

The ecological interpretation of probabilistic networks

Francis Banville^{1,2,3} Tanya Strydom^{1,3} Chris Brimacombe⁴ Michael Catchen^{3,5} Dominique Gravel^{2,3}

Gracielle Higino² Kari Norman¹ Timothée Poisot^{1,3}

¹ Université de Montréal ² Université de Sherbrooke ³ Quebec Centre for Biodiversity Science ⁴ University of Toronto ⁵ McGill University

Correspondance to:

Francis Banville — francis.banville@umontreal.ca

The stochastic nature of ecological interactions has led biologists to adopt a probabilistic view of ecological networks. Representing species interactions probabilistically (how likely are they occurring?) as opposed to deterministically (are they occurring?) allows a better assessment of their spatiotemporal variability and accounts for inherent uncertainties in both observations and predictions. However, despite this growing interest in probabilistic networks, guidelines for the estimation and documentation of probabilistic interaction data are still lacking. This is concerning given that our biological understanding of probabilistic interactions and their appropriate analyses depend on the often elusive methods and variables used for their estimation. We review how probabilistic interactions are defined at different spatial scales, from local interactions to regional networks (metawebs), with a strong emphasis on trophic interactions. These definitions are based on the distinction between the observation of an interaction at a specific time and space and its biological feasibility. We show that different network representations have different statistical properties when it comes to common ecological applications. Specifically, we argue that local probabilistic networks and metawebs differ in their spatial and temporal scaling of interactions, with potential interactions in metawebs remaining constant across spatial and temporal scales. This is in contrast with the taxonomic scaling of interactions, which does not qualitatively differ between both types of networks. We suggest two approaches to inferring deterministic networks from probabilistic webs that account for these differences and argue that systematic biases arise when directly inferring local networks from subsets of metawebs. To support our arguments, we build a spatiotemporally explicit model of probabilistic interactions and develop different case studies using empirical data on host-parasite interactions in Europe. Overall, our results underscore the importance of clear metadata for

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

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



Last revision: *November 14, 2023*

1 Introduction

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

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

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

By accounting for the uncertainty of interactions, probabilistic networks provide a more realistic portrait of species interactions and network structure (i.e. community-level properties), which is a major driver of the functioning, dynamics, and resilience of ecosystems worldwide (Proulx *et al.* (2005), McCann (2007), McCann (2011), Rooney & McCann (2012)). Moreover, the application and development of computational methods in network ecology, often based on a probabilistic representation of interactions, can alleviate the sampling efforts required to document species interactions (Strydom *et al.* (2021)). For example, statistical models can be used to estimate the uncertainty of pairwise interactions (Cirtwill *et al.* (2019)) and the probability of missing (false negatives) and spurious (false positives) interactions (Guimerà & Sales-Pardo (2009)). Considering the high rate of false negatives in species interaction data due to the difficulty of witnessing rare interactions (Catchen *et al.* (2023)), these models could inform the identification of priority sampling locations of ecological networks where data collection would yield the most valuable information, thereby reducing errors. Optimization models for sampling locations have mostly found applications in biological systems that are not networked, such as identifying priority sampling sites for disease hotspots (Andrade-Pacheco *et al.* (2020)), but there is substantial promise in applying them to probabilistic ecological interactions. Statistical models can also be used to generate network predictions without prior knowledge of pairwise interactions, for instance using body size (Petchey *et al.* (2008), Gravel *et al.* (2013)), phylogeny (Elmasri *et al.* (2020), Strydom *et al.* (2022)), or a combination of niche and neutral processes (Bartomeus *et al.* (2016), Pomeranz *et al.* (2019)) for inference. Topological null models (e.g., Bascompte *et al.* (2003), Fortuna & Bascompte (2006)), which can be used to generate underlying distributions of network measures for null hypothesis significance testing, are other examples of common probabilistic network models. Many measures have been developed to describe the structure (Poisot *et al.* (2016)) and diversity (Ohlmann *et al.* (2019), Godsoe *et al.* (2022)) of probabilistic networks. These models and measures support the use of this approach for the study of a wide range of ecological questions, from making better predictions of species distribution (Cazelles *et al.* (2016)) to forecasting the impact of climate change on ecological networks (Gilman *et al.* (2010)).

The lack of clear guidelines on the use of probabilistic interaction data is worrisome (Brimacombe *et al.*

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

84 **Probabilistic representations of interactions**

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

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

99 Pairwise interactions: the building blocks of ecological networks

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

210 **Metawebs: regional catalogs of interactions**

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

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

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

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

230 Every pair of taxa that have confidently been observed to interact at least once can be given a probability of 1
231 (i.e., $P_M(i \rightarrow j) = 1$) since we know that they *can* interact. This is not the case in local probabilistic networks,
232 in which interaction events usually remain stochastic (i.e., $P_N(i \rightarrow j) < 1$) after empirically observing
233 interactions because of their intrinsic spatiotemporal variability. Similarly, although rare interactions typically
234 have low probabilities in local networks, they can have high probabilities in metawebs when the traits of both
235 taxa are congruent. On the other hand, interactions that were never observed can have low probability values in
236 both metawebs and local networks, going as low as 0 for forbidden links. However, because of observation

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

245 Properties of probabilistic networks

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

253 Host-parasite network data

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

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

266 [Figure 1 about here.]

