

¹ **Title:** Bringing Elton and Grinnell together: a quantitative framework to represent
² the biogeography of ecological interactions

³ **Authors:** Dominique Gravel^{1,2,*}, Timothée Poisot^{1,2,3}, Néo Martinez⁴, Jennifer A.
⁴ Dunne⁵, Jason Tylianakis³, Tomas Roslin⁶, Tommy Nyman⁶, Ben Baiser⁷, Daniel B.
⁵ Stouffer³

⁶
⁷ 1: Canada Research Chair in Integrative Ecology. Département de biologie, Uni-
⁸ versité de Sherbrooke, 2500 Boulevard l'Université, Sherbrooke (Québec). J1K 2R1

⁹
¹⁰ 2: Québec Centre for Biodiversity Sciences

¹¹
¹² 3: University of Canterbury at Christchurch, School of Biological Sciences,

¹³
¹⁴ 4:

¹⁵
¹⁶ 5:

¹⁷
¹⁸ 6:

¹⁹
²⁰ 7:

²¹
²² **Keywords:** networks, spatial ecology, co-occurrence, probability of interaction

²³
²⁴ **Words in the abstract:**

²⁵ **Words in the main text:**

²⁶ **Words in the legends:**

²⁷ **Figures:**

²⁸ **Tables:**

²⁹ **References:**

³⁰ **Abstract**

³¹ **Introduction**

³² Community ecology is defined in most textbooks as *the study of the interactions that*
³³ *determine the distribution and abundance of organisms* (Krebs2001). Despite a general
³⁴ concensus on this definition (Scheiner2007), it is surprising that most research on the
³⁵ variation of community structure has focused mostly on the turnover of species com-
³⁶ position (Anderson2011), neglecting variation in the way they interact with each other
³⁷ (McGill2006). Not surprisingly given this fact, biogeographers are still struggling
³⁸ to figure out if interactions do impact species distribution (Wisz2012; Kissling2012).
³⁹ There has been a step ahead with recent methodological improvements accounting
⁴⁰ for interactions in species distribution models (Pollock2014; Pelissier2013), but those
⁴¹ remain nonetheless a 'species- based' approach to communities where interactions are
⁴² fixed covariates affecting distribution. Here we plea for a more integrated approach
⁴³ of species and interaction distribution.

⁴⁴ The problem of community assembly is often formulated as *how do we sample a*
⁴⁵ *regional pool of species to constitute a local community?* This question could be rewritten
⁴⁶ to address the problem of network assembly, as *how do we sample a regional pool of*
⁴⁷ *interactions to constitute a local interaction network?*. An illustration of this problem
⁴⁸ for a food web is provided at Figure 1. The metaweb represents potential interactions
⁴⁹ among all species that could be found in a given area. In this particular case, there are
⁵⁰ 274 nodes, and 1077 links among plants, herbivores and parasitoids from Northern
⁵¹ Europe. An instance of a local community is also illustrated, with 44 nodes and 160
⁵² interactions. Only 51.2% of all potential interactions are realized. Our objective here
⁵³ is to provide a conceptual framework to explain the sampling of the regional pool
⁵⁴ of interactions, along with a quantitative method to predict it. The problem could
⁵⁵ be formalized by sequentially understanding why only a fraction of the species are
⁵⁶ locally co-occurring, and after why these species are interacting or not.

57 There are multiple causes to spatial turnover in community composition. The first
58 and most studied driver is the effect of variation in the abiotic environment on species
59 performance. Combined with specific responses in demography, it generates varia-
60 tion among localities by locally selecting the fittest species (Leibold2004). Stochastic-
61 ity additionnaly plays a role, either because of the inherently unpredictable nature of
62 colonization and extinction events, or because of strong non-linear feedbacks generat-
63 ing alternative transients and equilibrium (Chase2007; Vellend2014). The analysis of
64 community turnover is usually performed with data represented in a table with rows
65 corresponding to sites (or measurements) and the columns to species. Beta- diversity
66 metrics quantify the variance of this community data (Legendre2005). Traditionnal
67 approaches rely on measures of the dissimilarity between communities, using indices
68 such as the Jaccard or the Bray Curtis measures of dissimilarity. Recent methods de-
69 compose the total variation of the community data into species and site contributions
70 to beta diversity (Legendre2013). Even though these methods compare whole list of
71 species between sites, or measurements, they remain fundamentally 'species-based'
72 since they are all based on the within column variation. None of them consider the
73 variation of associations (pairs or higher order motifs) explicitly.

74 The niche is by far the dominant concept to explain species distribution and com-
75 munity assembly, from the local to the global scale. Following Hutchinson, the niche
76 is viewed as the set of environmental conditions allowing a population to establish
77 and maintain a population (see also Holt2009). Community turnover arises following
78 successive species replacement along an environmental gradient, in agreement with
79 a Gleasonian view of communities (Gleason). The concept is straightforward to op-
80 erationalize with species distribution models, as it maps naturally on the available
81 data (both distribution and environmental data) and a vast array of statistical tools
82 representing it (e.g. Biomod, MaxEnt). It is however much harder to account for

83 ecological interactions in this approach (Peterson2011). These are often viewed as ex-
84 ternalities, constraining or expanding the range of environmental conditions required
85 for a species to maintain a population (Pulliam2000; Soberon2007).

