

# Food web reconstruction through phylogenetic transfer of low-rank network representation

Tanya Strydom<sup>1,2,‡</sup> Salomé Bouskila<sup>1,‡</sup> Francis Banville<sup>1,3,2</sup> Ceres Barros<sup>4</sup> Dominique Caron<sup>5,2</sup>  
Maxwell J Farrell<sup>6</sup> Marie-Josée Fortin<sup>6</sup> Victoria Hemming<sup>4</sup> Benjamin Mercier<sup>3,2</sup> Laura  
J. Pollock<sup>5,2</sup> Rogini Runghen<sup>7</sup> Giulio V. Dalla Riva<sup>8</sup> Timothée Poisot<sup>1,2</sup>

<sup>1</sup> Département de Sciences Biologiques, Université de Montréal, Montréal, Canada    <sup>2</sup> Québec Centre for Biodiversity Sciences, Montréal, Canada    <sup>3</sup> Université de Sherbrooke, Sherbrooke, Canada

<sup>4</sup> Department of Forest Resources Management, University of British Columbia, Vancouver, Canada

<sup>5</sup> Department of Biology, McGill University, Montréal, Canada    <sup>6</sup> Department of Ecology & Evolutionary Biology, University of Toronto, Toronto, Canada    <sup>7</sup> Centre for Integrative Ecology, School of Biological Sciences, University of Canterbury, Canterbury, New Zealand    <sup>8</sup> School of Mathematics and Statistics, University of Canterbury, Canterbury, New Zealand

<sup>‡</sup> These authors contributed equally to the work

## Correspondance to:

Timothée Poisot — timothee.poisot@umontreal.ca

Despite their importance in many ecological processes, collecting data and information on ecological interactions, and therefore species interaction networks, is an exceedingly challenging task. For this reason, large parts of the world have a data deficit when it comes to species interactions, and how the resulting networks are structured. As data collection alone is unlikely to be sufficient at filling these global gaps, community ecologists must adopt predictive methods. In this contribution, we develop such a method, relying on graph embedding (the extraction of explanatory latent variables from known graph structures) and transfer learning (the application of previous solutions to novel problems with limited predictors overlap) in order to assemble a predicted list of trophic interactions between mammals of Canada. This interaction list is derived from extensive knowledge of the mammalian food web of Europe, despite the fact that there are fewer than 5% of common species between the two locations. The results of the predictive model are compared against databases of recorded pairwise interactions, showing that we correctly recover over 95% of known interactions. We provide guidance on how this method can be adapted by substituting some approaches or predictors in order to make it more generally applicable.

## <sup>1</sup> Introduction

<sup>2</sup> There are two core challenges we are faced with in furthering our understanding of ecological networks  
<sup>3</sup> across space, particularly at macro-ecologically relevant scales (e.g. Trøjelsgaard & Olesen 2016). First,  
<sup>4</sup> networks within a location are difficult to sample properly (Jordano 2016a, b), resulting in a widespread  
<sup>5</sup> “Eltonian shortfall” (Hortal *et al.* 2015), *i.e.* a lack of knowledge about inter and intra specific  
<sup>6</sup> relationships. This first challenge has been, in large part, addressed by the recent emergence of a suite of  
<sup>7</sup> methods aiming to predict interactions within *existing* networks, many of which are reviewed in Strydom  
<sup>8</sup> *et al.* (2021a). Second, recent analyses based on collected data (Poisot *et al.* 2021a) or metadata (Cameron  
<sup>9</sup> *et al.* 2019) highlight that ecological networks are currently studied in a biased subset of space and  
<sup>10</sup> bioclimates, which impedes our ability to generalize any local understanding of network structure.  
<sup>11</sup> Meaning that, although the framework to address incompleteness *within* networks exists, there would still  
<sup>12</sup> be regions for which, due to a *lack* of local interaction data, we are unable to infer potential species  
<sup>13</sup> interactions. Having a general solution for inferring a *plausible* metaweb (despite the unavailability of  
<sup>14</sup> interaction data) could be the catalyst for significant breakthroughs in our ability to start thinking about  
<sup>15</sup> species interactions networks over large spatial scales.  
<sup>16</sup> Here, we present a general method for the transfer learning of network representations, relying on the  
<sup>17</sup> similarities of species in a biologically/ecologically relevant proxy space (e.g. shared morphology or  
<sup>18</sup> ancestry). Transfer learning is a machine learning methodology that uses the knowledge gained from  
<sup>19</sup> solving one problem and applying it to a related (destination) problem (Pan & Yang 2010; Torrey & Shavlik  
<sup>20</sup> 2010). In this instance, we solve the problem of predicting trophic interactions between species, based on  
<sup>21</sup> knowledge extracted from another species pool for which interactions are known by using phylogenetic  
<sup>22</sup> structure as a medium for transfer. This allows us to construct a *probabilistic* metaweb for a community  
<sup>23</sup> for which we have *no* prior trophic interaction data for the desired species pool. Our methodology is  
<sup>24</sup> outlined in fig. 1, where we provide an illustration based on learning the embedding of a metaweb of  
<sup>25</sup> trophic interactions for European mammals (known interactions; Maiorano *et al.* 2020b, a) and, based on  
<sup>26</sup> phylogenetic relationships between mammals globally (*i.e.*, phylogenetic tree Upham *et al.* 2019), infer a  
<sup>27</sup> metaweb for the Canadian mammalian species pool (interactions are treated as unknown in this instance).

<sup>28</sup>

[Figure 1 about here.]

29 There is a plurality of measures of species similarities that can be used for metaweb reconstruction (see e.g.  
30 Morales-Castilla *et al.* 2015); however, phylogenetic proximity has several desirable properties when  
31 working at large scales. Gerhold *et al.* (2015) made the point that phylogenetic signal captures  
32 diversification of characters (large macro-evolutionary process), but not necessarily community assembly  
33 (fine ecological process); Dormann *et al.* (2010) previously found very similar conclusions. Interactions  
34 tend reflect a phylogenetic signal because they have a conserved pattern of evolutionary convergence that  
35 encompasses a wide range of ecological and evolutionary mechanisms (Cavender-Bares *et al.* 2009;  
36 Mouquet *et al.* 2012), and - most importantly - retain this signal even when it is not detectable at the  
37 community scale (Hutchinson *et al.* 2017; Poisot & Stouffer 2018). Finally, species interactions at  
38 macro-ecological scales seem to respond mostly to macro-evolutionary processes (Price 2003); which is  
39 evidenced by the presence of conserved backbones in food webs (Dalla Riva & Stouffer 2016), strong  
40 evolutionary signature on prey choice (Stouffer *et al.* 2012), and strong phylogenetic signature in food web  
41 intervality (Eklöf & Stouffer 2016). Phylogenetic reconstruction has also previously been used within the  
42 context of ecological networks, namely understanding ancestral plant-insect interactions (Braga *et al.*  
43 2021). Taken together, these considerations suggest that phylogenies can reliably be used to transfer  
44 knowledge on species interactions.

45 Our case study shows that phylogenetic transfer learning is indeed an effective approach to predict the  
46 Canadian mammalian metaweb. This showcases that although the components (species) that make up  
47 the Canadian and European communities may be *minimally* shared, if the medium (proxy space) selected  
48 in the transfer step is biologically plausible, we can still effectively learn from the known network and  
49 make biologically relevant predictions of interactions. It should be reiterated that the framework  
50 presented in fig. 1 is amenable to changes; notably, the measure of similarity may not be phylogeny, and  
51 can be replaced by information on foraging (Beckerman *et al.* 2006), cell-level mechanisms (Boeckaerts *et*  
52 *al.* 2021), or a combination of traits and phylogenetic structure (Stock 2021).

## 53 Data used for the case study

54 We use data from the European metaweb assembled by Maiorano *et al.* (2020b), following the definition of  
55 the metaweb first introduced by Dunne (2006), *i.e.* an inventory of all possible interactions within species  
56 from a spatially delimited pool. Notably the metaweb is not a prediction of the food web at any specific

57 locale within the frontiers of the species pool – in fact, these local food webs are expected to have a subset  
58 of both the species and the interactions of their metaweb (Poisot *et al.* 2012). This being said, as the  
59 metaweb represents the total of functional, phylogenetic, and macroecological processes (Morales-Castilla  
60 *et al.* 2015), it is thus still worthy of ecological attention. We deduced the subgraph corresponding to all  
61 mammals by matching species names in the original network to the GBIF taxonomic backbone (GBIF  
62 Secretariat 2021) and retaining all those who matched to mammals. This serves a dual purpose 1) to  
63 extract only mammals from the European network and 2) to match and standardize species names when  
64 aggregating the different data sources further downstream (which is an important consideration when  
65 combining datasets (Grenié *et al.* 2021)). All nodes had valid matches to GBIF at this step, and so this  
66 backbone is used for all name reconciliation steps as outlined below.

