri *et al.* (2020), Strydom *et al.* (2022)), or a combination of niche and neutral processes (Bartomeus *et*  
71 *al.* (2016), Pomeranz *et al.* (2019)) for inference. Topological null models, which generate probabilistic  
72 networks by preserving chosen characteristics of the binary adjacency matrix while intentionally omitting  
73 others (Bascompte *et al.* (2003), Fortuna & Bascompte (2006)), serve as other examples of common  
74 probabilistic network models. Null models can be used to produce underlying distributions of network measures  
75 for null hypothesis significance testing. Many measures have been developed to describe the structure (Poisot *et*  
76 *al.* (2016)) and diversity (Ohlmann *et al.* (2019), Godsoe *et al.* (2022)) of probabilistic webs. These models and  
77 measures support the use of this approach for the study of a wide range of ecological questions, from making  
78 better predictions of species distribution (Cazelles *et al.* (2016)) to forecasting the impact of climate change on  
79 ecological webs (Gilman *et al.* (2010)).

80 The lack of clear guidelines on the use of probabilistic interaction data is worrisome, both for data producers  
81 and re-users who generate and manipulate these numbers. This is concerning because sampling strategies and  
82 decisions regarding network construction can affect our understanding of network properties (Brimacombe *et*  
83 *al.* (2023)). Besides methodological difficulties that may arise when assessing probabilistic interactions, a  
84 precise definition of probabilistic interactions appears to be lacking, making the estimation and use of these data  
85 more difficult. We aim to take a step back by outlining different ways in which probabilistic interactions are  
86 defined and used in network ecology. We distinguish two broad categories of probabilistic webs that necessitate  
87 distinct approaches when applied to key ecological questions: local webs describing probabilities of realized  
88 interactions, and regional webs (metawebs) describing probabilities of potential interactions. We highlight the  
89 distinctions in the ecological meaning of these two representations and show that they yield different statistical  
90 outcomes regarding e.g. the spatial and temporal scaling of interactions and the prediction of binary webs

91 across space. Moreover, there is currently no metadata standard that could guide the documentation of all types  
92 of probabilistic interactions (although see e.g., Salim *et al.* (2022) who discuss data standards for deterministic  
93 mutualistic webs). Well-defined metadata for probabilistic webs would support more adequate manipulation  
94 and integration of interaction data from different sources and guard against possible misinterpretations arising  
95 from ambiguous definitions of probabilistic networks. These metadata should outline the nature (i.e., local or  
96 regional) and type (e.g., predatory or pollination) of the interactions, provide information regarding the  
97 taxonomic level, identities, and characteristics (e.g., life stages) of the individuals involved in an interaction,  
98 present the mathematical formulation of probabilities, including clearly identified conditional variables (e.g.,  
99 spatial and temporal scales), and describe the methods and contexts (e.g., location, time, environmental  
100 conditions) in which interactions were estimated. Inadequately documented probabilistic interaction data should  
101 be used with caution when analyzing ecological webs. Our observations and advice can be applied to many  
102 types of ecological networks, from food webs to host-virus networks. Indeed, excluding networks of indirect  
103 interactions such as competition and facilitation networks (Kéfi *et al.* (2015), Kéfi *et al.* (2016)), most  
104 ecological webs describe probabilities of direct interactions, which are conceptually and mathematically  
105 analogous regardless of their biological type (e.g., trophic and parasitic interactions).

## 106 **Probabilistic representations of interactions**

107 One of the first aspects to take into consideration when estimating or interpreting probabilities of interactions is  
108 knowing if they describe potential or realized interactions. A potential (regional) interaction is defined as the  
109 biological capacity of two taxa to interact (i.e., the probability that they *can* theoretically interact) whereas a  
110 realized (local) interaction is the materialization or observation of this interaction in a well-defined space and  
111 time (i.e., the probability that they interact locally). Here, we use the terms *metaweb* (Dunne (2006)) to  
112 designate regional webs of potential interactions and *local webs* (Poisot *et al.* (2012)) for those of realized  
113 interactions. Metawebs are the network analogs of the species pool, where local webs originate from a subset of  
114 both species (nodes) and interactions (edges) of the regional metaweb (Saravia *et al.* (2022)). Without clear  
115 documentation, it can be challenging to know if published probabilistic webs describe local or regional  
116 interactions (tbl. 1 provides examples of studies employing both types of probabilistic networks), or if so-called  
117 probabilities are in reality *interaction scores* (i.e., non-probabilistic quantitative interactions). When  
118 probabilistic regional interactions are used and interpreted incorrectly as local interactions (and conversely),

119 this may generate misleading findings during data analysis. We believe that a better understanding of the  
120 differences, similarities, and relationships between these two probabilistic representations of ecological webs  
121 would alleviate interpretation errors and facilitate a more adequate utilization of interaction data.

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

123 Local and metawebs, like any type of network, are made of nodes and edges that can be represented at different  
124 levels of organization and precision. The basic units of ecological networks are individuals that interact with  
125 each other (e.g., by predation in food webs, Elton (2001)), forming individual-based networks (Melián *et al.*  
126 (2011)). The aggregation of these individuals into more or less homogeneous groups (e.g., populations, species,  
127 families, feeding guilds) allows us to represent nodes at broader taxonomic scales, which affects our  
128 interpretation of the properties of these systems (Guimarães (2020), Hemprich-Bennett *et al.* (2021)).

129 Edges linking nodes can describe a variety of interaction measures. Ecologists have traditionally represented  
130 interactions as binary objects that were considered realized after observing at least one individual from group  $i$   
131 interact with at least another individual from group  $j$ . In a binary adjacency matrix  $B$ , the presence or absence  
132 of an interaction  $B_{i \rightarrow j}$  between two taxa can be viewed as the result of a Bernoulli process

133  $B_{i \rightarrow j} \sim \text{Bernoulli}(P(B_{i \rightarrow j}))$ , with  $P(B_{i \rightarrow j})$  being the probability of interaction that characterizes our limited  
134 knowledge of the system and its intrinsic spatiotemporal variability. In probabilistic networks,  $P(B_{i \rightarrow j})$  are edge  
135 values, and the only two possible outcomes are the presence ( $B_{i \rightarrow j} = 1$ ) or absence ( $B_{i \rightarrow j} = 0$ ) of an interaction  
136 between each pair of nodes. Depending on the type of probabilistic network (local or metaweb), the  
137 mathematical formulation and interpretation of stochastic parameters like  $P(B_{i \rightarrow j})$  can be linked to  
138 environmental and biological factors such as species relative abundance, traits, area, and time (tbl. 1), for  
139 example using logistic regression with a logit link function with continuous explanatory variables. Predicting  
140 the number of local webs in which the interaction occurs can be achieved by using a Binomial distribution,  
141 assuming a constant probability of interaction and independence between networks (trials). When considering  
142 uncertainties around the estimation of  $P(B_{i \rightarrow j})$ , a Beta distribution can also be used to encompass all possible  
143 probability values. In that case, a Beta-Binomial distribution can be used to predict the number of networks in  
144 which the interaction occurs. Observing an interaction between two taxa at a given location and time provides  
145 important information that can be used to update previous estimates of  $P(B_{i \rightarrow j})$ , informing us on the biological  
146 capacity of both taxa to interact and the environmental conditions that enabled them to interact locally.

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

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

177 addressing the challenges in interpreting and using interaction probabilities in Bernoulli distributions, in both  
178 probabilistic local and metawebs.

179 **Local webs: communities interacting in space and time**

180 Probabilistic local webs describe how likely taxa are to interact at a given location and time period (i.e.,  
181 interactions are contingent upon the environmental and biological conditions of the community). In local webs,  
182 edges commonly represent our degree of belief that two taxa interact in nature, but can also represent the  
183 probability of *observing* this interaction (Catchen *et al.* (2023)). For example, Gravel *et al.* (2019) used a  
184 dataset of binary local European food webs of willow-galling sawflies and their natural enemies, all referenced  
185 in space and time and consisting of similar species, to infer the probabilities of locally observing interactions  
186 between co-occurring species. This was achieved by situating local webs within the context of environmental  
187 variables, i.e. by including temperature and precipitation as conditional variables in some of their models.

188 We define space as the collection of geographic coordinates  $(x, y, z)$ , with  $(x, y)$  representing longitude and  
189 latitude coordinates, and  $z$  denoting either altitudes or depths. These point coordinates delineate the spatial  
190 boundaries of the system, which may be portrayed as a polyhedron. Ecological interactions may exhibit  
191 variations along altitudinal gradients, as evidenced by changes in hummingbird-plant interactions (Weinstein &  
192 Graham (2017b) and Weinstein & Graham (2017a)) and mosquito biting rates (e.g., Kulkarni *et al.* (2006)) at  
193 different elevations. In contrast, time is defined as the specific time period within which interactions were either  
194 observed or predicted. With these definitions in mind, space and time can be conceptualized as distinct patches  
195 or time segments. Treating them as discrete dimensions aligns with the common sampling methods of  
196 ecological webs and provides actual probabilities of interactions across space and time. This differs from the  
197 approach of treating them as continuous variables, where edge values represent probability densities (i.e.,  
198 relative likelihoods of interactions occurring at infinitesimal locations and instants in time) rather than discrete  
199 probability values (which can be obtained by integrating probability densities). By employing discrete locations  
200 and time periods, we can quantify an area (or volume)  $A$  and a duration  $t$ , which can be readily used in  
201 spatiotemporal analyses of ecological networks. For example, when studying network-area relationships (NAR,  
202 Galiana *et al.* (2018)), we anticipate that local probabilities of interactions scale positively with area and  
203 duration because taxa have more opportunities to interact as these dimensions expand.

