

# Deciphering probabilistic species interaction networks

Francis Banville<sup>1,2,3</sup> Tanya Strydom<sup>1,3</sup> Penelope S. A. Blyth<sup>4</sup> Chris Brimacombe<sup>5</sup> Michael Catchen<sup>3,6</sup>  
Gabriel Dansereau<sup>1,3</sup> 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

Representing species interactions probabilistically (how likely are they to occur?) as opposed to deterministically (are they occurring?) expresses uncertainties in our knowledge of interactions. The sources of uncertainty captured by interaction probabilities depend on the method used to evaluate them: uncertainty of predictive models, subjective assessment of experts, or empirical measurement of interaction spatiotemporal variability. However, guidelines for the estimation and documentation of probabilistic interaction data are still lacking. This is concerning because our understanding and analysis of interaction probabilities depend on their sometimes elusive definition and uncertainty sources. We review how probabilistic interactions are defined at different spatial scales, from local interactions to regional networks (metawebs), with a strong emphasis on host-parasite and trophic (predatory and herbivory) interactions. These definitions are based on the distinction between the realization of an interaction at a specific time and space (local) and its biological feasibility (regional). Through case studies using host-parasite interactions in Europe, we illustrate how these two network representations differ in their statistical properties, specifically: how local networks and metawebs differ in their spatial and temporal scaling of probabilistic interactions, but not in their taxonomic scaling. We present two approaches to inferring binary interactions from probabilistic ones that account for these differences and show that systematic biases arise when directly inferring local networks from metawebs. Our results underscore the importance of more rigorous descriptions of probabilistic species interaction networks that specify their type of interaction (local or regional) and uncertainty sources.

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



Last revision: April 21, 2024

## **1 Introduction**

2 As we try to navigate global biodiversity change, filling in knowledge gaps about biodiversity becomes  
3 instrumental to monitoring and mitigating those changes (Abrego *et al.* 2021; Gonzalez & Londoño 2022;  
4 Hortal *et al.* 2015). However, cataloging species, populations and, in particular, ecological interactions (e.g.,  
5 predation, parasitism, and pollination) is a substantial challenge (Pascual *et al.* 2006; Polis 1991). There are  
6 methodological and biological constraints that hinder our ability to document species interactions, inevitably  
7 leading to uncertainty in our knowledge of interactions. For example, the spatial and temporal uncoupling of  
8 species (e.g., nocturnal and diurnal species coexisting in the same space with different daily activity timings,  
9 Jordano 1987) and the large number of rare and cryptic interactions in a community, contribute to these  
10 knowledge gaps (Jordano 2016).

11 More generally, several conditions must be satisfied for an interaction to be observed locally. First, both species  
12 must have overlapping geographic ranges, i.e. they must co-occur within the region of interest (Cazelles *et al.*  
13 2016; Morales-Castilla *et al.* 2015). Second, they must have some probability of meeting (Poisot *et al.* 2015).  
14 Probabilities of interspecific encounters are typically low, especially for rare species with low abundances.  
15 While species' absolute abundances may impact interaction frequencies (Vázquez *et al.* 2007), encounter  
16 probabilities are determined by their relative abundances (Canard *et al.* 2012). The probability that species  
17 meet also depends on their biology, such as their phenology (Olesen *et al.* 2010; Singer & McBride 2012) and  
18 discoverability (Broom & Ruxton 2005). Finally, when species do come into contact, an interaction occurs only  
19 if their traits are locally compatible (Poisot *et al.* 2015), including their phenotypes (Bolnick *et al.* 2011; Gravel  
20 *et al.* 2013; Stouffer *et al.* 2011) and behavior (Choh *et al.* 2012; Pulliam 1974). Interactions may also be  
21 influenced by a third species (e.g., a more profitable prey species, Golubski & Abrams 2011; Sanders & van  
22 Veen 2012). Documenting the location and timing of interactions becomes even more difficult when accounting  
23 for the spatiotemporal variability of ecological interactions (Poisot *et al.* 2012, 2015). Environmental factors,  
24 such as temperature (Angilletta *et al.* 2004), drought (Woodward *et al.* 2012), climate change (Araujo *et al.*  
25 2011; Gilman *et al.* 2010; Woodward *et al.* 2010), and habitat modifications (Tylianakis *et al.* 2007), contribute  
26 to this spatiotemporal variability by impacting species abundance and traits. Even under favorable  
27 circumstances, there remains a possibility that the interaction does not occur locally, either due to the intricate  
28 nature of the system or simply by chance. If it does occur, it might go undetected, particularly if it happens  
29 infrequently. In this context, it is unsurprising that our knowledge of ecological interactions remains limited

30 (Hortal *et al.* 2015) despite extensive biodiversity data collection (Schmeller *et al.* 2015).

31 Knowing the biological capacity of two species to interact directly (via e.g., trophic interactions) is necessary

32 but not sufficient for inferring their interaction at a specific time and space. The recognition of the intrinsic

33 variability of species interactions has led ecologists to expand their representation of ecological networks to

34 include a probabilistic view of interactions (Dallas *et al.* 2017; Fu *et al.* 2021; Poisot *et al.* 2016). This allows

35 filling in the Eltonian shortfall (Hortal *et al.* 2015) by modeling the probability of occurrence of interactions

36 (e.g., Gravel *et al.* 2019), which can be an important tool for directing efforts and taking action (Carlson *et al.*

37 2021), especially in places where access and resources for research are scarce. The probabilistic representation

38 of interactions has been applied to direct interactions, which are conceptually and mathematically analogous

39 regardless of their biological type (e.g., predation and pollination). This is in contrast with indirect interactions

40 (e.g., interspecific competition), which arise from distinct ecological processes and are often not directly

41 observable (Kéfi *et al.* 2015, 2016).

42 Representing direct interactions probabilistically can capture the spatiotemporal variability of the

43 aforementioned ecological processes and the uncertainty in our knowledge of interactions. Networks of

44 probabilistic interactions, within a Bayesian perspective, express our degree of belief (or confidence) regarding

45 the occurrence or observation of interactions. In contrast, interactions are simply regarded as either occurring or

46 not in networks of deterministic binary interactions. Based on the scale at which they are estimated, interaction

47 probabilities may reflect our level of confidence in whether interactions will be observed, realized locally, or

48 biologically feasible. As an illustration, we could outline a situation in which there is a 50% certainty that an

49 interaction occurs 50% of the time. Our level of confidence should be more definitive (approaching either 0 or

50 1) as we extend our sampling to a broader area and over a longer time period, thereby diminishing the

51 uncertainty of our knowledge of interactions (but not necessarily the estimation of their variability). In the

52 broadest sense, binary interactions are also a type of probabilistic interaction, in which the numerical value of

53 an interaction is restrained to 0 (non-occurring) or 1 (occurring). In networks of probabilistic interactions, only

54 forbidden interactions (i.e., interactions prohibited by biological traits or species absence, Jordano *et al.* 2003;

55 Olesen *et al.* 2010) have a probability value of zero, provided that intraspecific trait variability is considered

56 (Gonzalez-Varo & Traveset 2016).

57 By accounting for the uncertainty of interactions, networks of probabilistic interactions may provide a more

58 realistic portrait of species interactions. However, how the uncertainty of pairwise interactions propagates to

59 network structure (i.e., community-level properties driving the functioning, dynamics, and resilience of

60 ecosystems, McCann 2007; McCann 2011; Proulx *et al.* 2005; Rooney & McCann 2012) remains to be  
61 elucidated. The application and development of computational methods in network ecology, often based on a  
62 probabilistic representation of interactions, can alleviate (and guide) the sampling efforts required to document  
63 species interactions (Strydom *et al.* 2021). For example, statistical models can be used to estimate the  
64 uncertainty of pairwise interactions (Cirtwill *et al.* 2019) and the probability of missing (false negatives) and  
65 spurious (false positives) interactions (Guimerà & Sales-Pardo 2009). Statistical models can generate  
66 predictions of ecological networks without prior knowledge of pairwise interactions. They may do so using  
67 body size (Gravel *et al.* 2013; Petchey *et al.* 2008), phylogeny (Elmasri *et al.* 2020; Strydom *et al.* 2022), or a  
68 combination of niche and neutral processes (Bartomeus *et al.* 2016; Pomeranz *et al.* 2019) for inference.  
69 Topological null models, which generate networks of probabilistic interactions by preserving chosen  
70 characteristics of the adjacency matrix of binary interactions while intentionally omitting others (Bascompte *et*  
71 *al.* 2003; Fortuna & Bascompte 2006), are examples of common probabilistic interaction models. Null models  
72 can be used to produce underlying distributions of network measures for null hypothesis significance testing.  
73 Many measures have been developed to describe the structure (Poisot *et al.* 2016) and diversity (Godsoe *et al.*  
74 2022; Ohlmann *et al.* 2019) of probabilistic interaction networks. These models and measures support the use  
75 of this approach for the study of a wide range of ecological questions, from making better predictions of species  
76 distribution (Cazelles *et al.* 2016) to forecasting the impact of climate change on ecological networks (Gilman  
77 *et al.* 2010).  
78 Yet, a precise definition of probabilistic interactions appears to be lacking, making the estimation and use of  
79 these data more difficult. In this manuscript, we aim to take a step back by outlining different ways in which  
80 probabilistic interactions are defined and used in network ecology. We distinguish two broad categories of  
81 probabilistic interaction networks that necessitate distinct approaches: local networks describing probabilities  
82 of realized interactions, and regional networks (metawebs) describing probabilities of potential interactions. We  
83 highlight the distinctions in the ecological meaning of these two representations and examine some of their  
84 properties and relationships (particularly with space, time, and between each other) through empirical case  
85 studies. Moreover, the lack of clear guidelines on the use of probabilistic interaction data is worrisome, as it  
86 affects both data producers and re-users who generate and manipulate these numbers. This is concerning  
87 because sampling strategies and decisions regarding network construction can affect our understanding of  
88 network properties (Brimacombe *et al.* 2023). There is currently no reporting standard that could guide the  
89 documentation of all types of probabilistic interactions (although see e.g., Salim *et al.* 2022 who discuss data

90 standards for deterministic mutualistic networks). Well-defined reporting standards for probabilistic interactions  
91 would support more adequate manipulation and integration of interaction data from different sources and guard  
92 against possible misinterpretations arising from ambiguous definitions of probabilistic interaction networks.  
93 This documentation should outline the nature (i.e., local or regional) and type (e.g., predatory or pollination) of  
94 the interactions, provide information regarding the taxonomic level, identities, and characteristics (e.g., life  
95 stages) of the individuals involved in an interaction, present the mathematical formulation of probabilities,  
96 including clearly identified conditional variables (e.g., spatial and temporal scales), and describe the methods  
97 and contexts (e.g., location, time, environmental conditions) in which interactions were estimated. Inadequately  
98 documented probabilistic interaction data should be used with caution when analyzing ecological networks.  
99 The broad principles underlying our findings remain relevant and applicable across diverse ecological contexts.