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

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

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

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

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

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

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

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

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

301 Taxonomic scaling of interactions

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

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

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

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

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

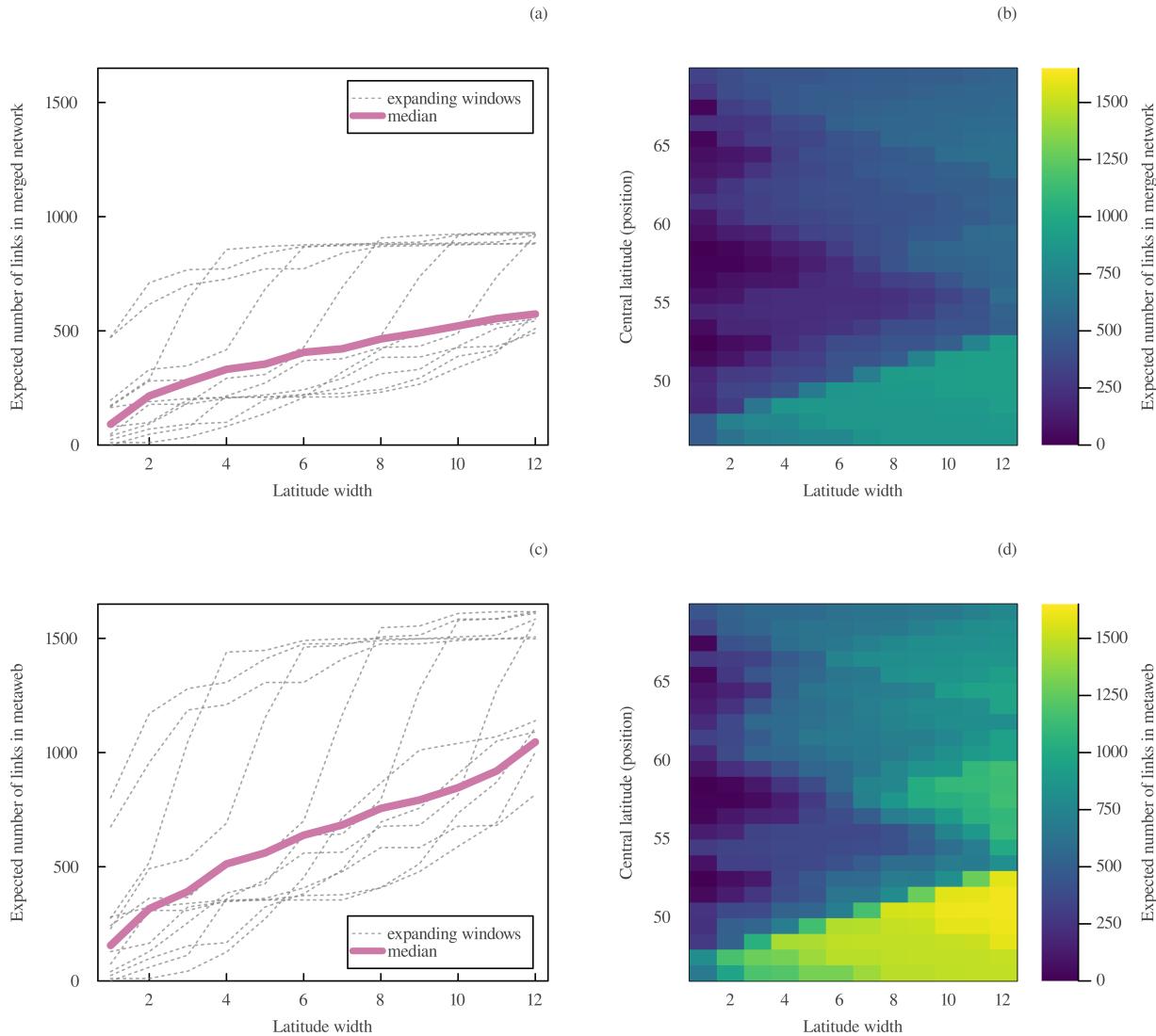
327 Ideally, our biological interpretation of probabilistic interactions should remain consistent across a network
328 even if it incorporates heterogenous levels of organization (e.g., a network whose nodes are composed of both
329 species and trophic species). This is common in ecological networks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

343 **Spatial and temporal scaling of interactions**

344 The study of network-area relationships (NARs) and interaction accumulation curves (IACs) are important
345 realms of research in network ecology. Firstly, NARs describe the scaling of network properties (such as
346 modularity and trophic chain lengths) with spatial scale (e.g., Wood *et al.* (2015), Galiana *et al.* (2018)). The
347 variation of network structure across spatial scales can be the result of the scaling of species richness
348 (species-area relationships, SARs) and the number of interactions (Brose *et al.* (2004)) with the area sampled,
349 but can also be due to ecological processes (e.g., spatial variability in community composition and extinction
350 and colonization events) occurring at different scales (Galiana *et al.* (2018)). Secondly, IACs describe the
351 scaling of the number of interactions observed with sampling effort (Jordano (2016)), which can impact many
352 measures of network structure such as connectance (McLeod *et al.* (2021b)). Beyond sampling effort, the
353 temporal scaling of interactions describes how network structure changes with the duration of the network
354 (Poisot *et al.* (2012)) given that different interactions are occurring through time. Because probabilistic local
355 networks can assess the spatiotemporal variability of interactions (eq. 2), they are useful tools to develop and
356 test different equations of the scaling of network structure with space and time.

357 Metawebs and local networks intrinsically differ in their relationship to spatial and temporal scales. On one
358 hand, because probabilistic metawebs are context-independent, potential interactions do not scale with space
359 and time since they depend solely on the biological capacity of the two taxa to interact, regardless of any
360 particular environmental conditions. This suggests that the probability that two taxa can potentially interact
361 should theoretically be the same in all metawebs in which they are present, provided that the data and methods
362 used to estimate them are the same. As a result, if a potential network M_1 is subsampled from a regional
363 metaweb M_0 to represent the potential interactions between a subset of taxa found in a smaller region, their
364 probabilities of interaction should be identical regardless of scale, i.e. $P_{M_1}(i \rightarrow j) = P_{M_0}(i \rightarrow j)$. On the other
365 hand, local ecological networks scale both spatially and temporally since interactions have more opportunities to
366 be realized and observed in a larger area and duration (e.g., more individuals, more trait variations, more chance
367 of an encounter, McLeod *et al.* (2020)). For example, if a local probabilistic network N_1 of area A_1 is obtained
368 from a larger network N_0 of area A_0 , with A_1 being completely nested within A_0 , interaction probabilities should
369 be smaller in the smaller network, i.e. $P_{N_1}(i \rightarrow j|A_1 < A_0) \leq P_{N_0}(i \rightarrow j|A_0)$. However, if A_1 and A_0 are disjoint,
370 interaction probabilities could be incidentally higher in the smaller area depending on their environmental and
371 biological conditions. Similarly, interaction probabilities should be smaller in networks of shorter durations if
372 time intervals are nested. Statistical models, such as network-area relationships and interaction accumulation

373 curves, can be used to generate estimates of local probabilities of interactions across spatial and temporal scales.



374

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

376 Predicting local networks across time and space is an important goal of network ecology (Strydom *et al.*
377 (2021)). Indeed, in a context of scarcity of interaction data (Jordano (2016)), ecologists must rely on predictive
378 models to reconstruct networks at fine spatial and temporal scales. For example, local ecological networks
379 could be reconstructed using real-time biomonitoring data and adequate numerical models (Bohan *et al.*
380 (2017)), which could pave the way for fine-scale studies of ecosystem functioning and dynamics. Besides
381 predictive models, statistical models can also be built to describe parameters of interest, such as probabilities of
382 interactions. In that case, parameter values provide valuable ecological information in their own rights.