86 The network approach proposes a convenient formalism to represent the structure
87 of local communities. Species are represented as nodes and interactions by links. The
88 data could also be represented by matrices, with each species in rows and columns,
89 and the entries representing the occurrence or the intensity of an interaction. Stud-
90 ies of network diversity are mostly concerned by the distribution of the interactions
91 within a location and not so much by the variation among locations (Dunne2005; Bas-
92 compte2007; Ings2007; Kefi2012). Network complexity is computed as the number
93 of interactions in the case of binary networks or interaction diversity in the case of
94 quantitative networks (Bersier2002). However, there is now evidence that ecological
95 interactions do also vary in space and time (Poisot2012; Trojelsgaard2015). The vari-
96 ability of community structure in this situation arises from the turnover of species
97 composition, along with the turnover of the interactions among pairs of species.
98 The occurrence and intensity of interactions could vary because of the environment,
99 species abundance and higher order interactions (Poisot2015a). The variation in com-
100 munity composition is often independent of the variation of ecological interactions,
101 suggesting these two components of network variability respond to different drivers
102 (Poisot2012).

103 Interestingly, the ecological network literature also has its own 'niche model' to
104 position a species in a community (Williams2000). The niche of a species in this
105 context represents the multidimensional space that could represent all of its interac-
106 tions. Each species is characterized by a niche position, an optimum and a range
107 over 3 to 5 different niche axes (Williams2000; Eklof2013). The niche model has been
108 successful at explaining the complexity of a variety of networks, from food webs to

109 plant-pollinator systems. The conceptual framework is however limited to local com-
110 munities and does not provide any explanation to the turnover of network structure
111 along environmental gradients.

112 Here we adopt the view that a community structure is best represented as an
113 ecological network of interactions and develop a theory to explain its turnover in
114 space and time. We propose a new description of the niche that integrates the effect
115 of the environment on species distribution and on ecological interactions. We first
116 present the conceptual framework and then formalize it mathematically with a prob-
117 abilistic approach to the sampling of the regional pool of interactions. We apply the
118 framework to study the spatial variation of host-parasite interactions across Europe.
119 We find that the variation of the environment causes both species and interaction
120 turnover. The network structures changes systematically across the latitudinal gra-
121 dient, with a peak of connectance at intermediate latitudes. At the pairwise level,
122 the statistical approach could be conceived as an interaction distribution model. At
123 the community level, the approach provides a likelihood based method to compare
124 different hypotheses of network turnover.

125 **The integrated niche**

126 Correctly describing the niche is key to understand turnover in community structure.
127 Despite several attempts to refresh the conceptual basis of what ecological niches
128 are, ecologists have not moved far past the "n-dimensional hypervolume" formalism
129 introduced by Hutchinson. Despite its intuitive interpretation and translation into
130 species distribution models (Boulangeat2012; Blonder2014), the concept has been con-
131 stantly criticized (Hardin1960; Peters1991; Chase2003; Silvertown2004; Soberon2007)
132 and several attempts have been made to expand and reinforce it.

133 Part of the problem surrounding the definition of the niche has been clarified
134 with the distinction between Eltonian and Grinnellian definitions (Chase2003). The
135 Grinnellian dimension of the niche is the set of environmental conditions required
136 for a species to maintain a population in a location. The Grinnellian niche is the most
137 intuitive to apply and is the conceptual backbone of species distribution models. The
138 Eltonian niche on the other hand is the effect of a species on its environment. This
139 aspect of the niche is well known by community ecologists, but is trickier to turn into
140 predictive models. Nonetheless, the development of the niche model of food web
141 structure (Williams2000) and its parameterization (Williams2010; Gravel2013) made it
142 more operational.

143 These perspectives are rather orthogonal to each other and thus lead to consider-
144 able confusion in the literature (McIntyre). Chase2003 attempted to reconcile them
145 in their definition of the niche: *[The niche is] the joint description of the environmental*
146 *conditions that allow a species to satisfy its minimum requirements so that the birth rate of a*
147 *local population is equal or greater than its death rate along with the set of per capita effects*
148 *of that species on these environmental conditions.* Their representation merges zero-net
149 growth isolines figuring the Grinnellian niche (when do the population persists) with
150 impact vectors figuring the Eltonian niche (what is the per capita impact). While this
151 representation has been very influential in community ecology at the local scale (the
152 resource-ratio theory of coexistence - Tilman1982), it remains impracticable at the
153 large spatial scale because of the difficulties to measure it. The absence of any math-
154 ematical representation of the niche that could easily be fit to ecological data perhaps
155 explain why biogeographers are still struggling to develop species distribution mod-
156 els taking into account ecological interactions.

157 We propose to integrate the two perspectives of the niche with a visual repre-
158 sentation of both components. The underlying rationale is that, in addition to the

159 environmental constraints on demographic performance, any organism requires re-
160 sources to sustain its metabolic demand and reproduction. Abiotic environmental
161 axes are any non-consumable factors affecting the demographic performance of an
162 organism. Alternatively, the resource axes are traits of the resources allowing inter-
163 actions with the consumer. The niche should therefore be viewed as the set of abiotic
164 environmental factors (the Grinnelian component) along with the set of traits (the
165 Eltonian component) allowing a sustainable population to establish and maintain at
166 a location. Accordingly, each species could be characterized by an optimal position
167 in both the environmental (x-axis) and the trait (y-axis) plane. The integrated niche is
168 then the hypervolume where interactions could occur and sustain a population. This
169 approach radically change the representation of the niche, putting species distribution
170 and ecological interactions in the same formalism.