67 The European metaweb represents the knowledge we want to learn and transfer; the phylogenetic  
68 similarity of mammals here represents the information for transfer. We used the mammalian consensus  
69 supertree by Upham *et al.* (2019), for which all approximatively 6000 names have been similarly matched  
70 to their GBIF valid names. This step allows us to place each node of the mammalian European metaweb  
71 in the phylogeny.

72 The destination problem to which we want to transfer knowledge is the trophic interactions between  
73 mammals in Canada. We obtained the list of extant species from the IUCN checklist, and selected the  
74 terrestrial and semi-aquatic species (this corresponds to the same selection that was applied by Maiorano  
75 *et al.* (2020b) in the European metaweb). The IUCN names were, as previously, reconciled against GBIF to  
76 have an exact match to the taxonomy.

77 After taxonomic cleaning and reconciliation as outlined in the following sections, the mammalian  
78 European metaweb has 260 species, and the Canadian species pool has 163; of these, 17 (about 4% of the  
79 total) are shared, and 89 species from Canada (54%) had at least one congeneric species in Europe. The  
80 similarity for both species pools predictably increases with higher taxonomic order, with 19% of shared  
81 genera, 47% of shared families, and 75% of shared orders; for the last point, Canada and Europe each had a  
82 single unique order (*Didelphimorphia* for Canada, *Erinaceomorpha* for Europe).

83 In the following sections, we describe the representational learning step applied to European data, the  
84 transfer step through phylogenetic similarity, and the generation of a probabilistic metaweb for the  
85 destination species pool.

86 **Method description**

87 The crux of the method is the transfer of knowledge of a known network, in order to predict interactions  
88 between species from another location. In fig. 1, we give a high-level overview of the approach; in the  
89 example around which this manuscript is built (leveraging detailed knowledge about binary trophic  
90 interactions between Mammalia in Europe to predict the less known trophic interactions between closely  
91 phylogenetically related Mammalia in Canada), we use a series of specific steps for network embedding,  
92 trait inference, network prediction and thresholding.

93 Specifically, our approach can be summarized as follows: from the known network in Europe, we use a  
94 truncated Singular Value Decomposition (t-SVD; Halko *et al.* 2011) to generate latent traits representing a  
95 low-dimensional embedding of the network; these traits give an unbiased estimate of the node's position  
96 in the latent feature spaces. Then, we map these latent traits onto a reference phylogeny (other  
97 distance-based measures of species proximity that allow for the inference of features in the latent space  
98 can be used, for example the dissimilarity in functional traits). Based on the reconstructed latent traits for  
99 species in the destination species pool, a Random Dot Product Graph model (hereafter RDPG; Young &  
100 Scheinerman 2007) predicts the interaction between species through a function of the nodes' features  
101 through matrix multiplication. Thus, from latent traits and node position, we can infer interactions.

102 **Implementation and code availability**

103 The entire pipeline is implemented in *Julia* 1.6 (Bezanson *et al.* 2017) and is available under the  
104 permissive MIT License at <https://osf.io/2zwqm/>. The taxonomic cleanup steps are done using GBIF.jl  
105 (Dansereau & Poisot 2021). The network embedding and analysis is done using EcologicalNetworks.jl  
106 (Poisot *et al.* 2019; Banville *et al.* 2021). The phylogenetic simulations are done using PhyloNetworks.jl  
107 (Solís-Lemus *et al.* 2017) and Phylo.jl (Reeve *et al.* 2016). A complete Project.toml file specifying the  
108 full tree of dependencies is available alongside the code. This material also includes a fully annotated copy  
109 of the entire code required to run this project (describing both the intent of the code and discussing some  
110 technical implementation details), a vignette for every step of the process, and a series of Jupyter  
111 notebooks with the text and code. The pipeline can be executed on a laptop in a matter of minutes, and  
112 therefore does not require extensive computational power.

113 **Step 1: Learning the origin network representation**

114 The first step in transfer learning is to learn the structure of the original dataset. In order to do so, we rely  
115 on an approach inspired from representational learning, where we learn a *representation* of the metaweb  
116 (in the form of the latent subspaces), rather than a list of interactions (species *a* eats *b*). This approach is  
117 conceptually different from other metaweb-scale predictions (e.g. Albouy *et al.* 2019), in that the metaweb  
118 representation is easily transferable. Specifically, we use RDPG to create a number of latent variables that  
119 can be combined into an approximation of the network adjacency matrix. RDPG results are known to  
120 have strong phylogenetic signal, and to capture the evolutionary backbone of food webs (Dalla Riva &  
121 Stouffer 2016). In addition, recent advances show that the latent variables produced this way can be used  
122 to predict *de novo* network edges (*i.e.* interactions; Runghen *et al.* 2021).

123 The latent variables are created by performing a truncated Singular Value Decomposition (t-SVD) on the  
124 adjacency matrix. SVD is an appropriate embedding of ecological networks, which has recently been  
125 shown to both capture their complex, emerging properties (Strydom *et al.* 2021b) and to allow highly  
126 accurate prediction of the interactions within a single network (Poisot *et al.* 2021b). Under SVD, an  
127 adjacency matrix  $\mathbf{A}$  (where  $\mathbf{A}_{m,n} \in \mathbb{B}$  where 1 indicates predation and 0 an absence thereof) is  
128 decomposed into three components resulting in  $\mathbf{A} = \mathbf{L}\Sigma\mathbf{R}'$ . Here,  $\Sigma$  is a  $m \times n$  diagonal matrix and  
129 contains only singular ( $\sigma$ ) values along its diagonal,  $\mathbf{L}$  is a  $m \times m$  unitary matrix, and  $\mathbf{R}'$  a  $n \times n$  unitary  
130 matrix. Truncating the SVD removes additional noise in the dataset by omitting non-zero and/or smaller  
131  $\sigma$  values from  $\Sigma$  using the rank of the matrix. Under a t-SVD  $\mathbf{A}_{m,n}$  is decomposed so that  $\Sigma$  is a square  $r \times r$   
132 diagonal matrix (whith  $1 \leq r \leq r_{full}$  where  $r_{full}$  is the full rank of  $\mathbf{A}$  and  $r$  the rank at which we truncate  
133 the matrix) containing only non-zero  $\sigma$  values. Additionally,  $\mathbf{L}$  is now a  $m \times r$  semi unitary matrix and  $\mathbf{R}'$  a  
134  $n \times r$  semi-unitary matrix.

135 The specific rank at which the SVD ought to be truncated is a difficult question. The purpose of SVD is to  
136 remove the noise (expressed at high dimensions) and to focus on the signal, (expressed at low dimensions).  
137 In datasets with a clear signal/noise demarcation, a scree plot of  $\Sigma$  can show a sharp drop at the rank where  
138 noise starts (Zhu & Ghodsi 2006). Because the European metaweb is almost entirely known, the amount  
139 of noise (uncertainty) is low; this is reflected in fig. 2 (left), where the scree plot shows no important drop,  
140 and in fig. 2 (right) where the proportion of variance explained increases smoothly at higher dimensions.  
141 For this reason, we default back to a threshold that explains 60% of the variance in the underlying data,

142 corresponding to 12 dimensions - *i.e.* a tradeoff between accuracy and a reduced number of features.

143 A RDPG estimates the probability of observing interactions between nodes (species) as a function of the  
144 nodes' latent variables. The latent variables used for the RDPG, called the left and right subspaces, are  
145 defined as  $\mathcal{L} = \mathbf{L}\sqrt{\Sigma}$ , and  $\mathcal{R} = \sqrt{\Sigma}\mathbf{R}$  – using the full rank of  $\mathbf{A}$ ,  $\mathcal{L}\mathcal{R}' = \mathbf{A}$ , and using any smaller rank  
146 results in  $\mathcal{L}\mathcal{R}' \approx \mathbf{A}$ . Using a rank of 1 for the t-SVD provides a first-order approximation of the network.

147 [Figure 2 about here.]

148 Because RDPG relies on matrix multiplication, the higher dimensions essentially serve to make specific  
149 interactions converge towards 0 or 1; therefore, for reasonably low ranks, there is no guarantee that the  
150 values in the reconstructed network will be within the unit range. In order to determine what constitutes  
151 an appropriate threshold for probability, we performed the RDPG approach on the European metaweb,  
152 and evaluated the probability threshold by treating this as a binary classification problem, specifically  
153 assuming that both 0 and 1 in the European metaweb are all true. Given the methodological details given  
154 in Maiorano *et al.* (2020b) and O'Connor *et al.* (2020), this seems like a reasonable assumption, although  
155 one that does not hold for all metawebs. We used the thresholding approach presented in Poisot *et al.*  
156 (2021b), and picked a cutoff that maximized Youden's *J* statistic (a measure of the informedness (trust) of  
157 predictions; Youden (1950)); the resulting cutoff was 0.22, and gave an accuracy above 0.99.

158 The left and right subspaces for the European metaweb, accompanied by the threshold for prediction,  
159 represent the knowledge we seek to transfer. In the next section, we explain how we rely on phylogenetic  
160 similarity to do so.