383 Different types of models (e.g., Bayesian and machine learning models) of ecological interactions have been
 384 built for predictive and descriptive purposes (Strydom *et al.* (2021)). Representing interactions probabilistically
 385 reflects the uncertainty of these models, which is usually represented in terms of probability distributions. Here
 386 we show how to build a simple generative mechanistic model of probabilistic interactions that takes into account
 387 their inherent spatiotemporal variability, i.e. a spatiotemporally-explicit model. Our model is not suitable for
 388 potential interactions, which are scale-independent. Rather, it could prove useful for predicting local
 389 interactions across time and space by generating new interaction data after parameter inference.

390 As stated by eq. 1, the probability that two taxa i and j interact at a given location (x, y) is given by the product
 391 of their probability of interaction given co-occurrence and their probability of co-occurrence. First, their
 392 probability of co-occurrence is given by their respective probabilities of occurrence $P_i(x, y)$ and $P_j(x, y)$ and the
 393 strength of association γ between their occurrence and co-occurrence (Cazelles *et al.* (2016)):

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

394 When $\gamma > 1$, the geographic distributions of both taxa are positively associated, which implies that the
 395 occurrence of one taxon increases the probability of occurrence of the other. In empirical networks, $\gamma > 1$ for
 396 most species pairs (Catchen *et al.* (2023)). The co-occurrence of both taxa is the result of a Bernoulli trial

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

397 Second, the probability of interaction given co-occurrence can be made temporally explicit by modeling it as a
 398 Poisson process with rate λ . This parameter corresponds to the expected frequency of interaction between both
 399 taxa in a given time period and its value can be estimated using prior data on interaction strengths, if available.
 400 Specifically, the probability that two co-occurring taxa interact during a time period t_0 is:

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

401 which approaches 1 when $t_0 \rightarrow \infty$.
 402 The realization of the interaction between i and j is the result of a Bernoulli trial with probability $P_N(i \rightarrow j)$. A
 403 Bayesian inference model can thus be built from the previous equations to estimate the value of the λ parameter

404 and generate new interaction data:

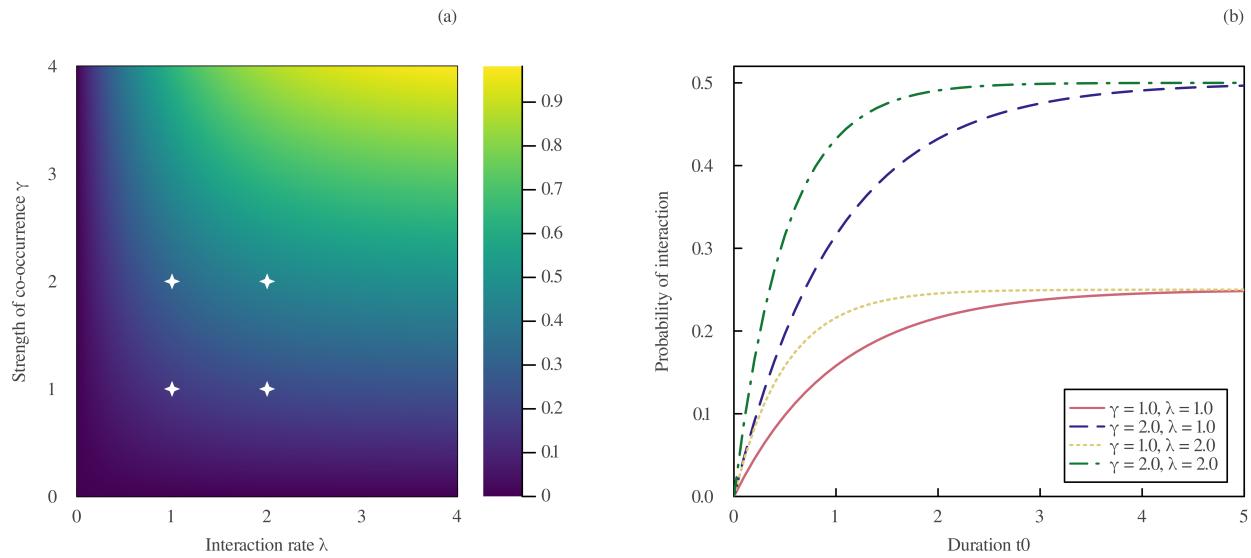
$$I \sim \text{Bernoulli}(P_N(i \rightarrow j)) \quad (10)$$

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

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

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

408 This simple model can be customized in many ways, e.g. by linking λ with given environmental variables or by
409 explicitly modeling observation errors (i.e., the probability of false negatives and false positives).



410

411 Sampling binary networks

412 Probabilistic networks can be used to predict binary interactions through random draws. This can be useful
413 when analyzing the structure of a probabilistic network when analytical probabilistic measures are lacking. A
414 binary network can be generated from independent Bernoulli trials for each interaction. The distribution of a
415 network's property can then be obtained after measuring the structure of all randomly generated networks
416 (Poisot *et al.* (2016)). Doing so allows us to represent the variability of network structure, although possibly
417 with biases when connectance is low (Poisot & Gravel (2014), Chagnon (2015)). When binary networks are
418 generated under a null model, this method can be used for null hypothesis significance testing, in which case the

