

<sup>1</sup> **Title:** Bringing Elton and Grinnell together: a quantitative framework to represent  
<sup>2</sup> the biogeography of ecological interaction networks

<sup>3</sup>

## <sup>4</sup> Abstract

<sup>5</sup> Biogeography has traditionally focused on the spatial distribution and abundance of  
<sup>6</sup> species. Both are driven by the way species interact with one another, but only re-  
<sup>7</sup> cently community ecologists realized these interactions vary across time and space and  
<sup>8</sup> need to be documented. Here, we call for an integrated approach, adopting the view  
<sup>9</sup> that community structure is best represented as a network of ecological interactions, and  
<sup>10</sup> show how it translates to biogeography questions. We propose that the ecological niche  
<sup>11</sup> should encompass the effect of the environment on species distribution (the Grinnellian  
<sup>12</sup> dimension of the niche) and on the ecological interactions among them (the Eltonian  
<sup>13</sup> dimension). Starting from this concept, we develop a quantitative theory to explain  
<sup>14</sup> turnover of interactions in space and time – *i.e.* a novel approach to interaction distribu-  
<sup>15</sup> tion modeling. We apply this framework to host–parasite interactions across Europe and  
<sup>16</sup> find that two aspects of the environment (temperature and precipitation) exert a strong  
<sup>17</sup> imprint on species co-occurrence, but not on species interactions. Even where species co-  
<sup>18</sup> occur, interaction proves to be stochastic rather than deterministic, adding to variation  
<sup>19</sup> in realized network structure. We also find that a large majority of host-parasite pairs  
<sup>20</sup> are never found together, thus precluding any inferences regarding their probability to in-  
<sup>21</sup> teract. This first attempt to explain variation of network structure at large spatial scales  
<sup>22</sup> opens new perspectives at the interface of species distribution modeling and community  
<sup>23</sup> ecology.

<sup>24</sup> **Keywords:** networks, spatial ecology, co-occurrence, probability of interaction

<sup>25</sup>

## <sup>26</sup> Introduction

<sup>27</sup> Community ecology is *the study of the interactions that determine the distribution and*  
<sup>28</sup> *abundance of organisms* (Krebs, 2009). Despite a general consensus on this definition  
<sup>29</sup> (Scheiner & Willig, 2007), research on variation in community structure (beta-diversity)  
<sup>30</sup> has mostly focused on the spatial and temporal turnover of species composition (Anderson  
<sup>31</sup> *et al.*, 2011). Such research usually neglects variation in the way species interact with  
<sup>32</sup> each other, despite accumulating empirical evidence that is a major source of diversity  
<sup>33</sup> (Poisot *et al.*, 2015). Given this omission, it is perhaps not surprising that biogeographers  
<sup>34</sup> are still struggling to establish whether interactions actually impacts the distribution of  
<sup>35</sup> species at large spatial scales (Kissling *et al.*, 2012). An interaction is conceived as  
<sup>36</sup> the direct effect of the action of one species on the demography on another species;  
<sup>37</sup> whether it scales up and impact the (co-)distribution of species remains matter of debate  
<sup>38</sup> at the moment (Wisz *et al.*, 2013; González-Salazar *et al.*, 2013; Cazelles, 2016; Harris,  
<sup>39</sup> 2016; Godsoe *et al.*, 2017). Treating interactions as fixed events nonetheless neglects a  
<sup>40</sup> large part of the complexity of empirical communities, and will most likely deliver biased  
<sup>41</sup> metrics (Poisot *et al.*, 2016b). Recent attempts at accounting for interactions in species  
<sup>42</sup> distribution models (Pollock *et al.*, 2014; Pellissier *et al.*, 2013; Ovaskainen *et al.*, 2017)  
<sup>43</sup> have brought some methodological advances, but are not sufficient for two reasons. First,  
<sup>44</sup> these techniques are still based on a ‘species-based’ approach to communities, where  
<sup>45</sup> interactions are merely treated as fixed covariates affecting distribution. Second, they fail  
<sup>46</sup> to provide a conceptual step forward, both in their treatment of interactions and in the  
<sup>47</sup> quality of the predictions they make.

<sup>48</sup> Network approaches offer a convenient representation of communities because they  
<sup>49</sup> simultaneously account for species composition and their interactions. Species are repre-  
<sup>50</sup> sented as nodes, so that networks already encompass all the information used by current  
<sup>51</sup> approaches of species distribution modeling; in addition, interactions are represented

52 by links, so that networks provide additional information on community structure. To  
53 date, studies of network diversity have mostly been concerned with the distribution of  
54 interactions within locations, and less so with variation among locations (Dunne, 2006;  
55 Bascompte & Jordano, 2007; Ings *et al.*, 2009; Kéfi *et al.*, 2012). There is, however, am-  
56 ple evidence that interaction networks vary in space and time (Laliberté & Tylianakis,  
57 2010; Poisot *et al.*, 2012; Schleuning *et al.*, 2012; Albouy *et al.*, 2014; Poisot *et al.*, 2016c;  
58 Trøjelsgaard *et al.*, 2015), even though there is no common framework with which to gen-  
59 eralize these results. Metacommunity theory provides explanations for variation in the  
60 distribution of the nodes (Gravel *et al.*, 2011; Pillai *et al.*, 2011; Cazelles *et al.*, 2015), but  
61 there are no such explanation to the variation of node and link occurrences. Consequently,  
62 we need theory to formalize these observations, as it is the only way towards fulfilling  
63 the goal of community ecology: providing cogent predictions about, and understanding  
64 of, the structure of ecological communities.

65 Given the historically different approaches to modelling the distributions of species  
66 vs. interactions, there is a need to bring the two together. Here, we offer an integrated  
67 approach to do so, adopting the view that community structure is best represented as  
68 a network of ecological interactions. Based on this idea, we propose a description of  
69 the ecological niche that integrates the effect of the environment on species distribution  
70 and on the ecological interactions among them. Building on this concept, we develop  
71 a quantitative theory to explain turnover of interactions in space and time. We first  
72 present the concept and then formalize it mathematically, using a probabilistic model  
73 to represent the sampling of the regional pool of interactions. At the level of species  
74 pairs, the statistical approach could be conceived as an interaction distribution model.  
75 At the community level, the approach provides a likelihood-based method to compare  
76 different hypotheses of network turnover. As an illustrative example, we apply this novel  
77 framework to a large data set on host-parasite interactions across Europe and find that

78 two aspects of the environment (temperature and precipitation) exert a strong imprint  
79 on species co-occurrence, but not on species interactions. The network structure changes  
80 systematically across the latitudinal gradient, with a peak of connectance at intermediate  
81 latitudes.

## 82 The two dimensions of community structure

83 The problem of community assembly is often formulated as *how are species sampled from*  
84 *a regional pool to constitute a local community* (Götzenberger *et al.*, 2012)? This question  
85 could be rewritten to address the problem of network assembly, as *how do samples from*  
86 *a regional pool of interactions constitute a local interaction network?* An illustration of  
87 this problem for a food web is provided in Fig. 1. The regional pool of interactions,  
88 the *metaweb*, represents potential interactions among all species that could be found in a  
89 given area. In this particular case, there are 275 nodes, and 1173 links among the plants  
90 (52 nodes), herbivores (96 nodes), and parasitoids (127 nodes) from Northern Europe. An  
91 instance of a local community is also illustrated, with 45 nodes and 93 interactions. Only  
92 55.0% of all potential interactions (plant-herbivore or herbivore-parasitoid combinations)  
93 are realized in this local network, revealing the stochastic nature of ecological interactions.  
94 Our objective here is to provide theory to explain the sampling of the regional pool of  
95 interactions, along with a quantitative method to predict it. The problem could be  
96 formalized sequentially by understanding first why only a fraction of the species co-occur  
97 locally and second why these species do or do not interact.

98 There are multiple causes of spatial turnover of species co-occurrence. The first and  
99 most-studied driver is the effect of variation in the abiotic environment on species perfor-  
100 mance. Combined with specific responses in demography, it generates variation among  
101 sites by selecting the locally fittest species (Leibold *et al.*, 2004). Stochasticity plays

an additional role, either because colonization and extinction events (Hanski, 1999) are inherently unpredictable or because strong non-linear feedbacks in community dynamics generate alternative transients and equilibria (Chase, 2007; Vellend *et al.*, 2014). Interaction themselves may impact co-distribution, as hypothesized long ago by Diamond (1975). Analyses of community turnover are usually performed with data represented in a table with rows corresponding to sites (or measurements) and columns to species. Metrics of beta diversity quantify the variance of this community data (Legendre *et al.*, 2005). Traditional approaches rely on measures of dissimilarity among communities, such as the Jaccard or Bray–Curtis indices. More recent approaches decompose total variation of the community data into species and site contributions to beta diversity (Legendre & De Cáceres, 2013), and further partition it into dissimilarity due to changes in species richness and dissimilarity due to actual species turnover (Baselga, 2010). Even though all of these methods compare whole lists of species among sites or measurements, they remain fundamentally "species-based", since they report variation within columns. None of them explicitly considers variation of associations (i.e., of pairs or higher-order motifs – Stouffer *et al.* 2007). Some previous studies have considered how species distribution could be influenced by the joint effects of the abiotic and biotic environment (Stephens & Heau, 2009; González-Salazar *et al.*, 2013; Cazelles *et al.*, 2015; Ovaskainen *et al.*, 2017), here we inverse the problem and describe how the distribution of biotic interactions is influenced by species distribution and the environment.

We are now getting a better understanding of interaction turnover. As mentioned above, in the network approach to community structure, species and interactions are represented by nodes and links, respectively. Associations can also be represented by matrices in which entries represent the occurrence or intensity of interactions among species (rows and columns). Network complexity is then computed as the number of interactions (in the case of binary networks) or interaction diversity (in the case of quantitative net-

128 works, Bersier *et al.* 2002). Variability in community structure consequently arises from  
129 the turnover of species composition, along with turnover of interactions among pairs of  
130 species. The occurrence and intensity of interactions could vary because of the environ-  
131 ment, species abundance, and higher-order ecological interactions (Poisot *et al.*, 2015).  
132 Variation in community composition can be independent of variation of ecological inter-  
133 actions, suggesting that species and interaction distribution may well respond to different  
134 drivers (Poisot *et al.*, 2012).

135 The "niche" is by far the dominant concept invoked to explain species distributions  
136 and community assembly, from the local to the global scale. Following Hutchinson (1957),  
137 the niche is viewed as the set of environmental conditions allowing a population to es-  
138 tablish and persist (see also Holt 2009). In other words, the niche is the location in  
139 that multidimensional space allowing a species to have a positive growth rate when rare  
140 (Godsoe *et al.*, 2017). Community turnover arises as a result of successive replacement of  
141 species along an environmental gradient, in agreement with the Gleasonian view of com-  
142 munities (Gleason, 1926). The concept is straightforward to put into practice with species  
143 distribution models, as it maps naturally on available distributional and environmental  
144 data. Consequently, a vast array of statistical tools have been developed to implement it  
145 (e.g. BIOMOD Thuiller 2003, MaxEnt Phillips *et al.* 2006). It is however much harder  
146 to account for ecological interactions within this approach (Townsend *et al.*, 2011). As  
147 such, these interactions are often viewed as externalities constraining or expanding the  
148 range of environmental conditions required for a species to maintain a viable population  
149 (Pulliam, 2000; Soberón, 2007).