## 100 **Probabilistic representations of interactions**

101 Consider a scenario where an avian predator has just established itself in a northern habitat home to a small  
102 rodent. Suppose these species have never co-occurred before, and as a result, their interaction has not been  
103 previously observed. What is the probability that the rodent is part of the diet of the avian predator, or put  
104 differently, what is the probability that they interact? Answering this question requires some clarification, as  
105 there are multiple ways to interpret and calculate interaction probabilities. We could calculate the probability  
106 that the traits of these species match, i.e. that the avian predator possesses the biological attributes to capture  
107 and consume the rodent. We could also calculate the probability that their traits support an interaction under the  
108 typical environmental conditions of the new habitat. For example, because avian predators hunt by sight,  
109 predation could be possible in the absence of snow but highly improbable when snow is present, as rodents may  
110 use it as a shelter to hide from predators. Finally, we could calculate the probability that the avian predator will  
111 consume the rodent at *that* particular location, for which the spatial and temporal boundaries need to be  
112 specified. The estimation of interaction probabilities, whether through predictive models or prior distributions,  
113 hinges on our understanding of these probabilities and the specific ecological processes we aim to capture.

114 An important aspect to consider when using probabilities of interactions is knowing if they describe potential or  
115 realized interactions, as these two types of interactions have distinct conceptual underpinnings and sources of  
116 uncertainty. A potential (regional) interaction is defined as the biological capacity of two taxa to interact (i.e.,  
117 the probability that they interact if they were to encounter each other and given sufficient time) whereas a

realized (local) interaction is the occurrence or observation of this interaction in a well-defined space and time (i.e., the probability that they interact locally). For two co-occurring taxa and over infinite time, the probability of local interaction is equivalent to the probability of regional (potential) interaction. We use the terms *metaweb* (Dunne 2006) to designate regional networks of potential interactions and *local networks* (Poisot *et al.* 2012) for those of realized interactions. Metawebs are the network analogs of the species pool, where local networks originate from a subset of both species (nodes) and interactions (edges) of the regional metaweb (Saravia *et al.* 2022). Without clear documentation, it can be challenging to know if published probabilistic interaction networks describe local or regional interactions (Tbl. 1 provides examples of studies employing both types of probabilistic interaction networks), or if so-called probabilities are in reality a form of interaction score. When probabilistic regional interactions are used and interpreted incorrectly as local interactions (and conversely), this may generate misleading findings during data analysis. We believe that a better understanding of probabilistic local and regional interactions would prevent interpretation errors (e.g., when studying network-area relationships with metawebs or local networks) and facilitate a more adequate use of interaction data. In the following sections, we delve into these distinctions as we scale up from pairwise interactions to higher-level representations of ecological networks, i.e. local networks and metawebs.

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

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

Ecologists have traditionally represented interactions (edges) as binary objects that were considered realized after observing at least one individual from group  $i$  interact with at least another individual from group  $j$ . In an adjacency matrix  $B$  of binary interactions, the presence or absence of an interaction  $B_{i \rightarrow j}$  between two taxa can be viewed as the result of a Bernoulli trial  $B_{i \rightarrow j} \sim \text{Bernoulli}(P(B_{i \rightarrow j}))$ , with  $P(B_{i \rightarrow j})$  being the probability of interaction. This interaction probability characterizes our limited ecological knowledge and/or the intrinsic spatiotemporal variability of interactions. It may be estimated through predictive models (e.g., those based on biological traits and species abundances) or expert (prior) knowledge about the interaction. In networks of

147 probabilistic interactions, the edge values  $P(B_{i \rightarrow j})$  are probabilistic events whose only two possible outcomes  
148 are the presence ( $B_{i \rightarrow j} = 1$ ) or absence ( $B_{i \rightarrow j} = 0$ ) of an interaction between each pair of nodes. Depending on  
149 the type of probabilistic interaction network (local or metaweb), the mathematical formulation and interpretation  
150 of stochastic parameters like  $P(B_{i \rightarrow j})$  can be linked to environmental and biological factors such as species  
151 abundance, traits, area, and time, for example using logistic regression with continuous explanatory variables.

152 Predicting the number of local networks in which the interaction between two given taxa occurs can be achieved  
153 by using a Binomial distribution, assuming a constant interaction probability and independence between  
154 networks (trials). When considering uncertainties around the estimation of  $P(B_{i \rightarrow j})$  or its spatiotemporal  
155 variability, a Beta distribution may be used to represent the relative likelihood of different probability values.  
156 For example, when calculating the probability of interaction between two taxa based on their local abundances,  
157 any spatiotemporal fluctuations in their abundances would introduce variability in the interaction probability at  
158 the local scale. If we take into account the stochasticity of the interaction probability, a Beta-Binomial  
159 distribution can be used to predict the number of networks in which the interaction occurs. Empirically  
160 observing an interaction between two taxa at a given location and time provides important information that can  
161 be used to update previous estimates of  $P(B_{i \rightarrow j})$ , informing us on the biological capacity of both taxa to interact  
162 and the environmental conditions that enabled them to interact locally.

163 Even though binary interaction networks constitute a highly valuable source of ecological information (Pascual  
164 *et al.* 2006), they overlook interaction strengths. Represented in a quantitative adjacency matrix  $W$ , interaction  
165 strengths better describe the energy flows, demographic impacts or frequencies of interactions between nodes  
166 (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  
167 measure. For example, they may represent local interaction rates between pairs of taxa (e.g., the flower-visiting  
168 rates of pollinators in a mutualistic network, Herrera 1989). When interaction strengths characterize predation  
169 pressure on prey, they can serve as parameters in a Lotka-Volterra model (e.g., Emmerson & Raffaelli 2004).

170 The extra amount of ecological information in quantitative networks typically comes at a cost of greater  
171 sampling effort and data volume in predictive models (Strydom *et al.* 2021), which can lead to relatively high  
172 levels of uncertainties when inferring quantitative networks with limited data.

173 As for binary interaction networks, the uncertainty and variability of interaction strengths can be represented  
174 probabilistically. However, the need to estimate the probability distribution of interaction strengths makes the  
175 inference of probabilities more challenging in quantitative networks compared to binary interaction networks.  
176 Interaction strengths can follow many probability distributions depending on the method. For instance, they can

follow a Poisson distribution  $W_{i \rightarrow j} \sim \text{Poisson}(\lambda_{i \rightarrow j})$  when predicting the number 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 networks due to their typically high sparseness (Jordano 2016). Because of the methodological difficulties typically encountered when building deterministic quantitative networks (which are only partially mitigated by models such as Ecopath, Plagányi & Butterworth 2004), binary interaction networks, which are easier to sample (Jordano 2016) and predict (Strydom *et al.* 2021), have been more frequently studied and modeled. Moreover, most published probabilistic interaction networks and methods describe 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 networks first. For these reasons, our primary focus is interpreting and using interaction probabilities in Bernoulli distributions, in both local networks and metawebs.

## Local networks: communities interacting in space and time

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

We define space as the collection of geographic coordinates  $(x, y, z)$ , with  $(x, y)$  representing longitude and latitude coordinates, and  $z$  denoting either altitudes or depths. These point coordinates delineate the spatial boundaries of the system, which may be portrayed as a polyhedron. Ecological interactions may exhibit variations along latitudinal and altitudinal gradients, as evidenced by changes in hummingbird-plant interactions (Weinstein & Graham 2017a, b) and mosquito biting rates (e.g., Kulkarni *et al.* 2006) at different elevations. In contrast, time is defined as the specific time period within which interactions were either observed or predicted. Even though space and time are continuous variables that should yield probability *densities* of

interactions (i.e., relative likelihoods of interactions occurring at infinitesimal locations and instants in time), these definitions enable them to be conceptualized as distinct patches and time segments. Treating space and time as discrete dimensions aligns with the common sampling methods of ecological networks and provides probabilities of interactions, which can be obtained by integrating probability densities over space and time. Furthermore, we can quantify both an area  $A$  and a duration  $t$ , which can be readily used in spatiotemporal analyses of ecological networks. When studying network-area relationships (NAR, Galiana *et al.* 2018), we anticipate that local probabilities of interactions scale positively with area and duration because taxa have more opportunities to interact as these dimensions expand.

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

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

Other important factors that can impact interaction probabilities at the local scale are taxa local abundances  $N_{i,k}$  and  $N_{j,k}$ , which affect encounter probabilities (Canard *et al.* 2012), and local traits distributions  $T_{i,k}$  and  $T_{j,k}$  (e.g., body mass, presence of given anatomical features, host resistance), which determine the ability of individuals to interact after encountering each other (Poisot *et al.* 2015). Moreover, local interactions may be conditional on local environmental factors such as temperature (Angilletta *et al.* 2004), precipitation (Woodward *et al.* 2012), habitat structure (Klecka & Boukal 2014), and presence of other taxa in the network (Kéfi *et al.* 2012; Pilosof *et al.* 2017), as described above. Here, we use the variable  $E_k$  to describe the local environmental context in which interaction probabilities were estimated. For example, in a mesocosm

233 experiment estimating interaction probabilities between predators and prey with and without shelters,  $E_k$  would  
234 represent the presence or absence of these shelters. Like co-occurrence,  $E_k$  can also be modeled  
235 probabilistically when the stochasticity or uncertainty of environmental factors is considered.  $E_k$  represents all  
236 environmental variables that were taken into consideration when measuring interaction probabilities; it is a  
237 subset of all environmental factors acting on ecological interactions. Finally, local interaction probabilities may  
238 be conditioned on higher-level properties of the network, which we denote by  $f(L)$ . Many topological null  
239 models (i.e., statistical models that randomize interactions by retaining certain properties of the network while  
240 excluding others) provide interaction probabilities from selected measures of network structure, such as  
241 connectance (Fortuna & Bascompte 2006) and the degree distribution (Bascompte *et al.* 2003).