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

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

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

Other important factors that can impact interaction probabilities at the local scale are taxa relative abundance (Canard *et al.* (2012)) and traits (Poisot *et al.* (2015)), as well as environmental factors such as temperature (Angilletta *et al.* (2004)), precipitation (Woodward *et al.* (2012)), habitat structure (Klecka & Boukal (2014)), and presence of other interacting taxa in the network (Pilosof *et al.* (2017), Kéfi *et al.* (2012)), as described above. Here, we use the variable  $\Omega_{x,y,z,t}$  (hereafter simply  $\Omega$ ) to describe the biological and ecological context in which interaction probabilities were estimated. For example, if a research team conducts a mesocosm experiment to estimate interaction probabilities between predators and prey with and without shelters (a place that offers refuge and protection for prey, shielding them from predators),  $\Omega$  would represent the presence or absence of these shelters. Like co-occurrence,  $\Omega$  can also be modeled probabilistically when the stochasticity or uncertainty of environmental and biological factors is considered. In sum,  $\Omega$  represents all ecological and biological variables that were taken into consideration when measuring interaction probabilities and is, therefore, a subset of all factors impacting ecological interactions. It is a vector of random variables whose values determine the overall uncertainty of an interaction.

The probability that two taxa  $i$  and  $j$  interact in a local web  $L$  can thus be conditional on the area (or volume)  $A$ , the time interval  $t$ , their co-occurrence  $C_{i,j}$  and chosen environmental and biological conditions  $\Omega$ . Although these variables are associated with distinct questions or mechanisms related to ecological interactions, they may covary with each other, such as the possible dependence of  $C_{i,j}$  and  $\Omega$  on spatial and temporal scales. When estimating interaction probabilities using e.g. a generalized linear model with multiple explanatory variables that might not be independent, it may become important to address collinearity. Using variable selection

232 techniques, for instance, may be necessary before fitting the model to mitigate this issue. The probability of  
 233 local interaction is described by the following expression when all these conditional variables are included:

$$P(L_{i \rightarrow j} | A, t, C_{i,j}, \Omega) \quad (2)$$

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

**Table 1: Mathematical expression 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 web  $L$  of realized interactions. Each expression 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 webs. 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.

Expression	Type	Outcome	Reference
$P(M_{i \rightarrow j})$	regional	biological feasibility of the interaction	Strydom <i>et al.</i> (2022)
$P(L_{i \rightarrow j})$	local	realization of the interaction	Fortuna & Bascompte (2006) (null model)
$P(L_{i \rightarrow j} A)$	local	realization of the interaction in a given area or volume	Galiana <i>et al.</i> (2018) *
$P(L_{i \rightarrow j} t)$	local	realization of the interaction during a given time period	Weinstein & Graham (2017a)

Expression	Type	Outcome	Reference
$P(L_{i \rightarrow j}   C_{i,j})$	local	realization of the interaction given that the taxa co-occur	Gravel <i>et al.</i> (2019)
$P(L_{i \rightarrow j}   \Omega)$	local	realization of the interaction given environmental conditions	Gravel <i>et al.</i> (2019) (temperature and precipitation)
$P(L_{i \rightarrow j}   M_{i \rightarrow j})$	local	realization of the interaction given that the taxa can biologically interact	this study

## 246 Metawebs: regional catalogs of interactions

247 Metawebs are networks of potential interactions that have been designed for broad spatial, temporal, and  
 248 taxonomic scales (e.g., species food webs at the continental scale). They represent the probability that taxa can  
 249 biologically interact regardless of their co-occurrence and local environmental conditions. Regional interactions  
 250 are by definition context-independent, i.e. they are not measured at a specific location and time. In contrast with  
 251 probabilistic local webs, which represent the stochasticity of interactions occurring in nature, probabilistic  
 252 metawebs measure our degree of belief in the capacity of two taxa to interact (i.e., the probability that their  
 253 traits could support an interaction in the right conditions). Consequently, although *neutrally* forbidden  
 254 interactions (i.e., improbable interactions between rare species, Canard *et al.* (2012)) tend to have low  
 255 probability values in local webs, they may exhibit a higher probability in the metaweb. Potential interactions  
 256 describe the probability that there exists at least one combination of phenotypes of taxa  $i$  and  $j$  that can interact  
 257 with each other if they were to encounter. For example, a piscivorous species may have the capacity to interact  
 258 with a potential prey species if they share at least one matching combination of body sizes, such as an adult  
 259 predator being able to consume juvenile prey. To reduce any biases in our calculation of this probability, it is  
 260 crucial to ensure that the set of traits sampled or considered accurately reflects the overall trait distribution in  
 261 both taxa. This enhances our confidence in concluding that the taxa cannot interact when no positive  
 262 observation has been made and increases our capacity to detect an interaction when it is biologically feasible. In  
 263 addition to pairwise interactions, the extent of sampling effort also contributes to shaping our assessment of  
 264 metaweb properties, as sampling a larger number of local webs allows us to capture more regional interactions  
 265 (McLeod *et al.* (2021)). A probability of potential interaction in a metaweb  $M$  can be expressed as

$$P(M_{i \rightarrow j}), \quad (3)$$

which, in contrast with local webs, is never conditional on any spatial, temporal, co-occurrence or environmental variables (tbl. 1).

Starting from a selected set of taxa, which are usually distributed within a broad region of interest, metawebs can be built using different data sources, including literature review (e.g., Maiorano2020Tetraeu), aggregated local web data (e.g., Gravel *et al.* (2019), Saravia *et al.* (2022)), and predictive models (e.g., Strydom *et al.* (2022), Shaw *et al.* (2024)). Every pair of taxa that have confidently been observed to interact at least once can be given a probability of 1 (i.e.,  $P(M_{i \rightarrow j}) = 1$ ) since we know that they *can* interact. This is not the case in local probabilistic webs, in which interaction events usually remain stochastic (i.e.,  $P(L_{i \rightarrow j}) < 1$ ) after empirically observing interactions because of their intrinsic spatiotemporal variability. Similarly, although rare interactions typically have low probabilities in local webs, they can have high probabilities in metawebs when the traits of both taxa are congruent. On the other hand, interactions that were never observed can have low probability values in both local and metawebs, going as low as 0 for forbidden links (i.e., interactions forbidden by biological traits). However, many observations of interactions are in reality false positives because of observation errors due to taxonomic misidentifications and ecological misinterpretations (e.g., due to cryptic species and interactions, Pringle & Hutchinson (2020)). Likewise, forbidden interactions can be false negatives in metawebs, e.g. if they have been assessed for specific phenotypes, locations or time. Employing Bayesian models, whether they are mechanistic or phenomenological, could improve the accuracy of our estimation of interaction probabilities within both local and regional webs (e.g., Bartomeus *et al.* (2016), Cirtwill *et al.* (2019)). This improvement is achieved in metawebs by leveraging prior information regarding the feasibility of interactions along with the empirical data on observed interactions.

## Properties of probabilistic webs

Probabilistic local and metawebs differ in their type of interactions (i.e., realized or potential) and in the conditional variables upon which interaction values depend. These differences are significant as they influence the characteristics of probabilistic networks. Neglecting to consider them may result in misleading results and interpretation errors when analyzing the properties of probabilistic webs, which could be particularly

291 problematic when addressing crucial ecological questions about networks. Here we compare the characteristics  
292 of local and metawebs through the presentation of four applications of probabilistic interactions: (1) describing  
293 their spatial and temporal scaling, (2) describing their taxonomic scaling, (3) sampling for binary webs, and (4)  
294 reconstructing probabilistic local webs from metawebs. All code and data to reproduce these analyses are  
295 available at the Open Science Framework (TBD).

## 296 Host-parasite network data

297 We use the collection of tripartite host-parasite networks sampled across Europe, created by Kopelke *et al.*  
298 (2017), in the following case studies. This dataset contains well-resolved binary local interactions between  
299 willows (52 species), willow-galling sawflies (96 species), and their parasitoids (126 species). Given its  
300 replicated webs spanning large spatiotemporal scales, this dataset is well-suited for analyzing a variety of  
301 ecological hypotheses and processes. Out of a total of 374 local webs, we retained those containing at least 5  
302 species, resulting in a set of 233 georeferenced local webs (networks sampled within areas of 0.1 to 0.3 km<sup>2</sup>  
303 during June and/or July spanning 29 years). We built a binary metaweb by aggregating all local interactions,  
304 which gave us a regional web composed of 274 species and 1080 interactions. In the first two panels of fig. 1,  
305 we show how the dissimilarity of interactions between common species ( $\beta_{OS}$ ) and the dissimilarity in species  
306 composition ( $\beta_S$ ) between the metaweb and aggregated local webs (Poisot *et al.* (2012)) vary with the number  
307 of sampled local webs. This shows that networks of local interactions are highly dissimilar from the metaweb,  
308 both in terms of species and interactions, especially when only a limited number of sites has been sampled.  
309 Both dissimilarity indices were calculated based on the number of items shared by the two webs ( $c_{LM}$ ) and the  
310 number of items unique to the metaweb ( $u_M$ ) and to the aggregated local web ( $u_L$ ). The  $\beta_S$  dissimilarity index  
311 uses species (nodes) as items being compared, while the  $\beta_{OS}$  index assesses dissimilarity based on interactions  
312 between shared species (Poisot *et al.* (2012)). Both indices were calculated following the  $\beta_W$  index of  
313 Whittaker (1960):

$$\beta_W = \frac{c_{LM} + u_L + u_M}{(2c_{LM} + u_L + u_M)/2} - 1. \quad (4)$$

314 [Figure 1 about here.]