171 The limits of the niches could be independent of each other (as in the exam-
172 ple at Fig. 2), alternatively interact. For instance, the optimimal prey body size
173 for predatory fishes could reduce with increasing temperature (Lelong2015), which
174 would make diet boundaries functions of the environment. The other way around,
175 we could also consider that the growth rate of the predator could change with the
176 body size of the preys it feeds on, thereby altering the environmental boundaries.

₁₇₇ **A probabilistic representation of ecological interactions**

₁₇₈ **networks in space**

₁₇₉ We now formalize the integrated niche with a probabilistic approach to interactions
₁₈₀ and distribution. We seek to represent the probability an interaction between species i
₁₈₁ and j occurs at location y . We define L_{ijy} as a stochastic variable and are looking at the
₁₈₂ probability this event occurs, $P(L_{ijy})$. The occurrence of an interaction is dependent on
₁₈₃ the co-occurrence of species i and j . This argument might seem trivial at first, but the
₁₈₄ explicit consideration of this condition in the probabilistic representation of ecological
₁₈₅ interactions will prove fundamental to understand their variation. We thus define X_{iy}
₁₈₆ as a stochastic variable representing the occurrence of a species i at location y , and
₁₈₇ similarly X_{jy} the co-occurrence of species i and j . The quantity we seek to understand
₁₈₈ is the probability of a joint event:

$$P(X_{i,y}, X_{j,y}, L_{ij,y}) \quad (1)$$

₁₈₉ Or simply said, the probability of observing both species i and j , and an interaction
₁₉₀ from i to j . This probability could be decomposed in two parts using the product rule
₁₉₁ of probabilities:

$$P(X_{iy}, X_{jy}, L_{ijy}) = P(X_{iy}, X_{jy}|E_y)P(L_{ijy}|\mathbf{T}, X_{iy}, X_{jy}, E_y) \quad (2)$$

₁₉₂ The left term is the probability of observing the two species co- occurring at lo-
₁₉₃ cation y . It corresponds to the Grinnelian dimension of the niche. The right term
₁₉₄ is a conditional probability, representing the probability that an interaction occurs
₁₉₅ between species i and j , given their set of traits \mathbf{T} and they are co- occurring. It
₁₉₆ is referred as the metaweb and corresponds to the Eltonian dimension of the niche

described above. We will see below how this formalism could be directly fitted to empirical data. But before turning to an application, we will discuss the interpretation of different variants of these two terms.

Variants of co-occurrence

There are several variants to the co-occurrence probability representing different hypotheses about the temporal and spatial variation in network structure (see the explicit formulations at Table 1). The simplest model relates co-occurrence probability directly to the environment, $P(X_{ijy}|E_y)$. In this situation there is no underlying assumption about the ecological processes responsible for co-occurrence. It could arise because of the impact of ecological interactions on distribution (Pollock2014) or alternatively because of share environmental requirements. In the former case, species are not independent to each other and the conditional occurrence must be accounted for explicitly, $P(X_{ijy}|E_y) = P(X_i|E_y, X_j)P(X_j|E_y)$. In the later case, species are independent and only the marginal occurrence must be accounted for, $P(X_{ijy}|E_y) = P(X_i|E_y)P(X_j|E_y)$.

The co-occurrence probability itself could be dependent on ecological interactions. Direct pairwise interactions such as competition, facilitation and predation have long been studied for their impact on co-distribution (such as in the cases studied by Diamond1976, Connor1980, Gotelli2000 etc.). Second and higher order interactions (e.g. trophic cascade) could also impact co-occurrence. Co-occurrence in ecological networks is however a topic of its own, influenced by the degree distribution and species richness (Cazelles2015). Almost only first order and second order interactions do impact co-occurrence. The covariance of interacting species to an environmental gradient also influences co-occurrence (Cazelles2016). Because of the complexity of relating co-occurrence to the interaction network structure, we will focus here on the

222 variation of interactions and not on the distribution, and leave this issue for Discus-
223 sion and future research.

224 **Variants of the metaweb**

225 There are also variants of the metaweb. First, most documented metawebs have thus
226 far considered that ecological interactions are deterministic, not probabilistic (e.g.
227 Havens1992; Woods2015). Species are assumed to interact if they are found together
228 in a location, independently of their abundance and the environment. In other words,
229 $P(L_{ijy}) = 1$ if $X_{ijy} = 1$, and 0 otherwise. This approach might be a reasonable ap-
230 proximation when the sampling and inference scales are large enough so that prob-
231 abilities of observing at least one interaction converges to unity and that the only
232 variation of networks considered arises from species distribution.

233 Ecological interactions could also vary with the environment, such as $P(L_{ijy}|E_y)$.
234 Although it is not common to see a conditional representation of ecological interac-
235 tions, experimental studies of pairwise interactions revealing their sensitivity to the
236 environment are common (REF). For instance, it has been documented that the pre-
237 dation risks of shorebirds do vary at the continental scale, from the south to the north
238 (REF). The effect of the environment on interactions propagate up the community
239 and influence network structure (REF).