150 Interestingly, the food web literature also has its own "niche model" to position a  
151 species in a community (Williams & Martinez, 2000), generalized later to other types of  
152 interaction networks (Eklöf *et al.*, 2013). The niche of a species in this context represents  
153 the multidimensional space of all of its interactions. Each species is characterized by a

niche position, an optimum and a range over three to five different niche axes (Williams & Martinez, 2000; Eklöf *et al.*, 2013). The niche model of food web structure and its variants have successfully explained the complexity of a variety of networks, from food webs to plant-pollinator systems (Allesina *et al.*, 2008; Williams *et al.*, 2010; Eklöf *et al.*, 2013). This conceptual framework is, however, limited to local communities, and does not provide any explanation for the turnover of network structure along environmental gradients.

## The integrated niche

Despite several attempts to update the concept of the ecological niche, ecologists have not moved far beyond the "n-dimensional hypervolume" defined by Hutchinson (1957). Despite its intuitive interpretation and easy translation into species distribution models (Boulangeat *et al.*, 2012; Blonder *et al.*, 2014), the concept has been frequently criticized (Hardin, 1960; Peters, 1991; Silvertown, 2004), and several attempts have been made to expand and improve it (Pulliam, 2000; Chase & Leibold, 2003; Soberón, 2007; Holt, 2009; McInerny & Etienne, 2012b).

Part of the problem surrounding the niche concept has been clarified with the distinction between Eltonian and Grinnellian definitions (Chase & Leibold, 2003). The Grinnellian dimension of the niche is the set of abiotic environmental conditions required for a species to maintain a population in a location. The Grinnellian niche is intuitive to apply, and constitutes the conceptual backbone of species distribution models. The Eltonian niche, on the other hand, is 'the place of a species in its biotic environment, its relations to food and enemies'. While this aspect of the niche is well known by community ecologists, it is trickier to turn into predictive models. Nonetheless, the development of the niche model of food web structure (Williams & Martinez, 2000) and its parameteri-

178 zation using functional traits (Gravel *et al.*, 2013; Bartomeus *et al.*, 2016) made it more  
179 operational.

180 These perspectives are rather orthogonal to each other, and this has resulted in con-  
181 siderable confusion in the literature (McInerny & Etienne, 2012a). Chase & Leibold 2003  
182 attempted to reconcile with the following definition: "*[The niche is] the joint description*  
183 *of the environmental conditions that allow a species to satisfy its minimum requirements*  
184 *so that the birth rate of a local population is equal to or greater than its death rate along*  
185 *with the set of per capita effects of that species on these environmental conditions*". Their  
186 representation merges zero-net-growth isolines delimiting the Grinnellian niche ("when  
187 does the population persist?") with impact vectors delimiting the Eltonian niche ("what is  
188 the per-capita impact?"). While this representation has been very influential in local-scale  
189 community ecology (the resource-ratio theory of coexistence, Tilman 1982), it remains  
190 impractical at larger spatial scales because of the difficulties in measuring it. The absence  
191 of any mathematical representation of the niche that can be easily fit to ecological data  
192 may explain why biogeographers are still struggling to develop species distribution mod-  
193 els that also consider ecological interactions. Thus, a more integrative description of the  
194 niche will be key to understand spatial and temporal turnover in community structure.

195 We propose to integrate the two perspectives of the niche using a visual representa-  
196 tion of both components (Fig. 2). The underlying rationale is that, in addition to the  
197 environmental constraints on demographic performance (Fig. 2, top panel), any organism  
198 requires resources to meet its metabolic demands and to sustain reproduction (Fig. 2,  
199 nodes in network of bottom panel). Abiotic environmental axes are any non-consumable  
200 factors affecting the demographic performance of an organism. Alternatively, the resource  
201 axes are traits of the resources that allow interactions with the consumers. By definition  
202 of what an interaction is, these axes also influence the growth rate. The niche can there-  
203 fore be viewed as the set of abiotic environmental conditions (the Grinnellian component)

204 along with the set of traits (the Eltonian component) that allow a population to establish  
205 and to grow at a location. This visual representation make operational the theoretical  
206 interpretation of the niche and species ranges by Godsoe *et al.* (2017). Accordingly, each  
207 species can be characterized by an optimal position along both the environmental (x-  
208 axis) and the trait (y-axis) plane. The integrated niche is then the hypervolume where  
209 interactions can occur and where a population has a positive growth rate.

210 This approach changes the representation of the niche, putting species distributions  
211 and ecological interactions into the same formalism. Moreover, it allows the limits of  
212 the niche axes to be independent of each other (as in the example in Fig. 2), or to  
213 interact. Some of these axes may not be independent. For instance, the optimal prey  
214 size for predatory fishes could decline with increasing temperature (Gibert & DeLong,  
215 2014), which would make diet boundaries functions of the environment. Alternatively,  
216 we could also consider that the growth rate of the predator changes with the size of its  
217 prey items, thereby altering the environmental boundaries and changing the shape of the  
218 niche in the middle of this visual representation. It is also important that this conceptual  
219 representation depicts a reality that could be highly dimensional and sometimes difficult  
220 to represent statistically (Clark *et al.*, 2007).

## 221 A probabilistic representation of interaction networks in 222 space

223 We now formalize the integrated niche with a probabilistic approach to interactions and  
224 distributions. In particular, we seek to represent the probability that an interaction  
225 between species  $i$  and  $j$  occurs at location  $y$ . We define  $L_{ijy}$  as a stochastic process taking  
226 a value of 1 when an interaction occurs and a value of 0 when it does not, and focus on  
227 the probability that this event occurs,  $P(L_{ijy} = 1)$ , over a given sampling interval and

space. We note that the occurrence of an interaction is dependent on the co-occurrence of species  $i$  and  $j$ . This argument might seem trivial at first, but the explicit consideration of this condition in the probabilistic representation of ecological interactions will prove instrumental to understanding their variation. We note this phenomenon we aim to describe differs fundamentally from the effect that interactions do have on co-occurrence (Cazelles *et al.*, 2016). We similarly define  $X_{iy}$  as a stochastic process representing the occurrence of species  $i$  at location  $y$ , taking a value of 1 when the species occurs and a value of 0 when it is absent. The quantity we seek to understand is the probability of a joint event, conditional on the set of environmental conditions  $E_y$ :

$$P(X_{iy}, X_{jy}, L_{ijy} | E_y) \quad (1)$$

Or simply said, the probability of observing both species  $i$  and  $j$  plus an interaction between  $i$  and  $j$  given the conditions  $E_y$  at location  $y$ . This probability could be decomposed into two parts using the product rule of probabilities:

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

The first term on the right-hand side of the equation is the probability of observing the two species co-occurring at location  $y$ . It corresponds to the Grinnellian dimension of the niche. The second term represents the probability that an interaction occurs between species  $i$  and  $j$ , given that they are co-occurring. This predicate can be refined using information on trait distribution and trait matching rules (Bartomeus *et al.*, 2016). In short, here we avoid specific references to traits and considered that interactions are provided in the data. But one could develop the model further to integrated them, inspired by the framework in Gravel *et al.* 2016. They proposed to represent the occurrence of an interaction between species as a joint probability event  $P(L_{ij}, \mathbf{T}_i, \mathbf{T}_j | E)$ , where  $\mathbf{T}_i$  and

249  $\mathbf{T}_j$  are trait vectors for species  $i$  and  $j$  respectively. Then after, one may also consider  
250 how these traits are related to the environment and co-occur in order to develop a trait-  
251 based representation of the spatial variation of interactions. Above, we referred to this  
252 entity as the "metaweb" and it corresponds to the Eltonian dimension of the niche.  
253 Below, we will see how this formalism can be directly fit to empirical data, restricting  
254 ourselves to occurrences only and referring to the above citations for trait-matching and  
255 trait-environment relationships. But before turning to an application, we will discuss the  
256 interpretation of different variants of these two terms

257 **Variants of co-occurrence**

258 There are several variants to the co-occurrence probability, representing different hy-  
259 potheses concerning spatial variation in network structure (see the explicit formulations  
260 in Table 1). The simplest model relates the probability of co-occurrence directly to the  
261 environment,  $P(X_{iy}, X_{jy}|E_y)$ . In this situation, there are no underlying assumptions  
262 about the ecological processes responsible for co-occurrence. Spatial associations be-  
263 tween species could arise because interactions constrain distribution, where in such case  
264 the co-occurrence would be conditional on  $L$ , or, alternatively, because of environmental  
265 requirements shared between  $i$  and  $j$  (Pollock *et al.*, 2014; Cazelles *et al.*, 2016). In the  
266 former case, species are not independent of each other and the conditional occurrence  
267 must be accounted for explicitly,  $P(X_{iy}, X_{jy}|E_y) = P(X_{iy}|E_y, X_{jy})P(X_{jy}|E_y)$ . In the lat-  
268 ter case, species are independent, and only the marginal occurrence must be accounted  
269 for,  $P(X_{ijy}|E_y) = P(X_{iy}|E_y)P(X_{jy}|E_y)$ .

270 The co-occurrence probability itself could depend on ecological interactions. This  
271 should be viewed as the realized component of the niche (i.e. the distribution when  
272 accounting for species interactions). Direct pairwise interactions such as competition,  
273 facilitation, and predation have long been studied for their impact on co-distribution

(e.g. Diamond 1975; Connor & Simberloff 1979; Gotelli 2000). Second- and higher-order interactions (e.g. trophic cascades) could also affect co-occurrence (Harris, 2016; Staniczenko *et al.*, 2017). Co-occurrence of multiple species embedded in ecological networks is a topic of its own, however, and is influenced by both network topology and species richness (Cazelles *et al.*, 2016). Not only direct interactions influence co-occurrence, but indirect interactions do as well (e.g. plant species sharing an herbivore, or herbivores sharing parasitoids, could repel each other in space (Holt & Lawton, 1993)). The impact of direct interactions and first-order indirect interactions on co-occurrence tends to vanish with increasing species richness in the community (Cazelles *et al.*, 2016). Further, co-occurrence is also influenced by the covariance of interacting species to an environmental gradient (Cazelles *et al.*, 2015).

The interpretation that interactions may impact co-occurrence, and that the realization of an interaction requires co-occurrence are not mutually exclusive. But since here the object is to describe the variation in the occurrence of interactions, which absolutely requires co-occurrence, we will give attention to this part of the problem. Relating co-occurrence to the structure of interaction networks is a problem on its own (and vice-versa). We will therefore focus here on the variation of interactions and not on their distribution, and leave this specific issue for the Perspectives section and future research.

## Variants of the metaweb

There are also variants of the metaweb. First, most documented metawebs have thus far considered ecological interactions to be deterministic, rather than probabilistic (e.g. Havens 1992; Wood *et al.* 2015). Species are assumed to interact whenever they are found together in a location, independent of their local abundance and the local environment. In other words,  $P(L_{ijy}|X_{iy} = 1, X_{jy} = 1) = 1$ . This approach might be a reasonable approximation if the spatial or temporal scale of sampling and inference is so large that

299 the probability of observing at least one interaction converges to unity. In this scenario,  
300 network variation arises solely from species distributions.

301 Second, ecological interactions could also vary with the environment, so that  $P(L_{ijy}|E_y)$ .  
302 Although it is rare to see a conditional representation of pairwise ecological interactions,  
303 experimental studies have frequently revealed interactions to be sensitive to the environ-  
304 ment. For instance, McKinnon *et al.* (2010) showed that predation risks of shorebirds  
305 vary at the continental scale, decreasing from the south to the north. It is also common  
306 to see increasing top-down control with temperature (e.g. Shurin *et al.* 2012; Gray *et al.*  
307 2016). Effects of the environment on interactions (Gibert & DeLong, 2014) also propagate  
308 up the community and influence network structure (Tylianakis *et al.*, 2007; Woodward  
309 *et al.*, 2010; Petchey *et al.*, 2010).