## 161 **Steps 2 and 3: Transfer learning through phylogenetic relatedness**

162 In order to transfer the knowledge from the European metaweb to the Canadian species pool, we  
163 performed ancestral character estimation using a Brownian motion model, which is a conservative  
164 approach in the absence of strong hypotheses about the nature of phylogenetic signal in the network  
165 decomposition (Litsios & Salamin 2012). This uses the estimated feature vectors for the European  
166 mammals to create a state reconstruction for all species (conceptually something akin to a trait-based  
167 mammalian phylogeny using generality and vulnerability traits) and allows us to impute the missing  
168 (latent) trait data for the Canadian species that are not already in the European network; as we are focused

169 on predicting contemporary interactions, we only retained the values for the tips of the tree. We assumed  
170 that all traits (*i.e.* the feature vectors for the left and right subspaces) were independent, which is a  
171 reasonable assumption as every trait/dimension added to the t-SVD has an *additive* effect to the one before  
172 it. Note that the Upham *et al.* (2019) tree itself has some uncertainty associated to inner nodes of the  
173 phylogeny. In this case study, we have decided to not propagate this uncertainty, as it would complexify  
174 the process. The Brownian motion algorithm returns the *average* value of the trait, and its upper and  
175 lower bounds. Because we do not estimate other parameters of the traits' distributions, we considered that  
176 every species trait is represented as a uniform distribution between these bounds; in a situation where the  
177 algorithm would return point values for all simulations, one could in theory either estimate the  
178 parameters of a distribution for each tip, or draw randomly from the outputs. In all cases, the inferred left  
179 and right sub-spaces for the Canadian species pool ( $\hat{\mathcal{L}}$  and  $\hat{\mathcal{R}}$ ) have entries that are distributions,  
180 representing the range of values for a given species at a given dimension.  
  
181 These objects represent the transferred knowledge, which we can use for prediction of the Canadian  
182 metaweb.

#### 183 Step 4: Probabilistic prediction of the destination network

184 The phylogenetic reconstruction of  $\hat{\mathcal{L}}$  and  $\hat{\mathcal{R}}$  has an associated uncertainty, represented by the breadth of  
185 the uniform distribution associated to each of their entries. Therefore, we can use this information to  
186 assemble a *probabilistic* metaweb in the sense of Poisot *et al.* (2016), *i.e.* in which every interaction is  
187 represented as a single, independent, Bernoulli event of probability  $p$ .

188 [Figure 3 about here.]

189 Specifically, we have adopted the following approach. For every entry in  $\hat{\mathcal{L}}$  and  $\hat{\mathcal{R}}$ , we draw a value from  
190 its distribution. This results in one instance of the possible left ( $\hat{\ell}$ ) and right ( $\hat{r}$ ) subspaces for the  
191 Canadian metaweb. These can be multiplied, to produce one matrix of real values. Because the entries in  
192  $\hat{\ell}$  and  $\hat{r}$  are in the same space where  $\mathcal{L}$  and  $\mathcal{R}$  were originally predicted, it follows that the threshold  $\rho$   
193 estimated for the European metaweb also applies. We use this information to produce one random  
194 Canadian metaweb,  $N = \hat{\mathcal{L}}\hat{\mathcal{R}}' \geq \rho$ . As we can see in (fig. 3) the European and Canadian metawebs are  
195 structurally similar (as would be expected given the biogeographic similarities) and the two (left and right)  
196 subspaces are distinct *i.e.* capturing predation (generality) and prey (vulnerability) traits.

197 Because the intervals around some trait values can be broad (in fact, probably broader than what they  
198 would actually be, see e.g. Garland *et al.* 1999), we repeat the above process  $2 \times 10^5$  times, which results in  
199 a probabilistic metaweb  $P$ , where the probability of an interaction (here conveying our degree of trust that  
200 it exists given the inferred trait distributions) is given by the number of times where it appears across all  
201 random draws  $N$ , divided by the number of samples. An interaction with  $P_{i,j} = 1$  means that these two  
202 species were predicted to interact in all  $2 \times 10^5$  random draws.

203 **Data cleanup, discovery, validation, and thresholding**

204 Once the probabilistic metaweb for Canada has been produced, we followed a number of data inflation  
205 steps to finalize it. This step is external to the actual transfer learning framework but rather serves as a  
206 way to augment and validate the predicted metaweb.

207 [Figure 4 about here.]

208 First, we extracted the subgraph corresponding to the 17 species shared between the European and  
209 Canadian pools and replaced these interactions with a probability of 0 (non-interaction) or 1 (interaction),  
210 according to their value in the European metaweb. This represents a minute modification of the inferred  
211 network (about 0.8% of all species pairs from the Canadian web), but ensures that we are directly re-using  
212 knowledge from Europe.

213 Second, we looked for all species in the Canadian pool known to the Global Biotic Interactions (GLOBI)  
214 database (Poelen *et al.* 2014), and extracted their known interactions. Because GLOBI aggregates observed  
215 interactions, it is not a *networks* data source, and therefore the only information we can reliably extract  
216 from it is that a species pair *was reported to interact at least once*. This last statement should yet be taken  
217 with caution, as some sources in GLOBI (e.g. Thessen & Parr 2014) are produced through text analysis,  
218 and therefore may not document direct evidence of the interaction. Nevertheless, should the predictive  
219 model work, we would expect that a majority of interactions known to GLOBI would also be predicted.  
220 After performing this check, we set the probability of all interactions known to GLOBI (366 in total, 33 of  
221 which were not predicted by the model, for a success rate of 91%) to 1.

222 Finally, we downloaded the data from Strong & Leroux (2014), who mined various literature sources to  
223 identify trophic interactions in Newfoundland. This dataset documented 25 interactions between

224 mammals, only two of which were not part of our (Canada-level) predictions, resulting in a success rate of  
225 92%. These two interactions were added to our predicted metaweb with a probability of 1.

226 [Figure 5 about here.]

227 Because the confidence intervals on the inferred trait space are probably over-estimates, we decided to  
228 apply a thresholding step to the interactions after the data inflation (fig. 5). Cirtwill & Hambäck (2021)  
229 proposed a number of strategies to threshold probabilistic networks. Their methods assume the  
230 underlying data to be tag-based sequencing, which represents interactions as co-occurrences of predator  
231 and prey within the same tags; this is conceptually identical to our Bernoulli-trial based reconstruction of  
232 a probabilistic network. We performed a full analysis of the effect of various cutoffs, and as they either  
233 resulted in removing too few interactions, or removing enough interactions that species started to be  
234 disconnected from the network, we set this threshold for a probability equivalent to 0 to the largest  
235 possible value that still allowed all species to have at least one interaction with a non-zero probability. The  
236 need for this slight deviation from the Cirtwill & Hambäck (2021) method highlights the need for  
237 additional development on network thresholding.

238 **Results and discussion of the case study**

239 In fig. 5, we examine the effect of varying the cutoff on  $P(i \rightarrow j)$  on the number of links, species, and  
240 connectance. Determining a cutoff using the maximum curvature, or central difference approximation of  
241 the second order partial derivative, as suggested by e.g. Cirtwill & Hambäck (2021), results in species being  
242 lost, or almost all links being kept. We therefore settled on the value that allowed all species to remain  
243 with at least one interaction. This result, in and of itself, suggests that additional methodological  
244 developments for the thresholding of probabilistic networks are required.

245 [Figure 6 about here.]

246 The t-SVD embedding is able to learn relevant ecological features for the network. fig. 6 shows that the  
247 first rank correlates linearly with generality and vulnerability (Schoener 1989), i.e. the number of preys  
248 and predators. Importantly, this implies that a rank 1 approximation represents the configuration model

249 for the metaweb, *i.e.* a set of random networks generated from a given degree sequence (Park & Newman  
250 2004). Accounting for the probabilistic nature of the degrees, the rank 1 approximation also represents the  
251 soft configuration model (van der Hoorn *et al.* 2018). Both models are maximum entropy graph models  
252 (Garlaschelli *et al.* 2018), with sharp (all network realizations satisfy the specified degree sequence) and  
253 soft (network realizations satisfy the degree sequence on average) local constraints, respectively. The (soft)  
254 configuration model is an unbiased random graph model widely used by ecologists in the context of null  
255 hypothesis significance testing of network structure (*e.g.* Bascompte *et al.* 2003) and can provide  
256 informative priors for Bayesian inference of network structure (*e.g.* Young *et al.* 2021). It is noteworthy  
257 that for this metaweb, the relevant information was extracted at the first rank. Because the first rank  
258 corresponds to the leading singular value of the system, the results of fig. 6 have a straightforward  
259 interpretation: degree-based processes are the most important in structuring the mammalian food web.

