

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

Tanya Strydom^{1,2,‡} Salomé Bouskila^{1,‡} Francis Banville^{1,3,2} Ceres Barros⁴ Dominique Caron^{5,2}
Maxwell J Farrell⁶ Marie-Josée Fortin⁶ Victoria Hemming⁷ Benjamin Mercier^{3,2} Laura
J. Pollock^{5,2} Rogini Runghen⁸ Giulio V. Dalla Riva⁹ Timothée Poisot^{1,2}

¹ Département de Sciences Biologiques, Université de Montréal, Montréal, Canada ² Quebec Centre for Biodiversity Science, Montréal, Canada ³ Département de Biologie, Université de Sherbrooke, Sherbrooke, Canada ⁴ Department of Forest Resources Management, University of British Columbia, Vancouver, B.C., Canada ⁵ Department of Biology, McGill University, Montréal, Canada ⁶ Department of Ecology & Evolutionary Biology, University of Toronto, Toronto, Canada ⁷ Department of Forest and Conservation Sciences, University of British Columbia, Vancouver, Canada ⁸ Centre for Integrative Ecology, School of Biological Sciences, University of Canterbury, Canterbury, New Zealand ⁹ School of Mathematics and Statistics, University of Canterbury, Canterbury, New Zealand

[‡] These authors contributed equally to the work

Correspondance to:

Timothée Poisot — timothee.poisot@umontreal.ca

1. Despite their importance in many ecological processes, collecting data and information on ecological interactions 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, community ecologists must adopt predictive methods.
2. We present a methodological framework that uses graph embedding and transfer learning to assemble a predicted list of trophic interactions of a species pool for which their interactions are unknown. Specifically, we ‘learn’ the information from a known interaction network by inferring the latent traits of species and infer the latent traits of a species pool for which we have no *a priori* interaction data based on their phylogenetic relatedness to species from the known network. The latent traits can then be used to predict interactions and construct an interaction network.
3. Here we assembled a metaweb for Canadian mammals derived from interactions in the European food web, despite only 4% of common species being shared between the two locations. The results of the predictive model are compared against databases of recorded pairwise interactions, showing that we correctly recover 91% of known interactions.
4. The framework itself is robust even when the known network is incomplete or contains spurious interactions making it an ideal candidate as a tool for filling gaps when it comes to species interactions. We provide guidance on how this framework can be adapted by substituting some approaches or predictors in order to make it more generally applicable.

1 Introduction

2 There are two core challenges we are faced with in furthering our understanding of ecological networks
3 across space, particularly at macro-ecologically relevant scales (e.g. Trøjelsgaard & Olesen, 2016). First,
4 ecological networks within a location are difficult to sample properly (Jordano, 2016a, 2016b), resulting in
5 a widespread “Eltonian shortfall” (Hortal et al., 2015), *i.e.* a lack of knowledge about inter- and intra-
6 specific relationships. This first challenge has been, in large part, addressed by the recent emergence of a
7 suite of methods aiming to predict interactions within *existing* networks, many of which are reviewed in
8 Strydom, Catchen, et al. (2021). Second, recent analyses based on collected data (Poisot, Bergeron, et al.,
9 2021) or metadata (Cameron et al., 2019) highlight that ecological networks are currently studied in a
10 biased subset of space and bioclimates, which impedes our ability to generalize any local understanding of
11 network structure. Meaning that, although the framework to address incompleteness *within* networks
12 exists, there would still be regions for which, due to a *lack* of local interaction data, we are unable to infer
13 potential species interactions. Having a general solution for inferring *potential* interactions (despite the
14 unavailability of interaction data) could be the catalyst for significant breakthroughs in our ability to start
15 thinking about species interaction networks over large spatial scales. In a recent overview of the field of
16 ecological network prediction, Strydom, Catchen, et al. (2021) identified two challenges of interest to the
17 prediction of interactions at large scales. First, there is a relative scarcity of relevant data in most places
18 globally – paradoxically, this restricts our ability to infer interactions to locations where inference is
19 perhaps the least required; second, accurate predictions often demand accurate predictors, and the lack of
20 methods that can leverage small amount of data is a serious impediment to our predictive ability globally.

21 Here, we present a general method to infer potential trophic interactions, relying on the transfer learning
22 of network representations, specifically by using similarities of species in a biologically/ecologically
23 relevant proxy space (e.g. shared morphology or ancestry). Transfer learning is a machine learning
24 methodology that uses the knowledge gained from solving one problem and applying it to a related
25 (destination) problem (Pan & Yang, 2010; Torrey & Shavlik, 2010). In this instance, we solve the problem
26 of predicting trophic interactions between species, based on knowledge extracted from another species
27 pool for which interactions are known by using phylogenetic structure as a medium for transfer. There is a
28 plurality of measures of species similarities that can be used for inferring *potential* species interactions *i.e.*
29 metaweb reconstruction (see *e.g.* Morales-Castilla et al., 2015); however, phylogenetic proximity has