## 310 **Application: continental-scale variation of host-parasite 311 community structure**

312 We now turn to an illustration of our theory with the analysis of an empirical dataset  
313 of host-parasite networks sampled throughout the south-north environmental gradient in  
314 continental Europe (Kopelke *et al.*, 2017). Our objective here is to illustrate potential  
315 applications of the approach outlined above, rather than to specifically describe the bio-  
316 geography of this particular community. The focal system consists of local food webs of  
317 willows (genus *Salix*), their galling insects, and the natural enemies (parasitoids and in-  
318 quilines) of gallers. We ask: i) how much does network structure vary across the gradient,  
319 and ii) what is the primary driver of network turnover across the gradient?

320 **Data**

321 Communities of willows, gallers, and parasitoids are species-rich and widely distributed,  
322 with pronounced variation in community composition across space. The genus *Salix* in-  
323 cludes over 400 species, most of which are shrubs or small trees (Argus, 1997), and is  
324 common in moist habitats across the Northern Hemisphere (Skvortsov, 1999). Willows  
325 support a highly diverse community of herbivorous insects, with one of the main herbivore  
326 groups being gall-inducing sawflies (Hymenoptera: Tenthredinidae: Nematinae: Euurina  
327 (Kopelke, 1999). Gall formation is induced by sawfly females during oviposition, and  
328 includes marked manipulation of host-plant chemistry by the galler (Nyman & Julkunen-  
329 Tiitto, 2000). The enemy community of the gallers includes nearly 100 species belonging  
330 to 17 insect families of four orders (Kopelke, 2003). These encompass two main types:  
331 inquiline larvae (Coleoptera, Lepidoptera, Diptera, and Hymenoptera) feed primarily on  
332 gall tissue, but typically kill the galler larva in the process, while parasitoid larvae (repre-  
333 senting many families in Hymenoptera) kill the galler larvae by direct feeding (Kopelke,  
334 2003).

335 Local realizations of the willow-galler-parasitoid network were reconstructed from  
336 community samples collected between 1982 and 2010. During this period, willow galls  
337 were collected at 370 sites across Central and Northern Europe. In total, 52, 96 and  
338 127 *Salix*, galler and parasitoid and inquiline taxa were distinguished, respectively. The  
339 strength of this dataset is that observations were observed *in situ*, rather than inferred  
340 from expert knowledge or other sources of information, thereby allowing the analysis of  
341 their spatial variation. The drawback is that, because many species are rare, some of the  
342 pairs of species may have been observed co-occurring only a few times, or never. As a  
343 consequence, despite the extent of the sampling, there is significant uncertainty in the  
344 quantification of some of the links. The above described methodology explicitly aims at  
345 describing this uncertainty, and research should be done to develop methods to reduce it

346 (see Conclusion).

347 The current study represents the first analysis of the full data set from a spatial  
348 perspective. Full details about data collection and species identification are provided  
349 in (Kopelke *et al.*, 2017). Annual mean temperature and precipitation were obtained  
350 from WorldClim using the R package dismo (Hijmans, 2015) and GPS coordinates of the  
351 sampling locations. While other covariates could have also been considered to represent  
352 the high dimensionality of the niche, these two variables are likely representative of the  
353 most important axes of European climate, and are also more easily interpretable than  
354 reduced variables obtained, for example, by principal component analysis.

## 355 Analysis

356 Computing the probability of observing an interaction involves fitting a set of binomial  
357 models and collecting their estimated probabilities. For the sake of illustration, we con-  
358 sidered generalized linear models – although more flexible fitting algorithms (e.g. GAM  
359 or Random Forest) could equally well be used, as long as the algorithm can estimate the  
360 probability for each observation. The data consist of a simple (albeit large and sparse)  
361 table with the observation of each species,  $X_{iy}$  and  $X_{jy}$ , their co-occurrence,  $X_{ijy}$ , the  
362 observation of an interaction  $L_{ijy}$ , and environmental co- variates  $E_y$ . Thus, there is  
363 one row per pair of species per site. We considered that an absence of a record of an  
364 interaction between co-occurring species at a site means a true absence (see below for a  
365 discussion on this issue).

366 We compared three models for the co-occurrence probability. The first one directly  
367 represents the co-occurrence probability conditional on the local environment,  $P(X_{iy}, X_{jy}|E_y)$   
368 (models are listed in Tables 1 and 2). Hence, this model makes no assumption about the  
369 mechanisms driving co- occurrence for any given environment, and instead uses the in-  
370 formation directly available in the data. It thereby indirectly accounts for the effect of

interactions on co-occurrence, if there are any. The second model considers independent occurrence of species. In this case, we independently fit  $P(X_{iy}|E_y)$  and  $P(X_{jy}|E_y)$ , and we then take their product to derive the probability of co-occurrence. This model should be viewed as a null hypothesis with respect to the first model, since a comparison between the respective models will reveal if there is significant spatial association of the two species beyond a joint response to the shared environment (Cazelles *et al.*, 2016). Finally, the third model assumes that the probability of co-occurrence is independent of the environment and thus constant throughout the landscape. In other words,  $P(X_{iy}, X_{jy})$  is obtained by simply counting the number of observed co-occurrences divided by the total number of observations. Thus, the comparison between the first and third model allows us to test the hypothesis that co-occurrence is conditional on the environment. Whenever the environment was included as a covariate in the GLM, we considered a second-order polynomial term for both temperature and precipitation in order to account for optima in environmental conditions. There are consequently five parameters for the first model when fitting a given pair of species, 10 parameters for the second, and only one for the third model.

Following the same logic, we compared three models of the interaction probability. The first model conditions the interaction probability on the local environmental variables,  $P(L_{ijy}|X_{iy}, X_{jy}, E_y)$ . Consequently, the model was fit to the subset of the data where the two species co-occur. The second model fits the interaction probability independently of the local environmental variables,  $P(L_{ijy}|X_{iy}, X_{jy})$ . It corresponds to the number of times the two species were observed to interact when co-occurring, divided by the number of times that they co-occurred. The third model is an extreme case evaluated only to test the hypothesis that if two species are found to interact at least once, then they should interact whenever they co-occur,  $P(L_{ijy}|X_{iy}, X_{jy}) = 1$ . While not necessarily realistic, this model tests an assumption commonly invoked in the representation of local networks

397 from the knowledge of a deterministic metaweb. It represents potential interactions,  
398 rather than realized ones. There are consequently five parameters for the first model, a  
399 single parameter for the second model and no parameter to evaluate for the third model  
400 (where the interaction probability is fixed by the hypothesis).

401 We fitted the different models to each pair of species and recorded the predicted  
402 probabilities. The joint probability  $P(L_{ijy}, X_{iy}, X_{jy})$  was then computed from Eq. 2, and  
403 the likelihood of each observation  $D_{ijy}$  given the model  $\theta$  was computed as  $\mathcal{L}(D_{ijy}|\theta) =$   
404  $P(L_{ij}, X_{iy}, X_{jy})$  if an interaction was observed, and as  $\mathcal{L}(D_{ijy}|\theta) = 1 - P(L_{ij}, X_{iy}, X_{jy})$   
405 if no interaction was observed. The log-likelihood was summed over the entire dataset  
406 to compare the different models by AIC. We therefore evaluate the likelihood of all local  
407 networks, given the model. Not surprisingly, there was a very large number of species pairs  
408 for which this model could not be computed, as they simply never co-occurred. For these  
409 pairs, we have no information of the interaction probability, and they were consequently  
410 removed from the analysis. The log-likelihood reported across the entire dataset was  
411 therefore summed over all pairs of species observed to co-occur at least once. Interactions  
412 between the first (*Salix*) and second (gallers) trophic layers and those between the second  
413 and third (parasitoids) were considered separately. Finally, we used the full model (in  
414 which both co-occurrence and the interaction are conditional on the environment) to  
415 interpolate species distributions and interaction probabilities across the entire European  
416 continent. We reconstructed the expected network for each location in a 1 X 1 km grid  
417 and computed the probabilistic connectance following Poisot *et al.* (2016b).

418 All of the data are openly available in the database *mangal* (Poisot *et al.*, 2016a)  
419 and all R scripts for running the analyses, are provided in the github repository [https://github.com/DominiqueGravel/ms\\_probaweb](https://github.com/DominiqueGravel/ms_probaweb).

421 **Results**

422 Despite the extensive sampling, many pairs of species were observed to co- occur only a  
423 few times. This made it difficult to evaluate interaction probabilities with any reasonable  
424 confidence. Thus, we start with an example of a single pair of species selected because  
425 of its high number of co- occurrences ( $N_{ij} = 38$ ): the leaf folder *Phyllocolpa prussica* and  
426 the parasitoid *Chrysocharis elongata*. These two fairly abundant species were observed  
427  $N_i = 49$  and  $N_j = 121$  times, respectively, across the 370 sites, and they were found to  
428 interact with a marginal probability  $P(L_{ij}) = 0.55$ , which means they interacted at 21  
429 different locations. Here, a comparison of model fit (Table 2) reveals that conditioning  
430 the interaction probability on local environmental conditions adds no explanatory power  
431 beyond a model assuming the same probability of interaction anywhere in space (Model  
432 2 vs Model 3). Moreover, when the two species co-occur, the occurrence of the interac-  
433 tion was insensitive to the environment (Model 2 vs Model 3). Alternatively, climatic  
434 variables significantly impacted co-occurrence (Model 3 vs Model 4). The independent  
435 model performed worse than the non-random co- occurrence model (Model 3 vs Model  
436 6). The full model revealed that the greatest interaction probability occurred at interme-  
437 diate temperature and precipitation, simply because this is where the two species most  
438 frequently co-occur (Fig. 3). The probabilities of co-occurrence and interaction can be  
439 represented in space, where we find that the highest interaction probability occurred in  
440 Central Europe (Fig. 4).

441 We evaluated each model for all pairs of species in order to better understand the  
442 large-scale drivers of network turnover. The results were highly consistent among trophic  
443 levels (*Salix*-gallers and gallers-parasitoids; Table 3), despite large variability in the fit of  
444 the models to the different pairs of species. This variability was particularly important for  
445 the models of interactions given co-occurrence. Across all pairs of species, the conditional  
446 representation of interactions performed (Model 2) better than the marginal one (Model

447 1); that is, interactions did not occur systematically whenever the two species were found  
448 co-occurring. Hence, in addition to species turnover, the stochastic nature of interactions  
449 contributes to network variability. In total, we recorded 1,173 pairs of interactions, only  
450 290 of which occurred more than five times. Out of these 290 interactions, 143 were  
451 systematically detected whenever the two species co-occurred. In the instances when  
452 species co-occurred, the two environmental variables considered proved relatively poor  
453 predictors of their interactions (Model 2 vs Model 3). Not surprisingly, for both types  
454 of interactions (*Salix*-galler and galler-parasitoid), the log-likelihood increased when the  
455 environment was considered. However, the extra number of parameters exceeded the gain  
456 in log-likelihood and inflated AIC. Therefore, the most parsimonious model excluded the  
457 effect of the environment. On the basis of log-likelihood only, co-occurrence was non-  
458 independent for both *Salix*-galler and galler-parasitoid interactions. Thus, according to  
459 AIC, the best model was the one of non-random co-occurrence (Model 3 vs Model 6) for  
460 both types of interactions.