## 260 Discussion

261 One important aspect in which Europe and Canada differ (despite their comparable bioclimatic  
262 conditions) is the degree of the legacy of human impacts, which have been much longer in Europe.  
263 Nenzén *et al.* (2014) showed that even at small scales (the Iberian peninsula), mammal food webs retain  
264 the signal of both climate change and human activity, even when this human activity was orders of  
265 magnitude less important than it is now. Similarly, Yeakel *et al.* (2014) showed that changes in human  
266 occupation over several centuries can lead to food web collapse. Megafauna in particular seems to be very  
267 sensitive to human arrival (Pires *et al.* 2015). In short, there is well-substantiated support for the idea that  
268 human footprint affects more than the risk of species extinction (Marco *et al.* 2018), and can lead to  
269 changes in interaction structure. Yet, owing to the inherent plasticity of interactions, there have been  
270 documented instances of food webs undergoing rapid collapse/recovery cycles over short periods of time  
271 (Pedersen *et al.* 2017). The embedding of a network, in a sense, embeds its macro-evolutionary history,  
272 especially as RDPG captures ecological signal (Dalla Riva & Stouffer 2016); at this point, it is important to  
273 recall that a metaweb is intended as a catalogue of all possible interactions, which should then be filtered  
274 (Morales-Castilla *et al.* 2015). In practice (and in this instance) the reconstructed metaweb will predict  
275 interactions that are plausible based on the species' evolutionary history, however some interactions  
276 would not be realized due to human impact.

277 Cirtwill *et al.* (2019) previously made the point that network inference techniques based on Bayesian  
278 approaches would perform far better in the presence of an interaction-level informative prior; the  
279 desirable properties of such a prior would be that it is expressed as a probability, preferably representing a  
280 Bernoulli event, the value of which would be representative of relevant biological processes (probability of  
281 predation in this case). We argue that the probability returned at the very last step of our framework may  
282 serve as this informative prior; indeed, the output of our analysis can be used in subsequent steps, also  
283 possibly involving expert elicitation to validate some of the most strongly recommended interactions. One  
284 important *caveat* to keep in mind when working with interaction inference is that interactions can never  
285 really be true negatives (in the current state of our methodological framework and data collection  
286 limitations); this renders the task of validating a model through the usual application of binary  
287 classification statistics very difficult (although see Strydom *et al.* 2021a for a discussion of alternative  
288 suggestions). The other way through which our framework can be improved is by substituting the  
289 predictors that are used for transfer. For example, in the presence of information on species traits that are  
290 known to be predictive of species interactions, one might want to rely on functional rather than  
291 phylogenetic distances – in food webs, body size (and allometrically related variables) has been established  
292 as such a variable (Brose *et al.* 2006); the identification of relevant functional traits is facilitated by recent  
293 methodological developments (Rosado *et al.* 2013). It should be noted that Xing & Fayle (2021) highlight  
294 phylogenetic relatedness as one of the core components of network comparison at the global scale. In this  
295 case study, we have embedded the original metaweb using t-SVD, because it lends itself to a RDPC  
296 reconstruction, which is known to capture the consequences of evolutionary processes (Dalla Riva &  
297 Stouffer 2016); this being said, there are others ways to embed graphs (Cai *et al.* 2017; Arsov & Mirceva  
298 2019; Cao *et al.* 2019), which can be used as alternatives.

299 As Herbert (1965) rightfully pointed out, “[y]ou can’t draw neat lines around planet-wide problems”; in  
300 this regard, our approach must contend with two interesting problems. The first is the limit of the  
301 metaweb to embed and transfer. If the initial metaweb is too narrow in scope, notably from a taxonomic  
302 point of view, the chances of finding another area with enough related species to make a reliable inference  
303 decrease. This is notably true if the metaweb is assembled in an area with mostly endemic species.  
304 Conversely, the metaweb should be reliably filled, which assumes that the  $S^2$  interactions in a pool of  $S$   
305 species have been examined, either through literature surveys or expert elicitation. The second problem is  
306 to determine which area should be used to infer the new metaweb in, as this determines the species pool

307 that must be used. In our application, we focused on the mammals of Canada. The upside of this  
308 approach is that information at the country level is likely to be required by policy makers and stakeholders  
309 for their biodiversity assessment, as each country tends to set goals at the national level (Buxton *et al.*  
310 2021) for which quantitative instruments are designed (Turak *et al.* 2017), with specific strategies often  
311 enacted at smaller scales (Ray *et al.* 2021). Yet these national divisions, in large parts of the world, reflect  
312 nothing except for the legacy of settler colonialism, and operating under them must be done under the  
313 clear realization that they contributed to the ongoing biodiversity crisis (Adam 2014), can reinforce  
314 environmental injustice (Choudry 2013; Domínguez & Luoma 2020), and on Turtle Island especially, will  
315 probably end up being replaced by Indigenous principles of land management (Eichhorn *et al.* 2019;  
316 No'kmaq *et al.* 2021).

317 **Acknowledgements:** We acknowledge that this study was conducted on land within the traditional  
318 unceded territory of the Saint Lawrence Iroquoian, Anishinabewaki, Mohawk, Huron-Wendat, and  
319 Omàmiwininiwak nations. TP, TS, DC, and LP received funding from the Canadian Institute for Ecology &  
320 Evolution. FB is funded by the Institut de Valorisation des Données. TS, SB, and TP are funded by a  
321 donation from the Courtois Foundation. CB was awarded a Mitacs Elevate Fellowship no. IT12391, in  
322 partnership with fRI Research, and also acknowledges funding from Alberta Innovates and the Forest  
323 Resources Improvement Association of Alberta. M-JF acknowledges funding from NSERC Discovery  
324 Grant and NSERC CRC. RR is funded by New Zealand's Biological Heritage Ngā Koiora Tuku Iho  
325 National Science Challenge, administered by New Zealand Ministry of Business, Innovation, and  
326 Employment. BM is funded by the NSERC Alexander Graham Bell Canada Graduate Scholarship and the  
327 FRQNT master's scholarship. LP acknowledges funding from NSERC Discovery Grant (NSERC  
328 RGPIN-2019-05771). TP acknowledges financial support from NSERC through the Discovery Grants and  
329 Discovery Accelerator Supplement programs.

## 330 **References**

- 331 Adam, R. (2014). *Elephant treaties: The Colonial legacy of the biodiversity crisis*. UPNE.
- 332 Albouy, C., Archambault, P., Appeltans, W., Araújo, M.B., Beauchesne, D., Cazelles, K., *et al.* (2019). The  
333 marine fish food web is globally connected. *Nature Ecology & Evolution*, 3, 1153–1161.
- 334 Arsov, N. & Mirceva, G. (2019). Network Embedding: An Overview. *arXiv:1911.11726 [cs, stat]*.

- 335 Banville, F., Vissault, S. & Poisot, T. (2021). Mangal.jl and EcologicalNetworks.jl: Two complementary  
336 packages for analyzing ecological networks in Julia. *Journal of Open Source Software*, 6, 2721.
- 337 Bascompte, J., Jordano, P., Melian, C.J. & Olesen, J.M. (2003). The nested assembly of plant-animal  
338 mutualistic networks. *Proceedings of the National Academy of Sciences*, 100, 9383–9387.
- 339 Beckerman, A.P., Petchey, O.L. & Warren, P.H. (2006). Foraging biology predicts food web complexity.  
340 *Proceedings of the National Academy of Sciences*, 103, 13745–13749.
- 341 Bezanson, J., Edelman, A., Karpinski, S. & Shah, V. (2017). Julia: A Fresh Approach to Numerical  
342 Computing. *SIAM Review*, 59, 65–98.
- 343 Boeckaerts, D., Stock, M., Criels, B., Gerstmans, H., De Baets, B. & Briers, Y. (2021). Predicting  
344 bacteriophage hosts based on sequences of annotated receptor-binding proteins. *Scientific Reports*, 11,  
345 1467.
- 346 Braga, M.P., Janz, N., Nylin, S., Ronquist, F. & Landis, M.J. (2021). Phylogenetic reconstruction of ancestral  
347 ecological networks through time for pierid butterflies and their host plants. *Ecology Letters*, n/a.
- 348 Brose, U., Jonsson, T., Berlow, E.L., Warren, P., Banasek-Richter, C., Bersier, L.-F., et al. (2006).  
349 ConsumerResource Body-Size Relationships in Natural Food Webs. *Ecology*, 87, 2411–2417.
- 350 Buxton, R.T., Bennett, J.R., Reid, A.J., Shulman, C., Cooke, S.J., Francis, C.M., et al. (2021). Key  
351 information needs to move from knowledge to action for biodiversity conservation in Canada.  
352 *Biological Conservation*, 256, 108983.
- 353 Cai, H., Zheng, V.W. & Chang, K.C.-C. (2017). A Comprehensive Survey of Graph Embedding: Problems,  
354 Techniques and Applications. *arXiv preprint arXiv:1709.07604*.
- 355 Cameron, E.K., Sundqvist, M.K., Keith, S.A., CaraDonna, P.J., Mousing, E.A., Nilsson, K.A., et al. (2019).  
356 Uneven global distribution of food web studies under climate change. *Ecosphere*, 10, e02645.
- 357 Cao, R.-M., Liu, S.-Y. & Xu, X.-K. (2019). Network embedding for link prediction: The pitfall and  
358 improvement. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 29, 103102.
- 359 Cavender-Bares, J., Kozak, K.H., Fine, P.V.A. & Kembel, S.W. (2009). The merging of community ecology  
360 and phylogenetic biology. *Ecology Letters*, 12, 693–715.
- 361 Choudry, A. (2013). Saving biodiversity, for whom and for what? Conservation NGOs, complicity,  
362 colonialism and conquest in an era of capitalist globalization. In: *NGOization: Complicity*,