several desirable properties when working at large scales. Gerhold et al. (2015) made the point that phylogenetic signal captures diversification of characters (large macro-evolutionary process), but not necessarily community assembly (fine ecological process); Dormann et al. (2010) previously found very similar conclusions. Interactions tend to reflect a phylogenetic signal because they have a conserved pattern of evolutionary convergence that encompasses a wide range of ecological and evolutionary mechanisms (Cavender-Bares et al., 2009; Mouquet et al., 2012), and - most importantly - retain this signal even if it is obscured at the community scale due to e.g. local conditions (Hutchinson et al., 2017; Poisot & Stouffer, 2018). Finally, species interactions at macro-ecological scales seem to respond mostly to macro-evolutionary processes (Price, 2003); which is evidenced by the presence of conserved backbones in food webs (Dalla Riva & Stouffer, 2016; Mora et al., 2018), strong evolutionary signature on prey choice (Stouffer et al., 2012), and strong phylogenetic signature in food web intervalty (Eklöf & Stouffer, 2016). Phylogenetic reconstruction has also previously been used within the context of ecological networks, namely understanding ancestral plant-insect interactions (Braga et al., 2021). Taken together, these considerations suggest that phylogenies can reliably be used to transfer knowledge on species interactions.

[Figure 1 about here.]

In fig. 1, where we provide a methodological overview based on learning the embedding of a metaweb of trophic interactions for European mammals (known interactions; Maiorano et al., 2020a, 2020b) and, based on phylogenetic relationships between mammals globally (*i.e.*, phylogenetic tree Upham et al., 2019), infer a metaweb for the Canadian mammalian species pool (using only a species list *i.e.* interactions are ‘unknown’ in this instance). Following the definition of Dunne (2006), a metaweb is a network analogue to the regional species pool; specifically, it is an inventory of all *potential* interactions between species from a spatially delimited area (and so captures the γ diversity of interactions). The metaweb is, therefore, *not* a prediction of the food web at a specific locale within the spatial area it covers, and will have a different structure (notably by having a larger connectance; see *e.g.* Wood et al., 2015). These local food webs (which captures the α diversity of interactions) are a subset of the metaweb’s species and interactions, and have been called “metaweb realizations” (Poisot et al., 2015). Differences between local food web and their metaweb are due to chance, species abundance and co-occurrence, local environmental conditions, and local distribution of functional traits, among others.

Because the metaweb represents the joint effect of functional, phylogenetic, and macroecological

59 processes (Morales-Castilla et al., 2015), it holds valuable ecological information. Specifically, it is the
60 “upper bounds” on what the composition of the local networks can be (see e.g. McLeod et al., 2021). These
61 local networks, in turn, can be reconstructed given appropriate knowledge of local species composition,
62 providing information on structure of food webs at finer spatial scales. This has been done for example for
63 tree-galler-parasitoid systems (Gravel et al., 2018), fish trophic interactions (Albouy et al., 2019), tetrapod
64 trophic interactions (O’Connor et al., 2020), and crop-pest networks (Grünig et al., 2020). Whereas the
65 original metaweb definition, and indeed most past uses of metawebs, was based on the presence/absence
66 of interactions, we focus on *probabilistic* metawebs where interactions are represented as the chance of
67 success of a Bernoulli trial (see e.g. Poisot et al., 2016); therefore, not only does our method recommend
68 interactions that may exist, it gives each interaction a score, allowing us to properly weigh them.

69 Our case study shows that phylogenetic transfer learning is an effective approach to the generation of
70 probabilistic metawebs. This showcases that although the components (species) that make up the
71 Canadian and European communities may be *minimally* shared (the overall species overlap is less than
72 4%), if the medium (proxy space) selected in the transfer step is biologically plausible, we can still
73 effectively learn from the known network and make biologically relevant predictions of interactions.

74 Indeed, as we detail in the results, when validated against known but fractional data of trophic
75 interactions between Canadian mammals, our model achieves a predictive accuracy of approximately 91%.
76 It should be reiterated that the framework presented in fig. 1 is amenable to changes e.g. it is possible to
77 use distinct trees if working with distinct clades (such as pollination networks) or, alternatively, the
78 measure of similarity may not be phylogeny, and can be replaced by information on foraging (Beckerman
79 et al., 2006), cell-level mechanisms (Boeckaerts et al., 2021), or a combination of traits and phylogenetic
80 structure (Stock, 2021). Most importantly, although we focus on a trophic system, it is an established fact
81 that different (non-trophic) interactions do themselves interact with and influence the outcome of trophic
82 interactions (see e.g. Kawatsu et al., 2021; Kéfi et al., 2012). Future development of metaweb inference
83 techniques should cover the prediction of multiple interaction types.

84 **Method description**

85 The core point of our method is the transfer of knowledge of a known ecological network, in order to
86 predict interactions between species from another location for which the network is unknown (or partially

known). In fig. 1 the grey text boxes give a high-level overview of the approach; in the example around which this manuscript is built The method we develop is, ecologically speaking, a “black box,” i.e. an algorithm that can be understood mathematically, but whose component parts are not always directly tied to ecological processes. There is a growing realization in machine learning that (unintentional) black box algorithms are not necessarily a bad thing (Holm, 2019), as long as their constituent parts can be examined (which is the case with our method). But more importantly, data hold more information than we might think; as such, even algorithms that are disconnected from a model can make correct guesses most of the time (Halevy et al., 2009); in fact, in an instance of ecological forecasting of spatio-temporal systems, model-free approaches (i.e. drawing all of their information from the data) outperformed model-informed ones (Perretti et al., 2013).