461 Once we had selected the best model based on AIC (Model 3, Table 2), we used  
462 it to reconstruct the expected species richness, along with the most likely network for  
463 each location. Using this approach, we mapped the expected distribution of network  
464 properties across Europe (Fig. 5). For simplicity, we chose to consider connectance  
465 as our descriptor of network configuration, as this metric can be easily computed from  
466 probabilistic networks (Poisot *et al.*, 2016b) and is also a good proxy for many other  
467 network properties (Poisot & Gravel, 2014). Overall, we found a peak in *Salix*, gallers and  
468 parasitoid diversity in Northern Europe. The expected number of interactions roughly  
469 followed the distribution of species richness, but accumulated at a rate different from  
470 species numbers. Connectance likewise peaked in Northern Europe (Fig. 5).

## <sup>471</sup> Interpretation

<sup>472</sup> We have proposed that the representation of community structure and its variation in  
<sup>473</sup> space is best captured by the formalism of ecological networks, as both the distribution  
<sup>474</sup> of species and their interspecific interactions can then be accounted for. We consequently  
<sup>475</sup> revised the niche concept in order to integrate its abiotic and biotic components. This  
<sup>476</sup> integrated niche was represented visually with an ordination of species into an environ-  
<sup>477</sup> mental space and a trait space. The fundamental niche of a species is represented as  
<sup>478</sup> the set of environmental conditions and resources that allow a species to establish in a  
<sup>479</sup> location, thereby integrating the Eltonian and the Grinnellian components of the niche.  
<sup>480</sup> We then translated the concept mathematically by investigating the probability of the  
<sup>481</sup> joint occurrences of species and their interaction, which should be interpreted as an in-  
<sup>482</sup> teraction distribution model. We used this approach to characterize the turnover of the  
<sup>483</sup> structure of ecological interactions in a species-rich tri-trophic network across Western  
<sup>484</sup> Europe, finding that the primary driver of network variation is the turnover in species  
<sup>485</sup> composition.

<sup>486</sup> Applying the framework to our large data set on host-parasite interactions across  
<sup>487</sup> Europe revealed key features in the interaction between *Salix* taxa, their herbivores,  
<sup>488</sup> and the natural enemies of these herbivores. Consistent with a general increase in the  
<sup>489</sup> diversity of *Salix* towards boreal areas (Cronk *et al.*, 2015), overall species richness of  
<sup>490</sup> the networks increased towards the north. The distribution of *Salix* species richness  
<sup>491</sup> largely matched those of gallers and parasitoids. These observations within Europe are  
<sup>492</sup> also matched by the ones found at a global scale for *Salix* (Argus, 1997; Cronk *et al.*,  
<sup>493</sup> 2015; Wu *et al.*, 2015) and sawflies (Kouki *et al.*, 1994; Kouki, 1999). Species richness  
<sup>494</sup> in a common group of parasitic wasps, the Ichneumonidae, was originally presumed to  
<sup>495</sup> show a similar "reversed latitudinal gradient", but this observation has been recently  
<sup>496</sup> challenged by findings of rather high ichneumonid diversity in the tropics (Veijalainen

497 *et al.*, 2013). Nevertheless, ichneumonid subfamilies specifically associated with sawflies  
498 (Ctenopelmatinae, Tryphoninae) are clearly less diverse in the south.

499 Exactly what processes are responsible for the distribution of species richness at dif-  
500 ferent trophic levels is yet to be established (but see e.g. Roininen *et al.* 2005; Nyman  
501 *et al.* 2010; Leppänen *et al.* 2014), but as a net outcome of different latitudinal trends  
502 across trophic levels, the distribution of co-occurrence and therefore of potential inter-  
503 actions differed between the first and second layers of feeding links. The correlation  
504 between expected *Salix* and galler richness was 0.73, while it was 0.58 between gallers  
505 and their parasitoids. Therefore, the ratio of herbivores to *Salix* species is essentially  
506 constant across Europe, whereas each herbivore species is potentially attacked by a richer  
507 enemy community higher latitudes. Consequently, overall connectance peaks in Northern  
508 Europe (Fig. 5).

509 In terms of species interacting with each other, our analysis suggests that the en-  
510 vironment leaves a detectable imprint on species co-occurrence, but only a slight mark  
511 on the occurrence of realized links among species in a specific place: the probability of  
512 finding a given combination of species at a higher and a lower trophic level at the same  
513 site was clearly affected by the environment, whereas the probability of observing an in-  
514 teraction between the two was, given co-occurrence, not significantly so. The interactions  
515 were highly uncertain, as only few pairs of species were systematically interacting when  
516 co-occurring but most not. This applies to the example species *Phyllocolpa prussica* and  
517 *Chrysocharis elongata* (Figs. 3 and 4), but also to all species pairs more generally. For  
518 the example species pair, the full model revealed that the joint probability event (inter-  
519 action and co-occurrence) peaks at intermediate temperature and precipitation, simply  
520 because this is where the two species co-occur most often. This does not imply that  
521 species will always interact when they meet – although this is a basic assumption in most  
522 documented metawebs to date (e.g. Havens 1992; Wood *et al.* 2015). Rather, an inter-

523 action is better represented as a stochastic process whose probability is also influenced  
524 by the probability with which species co-occur. What we cannot reliably know is how  
525 this stochasticity splits into different sources of uncertainty – i.e., the extent to which  
526 a species at the higher trophic level runs into a species at the lower level co-occurring  
527 at the site, the extent to which this interaction is detected by an observer collecting a  
528 finite sample, or simply the uncertainty arising from incomplete description of a highly  
529 dimensional niche. Future work will be required to document the relative importance of  
530 these sources of uncertainty in the occurrence of interactions.

## 531 Perspectives

532 Evidence that the structure of ecological networks does vary across habitats (e.g. Tylianakis  
533 *et al.* 2007), over environmental gradients (Lurgi *et al.*, 2012) and in time (P. & A., 2014)  
534 is accumulating rapidly. It is not clear, however, to what extent the turnover of net-  
535 work structure is driven by a systematic change in species composition or of pairwise  
536 interactions (Poisot *et al.*, 2012, 2015). Our model comparison of host-parasite inter-  
537 actions revealed that most of the turnover is driven by species-specific responses to the  
538 environment, impacting species richness, and that co-occurrence was mostly indepen-  
539 dent. Further, the occurrence of interactions among host and parasite is highly stochastic  
540 even when both are present, and not predictable by the variables considered by us. We  
541 know that interactions vary with the environment in other systems, for instance, her-  
542 bivory (Shurin *et al.*, 2012; Baskett *et al.*, 2018) and predation (McKinnon *et al.*, 2010;  
543 Legagneux *et al.*, 2014) are often found to increase with temperature, resulting in spatial  
544 variation of trophic cascades (Gray *et al.*, 2016). What remains unclear, however, is the  
545 extent to which such variation is driven by a turnover of species composition along gradi-  
546 ents, or a turnover of the interactions. Here we found that interactions vary substantially

547 but non-predictably along the annual temperature and the precipitation gradient. The  
548 lack of detectable signal may be due to our choice of covariates. Indeed, a previous study  
549 focusing on a subset of the system studied here identified habitat characteristics as the  
550 primary drivers of interactions (Nyman *et al.*, 2015). New investigations with other sys-  
551 tems will thus be required to challenge this result. Under all circumstances, documenting  
552 the relationship between the environment and the occurrence of interactions at conti-  
553 nental scales is critical for understanding how large- scale variation of trophic regulation  
554 influences community dynamics and ecosystem functioning (Harfoot *et al.*, 2014).

555 The framework we provide complements methods to compute network beta- diversity  
556 presented by Poisot *et al.* 2012 and Timothée *et al.* 2018. The total network turnover is  
557 partitionned in interaction turnover and species turnover, which in our approach would  
558 correspond to the three terms of Eq. 2, respectively. One could easily derive the different  
559  $\beta$  of Poisot *et al.* 2012 using the fitted probabilities. For instance, for a pair of species  $i$   
560 and  $j$  and sites  $x$  and  $y$ , the expectation for the fraction  $a$  for the interactions would be  
561  $P(L_{ijx}|X_{ix}, X_{jx})P(L_{ijx}|X_{ix}, X_{jx})$ . The same could be computed for factions  $b$  and  $c$ , and  
562 therefore the expected beta-diversity for an entire network could easily be recomputed this  
563 way. The interesting feature of the approach we proposed is that the different partitions  
564 for network turnover could be investigated according to different hypotheses (as performed  
565 here), in order to evaluate for instance the role of environmental heterogeneity or species  
566 co-distribution on network turnover.

567 We restricted our analysis to the effect of co-occurrence on ecological interactions,  
568 neglecting the inverse of the problem. We did not investigate in depth the drivers of  
569 co-occurrence and simply took it for granted from the data. Co-occurrence was indeed  
570 different from the expectation of independent species distributions. It thus begs the  
571 question of whether, once environmental effects on species-specific distribution have been  
572 accounted for, interactions come with significant effects on co-occurrence? We could

573 rephrase this problem by asking whether the fundamental niche differs from the real-  
574 ized niche, and how this applies to our framework. For example, we have considered  
575 above simply the co-occurrence probability,  $P(X_{iy}, X_{jy}|E_y)$ , which could be expanded as  
576  $P(X_{iy}|X_{jy}, E_y)P(X_{jy}|E_y)$ . After some re-arrangement of Eq. 2, the marginal occurrence  
577 probability,  $P(X_{jy}|E_y)$ , could be considered as a species distribution model taking into  
578 account the interaction between these species (in the same spirit as done by González-  
579 Salazar *et al.* 2013). This derivation would however critically depend on a strong *a priori*  
580 expectation of the conditional probability of observing a species given the distribution  
581 of the other species. This assumption seems reasonable for some situations, such as a  
582 parasitoid species that requires a host to develop. On the other hand, Cazelles (2016)  
583 found that the strength of this association is often rather weak if not inexistant (for in-  
584 stance, with the example pair presented in Table 1). The lack of an association could  
585 simply arise when the parasitoid is generalist enough that it is not constrained to track  
586 the distribution of any single/given host (Cazelles *et al.*, 2015).

587 The approach we presented could easily be expanded to account for traits, in order  
588 to have a complete and more mechanistic representation of the niche. Both the Eltonian  
589 and Grinelian dimensions of the niche could be represented as well. As mentioned earlier,  
590 Gravel *et al.* 2016 proposed to represent the occurrence of interactions based on some trait-  
591 matching rules (given by a function of the type  $P(L_{ij})|\mathbf{T}_i, \mathbf{T}_j, E)$ ) and the co-distribution  
592 of traits (corresponding to  $P(\mathbf{T}_i, \mathbf{T}_j|E)$ ). The key insight from this formalism proposed  
593 by Gravel *et al.* 2016 is that trait co-distribution will impact the variation in network  
594 structure. As a result, it was found that network properties often co-vary with functional  
595 structure and functional diversity Idaline *et al.*, 2018. The next step in the development  
596 of this framework will be to develop the models for the trait co-distribution, for instance  
597 following Ovaskainen *et al.* 2017. This may facilitate the accessibility and the applicability  
598 of the framework since the basic unit of analysis is no more the species, but rather a

599 combination of traits, thereby facilitating the statistical analysis by sharing information  
600 among species.

601 At present, there is only indirect support for the hypothesis that interacting species  
602 are conditionally distributed, but this possibility should be the topic of theoretical inves-  
603 tigation with dynamical metacommunity models (Cazelles *et al.*, 2015) and empirical hy-  
604 pothesis testing. The impact of ecological interactions on the distribution of co-occurrence  
605 has been the topic of many publications since Diamond (1975) seminal study on competi-  
606 tion and "checkerboard" distribution, but pairwise approaches have only recently received  
607 attention (Veech, 2013). Whether two interacting species are more closely associated in  
608 space remains unclear, since most approaches based on null models consider community-  
609 level metrics (e.g. Gotelli 2000), such as the C-score, thereby making it hard to evaluate  
610 if specific interactions do indeed affect co-occurrence. The expansion of the framework  
611 we describe to account for the difference between the realized and the fundamental niche  
612 will therefore require further investigation of the impact of interactions on co-occurrence.