²⁴⁰ **Application: continental-scale variation of host-parasite**
²⁴¹ **community structure**

²⁴² In this section we provide an illustration of the framework with an empirical dataset
²⁴³ of host-parasitoid networks sampled throughout continental Europe along a south-
²⁴⁴ north gradient. The analysis targets networks composed of willows (genus *Salix*),
²⁴⁵ their galling insects, and the natural enemies of these gallers. The questions we
²⁴⁶ address with the framework and the dataset are: i) how much variation in the network
²⁴⁷ structure is there across the gradient and ii) what is the primary driver of network
²⁴⁸ turnover across the gradient?

²⁴⁹ **Data**

²⁵⁰ Communities of both willows and gallers are species-rich and widely distributed,
²⁵¹ with pronounced variation in community composition across space. The genus *Salix*
²⁵² includes over 400 species, most of which are shrubs or small trees (ref). The genus
²⁵³ is common in most habitats across the Northern Hemisphere (ref). Willows support
²⁵⁴ a highly diverse community of herbivorous insects, and one of the main herbivore
²⁵⁵ groups in this system are gall-inducing sawflies (Hymenoptera: Tenthredinidae: Ne-
²⁵⁶ matinae: Euurina (ref). Gall formation is induced by sawfly females during ovipo-
²⁵⁷ sition, and gall formation includes marked manipulation of plant chemistry by the
²⁵⁸ galler (ref). The enemy community of the gallers includes nearly 100 species belong-
²⁵⁹ ing to 17 insect families of four orders. These enemies encompass both inquilines and
²⁶⁰ parasitoids: inquiline larvae (Coleoptera, Lepidoptera, Diptera, and Hymenoptera)
²⁶¹ feed primarily on gall tissue, but typically kill the galler larva in the process. Par-
²⁶² asitoid larvae (representing many families in Hymenoptera) kill the galler larvae by
²⁶³ direct feeding (REF). In terms of associations between the trophic levels, phylogeny-

264 based comparative studies have demonstrated that galls represent "extended pheno-
265 types" of the gallers, meaning that gall form, location, and chemistry is determined
266 mainly by the galling insects and not by their host plants [ref]. Because galler par-
267 asitoids have to penetrate a protective wall of modified plant tissue in order to gain
268 access to their victims, gall morphology has been inferred to strongly affect the asso-
269 ciations between parasitoids and hosts (ref, ref). Thus, the set of parasitoids attacking
270 each host is presumptively constrained by the form, size, and thickness of its gall.

271 Local realizations of the willow–galler–enemy network were reconstructed from
272 samples collected between 1982 and 2010. Plant galls were collected by J.P. Kopelke
273 during this period at 374 sites across Central and Northern Europe. Sampling was
274 conducted in the summer months of June and/or July, during the latter stages of
275 larval development. Galler species were identified on the basis of willow host species
276 and gall morphology, as these are distinct for each sawfly species. At each site, galls
277 were randomly collected from several willow individuals in an area of about 0.1-
278 0.3 km². Most sites were visited only once, with a total of 641 site visits across the
279 374 sites. GPS coordinates were recorded for each location and the annual mean
280 temperature and annual precipitation were obtained from WorldClim. While other
281 covariates could have been considered, we figured that they are representative of
282 the most important axes of the European climate, and more easily interpretable than
283 reduced variables obtained by a PCA,

284 The methods used for rearing natural enemies from the galls have been previously
285 described by e.g. [ref, ref]. In brief, galls were opened to score the presence of galler
286 or parasitoid/inquiline larvae. Enemy larvae were classified to preliminary morphos-
287 pecies, and the identity of each morphospecies was determined by connecting them
288 to adults emerging after hibernation. The galls were reared by storing single galls in
289 small glass tubes (Kopelke 1985a, 1994a, 1999, 2003a, b). Hibernation of galls con-

taining parasitoids took place either within the glass tubes or between blotting paper in flowerpots filled with clay granulate or a mixture of peat dust and sand. These pots were stored over the winter in a roof garden and/or in a climatic chamber. In most cases, the matching of larval morphospecies with adult individuals emerging from the rearings allowed the identification of the natural enemies to the species level. Nonetheless, in some cases, individuals could only be identified to one of the (super)families Braconidae, Ichneumonidae, and Chalcidoidea. This was particularly the case when only remains of faeces, vacant cocoons of parasitoids, and/or dead host larvae were found, as was the case when parasitoids had already emerged from the gall. As a result, the largest taxon in the data set, "Chalcidoidea indeterminate," represents a superfamily of very small parasitoids that are hard to distinguish.

In total, the sampling targeted 52 Salix taxa, yielding 146,622 galls for dissection and rearing. These galls represented 96 galler species, and yielded 42,133 individually-identified parasitoids and inquilines. Of these, 25,170 (60%) could be identified to the species level. Overall, 126 parasitoid and inquiline taxa were distinguished in the material. Data on host associations within subsets of this material have been previously reported by (Kopelke ref ref b) and by Nyman et al. 2007, whereas the current study represents the first analysis of the full data set from a spatial perspective.

Analysis

Computing the probability of observing an interaction involves fitting a set of binomial models and collecting their estimated probabilities. We considered second-order generalized linear models for the sake of the illustration, but more sophisticated fitting algorithms (e.g. GAM or Random Forest) could be used as well, as long as the algorithm can provide an estimated probability for each observation. The data consists of a simple (albeit large) table with the observation of each species, X_i and X_j ,

315 their co-occurrence, X_{ij} , the observation of an interaction L_{ij} , and environmental co-
316 variates. There is one row per pair of species per site. We considered that an ab-
317 sence of a record of an interaction between co-occurring species at a site means a true
318 absence.