242 The probability that two taxa  $i$  and  $j$  interact in a local network  $L$  can thus be conditioned on their co-occurrence  
243  $X_{i,j,k}$  (or more explicitly on their occurrences  $X_{i,k}$  and  $X_{j,k}$ ), local abundances  $N_{i,k}$  and  $N_{j,k}$ , local traits  
244 distributions  $T_{i,k}$  and  $T_{j,k}$ , local environmental conditions  $E_k$ , network area (or volume)  $A$ , time interval  $t$ , and  
245 network properties  $f(L)$ .

246 Although these variables correspond to distinct ecological inquiries or mechanisms related to ecological  
247 interactions, they may covary with each other, such as the possible dependence of  $X_{i,j,k}$  and  $E_k$  on spatial and  
248 temporal scales. When estimating interaction probabilities using e.g. a generalized linear model with multiple  
249 explanatory variables that might not be independent, it may become important to address collinearity. In such a  
250 case, to mitigate this issue, it may be necessary to use variable selection techniques before fitting the model to  
251 data. The probability that a local interaction is realized is described by the following expression when all these  
252 conditional variables are included:

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

253 The representation of the local context in which probabilities are estimated and the variables that should be  
254 taken into consideration depend on the study system, the objectives of the study, and the resources available to  
255 the researchers. In other words, these variables do not systematically need to be accounted for. For example, in  
256 Gravel *et al.* (2019), for the purpose of model comparison, some models incorporated environmental variables  
257 as conditions of interaction probabilities, while others did not. When accounted for, these variables should be  
258 clearly described in the documentation of the data (Brimacombe *et al.* 2023), preferentially in mathematical  
259 terms to avoid any confusion in their interpretation and to limit manipulation errors during their re-use. For

260 instance, ecologists should be explicit about their consideration of co-occurrence in their estimation of local  
 261 interaction probabilities. Indeed, it is important to specify if probability values are conditional  
 262 ( $P(L_{i \rightarrow j}|X_{i,j,k} = 1)$ ) or not ( $P(L_{i \rightarrow j})$ ) on co-occurrence, as this can change the interpretation of the data. In  
 263 Tbl. 1, we present examples of studies that used these different formulations of probabilistic interactions and  
 264 conditional variables. We have included the probability of empirically observing an interaction that is realized  
 265 locally  $P(O_{i \rightarrow j}|L_{i \rightarrow j})$  to underscore the distinction between local observations and actual realizations of  
 266 interactions, even though the focus of this manuscript is not on the observation of interactions.

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

Expression	Type	Outcome	Variability sources	Reference
$P(L_{i \rightarrow j} X_{i,k}, X_{j,k}, \dots)$	local	realization of the interaction given taxa co-occurrence	spatiotemporal variability	Gravel <i>et al.</i> (2019)
$P(L_{i \rightarrow j} N_{i,k}, N_{j,k}, \dots)$	local	realization of the interaction given taxa abundances	neutral models	Canard <i>et al.</i> (2014)
$P(L_{i \rightarrow j} T_{i,k}, T_{j,k}, \dots)$	local	realization of the interaction given local traits	trait matching models	Gravel <i>et al.</i> (2016)
$P(L_{i \rightarrow j} E_k, \dots)$	local	realization of the interaction given local environmental conditions	environmental- based models	Gravel <i>et al.</i> (2019) (temperature and precipitation)
$P(L_{i \rightarrow j} A, \dots)$	local	realization of the interaction in a given area or volume	spatial models	Galiana <i>et al.</i> (2018) *
$P(L_{i \rightarrow j} t, \dots)$	local	realization of the interaction during a given time period	temporal models	Weinstein & Graham (2017a)
$P(L_{i \rightarrow j} f(L), \dots)$	local	realization of the interaction given network structure	topological null models	Fortuna & Bascompte (2006) (connectance)

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

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

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

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

## Metawebs: regional catalogs of interactions

Metawebs (Dunne 2006) are networks of potential interactions over broad spatial, temporal, and taxonomic scales (e.g., food webs at the continental scale). Potential interactions describe the biological capacity of taxa to interact, which is typically assessed at the regional scale. Metawebs of probabilistic interactions are particularly

useful in situations where there is uncertainty in the ability of taxa to interact. This uncertainty frequently arises due to insufficient interaction data, especially for taxa that have not yet been observed to co-occur, and uncertainties in trait-matching models. As data accumulates, interactions in metawebs should tend towards binarity, approaching probability values of 0 (repeatedly failing to observe an interaction between two co-occurring taxa) and 1 (observing an interaction at least once). The extent of sampling effort thus influences our evaluation of probabilities of potential interactions, as sampling over a larger area or for a longer duration enables us to capture a greater number of regional interactions (McLeod *et al.* 2021). However, in contrast with local networks of probabilistic interactions, which describe local interaction stochasticity, regional interactions are not evaluated for any particular local context. Although *neutrally* forbidden interactions (i.e., between rare species, Canard *et al.* 2012) tend to have low probability values in local networks, they may have higher probabilities in the metaweb (i.e., when species' biological traits can support an interaction if they were to encounter each other). Likewise, non-co-occurring taxa may have a non-zero probability of interaction in the metaweb.

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

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

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

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

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

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

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

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

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

325 This is because the biological feasibility of an interaction is a prerequisite for its ecological feasibility.  
326 Biological feasibility is necessary but not sufficient for an interaction to be ecologically feasible. Our discussion

327 of metawebs will focus on the biological feasibility of interactions since most methods developed for inferring  
328 probabilities of regional interactions do not explicitly take into account environmental conditions (e.g., Strydom  
329 *et al.* 2022).

330 Starting from a selected set of taxa, which are usually distributed within a broad region of interest, metawebs  
331 can be built using different data sources, including literature review (e.g., Maiorano *et al.* 2020), aggregated  
332 interaction data (e.g., Gravel *et al.* 2019; Saravia *et al.* 2022), trait-matching models (e.g., Shaw *et al.* 2024;  
333 Strydom *et al.* 2022), and expert knowledge. Every pair of taxa that have confidently been observed to interact  
334 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  
335 differs from local networks of probabilistic interactions, where interaction events may remain stochastic (i.e.,  
336  $P(L_{i \rightarrow j}) < 1$ ) even after empirically observing interactions due to their spatiotemporal variability. Interactions  
337 that were never observed typically have low probability values in local networks and vary from low to high  
338 values in metawebs, contingent upon taxa traits distributions (reaching 0 for forbidden links).

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

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

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

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

354

The quick brown fox jumps right over the lazy dog. the quick brown fox jumps right over the lazy dog. the quick brown fox jumps right over the lazy dog. the quick brown fox jumps right over the lazy dog. the quick brown fox jumps right over the lazy dog. the quick brown fox jumps right over the lazy dog. the quick brown fox jumps right over the lazy dog. the quick brown fox jumps right over the lazy dog.

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

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

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

$$P(X_{i,k}, X_{j,k}) = P(X_{i,k}|X_{j,k})P(X_{j,k}) = P(X_{i,k})P(X_{j,k})\gamma. \quad (9)$$

376 When  $\gamma > 1$ , it signifies a positive association in the geographic distributions of both taxa, indicating that the

377 presence of one taxon enhances the probability of occurrence of the other. Attractions may be the result of  
 378 positive interactions (e.g., mutualism) or positive/negative interactions (e.g., antagonism between predators and  
 379 prey, Cazelles *et al.* 2016). In empirical networks,  $\gamma > 1$  holds true for the majority of species pairs (Catchen *et*  
 380 *al.* 2023). In contrast, repulsions ( $\gamma < 1$ ) may be caused e.g. by strong interspecific competition (Cazelles *et al.*  
 381 2016). We model the co-occurrence  $X_{i,j,k}$  of both taxa as the outcome of a Bernoulli trial

$$X_{i,j,k} \sim \text{Bernoulli}(P(X_{i,k}, X_{j,k})). \quad (10)$$

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

$$P(L_{i \rightarrow j} | X_{i,j,k} = 1) = 1 - e^{-\lambda_k t_0}, \quad (11)$$

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

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

$$L_{i \rightarrow j} \sim \text{Bernoulli}(P(L_{i \rightarrow j})) \quad (12)$$

$$P(L_{i \rightarrow j}) = P(X_{i,k})P(X_{j,k})\gamma(1 - e^{-\lambda_k t_0}) \quad (13)$$

$$\gamma \sim \text{Gamma}(2, 0.5) \quad (14)$$

$$\lambda_k \sim \text{Exponential}(2) \quad (15)$$

394 In Fig. 1, we show the variation in the probability of interaction under different parameter values. In the right  
395 panel, we notice that, irrespective of the interaction rate  $\lambda_k$ , the probability of interaction converges toward an  
396 asymptote determined by the probability of co-occurrence  $P(X_{i,j,k})$  (Eq. 9). This model can be customized in  
397 different ways, such as by linking  $\lambda_k$  with specific environmental variables or explicitly incorporating  
398 observation errors (i.e., the probabilities of false negatives and false positives).

399 [Figure 1 about here.]

## 400 Properties of probabilistic interaction networks

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

### 412 Application 1: Dissimilarity of binary and probabilistic interaction networks

413 We use the collection of tripartite host-parasite networks sampled across Europe, created by Kopelke *et al.*  
414 (2017), in the following case studies. This dataset contains well-resolved binary local interactions between  
415 willows (52 species), willow-galling sawflies (96 species), and their parasitoids (126 species). Given its  
416 replicated networks spanning large spatiotemporal scales, this dataset is well-suited for analyzing a variety of

ecological hypotheses and processes. Out of a total of 374 local networks, we retained those containing at least 5 species, resulting in a set of 233 georeferenced local networks (networks sampled within areas of 0.1 to 0.3 km<sup>2</sup> during June and/or July spanning 29 years). We built a metaweb of binary interactions by aggregating all local interactions, which gave us a regional network composed of 274 species and 1080 interactions. In Fig. 2, we show how the dissimilarity between the metaweb and the aggregated local networks varies with the number of sampled local networks. To do so, we randomly selected one local network of binary interactions and sequentially sampled additional networks while aggregating both their species and interactions. Next, we compared the metaweb and the aggregated local networks using the dissimilarity of interactions between common species ( $\beta_{OS}$ , Fig. 2a) and the dissimilarity in species composition ( $\beta_S$ , Fig. 2b) (Poisot *et al.* 2012). We repeated this sampling process one hundred times and highlighted the median dissimilarity values across simulations, as well as the 50% and 95% percentile intervals. This shows that networks of local interactions are highly dissimilar from the metaweb, both in terms of species and interactions, especially when only a limited number of sites has been sampled. Both dissimilarity indices were calculated based on the number of items shared by the two networks ( $c_{LM}$ ) and the number of items unique to the metaweb ( $u_M$ ) and to the aggregated local network ( $u_L$ ). The  $\beta_S$  dissimilarity index uses species (nodes) as items being compared, while the  $\beta_{OS}$  index assesses dissimilarity based on interactions between shared species (Poisot *et al.* 2012). Both indices were calculated following the  $\beta_W$  index of Whittaker (1960):

$$\beta_W = \frac{c_{LM} + u_L + u_M}{(2c_{LM} + u_L + u_M)/2} - 1. \quad (16)$$

[Figure 2 about here.]