- 363        *contradictions and prospects*. Bloomsbury Publishing, pp. 24–44.
- 364    Cirtwill, A.R., Eklf, A., Roslin, T., Wootton, K. & Gravel, D. (2019). A quantitative framework for  
365        investigating the reliability of empirical network construction. *Methods in Ecology and Evolution*, 0.
- 366    Cirtwill, A.R. & Hambäck, P. (2021). Building food networks from molecular data: Bayesian or  
367        fixed-number thresholds for including links. *Basic and Applied Ecology*, 50, 67–76.
- 368    Dalla Riva, G.V. & Stouffer, D.B. (2016). Exploring the evolutionary signature of food webs' backbones  
369        using functional traits. *Oikos*, 125, 446–456.
- 370    Dansereau, G. & Poisot, T. (2021). SimpleSDMLayers.jl and GBIF.jl: A Framework for Species Distribution  
371        Modeling in Julia. *Journal of Open Source Software*, 6, 2872.
- 372    Domínguez, L. & Luoma, C. (2020). Decolonising Conservation Policy: How Colonial Land and  
373        Conservation Ideologies Persist and Perpetuate Indigenous Injustices at the Expense of the  
374        Environment. *Land*, 9, 65.
- 375    Dormann, C.F., Gruber, B., Winter, M. & Herrmann, D. (2010). Evolution of climate niches in European  
376        mammals? *Biology Letters*, 6, 229–232.
- 377    Dunne, J.A. (2006). The Network Structure of Food Webs. In: *Ecological networks: Linking structure and*  
378        *dynamics* (eds. Dunne, J.A. & Pascual, M.). Oxford University Press, pp. 27–86.
- 379    Eichhorn, M.P., Baker, K. & Griffiths, M. (2019). Steps towards decolonising biogeography. *Frontiers of*  
380        *Biogeography*, 12, 1–7.
- 381    Eklöf, A. & Stouffer, D.B. (2016). The phylogenetic component of food web structure and intervality.  
382        *Theoretical Ecology*, 9, 107–115.
- 383    Garland, T., JR., Midford, P.E. & Ives, A.R. (1999). An Introduction to Phylogenetically Based Statistical  
384        Methods, with a New Method for Confidence Intervals on Ancestral Values1. *American Zoologist*, 39,  
385        374–388.
- 386    Garlaschelli, D., Hollander, F. den & Roccaverde, A. (2018). Covariance structure behind breaking of  
387        ensemble equivalence in random graphs. *Journal of Statistical Physics*, 173, 644–662.
- 388    GBIF Secretariat. (2021). GBIF Backbone Taxonomy.
- 389    Gerhold, P., Cahill, J.F., Winter, M., Bartish, I.V. & Prinzing, A. (2015). Phylogenetic patterns are not

- 390 proxies of community assembly mechanisms (they are far better). *Functional Ecology*, 29, 600–614.
- 391 Grenié, M., Berti, E., Carvajal-Quintero, J.D., Winter, M. & Sagouis, A. (2021). Harmonizing taxon names  
392 in biodiversity data: A review of tools, databases, and best practices.
- 393 Halko, N., Martinsson, P.G. & Tropp, J.A. (2011). Finding Structure with Randomness: Probabilistic  
394 Algorithms for Constructing Approximate Matrix Decompositions. *SIAM Review*, 53, 217–288.
- 395 Herbert, F. (1965). *Dune*. First. Chilton Book Company, Philadelphia.
- 396 Hortal, J., de Bello, F., Diniz-Filho, J.A.F., Lewinsohn, T.M., Lobo, J.M. & Ladle, R.J. (2015). Seven  
397 Shortfalls that Beset Large-Scale Knowledge of Biodiversity. *Annual Review of Ecology, Evolution, and*  
398 *Systematics*, 46, 523–549.
- 399 Hutchinson, M.C., Cagua, E.F. & Stouffer, D.B. (2017). Cophylogenetic signal is detectable in pollination  
400 interactions across ecological scales. *Ecology*, n/a–n/a.
- 401 Jordano, P. (2016a). Chasing Ecological Interactions. *PLOS Biol*, 14, e1002559.
- 402 Jordano, P. (2016b). Sampling networks of ecological interactions. *Functional Ecology*, 30, 1883–1893.
- 403 Litsios, G. & Salamin, N. (2012). Effects of Phylogenetic Signal on Ancestral State Reconstruction.  
404 *Systematic Biology*, 61, 533–538.
- 405 Maiorano, L., Montemaggioli, A., Ficetola, G.F., O'Connor, L. & Thuiller, W. (2020a). Data from: Tetra-EU  
406 1.0: A species-level trophic meta-web of European tetrapods.
- 407 Maiorano, L., Montemaggiori, A., Ficetola, G.F., O'Connor, L. & Thuiller, W. (2020b). TETRA-EU 1.0: A  
408 species-level trophic metaweb of European tetrapods. *Global Ecology and Biogeography*, 29, 1452–1457.
- 409 Marco, M.D., Venter, O., Possingham, H.P. & Watson, J.E.M. (2018). Changes in human footprint drive  
410 changes in species extinction risk. *Nature Communications*, 9, 4621.
- 411 Morales-Castilla, I., Matias, M.G., Gravel, D. & Araújo, M.B. (2015). Inferring biotic interactions from  
412 proxies. *Trends in Ecology & Evolution*, 30, 347–356.
- 413 Mouquet, N., Devictor, V., Meynard, C.N., Munoz, F., Bersier, L.-F., Chave, J., et al. (2012).  
414 Ecophylogenetics: Advances and perspectives. *Biological Reviews*, 87, 769–785.
- 415 Nenzén, H.K., Montoya, D. & Varela, S. (2014). The Impact of 850,000 Years of Climate Changes on the  
416 Structure and Dynamics of Mammal Food Webs. *PLOS ONE*, 9, e106651.

- 417 No'kmaq, M., Marshall, A., Beazley, K.F., Hum, J., joudry, shalan, Papadopoulos, A., *et al.* (2021).  
418 "Awakening the sleeping giant": Re-Indigenization principles for transforming biodiversity  
419 conservation in Canada and beyond. *FACETS*, 6, 839–869.
- 420 O'Connor, L.M.J., Pollock, L.J., Braga, J., Ficetola, G.F., Maiorano, L., Martinez-Almoyna, C., *et al.* (2020).  
421 Unveiling the food webs of tetrapods across Europe through the prism of the Eltonian niche. *Journal of*  
422 *Biogeography*, 47, 181–192.
- 423 Pan, S.J. & Yang, Q. (2010). A Survey on Transfer Learning. *IEEE Transactions on Knowledge and Data*  
424 *Engineering*, 22, 1345–1359.
- 425 Park, J. & Newman, M.E.J. (2004). Statistical mechanics of networks. *Physical Review E*, 70, 066117.
- 426 Pedersen, E.J., Thompson, P.L., Ball, R.A., Fortin, M.-J., Gouhier, T.C., Link, H., *et al.* (2017). Signatures of  
427 the collapse and incipient recovery of an overexploited marine ecosystem. *Royal Society Open Science*,  
428 4, 170215.
- 429 Pires, M.M., Koch, P.L., Fariña, R.A., de Aguiar, M.A.M., dos Reis, S.F. & Guimarães, P.R. (2015).  
430 Pleistocene megafaunal interaction networks became more vulnerable after human arrival.  
431 *Proceedings of the Royal Society B: Biological Sciences*, 282, 20151367.
- 432 Poelen, J.H., Simons, J.D. & Mungall, C.J. (2014). Global biotic interactions: An open infrastructure to  
433 share and analyze species-interaction datasets. *Ecological Informatics*, 24, 148–159.
- 434 Poisot, T., Belisle, Z., Hoebeke, L., Stock, M. & Szefer, P. (2019). EcologicalNetworks.jl - analysing  
435 ecological networks. *Ecography*.
- 436 Poisot, T., Bergeron, G., Cazelles, K., Dallas, T., Gravel, D., MacDonald, A., *et al.* (2021a). Global  
437 knowledge gaps in species interaction networks data. *Journal of Biogeography*, n/a.
- 438 Poisot, T., Canard, E., Mouillot, D., Mouquet, N. & Gravel, D. (2012). The dissimilarity of species  
439 interaction networks. *Ecology Letters*, 15, 1353–1361.
- 440 Poisot, T., Cirtwill, A.R., Cazelles, K., Gravel, D., Fortin, M.-J. & Stouffer, D.B. (2016). The structure of  
441 probabilistic networks. *Methods in Ecology and Evolution*, 7, 303–312.
- 442 Poisot, T., Ouellet, M.-A., Mollentze, N., Farrell, M.J., Becker, D.J., Albery, G.F., *et al.* (2021b). Imputing the  
443 mammalian virome with linear filtering and singular value decomposition. *arXiv:2105.14973 [q-bio]*.