319 We compared three models for the co-occurrence probability. The first model fits
320 directly the co-occurrence probability conditional on the local environment, $P(X_{ijz}|E_z)$.
321 This model therefore does not make any hypothesis about the mechanisms driving
322 co-occurrence for any given environment and use the information directly available
323 in the data. The second models consider independent co-occurrence of species. In
324 this case, we fit independently $P(X_{iz}|E_z)$ and $P(X_{jz}|E_z)$ and take their product to
325 derive co-occurrence probability. This model should be viewed as a null hypothesis
326 with respect to the first model since their comparison reveals if there is significant
327 spatial association of the two species once considering the response to the environ-
328 ment (Cazelles2016). Finally the third model considers that co-occurrence probability
329 is independent of the environment and thus constant throughout the landscape. In
330 other words, $P(X_{ijz})$ is obtained by simply counting the number of observed co-
331 occurrences, divided by the total number of observations. The comparison between
332 the first and third model allows testing the hypothesis that co-occurrence is condi-
333 tional on the environment. Whenever the environment is included as a co-variate in
334 the glm, we considered a second-order polynomial response for both the temperature
335 and precipitations. There are consequently 5 parameters for the first model when
336 fitting a given pair of species, 10 parameters for the second model and only 1 for the
337 third model.

338 Following the same logic, we compared three models for the interaction prob-
339 ability. The first model fits the interaction probability conditional on the local en-
340 vironmental variables, $P(L_{ijz}|E_z, X_i, X_j)$. Consequently, the model was fitted on a

341 subset of the data, when the two species are found co- occurring. The second model
342 fits the interaction probability independently of the local environmental variables,
343 $P(L_{ijz}|X_i, X_j)$. It corresponds to the number of times the two species were observed
344 interacting when co-occurring, divided by the number of times they co-occurred.
345 The third model is an extreme case performed only to test the hypothesis that if two
346 species are found to interact at least once, then they should interact whenever they co-
347 occur, $P(L_{ijz}|X_i, X_j) = 1$. While not necessarily realistic, this model tests an hypothesis
348 that is commonly done in the representation of local networks from knowledge
349 of a deterministic metaweb (such as in Havens1992; Piechnik2008; Wood2015). There
350 are consequently 5 parameters for the first model, a single parameter for the second
351 model and no parameter to evaluate for the third model (the interaction probability
352 is fixed by hypothesis).

353 The different models were fitted to each pair of species and the fitted probabilities
354 were recorded. The joint probability $P(L_{ij}, X_i, X_j)$ was then computed from Eq. 2
355 and the likelihood of each observation was computed as $L(\theta|D) = P(L_{ij}, X_i, X_j)$ if
356 an interaction was observed and $L(\theta|D) = 1 - P(L_{ij}, X_i, X_j)$ if no interaction was
357 observed. The log- likelihood was summed over the entire dataset to compare the
358 different models by AIC. Not surprisingly, it was impossible to compute this model
359 for a very large number of pairs of species because they never co-occurred. These
360 pairs were removed from the analysis because the co-occurrence probability is null
361 and we have no information for the interaction probability. The reported likelihood
362 across the entire dataset is summed over all pairs of species that were observed co-
363 occurring at least one time. We considered separately the salix-galler and the galler-
364 parasitoid interactions.

365 Finally, we used the full model (both the co-occurrence and the interaction are
366 conditional on the environment) interpolate species distribution and the interaction

367 probability across the entire Europe. We reconstructed the expected network for each
368 location in a 1km X 1km grid. We then after computed the probabilistic connectance
369 following Poisot2015b.

370 All of the data are openly available in the database *mangal* (Poisot2015) and all R
371 scripts for querying and pre-processing the data, along with the analysis are provided
372 in supplementary material.

373 Results

374 We start with an example for a single pair of species that we selected because of a
375 sufficiently large number of times they were found co-occurring ($N_{ij} = 41$). Despite
376 the extent of the sampling, many pairs of species are found co-occurring only a few
377 times, making it hard to evaluate interaction probabilities with a reasonable confi-
378 dence interval. This particular example involves the interaction between *Phyllocolpa*
379 *plicalapporum* and *Pediobius saulius*, two fairly abundant species, observed respec-
380 tively $N_i = 53$ and $N_j = 129$ times across the 374 sites. These two species are found
381 interacting with marginal probability $P(L_{ij}) = 0.73$, which means they were found
382 interacting at 30 different locations. The model comparison (Table 1) reveals that the
383 interaction probability conditional on the co- occurrence better explain the distribu-
384 tion (Model 1 vs Model 2). The probabilistic representation of the metaweb yields
385 a much better fit to the data than the deterministic version. When the two species
386 co-occur, the occurrence of the interaction is insensitive to the environment (Model
387 2 vs Model 3). Alternatively, climatic variables significantly impact co- occurrence
388 (Model 3 vs Model 4). The neutral model performs worst than the non-random co-
389 occurrence model (Model 3 vs Model 6). The full model reveals that the greatest
390 interaction probability occurs at intermediate temperature and precipitations, simply
391 because this is where the two species are found co- occurring the most often (Fig.

392 3). The co-occurrence and the interaction probabilities could be represented in space,
393 where we find that the highest interaction probability occurs in central Europe (Fig.
394 4).