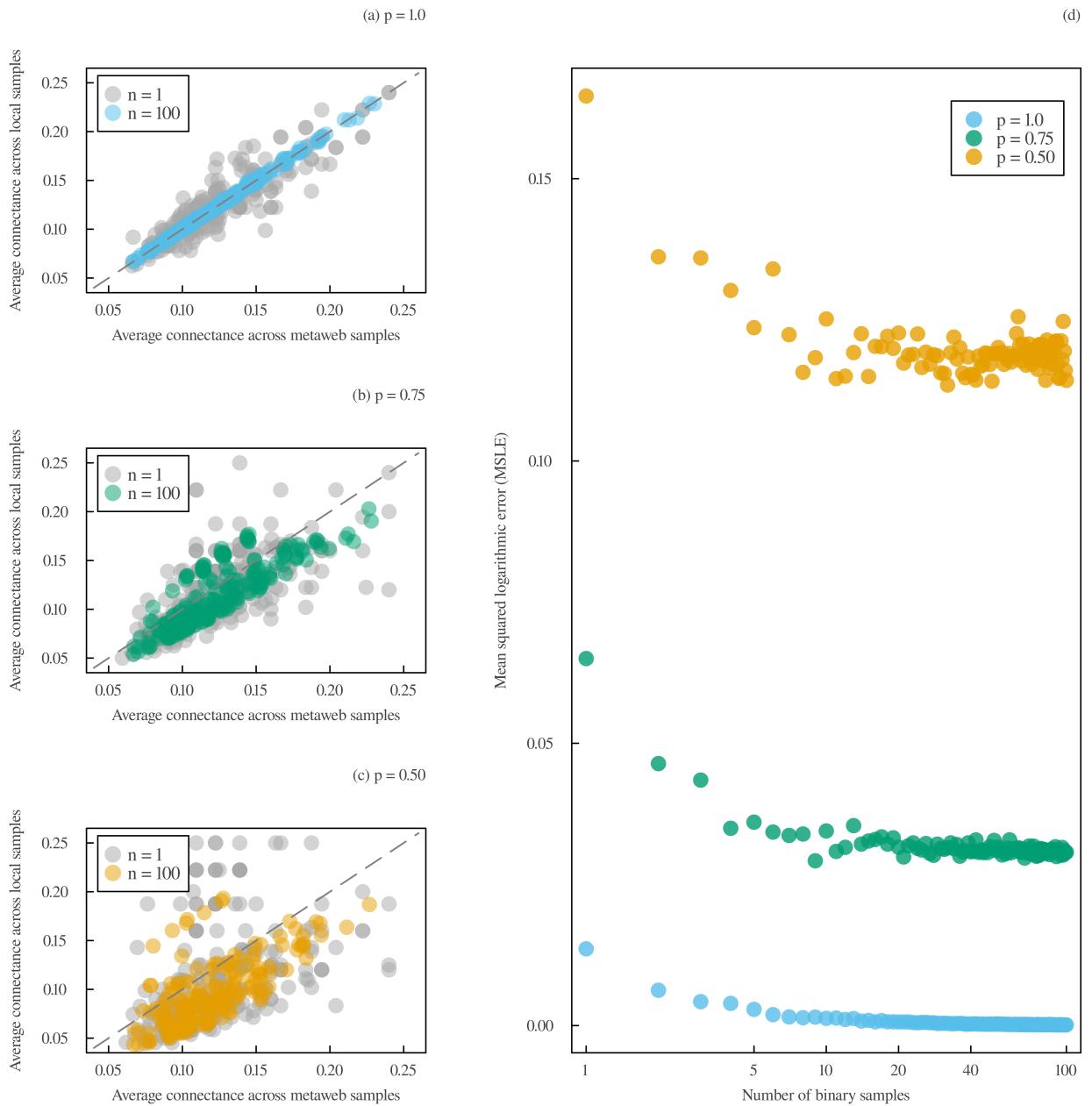
419 observed measure is compared to a null distribution (e.g., Bascompte *et al.* (2003)). Furthermore, randomly
420 generating binary networks across space and time can help us visualize the spatiotemporal variability of
421 network structure and test ecological hypotheses on interactions across large spatial and temporal scales.

422 There are at least two different approaches when sampling binary networks from probabilistic webs across
423 space, e.g. if we want to predict a binary network for each of $n \times n$ grid cells. The first approach is to conduct a
424 single Bernoulli trial for each pair of taxa found in the region of interest based on their regional probability of
425 interaction. As a result, each pair of taxa that are predicted to interact in the regional network realization will
426 interact in all of the n^2 networks in which they co-occur. This sampling technique is best used with potential
427 interactions that have no spatial variation. The second approach is to independently sample each of the n^2
428 networks. In practice, this can be done by generating a different probabilistic network for each grid cell.
429 Depending on how they were obtained, these networks can differ in their taxa composition (nodes) and/or
430 interaction probabilities (edges). Then, binary networks can be independently sampled for each grid cell.
431 Because this method generates spatial variation in binary interactions, it is best used with local interactions.

432 The choice of sampling approach has an impact on the selection of grid cell size. In the first approach,
433 interactions will be the same regardless of cell size because interactions are sampled only once from the
434 regional network. However, in the second approach, local interaction probabilities depend on the network area.
435 For example, let N_1 and N_2 be networks of area $\frac{1}{2}A_0$ nested within A_0 and disjoint from each other, i.e. two
436 contiguous cells that form N_0 . If N_1 and N_2 are independent (which is rarely the case in reality because of
437 spatial auto-correlation), the probability that two taxa i and j interact in N_0 is given by:

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

438 Because of its larger area, the probability that the two taxa interact in N_0 is higher than in N_1 and N_2 . When
439 sampling binary interactions from local networks, it is thus important to sample at the same spatial scale as the
440 one for which probabilities were estimated. Otherwise, interaction probabilities must be adjusted to correspond
441 to the targeted cell size and avoid systematic biases in prediction.



442

443 Prediction of local networks from metawebs

444 Metawebs are an important source of ecological information that can be leveraged for predicting local networks.
 445 Indeed, binary local networks can be reconstructed by selecting a subset of both taxa and interactions from the
 446 metaweb (Dunne (2006)). This implies that there are always more interactions in a metaweb compared to its
 447 corresponding local networks. In practice, because a community's composition is arguably easier to sample or
 448 predict than its interactions, the list of taxa to select can be known empirically or obtained using range maps or

449 species distribution models. The challenge is rather to choose which interactions to select from the metaweb,
450 which could require more sophisticated statistical models and ecological knowledge. Because the structure of
451 local networks is constrained by the one of the metaweb (Saravia *et al.* (2022)), inferring their structure from
452 the metaweb before predicting local pairwise interactions could prove promising (Strydom *et al.* (2021)).

453 When building probabilistic local networks from a probabilistic metaweb, interaction probabilities decrease.
454 This is because two taxa must first have the capacity to interact (i.e., $M_{i,j}^* = 1$ in the binary metaweb M^*) before
455 interacting locally:

$$P_N(i \rightarrow j) = P_M(i \rightarrow j) \times P_N(i \rightarrow j | M_{i,j}^* = 1).$$

456 Therefore, inferring local networks from their metaweb keeping the same values of interaction probability
457 would generate systematic biases in the prediction. In that case, these networks would instead represent
458 smaller-scale metawebs of potential interactions, which could be misinterpreted as local interactions. As
459 suggested by McLeod *et al.* (2021b), even though metawebs do not capture the spatiotemporal variability of
460 interactions, they set the upper limit of local interactions. In other words, the probability that two taxa interact at
461 a given location and time is always lower or equal to the probability of their potential interaction:

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

462 Moreover, the probability that two taxa have the biological capacity to interact must be higher than the
463 probability that they will ever interact because they might never co-occur or encounter locally. More precisely,
464 the accumulated probability of realized interactions across all spatial, temporal, and environmental conditions
465 must be lower than the probability of potential interaction, i.e.