Next, we investigate the differences in scaling between networks of probabilistic or binary interactions (Fig. 2c-d). We converted these binary interaction networks into probabilistic ones using models based on simple assumptions. Our models are not designed to estimate the exact values of probabilistic interactions. Instead, their purpose is to create plausible networks that serve as illustrative examples to highlight distinctions between local networks and metawebs of probabilistic interactions. We created two metawebs of probabilistic interactions by employing constant false positive and false negative rates for all regional interactions. In the first metaweb, we set both false positive and false negative rates to zero to prevent artificially inflating the total number of links, enabling a more accurate comparison with binary interaction networks. This gave us a probability of regional interaction of 1 when at least one interaction has been observed locally and of 0 in the

444 absence of any observed interaction between a given pair of species (Fig. 2c-d). In the second metaweb (used in  
 445 the next applications), we introduced a 5% false positive rate to account for spurious interactions and a 10%  
 446 false negative rate to address the elevated occurrence of missing interactions in ecological networks (Catchen *et*  
 447 *al.* 2023). We believe these rates represent reasonable estimates of missing and spurious potential interactions,  
 448 but confirming their accuracy is challenging due to the unavailability of data on the actual feasibility of  
 449 interaction. Observed interactions were thus given a probability of regional interaction of 95%, whereas  
 450 unobserved ones were assigned a probability of 10%.

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

$$P(L_{i \rightarrow j}) = P(L_{i \rightarrow j}|M_{i \rightarrow j} = 1) \times P(M_{i \rightarrow j} = 1). \quad (17)$$

456 We built local networks of probabilistic interactions using the taxa found in the empirical local networks and  
 457 attributing pairwise interaction probabilities based on the metawebs of probabilistic interactions and a constant  
 458 value of  $P(L_{i \rightarrow j}|M_{i \rightarrow j})$  across interactions. Low values of  $P(L_{i \rightarrow j}|M_{i \rightarrow j})$  indicate that feasible interactions rarely  
 459 occur locally, intermediate values around 50% suggest considerable spatiotemporal variability, while high  
 460 values indicate that regional interactions are nearly always realized locally. Following Eq. 17, the local  
 461 probability of interaction between a given pair of taxa consistently remained equal to or below their probability  
 462 of regional interaction.

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

$$P(L_{1,2,i \rightarrow j}|M_{i \rightarrow j}) = 1 - [1 - P(L_{1,i \rightarrow j}|M_{i \rightarrow j})] \times [1 - P(L_{2,i \rightarrow j}|M_{i \rightarrow j})], \quad (18)$$

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

474 By comparing the scaling relationships observed in networks of binary and probabilistic interactions; Fig. 2  
 475 illustrates that high values of  $P(L_{i \rightarrow j}|M_{i \rightarrow j})$  lead to systematic overestimations in the number of links and  
 476 connectance, especially when  $P(L_{i \rightarrow j}|M_{i \rightarrow j}) = 1$  (corresponding to the scenario where local probabilities of  
 477 interactions are equivalent to the probabilities of regional interactions). However, these biases tend to diminish  
 478 as the number of sampled networks increases.

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

480 Network-Area Relationships document the scaling of network properties (such as modularity and trophic chain  
 481 lengths) with spatial scale (e.g., Galiana *et al.* 2018; Wood *et al.* 2015). The variations in network structure  
 482 across spatial scales may stem from the scaling of species richness (species-area relationships, SARs), the  
 483 number of interactions (Brose *et al.* 2004), and many other higher-level properties of the system (e.g.,  
 484 environmental heterogeneity, Thompson & Townsend 2005) with the sampled area. Additionally, ecological  
 485 processes occurring at distinct spatial scales, such as the spatial variability in local community composition  
 486 resulting from different sequences of extinction and colonization events, can also contribute to this variation  
 487 (Galiana *et al.* 2018). Next, interaction accumulation curves describe the scaling of the number of observed  
 488 interactions with sampling effort (Jordano 2016). Sampling effort, which may correspond to the duration of the  
 489 sampling period used to construct the network, can impact connectance (Bersier *et al.* 1999) and various  
 490 measures of network structure (Banašek-Richter *et al.* 2004; McLeod *et al.* 2021). Apart from sampling effort,  
 491 the temporal scaling of interactions also elucidates how network structure changes with the temporal resolution  
 492 of the network, acknowledging that distinct interactions take place over time, ranging from short-term  
 493 fluctuations of interactions to long-term trends. As local networks of probabilistic interactions may explicitly  
 494 account for the spatiotemporal variability of interactions (as shown in Box 1), they offer a distinct approach to  
 495 investigating the scaling of network structure with space and time, in contrast to networks of binary and  
 496 quantitative interactions, by making the stochasticity of interactions the focal point of the modeling process.

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

510 On the other hand, local interactions scale both spatially and temporally, given that they have more opportunities  
511 to be realized and observed in larger areas and longer durations. This is attributed to factors such as a higher  
512 number of individuals, greater trait variations, and increased opportunities for encounters, as highlighted by  
513 McLeod *et al.* (2020). For example, if a local network of probabilistic interactions  $L_1$  with an area  $A_1$  is derived  
514 from a larger network  $L_0$  with an area  $A_0$ , and  $A_1$  is entirely nested within  $A_0$ , interaction probabilities should  
515 be lower in the smaller network, 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,  
516 interaction probabilities could be higher in the smaller area, contingent upon their environmental and biological  
517 conditions. Likewise, interaction probabilities are expected to be lower in networks with shorter durations when  
518 time intervals are nested. In Fig. 3, we show how the expected number of local host-parasite interactions scales  
519 with area, represented as an expanding latitudinal window, in comparison with regional interactions. Even  
520 though we employed local probabilities of interactions equal to regional interactions for the purpose of  
521 comparison (i.e., using  $P(L_{i \rightarrow j}|M_{i \rightarrow j}) = 1$  here), we notice that the total number of regional interactions scales  
522 more rapidly than local interactions. This is because numerous regional interactions involve species that never  
523 co-occur, and as a result, these interactions are not captured in local networks.

524

[Figure 3 about here.]

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

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

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

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

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

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

555 Ideally, our biological interpretation of probabilistic interactions should remain consistent across a network  
556 even if it incorporates heterogeneous levels of organization, such as a network whose nodes represent both  
557 species and trophic species (groups of species sharing similar predators and prey). This is common in  
558 ecological networks where taxonomic resolution is typically low (Hemprich-Bennett *et al.* 2021; Vázquez *et al.*  
559 2022). Interaction probabilities at broader taxonomic scales should be based on probabilities of interactions  
560 between individuals, for both local networks and metawebs. For instance, in local individual-based food webs,  
561 the probability that two individuals interact reflects our degree of belief that one individual will consume the  
562 other. Similarly, in local species-based food webs, the probability that two species interact represents our degree  
563 of belief that at least one individual from the predator species consumes at least another individual from the  
564 prey species. Furthermore, in local clade-based food webs, the probability that two clades interact represents  
565 our degree of belief that at least two species from these clades interact with each other or, equivalently, that at  
566 least two individuals from these clades interact with each other. Fundamentally, the taxonomic scaling of  
567 interactions involves aggregating interactions between individuals into larger groups, which may exhibit varying  
568 degrees of homogeneity based on the organisms and the study system. In that regard, taxonomic scaling is  
569 analogous to the spatial and temporal scaling of interactions, as they all represent different ways to aggregate  
570 individuals into broader groups (either spatially, temporally, or taxonomically).

## 571 Application 4: Sampling for binary interaction networks

572 The prediction of binary interactions through Bernoulli trials is an important application of probabilistic  
573 interaction networks. This approach proves beneficial for analyzing the structural characteristics of probabilistic  
574 interaction networks, particularly in the absence of specific analytical formulas for measures describing the  
575 structure of the complete network (Poisot *et al.* 2016). By performing independent Bernoulli trials for each  
576 interaction, a network of binary interactions may be generated. A probability distribution of network properties  
577 can be obtained by measuring network structure across multiple randomly generated networks (Poisot *et al.*  
578 2016). This method enables the representation of the variability of network structure, albeit with possible biases  
579 when connectance is low (Chagnon 2015; Poisot & Gravel 2014). Employing this strategy to generate binary  
580 interaction networks under a null model facilitates null hypothesis significance testing, wherein the observed

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

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

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

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

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

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

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

In Fig. 4, we compare the average connectance of binary interaction networks resulting from these two sampling techniques. We drew regional and local interactions from our host-parasite networks of probabilistic

605 interactions, generating a number of binary interaction network realizations for each site in the dataset. These  
606 two sampling approaches yield different outcomes, particularly for lower values of  $P(L_{i \rightarrow j} | M_{i \rightarrow j})$ , which  
607 represent instances where regional interactions do not consistently manifest locally. Small discrepancies  
608 between these techniques are also apparent when we equate the probability of local interaction to the probability  
609 of regional interaction (i.e., when using  $P(L_{i \rightarrow j} | M_{i \rightarrow j}) = 1.0$  in Eq. 17, Fig. 4a), especially when the number of  
610 simulations of binary interaction networks for each location is low ( $n = 1$ ). As anticipated, we observe that  
611 sampling binary interactions from the metaweb tends to overestimate connectance on average compared to  
612 sampling them from local networks (Fig. 4a-c). Furthermore, we observe an increase in the variability of  
613 connectance when employing a single simulation (Fig. 4, gray markers), which in our opinion is a more tangible  
614 representation of the process leading to the realization of local interactions in nature.

615 [Figure 4 about here.]