97 Data used for the case study

We use data from the European metaweb assembled by Maiorano et al. (2020a). This was assembled using data extracted from scientific literature (including published papers, books, and grey literature) from the last 50 years and includes all terrestrial tetrapods (mammals, breeding birds, reptiles and amphibians) occurring on the European sub-continent (and Turkey) - with the caveat that only species introduced in historical times and currently naturalized being included. The European metaweb was filtered using the Global Biodiversity Information Facility (GBIF) taxonomic backbone (GBIF Secretariat, 2021) so as to contain only terrestrial and semi-aquatic mammals. As all species had valid matches to the GBIF taxonomy it was used as the backbone for the remaining reconciliation steps namely, the mammalian consensus supertree by Upham et al. (2019) (which is used for the knowledge transfer step) and for the Canadian species list—which was extracted from the International Union for Conservation of Nature (IUCN) checklist, and corresponds to the same selection criteria that was applied by Maiorano et al. (2020a) in the European metaweb. After taxonomic cleaning and reconciliation the mammalian European metaweb has 260 species, and the Canadian species pool 163; of these, 17 (about 4% of the total) are shared, and 89 species from Canada (54%) had at least one congeneric species in Europe. The similarity for both species pools predictably increases with higher taxonomic order, with 19% of shared genera, 47% of shared families, and 75% of shared orders; for the last point, Canada and Europe each had a single unique order (*Didelphimorphia* for Canada, *Erinaceomorpha* for Europe).

115 **Implementation and code availability**

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

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

127 The first step in transfer learning is to learn the structure of the original dataset. In order to do so, we rely
128 on an approach inspired from representational learning, where we learn a *representation* of the metaweb
129 (in the form of the latent subspaces), rather than a list of interactions (species *a* eats *b*). This approach is
130 conceptually different from other metaweb-scale predictions (e.g. Albouy et al., 2019), in that the metaweb
131 representation is easily transferable. Specifically, we use Random Dot Product Graph model (hereafter
132 RDPG; S. J. Young & Scheinerman, 2007) to create a number of latent variables that can be combined into
133 an approximation of the network adjacency matrix. RDPG is known to capture the evolutionary backbone
134 of food webs (Dalla Riva & Stouffer, 2016), resulting in strong phylogenetic signal in RDPG results; in
135 other words, the latent variables of an RDPG can be mapped onto a phylogenetic tree, and
136 phylogenetically similar predators should share phylogenetically similar preys. In addition, recent
137 advances show that the latent variables produced this way can be used to predict *de novo* interactions.
138 Interestingly, the latent variables do not need to be produced by decomposing the network itself; in a
139 recent contribution, Runghen et al. (2021) showed that deep artificial neural networks are able to
140 reconstruct the left and right subspaces of an RDPG, in order to predict human movement networks from
141 individual/location metadata and opens up the possibility of using additional metadata as predictors.
142 The latent variables are created by performing a truncated Singular Value Decomposition (t-SVD; Halko et

143 al., 2011) on the adjacency matrix. SVD is an appropriate embedding of ecological networks, which has
144 recently been shown to both capture their complex, emerging properties (Strydom, Dalla Riva, et al., 2021)
145 and to allow highly accurate prediction of the interactions within a single network (Poisot, Ouellet, et al.,
146 2021). Under SVD, an adjacency matrix \mathbf{A} (where $\mathbf{A}_{m,n} \in \mathbb{B}$ where 1 indicates predation and 0 an absence
147 thereof) is decomposed into three components resulting in $\mathbf{A} = \mathbf{U}\Sigma\mathbf{V}'$. Here, Σ is a $m \times n$ diagonal matrix
148 and contains only singular (σ) values along its diagonal, \mathbf{U} is a $m \times m$ unitary matrix, and \mathbf{V}' a $n \times n$
149 unitary matrix. Truncating the SVD removes additional noise in the dataset by omitting non-zero and/or
150 smaller σ values from Σ using the rank of the matrix. Under a t-SVD $\mathbf{A}_{m,n}$ is decomposed so that Σ is a
151 square $r \times r$ diagonal matrix (with $1 \leq r \leq r_{full}$ where r_{full} is the full rank of \mathbf{A} and r the rank at which we
152 truncate the matrix) containing only non-zero σ values. Additionally, \mathbf{U} is now a $m \times r$ semi-unitary
153 matrix and \mathbf{V}' a $n \times r$ semi-unitary matrix.

154 The specific rank at which the SVD ought to be truncated is a difficult question. The purpose of SVD is to
155 remove the noise (expressed at high dimensions) and to focus on the signal, (expressed at low dimensions).
156 In datasets with a clear signal/noise demarcation, a scree plot of Σ can show a sharp drop at the rank where
157 noise starts (Zhu & Ghodsi, 2006). Because the European metaweb is almost entirely known, the amount
158 of noise (uncertainty) is low; this is reflected in fig. 2 (left), where the scree plot shows no important drop,
159 and in fig. 2 (right) where the proportion of variance explained increases smoothly at higher dimensions.
160 For this reason, we default back to a threshold that explains 60% of the variance in the underlying data,
161 corresponding to 12 dimensions - *i.e.* a tradeoff between accuracy and a reduced number of features.
162 An RDPG estimates the probability of observing interactions between nodes (species) as a function of the
163 nodes' latent variables, and is a way to turn an SVD (which decompose one matrix into three) into two
164 matrices that can be multiplied to provide an approximation of the network. The latent variables used for
165 the RDPG, called the left and right subspaces, are defined as $\mathcal{L} = \mathbf{U}\sqrt{\Sigma}$, and $\mathcal{R} = \sqrt{\Sigma}\mathbf{V}'$ – using the full
166 rank of \mathbf{A} , $\mathcal{L}\mathcal{R} = \mathbf{A}$, and using any smaller rank results in $\mathcal{L}\mathcal{R} \approx \mathbf{A}$. Using a rank of 1 for the t-SVD
167 provides a first-order approximation of the network. One advantage of using an RDPG rather than an SVD
168 is that the number of components to estimate decreases; notably, one does not have to estimate the
169 singular values of the SVD. Furthermore, the two subspaces can be directly multiplied to yield a network.