- 444 Poisot, T. & Stouffer, D.B. (2018). Interactions retain the co-phylogenetic matching that communities lost.  
445      *Oikos*, 127, 230–238.
- 446 Price, P.W. (2003). *Macroevolutionary theory on macroecological patterns*. Cambridge University Press.
- 447 Ray, J.C., Grimm, J. & Olive, A. (2021). The biodiversity crisis in Canada: Failures and challenges of  
448      federal and sub-national strategic and legal frameworks. *FACETS*, 6, 1044–1068.
- 449 Reeve, R., Leinster, T., Cobbold, C.A., Thompson, J., Brummitt, N., Mitchell, S.N., *et al.* (2016). How to  
450      partition diversity. *arXiv:1404.6520 [q-bio]*.
- 451 Rosado, B.H.P., Dias, A. & de Mattos, E. (2013). Going Back to Basics: Importance of Ecophysiology when  
452      Choosing Functional Traits for Studying Communities and Ecosystems. *Natureza & conservação*  
453      *revista brasileira de conservação da natureza*, 11, 15–22.
- 454 Runghen, R., Stouffer, D.B. & Dalla Riva, G.V. (2021). Exploiting node metadata to predict interactions in  
455      large networks using graph embedding and neural networks.
- 456 Schoener, T.W. (1989). Food webs from the small to the large. *Ecology*, 70, 1559–1589.
- 457 Solís-Lemus, C., Bastide, P. & Ané, C. (2017). PhyloNetworks: A Package for Phylogenetic Networks.  
458      *Molecular Biology and Evolution*, 34, 3292–3298.
- 459 Stock, M. (2021). Pairwise learning for predicting pollination interactions based on traits and phylogeny.  
460      *Ecological Modelling*, 14.
- 461 Stouffer, D.B., Sales-Pardo, M., Sirer, M.I. & Bascompte, J. (2012). Evolutionary Conservation of Species'  
462      Roles in Food Webs. *Science*, 335, 1489–1492.
- 463 Strong, J.S. & Leroux, S.J. (2014). Impact of Non-Native Terrestrial Mammals on the Structure of the  
464      Terrestrial Mammal Food Web of Newfoundland, Canada. *PLOS ONE*, 9, e106264.
- 465 Strydom, T., Catchen, M.D., Banville, F., Caron, D., Dansereau, G., Desjardins-Proulx, P., *et al.* (2021a). A  
466      *Roadmap Toward Predicting Species Interaction Networks (Across Space and Time)* (Preprint).  
467      EcoEvoRxiv.
- 468 Strydom, T., Dalla Riva, G.V. & Poisot, T. (2021b). SVD Entropy Reveals the High Complexity of Ecological  
469      Networks. *Frontiers in Ecology and Evolution*, 9.
- 470 Thessen, A.E. & Parr, C.S. (2014). Knowledge extraction and semantic annotation of text from the

- 471 encyclopedia of life. *PLoS one*, 9, e89550.
- 472 Torrey, L. & Shavlik, J. (2010). Transfer learning. In: *Handbook of research on machine learning*  
473 *applications and trends: Algorithms, methods, and techniques*. IGI global, pp. 242–264.
- 474 Trøjelsgaard, K. & Olesen, J.M. (2016). Ecological networks in motion: Micro- and macroscopic variability  
475 across scales. *Functional Ecology*, 30, 1926–1935.
- 476 Turak, E., Brazill-Boast, J., Cooney, T., Drielsma, M., DelaCruz, J., Dunkerley, G., et al. (2017). Using the  
477 essential biodiversity variables framework to measure biodiversity change at national scale. *Biological*  
478 *Conservation*, SI:Measures of biodiversity, 213, 264–271.
- 479 Upham, N.S., Esselstyn, J.A. & Jetz, W. (2019). Inferring the mammal tree: Species-level sets of  
480 phylogenies for questions in ecology, evolution, and conservation. *PLOS Biology*, 17, e3000494.
- 481 van der Hoorn, P., Lippner, G. & Krioukov, D. (2018). Sparse Maximum-Entropy Random Graphs with a  
482 Given Power-Law Degree Distribution. *Journal of Statistical Physics*, 173, 806–844.
- 483 Xing, S. & Fayle, T.M. (2021). The rise of ecological network meta-analyses: Problems and prospects.  
484 *Global Ecology and Conservation*, 30, e01805.
- 485 Yeakel, J.D., Pires, M.M., Rudolf, L., Dominy, N.J., Koch, P.L., Guimarães, P.R., et al. (2014). Collapse of an  
486 ecological network in Ancient Egypt. *PNAS*, 111, 14472–14477.
- 487 Youden, W.J. (1950). Index for rating diagnostic tests. *Cancer*, 3, 32–35.
- 488 Young, J.-G., Cantwell, G.T. & Newman, M.E.J. (2021). Bayesian inference of network structure from  
489 unreliable data. *Journal of Complex Networks*, 8.
- 490 Young, S.J. & Scheinerman, E.R. (2007). Random Dot Product Graph Models for Social Networks. In:  
491 *Algorithms and Models for the Web-Graph*, Lecture Notes in Computer Science (eds. Bonato, A. &  
492 Chung, F.R.K.). Springer, Berlin, Heidelberg, pp. 138–149.
- 493 Zhu, M. & Ghodsi, A. (2006). Automatic dimensionality selection from the scree plot via the use of profile  
494 likelihood. *Computational Statistics & Data Analysis*, 51, 918–930.

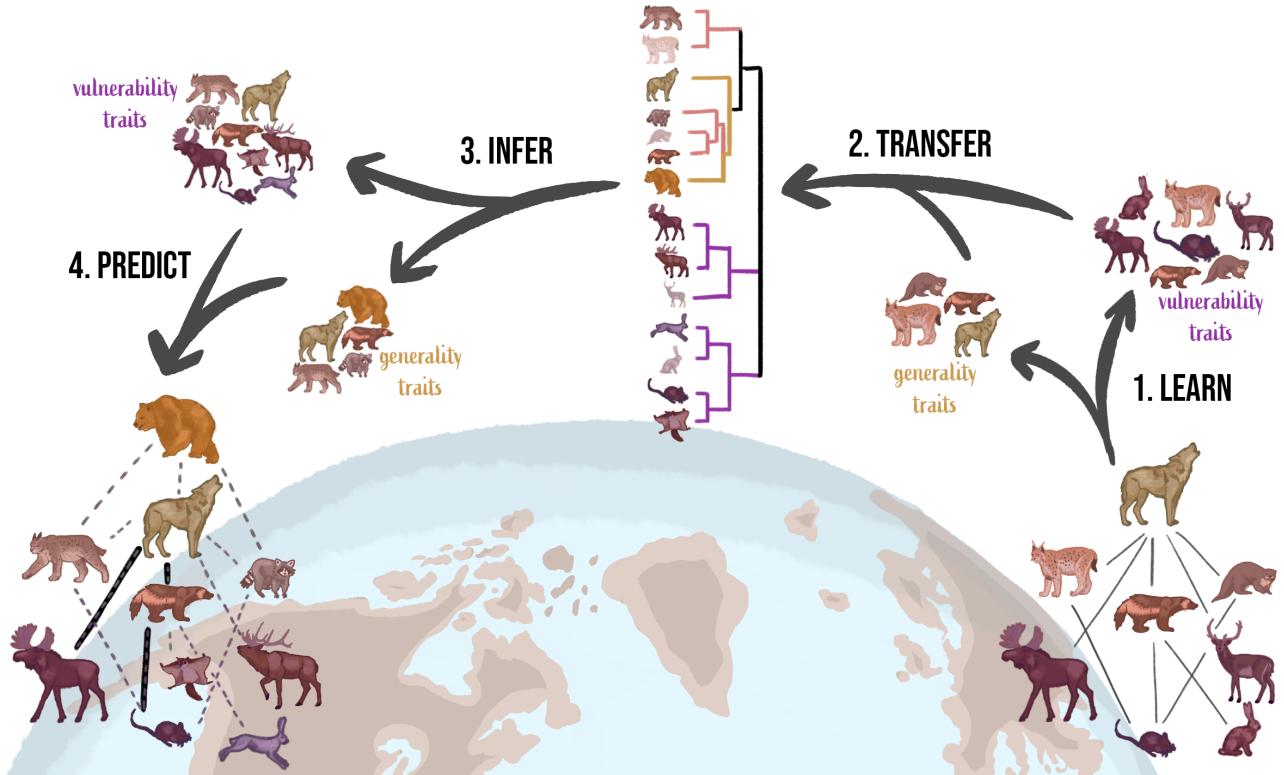


Figure 1: Overview of the phylogenetic transfer learning (and prediction) of species interaction networks. Starting from an initial, known, network, we learn its representation through a graph embedding step (here, a truncated Singular Value Decomposition; Step 1), yielding a series of latent traits (vulnerability traits representing species at the lower trophic-level and generality traits representing species at higher trophic-levels; *sensu* Schoener (1989)); second, for the destination species pool, we perform ancestral character estimation using a phylogeny (here, using a Brownian model for the latent traits; Step 2); we then sample from the reconstructed distribution of latent traits (Step 3) to generate a probabilistic metaweb at the destination (here, assuming a uniform distribution of traits), and threshold it to yield the final list of interactions (Step 4).