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

466 Conclusion

467 In this contribution, we underline the importance of network metadata for adequately interpreting and
468 manipulating probabilistic interaction data. The mathematical representation of probabilities and their statistical

469 properties depend on the type of interactions (local or potential) and the conditions in which they were
470 estimated. We showed that probabilistic local networks and metawebs differ in their relationship to spatial and
471 temporal scales, with potential interactions remaining consistent across scales. In contrast with metawebs, local
472 interactions are measured in a specific context (e.g., in a given area, time, and biological and environmental
473 conditions) and depend on taxa co-occurrence. These important conceptual differences bring to light the need
474 to use probabilistic data with caution, for instance when generating binary network realizations across space and
475 predicting local networks from metawebs. Clear metadata describing the type of interaction and the variables
476 used in their estimation are required to ensure adequate data manipulation. Better data practices and
477 foundations for probabilistic thinking in network ecology could enable more reliable assessments of the
478 spatiotemporal variability and uncertainty of biotic interactions.

479 It is essential to enhance our comprehension of both potential and local interactions, especially considering the
480 current scarcity of interaction data. However, while sampling biological communities does decrease the
481 uncertainty of interactions by accumulating evidence for their feasibility and local realization, there is a limit to
482 how much we can diminish their inherent randomness. In metawebs, probabilities reflect our limited knowledge
483 of interactions (i.e., our degree of belief that interactions are feasible), which is expected to improve with a
484 larger volume of data. We should anticipate that potential interactions will become more definitive (with
485 probabilities approaching 0 or 1) as we investigate various conditions, including different combinations of
486 species traits. Conversely, in the case of local networks, which can be seen as random instances of metawebs,
487 randomness cannot be as reduced. In local interactions, probabilities can be divided into a component
488 representing uncertainty and another representing spatiotemporal variability. Owing to environmental
489 heterogeneity, there will invariably be instances in which an interaction occurs and others in which it does not,
490 across different times and locations, irrespective of the extent to which we can improve our knowledge of its
491 biological feasibility and the local conditions that facilitate its occurrence. When local networks depict
492 probabilities of observing interactions rather than just their actual occurrence, we must also consider the
493 observation variability as an additional source of randomness. Every ecological process is stochastic but there is
494 also a possibility that a phenomenon goes undetected. Quantifying and partitioning this stochasticity will enable
495 us to make more accurate predictions about ecological interactions at various spatial and temporal scales.

496 **Acknowledgment**

497 We acknowledge that this study was conducted on land within the traditional unceded territory of the Saint
498 Lawrence Iroquoian, Anishinabewaki, Mohawk, Huron-Wendat, and Omàmiwininiwak nations. This work was
499 supported by the Institute for Data Valorisation (IVADO) and the Natural Sciences and Engineering Research
500 Council of Canada (NSERC) Collaborative Research and Training Experience (CREATE) program, through the
501 Computational Biodiversity Science and Services (BIOS²) program. A special thanks to all members of the
502 Black Holes and Revelations working group (organized by BIOS²) for their insightful discussions and valuable
503 feedback on this manuscript.

504 **References**

- 505 Andrade-Pacheco, R., Rerolle, F., Lemoine, J., Hernandez, L., Meïté, A., Juziwelo, L., *et al.* (2020). [Finding hotspots: Development of an adaptive spatial sampling approach](#). *Scientific Reports*, 10, 10939.
- 506 Angilletta, M.J., Jr., Steury, T.D. & Sears, M.W. (2004). [Temperature, Growth Rate, and Body Size in Ectotherms: Fitting Pieces of a Life-History Puzzle1](#). *Integrative and Comparative Biology*, 44, 498–509.
- 507 Araujo, M.B., Rozenfeld, A., Rahbek, C. & Marquet, P.A. (2011). [Using species co-occurrence networks to assess the impacts of climate change](#). *Ecography*, 34, 897–908.
- 508 Bartomeus, I., Gravel, D., Tylianakis, J.M., Aizen, M.A., Dickie, I.A. & Bernard-Verdier, M. (2016). [A common framework for identifying linkage rules across different types of interactions](#). *Functional Ecology*, 30, 1894–1903.
- 509 Bascompte, J., Jordano, P., Melián, C.J. & Olesen, J.M. (2003). [The nested assembly of plantanimal mutualistic networks](#). *Proceedings of the National Academy of Sciences*, 100, 9383–9387.
- 510 Berlow, E.L., Neutel, A.-M., Cohen, J.E., De Ruiter, P.C., Ebenman, B., Emmerson, M., *et al.* (2004). [Interaction strengths in food webs: Issues and opportunities](#). *Journal of Animal Ecology*, 73, 585–598.
- 511 Blanchet, F.G., Cazelles, K. & Gravel, D. (2020). [Co-occurrence is not evidence of ecological interactions](#). *Ecology Letters*, 23, 1050–1063.
- 512 Bohan, D.A., Vacher, C., Tamaddoni-Nezhad, A., Raybould, A., Dumbrell, A.J. & Woodward, G. (2017). [Next-Generation Global Biomonitoring: Large-scale, Automated Reconstruction of Ecological Networks](#).

- 522 *Trends in Ecology & Evolution*, 32, 477–487.
- 523 Bolnick, D.I., Amarasekare, P., Araújo, M.S., Bürger, R., Levine, J.M., Novak, M., *et al.* (2011). Why
524 intraspecific trait variation matters in community ecology. *Trends in Ecology & Evolution*, 26, 183–192.
- 525 Borrett, S.R. & Scharler, U.M. (2019). Walk partitions of flow in Ecological Network Analysis: Review and
526 synthesis of methods and indicators. *Ecological Indicators*, 106, 105451.
- 527 Boulangeat, I., Gravel, D. & Thuiller, W. (2012). Accounting for dispersal and biotic interactions to disentangle
528 the drivers of species distributions and their abundances. *Ecology Letters*, 15, 584–593.
- 529 Brimacombe, C., Bodner, K., Michalska-Smith, M., Poisot, T. & Fortin, M.-J. (2023). Shortcomings of reusing
530 species interaction networks created by different sets of researchers. *PLOS Biology*, 21, e3002068.
- 531 Brose, U., Ostling, A., Harrison, K. & Martinez, N.D. (2004). Unified spatial scaling of species and their
532 trophic interactions. *Nature*, 428, 167–171.
- 533 Canard, E., Mouquet, N., Marescot, L., Gaston, K.J., Gravel, D. & Mouillot, D. (2012). Emergence of
534 Structural Patterns in Neutral Trophic Networks. *PLOS ONE*, 7, e38295.
- 535 Catchen, M.D., Poisot, T., Pollock, L.J. & Gonzalez, A. (2023). The missing link: Discerning true from false
536 negatives when sampling species interaction networks.
- 537 Cazelles, K., Araujo, M.B., Mouquet, N. & Gravel, D. (2016). A theory for species co-occurrence in interaction
538 networks. *Theoretical Ecology*, 9, 39–48.
- 539 Chagnon, P.-L. (2015). Characterizing topology of ecological networks along gradients: The limits of metrics'
540 standardization. *Ecological Complexity*, 22, 36–39.
- 541 Choh, Y., Ignacio, M., Sabelis, M.W. & Janssen, A. (2012). Predator-prey role reversals, juvenile experience
542 and adult antipredator behaviour. *Scientific Reports*, 2, 728.
- 543 Cirtwill, A.R., Eklöf, A., Roslin, T., Wootton, K. & Gravel, D. (2019). A quantitative framework for
544 investigating the reliability of empirical network construction. *Methods in Ecology and Evolution*, 10,
545 902–911.
- 546 Dunne, J.A. (2006). The Network Structure of Food Webs. In: *Ecological networks: Linking structure and*
547 *dynamics* (eds. Dunne, J.A. & Pascual, M.). Oxford University Press, pp. 27–86.