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

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

622 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  
623 sampling binary interactions from local networks, it is crucial to sample at the same spatial scale for which  
624 probabilities were estimated. Otherwise, interaction probabilities must be adjusted to align with the intended  
625 cell size, preventing systematic biases in predictions.

## 626 Application 5: Prediction of local networks from metawebs

627 Metawebs serve as a valuable source of ecological information for predicting local networks across time and  
628 space. Local networks of binary interactions can be reconstructed by selecting a subset of taxa and interactions  
629 from the metaweb (Dunne 2006). This implies that metawebs consistently contain more interactions than their

630 corresponding local networks, even though their connectance is usually much smaller than the ones of local  
 631 networks (Gravel *et al.* 2011). Determining the list of taxa to select can be achieved empirically or through  
 632 numerical methods like range maps or species distribution models. As species composition is arguably less  
 633 difficult to sample or predict than pairwise interactions, the primary challenge lies in deciding which  
 634 interactions to select from the metaweb, a task that may necessitate advanced statistical models and ecological  
 635 expertise. Inferring the structure of local networks from the metaweb before predicting specific local pairwise  
 636 interactions could hold promise (Strydom *et al.* 2021), considering that the structure of local networks is  
 637 constrained by the metaweb (Saravia *et al.* 2022).

638 Building local networks of probabilistic interactions from a metaweb of probabilistic interactions involves a  
 639 reduction in the value of pairwise interaction probabilities. This decrease is attributed to the prerequisite that  
 640 two taxa must initially possess the capacity to interact before engaging in local interactions (Eq. 17). Therefore,  
 641 inferring local networks from their metaweb while maintaining identical interaction probability values would  
 642 introduce systematic biases into the predictions. In such cases, these networks would essentially represent  
 643 smaller-scale metawebs of potential interactions, possibly leading to misinterpretations by being perceived as  
 644 local interactions. As proposed by McLeod *et al.* (2021), although metawebs do not capture the spatiotemporal  
 645 variability of interactions, they establish an upper limit for local interactions (similarly for metawebs of  
 646 probabilistic interactions, Strydom *et al.* 2023). In other words, the probability of two taxa interacting at a  
 647 specific location and time is consistently lower or equal to the probability of their regional interaction:

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

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

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

652 Estimating more precisely the probability  $P(L_{i \rightarrow j} | M_{i \rightarrow j})$  that two taxa interact locally if they can potentially  
 653 interact allows for improved predictions of local networks from the metaweb of probabilistic interactions. This

654 task is challenging due to the variability of this probability across space and time, as well as its variability  
655 across pairwise interactions within a network. Using simple models of  $P(L_{i \rightarrow j} | M_{i \rightarrow j})$ , as demonstrated in our  
656 case studies, represents an initial step toward the overarching objective of reconstructing local networks from  
657 metawebs.

## 658 Conclusion

659 In this contribution, we underline the importance of network metadata for adequately interpreting and  
660 manipulating probabilistic interaction data. The mathematical representation of probabilities and their statistical  
661 properties depend on the type of interactions (local or regional) and the conditions under which these  
662 interactions were evaluated. We showed that local networks and metawebs of probabilistic interactions differ in  
663 their relationship to spatial and temporal scales, with regional interactions remaining consistent across scales.

664 In contrast with metawebs, local interactions are measured in a specific context (e.g., in a given area, time, and  
665 biological and environmental conditions) and depend on taxa co-occurrence. These differences bring to light  
666 the need to use probabilistic data with caution, for instance when generating network realizations of binary  
667 interactions across space and predicting local networks from metawebs. Clear metadata describing the type of  
668 interaction and the variables used in their estimation are required to ensure adequate data manipulation. Sound  
669 data practices and foundations for probabilistic thinking in network ecology could facilitate reliable assessments  
670 of the spatiotemporal variability and uncertainty of biotic interactions.

671 Estimating local interaction probabilities independently for each taxa pair and assembling them into a network  
672 of probabilistic interactions comes with limitations. Predicting local networks of binary interactions based on  
673 these interaction probabilities assumes independence among interactions, a condition seldom respected in  
674 practice (Golubski & Abrams 2011). Relaxing this assumption is the next logical step in the stochastic  
675 representation of interactions. A more accurate representation of the stochasticity of ecological networks  
676 involves creating *probabilistic networks* ( $P(L|...)$  and  $P(M|...)$ ), rather than networks of *probabilistic*  
677 *interactions* ( $P(L_{i \rightarrow j}|...)$  and  $P(M_{i \rightarrow j}|...)$ ). Probabilistic networks describe the probability that a particular  
678 network of binary (or quantitative) interactions (its whole adjacency matrix) is realized. For example, Young *et*  
679 *al.* (2021) used a Bayesian approach to estimate the probability of different plant-pollinator network structures  
680 derived from imperfect observational data. A probability distribution of ecological networks may also be  
681 derived using the principle of maximum entropy given some structural constrained (e.g., Cimini *et al.* 2019;

682 Park & Newman 2004). Regardless of the method employed, generating probabilistic local networks, bypassing  
683 the need to independently estimate local interaction probabilities, could lead to more accurate predictions of  
684 local networks. Furthermore, probabilistic networks could serve as an alternative to null hypothesis significance  
685 testing when comparing the structure of a local network to some random expectations or, as done in Pellissier *et*  
686 *al.* (2018), to the metaweb. These random expectations are typically derived by performing a series of Bernoulli  
687 trials on probabilistic interactions, assuming independence, to generate a distribution of networks of binary  
688 interactions and then calculate their structure (Poisot *et al.* 2016). One could for instance compare the  
689 likelihood of an observed network to the one of the most likely network structure according to the probabilistic  
690 network distribution, thereby directly obtaining a measure of discrepancy of the empirical network. Generating  
691 probabilistic ecological networks represents a tangible challenge, one that, in the coming years, promises to  
692 unlock doors to more advanced and adequate analyses of ecological networks.

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

711 **Acknowledgment**

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

719 **References**

- 720 Abrego, N., Roslin, T., Huotari, T., Ji, Y., Schmidt, N.M., Wang, J., *et al.* (2021). [Accounting for species](#)  
721 [interactions is necessary for predicting how arctic arthropod communities respond to climate change.](#)  
722 *Ecography*, 44, 885–896.
- 723 Angilletta, M.J., Jr., Steury, T.D. & Sears, M.W. (2004). [Temperature, Growth Rate, and Body Size in](#)  
724 [Ectotherms: Fitting Pieces of a Life-History Puzzle](#). *Integrative and Comparative Biology*, 44, 498–509.
- 725 Araujo, M.B., Rozenfeld, A., Rahbek, C. & Marquet, P.A. (2011). [Using species co-occurrence networks to](#)  
726 [assess the impacts of climate change](#). *Ecography*, 34, 897–908.
- 727 Banašek-Richter, C., Cattin, M.-F. & Bersier, L.-F. (2004). [Sampling effects and the robustness of quantitative](#)  
728 [and qualitative food-web descriptors](#). *Journal of Theoretical Biology*, 226, 23–32.
- 729 Bartomeus, I., Gravel, D., Tylianakis, J.M., Aizen, M.A., Dickie, I.A. & Bernard-Verdier, M. (2016). [A](#)  
730 [common framework for identifying linkage rules across different types of interactions](#). *Functional Ecology*,  
731 30, 1894–1903.
- 732 Bascompte, J., Jordano, P., Melián, C.J. & Olesen, J.M. (2003). [The nested assembly of plant–animal](#)  
733 [mutualistic networks](#). *Proceedings of the National Academy of Sciences*, 100, 9383–9387.
- 734 Berg, M.P. & Ellers, J. (2010). [Trait plasticity in species interactions: A driving force of community dynamics](#).  
735 *Evolutionary Ecology*, 24, 617–629.

- 736 Berlow, E.L., Neutel, A.-M., Cohen, J.E., De Ruiter, P.C., Ebenman, B., Emmerson, M., *et al.* (2004).  
737 **Interaction strengths in food webs: Issues and opportunities.** *Journal of Animal Ecology*, 73, 585–598.
- 738 Bersier, L.-F., Dixon, P. & Sugihara, G. (1999). **Scale-Invariant or Scale-Dependent Behavior of the Link**  
739 **Density Property in Food Webs: A Matter of Sampling Effort?** *The American Naturalist*, 153, 676–682.
- 740 Bohan, D.A., Vacher, C., Tamaddoni-Nezhad, A., Raybould, A., Dumbrell, A.J. & Woodward, G. (2017).  
741 **Next-Generation Global Biomonitoring: Large-scale, Automated Reconstruction of Ecological Networks.**  
742 *Trends in Ecology & Evolution*, 32, 477–487.
- 743 Bolnick, D.I., Amarasekare, P., Araújo, M.S., Bürger, R., Levine, J.M., Novak, M., *et al.* (2011). **Why**  
744 **intraspecific trait variation matters in community ecology.** *Trends in Ecology & Evolution*, 26, 183–192.
- 745 Borrett, S.R. & Scharler, U.M. (2019). **Walk partitions of flow in Ecological Network Analysis: Review and**  
746 **synthesis of methods and indicators.** *Ecological Indicators*, 106, 105451.
- 747 Boulangeat, I., Gravel, D. & Thuiller, W. (2012). **Accounting for dispersal and biotic interactions to disentangle**  
748 **the drivers of species distributions and their abundances.** *Ecology Letters*, 15, 584–593.
- 749 Brimacombe, C., Bodner, K., Michalska-Smith, M., Poisot, T. & Fortin, M.-J. (2023). **Shortcomings of reusing**  
750 **species interaction networks created by different sets of researchers.** *PLOS Biology*, 21, e3002068.
- 751 Broom, M. & Ruxton, G.D. (2005). **You can run—or you can hide: Optimal strategies for cryptic prey against**  
752 **pursuit predators.** *Behavioral Ecology*, 16, 534–540.
- 753 Brose, U., Ostling, A., Harrison, K. & Martinez, N.D. (2004). **Unified spatial scaling of species and their**  
754 **trophic interactions.** *Nature*, 428, 167–171.
- 755 Burnham, K.P. & Anderson, D.R. (2004). **Multimodel Inference: Understanding AIC and BIC in Model**  
756 **Selection.** *Sociological Methods & Research*, 33, 261–304.
- 757 Canard, E.F., Mouquet, N., Mouillot, D., Stanko, M., Miklisova, D. & Gravel, D. (2014). **Empirical Evaluation**  
758 **of Neutral Interactions in Host-Parasite Networks.** *The American Naturalist*, 183, 468–479.
- 759 Canard, E., Mouquet, N., Marescot, L., Gaston, K.J., Gravel, D. & Mouillot, D. (2012). **Emergence of**  
760 **Structural Patterns in Neutral Trophic Networks.** *PLOS ONE*, 7, e38295.
- 761 Carlson, C.J., Farrell, M.J., Grange, Z., Han, B.A., Mollentze, N., Phelan, A.L., *et al.* (2021). **The future of**  
762 **zoonotic risk prediction.** *Philosophical Transactions of the Royal Society B: Biological Sciences*, 376,  
763 20200358.