613 Ecological networks are known to be extremely sparse, *i.e.* they have far more absences  
614 than presences of interactions. Absences of interactions, however, can come from different  
615 sources. The fact that unequal sampling at the local scale can affect our understanding  
616 of network structure is well documented (Martinez *et al.*, 1999). In a spatial context,  
617 however, some interactions may be undocumented simply because the species involved  
618 have never been observed to co-occur. Although these cases are reported as a lack of  
619 interactions, in actuality we cannot make any reliable inference from them: since the  
620 species have never been observed together, it remains possible that they would interact  
621 if they did. A fundamentally different category of absences of interactions are then those  
622 reported after multiple observations of species co-occurrence. Thus, to gain confidence  
623 that the probability of an interaction is low, extensive sampling (that is, several records  
624 of co- occurrence) is needed. Generally, our confidence that the interaction is indeed

impossible will increase with the number of observations of the species pair. Seeing that this is essentially a Bernoulli process (the probability that the species will interact given their presence), the breadth of the confidence interval is expected to saturate after a fixed number of observations, which can be set as a threshold above which a species pair has finally been observed "often enough". For instance, if two species are observed together at  $k = 10$  locations but they are never found interacting ( $n = 0$ ), this means that the maximum likelihood estimate for the probability of this binomial distribution is  $p = 0$ . Using the Clopper-Pearson method to compute the 95% confidence interval, we find that the real value of the probability of the interaction occurring is somewhere between 0 and 0.31. This interval is rather large, indicating that a significant amount of sampling is required to have credible interaction networks. Fortunately, there are bayesian methods could be used to complement available information and reduce this uncertainty (Cirtwill *et al.*, 2018).

## Conclusion

Our representation of spatial variation of community structure offers a new approach for the study of the biogeography of ecological networks. We see the following key challenges and opportunities ahead in this exciting area of research:

**1. New generation of network data.** Investigating spatial variation of network structure will require high quality and highly replicated network data. We have investigated one the most comprehensive spatial network datasets we are aware of and nonetheless found immense gaps of knowledge in its resolution. Species richness accumulates much faster than observations of ecological interactions (Poisot *et al.*, 2012). Each pair of species must be observed several times in order to obtain reliable estimates of their interaction probability.

649       **2. Estimation of the reliability of interactions.** We need quantitative tools  
650       to estimate the confidence intervals around inferred interaction probabilities, as well as  
651       estimators of the frequency of false absences. Bayesian methods are promising to that  
652       end because we could use information on the target species (e.g. if they are known as  
653       specialists or generalists) to provide prior estimates of the interaction probability.

654       **3. From interaction probabilities to a distribution of network properties.**  
655       Metrics are available to analyse the structure of probabilistic networks (Poisot *et al.*,  
656       2016b). These metrics are useful as first approximation, but they assume independence  
657       among interactions. This might not be the case in nature because of the role of co-  
658       occurrence and shared environmental requirements. We also need to better understand  
659       the distribution of network properties arising from probabilistic interactions.

660       **4. Investigation of the environmental-dependence of ecological interactions.**  
661       There is evidence that interactions can vary in space, but this problem has not been  
662       investigated in a systematic fashion. The paucity of currently available data precludes an  
663       extensive analysis of this question at present.

664       **5. Effects of ecological interactions on co-occurrence.** We have intentionally  
665       omitted the feedback of ecological interactions on co-occurrence in this framework. As  
666       abundance can impact the occurrence of interactions, and, conversely since interactions  
667       impact abundance (Canard *et al.*, 2014), we could reasonably expect that interactions  
668       will also influence co-occurrence. Theory in this regard does exist for simple three-species  
669       modules (Cazelles *et al.*, 2015), but its extension to entire co-occurrence networks will  
670       prove critical in the future, especially given the interest in using co- occurrence to infer  
671       ecological interactions (Morales-Castilla *et al.*, 2015; Morueta-Holme *et al.*, 2016).

## 672 Acknowledgements

673 This is a contribution to the working groups *Continental-scale variation of ecological net-*  
674 *works* supported by the Canadian Institute for Ecology and Evolution), and the *Next*  
675 *Generation Data, Models, and Theory Working Group*, supported by the Santa Fe Insti-  
676 tute, the Betsy and Jesse Fink Foundation, the ASU-SFI Center for Biosocial Complex  
677 Systems, and NSF Grant PHY-1240192. DG also acknowledges financial support from  
678 NSERC- Discovery grant program and Canada Research Chair program.

679 **References**

- 680 Albouy, C., Velez, L., Coll, M., Colloca, F., Le Loc'h, F., Mouillot, D. & Gravel, D. (2014)  
681 From projected species distribution to food-web structure under climate change. *Global  
682 Change Biology* **20**, 730–741.
- 683 Allesina, S., Alonso, D. & Pascual, M. (2008) A general model for food web structure.  
684 *Science* **320**, 658–61.
- 685 Anderson, M.J., Crist, T.O., Chase, J.M., Vellend, M., Inouye, B.D., Freestone, A.L.,  
686 Sanders, N.J., Cornell, H.V., Comita, L.S., Davies, K.F., Harrison, S.P., Kraft, N.J.B.,  
687 Stegen, J.C. & Swenson, N.G. (2011) Navigating the multiple meanings of beta-  
688 diversity: a roadmap for the practicing ecologist. *Ecology Letters* **14**, 19–28.
- 689 Argus, G. (1997) Infrageneric classification of Salix (Salicaceae) in the New World. *Sys-  
690 tematic Botany Monographs* **52**, 1–121.
- 691 Bartomeus, I., Gravel, D., Tylianakis, J., Aizen, M., Dickie, I. & Bernard-Verdier, M.  
692 (2016) A common framework for identifying linkage rules across different types of in-  
693 teractions. *Functional Ecology* **30**, 1894–1903.
- 694 Bascompte, J. & Jordano, P. (2007) Plant-Animal Mutualistic Networks: The Architec-  
695 ture of Biodiversity. *Annual Review of Ecology, Evolution, and Systematics* **38**, 567–  
696 593.
- 697 Baselga, A. (2010) Partitioning the turnover and nestedness components of beta diversity.  
698 *Global Ecology and Biogeography* **19**, 134–143.
- 699 Baskett, C.A., Schemske, D.W. & Novotny, V. (2018) Latitudinal patterns of herbivore  
700 pressure in a temperate herb support the biotic interactions hypothesis. *Ecology Letters*  
701 **21**, 578–587.

- 702 Bersier, L.F., Banašek-Richter, C. & Cattin, M.F. (2002) Quantitative descriptors of  
703 food-web matrices. *Ecology* **83**, 2394–2407.
- 704 Blonder, B., Lamanna, C., Violle, C. & Enquist, B.J. (2014) The n-dimensional hyper-  
705 volume. *Global Ecology and Biogeography* **23**, 595–609.
- 706 Boulangeat, I., Gravel, D. & Thuiller, W. (2012) Accounting for dispersal and biotic  
707 interactions to disentangle the drivers of species distributions and their abundances.  
708 *Ecology Letters* **15**, 584–593.
- 709 Canard, E., Mouquet, N., Mouillot, D., Stanko, M., Miklisova, D. & Gravel, D. (2014)  
710 Empirical evaluation of neutral interactions in host-parasite networks. *The American  
711 naturalist* **183**, 468–79.
- 712 Cazelles, K. (2016) *Influence des interactions biotiques sur la répartition géographique des  
713 espèces*. PhD. Dissertation. Université du Québec à Rimouski, Rimouski.
- 714 Cazelles, K., Araújo, M.B., Mouquet, N. & Gravel, D. (2016) A theory for species co-  
715 occurrence in interaction networks. *Theoretical Ecology* **9**, 39–48.
- 716 Cazelles, K., Mouquet, N., Mouillot, D. & Gravel, D. (2015) On the integration of biotic  
717 interaction and environmental constraints at the biogeographical scale. *Ecography* **39**,  
718 921–931.
- 719 Chase, J. & Leibold, M. (2003) *Ecological niches*. Chicago University Press, Chicago.
- 720 Chase, J.M. (2007) Drought mediates the importance of stochastic community assembly.  
721 *Proceedings of the National Academy of Sciences of the United States of America* **104**,  
722 17430–4.
- 723 Cirtwill, A., Eklöf, A., Roslin, T., Wootton, K. & Gravel, D. (2018) A quantitative  
724 framework for investigating the reliability of network construction. *bioRxiv* .

- 725 Clark, J.S., Dietze, M., Chakraborty, S., Agarwal, P.K., Ibanez, I., LaDeau, S. & Wolosin,  
726 M. (2007) Resolving the biodiversity paradox. *Ecology Letters* **10**, 647–659.
- 727 Connor, E. & Simberloff, D. (1979) The assembly of species communities: chance or  
728 competition? *Ecology* **60**, 1132–1140.
- 729 Cronk, Q., Ruzzier, E., Belyaeva, I. & Percy, D. (2015) Salix transect of Europe: lati-  
730 tudinal patterns in willow diversity from Greece to arctic Norway. *Biodiversity Data*  
731 *Journal* **3**, e6258.
- 732 Diamond, J. (1975) Assembly of species communities. *Ecology and evolution of communi-*  
733 *ties* (eds. M. Cody & J. Diamond), pp. 342–444, Harvard University Press, Cambridge.
- 734 Dunne, J.A. (2006) The network structure of food webs. *Ecological networks: Linking*  
735 *structure and dynamics* (eds. M. Pascual & J.A. Dunne), pp. 27–86, Oxford University  
736 Press, Oxford.
- 737 Eklöf, A., Jacob, U., Kopp, J., Bosch, J., Castro-Urgal, R., Chacoff, N.P., Dalsgaard, B.,  
738 de Sassi, C., Galetti, M., Guimarães, P.R., Lomáscolo, S.B., Martín González, A.M.,  
739 Pizo, M.A., Rader, R., Rodrigo, A., Tylianakis, J.M., Vázquez, D.P. & Allesina, S.  
740 (2013) The dimensionality of ecological networks. *Ecology Letters* **16**, 577–583.
- 741 Gibert, J.P. & DeLong, J.P. (2014) Temperature alters food web body-size structure.  
742 *Biology Letters* **10**.
- 743 Gleason, H. (1926) The individualistic concept of the plant association. *Bulletin of the*  
744 *Torrey Botanical Club* **53**, 7–26.
- 745 Godsoe, W., Jankowski, J., Holt, R.D. & Gravel, D. (2017) Integrating biogeography with  
746 contemporary niche theory. *Trends in Ecology and Evolution* **32**, 488–499.