315 We converted these binary networks into probabilistic ones using models based on simple assumptions. Our  
316 models are not designed to estimate the exact values of probabilistic interactions. Instead, their purpose is to

317 create plausible networks that serve as illustrative examples to highlight distinctions between probabilistic local  
318 and metawebs. We created two probabilistic metawebs by employing constant false positive and false negative  
319 rates for all regional interactions. In the first metaweb, we set both false positive and false negative rates to zero  
320 to prevent artificially inflating the total number of links, enabling a more accurate comparison with binary webs.  
321 This gave us a probability of regional interaction of 1 when at least one interaction has been observed and of 0  
322 in the absence of any observed interaction between a given pair of species. In the second metaweb, we  
323 introduced a 5% false positive rate to account for spurious interactions and a 10% false negative rate to address  
324 the elevated occurrence of missing interactions in ecological networks (Catchen *et al.* (2023)). We believe these  
325 rates represent reasonable estimates of missing and spurious potential interactions, but confirming their  
326 accuracy is challenging due to the unavailability of data on the actual feasibility of interaction. Observed  
327 interactions were thus given a probability of regional interaction of 95%, whereas unobserved ones were  
328 assigned a probability of 10%.

329 To build probabilistic local webs, we first recognize that local interactions must initially be biologically feasible  
330 before occurring at a specific time and space. A local probability of interaction  $P(L_{i \rightarrow j})$  can be expressed as the  
331 product of the probability of local interaction given that the two taxa can potentially interact  $P(L_{i \rightarrow j}|M_{i \rightarrow j} = 1)$ ,  
332 which we denote as  $P(L|M)$  for the sake of simplicity, with their probability of regional interaction  $P(M_{i \rightarrow j})$ :

$$P(L_{i \rightarrow j}) = P(L_{i \rightarrow j}|M_{i \rightarrow j} = 1) \times P(M_{i \rightarrow j} = 1). \quad (5)$$

333 We built the probabilistic local webs from the binary ones by using the probabilistic metawebs and a constant  
334 value of  $P(L|M)$  across interactions. Low values of  $P(L|M)$  indicate that feasible interactions rarely occur  
335 locally, intermediate values around 50% suggest considerable spatiotemporal variability, while high values  
336 indicate that regional interactions are nearly always realized locally. Following eq. 5, the local probability of  
337 interaction between a given pair of taxa consistently remained equal to or below their probability of regional  
338 interaction.

339 In the last two panels of fig. 1, we show how the aggregated number of links and connectance (i.e., the  
340 proportion of all of the non-forbidden links that are realized) scale with the number of sampled local  
341 probabilistic webs, according to different values of  $P(L|M)$ . When aggregating local probabilistic webs, the  
342 constancy of the probability of regional interaction across the entire study area means that any rise in the  
343 probability of local interaction is solely attributable to an increase in  $P(L|M)$ . The probability  $P(L_{1,2}|M)$  of

344 local interaction among potentially interacting species in an aggregated web  $L_{1,2}$  is obtained by:

$$P(L_{1,2}|M) = 1 - (1 - P(L_1|M)) \times (1 - P(L_2|M)), \quad (6)$$

345 where  $P(L_1|M)$  and  $P(L_2|M)$  are the probabilities of local interaction among two potentially interacting species  
346 in the subnetworks  $L_1$  and  $L_2$ , respectively. This equation provides the probability that the interaction is realized  
347 in either (1) exclusively the local web  $L_1$ , (2) exclusively the local web  $L_2$  or (3) both local webs, given that the  
348 two taxa have the biological capacity to interact.

349 By comparing the scaling relationships observed in binary and probabilistic webs, fig. 1 illustrates that high  
350 values of  $P(L|M)$  lead to systematic overestimations in the number of links and connectance, especially when  
351  $P(L|M) = 1$  (corresponding to the scenario where local probabilities of interactions are equivalent to the  
352 probabilities of regional interactions). However, these biases tend to diminish as the number of sampled webs  
353 increases.

### 354 Spatial and temporal scaling of interactions

355 The investigation of network-area relationships and interaction accumulation curves is an important area of  
356 research in network ecology. First, network-area relationships elucidate the scaling of network properties (such  
357 as modularity and trophic chain lengths) with spatial scale (e.g., Wood *et al.* (2015), Galiana *et al.* (2018)). The  
358 variations in network structure across spatial scales may stem from the scaling of species richness (species-area  
359 relationships, SARs) and the number of interactions (Brose *et al.* (2004)) with the sampled area. Additionally,  
360 ecological processes unfolding at distinct spatial scales, such as the spatial variability in local community  
361 composition resulting from different sequences of extinction and colonization events, can also contribute to this  
362 variation (Galiana *et al.* (2018)). Next, interaction accumulation curves describe the scaling of the number of  
363 observed interactions with sampling effort (Jordano (2016)). Sampling effort, which may correspond to the  
364 duration of the sampling period used to construct the network, can impact connectance (Bersier *et al.* (1999))  
365 and various measures of network structure (Banašek-Richter *et al.* (2004), McLeod *et al.* (2021)). Apart from  
366 sampling effort, the temporal scaling of interactions also elucidates how network structure changes with the  
367 temporal resolution of the network (Poisot *et al.* (2012)), acknowledging that distinct interactions take place  
368 over time, ranging from short-term fluctuations of interactions to long-term trends. As probabilistic local webs  
369 may explicitly account for the spatiotemporal variability of interactions, they offer a distinct approach to

370 investigating the scaling of network structure with space and time, in contrast to binary and quantitative webs,  
371 by making the stochasticity of interactions the focal point of the modeling process.

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

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

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

401 Predicting local webs across time and space is a pivotal goal of network ecology, especially given the scarcity of  
 402 interaction data (Strydom *et al.* (2021)). Ecologists may resort to predictive models (e.g., generative Bayesian  
 403 and machine learning models) to reconstruct networks at fine spatial and temporal scales with limited  
 404 interaction data. For instance, real-time biomonitoring data coupled with appropriate numerical models (Bohan  
 405 *et al.* (2017)) can be employed to reconstruct local ecological webs, opening avenues for in-depth studies on  
 406 local ecosystem functioning and dynamics. The probabilistic representation of interactions acknowledges the  
 407 inherent uncertainty in these models, typically expressed through probability distributions. We introduce and  
 408 develop a simple generative mechanistic model for probabilistic local interactions that takes into consideration  
 409 their spatiotemporal variability (i.e. a spatiotemporally explicit model of local interactions). It is essential to  
 410 note that our model is not designed for regional interactions, which are scale-independent. Rather, it could  
 411 prove valuable for predicting local interactions across time and space by generating new interaction data  
 412 following parameter inference.

413 As indicated by eq. 1, the probability that two taxa  $i$  and  $j$  interact at a specific location  $(x, y, z)$  may be  
 414 determined by the product of their probability of interaction given co-occurrence and their probability of  
 415 co-occurrence. First, their co-occurrence probability can be calculated using their individual occurrence  
 416 probabilities  $P(X_i)$  and  $P(X_j)$ , along with the strength of association  $\gamma$  between their occurrences and  
 417 co-occurrence (Cazelles *et al.* (2016)):

$$P(C_{i,j}) = P(X_i, X_j) = P(X_i)P(X_j)\gamma. \quad (7)$$

418 When  $\gamma > 1$ , it signifies a positive association in the geographic distributions of both taxa, indicating that the  
 419 presence of one taxon enhances the probability of occurrence of the other. Attractions may be the result of  
 420 positive interactions (e.g., mutualism) or positive/negative interactions (e.g., antagonism between predators and  
 421 prey, Cazelles *et al.* (2016)). In empirical webs,  $\gamma > 1$  holds true for the majority of species pairs (Catchen *et*  
 422 *al.* (2023)). In contrast, repulsions ( $\gamma < 1$ ) may be caused by strong interspecific competition (Cazelles *et al.*  
 423 (2016)). We model the co-occurrence of both taxa as the outcome of a Bernoulli trial

$$C_{i,j} \sim \text{Bernoulli}(P(X_i, X_j)). \quad (8)$$

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

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

429 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  
 430 instance, if the duration  $t_0$  is measured in months,  $\lambda$  denote the expected number of interactions per month.  
 431 The occurrence of an interaction between  $i$  and  $j$  can be modeled as a Bernoulli trial with a probability of  
 432  $P(L_{i \rightarrow j})$ . Consequently, a Bayesian inference model can be built based on the preceding equations to estimate  
 433 the value of the  $\lambda$  and  $\gamma$  parameters and generate novel interaction data:

$$L_{i \rightarrow j} \sim \text{Bernoulli}(P(L_{i \rightarrow j})) \quad (10)$$

$$P(L_{i \rightarrow j}) = P(X_i)P(X_j)\gamma(1 - e^{-\lambda t_0}) \quad (11)$$

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

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

434 In fig. 3, we show the variation in the probability of interaction under different parameter values. In the right  
 435 panel, we notice that, irrespective of the interaction rate  $\lambda$ , the probability of interaction converges toward an  
 436 asymptote determined by the co-occurrence  $P(C_{i,j})$  (eq. 7). This model can be customized in different ways,

437 such as by linking  $\lambda$  with specific environmental variables or explicitly incorporating observation errors (i.e.,  
438 the probabilities of false negatives and false positives).