170 [Figure 2 about here.]
171 Because RDPG relies on matrix multiplication, the higher dimensions essentially serve to make specific

interactions converge towards 0 or 1; therefore, for reasonably low ranks, there is no guarantee that the values in the reconstructed network will be within the unit range. In order to determine what constitutes an appropriate threshold for probability, we performed the RDPG approach on the European metaweb, and evaluated the probability threshold by treating this as a binary classification problem, specifically assuming that both 0 and 1 in the European metaweb are all true. Given the methodological details given in Maiorano et al. (2020a) and O'Connor et al. (2020), this seems like a reasonable assumption, although one that does not hold for all metawebs. We used the thresholding approach presented in Poisot, Ouellet, et al. (2021), and picked a cutoff that maximized Youden's J statistic (a measure of the informedness (trust) of predictions; Youden (1950)); the resulting cutoff was 0.22, and gave an accuracy above 0.99. In Supp. Mat. 1, we provide several lines of evidence that using the entire network to estimate the threshold does not lead to overfitting; that using a subset of species would yield the same threshold; that decreasing the quality of the original data by adding or removing interactions would minimally affect the predictive accuracy of RDPG applied to the European metaweb; and that the networks reconstructed from artificially modified data are reconstructed with the correct ecological properties.

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

Steps 2 and 3: Transfer learning through phylogenetic relatedness

In order to transfer the knowledge from the European metaweb to the Canadian species pool, we performed ancestral character estimation using a Brownian motion model, which is a conservative approach in the absence of strong hypotheses about the nature of phylogenetic signal in the network decomposition (Litsios & Salamin, 2012). This uses the estimated feature vectors for the European mammals to create a state reconstruction for all species (conceptually something akin to a trait-based mammalian phylogeny using latent generality and vulnerability traits) and allows us to impute the missing (latent) trait data for the Canadian species that are not already in the European network; as we are focused on predicting contemporary interactions, we only retained the values for the tips of the tree. We assumed that all traits (*i.e.* the feature vectors for the left and right subspaces) were independent, which is a reasonable assumption as every trait/dimension added to the t-SVD has an *additive* effect to the one before it. Note that the Upham et al. (2019) tree itself has some uncertainty associated to inner nodes of

201 the phylogeny. In this case study, we have decided to not propagate this uncertainty, as it would
202 complexify the process. The Brownian motion algorithm returns the *average* value of the trait, and its
203 upper and lower bounds. Because we do not estimate other parameters of the traits' distributions, we
204 considered that every species trait is represented as a uniform distribution between these bounds. The
205 choice of the uniform distribution was made because the algorithm returns a minimum and maximum
206 point estimate for the value, and given this information, the uniform distribution is the one with
207 maximum entropy. Had all mean parameters estimates been positive, the exponential distribution would
208 have been an alternative, but this is not the case for the subspaces of an RDPG. In order to examine the
209 consequences of the choice of distribution, we estimated the variance per latent variable per node to use a
210 Normal distribution; as we show in Supp. Mat. 2, this decision results in dramatically over-estimating the
211 number and probability of interactions, and therefore we keep the discussions in the main text to the
212 uniform case. The inferred left and right subspaces for the Canadian species pool ($\hat{\mathcal{L}}$ and $\hat{\mathcal{R}}$) have entries
213 that are distributions, representing the range of values for a given species at a given dimension. These
214 objects represent the transferred knowledge, which we can use for prediction of the Canadian metaweb.

215 **Step 4: Probabilistic prediction of the destination network**

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

220 [Figure 3 about here.]

221 Specifically, we have adopted the following approach. For every entry in $\hat{\mathcal{L}}$ and $\hat{\mathcal{R}}$, we draw a value from
222 its distribution. This results in one instance of the possible left ($\hat{\ell}$) and right (\hat{r}) subspaces for the
223 Canadian metaweb. These can be multiplied, to produce one matrix of real values. Because the entries in
224 $\hat{\ell}$ and \hat{r} are in the same space where \mathcal{L} and \mathcal{R} were originally predicted, it follows that the threshold ρ
225 estimated for the European metaweb also applies. We use this information to produce one random
226 Canadian metaweb, $N = \hat{\mathcal{L}}\hat{\mathcal{R}}' \geq \rho$. As we can see in (fig. 3), the European and Canadian metawebs are
227 structurally similar (as would be expected given the biogeographic similarities) and the two (left and right)
228 subspaces are distinct *i.e.* capturing predation (generality) and prey (vulnerability) latent traits.