- 747 González-Salazar, C., Stephens, C.R. & Marquet, P.A. (2013) Comparing the relative con-  
748 tributions of biotic and abiotic factors as mediators of species' distributions. *Ecological*  
749 *Modelling* **248**, 57 – 70.
- 750 Gotelli, N.J. (2000) Null Model Analysis of Species Co-Occurrence Patterns. *Ecology* **81**,  
751 2606.
- 752 Götzenberger, L., de Bello, F., Bråthen, K.A., Davison, J., Dubuis, A., Guisan, A., Lepš,  
753 J., Lindborg, R., Moora, M., Pärtel, M., Pellissier, L., Pottier, J., Vittoz, P., Zobel,  
754 K. & Zobel, M. (2012) Ecological assembly rules in plant communities—approaches,  
755 patterns and prospects. *Biological reviews of the Cambridge Philosophical Society* **87**,  
756 111–27.
- 757 Gravel, D., Albouy, C. & Thuiller, W. (2016) The meaning of functional trait composition  
758 of food webs for ecosystem functioning. *Philosophical Transactions of the Royal Society*  
759 *of London B: Biological Sciences* **371**.
- 760 Gravel, D., Massol, F., Canard, E., Mouillot, D. & Mouquet, N. (2011) Trophic theory  
761 of island biogeography. *Ecology Letters* **14**, 1010–6.
- 762 Gravel, D., Poisot, T., Albouy, C., Velez, L. & Mouillot, D. (2013) Inferring food web  
763 structure from predator-prey body size relationships. *Methods in Ecology and Evolution*  
764 **4**, 1083–1090.
- 765 Gray, S.M., Poisot, T., Harvey, E., Mouquet, N., Miller, T.E. & Gravel, D. (2016) Temper-  
766 ature and trophic structure are driving microbial productivity along a biogeographical  
767 gradient. *Ecography* **39**, 981–996.
- 768 Hanski, I. (1999) *Metapopulation Ecology*. Oxford University Press, Oxford.
- 769 Hardin, G. (1960) The competitive exclusion principle. *Science* **131**, 1292–1297.

- 770 Harfoot, M.B.J., Newbold, T., Tittensor, D.P., Emmott, S., Hutton, J., Lyutsarev, V.,  
771 Smith, M.J., Scharlemann, J.P.W. & Purves, D.W. (2014) Emergent global patterns of  
772 ecosystem structure and function from a mechanistic general ecosystem model. *PLoS*  
773 *Biol* **12**, 1–24.
- 774 Harris, D.J. (2016) Inferring species interactions from co-occurrence data with markov  
775 networks. *Ecology* **97**, 3308–3314.
- 776 Havens, K. (1992) Scale and structure in natural food webs. *Science* **257**, 1107–1109.
- 777 Hijmans, R.J. (2015) *raster: Geographic Data Analysis and Modeling*. R package version  
778 2.5-2.
- 779 Holt, R.D. (2009) Bringing the Hutchinsonian niche into the 21st century : Ecological  
780 and evolutionary perspectives. *Proceedings of the National Academy of Sciences* **106**,  
781 19659–19665.
- 782 Holt, R.D. & Lawton, J.H. (1993) Apparent competition and enemy-free space in insect  
783 host-parasitoid communities. *The American Naturalist* **142**, 623–645.
- 784 Hutchinson, G. (1957) Concluding remarks. *Cold Spring Harbour Symposium* **22**, 415–  
785 427.
- 786 Idaline, L., Isabelle, A., Christoph, D., Ulrich, B., Isabelle, B. & Dominique, G. (2018)  
787 Species traits as drivers of food web structure. *Oikos* **127**, 316–326.
- 788 Ings, T.C., Montoya, J.M., Bascompte, J., Blüthgen, N., Brown, L., Dormann, C.F.,  
789 Edwards, F., Figueroa, D., Jacob, U., Jones, J.I., Lauridsen, R.B., Ledger, M.E.,  
790 Lewis, H.M., Olesen, J.M., van Veen, F.J.F., Warren, P.H. & Woodward, G. (2009)  
791 Ecological networks—beyond food webs. *The Journal of animal ecology* **78**, 253–269.

- 792 Kéfi, S., Berlow, E.L., Wieters, E.a., Navarrete, S.a., Petchey, O.L., Wood, S.a., Boit, A.,  
793 Joppa, L.N., Lafferty, K.D., Williams, R.J., Martinez, N.D., Menge, B.a., Blanchette,  
794 C.a., Iles, A.C. & Brose, U. (2012) More than a meal... integrating non-feeding inter-  
795 actions into food webs. *Ecology Letters* pp. 291–300.
- 796 Kissling, W.D., Dormann, C.F., Groeneveld, J., Hickler, T., Kühn, I., McInerny, G.J.,  
797 Montoya, J.M., Römermann, C., Schiffers, K., Schurr, F.M., Singer, A., Svenning,  
798 J.C., Zimmermann, N.E. & O'Hara, R.B. (2012) Towards novel approaches to mod-  
799 ellng biotic interactions in multispecies assemblages at large spatial extents. *Journal*  
800 *of Biogeography* **39**, 2163–2178.
- 801 Kopelke, J., Nyman, T., Cazelles, K., Gravel, D., Vissault, S. & Roslin, T. (2017) Food-  
802 web structure of willow-galling sawflies and their natural enemies across europe. *Ecology*  
803 **98**, 1730–1730.
- 804 Kopelke, J.P. (1999) Gallenerzeugende Blattwespen Europas - Taxonomische Grundlagen,  
805 Biologie und Ökologie (Tenthredinidae: Nematinae: Euura, Phyllocolpa, Pontania).  
806 *Courier Forschungsinstitut Senckenberg* **212**, 1–183.
- 807 Kopelke, J.P. (2003) Natural enemies of gall-forming sawflies on willows (*Salix* spp.).  
808 *Entomologia Generalis* **26**, 277–312.
- 809 Kouki, J. (1999) Latitudinal gradients in species richness in northern areas: some excep-  
810 tional patterns. *Ecological Bulletins* **47**, 30–37.
- 811 Kouki, J., Niemelä, P. & Viitalsaari, M. (1994) ) Reversed latitudinal gradient in species  
812 richness of sawflies (Hymenoptera, Symphyta). *Annales Zoologici Fennici* **31**, 83–88.
- 813 Krebs, C. (2009) *Ecology: The Experimental Analysis of Distribution and Abundance*. 6th  
814 ed. Pearson, San Francisco.

- 815 Laliberté, E. & Tylianakis, J.M. (2010) Deforestation homogenizes tropical parasitoid-  
816 host networks. *Ecology* **91**, 1740–1747.
- 817 Legagneux, P., Gauthier, G., Lecomte, N., Schmidt, N.M., Reid, D., Cadieux, M.C.,  
818 Berteaux, D., Bêty, J., Krebs, C.J., Ims, R.a., Yoccoz, N.G., Morrison, R.I.G., Leroux,  
819 S.J., Loreau, M. & Gravel, D. (2014) Arctic ecosystem structure and functioning shaped  
820 by climate and herbivore body size. *Nature Climate Change* **E2168**, 1–5.
- 821 Legendre, P., Borcard, D. & Peres-Neto, P.R. (2005) Analyzing beta diversity: partition-  
822 ing the spatial variation of composition data. *Ecological Monographs* **75**, 435–450.
- 823 Legendre, P. & De Cáceres, M. (2013) Beta diversity as the variance of community data:  
824 dissimilarity coefficients and partitioning. *Ecology Letters* **16**, 951–963.
- 825 Leibold, M., Holyoak, M., Mouquet, N., Amarasekare, P., Chase, J.M., Hoopes, M.,  
826 Holt, R.D., Shurin, J.B., Law, R., Tilman, D., Loreau, M. & Gonzalez, A. (2004) The  
827 metacommunity concept : a framework for multi-scale community ecology. *Ecology*  
828 *Letters* **7**, 601–613.
- 829 Leppänen, S., Malm, T., Värri, K. & Nyman, T. (2014) A comparative analysis of genetic  
830 differentiation across six shared willow host species in leaf- and bud-galling sawflies.  
831 *PLoS ONE* **9**, e116286.
- 832 Lurgi, M., López, B.C. & Montoya, J.M. (2012) Novel communities from climate change.  
833 *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*  
834 **367**, 2913–22.
- 835 Martinez, N.D., Hawkins, B.a., Dawah, H.A. & Feifarek, B.P. (1999) Effects of Sampling  
836 Effort on Characterization of Food-Web Structure. *Ecology* **80**, 1044–1055.
- 837 McInerny, G.J. & Etienne, R.S. (2012a) Ditch the niche - is the niche a useful concept in  
838 ecology or species distribution modelling? *Journal of Biogeography* **39**, 2096–2102.

- 839 McInerny, G.J. & Etienne, R.S. (2012b) Stitch the niche - a practical philosophy and  
840 visual schematic for the niche concept. *Journal of Biogeography* **39**, 2103–2111.
- 841 McKinnon, L., Smith, P.A., Nol, E., Martin, J.L., Doyle, F.I., Abraham, K.F., Gilchrist,  
842 H.G., Morrison, R.I.G. & Béty, J. (2010) Lower predation risk for migratory birds at  
843 high latitudes. *Science* **327**, 326–327.
- 844 Morales-Castilla, I., Matias, M.G., Gravel, D. & Araújo, M.B. (2015) Inferring biotic  
845 interactions from proxies. *Trends in Ecology & Evolution* **30**, 347–356.
- 846 Morueta-Holme, N., Blonder, B., Sandel, B., McGill, B.J., Peet, R.K., Ott, J.E., Viole,  
847 C., Enquist, B.J., Jørgensen, P.M. & Svenning, J.C. (2016) A network approach for  
848 inferring species associations from co-occurrence data. *Ecography* **39**, 1139–1150.
- 849 Nyman, T. & Julkunen-Tiitto, R. (2000) Manipulation of the phenolic chemistry of wil-  
850 lows by gall-inducing sawflies. *Proceedings of the National Academy of Sciences* **97**,  
851 13184–13187.
- 852 Nyman, T., Leppänen, S.A., Várkonyi, G., Shaw, M.R., Koivisto, R., Barstad, T.E.,  
853 Vikberg, V. & Roininen, H. (2015) Determinants of parasitoid communities of willow-  
854 galling sawflies: habitat overrides physiology, host plant and space. *Molecular Ecology*  
855 **24**, 5059–5074.
- 856 Nyman, T., Vikberg, V., Smith, D. & BoevéJ-L (2010) How common is ecological speci-  
857 ation in plant-feeding insects? A “Higher” Nematinae perspective. *BMC Evolutionary  
858 Biology* **10**, e266.
- 859 Ovaskainen, O., Tikhonov, G., Norberg, A., Blanchet, F.G., Duan, L., Dunson, D., Roslin,  
860 T., Abrego, N. & Chave, J. (2017) How to make more out of community data? a  
861 conceptual framework and its implementation as models and software. *Ecology Letters*  
862 **20**, 561–576.

- 863 P., S.M. & A., B.L. (2014) Partitioning interaction turnover among alpine pollination  
864 networks: spatial, temporal, and environmental patterns. *Ecosphere* **5**, art149.
- 865 Pellissier, L., Rohr, R., Ndiribe, C., Pradervand, J.N., Salamin, N., Guisan, A. & Wisz, M.  
866 (2013) Combining food web and species distribution models for improved community  
867 projections. *Ecology and Evolution* **3**, 4572–4583.
- 868 Petchey, O.L., Brose, U. & Rall, B.C. (2010) Predicting the effects of temperature on food  
869 web connectance. *Philosophical transactions of the Royal Society of London. Series B,  
870 Biological sciences* **365**, 2081–91.
- 871 Peters, R. (1991) *A critique for ecology*. Cambridge University Press, Cambridge.
- 872 Phillips, S.J., Anderson, R.P. & Schapire, R.E. (2006) Maximum entropy modeling of  
873 species geographic distributions. *Ecological Modelling* **190**, 231 – 259.
- 874 Pillai, P., Gonzalez, A. & Loreau, M. (2011) Metacommunity theory explains the emer-  
875 gence of food web complexity. *Proceedings of the National Academy of Sciences of the  
876 United States of America* **108**, 19293–19298.
- 877 Poisot, T., Baiser, B., Dunne, J.A., Kéfi, S., Massol, F., Mouquet, N., Romanuk, T.N.,  
878 Stouffer, D.B., Wood, S.A. & Gravel, D. (2016a) mangal – making ecological network  
879 analysis simple. *Ecography* **39**, 384–390.
- 880 Poisot, T., Canard, E., Mouillot, D., Mouquet, N. & Gravel, D. (2012) The dissimilarity  
881 of species interaction networks. *Ecology Letters* **15**, 1353–1361.
- 882 Poisot, T., Cirtwill, A.R., Cazelles, K., Gravel, D., Fortin, M.J. & Stouffer, D.B. (2016b)  
883 The structure of probabilistic networks. *Methods in Ecology and Evolution* **7**, 303–312.
- 884 Poisot, T. & Gravel, D. (2014) When is an ecological network complex? connectance  
885 drives degree distribution and emerging network properties. *PeerJ* **2**, e251.