439 [Figure 3 about here.]

440 **Taxonomic scaling of interactions**

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

449 There are no inherent differences in the taxonomic scaling between local and metawebs, as only the nodes are  
450 defined taxonomically. In other words, the probability values of edges in both local (eq. 2) and metawebs (eq. 3)  
451 are not conditional on any taxonomic scale. The phylogenetic scale is tied to the definition of the event itself  
452 (i.e., the interaction between two taxa), not to the conditional variables. In both types of webs, transitioning to a  
453 broader level of organization (e.g., from a species-level web  $S$  to a genus-level web  $G$ ) can be accomplished  
454 directly by using probabilities from finer scales. For example, in a network with  $n_1$  species from genus  $g_1$  and  
455  $n_2$  species from genus  $g_2$ , one can compute the probability that at least one species from genus  $g_1$  interacts with  
456 at least one species from genus  $g_2$  as follows:

$$P(G_{g_1 \rightarrow g_2}) = 1 - \prod_{i=1}^{n_1} \prod_{j=1}^{n_2} (1 - P(S_{g_{1i} \rightarrow g_{2j}})), \quad (14)$$

457 where  $g_{1i}$  and  $g_{2j}$  are the species of the corresponding genus. If it is known that at least two of these species  
458 interact (i.e.,  $P(S_{g_{1i} \rightarrow g_{2j}}) = 1$  for at least one pair of  $(g_{1i}, g_{2j})$ ), it implies a probability of genus interaction  
459 equal to 1. Canard *et al.* (2012) built a species-based network following a similar approach, by using simulated  
460 interactions between individuals derived from a neutral model (i.e., a model that assumed ecological  
461 equivalence among individuals). In contrast, a more sophisticated approach is necessary when transitioning

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

466 Ideally, our biological interpretation of probabilistic interactions should remain consistent across a network  
467 even if it incorporates heterogeneous levels of organization, such as a network whose nodes represent both  
468 species and trophic species (groups of species sharing similar predators and prey). This is common in  
469 ecological webs where taxonomic resolution is typically low (Hemprich-Bennett *et al.* (2021), Vázquez§ *et al.*  
470 (2022)). Interaction probabilities at broader taxonomic scales should be based on probabilities of interactions  
471 between individuals, for both local and metawebs. For instance, in local individual-based food webs, the  
472 probability that two individuals interact reflects our degree of belief that one individual will consume the other.  
473 Similarly, in local species-based food webs, the probability that two species interact represents our degree of  
474 belief that at least one individual from the predator species consumes at least another individual from the prey  
475 species. Furthermore, in local clade-based food webs, the probability that two clades interact represents our  
476 degree of belief that at least two species from these clades interact with each other or, equivalently, that at least  
477 two individuals from these clades interact with each other. Fundamentally, the taxonomic scaling of interactions  
478 involves aggregating interactions between individuals into larger groups, which may exhibit varying degrees of  
479 homogeneity based on the organisms and the study system. In that regard, taxonomic scaling is analogous to the  
480 spatial and temporal scaling of interactions, as they all represent different ways to aggregate individuals into  
481 broader groups (either spatially, temporally, or taxonomically).

## 482 Sampling for binary webs

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

491 compared against the simulated distribution (e.g., Bascompte *et al.* (2003)). Additionally, the random  
492 generation of binary networks, from a probabilistic web that accounts for the spatiotemporal variability of  
493 interactions, may effectively capture network structure across space and time. This facilitates the investigation  
494 of ecological hypotheses about interactions at broad spatial and temporal scales.

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

504 In fig. 4, we compare the average connectance of binary networks resulting from these two sampling techniques,  
505 where regional and local interactions are drawn from our host-parasite probabilistic networks, generating a  
506 number of binary web realizations for each site in the dataset. These two sampling approaches yield different  
507 outcomes, particularly for lower values of  $P(L|M)$ , which denote instances where regional interactions do not  
508 consistently manifest locally. Small discrepancies are also apparent between these techniques when we equate  
509 the probability of local interaction to the probability of regional interaction (i.e., when using  $P(L|M) = 1.0$  in  
510 eq. 5), especially when the number of binary network samples for each location is low. As anticipated, we  
511 observe that sampling binary interactions from the metaweb tends to overestimate connectance on average  
512 compared to sampling them from local webs. Furthermore, we observe an increase in the variability of  
513 connectance when employing a single sample, representing what we consider as a more tangible process  
514 leading to the realization of local and regional interactions in nature.

515 [Figure 4 about here.]

516 The choice of a sampling approach can influence the selection of grid cell size when delineating local  
517 communities within a broader region of interest. In the first approach, pairwise interactions remain constant  
518 irrespective of cell size since they are sampled only once from the metaweb. However, in the second approach,  
519 local interaction probabilities are contingent on the network area. For instance, consider the local webs  $L_1$  and

520  $L_2$  with an area of  $\frac{1}{2}A_0$ , both nested within  $A_0$  but disjoint from each other, forming  $L_0$ . If we treat  $L_1$  and  $L_2$  as  
521 independent, the probability of interaction between taxa  $i$  and  $j$  in  $L_0$  is given by:

$$P(L_{0,i \rightarrow j}) = 1 - (1 - P(L_{1,i \rightarrow j}) \times P(L_{2,i \rightarrow j})). \quad (15)$$

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

## 526 Prediction of local webs from metawebs

527 Metawebs serve as a valuable source of ecological information for predicting local webs across time and space.  
528 Binary local webs can be reconstructed by selecting a subset of taxa and interactions from the metaweb (Dunne  
529 (2006)). This implies that metawebs consistently contain more interactions than their corresponding local webs.  
530 Determining the list of taxa to select can be achieved empirically or through methods like range maps or species  
531 distribution models. As the species composition of a community is arguably less difficult to sample or predict  
532 than its interactions, the primary challenge lies in deciding which interactions to select from the metaweb, a task  
533 that may necessitate advanced statistical models and ecological expertise. Inferring the structure of local webs  
534 from the metaweb before predicting specific local pairwise interactions could hold promise (Strydom *et al.*  
535 (2021)), considering that the structure of local webs is constrained by the metaweb (Saravia *et al.* (2022)).

536 Building probabilistic local webs from a probabilistic metaweb involves a reduction in the value of pairwise  
537 interaction probabilities. This decrease is attributed to the prerequisite that two taxa must initially possess the  
538 capacity to interact before engaging in local interactions (eq. 5). Therefore, inferring local webs from their  
539 metaweb while maintaining identical interaction probability values would introduce systematic biases into the  
540 predictions. In such cases, these networks would essentially represent smaller-scale metawebs of potential  
541 interactions, possibly leading to misinterpretations by being perceived as local interactions. As proposed by  
542 McLeod *et al.* (2021), although metawebs do not capture the spatiotemporal variability of interactions, they  
543 establish an upper limit for local interactions. In other words, the probability of two taxa interacting at a specific  
544 location and time is consistently lower or equal to the probability of their regional interaction:

$$P(L_{i \rightarrow j}|A, t, C, \Omega) \leq P(M_{i \rightarrow j}). \quad (16)$$

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

$$\int_{\Omega} \int_A \int_t P(L_{i \rightarrow j}|A, t, \Omega) dt dA d\Omega \leq P(M_{i \rightarrow j}). \quad (17)$$

549 Estimating more precisely the probability  $P(L|M)$  that two taxa interact locally if they can potentially interact  
 550 allows for improved predictions of local webs from a probabilistic metaweb. This task is challenging due to the  
 551 variability of this probability across space and time, as well as its variability across pairwise interactions within  
 552 a network. Using simple models of  $P(L|M)$ , as demonstrated in our case studies, represents an initial step  
 553 toward the overarching objective of reconstructing local webs from metawebs.

## 554 Conclusion

555 In this contribution, we underline the importance of network metadata for adequately interpreting and  
 556 manipulating probabilistic interaction data. The mathematical representation of probabilities and their statistical  
 557 properties depend on the type of interactions (local or regional) and the conditions under which these  
 558 interactions were evaluated. We showed that probabilistic local and metawebs differ in their relationship to  
 559 spatial and temporal scales, with regional interactions remaining consistent across scales. In contrast with  
 560 metawebs, local interactions are measured in a specific context (e.g., in a given area, time, and biological and  
 561 environmental conditions) and depend on taxa co-occurrence. These important conceptual differences bring to  
 562 light the need to use probabilistic data with caution, for instance when generating binary network realizations  
 563 across space and predicting local webs from metawebs. Clear metadata describing the type of interaction and  
 564 the variables used in their estimation are required to ensure adequate data manipulation. Better data practices  
 565 and foundations for probabilistic thinking in network ecology could enable more reliable assessments of the  
 566 spatiotemporal variability and uncertainty of biotic interactions.  
 567 It is essential to enhance our comprehension of both regional and local interactions, especially considering the

568 current scarcity of interaction data. However, while sampling biological communities does decrease the  
569 uncertainty of interactions by accumulating evidence for their feasibility and local realization, there is a limit to  
570 how much we can diminish their inherent randomness. In metawebs, probabilities reflect our limited knowledge  
571 of interactions (i.e., our degree of belief that interactions are feasible), which is expected to improve with a  
572 larger volume of data. We should anticipate that regional interactions will become more definitive (with  
573 probabilities approaching 0 or 1) as we investigate various conditions, including different combinations of  
574 species traits. Conversely, in the case of local webs, which can be seen as random instances of metawebs,  
575 randomness cannot be as reduced. In local interactions, probabilities can be divided into a component  
576 representing uncertainty and another representing spatiotemporal variability. Owing to environmental  
577 heterogeneity, there will invariably be instances in which an interaction occurs and others in which it does not,  
578 across different times and locations, irrespective of the extent to which we can improve our knowledge of its  
579 biological feasibility and the local conditions that facilitate its occurrence. When local webs depict probabilities  
580 of observing interactions rather than just their actual occurrence, we must also consider the observation  
581 variability as an additional source of randomness. Every ecological process is stochastic but there is also a  
582 possibility that a phenomenon goes undetected. Quantifying and partitioning this stochasticity will enable us to  
583 make more accurate predictions about ecological interactions at various spatial and temporal scales. This will  
584 prove to be of vital importance as our time to understand nature runs out, especially at the places where the  
585 impacts of climate change and habitat loss hit harder.