395 We did evaluated each model for all pairs of species in order to better understand
396 the large scale drivers of network turnover. Salix-gallers and gallers-parasitoids were
397 analyzed separately (Table 2). The results are comparable, albeit some very minor
398 details. We do find that across all pairs of species, the probabilistic representation of
399 interactions again does better than the deterministic (Model 1 vs Model 2). Interac-
400 tions do not happen systematically whenever the two species are found co-occurring,
401 meaning that the stochastic nature of interactions contribute to network variability
402 in addition to species turnover. There are 1077 recorded pairs of interactions, with
403 only 224 of them occurring less than 5 times. Out of these 224 interactions, only 77
404 are found systematically whenever the two species do co-occur. Even though inter-
405 actions are better represented probabilistically, the two environmental variables that
406 were considered are pretty poor predictors of their occurrence (Model 2 vs Model
407 3). Not surprisingly, the likelihood increases for both types of interactions when the
408 environment is considered. The extra number of parameters however exceed the gain
409 in likelihood, and therefore the best model excludes the effect of the environment.

410 According to the log-likelihood only, the co-occurrence is non-neutral for salix-
411 galler interactions, while it is neutral for the galler-parasitoid interaction. However,
412 the gain in log-likelihood for the neutral model of galler- parasitoid co-occurrence is
413 inferior to the extra number of parameters (twice as many since two species distribu-
414 tion models are fitted instead of just one), which has for consequence that the best
415 model according to AIC has non-random co-occurrence (Model 3 vs Model 6), for
416 both types of interactions.

417 The approach we present not only has implications for understanding the bio-

418 geography of pairwise interactions and interaction networks, but also the quality of
419 the evaluation of metawebs. We investigated the reliability of the estimated metaweb
420 across the entire dataset. As mentionned above, across the 32 412 pairs of species,
421 only 1077 pairs are interacting at least at a single location, for a connectance of 0.03.
422 However, only 8437 species are found co- occurring at least one time across all lo-
423 cations. There are consequently 23975 gaps of information in the metaweb (74.0%
424 - see Fig. 5). Given that we do not know if the non co-occurring species do in-
425 deed co-occur, it means that a more appropriate estimate of connectance would be
426 $C = 1077/8437 = 0.128$. This result reveals that the evaluation of the sampling
427 quality of ecological networks is a problem on its own that worths further attention.

428 Once we selected the best model (model 3, Table 2), we were able to reconstruct the
429 expected species richness across Europe, along with the most likely network for each
430 location, and therefore map the expected distribution of network properties (Fig. 6).
431 We simply considered connectance, as it could be easily computed from probabilistic
432 networks (Poisot2015b) and is also a good proxy for many other network properties
433 (Poisot2014). The diversity of *Salix* tends to increase toward boreal areas, and we
434 consequently find a peak in diversity in northern Europe. The distribution of the
435 expected number of interactions follows the distribution of species richness, but not
436 at the same rate. Consequently, connectance is peaking in central Europe and in
437 England.

438 **Discussion**

439 **Summary of the framework**

440

441 **Toward interaction distribution models**

442

443 **What are the drivers of network variation in space?**

444

445 **Forecasting network structure under global change**

446

447 **Investigating the realized niche and the impact of biotic interactions on distri-**
448 **bution**

449

450 **Guidance for empirical studies**

451

452 Conclusion

453 Research agenda on probabilistic interactions

454

455 • Need a new type of data

456 • Need to investigate the reliability of interactions

457 • From interaction probabilities to a distribution of network properties

458 • Deeper investigations of the environmental dependence of interactions

459 • Trait-based approach to interactions

460 • The effect of interaction on co-occurrence

⁴⁶¹ Acknowledgements

⁴⁶² This is a contribution to the working groups *Networks over ecological gradients* (Santa
⁴⁶³ Fe Institute) and *Continental-scale variation of ecological networks* (Canadian Institute for
⁴⁶⁴ Ecology and Evolution).

465 **Figure legends**

466 **Figure 1**

467 **Non-random sampling of the metaweb.** Network assembly could be viewed as a
468 sampling process of the regional pool of potential interactions. Species are sampled
469 first (indicated by colored nodes) and among the present species in the local network,
470 only some interactions are occurring (indicated by colored links). The challenge we
471 address with the quantitative framework proposed is to adequately characterize this
472 sampling process. The sampling of the metaweb is illustrated with a local interaction
473 network among Salix, gallers and parasitoids. Here, the metaweb was constructed
474 by aggregating observed interactions across 374 local networks. The color nodes
475 represent the species that were found in the most diverse of these 374 local networks.

476 **Figure 2**

477 **Visual representation of the integrated niche.** We represent visually the integration
478 between two views of the niche. In biogeography, the niche is considered the set of
479 environmental conditions where the intrinsic growth rate r is positive. The horizontal
480 axis represents an environmental gradient impacting the growth of the focal species
481 (in red). The location of the different species along this gradient represent their op-
482 timum, and the vertical dotted lines represent the limits of the [Grinnelian] niche of
483 the focal species. In food web ecology, the [Eltonian] niche represents the location of
484 a species in the food web, as determined by its preys and its predators. The vertical
485 axis represents a niche gradient, presumably a trait such as body size. The location of
486 each species along this gradient represent their niche position. The focal species will
487 feed on the different preys whose niche location falls within a given interval around
488 the optimum, represented by the horizontal dotted lines. The integrated niche [Grin-

nelian & Eltonian] corresponds to the square in the middle where an interaction is possible. According to our probabilistic framework, the central square represents the area where the joint probability of observing interactions and co-occurrence is positive.