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

235 It must be noted that despite bringing in a large amount of information from the European species pool
236 and interactions, the Canadian metaweb has distinct structural properties. Following an approach similar
237 to Vermaat et al. (2009), we show in Supp. Mat. 3 that not only can we observe differences in a
238 multivariate space between the European and Canadian metaweb, we can also observe differences in the
239 same space between random subgraphs from these networks. These results line up with the studies
240 spatializing metawebs that have been discussed in the introduction: changes in the species pool are
241 driving local structural changes in the networks.

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

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

246 [Figure 4 about here.]

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

252 Second, we looked for all species in the Canadian pool known to the Global Biotic Interactions (GLoBI)
253 database (Poelen et al., 2014), and extracted their known interactions. Because GLoBI aggregates observed
254 interactions, it is not a *networks* data source, and therefore the only information we can reliably extract
255 from it is that a species pair *was reported to interact at least once*. This last statement should yet be taken

256 with caution, as some sources in GLoBI (e.g. Thessen & Parr, 2014) are produced through text analysis,
257 and therefore may not document direct evidence of the interaction. Nevertheless, should the predictive
258 model work, we would expect that a majority of interactions known to GLoBI would also be predicted. We
259 retrieved 366 interactions between mammals from the Canadian species pool from GLoBI, 33 of which
260 were not predicted by the model; this results in a success rate of 91%. After performing this check, we set
261 the probability of all interactions known to GLoBI to 1.

262 Finally, we downloaded the data from Strong & Leroux (2014), who mined various literature sources to
263 identify trophic interactions in Newfoundland. This dataset documented 25 interactions between
264 mammals, only two of which were not part of our (Canada-level) predictions, resulting in a success rate of
265 92%. These two interactions were added to our predicted metaweb with a probability of 1. A table listing
266 all interactions in the predicted Canadian metaweb can be found in the supplementary material.

267 [Figure 5 about here.]

268 Because the confidence intervals on the inferred trait space are probably over-estimates, we decided to
269 apply a thresholding step to the interactions after the data inflation (see fig. 5 showing the effect of varying
270 the cutoff on $P(i \rightarrow j)$). Cirtwill & Hambäck (2021) proposed a number of strategies to threshold
271 probabilistic networks. Their methods assume the underlying data to be tag-based sequencing, which
272 represents interactions as co-occurrences of predator and prey within the same tags; this is conceptually
273 identical to our Bernoulli-trial based reconstruction of a probabilistic network. We performed a full
274 analysis of the effect of various cutoffs, and as they either resulted in removing too few interactions, or
275 removing enough interactions that species started to be disconnected from the network, we set this
276 threshold for a probability equivalent to 0 to the largest possible value that still allowed all species to have
277 at least one interaction with a non-zero probability. The need for this slight deviation from the Cirtwill &
278 Hambäck (2021) method highlights the need for additional development on network thresholding.

279 Results and discussion of the case study

280 [Figure 6 about here.]

281 The t-SVD embedding is able to learn relevant ecological features for the network. fig. 6 shows that the
282 first rank correlates linearly with generality and vulnerability (Schoener, 1989), *i.e.* the number of preys

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

296 Discussion

One important aspect in which Europe and Canada differ (despite their comparable bioclimatic conditions) is the degree of the legacy of human impacts, which have been much longer in Europe. Nenzén et al. (2014) showed that even at small scales (the Iberian peninsula), mammal food webs retain the signal of both past climate change and human activity, even when this human activity was orders of magnitude less important than it is now. Similarly, Yeakel et al. (2014) showed that changes in human occupation over several centuries can lead to food web collapse. Megafauna in particular seems to be very sensitive to human arrival (Pires et al., 2015). In short, there is well-substantiated support for the idea that human footprint affects more than the risk of species extinction (Marco et al., 2018), and can lead to changes in interaction structure. Yet, owing to the inherent plasticity of interactions, there have been documented instances of food webs undergoing rapid collapse/recovery cycles over short periods of time (Pedersen et al., 2017). The embedding of a network, in a sense, embeds its macro-evolutionary history, especially as RDPG captures ecological signal (Dalla Riva & Stouffer, 2016); at this point, it is important to recall that a metaweb is intended as a catalogue of all potential interactions, which should then be filtered (Morales-Castilla et al., 2015). In practice (and in this instance) the reconstructed metaweb will predict

311 interactions that are plausible based on the species' evolutionary history, however some interactions
312 would/would not be realized due to human impact.

313 Dallas et al. (2017) suggested that most links in ecological networks may be cryptic, *i.e.* uncommon or
314 otherwise hard to observe. This argument essentially echoes Jordano (2016b): the sampling of ecological
315 interactions is difficult because it requires first the joint observation of two species, and then the
316 observation of their interaction. In addition, it is generally expected that weak or rare links would be more
317 common in networks (Csermely, 2004), compared to strong, persistent links; this is notably the case in
318 food chains, wherein many weaker links are key to the stability of a system (Neutel et al., 2002). In the
319 light of these observations, the results in fig. 4 are not particularly surprising: we expect to see a surge in
320 these low-probability interactions under a model that has a good predictive accuracy. Because the
321 predictions we generate are by design probabilistic, then one can weigh these rare links appropriately. In a
322 sense, that most ecological interactions are elusive can call for a slightly different approach to sampling:
323 once the common interactions are documented, the effort required in documenting each rare interaction
324 may increase exponentially. Recent proposals suggest that machine learning algorithms, in these
325 situations, can act as data generators (Hoffmann et al., 2019): in this perspective, high quality
326 observational data can be supplemented with synthetic data coming from predictive models, which
327 increases the volume of information available for inference. Indeed, Strydom, Catchen, et al. (2021)
328 suggested that knowing the metaweb may render the prediction of local networks easier, because it fixes
329 an "upper bound" on which interactions can exist; indeed, with a probabilistic metaweb, we can consider
330 that the metaweb represents an aggregation of informative priors on the interactions.