## 586 Acknowledgment

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

594 **References**

- 595 Abrego, N., Roslin, T., Huotari, T., Ji, Y., Schmidt, N.M., Wang, J., *et al.* (2021). Accounting for species  
596 interactions is necessary for predicting how arctic arthropod communities respond to climate change.  
597 *Ecography*, 44, 885–896.
- 598 Andrade-Pacheco, R., Rerolle, F., Lemoine, J., Hernandez, L., Meité, A., Juziwelo, L., *et al.* (2020). Finding  
599 hotspots: Development of an adaptive spatial sampling approach. *Scientific Reports*, 10, 10939.
- 600 Angilletta, M.J., Jr., Steury, T.D. & Sears, M.W. (2004). Temperature, Growth Rate, and Body Size in  
601 Ectotherms: Fitting Pieces of a Life-History Puzzle<sup>1</sup>. *Integrative and Comparative Biology*, 44, 498–509.
- 602 Araujo, M.B., Rozenfeld, A., Rahbek, C. & Marquet, P.A. (2011). Using species co-occurrence networks to  
603 assess the impacts of climate change. *Ecography*, 34, 897–908.
- 604 Banašek-Richter, C., Cattin, M.-F. & Bersier, L.-F. (2004). Sampling effects and the robustness of quantitative  
605 and qualitative food-web descriptors. *Journal of Theoretical Biology*, 226, 23–32.
- 606 Bartomeus, I., Gravel, D., Tylianakis, J.M., Aizen, M.A., Dickie, I.A. & Bernard-Verdier, M. (2016). A  
607 common framework for identifying linkage rules across different types of interactions. *Functional Ecology*,  
608 30, 1894–1903.
- 609 Bascompte, J., Jordano, P., Melián, C.J. & Olesen, J.M. (2003). The nested assembly of plantanimal mutualistic  
610 networks. *Proceedings of the National Academy of Sciences*, 100, 9383–9387.
- 611 Berlow, E.L., Neutel, A.-M., Cohen, J.E., De Ruiter, P.C., Ebenman, B., Emmerson, M., *et al.* (2004).  
612 Interaction strengths in food webs: Issues and opportunities. *Journal of Animal Ecology*, 73, 585–598.
- 613 Bersier, L.-F., Dixon, P. & Sugihara, G. (1999). Scale-Invariant or Scale-Dependent Behavior of the Link  
614 Density Property in Food Webs: A Matter of Sampling Effort? *The American Naturalist*, 153, 676–682.
- 615 Blanchet, F.G., Cazelles, K. & Gravel, D. (2020). Co-occurrence is not evidence of ecological interactions.  
616 *Ecology Letters*, 23, 1050–1063.
- 617 Bohan, D.A., Vacher, C., Tamaddoni-Nezhad, A., Raybould, A., Dumbrell, A.J. & Woodward, G. (2017).  
618 Next-Generation Global Biomonitoring: Large-scale, Automated Reconstruction of Ecological Networks.  
619 *Trends in Ecology & Evolution*, 32, 477–487.

- 620 Bolnick, D.I., Amarasekare, P., Araújo, M.S., Bürger, R., Levine, J.M., Novak, M., *et al.* (2011). Why  
621 **intraspecific trait variation matters in community ecology.** *Trends in Ecology & Evolution*, 26, 183–192.
- 622 Borrett, S.R. & Scharler, U.M. (2019). **Walk partitions of flow in Ecological Network Analysis: Review and**  
623 **synthesis of methods and indicators.** *Ecological Indicators*, 106, 105451.
- 624 Boulangeat, I., Gravel, D. & Thuiller, W. (2012). **Accounting for dispersal and biotic interactions to disentangle**  
625 **the drivers of species distributions and their abundances.** *Ecology Letters*, 15, 584–593.
- 626 Brimacombe, C., Bodner, K., Michalska-Smith, M., Poisot, T. & Fortin, M.-J. (2023). **Shortcomings of reusing**  
627 **species interaction networks created by different sets of researchers.** *PLOS Biology*, 21, e3002068.
- 628 Broom, M. & Ruxton, G.D. (2005). **You can run or you can hide: Optimal strategies for cryptic prey against**  
629 **pursuit predators.** *Behavioral Ecology*, 16, 534–540.
- 630 Brose, U., Ostling, A., Harrison, K. & Martinez, N.D. (2004). **Unified spatial scaling of species and their**  
631 **trophic interactions.** *Nature*, 428, 167–171.
- 632 Canard, E., Mouquet, N., Marescot, L., Gaston, K.J., Gravel, D. & Mouillot, D. (2012). **Emergence of**  
633 **Structural Patterns in Neutral Trophic Networks.** *PLOS ONE*, 7, e38295.
- 634 Catchen, M.D., Poisot, T., Pollock, L.J. & Gonzalez, A. (2023). **The missing link: Discerning true from false**  
635 **negatives when sampling species interaction networks.**
- 636 Cazelles, K., Araujo, M.B., Mouquet, N. & Gravel, D. (2016). **A theory for species co-occurrence in interaction**  
637 **networks.** *Theoretical Ecology*, 9, 39–48.
- 638 Chagnon, P.-L. (2015). **Characterizing topology of ecological networks along gradients: The limits of metrics'**  
639 **standardization.** *Ecological Complexity*, 22, 36–39.
- 640 Choh, Y., Ignacio, M., Sabelis, M.W. & Janssen, A. (2012). **Predator-prey role reversals, juvenile experience**  
641 **and adult antipredator behaviour.** *Scientific Reports*, 2, 728.
- 642 Cirtwill, A.R., Eklöf, A., Roslin, T., Wootton, K. & Gravel, D. (2019). **A quantitative framework for**  
643 **investigating the reliability of empirical network construction.** *Methods in Ecology and Evolution*, 10,  
644 902–911.
- 645 Dallas, T., Park, A.W. & Drake, J.M. (2017). **Predicting cryptic links in host-parasite networks.** *PLOS*  
646 **Computational Biology**, 13, e1005557.

- 647 Dunne, J.A. (2006). The Network Structure of Food Webs. In: *Ecological networks: Linking structure and*  
648 *dynamics* (eds. Dunne, J.A. & Pascual, M.). Oxford University Press, pp. 27–86.
- 649 Elmasri, M., Farrell, M.J., Davies, T.J. & Stephens, D.A. (2020). [A hierarchical Bayesian model for predicting](#)  
650 [ecological interactions using scaled evolutionary relationships](#). *The Annals of Applied Statistics*, 14,  
651 221–240.
- 652 Elton, C.S. (2001). *Animal Ecology*. University of Chicago Press, Chicago, IL.
- 653 Emmerson, M.C. & Raffaelli, D. (2004). [Predator-prey body size, interaction strength and the stability of a real](#)  
654 [food web](#). *Journal of Animal Ecology*, 73, 399–409.
- 655 Fortuna, M.A. & Bascompte, J. (2006). [Habitat loss and the structure of plant-animal mutualistic networks](#).  
656 *Ecology Letters*, 9, 281–286.
- 657 Fu, X., Seo, E., Clarke, J. & Hutchinson, R.A. (2021). [Link Prediction Under Imperfect Detection: Collaborative Filtering for Ecological Networks](#). *IEEE Transactions on Knowledge and Data Engineering*,  
658 33, 3117–3128.
- 660 Galiana, N., Lurgi, M., Claramunt-López, B., Fortin, M.-J., Leroux, S., Cazelles, K., et al. (2018). [The spatial](#)  
661 [scaling of species interaction networks](#). *Nature Ecology & Evolution*, 2, 782–790.
- 662 Gilman, S.E., Urban, M.C., Tewksbury, J., Gilchrist, G.W. & Holt, R.D. (2010). [A framework for community](#)  
663 [interactions under climate change](#). *Trends in Ecology & Evolution*, 25, 325–331.
- 664 Godsoe, W., Murray, R. & Iritani, R. (2022). [Species interactions and diversity: A unified framework using Hill](#)  
665 [numbers](#). *Oikos*, n/a, e09282.
- 666 Golubski, A.J. & Abrams, P.A. (2011). [Modifying modifiers: What happens when interspecific interactions](#)  
667 [interact?](#) *Journal of Animal Ecology*, 80, 1097–1108.
- 668 Gomez, J.M., Verdu, M. & Perfectti, F. (2010). [Ecological interactions are evolutionarily conserved across the](#)  
669 [entire tree of life](#). *Nature*, 465, 918–U6.
- 670 Gonzalez, A. & Londoño, M.C. (2022). [Monitor biodiversity for action](#). *Science*, 378, 1147–1147.
- 671 Gonzalez-Varo, J.P. & Traveset, A. (2016). [The Labile Limits of Forbidden Interactions](#). *Trends in Ecology &*  
672 *Evolution*, 31, 700–710.