- 886 Poisot, T., Gravel, D., Leroux, S., Wood, S.A., Fortin, M.J., Baiser, B., Cirtwill, A.R.,  
887 Araújo, M.B. & Stouffer, D.B. (2016c) Synthetic datasets and community tools for the  
888 rapid testing of ecological hypotheses. *Ecography* **39**, 402–408.
- 889 Poisot, T., Stouffer, D.B. & Gravel, D. (2015) Beyond species: why ecological interactions  
890 vary through space and time. *Oikos* **124**, 243–251.
- 891 Pollock, L.J., Tingley, R., Morris, W.K., Golding, N., O'Hara, R.B., Parris, K.M., Vesk,  
892 P.A. & McCarthy, M.A. (2014) Understanding co-occurrence by modelling species si-  
893 multaneously with a joint species distribution model (jsdm). *Methods in Ecology and*  
894 *Evolution* **5**, 397–406.
- 895 Pulliam, H.R. (2000) On the relationship between niche and distribution. *Ecology Letters*  
896 **3**, 349–361.
- 897 Roininen, H., Nyman, T. & Zinovjev, A. (2005) Biology, ecology, and evolution of gall-  
898 inducing sawflies (Hymenoptera: Tenthredinidae and Xyelidae). *Biology, Ecology, and*  
899 *Evolution of Gall-Inducing Arthropods* (eds. A. Raman, C.W. Schaefer & T.M. With-  
900 ers), pp. 467–494, Science Publishers, Inc.
- 901 Scheiner, S.M. & Willig, M.R. (2007) A general theory of ecology. *Vital And Health*  
902 *Statistics. Series 20 Data From The National Vitalstatistics System Vital Health Stat*  
903 *20 Data Natl Vital Sta .*
- 904 Schleuning, M., Fründ, J., Klein, A.M., Abrahamczyk, S., Alarcón, R., Albrecht, M., An-  
905 dersson, G., Bazarian, S., Böhning-Gaese, K., Bommarco, R., Dalsgaard, B., Dehling,  
906 D., Gotlieb, A., Hagen, M., Hickler, T., Holzschuh, A., Kaiser-Bunbury, C., Kreft,  
907 H., Morris, R., Sandel, B., Sutherland, W., Svenning, J.C., Tscharntke, T., Watts,  
908 S., Weiner, C., Werner, M., Williams, N., Winqvist, C., Dormann, C. & Blüthgen,

- 909 N. (2012) Specialization of mutualistic interaction networks decreases toward tropical  
910 latitudes. *Current Biology* **22**, 1925 – 1931.
- 911 Shurin, J.B., Clasen, J.L., Greig, H.S., Kratina, P. & Thompson, P.L. (2012) Warming  
912 shifts top-down and bottom-up control of pond food web structure and function.  
913 *Philosophical Transactions of the Royal Society of London B: Biological Sciences* **367**,  
914 3008–3017.
- 915 Silvertown, J. (2004) Plant coexistence and the niche. *Trends in Ecology and Evolution*  
916 **19**, 605–611.
- 917 Skvortsov, A.K. (1999) Willows of Russia and adjacent countries. Taxonomical and geo-  
918 graphical revision. *Univ. Joensuu Fac. Math. Natl. Sci. Rep. Ser* **39**, 1–307.
- 919 Soberón, J. (2007) Grinnellian and eltonian niches and geographic distributions of species.  
920 *Ecology Letters* **10**, 1115–1123.
- 921 Staniczenko, P.P., Sivasubramaniam, P., Suttle, K.B., Pearson, R.G. & Lawler, J. (2017)  
922 Linking macroecology and community ecology: refining predictions of species distribu-  
923 tions using biotic interaction networks. *Ecology Letters* **20**, 693–707.
- 924 Stephens, C. & Heau, J. (2009) Using biotic interaction networks for prediction in biodi-  
925 versity and emerging diseases. *PloS one* **4**, e5725.
- 926 Stouffer, D.B., Camacho, J., Jiang, W. & Amaral, L.a.N. (2007) Evidence for the existence  
927 of a robust pattern of prey selection in food webs. *Proceedings of the Royal Society B:  
928 Biological Sciences* **274**, 1931–1940.
- 929 Thuiller, W. (2003) Biomod – optimizing predictions of species distributions and project-  
930 ing potential future shifts under global change. *Global Change Biology* **9**, 1353–1362.

- 931 Tilman, D. (1982) *Resource competition and community structure*. Princeton University  
932 Press, Princeton.
- 933 Timothée, P., Cynthia, G., MarieJosée, F., Dominique, G. & Pierre, L. (2018) Hosts,  
934 parasites and their interactions respond to different climatic variables. *Global Ecology*  
935 and Biogeography
- 936 Townsend, P.A., Soberón, S., Pearson, R.G., Anderson, R.P., Martínez-Meyer, E., Miguel,  
937 N. & Araújo, M.B. (2011) *Ecological Niches and Geographic Distributions*. Princeton  
938 University Press, Princeton.
- 939 Trøjelsgaard, K., Jordano, P., Carstensen, D.W. & Olesen, J.M. (2015) Geographical  
940 variation in mutualistic networks: similarity, turnover and partner fidelity. *Proceedings*  
941 of the Royal Society of London B: Biological Sciences
- 942 Tylianakis, J.M., Tscharntke, T. & Lewis, O.T. (2007) Habitat modification alters the  
943 structure of tropical host-parasitoid food webs. *Nature*
- 944 Veech, J.A. (2013) A probabilistic model for analysing species co-occurrence. *Global Ecol-*  
945 *ogy and Biogeography*
- 946 Veijalainen, A., Sääksjärvi, I., Erwin, T., Gómez, I. & Longino, J. (2013) Subfamily  
947 composition of Ichneumonidae (Hymenoptera) from western Amazonia: insights into  
948 diversity of tropical parasitoid wasps. *Insect Conservation and Diversity*
- 949 Vellend, M., Srivastava, D.S., Anderson, K.M., Brown, C.D., Jankowski, J.E., Kleyn-  
950 hans, E.J., Kraft, N.J.B., Letaw, A.D., Macdonald, A.A.M., Maclean, J.E., Myers-  
951 Smith, I.H., Norris, A.R. & Xue, X. (2014) Assessing the relative importance of neutral  
952 stochasticity in ecological communities. *Oikos*
- 953 Williams, R. & Martinez, N. (2000) Simple rules yield complex food webs. *Nature*
- 954 180–183.

- 955 Williams, R.J., Anandanadesan, A. & Purves, D. (2010) The probabilistic niche model  
956 reveals the niche structure and role of body size in a complex food web. *PloS One* **5**,  
957 e12092.
- 958 Wisz, M.S., Pottier, J., Kissling, W.D., Pellissier, L., Lenoir, J., Damgaard, C.F., Dor-  
959 mann, C.F., Forchhammer, M.C., Grytnes, J.A., Guisan, A., Heikkinen, R.K., Høye,  
960 T.T., Kühn, I., Luoto, M., Maiorano, L., Nilsson, M.C., Normand, S., Ockinger, E.,  
961 Schmidt, N.M., Termansen, M., Timmermann, A., Wardle, D.a., Aastrup, P. & Sven-  
962 ning, J.C. (2013) The role of biotic interactions in shaping distributions and realised  
963 assemblages of species: implications for species distribution modelling. *Biological Re-  
views* **88**, 15–30.
- 964
- 965 Wood, S.A., Russell, R., Hanson, D., Williams, R.J. & Dunne, J.A. (2015) Effects of  
966 spatial scale of sampling on food web structure. *Ecology and Evolution* **5**, 3769–3782.
- 967 Woodward, G., Perkins, D.M. & Brown, L.E. (2010) Climate change and freshwater  
968 ecosystems: impacts across multiple levels of organization. *Philosophical transactions  
969 of the Royal Society of London. Series B, Biological sciences* **365**, 2093–106.
- 970 Wu, J., Nyman, T., Wang, D.C., Argus, G., Yang, Y.P. & Chen, J.H. (2015) Phylogeny  
971 of Salix subgenus Salix s.l. (Salicaceae): delimitation, biogeography, and reticulate  
972 evolution. *BMC Evolutionary Biology* **15**, e31.

Table 1: Example of data used to perform the analysis. Traditional species distribution models represent the occurrence of single species as a function of the environment, with models such as  $X_i \text{ Temp}$  performed one species after the other. Some also tried to represent both the abiotic and biotic niche, with models such as  $X_i \text{ Temp} + X_j$  (e.g. González-Salazar *et al.* 2013), while the new joint species distribution models (Ovaskainen *et al.*, 2017) consider the entire community at once ( $\mathbf{X} \text{ Temp}$ ) and aims at representing co-distribution ( $X_{ij}$ ). Here the focus is on the occurrence of interactions,  $L_{ij}$ . The full model (see model 3, Tables 1 and 2) considers the result of two submodels, one representing the occurrence of interactions given co-occurrence,  $L_{ij} \text{ Temp}$  (evaluated only on the subset of data where  $X_{ij} = 1$ ), and the other representing the co-occurrence,  $X_{ij} \text{ Temp}$ .

Location	$Sp_i$	$Sp_j$	$X_i$	$X_j$	$X_{ij}$	$L_{ij}$	Temp
1	A	B	0	0	0	0	20
2	A	B	1	1	1	0	21
3	A	B	1	0	0	0	22
4	A	B	0	0	0	0	23
5	A	B	1	1	1	1	24
1	A	C	0	1	0	0	20
2	A	C	1	0	0	0	21
3	A	C	1	1	1	1	22
4	A	C	0	0	0	0	23
5	A	C	1	0	0	0	24

Table 2: Summary of model comparison for the interaction between the leaf folder *Phyllocolpa prussica*) and the parasitoid *Chrysocharis elongata*

. Note that  $E$  is a short notation including both temperature and precipitation.

#	Metaweb model	Co-occurrence model	LL	npars	AIC
1	$P(L_{ijy})$	$P(X_{iy}, X_{jy} E_y)$	-71.1	6	154.2
2	$P(L_{ijy} X_{iy}, X_{jy})$	$P(X_{iy}, X_{jy} E_y)$	-65.7	6	143.4
3	$P(L_{ijy} X_{iy}, X_{jy}, E_y)$	$P(X_{iy}, X_{jy} E_y)$	-65.6	10	151.3
4	$P(L_{ijy} X_{iy}, X_{jy}, E_y)$	$P(X_{iy}, X_{jy})$	-84.5	6	183
5	$P(L_{ijy} X_{iy}, X_{jy}, E_y)$	$P(X_{iy})P(X_{jy})$	-80.7	7	173.4
6	$P(L_{ijy} X_{iy}, X_{jy}, E_y)$	$P(X_{iy} E_y)P(X_{jy} E_y)$	-68.8	15	167.6

Table 3: Summary of model comparison for the interaction across all pairs of *Salix*, gallers and parasitoids.