331 Related to the last point, Cirtwill et al. (2019) showed that network inference techniques based on
332 Bayesian approaches would perform far better in the presence of an interaction-level informative prior;
333 the desirable properties of such a prior would be that it is expressed as a probability, preferably
334 representing a Bernoulli event, the value of which would be representative of relevant biological processes
335 (probability of predation in this case). We argue that the probability returned at the very last step of our
336 framework may serve as this informative prior; indeed, the output of our analysis can be used in
337 subsequent steps, also possibly involving expert elicitation to validate some of the most strongly
338 recommended interactions. One important *caveat* to keep in mind when working with interaction
339 inference is that interactions can never really be true negatives (in the current state of our methodological
340 framework and data collection limitations); this renders the task of validating a model through the usual

application of binary classification statistics very difficult (although see Strydom, Catchen, et al., 2021 for a discussion of alternative suggestions). The other way through which our framework can be improved is by substituting the predictors that are used for transfer. For example, in the presence of information on species traits that are known to be predictive of species interactions, one might want to rely on functional rather than phylogenetic distances – in food webs, body size (and allometrically related variables) has been established as such a variable (Brose et al., 2006); the identification of relevant functional traits is facilitated by recent methodological developments (Rosado et al., 2013). It should be noted that Xing & Fayle (2021) highlight phylogenetic relatedness as one of the core components of network comparison at the global scale. In this case study, we have embedded the original metaweb using t-SVD, because it lends itself to an RDPG reconstruction, which is known to capture the consequences of evolutionary processes (Dalla Riva & Stouffer, 2016); this being said, there are other ways to embed graphs (Arsov & Mirceva, 2019; Cai et al., 2017; Cao et al., 2019), which can be used as alternatives.

As Herbert (1965) rightfully pointed out, “[y]ou can’t draw neat lines around planet-wide problems”; in this regard, our approach (and indeed, any inference of a metaweb at large scales) must contend with several interesting and interwoven families of problems. The first is the limit of the metaweb to embed and transfer. If the initial metaweb is too narrow in scope, notably from a taxonomic point of view, the chances of finding another area with enough related species to make a reliable inference decreases; this would likely be indicated by large confidence intervals during ancestral character estimation, but the lack of well documented metawebs is currently preventing the development of more concrete guidelines. The question of phylogenetic relatedness and dispersal is notably true if the metaweb is assembled in an area with mostly endemic species, and as with every predictive algorithm, there is room for the application of our best ecological judgement. Conversely, the metaweb should be reliably filled, which assumes that the S^2 interactions in a pool of S species have been examined, either through literature surveys or expert elicitation. Supp. Mat. 1 provides some guidance as to the type of sampling effort that should be prioritized. While RDPG was able to maintain very high predictive power when interactions were missing, the addition of false positive interactions was immediately detected; this suggests that it may be appropriate to err on the side of “too many” interactions when constructing the initial metaweb to be transferred. The second series of problems are related to determining which area should be used to infer the new metaweb in, as this determines the species pool that must be used. In our application, we focused on the mammals of Canada. The upside of this approach is that information at the country level is likely

371 to be required by policy makers and stakeholders for their biodiversity assessment, as each country tends
372 to set goals at the national level (Buxton et al., 2021) for which quantitative instruments are designed
373 (Turak et al., 2017), with specific strategies often enacted at smaller scales (Ray et al., 2021). And yet, we
374 do not really have a satisfying answer to the question of “where does a food web stop”; the current most
375 satisfying solutions involve examining the spatial consistency of network area relationships (Fortin et al.,
376 2021; see e.g. Galiana et al., 2018, 2019, 2021), which is of course impossible in the absence of enough
377 information about the network itself. This suggests that an *a posteriori* refinement of the results may be
378 required, based on a downscaling of the metaweb. The final family of problems relates less to the
379 availability of data or quantitative tools, and more to the praxis of spatial ecology. Operating under the
380 context of national divisions, in large parts of the world, reflects nothing more than the legacy of settler
381 colonialism. Indeed, the use of ecological data is not an apolitical act (Nost & Goldstein, 2021), as data
382 infrastructures tend to be designed to answer questions within national boundaries, and their use both
383 draws upon and reinforces territorial statecraft; as per Machen & Nost (2021), this is particularly true
384 when the output of “algorithmic thinking” (e.g. relying on machine learning to generate knowledge) can
385 be re-used for governance (e.g. enacting conservation decisions at the national scale). We therefore
386 recognize that methods such as we propose operate under the framework that contributed to the ongoing
387 biodiversity crisis (Adam, 2014), reinforced environmental injustice (Choudry, 2013; Domínguez &
388 Luoma, 2020), and on Turtle Island especially, should be replaced by Indigenous principles of land
389 management (Eichhorn et al., 2019; No’kmaq et al., 2021). As we see AI/ML being increasingly mobilized
390 to generate knowledge that is lacking for conservation decisions (e.g. Lamba et al., 2019; Mosebo
391 Fernandes et al., 2020), our discussion of these tools need to go beyond the technical, and into the
392 governance consequences they can have.