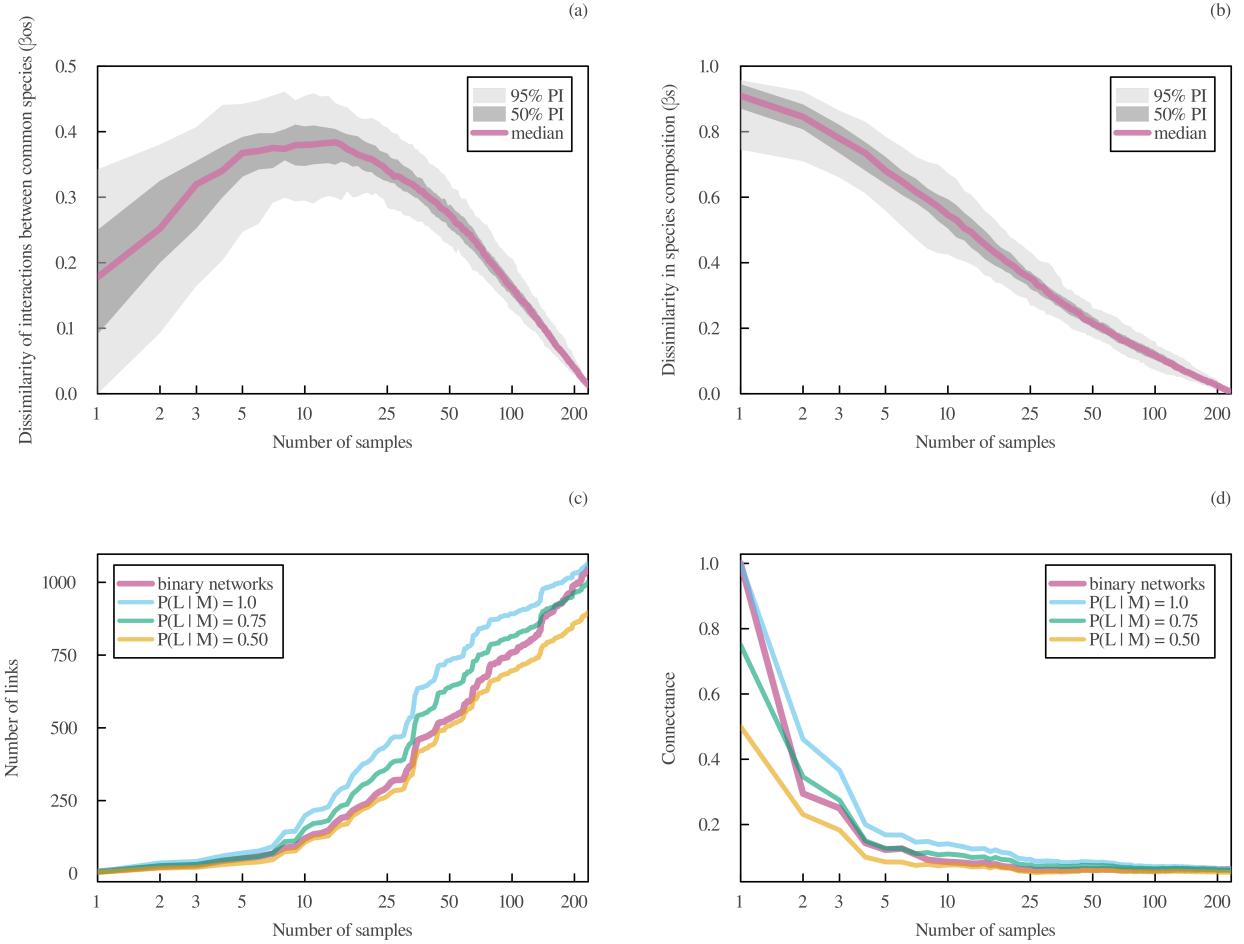
- 673 Gravel, D., Baiser, B., Dunne, J.A., Kopalke, J.-P., Martinez, N.D., Nyman, T., *et al.* (2019). Bringing Elton  
674 and Grinnell together: A quantitative framework to represent the biogeography of ecological interaction  
675 networks. *Ecography*, 42, 401–415.
- 676 Gravel, D., Poisot, T., Albouy, C., Velez, L. & Mouillot, D. (2013). Inferring food web structure from  
677 predator-prey body size relationships. *Methods in Ecology and Evolution*, 4, 1083–1090.
- 678 Guimarães, P.R. (2020). The Structure of Ecological Networks Across Levels of Organization. *Annual Review*  
679 of Ecology, Evolution, and Systematics, 51, 433–460.
- 680 Guimerà, R. & Sales-Pardo, M. (2009). Missing and spurious interactions and the reconstruction of complex  
681 networks. *Proceedings of the National Academy of Sciences*, 106, 22073–22078.
- 682 Hemprich-Bennett, D.R., Oliveira, H.F.M., Le Comber, S.C., Rossiter, S.J. & Clare, E.L. (2021). Assessing the  
683 impact of taxon resolution on network structure. *Ecology*, 102, e03256.
- 684 Herrera, C.M. (1989). Pollinator abundance, morphology, and flower visitation rate: Analysis of the “quantity”  
685 component in a plant-pollinator system. *Oecologia*, 80, 241–248.
- 686 Hortal, J., de Bello, F., Diniz-Filho, J.A.F., Lewinsohn, T.M., Lobo, J.M. & Ladle, R.J. (2015). Seven Shortfalls  
687 that Beset Large-Scale Knowledge of Biodiversity. *Annual Review of Ecology, Evolution, and Systematics*,  
688 46, 523–549.
- 689 Jordano, P. (1987). Patterns of Mutualistic Interactions in Pollination and Seed Dispersal: Connectance,  
690 Dependence Asymmetries, and Coevolution. *The American Naturalist*, 129, 657–677.
- 691 Jordano, P. (2016). Sampling networks of ecological interactions. *Functional Ecology*, 30, 1883–1893.
- 692 Jordano, P., Bascompte, J. & Olesen, J.M. (2003). Invariant properties in coevolutionary networks of  
693 plant-animal interactions. *Ecology Letters*, 6, 69–81.
- 694 Kéfi, S., Berlow, E.L., Wieters, E.A., Joppa, L.N., Wood, S.A., Brose, U., *et al.* (2015). Network structure  
695 beyond food webs: Mapping non-trophic and trophic interactions on Chilean rocky shores. *Ecology*, 96,  
696 291–303.
- 697 Kéfi, S., Berlow, E.L., Wieters, E.A., Navarrete, S.A., Petchey, O.L., Wood, S.A., *et al.* (2012). More than a  
698 meal... integrating non-feeding interactions into food webs. *Ecology Letters*, 15, 291–300.
- 699 Kéfi, S., Miele, V., Wieters, E.A., Navarrete, S.A. & Berlow, E.L. (2016). How Structured Is the Entangled  
700 Bank? The Surprisingly Simple Organization of Multiplex Ecological Networks Leads to Increased

- 701 Persistence and Resilience. *PLOS Biology*, 14, e1002527.
- 702 Klecka, J. & Boukal, D.S. (2014). The effect of habitat structure on prey mortality depends on predator and  
703 prey microhabitat use. *Oecologia*, 176, 183–191.
- 704 Kopelke, J.-P., Nyman, T., Cazelles, K., Gravel, D., Vissault, S. & Roslin, T. (2017). Food-web structure of  
705 willow-galling sawflies and their natural enemies across Europe. *Ecology*, 98, 1730–1730.
- 706 Kulkarni, M.A., Kweka, E., Nyale, E., Lyatuu, E., Mosha, F.W., Chandramohan, D., et al. (2006).  
707 Entomological Evaluation of Malaria Vectors at Different Altitudes in Hai District, Northeastern Tanzania.  
708 *Journal of Medical Entomology*, 43, 580–588.
- 709 McCann, K. (2007). Protecting biostructure. *Nature*, 446, 29–29.
- 710 McCann, K.S. (2011). Food webs (MPB-50). In: *Food Webs (MPB-50)*. Princeton University Press.
- 711 McLeod, A., Leroux, S.J., Gravel, D., Chu, C., Cirtwill, A.R., Fortin, M.-J., et al. (2021). Sampling and  
712 asymptotic network properties of spatial multi-trophic networks. *Oikos*, n/a.
- 713 McLeod, A.M., Leroux, S.J. & Chu, C. (2020). Effects of species traits, motif profiles, and environment on  
714 spatial variation in multi-trophic antagonistic networks. *Ecosphere*, 11, e03018.
- 715 Melián, C.J., Vilas, C., Baldó, F., González-Ortegón, E., Drake, P. & Williams, R.J. (2011). Eco-evolutionary  
716 Dynamics of Individual-Based Food Webs. In: *Advances in Ecological Research*, The Role of Body Size in  
717 Multispecies Systems (ed. Belgrano, A.). Academic Press, pp. 225–268.
- 718 Ohlmann, M., Miele, V., Dray, S., Chalmandrier, L., O'Connor, L. & Thuiller, W. (2019). Diversity indices for  
719 ecological networks: A unifying framework using Hill numbers. *Ecology Letters*, 22, 737–747.
- 720 Olesen, J.M., Bascompte, J., Dupont, Y.L., Elberling, H., Rasmussen, C. & Jordano, P. (2010). Missing and  
721 forbidden links in mutualistic networks. *Proceedings of the Royal Society B: Biological Sciences*, 278,  
722 725–732.
- 723 Pascual, M., Dunne, J.A. & Dunne, J.A. (2006). *Ecological Networks: Linking Structure to Dynamics in Food  
724 Webs*. Oxford University Press, USA.
- 725 Petchey, O.L., Beckerman, A.P., Riede, J.O. & Warren, P.H. (2008). Size, foraging, and food web structure.  
726 *Proceedings of the National Academy of Sciences*, 105, 4191–4196.