- 548 Elmasri, M., Farrell, M.J., Davies, T.J. & Stephens, D.A. (2020). **A hierarchical Bayesian model for predicting**
549 **ecological interactions using scaled evolutionary relationships.** *The Annals of Applied Statistics*, 14,
550 221–240.
- 551 Elton, C.S. (2001). *Animal Ecology*. University of Chicago Press, Chicago, IL.
- 552 Emmerson, M.C. & Raffaelli, D. (2004). **Predatorprey body size, interaction strength and the stability of a real**
553 **food web.** *Journal of Animal Ecology*, 73, 399–409.
- 554 Fortuna, M.A. & Bascompte, J. (2006). **Habitat loss and the structure of plantanimal mutualistic networks.**
555 *Ecology Letters*, 9, 281–286.
- 556 Galiana, N., Lurgi, M., Claramunt-López, B., Fortin, M.-J., Leroux, S., Cazelles, K., et al. (2018). **The spatial**
557 **scaling of species interaction networks.** *Nature Ecology & Evolution*, 2, 782–790.
- 558 Gilman, S.E., Urban, M.C., Tewksbury, J., Gilchrist, G.W. & Holt, R.D. (2010). **A framework for community**
559 **interactions under climate change.** *Trends in Ecology & Evolution*, 25, 325–331.
- 560 Godsoe, W., Murray, R. & Iritani, R. (2022). **Species interactions and diversity: A unified framework using Hill**
561 **numbers.** *Oikos*, n/a, e09282.
- 562 Golubski, A.J. & Abrams, P.A. (2011). **Modifying modifiers: What happens when interspecific interactions**
563 **interact?** *Journal of Animal Ecology*, 80, 1097–1108.
- 564 Gomez, J.M., Verdu, M. & Perfectti, F. (2010). **Ecological interactions are evolutionarily conserved across the**
565 **entire tree of life.** *Nature*, 465, 918–U6.
- 566 Gonzalez-Varo, J.P. & Traveset, A. (2016). **The Labile Limits of Forbidden Interactions.** *Trends in Ecology &*
567 *Evolution*, 31, 700–710.
- 568 Gravel, D., Baiser, B., Dunne, J.A., Kopelke, J.-P., Martinez, N.D., Nyman, T., et al. (2019). **Bringing Elton**
569 **and Grinnell together: A quantitative framework to represent the biogeography of ecological interaction**
570 **networks.** *Ecography*, 42, 401–415.
- 571 Gravel, D., Poisot, T., Albouy, C., Velez, L. & Mouillot, D. (2013). **Inferring food web structure from**
572 **predatorprey body size relationships.** *Methods in Ecology and Evolution*, 4, 1083–1090.
- 573 Guimarães, P.R. (2020). **The Structure of Ecological Networks Across Levels of Organization.** *Annual Review*
574 *of Ecology, Evolution, and Systematics*, 51, 433–460.

- 575 Guimerà, R. & Sales-Pardo, M. (2009). [Missing and spurious interactions and the reconstruction of complex](#)
576 [networks](#). *Proceedings of the National Academy of Sciences*, 106, 22073–22078.
- 577 Hemprich-Bennett, D.R., Oliveira, H.F.M., Le Comber, S.C., Rossiter, S.J. & Clare, E.L. (2021). [Assessing the](#)
578 [impact of taxon resolution on network structure](#). *Ecology*, 102, e03256.
- 579 Herrera, C.M. (1989). [Pollinator abundance, morphology, and flower visitation rate: Analysis of the “quantity”](#)
580 [component in a plant-pollinator system](#). *Oecologia*, 80, 241–248.
- 581 Hortal, J., de Bello, F., Diniz-Filho, J.A.F., Lewinsohn, T.M., Lobo, J.M. & Ladle, R.J. (2015). [Seven Shortfalls](#)
582 [that Beset Large-Scale Knowledge of Biodiversity](#). *Annual Review of Ecology, Evolution, and Systematics*,
583 46, 523–549.
- 584 Jordano, P. (1987). [Patterns of Mutualistic Interactions in Pollination and Seed Dispersal: Connectance,](#)
585 [Dependence Asymmetries, and Coevolution](#). *The American Naturalist*, 129, 657–677.
- 586 Jordano, P. (2016). [Sampling networks of ecological interactions](#). *Functional Ecology*, 30, 1883–1893.
- 587 Jordano, P., Bascompte, J. & Olesen, J.M. (2003). [Invariant properties in coevolutionary networks of](#)
588 [plantanimal interactions](#). *Ecology Letters*, 6, 69–81.
- 589 Kéfi, S., Berlow, E.L., Wieters, E.A., Joppa, L.N., Wood, S.A., Brose, U., *et al.* (2015). [Network structure](#)
590 [beyond food webs: Mapping non-trophic and trophic interactions on Chilean rocky shores](#). *Ecology*, 96,
591 291–303.
- 592 Kéfi, S., Berlow, E.L., Wieters, E.A., Navarrete, S.A., Petchey, O.L., Wood, S.A., *et al.* (2012). [More than a](#)
593 [meal... integrating non-feeding interactions into food webs](#). *Ecology Letters*, 15, 291–300.
- 594 Kéfi, S., Miele, V., Wieters, E.A., Navarrete, S.A. & Berlow, E.L. (2016). [How Structured Is the Entangled](#)
595 [Bank? The Surprisingly Simple Organization of Multiplex Ecological Networks Leads to Increased](#)
596 [Persistence and Resilience](#). *PLOS Biology*, 14, e1002527.
- 597 Klecka, J. & Boukal, D.S. (2014). [The effect of habitat structure on prey mortality depends on predator and](#)
598 [prey microhabitat use](#). *Oecologia*, 176, 183–191.
- 599 Kopelke, J.-P., Nyman, T., Cazelles, K., Gravel, D., Vissault, S. & Roslin, T. (2017). [Food-web structure of](#)
600 [willow-galling sawflies and their natural enemies across Europe](#). *Ecology*, 98, 1730–1730.
- 601 Kulkarni, M.A., Kweka, E., Nyale, E., Lyatuu, E., Mosha, F.W., Chandramohan, D., *et al.* (2006).
602 [Entomological Evaluation of Malaria Vectors at Different Altitudes in Hai District, Northeastern Tanzania](#).