393 **Acknowledgements:** We acknowledge that this study was conducted on land within the traditional
394 unceded territory of the Saint Lawrence Iroquoian, Anishinabewaki, Mohawk, Huron-Wendat, and
395 Omàmiwininiwak nations. TP, TS, DC, and LP received funding from the Canadian Institute for Ecology &
396 Evolution. FB is funded by the Institute for Data Valorization (IVADO). TS, SB, and TP are funded by a
397 donation from the Courtois Foundation. CB was awarded a Mitacs Elevate Fellowship no. IT12391, in
398 partnership with fRI Research, and also acknowledges funding from Alberta Innovates and the Forest
399 Resources Improvement Association of Alberta. M-JF acknowledges funding from NSERC Discovery
400 Grant and NSERC CRC. RR is funded by New Zealand’s Biological Heritage Ngā Koiora Tuku Iho

401 National Science Challenge, administered by New Zealand Ministry of Business, Innovation, and
402 Employment. BM is funded by the NSERC Alexander Graham Bell Canada Graduate Scholarship and the
403 FRQNT master's scholarship. LP acknowledges funding from NSERC Discovery Grant (NSERC
404 RGPIN-2019-05771). TP acknowledges financial support from NSERC through the Discovery Grants and
405 Discovery Accelerator Supplement programs.

406 **Conflict of interest:** The authors have no conflict interests to disclose

407 **Authors' contributions:** TS, SB, and TP designed the study and performed the analysis; GVDR, MF, and
408 RR provided additional feedback on the analyses. DC, BM, and FB helped with data collection. All
409 authors contributed to writing and editing the manuscript.

410 **Data availability:** All code and data used in this manuscript is publicly available and archived on OSF
411 <https://osf.io/2zwqm/> and is currently referenced in the manuscript.

412 References

- 413 Adam, R. (2014). *Elephant treaties: The Colonial legacy of the biodiversity crisis*. UPNE.
- 414 Albouy, C., Archambault, P., Appeltans, W., Araújo, M. B., Beauchesne, D., Cazelles, K., Cirtwill, A. R.,
415 Fortin, M.-J., Galiana, N., Leroux, S. J., Pellissier, L., Poisot, T., Stouffer, D. B., Wood, S. A., & Gravel, D.
416 (2019). The marine fish food web is globally connected. *Nature Ecology & Evolution*, 3(8, 8),
417 1153–1161. <https://doi.org/10.1038/s41559-019-0950-y>
- 418 Arsov, N., & Mirceva, G. (2019, November 26). *Network Embedding: An Overview*.
419 <http://arxiv.org/abs/1911.11726>
- 420 Banville, F., Vissault, S., & Poisot, T. (2021). Mangal.jl and EcologicalNetworks.jl: Two complementary
421 packages for analyzing ecological networks in Julia. *Journal of Open Source Software*, 6(61), 2721.
422 <https://doi.org/10.21105/joss.02721>
- 423 Bascompte, J., Jordano, P., Melian, C. J., & Olesen, J. M. (2003). The nested assembly of plant-animal
424 mutualistic networks. *Proceedings of the National Academy of Sciences*, 100(16), 9383–9387.
425 <https://doi.org/10.1073/pnas.1633576100>
- 426 Beckerman, A. P., Petchey, O. L., & Warren, P. H. (2006). Foraging biology predicts food web complexity.

- 427 *Proceedings of the National Academy of Sciences*, 103(37), 13745–13749.
- 428 <https://doi.org/10.1073/pnas.0603039103>
- 429 Bezanson, J., Edelman, A., Karpinski, S., & Shah, V. (2017). Julia: A Fresh Approach to Numerical
430 Computing. *SIAM Review*, 59(1), 65–98. <https://doi.org/10.1137/141000671>
- 431 Boeckaerts, D., Stock, M., Criel, B., Gerstmans, H., De Baets, B., & Briers, Y. (2021). Predicting
432 bacteriophage hosts based on sequences of annotated receptor-binding proteins. *Scientific Reports*,
433 11(1, 1), 1467. <https://doi.org/10.1038/s41598-021-81063-4>
- 434 Braga, M. P., Janz, N., Nylin, S., Ronquist, F., & Landis, M. J. (2021). Phylogenetic reconstruction of
435 ancestral ecological networks through time for pierid butterflies and their host plants. *Ecology Letters*,
436 n/a(n/a). <https://doi.org/10.1111/ele.13842>
- 437 Brose, U., Jonsson, T., Berlow, E. L., Warren, P., Banasek-Richter, C., Bersier, L.-F., Blanchard, J. L., Brey,
438 T., Carpenter, S. R., Blandenier, M.-F. C., Cushing, L., Dawah, H. A., Dell, T., Edwards, F.,
439 Harper-Smith, S., Jacob, U., Ledger, M. E., Martinez, N. D., Memmott, J., ... Cohen, J. E. (2006).
440 ConsumerResource Body-Size Relationships in Natural Food Webs. *Ecology*, 87(10), 2411–2417.
441 [https://doi.org/10.1890/0012-9658\(2006\)87%5B2411:CBRINF%5D2.0.CO;2](https://doi.org/10.1890/0012-9658(2006)87%5B2411:CBRINF%5D2.0.CO;2)
- 442 Buxton, R. T., Bennett, J. R., Reid, A. J., Shulman, C., Cooke, S. J., Francis, C. M., Nyboer, E. A., Pritchard,
443 G., Binley, A. D., Avery-Gomm, S., Ban, N. C., Beazley, K. F., Bennett, E., Blight, L. K., Bortolotti, L. E.,
444 Camfield, A. F., Gadallah, F., Jacob, A. L., Naujokaitis-Lewis, I., ... Smith, P. A. (2021). Key
445 information needs to move from knowledge to action for biodiversity conservation in Canada.
446 *Biological Conservation*, 256, 108983. <https://doi.org/10.1016/j.biocon.2021.108983>
- 447 Cai, H., Zheng, V. W., & Chang, K. C.-C. (2017). *A Comprehensive Survey of Graph Embedding: Problems,*
448 *Techniques and Applications*. <http://arxiv.org/abs/1709.07604>
- 449 Cameron, E. K., Sundqvist, M. K., Keith, S. A., CaraDonna, P. J., Mousing, E. A., Nilsson, K. A., Metcalfe,
450 D. B., & Classen, A. T. (2019). Uneven global distribution of food web studies under climate change.
451 *Ecosphere*, 10(3), e02645. <https://doi.org/10.1002/ecs2.2645>
- 452 Cao, R.-M., Liu, S.-Y., & Xu, X.-K. (2019). Network embedding for link prediction: The pitfall and
453 improvement. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 29(10), 103102.
454 <https://doi.org/10.1063/1.5120724>