- 764 Castoe, T.A., Spencer, C.L. & Parkinson, C.L. (2007). Phylogeographic structure and historical demography of  
765 the western diamondback rattlesnake (*Crotalus Atrox*): A perspective on North American desert  
766 biogeography. *Molecular Phylogenetics and Evolution*, 42, 193–212.
- 767 Catchen, M.D., Poisot, T., Pollock, L.J. & Gonzalez, A. (2023). The missing link: Discerning true from false  
768 negatives when sampling species interaction networks.
- 769 Cazelles, K., Araujo, M.B., Mouquet, N. & Gravel, D. (2016). A theory for species co-occurrence in interaction  
770 networks. *Theoretical Ecology*, 9, 39–48.
- 771 Chagnon, P.-L. (2015). Characterizing topology of ecological networks along gradients: The limits of metrics'  
772 standardization. *Ecological Complexity*, 22, 36–39.
- 773 Choh, Y., Ignacio, M., Sabelis, M.W. & Janssen, A. (2012). Predator-prey role reversals, juvenile experience  
774 and adult antipredator behaviour. *Scientific Reports*, 2, 728.
- 775 Cimini, G., Squartini, T., Saracco, F., Garlaschelli, D., Gabrielli, A. & Caldarelli, G. (2019). The statistical  
776 physics of real-world networks. *Nature Reviews Physics*, 1, 58–71.
- 777 Cirtwill, A.R., Eklöf, A., Roslin, T., Wootton, K. & Gravel, D. (2019). A quantitative framework for  
778 investigating the reliability of empirical network construction. *Methods in Ecology and Evolution*, 10,  
779 902–911.
- 780 Dallas, T., Park, A.W. & Drake, J.M. (2017). Predicting cryptic links in host-parasite networks. *PLOS  
781 Computational Biology*, 13, e1005557.
- 782 Dunne, J.A. (2006). The Network Structure of Food Webs. In: *Ecological networks: Linking structure and  
783 dynamics* (eds. Dunne, J.A. & Pascual, M.). Oxford University Press, pp. 27–86.
- 784 Elmasri, M., Farrell, M.J., Davies, T.J. & Stephens, D.A. (2020). A hierarchical Bayesian model for predicting  
785 ecological interactions using scaled evolutionary relationships. *The Annals of Applied Statistics*, 14,  
786 221–240.
- 787 Elton, C.S. (2001). *Animal Ecology*. University of Chicago Press, Chicago, IL.
- 788 Emmerson, M.C. & Raffaelli, D. (2004). Predator–prey body size, interaction strength and the stability of a real  
789 food web. *Journal of Animal Ecology*, 73, 399–409.
- 790 Fedorov, V.B., Goropashnaya, A.V., Boeskorov, G.G. & Cook, J.A. (2008). Comparative phylogeography and  
791 demographic history of the wood lemming (*Myopus schisticolor*): Implications for late Quaternary history

- 792 of the taiga species in Eurasia. *Molecular Ecology*, 17, 598–610.
- 793 Fortuna, M.A. & Bascompte, J. (2006). Habitat loss and the structure of plant–animal mutualistic networks.  
794 *Ecology Letters*, 9, 281–286.
- 795 Fu, X., Seo, E., Clarke, J. & Hutchinson, R.A. (2021). Link Prediction Under Imperfect Detection:  
796 Collaborative Filtering for Ecological Networks. *IEEE Transactions on Knowledge and Data Engineering*,  
797 33, 3117–3128.
- 798 Galiana, N., Lurgi, M., Claramunt-López, B., Fortin, M.-J., Leroux, S., Cazelles, K., et al. (2018). The spatial  
799 scaling of species interaction networks. *Nature Ecology & Evolution*, 2, 782–790.
- 800 Gilman, S.E., Urban, M.C., Tewksbury, J., Gilchrist, G.W. & Holt, R.D. (2010). A framework for community  
801 interactions under climate change. *Trends in Ecology & Evolution*, 25, 325–331.
- 802 Godsoe, W., Murray, R. & Iritani, R. (2022). Species interactions and diversity: A unified framework using Hill  
803 numbers. *Oikos*, n/a, e09282.
- 804 Golubski, A.J. & Abrams, P.A. (2011). Modifying modifiers: What happens when interspecific interactions  
805 interact? *Journal of Animal Ecology*, 80, 1097–1108.
- 806 Gomez, J.M., Verdu, M. & Perfectti, F. (2010). Ecological interactions are evolutionarily conserved across the  
807 entire tree of life. *Nature*, 465, 918–U6.
- 808 Gonzalez, A. & Londoño, M.C. (2022). Monitor biodiversity for action. *Science*, 378, 1147–1147.
- 809 Gonzalez-Varo, J.P. & Traveset, A. (2016). The Labile Limits of Forbidden Interactions. *Trends in Ecology &*  
810 *Evolution*, 31, 700–710.
- 811 Gravel, D., Albouy, C. & Thuiller, W. (2016). The meaning of functional trait composition of food webs for  
812 ecosystem functioning. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 371,  
813 20150268.
- 814 Gravel, D., Baiser, B., Dunne, J.A., Kopelke, J.-P., Martinez, N.D., Nyman, T., et al. (2019). Bringing Elton  
815 and Grinnell together: A quantitative framework to represent the biogeography of ecological interaction  
816 networks. *Ecography*, 42, 401–415.
- 817 Gravel, D., Massol, F., Canard, E., Mouillot, D. & Mouquet, N. (2011). Trophic theory of island biogeography.  
818 *Ecology Letters*, 14, 1010–1016.

- 819 Gravel, D., Poisot, T., Albouy, C., Velez, L. & Mouillot, D. (2013). [Inferring food web structure from](#)  
820 [predator–prey body size relationships](#). *Methods in Ecology and Evolution*, 4, 1083–1090.
- 821 Guimarães, P.R. (2020). [The Structure of Ecological Networks Across Levels of Organization](#). *Annual Review*  
822 *of Ecology, Evolution, and Systematics*, 51, 433–460.
- 823 Guimerà, R. & Sales-Pardo, M. (2009). [Missing and spurious interactions and the reconstruction of complex](#)  
824 [networks](#). *Proceedings of the National Academy of Sciences*, 106, 22073–22078.
- 825 Hemprich-Bennett, D.R., Oliveira, H.F.M., Le Comber, S.C., Rossiter, S.J. & Clare, E.L. (2021). [Assessing the](#)  
826 [impact of taxon resolution on network structure](#). *Ecology*, 102, e03256.
- 827 Herrera, C.M. (1989). [Pollinator abundance, morphology, and flower visitation rate: Analysis of the “quantity”](#)  
828 [component in a plant-pollinator system](#). *Oecologia*, 80, 241–248.
- 829 Hortal, J., de Bello, F., Diniz-Filho, J.A.F., Lewinsohn, T.M., Lobo, J.M. & Ladle, R.J. (2015). [Seven Shortfalls](#)  
830 [that Beset Large-Scale Knowledge of Biodiversity](#). *Annual Review of Ecology, Evolution, and Systematics*,  
831 46, 523–549.
- 832 Jordano, P. (1987). [Patterns of Mutualistic Interactions in Pollination and Seed Dispersal: Connectance,](#)  
833 [Dependence Asymmetries, and Coevolution](#). *The American Naturalist*, 129, 657–677.
- 834 Jordano, P. (2016). [Sampling networks of ecological interactions](#). *Functional Ecology*, 30, 1883–1893.
- 835 Jordano, P., Bascompte, J. & Olesen, J.M. (2003). [Invariant properties in coevolutionary networks of](#)  
836 [plant–animal interactions](#). *Ecology Letters*, 6, 69–81.
- 837 Kausrud, K.L., Mysterud, A., Steen, H., Vik, J.O., Østbye, E., Cazelles, B., et al. (2008). [Linking climate](#)  
838 [change to lemming cycles](#). *Nature*, 456, 93–97.
- 839 Kéfi, S., Berlow, E.L., Wieters, E.A., Joppa, L.N., Wood, S.A., Brose, U., et al. (2015). [Network structure](#)  
840 [beyond food webs: Mapping non-trophic and trophic interactions on Chilean rocky shores](#). *Ecology*, 96,  
841 291–303.
- 842 Kéfi, S., Berlow, E.L., Wieters, E.A., Navarrete, S.A., Petchey, O.L., Wood, S.A., et al. (2012). [More than a](#)  
843 [meal... integrating non-feeding interactions into food webs](#). *Ecology Letters*, 15, 291–300.
- 844 Kéfi, S., Miele, V., Wieters, E.A., Navarrete, S.A. & Berlow, E.L. (2016). [How Structured Is the Entangled](#)  
845 [Bank? The Surprisingly Simple Organization of Multiplex Ecological Networks Leads to Increased](#)  
846 [Persistence and Resilience](#). *PLOS Biology*, 14, e1002527.