- 603 *Journal of Medical Entomology*, 43, 580–588.
- 604 McCann, K. (2007). Protecting biostructure. *Nature*, 446, 29–29.
- 605 McCann, K.S. (2011). Food webs (MPB-50). In: *Food Webs (MPB-50)*. Princeton University Press.
- 606 McLeod, A., Leroux, S.J., Gravel, D., Chu, C., Cirtwill, A.R., Fortin, M.-J., *et al.* (2021a). [Sampling and asymptotic network properties of spatial multi-trophic networks](#). *Oikos*, 130, 2250–2259.
- 608 McLeod, A., Leroux, S.J., Gravel, D., Chu, C., Cirtwill, A.R., Fortin, M.-J., *et al.* (2021b). [Sampling and asymptotic network properties of spatial multi-trophic networks](#). *Oikos*, n/a.
- 610 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.
- 612 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.
- 615 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.
- 617 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.
- 620 Pascual, M., Dunne, J.A. & Dunne, J.A. (2006). *Ecological Networks: Linking Structure to Dynamics in Food Webs*. Oxford University Press, USA.
- 622 Petchey, O.L., Beckerman, A.P., Riede, J.O. & Warren, P.H. (2008). [Size, foraging, and food web structure](#). *Proceedings of the National Academy of Sciences*, 105, 4191–4196.
- 624 Pilosof, S., Porter, M.A., Pascual, M. & Kéfi, S. (2017). [The multilayer nature of ecological networks](#). *Nature Ecology & Evolution*, 1, 1–9.
- 626 Poisot, T., Canard, E., Mouillot, D., Mouquet, N. & Gravel, D. (2012). [The dissimilarity of species interaction networks](#). *Ecology Letters*, 15, 1353–1361.
- 628 Poisot, T., Cirtwill, A.R., Cazelles, K., Gravel, D., Fortin, M.-J. & Stouffer, D.B. (2016). [The structure of probabilistic networks](#). *Methods in Ecology and Evolution*, 7, 303–312.

- 630 Poisot, T. & Gravel, D. (2014). When is an ecological network complex? Connectance drives degree
631 distribution and emerging network properties. *PeerJ*, 2, e251.
- 632 Poisot, T., Stouffer, D.B. & Gravel, D. (2015). Beyond species: Why ecological interaction networks vary
633 through space and time. *Oikos*, 124, 243–251.
- 634 Pollock, L.J., Tingley, R., Morris, W.K., Golding, N., O’Hara, R.B., Parris, K.M., et al. (2014). Understanding
635 co-occurrence by modelling species simultaneously with a Joint Species Distribution Model (JSDM).
636 *Methods in Ecology and Evolution*, 5, 397–406.
- 637 Pomeranz, J.P.F., Thompson, R.M., Poisot, T. & Harding, J.S. (2019). Inferring predator-prey interactions in
638 food webs. *Methods in Ecology and Evolution*, 10, 356–367.
- 639 Pringle, R.M. & Hutchinson, M.C. (2020). Resolving Food-Web Structure. *Annual Review of Ecology,*
640 *Evolution, and Systematics*, 51, 55–80.
- 641 Proulx, S.R., Promislow, D.E.L. & Phillips, P.C. (2005). Network thinking in ecology and evolution. *Trends in*
642 *Ecology & Evolution*, SPECIAL ISSUE: BUMPER BOOK REVIEW, 20, 345–353.
- 643 Pulliam, H.R. (1974). On the Theory of Optimal Diets. *The American Naturalist*, 108, 59–74.
- 644 Rooney, N. & McCann, K.S. (2012). Integrating food web diversity, structure and stability. *Trends in Ecology*
645 *& Evolution*, 27, 40–46.
- 646 Salim, J.A., Saraiva, A.M., Zermoglio, P.F., Agostini, K., Wolowski, M., Drucker, D.P., et al. (2022). Data
647 standardization of plant-pollinator interactions. *GigaScience*, 11, giac043.
- 648 Saravia, L.A., Marina, T.I., Kristensen, N.P., De Troch, M. & Momo, F.R. (2022). Ecological network
649 assembly: How the regional metaweb influences local food webs. *Journal of Animal Ecology*, 91, 630–642.
- 650 Schmeller, D.S., Julliard, R., Bellingham, P.J., Böhm, M., Brummitt, N., Chiarucci, A., et al. (2015). Towards a
651 global terrestrial species monitoring program. *Journal for Nature Conservation*, 25, 51–57.
- 652 Singer, M.C. & McBride, C.S. (2012). Geographic mosaics of species’ association: A definition and an
653 example driven by plant-insect phenological synchrony. *Ecology*, 93, 2658–2673.
- 654 Stouffer, D.B., Rezende, E.L. & Amaral, L.A.N. (2011). The role of body mass in diet contiguity and food-web
655 structure. *Journal of Animal Ecology*, 80, 632–639.