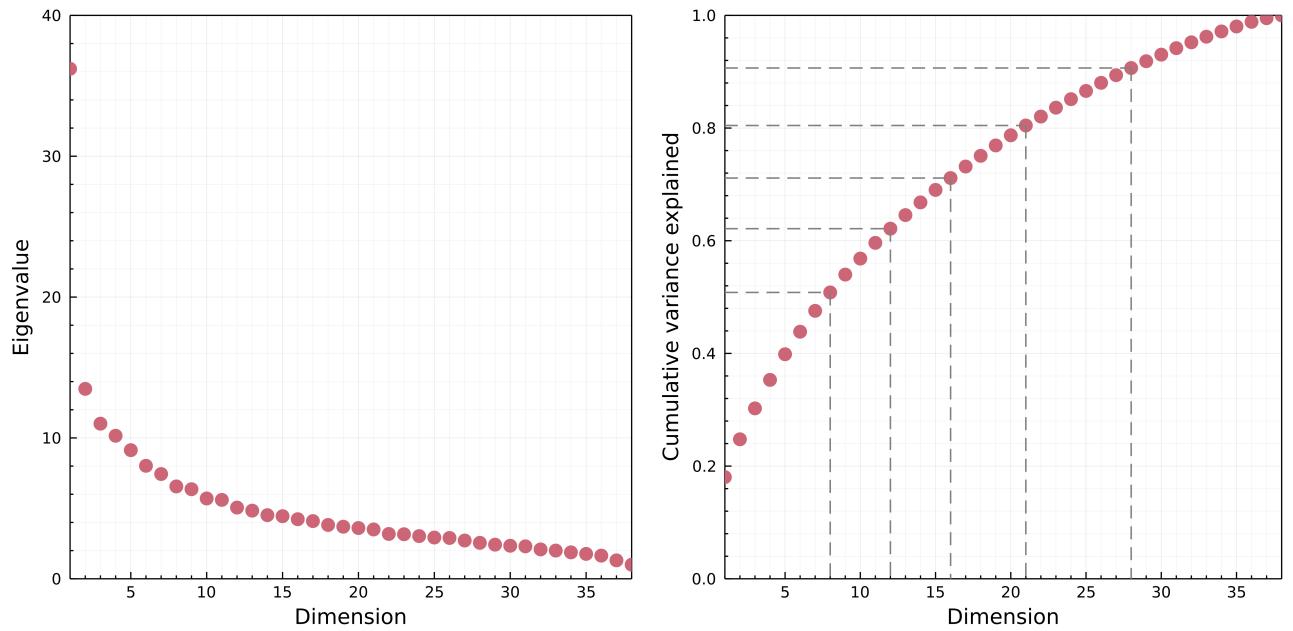


Figure 2: Left: representation of the screeplot of the singular values from the t-SVD on the European metaweb. The screeplot shows no obvious drop in the singular values that may be leveraged to automatically detect a minimal dimension for embedding, after e.g. Zhu & Ghodsi (2006). Right: cumulative fraction of variance explained by each dimension up to the rank of the European metaweb. The grey lines represent cutoffs at 50, 60... 90% of variance explained. For the rest of the analysis, we reverted to an arbitrary threshold of 60% of variance explained, which represented a good tradeoff between accuracy and reduced number of features.

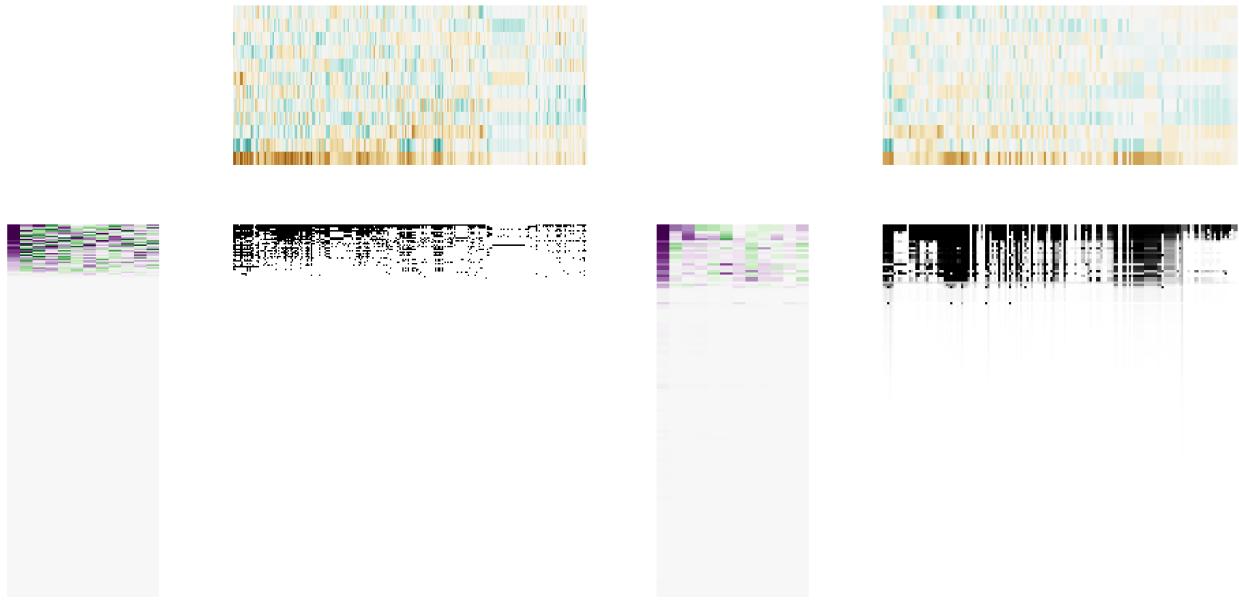


Figure 3: Visual representation of the left (green/purple) and right (green/brown) subspaces, alongside the adjacency matrix of the food web they encode (greyscale). The European metaweb is on the left, and the imputed Canadian metaweb (before data inflation) on the right. This figure illustrates how much structure the left sub-space captures. As we show in fig. 6, the species with a value of 0 in the left subspace are species without any prey.

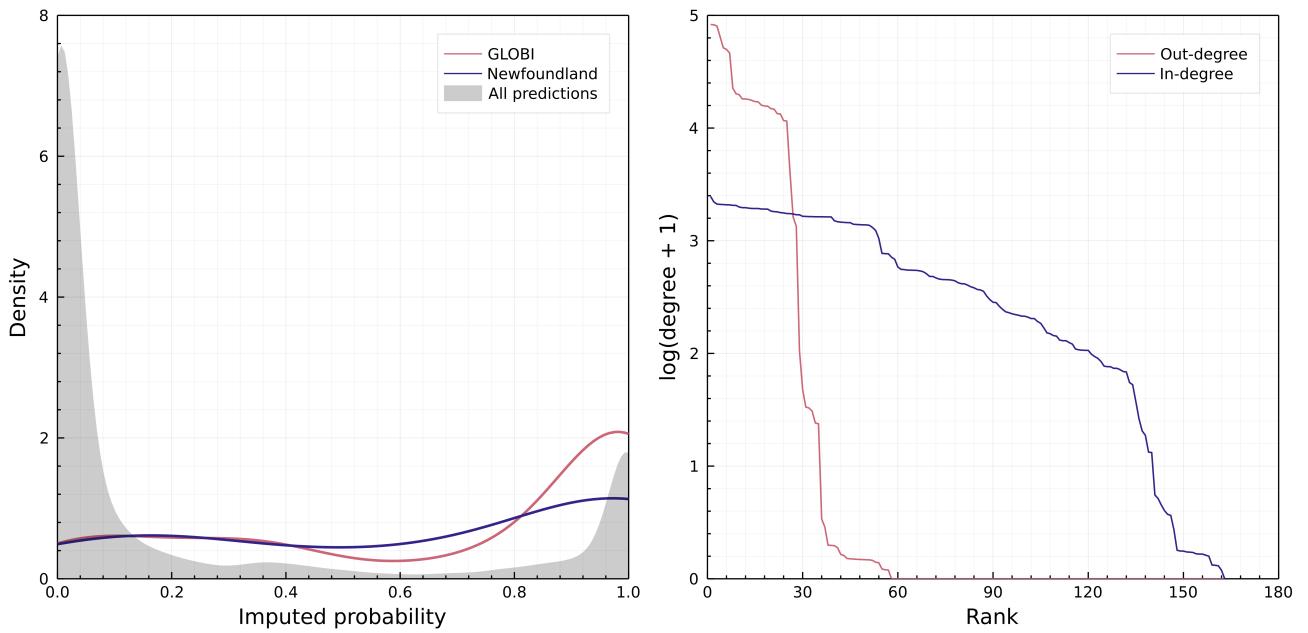


Figure 4: Left, comparison of the probabilities of interactions assigned by the model to all interactions (grey curve), the subset of interactions found in GLOBI (red), and in the Strong & Leroux (2014) Newfoundland dataset (blue). The model recovers more interaction with a low probability compared to data mining, which can suggest that collected datasets are biased towards more common or easy to identify interactions. Right, distribution of the in-degree and out-degree of the mammals from Canada in the reconstructed metaweb. This figure describes a flat, relatively short food web, in which there are few predators but a large number of preys.