- 847 Kissner, K.J., Forbes, M.R. & Secoy, D.M. (1997). [Rattling Behavior of Prairie Rattlesnakes \(\*Crotalus viridis\* viridis, Viperidae\) in Relation to Sex, Reproductive Status, Body Size, and Body Temperature](#). *Ethology*, 103, 1042–1050.
- 848
- 849
- 850 Klecka, J. & Boukal, D.S. (2014). [The effect of habitat structure on prey mortality depends on predator and prey microhabitat use](#). *Oecologia*, 176, 183–191.
- 851
- 852 Kopelke, J.-P., Nyman, T., Cazelles, K., Gravel, D., Vissault, S. & Roslin, T. (2017). [Food-web structure of willow-galling sawflies and their natural enemies across Europe](#). *Ecology*, 98, 1730–1730.
- 853
- 854 Kulkarni, M.A., Kweka, E., Nyale, E., Lyatuu, E., Mosha, F.W., Chandramohan, D., *et al.* (2006). [Entomological Evaluation of Malaria Vectors at Different Altitudes in Hai District, Northeastern Tanzania](#). *Journal of Medical Entomology*, 43, 580–588.
- 855
- 856
- 857 Maiorano, L., Montemaggioli, A., Ficetola, G.F., O'Connor, L. & Thuiller, W. (2020). [TETRA-EU 1.0: A species-level trophic metaweb of European tetrapods](#). *Global Ecology and Biogeography*, 29, 1452–1457.
- 858
- 859 McCann, K. (2007). Protecting biostructure. *Nature*, 446, 29–29.
- 860 McCann, K.S. (2011). Food webs (MPB-50). In: *Food Webs (MPB-50)*. Princeton University Press.
- 861 McLeod, A., Leroux, S.J., Gravel, D., Chu, C., Cirtwill, A.R., Fortin, M.-J., *et al.* (2021). [Sampling and asymptotic network properties of spatial multi-trophic networks](#). *Oikos*, n/a.
- 862
- 863 McLeod, A.M., Leroux, S.J. & Chu, C. (2020). [Effects of species traits, motif profiles, and environment on spatial variation in multi-trophic antagonistic networks](#). *Ecosphere*, 11, e03018.
- 864
- 865 Melián, C.J., Vilas, C., Baldó, F., González-Ortegón, E., Drake, P. & Williams, R.J. (2011). [Eco-evolutionary Dynamics of Individual-Based Food Webs](#). In: *Advances in Ecological Research*, The Role of Body Size in Multispecies Systems (ed. Belgrano, A.). Academic Press, pp. 225–268.
- 866
- 867
- 868 Morales-Castilla, I., Matias, M.G., Gravel, D. & Araújo, M.B. (2015). [Inferring biotic interactions from proxies](#). *Trends in Ecology & Evolution*, 30, 347–356.
- 869
- 870 Ohlmann, M., Miele, V., Dray, S., Chalmandrier, L., O'Connor, L. & Thuiller, W. (2019). [Diversity indices for ecological networks: A unifying framework using Hill numbers](#). *Ecology Letters*, 22, 737–747.
- 871
- 872 Olesen, J.M., Bascompte, J., Dupont, Y.L., Elberling, H., Rasmussen, C. & Jordano, P. (2010). [Missing and forbidden links in mutualistic networks](#). *Proceedings of the Royal Society B: Biological Sciences*, 278, 725–732.
- 873
- 874

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

- 901 Pringle, R.M. & Hutchinson, M.C. (2020). [Resolving Food-Web Structure](#). *Annual Review of Ecology, Evolution, and Systematics*, 51, 55–80.
- 903 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.
- 905 Pulliam, H.R. (1974). [On the Theory of Optimal Diets](#). *The American Naturalist*, 108, 59–74.
- 906 Rooney, N. & McCann, K.S. (2012). [Integrating food web diversity, structure and stability](#). *Trends in Ecology & Evolution*, 27, 40–46.
- 908 Salim, J.A., Saraiva, A.M., Zermoglio, P.F., Agostini, K., Wolowski, M., Drucker, D.P., *et al.* (2022). [Data standardization of plant–pollinator interactions](#). *GigaScience*, 11, giac043.
- 910 Sanders, D. & van Veen, F.J.F. (2012). [Indirect commensalism promotes persistence of secondary consumer species](#). *Biology Letters*, 8, 960–963.
- 912 Saravia, L.A., Marina, T.I., Kristensen, N.P., De Troch, M. & Momo, F.R. (2022). [Ecological network assembly: How the regional metaweb influences local food webs](#). *Journal of Animal Ecology*, 91, 630–642.
- 914 Schmeller, D.S., Julliard, R., Bellingham, P.J., Böhm, M., Brummitt, N., Chiarucci, A., *et al.* (2015). [Towards a global terrestrial species monitoring program](#). *Journal for Nature Conservation*, 25, 51–57.
- 916 Shaw, J.O., Dunhill, A.M., Beckerman, A.P., Dunne, J.A. & Hull, P.M. (2024). [A framework for reconstructing ancient food webs using functional trait data](#).
- 918 Singer, M.C. & McBride, C.S. (2012). [Geographic mosaics of species' association: A definition and an example driven by plant–insect phenological synchrony](#). *Ecology*, 93, 2658–2673.
- 920 Staniczenko, P.P.A., Sivasubramaniam, P., Suttle, K.B. & Pearson, R.G. (2017). [Linking macroecology and community ecology: Refining predictions of species distributions using biotic interaction networks](#). *Ecology Letters*, 20, 693–707.
- 923 Stouffer, D.B., Rezende, E.L. & Amaral, L.A.N. (2011). [The role of body mass in diet contiguity and food-web structure](#). *Journal of Animal Ecology*, 80, 632–639.
- 925 Strydom, T., Bouskila, S., Banville, F., Barros, C., Caron, D., Farrell, M.J., *et al.* (2022). [Food web reconstruction through phylogenetic transfer of low-rank network representation](#). *Methods in Ecology and Evolution*, 13.

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

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

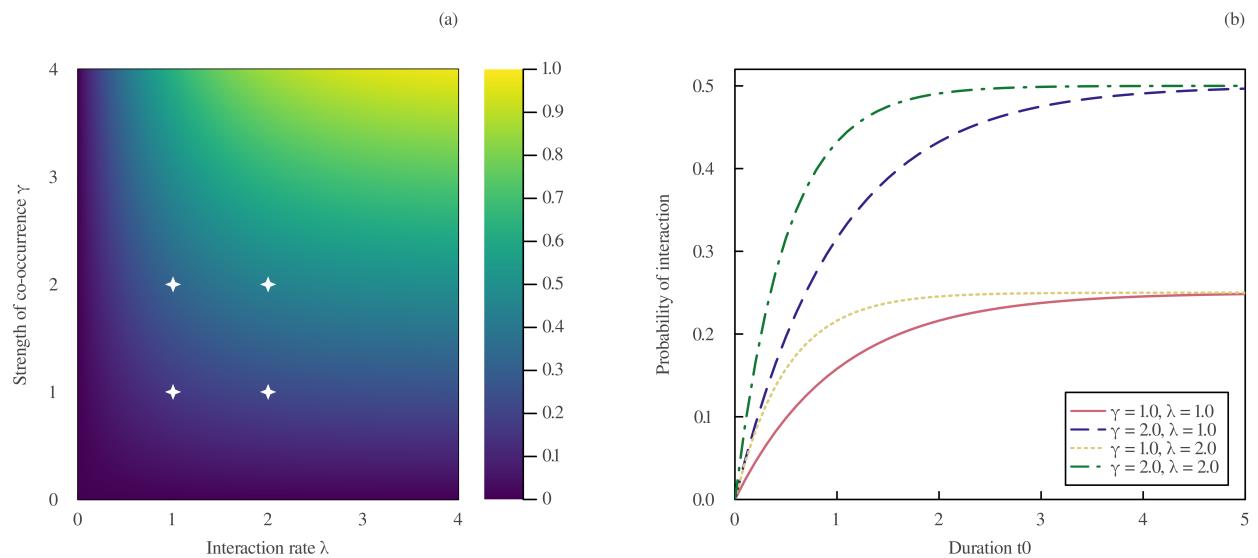
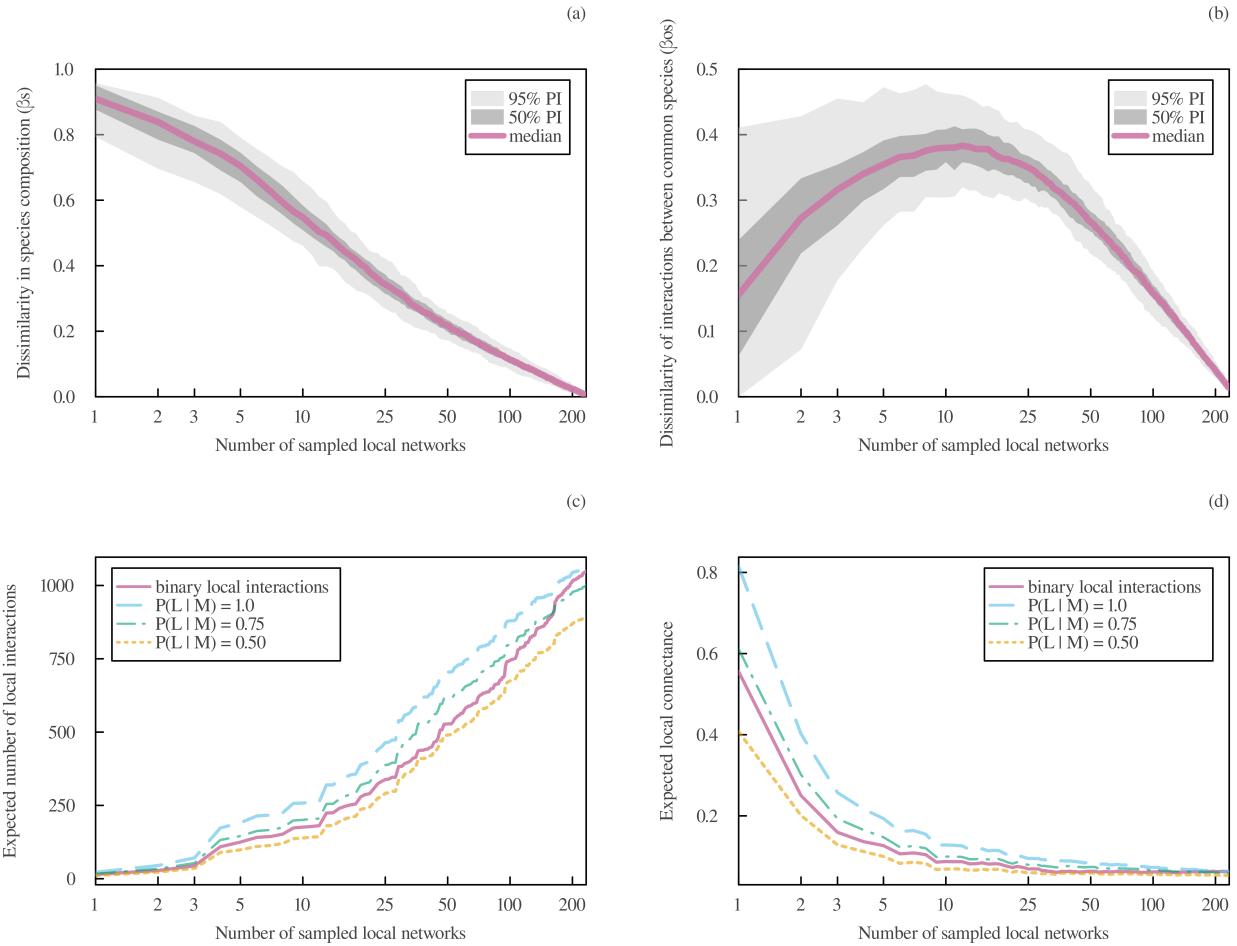
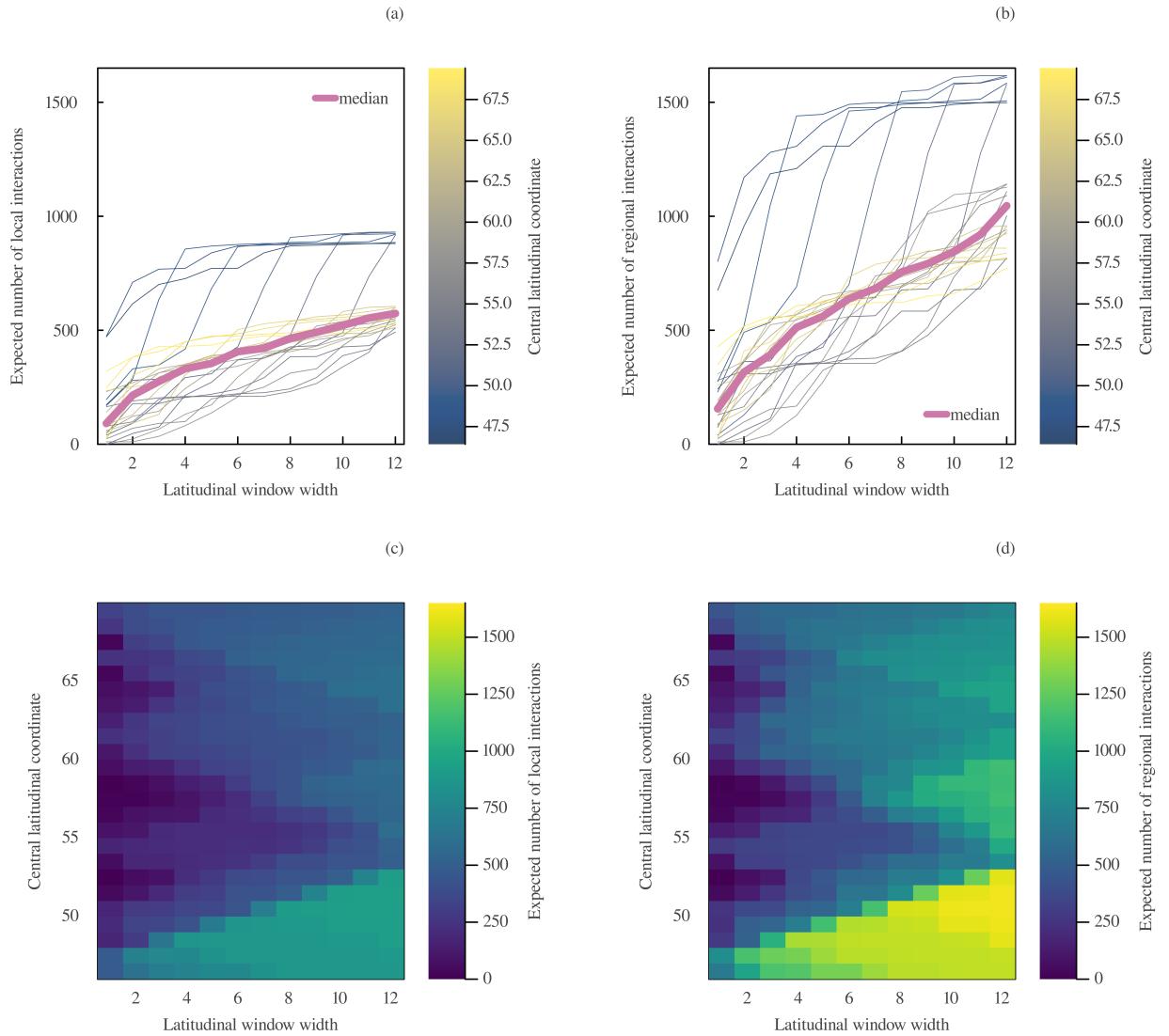