- 656 Strydom, T., Bouskila, S., Banville, F., Barros, C., Caron, D., Farrell, M.J., *et al.* (2022). [Food web](#)
657 [reconstruction through phylogenetic transfer of low-rank network representation](#). *Methods in Ecology and*
658 *Evolution*, 13.
- 659 Strydom, T., Catchen, M.D., Banville, F., Caron, D., Dansereau, G., Desjardins-Proulx, P., *et al.* (2021). [A](#)
660 [roadmap towards predicting species interaction networks \(across space and time\)](#). *Philosophical*
661 *Transactions of the Royal Society B-Biological Sciences*, 376, 20210063.
- 662 Tylianakis, J.M., Tscharntke, T. & Lewis, O.T. (2007). [Habitat modification alters the structure of tropical](#)
663 [hostparasitoid food webs](#). *Nature*, 445, 202–205.
- 664 Vázquez, D.P., Peralta, G., Cagnolo, L., Santos, M. & Igual, E. autores contribuyeron por. (2022). [Ecological](#)
665 [interaction networks. What we know, what we don't, and why it matters](#). *Ecología Austral*, 32, 670–697.
- 666 Weinstein, B.G. & Graham, C.H. (2017a). [On comparing traits and abundance for predicting species](#)
667 [interactions with imperfect detection](#). *Food Webs*, 11, 17–25.
- 668 Weinstein, B.G. & Graham, C.H. (2017b). [Persistent bill and corolla matching despite shifting temporal](#)
669 [resources in tropical hummingbird-plant interactions](#). *Ecology Letters*, 20, 326–335.
- 670 Wood, S.A., Russell, R., Hanson, D., Williams, R.J. & Dunne, J.A. (2015). [Effects of spatial scale of sampling](#)
671 [on food web structure](#). *Ecology and Evolution*, 5, 3769–3782.
- 672 Woodward, G., Benstead, J.P., Beveridge, O.S., Blanchard, J., Brey, T., Brown, L.E., *et al.* (2010). [Chapter 2 -](#)
673 [Ecological Networks in a Changing Climate](#). In: *Advances in Ecological Research*, Ecological Networks
674 (ed. Woodward, G.). Academic Press, pp. 71–138.
- 675 Woodward, G., Brown, L.E., Edwards, F.K., Hudson, L.N., Milner, A.M., Reuman, D.C., *et al.* (2012). [Climate](#)
676 [change impacts in multispecies systems: Drought alters food web size structure in a field experiment](#).
- 677 [Philosophical Transactions of the Royal Society B: Biological Sciences, 367, 2990–2997.](#)

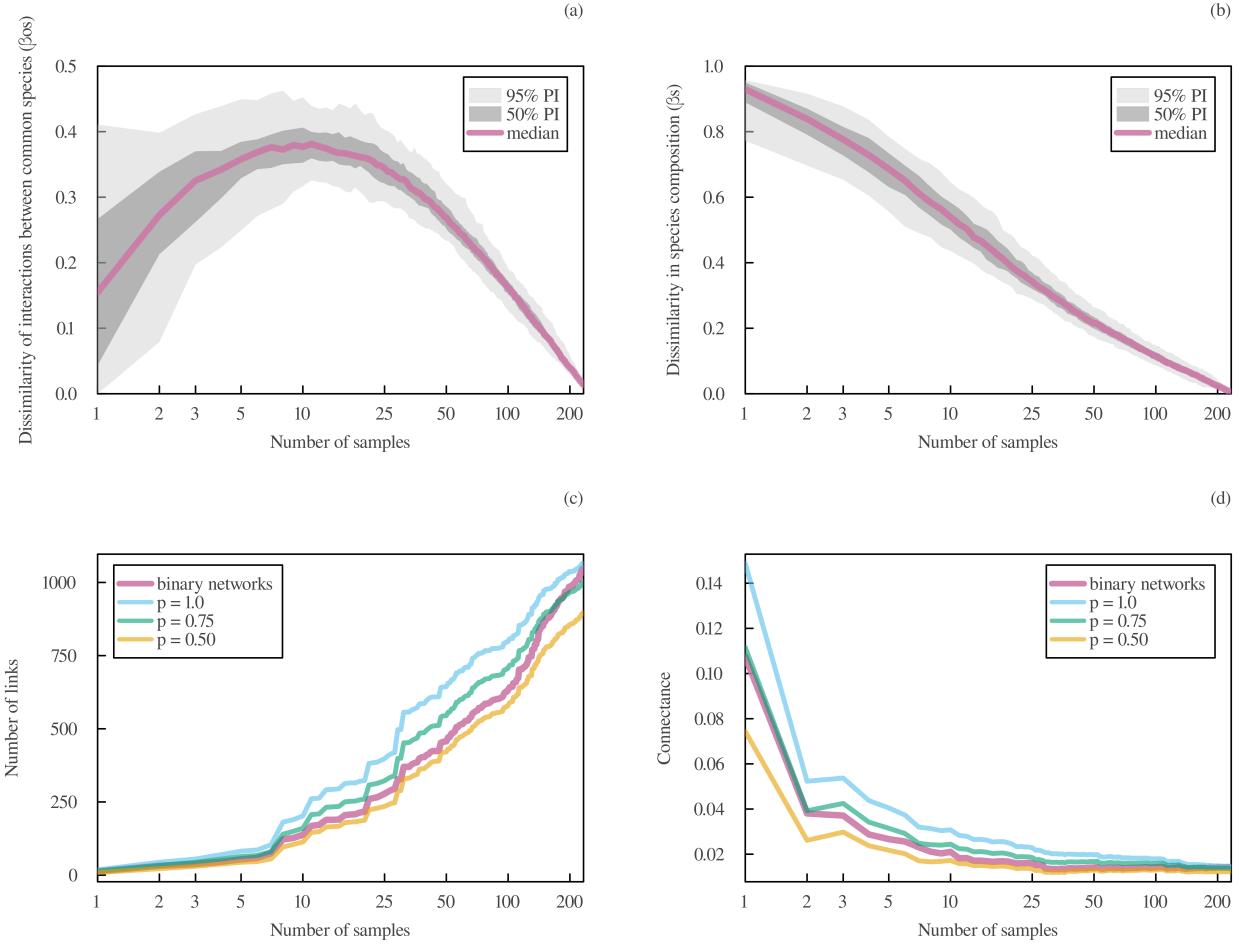


Figure 1: Network accumulation curves. (a) Dissimilarity of interactions between common species and (b) dissimilarity in species composition between aggregated local networks and the binary regional metaweb. Aggregated local networks were obtained by sequentially and randomly selecting a number of binary local networks and aggregating both their species and interactions. In both panels, the colored line represents the median dissimilarity across simulations and the grey areas cover the 50% and 95% percentile intervals. (c) Scaling of the number of links and (d) scaling of connectance with the number of sampled binary and probabilistic local networks. For a better comparison with binary networks, local probabilistic networks were derived from the probabilistic metaweb with a false positive and false negative rates of zero. We used a fixed value of p (the local probability of interaction among potentially interacting species) for each local network, which varied across curves. Aggregated probabilistic local networks were obtained by sequentially and randomly selecting a number of probabilistic local networks and aggregating both their species and interactions (with the value of p adjusting according to eq. 5.).