Figure 3

Probabilistic representation of the interaction probability between a leaf galler (*Phyllocolpa plicalapporum*) and a parasitoid (*Pediobius saulius*) across a temperature and a precipitation gradient. The representation is based on predictions from model 3 (see Table 1). For the left panel: open circles represent the absence of both species or of an interaction, the closed circles represent co-occurrence and other symbols the occurrence of only one of the two species. For the other two panels the open circles represent co-occurrence but an absence of interaction and the closed circles represent the occurrence of an interaction.

Figure 4

Probabilistic representation of the interaction probability between a leaf galler (*Phyllocolpa plicalapporum*) and a parasitoid (*Pediobius saulius*) across Europe.

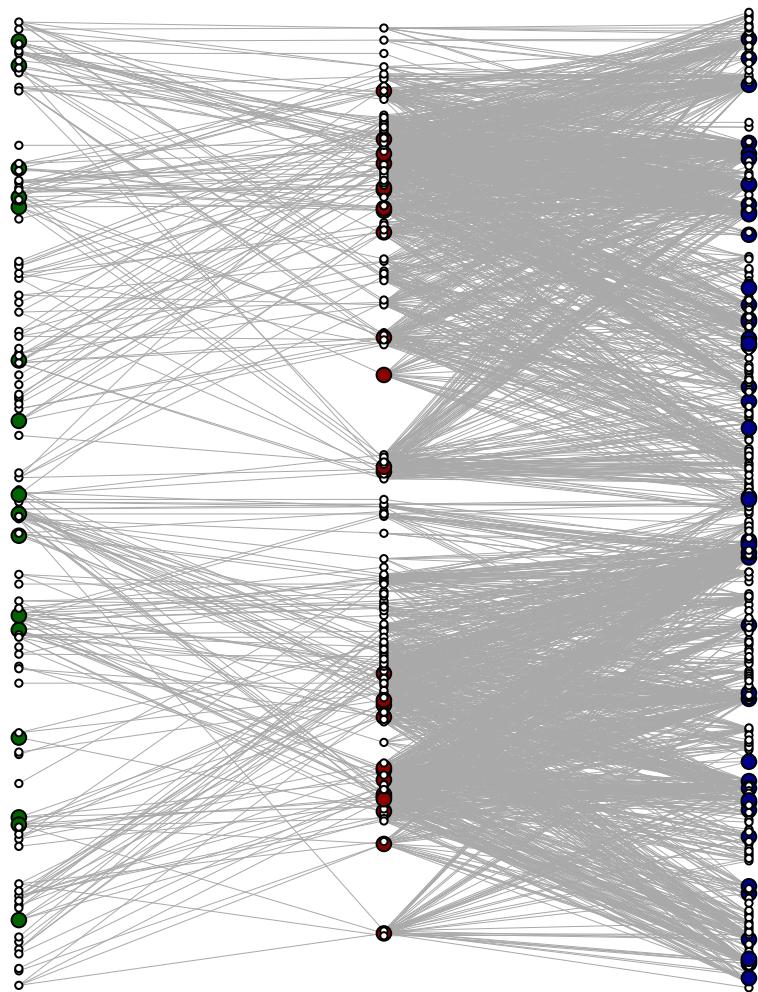
Figure 5

Representation of the A) metaweb and B) gaps of data in the metaweb. The salix and gallers were regrouped as 'victims' for the sake of the illustration, as the gallers and the parasitoids were regrouped as 'ennemis'.