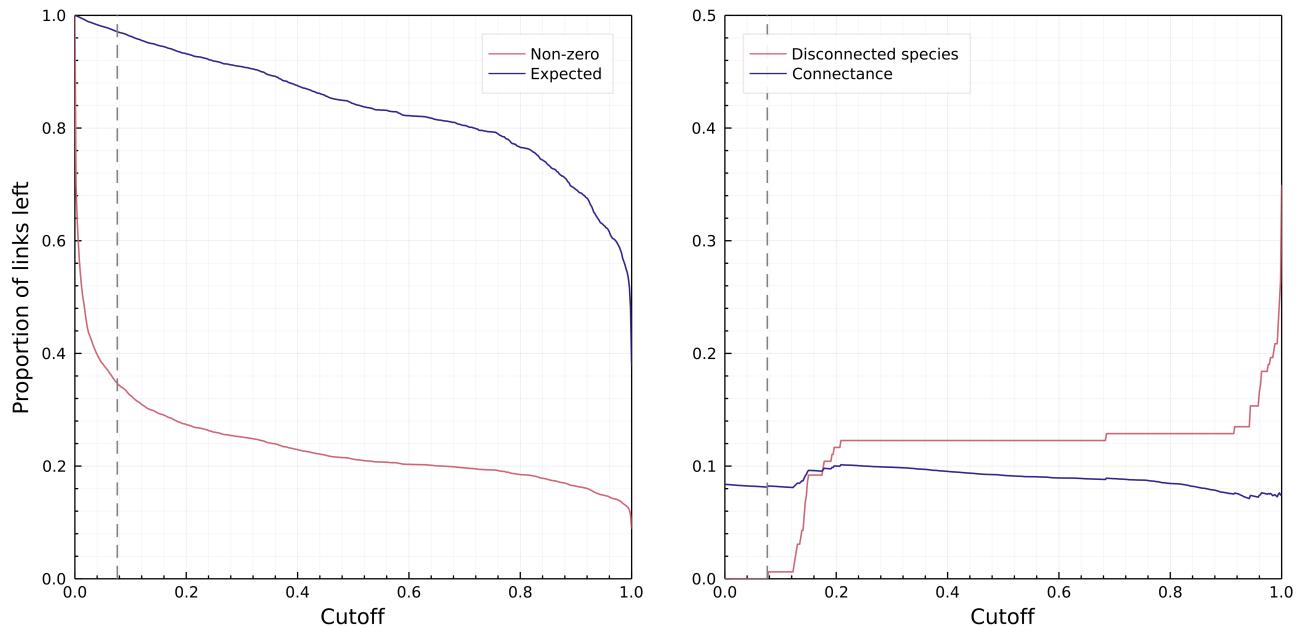


Figure 5: Left: effect of varying the cutoff for probabilities to be considered non-zero on the number of unique links and on  $\hat{L}$ , the probabilistic estimate of the number of links assuming that all interactions are independent. Right: effect of varying the cutoff on the number of disconnected species, and on network connectance. In both panels, the grey line indicates the cutoff  $P(i \rightarrow j) \approx 0.08$  that resulted in the first species losing all of its interactions.

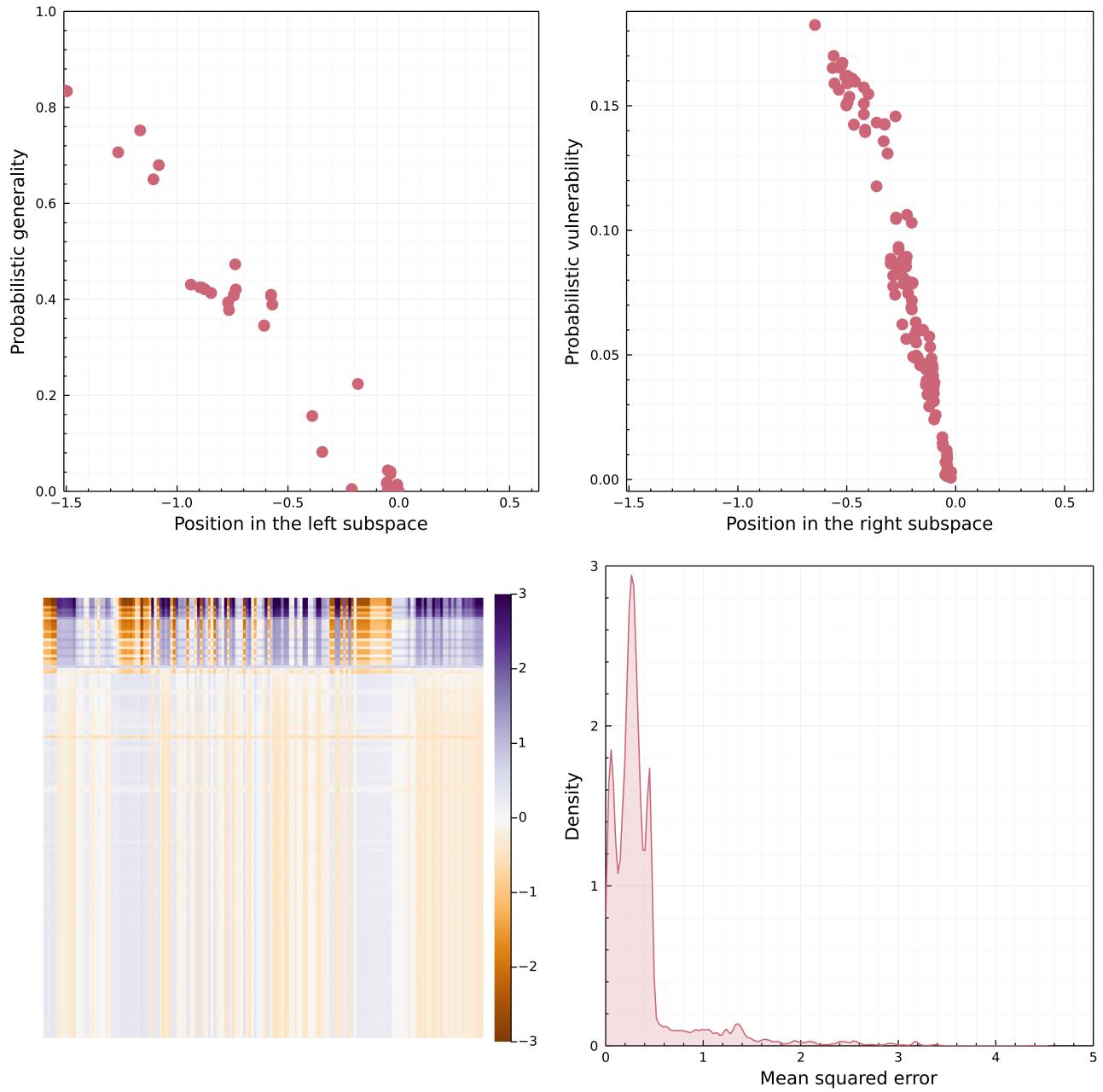


Figure 6: Top: biological significance of the first dimension. Left: there is a linear relationship between the values on the first dimension of the left subspace and the generality, *i.e.* the relative number of preys, *sensu* Schoener (1989). Species with a value of 0 in this subspace are at the bottom-most trophic level. Right: there is, similarly, a linear relationship between the position of a species on the first dimension of the right subspace and its vulnerability, *i.e.* the relative number of predators. Taken together, these two figures show that the first-order representation of this network would capture its degree distribution. Bottom: topological consequences of the first dimension. Left: differences in the z-score of the actual configuration model for the reconstructed network, and the prediction based only on the first dimension. Right: distribution of the differences in the left panel.