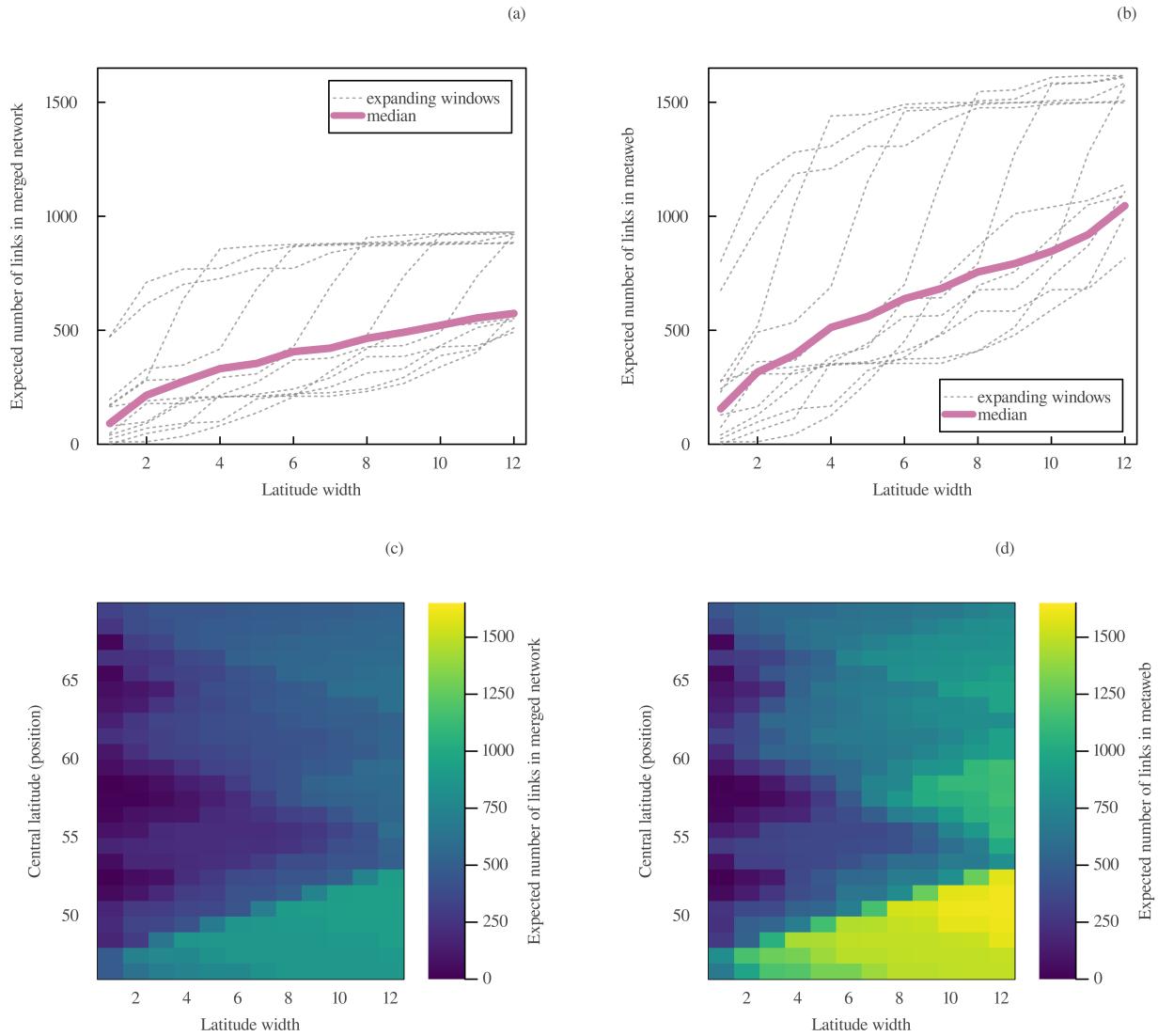
- 727 Pilosof, S., Porter, M.A., Pascual, M. & Kéfi, S. (2017). [The multilayer nature of ecological networks](#). *Nature Ecology & Evolution*, 1, 1–9.
- 728
- 729 Plagányi, É.E. & Butterworth, D.S. (2004). [A critical look at the potential of Ecopath with ecosim to assist in practical fisheries management](#). *African Journal of Marine Science*, 26, 261–287.
- 730
- 731 Poisot, T., Canard, E., Mouillot, D., Mouquet, N. & Gravel, D. (2012). [The dissimilarity of species interaction networks](#). *Ecology Letters*, 15, 1353–1361.
- 732
- 733 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.
- 734
- 735 Poisot, T. & Gravel, D. (2014). [When is an ecological network complex? Connectance drives degree distribution and emerging network properties](#). *PeerJ*, 2, e251.
- 736
- 737 Poisot, T., Stouffer, D.B. & Gravel, D. (2015). [Beyond species: Why ecological interaction networks vary through space and time](#). *Oikos*, 124, 243–251.
- 738
- 739 Polis, G.A. (1991). [Complex Trophic Interactions in Deserts: An Empirical Critique of Food-Web Theory](#). *The American Naturalist*, 138, 123–155.
- 740
- 741 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.
- 742
- 743
- 744 Pomeranz, J.P.F., Thompson, R.M., Poisot, T. & Harding, J.S. (2019). [Inferring predator-prey interactions in food webs](#). *Methods in Ecology and Evolution*, 10, 356–367.
- 745
- 746 Pringle, R.M. & Hutchinson, M.C. (2020). [Resolving Food-Web Structure](#). *Annual Review of Ecology, Evolution, and Systematics*, 51, 55–80.
- 747
- 748 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.
- 749
- 750 Pulliam, H.R. (1974). [On the Theory of Optimal Diets](#). *The American Naturalist*, 108, 59–74.
- 751 Rooney, N. & McCann, K.S. (2012). [Integrating food web diversity, structure and stability](#). *Trends in Ecology & Evolution*, 27, 40–46.
- 752

- 753 Salim, J.A., Saraiva, A.M., Zermoglio, P.F., Agostini, K., Wolowski, M., Drucker, D.P., *et al.* (2022). Data  
754 standardization of plantpollinator interactions. *GigaScience*, 11, giac043.
- 755 Sanders, D. & van Veen, F.J.F. (2012). Indirect commensalism promotes persistence of secondary consumer  
756 species. *Biology Letters*, 8, 960–963.
- 757 Saravia, L.A., Marina, T.I., Kristensen, N.P., De Troch, M. & Momo, F.R. (2022). Ecological network  
758 assembly: How the regional metaweb influences local food webs. *Journal of Animal Ecology*, 91, 630–642.
- 759 Schmeller, D.S., Julliard, R., Bellingham, P.J., Böhm, M., Brummitt, N., Chiarucci, A., *et al.* (2015). Towards a  
760 global terrestrial species monitoring program. *Journal for Nature Conservation*, 25, 51–57.
- 761 Shaw, J.O., Dunhill, A.M., Beckerman, A.P., Dunne, J.A. & Hull, P.M. (2024). A framework for reconstructing  
762 ancient food webs using functional trait data.
- 763 Singer, M.C. & McBride, C.S. (2012). Geographic mosaics of species' association: A definition and an  
764 example driven by plantinsect phenological synchrony. *Ecology*, 93, 2658–2673.
- 765 Stouffer, D.B., Rezende, E.L. & Amaral, L.A.N. (2011). The role of body mass in diet contiguity and food-web  
766 structure. *Journal of Animal Ecology*, 80, 632–639.
- 767 Strydom, T., Bouskila, S., Banville, F., Barros, C., Caron, D., Farrell, M.J., *et al.* (2022). Food web  
768 reconstruction through phylogenetic transfer of low-rank network representation. *Methods in Ecology and  
769 Evolution*, 13.
- 770 Strydom, T., Catchen, M.D., Banville, F., Caron, D., Dansereau, G., Desjardins-Proulx, P., *et al.* (2021). A  
771 roadmap towards predicting species interaction networks (across space and time). *Philosophical  
772 Transactions of the Royal Society B-Biological Sciences*, 376, 20210063.
- 773 Tylianakis, J.M., Tscharntke, T. & Lewis, O.T. (2007). Habitat modification alters the structure of tropical  
774 hostparasitoid food webs. *Nature*, 445, 202–205.
- 775 Vázquez§, D.P., Peralta§, G., Cagnolo, L., Santos, M. & Igual, §.E. autores contribuyeron por. (2022).  
776 Ecological interaction networks. What we know, what we don't, and why it matters. *Ecología Austral*, 32,  
777 670–697.
- 778 Weinstein, B.G. & Graham, C.H. (2017a). On comparing traits and abundance for predicting species  
779 interactions with imperfect detection. *Food Webs*, 11, 17–25.