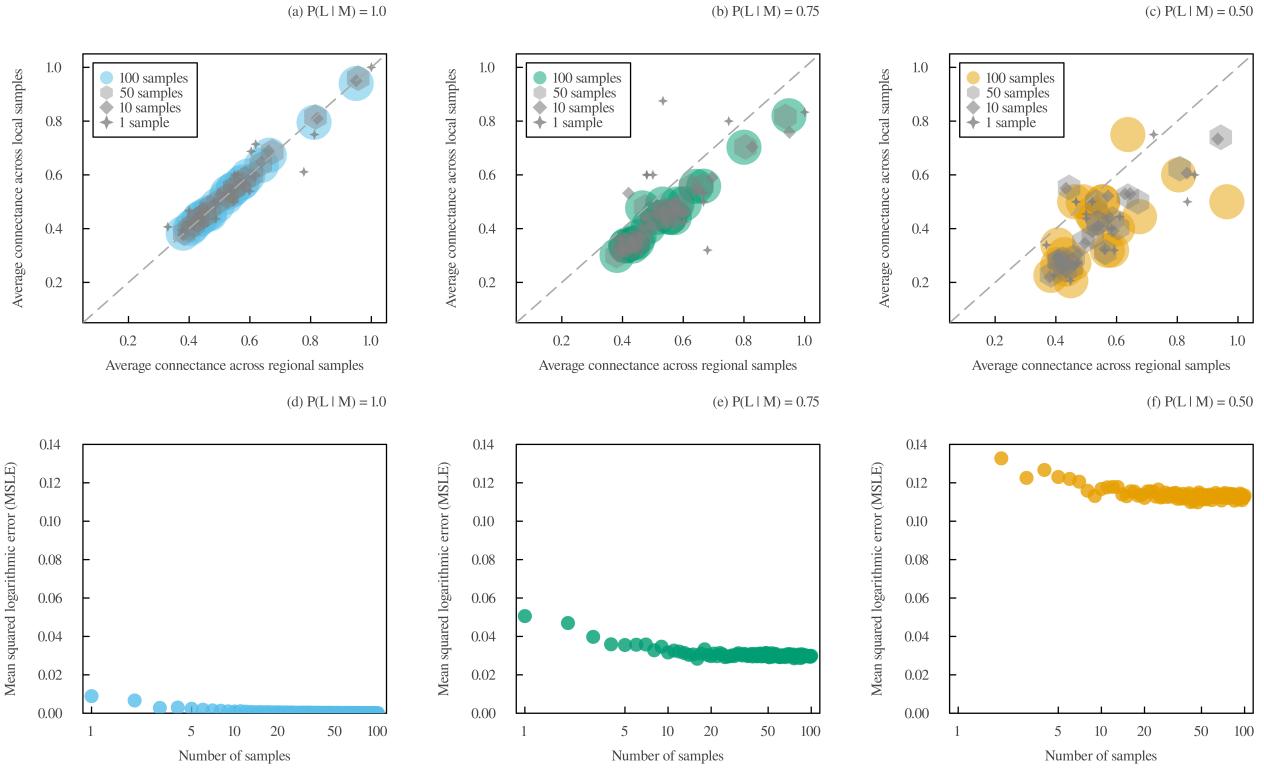
Figure 1: **Parameters of the spatiotemporally explicit model of interactions.** (a) Probability of local interaction given by the process model (Eq. 13) under different values of  $\lambda_k$  and  $\gamma$ , with  $t_0 = 1$ . The parameter values used in the right panel are denoted by the white stars. (b) Scaling of the probability of interaction with the duration parameter  $t_0$  in Eq. 13, for different values of  $\lambda_k$  and  $\gamma$ . In both panels, the marginal probabilities of occurrence  $P(X_{i,k})$  and  $P(X_{j,k})$  are set to a constant value of 0.5.



**Figure 2: Network accumulation curves.** (a) Dissimilarity of interactions between common species and (b) dissimilarity in species composition between aggregated local networks and the metaweb of binary host-parasite interactions. Aggregated local networks were obtained by sequentially and randomly selecting a number of local networks of binary interactions and aggregating both their species and interactions. In both panels, the colored line represents the median dissimilarity across simulations and the grey areas cover the 50% and 95% percentile intervals. (c) Scaling of the number of links and (d) scaling of connectance with the number of sampled binary and probabilistic local interaction networks. For a better comparison with binary interactions, local networks of probabilistic interactions were derived from the metaweb of probabilistic interactions with a false positive and false negative rate of zero. A specific value of  $P(L_{i \rightarrow j}|M_{i \rightarrow j})$  (the local probability of interaction among potentially interacting species, indices omitted in the figure for simplicity) was used for all local networks within a particular curve. Aggregated local networks of probabilistic interactions were obtained by sequentially and randomly selecting a number of local networks and aggregating both their species and interactions (with the value of  $P(L_{i \rightarrow j}|M_{i \rightarrow j})$  adjusting according to Eq. 18).



**Figure 3: Spatial scaling of interactions.** Expected number of host-parasite interactions in a network aggregating all (a) local and (b) regional probabilistic interactions within a latitudinal window of a given length. Every dashed curve corresponds to a different expanding window characterized by a different central latitude, with the colored solid line representing the median number of interactions. Heatmaps of the expected number of (c) local and (d) regional interactions found in windows of specified length and central latitudes. Probabilities of regional interactions were obtained with a false positive rate of 5% and a false negative rate of 10%. Local probabilistic interactions were derived from probabilistic regional interactions by setting the value of  $P(L_{i \rightarrow j} | M_{i \rightarrow j})$  (the local probability of interaction among potentially interacting species) to 1, ensuring a conservative comparison between aggregated local networks and metawebs. Aggregated local networks were obtained by aggregating both the species and interactions found within a particular latitudinal window, with the values of  $P(L_{i \rightarrow j} | M_{i \rightarrow j})$  remaining at their maximum value of 1 following Eq. 18.



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