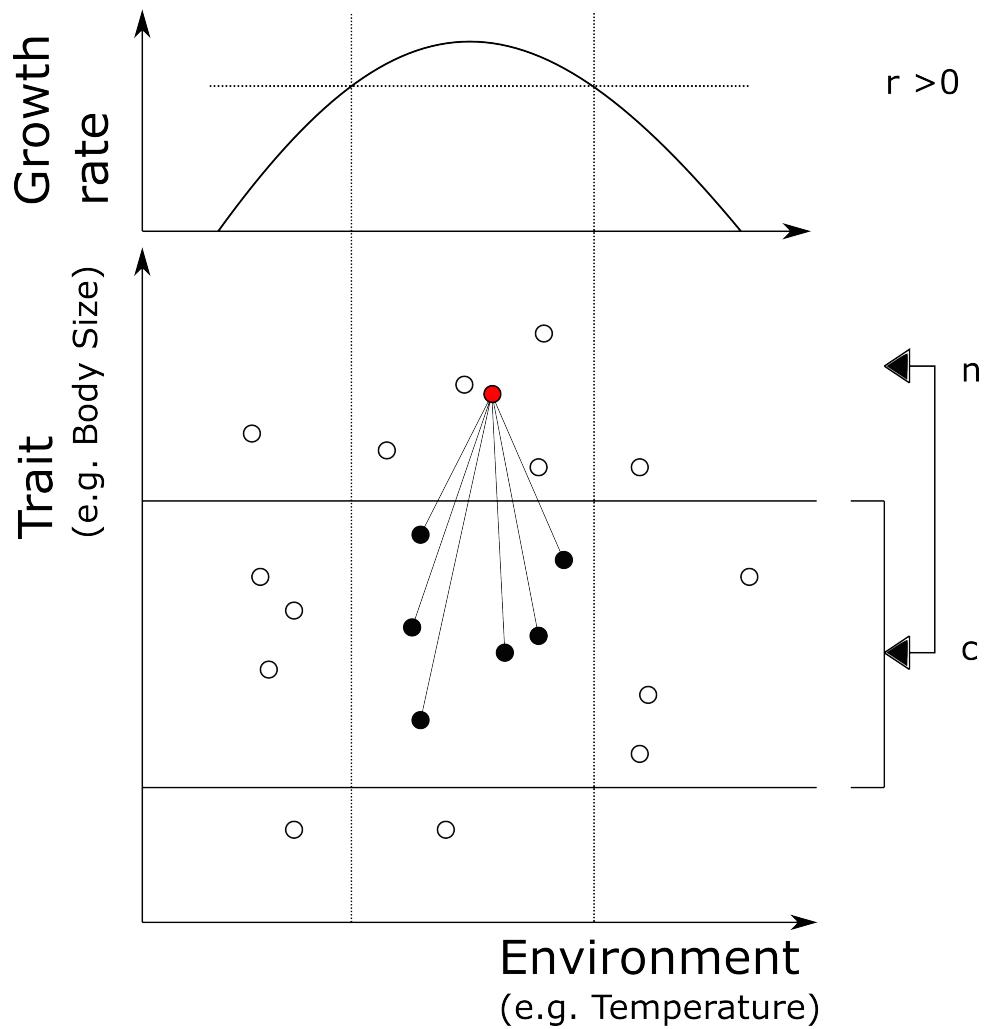
509 **Figure 6**

510 **Mapping the distribution of species richness, the number of links and connectance**
511 **across Europe.** The representation is based on predictions from model 3 (see Table
512 2).

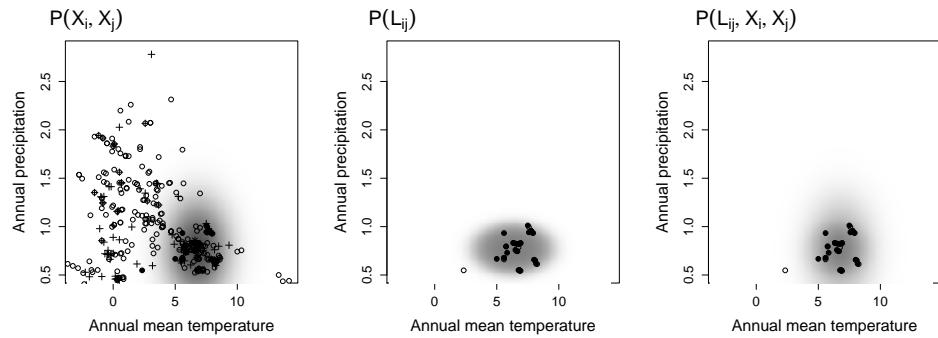
513 **Figure 1**



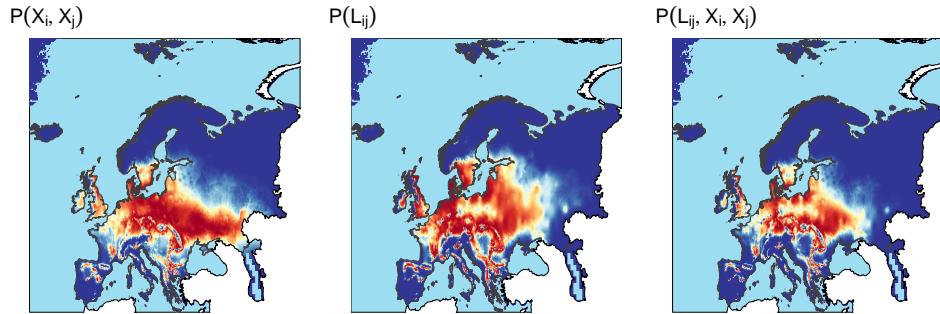
514 **Figure 2**



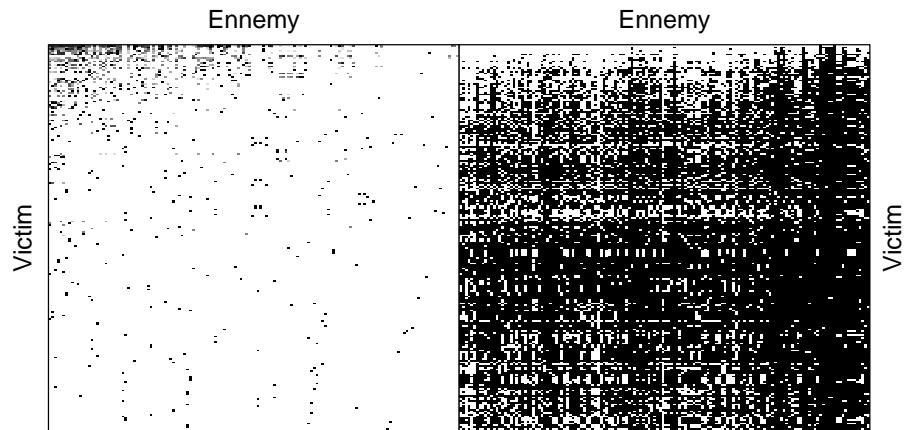
515 **Figure 3**



516 **Figure 4**



517 **Figure 5**



518 **Figure 6**