- 780 Weinstein, B.G. & Graham, C.H. (2017b). Persistent bill and corolla matching despite shifting temporal  
781 resources in tropical hummingbird-plant interactions. *Ecology Letters*, 20, 326–335.
- 782 Whittaker, R.H. (1960). Vegetation of the Siskiyou Mountains, Oregon and California. *Ecological  
783 Monographs*, 30, 279–338.
- 784 Wood, S.A., Russell, R., Hanson, D., Williams, R.J. & Dunne, J.A. (2015). Effects of spatial scale of sampling  
785 on food web structure. *Ecology and Evolution*, 5, 3769–3782.
- 786 Woodward, G., Benstead, J.P., Beveridge, O.S., Blanchard, J., Brey, T., Brown, L.E., et al. (2010). Chapter 2 -  
787 Ecological Networks in a Changing Climate. In: *Advances in Ecological Research*, Ecological Networks  
788 (ed. Woodward, G.). Academic Press, pp. 71–138.
- 789 Woodward, G., Brown, L.E., Edwards, F.K., Hudson, L.N., Milner, A.M., Reuman, D.C., et al. (2012). Climate  
790 change impacts in multispecies systems: Drought alters food web size structure in a field experiment.  
791 *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 webs and the binary metaweb of host-parasite interactions. Aggregated local webs were obtained by sequentially and randomly selecting a number of binary local webs 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 webs. For a better comparison with binary webs, local probabilistic webs were derived from the probabilistic metaweb with a false positive and false negative rate of zero. A specific value of  $P(L|M)$  (the local probability of interaction among potentially interacting species) was used for all local webs within a particular curve. Aggregated probabilistic local webs were obtained by sequentially and randomly selecting a number of probabilistic local webs and aggregating both their species and interactions (with the value of  $P(L|M)$  adjusting according to eq. 6).



**Figure 2: Spatial scaling of interactions.** Expected number of host-parasite interactions in a network aggregating all probabilistic (a) local and (b) regional interactions within a latitudinal window of a given length. Every dashed curve corresponds to a different expanding window characterized by a different central latitude, with the colored solid line representing the median number of interactions. Heatmaps of the expected number of (c) local and (d) regional interactions found in windows of specified length and central latitudes. Probabilities of regional interactions were obtained with a false positive rate of 5% and a false negative rate of 10%. Local probabilistic interactions were derived from the probabilistic metaweb by setting the value of  $P(L|M)$  (the local probability of interaction among potentially interacting species) to 1, ensuring a conservative comparison between aggregated local webs and metawebs. Aggregated local webs were obtained by aggregating both the species and interactions found within a particular latitudinal window, with the values of  $P(L|M)$  remaining at their maximum value of 1 following eq. 6.

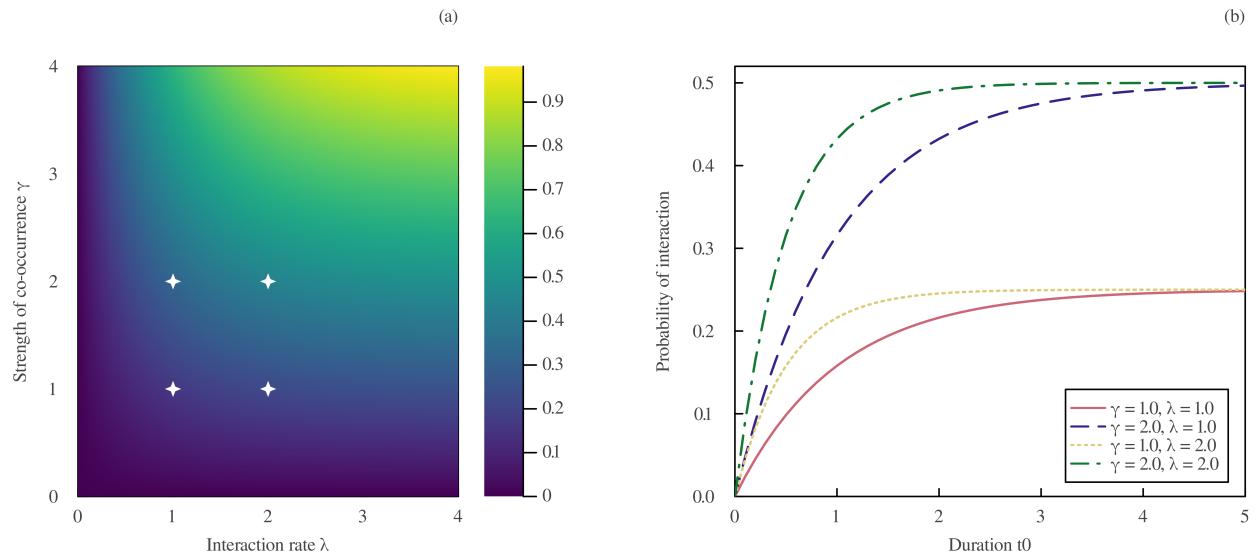
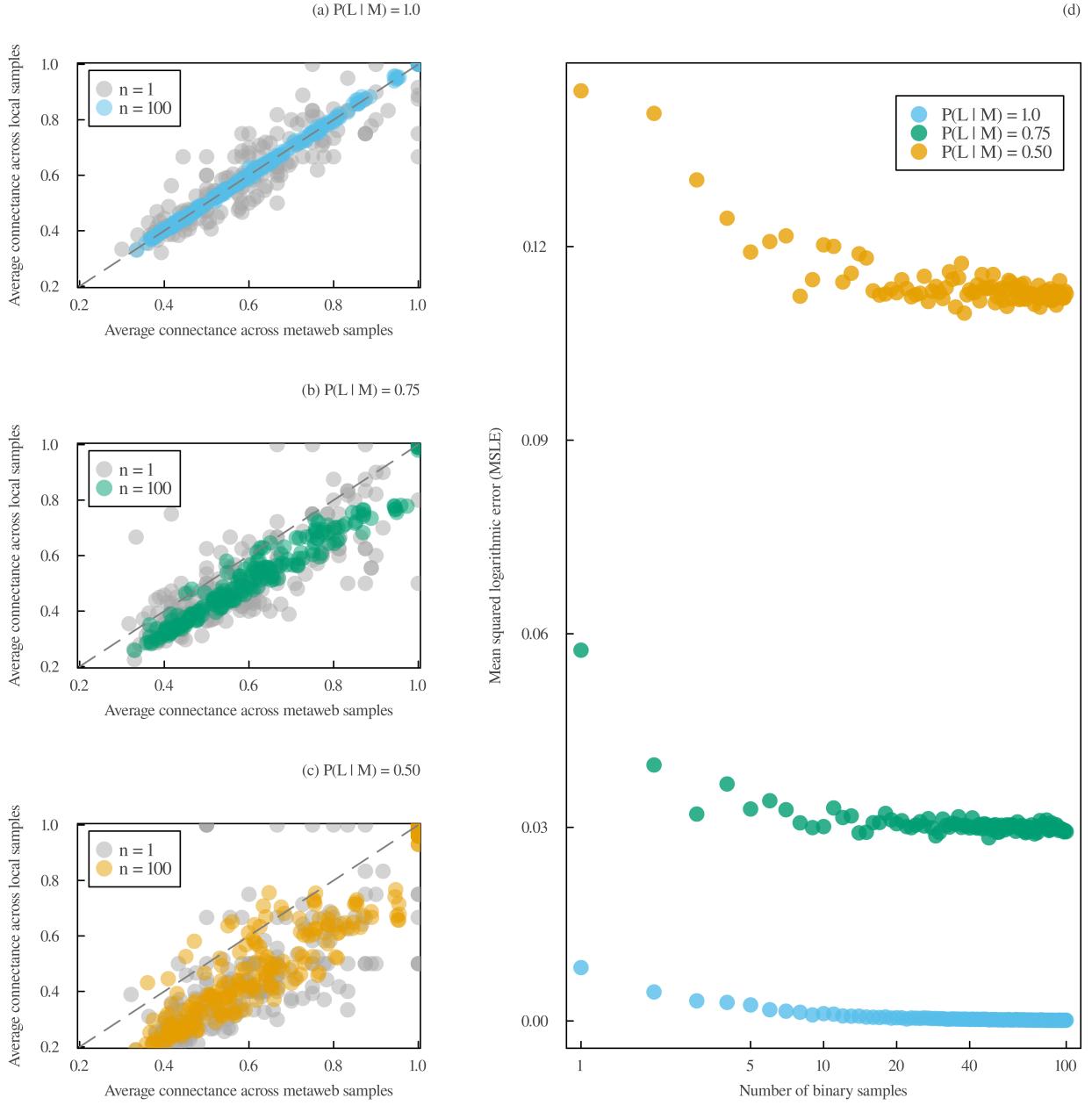


Figure 3: **Parameters of the spatiotemporally explicit model of interactions.** (a) Probability of local interaction given by the process model (eq. 11) under different values of  $\lambda$  and  $\gamma$ , with  $t_0 = 1$ . The parameter values used in the right panel are denoted by the white stars. (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 webs.** Comparison between the average connectance of binary network samples obtained from the probabilistic local and metawebs. Each dot corresponds to a different site. The local probability of interaction between potentially interacting species was set to three different values: (a)  $P(L|M) = 1.0$ , (b)  $P(L|M) = 0.75$ , and (c)  $P(L|M) = 0.50$ . 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  $P(L|M)$  used in panels a-c. Probabilities of regional interactions were obtained with a false positive rate of 5% and a false negative rate of 10%. Metaweb samples were obtained by randomly sampling binary interactions from the probabilistic metaweb, and then propagating this result to all local webs that include the species potentially engaged in the interactions. Local binary webs were generated by independently sampling binary interactions for each local probabilistic web.