Interaction	#	Metaweb model	Co-occurrence model	LL	npars	AIC
<i>Salix</i> -Galler	1	$P(L_{ijy})$	$P(X_{iy}, X_{jy} E_y)$	-6022.1	7548	27140.3
	2	$P(L_{ijy} X_{iy}, X_{jy})$	$P(X_{iy}, X_{jy} E_y)$	-5547.9	7548	26191.8
	3	$P(L_{ijy} X_{iy}, X_{jy}, E_y)$	$P(X_{iy}, X_{jy} E_y)$	-5364.0	12580	35888.0
	4	$P(L_{ijy} X_{iy}, X_{jy}, E_y)$	$P(X_{iy}, X_{jy})$	-5998.4	8806	30287.2
	5	$P(L_{ijy} X_{iy}, X_{jy}, E_y)$	$P(X_{iy})P(X_{jy})$	-6636.1	7548	27092.7
	6	$P(L_{ijy} X_{iy}, X_{jy}, E_y)$	$P(X_{iy} E_y)P(X_{jy} E_y)$	-6002.9	18870	49745.7
Galler-Parasitoid	1	$P(L_{ijy})$	$P(X_{iy}, X_{jy} E_y)$	-22068.1	19206	82548.2
	2	$P(L_{ijy} X_{iy}, X_{jy})$	$P(X_{iy}, X_{jy} E_y)$	-19504.8	19206	77421.6
	3	$P(L_{ijy} X_{iy}, X_{jy}, E_y)$	$P(X_{iy}, X_{jy} E_y)$	-20217.0	32010	104454.1
	4	$P(L_{ijy} X_{iy}, X_{jy}, E_y)$	$P(X_{iy}, X_{jy})$	-19591.3	22407	77594.5
	5	$P(L_{ijy} X_{iy}, X_{jy}, E_y)$	$P(X_{iy})P(X_{jy})$	-22491.5	19206	89796.9
	6	$P(L_{ijy} X_{iy}, X_{jy}, E_y)$	$P(X_{iy} E_y)P(X_{jy} E_y)$	-18936.9	48015	133903.7

<sub>973</sub> **Figure legends**

<sub>974</sub> **Figure 1**

<sub>975</sub> **Non-random sampling of the metaweb.** Network assembly can be viewed as a sam-  
<sub>976</sub> pling process of the regional pool of potential interactions. Species (indicated by colored  
<sub>977</sub> nodes) are sampled first, and among the species found in the local network, only some in-  
<sub>978</sub> teractions (indicated by blue links) occur. We characterize these sampling processes with  
<sub>979</sub> the quantitative framework proposed in this paper. As a concrete illustration of metaweb  
<sub>980</sub> sampling, we here show a local interaction network among *Salix* (bottom/green), gallers  
<sub>981</sub> (center/blue), and parasitoids (top/red). The metaweb was constructed by aggregat-  
<sub>982</sub> ing interactions observed across 370 local networks. Nodes were distributed to minimize  
<sub>983</sub> crossing of interactions.

<sub>984</sub> **Figure 2**

<sub>985</sub> **Visual representation of the integrated niche.** In biogeography, the niche is con-  
<sub>986</sub> sidered the set of environmental conditions where the intrinsic growth rate  $r$  is positive  
<sub>987</sub> (Holt, 2009). The horizontal axis represents an environmental gradient impacting the  
<sub>988</sub> growth of a focal species (red point). The location of each species (grey points) along  
<sub>989</sub> this gradient represents their optimum, and the vertical dotted lines represent the limits  
<sub>990</sub> of the Grinnellian niche of the focal species. In food web ecology, the Eltonian niche  
<sub>991</sub> represents the location of a species in the food web, as determined by its niche position  
<sub>992</sub> ( $n$ ) and its niche optimum ( $c$ ). The vertical axis represents a niche gradient, for example  
<sub>993</sub> a trait such as body size. The location of each species along this gradient represents their  
<sub>994</sub> niche position. The focal species will feed only on host/prey species occupying niche lo-  
<sub>995</sub> cations within a given interval around the optimum, represented by the horizontal lines.  
<sub>996</sub> The integrated Grinnellian and Eltonian niche corresponds to the square in the middle

997 where an interaction is possible owing to a match of traits and spatial distribution. The  
998 central square represents the area where the joint probability of observing co-occurrence  
999 and interactions is positive.

1000 **Figure 3**

1001 **Probabilistic representation of the interaction probability between a leaf folder**  
1002 (*Phyllocolpa prussica*) and a parasitoid (*Chrysocharis elongata*) across gradi-  
1003 ents of annual average temperature and annual precipitation. The colour gra-  
1004 dient represents the probability of observing the event, from 0 (white) to 1 (black). The  
1005 representation is based on predictions from Model 3 (see Table 1). In the left panel,  
1006 open circles represent the absence of both species, whereas closed circles represent co-  
1007 occurrence and plus signs the occurrence of only one of the two species. In the other  
1008 two panels, open circles represent co-occurrence but an absence of interaction and closed  
1009 circles the occurrence of an interaction.

1010 **Figure 4**

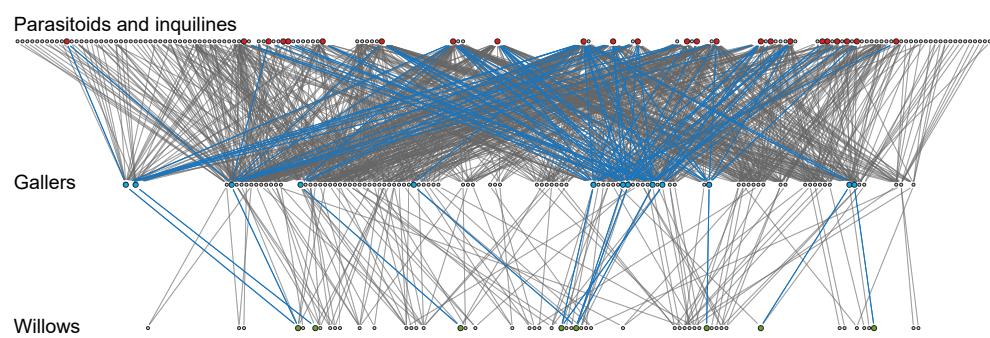
1011 **Probabilistic representation of the interaction probability between a leaf folder**  
1012 (*Phyllocolpa prussica*) and a parasitoid (*Chrysocharis elongata*) across Eu-  
1013 rope. Climate data are extracted from WorldClim over a 1km x 1km grid and predic-  
1014 tions made according to the model illustrated at Fig. 3. Colour gradient ranges from low  
1015 values (dark blue) to high values (yellow).

1016 **Figure 5**

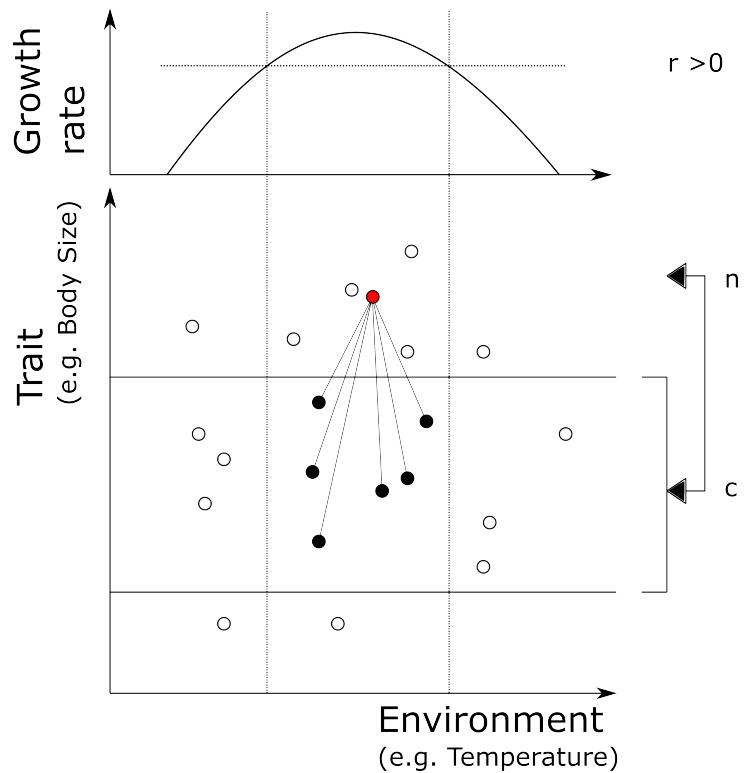
1017 **Mapping the distribution of species richness, the number of links and con-**  
1018 **nectance across Europe.** The representation is based on predictions from Model 3 (see  
1019 Table 2) using climate data extracted from WorldClim over a 1km x 1km grid. Species

richness is obtained by summation of individual occurrence probabilities, and link density by summation of interaction probabilities. Color gradient ranges from low values (dark blue) to high values (yellow).

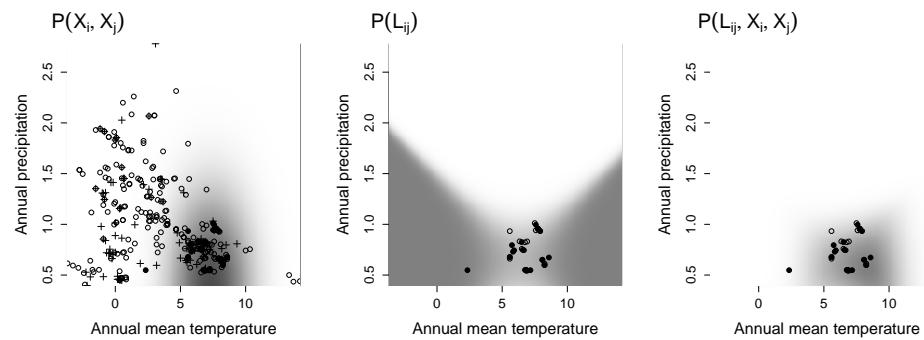
1023 **Figure 1**



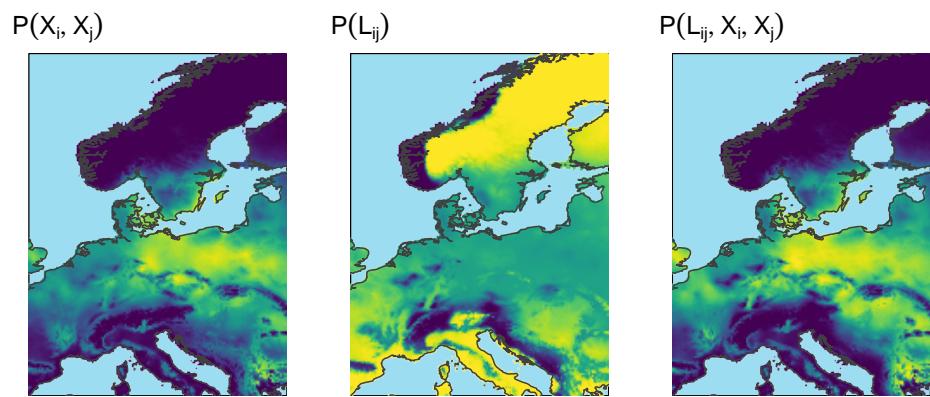
<sub>1024</sub> Figure 2



1025 **Figure 3**



<sub>1026</sub> **Figure 4**



<sub>1027</sub> **Figure 5**