- 455 Cavender-Bares, J., Kozak, K. H., Fine, P. V. A., & Kembel, S. W. (2009). The merging of community
456 ecology and phylogenetic biology. *Ecology Letters*, 12(7), 693–715.
457 <https://doi.org/10.1111/j.1461-0248.2009.01314.x>
- 458 Choudry, A. (2013). Saving biodiversity, for whom and for what? Conservation NGOs, complicity,
459 colonialism and conquest in an era of capitalist globalization. In *NGOization: Complicity,
460 contradictions and prospects* (pp. 24–44). Bloomsbury Publishing.
- 461 Cirtwill, A. R., Ekl, A., Roslin, T., Wootton, K., & Gravel, D. (2019). A quantitative framework for
462 investigating the reliability of empirical network construction. *Methods in Ecology and Evolution*, 0.
463 <https://doi.org/10.1111/2041-210X.13180>
- 464 Cirtwill, A. R., & Hambäck, P. (2021). Building food networks from molecular data: Bayesian or
465 fixed-number thresholds for including links. *Basic and Applied Ecology*, 50, 67–76.
466 <https://doi.org/10.1016/j.baae.2020.11.007>
- 467 Csermely, P. (2004). Strong links are important, but weak links stabilize them. *Trends in Biochemical
468 Sciences*, 29(7), 331–334. <https://doi.org/10.1016/j.tibs.2004.05.004>
- 469 Dalla Riva, G. V., & Stouffer, D. B. (2016). Exploring the evolutionary signature of food webs' backbones
470 using functional traits. *Oikos*, 125(4), 446–456. <https://doi.org/10.1111/oik.02305>
- 471 Dallas, T., Park, A. W., & Drake, J. M. (2017). Predicting cryptic links in host-parasite networks. *PLOS
472 Computational Biology*, 13(5), e1005557. <https://doi.org/10.1371/journal.pcbi.1005557>
- 473 Dansereau, G., & Poisot, T. (2021). SimpleSDMLayers.jl and GBIF.jl: A Framework for Species
474 Distribution Modeling in Julia. *Journal of Open Source Software*, 6(57), 2872.
475 <https://doi.org/10.21105/joss.02872>
- 476 Domínguez, L., & Luoma, C. (2020). Decolonising Conservation Policy: How Colonial Land and
477 Conservation Ideologies Persist and Perpetuate Indigenous Injustices at the Expense of the
478 Environment. *Land*, 9(3, 3), 65. <https://doi.org/10.3390/land9030065>
- 479 Dormann, C. F., Gruber, B., Winter, M., & Herrmann, D. (2010). Evolution of climate niches in European
480 mammals? *Biology Letters*, 6(2), 229–232. <https://doi.org/10.1098/rsbl.2009.0688>
- 481 Dunne, J. A. (2006). The Network Structure of Food Webs. In J. A. Dunne & M. Pascual (Eds.), *Ecological
482 networks: Linking structure and dynamics* (pp. 27–86). Oxford University Press.

- 483 Eichhorn, M. P., Baker, K., & Griffiths, M. (2019). Steps towards decolonising biogeography. *Frontiers of*
484 *Biogeography*, 12(1), 1–7. <https://doi.org/10.21425/F5FBG44795>
- 485 Eklöf, A., & Stouffer, D. B. (2016). The phylogenetic component of food web structure and intervality.
486 *Theoretical Ecology*, 9(1), 107–115. <https://doi.org/10.1007/s12080-015-0273-9>
- 487 Fortin, M.-J., Dale, M. R. T., & Brimacombe, C. (2021). Network ecology in dynamic landscapes.
488 *Proceedings of the Royal Society B: Biological Sciences*, 288(1949), rspb.2020.1889, 20201889.
489 <https://doi.org/10.1098/rspb.2020.1889>
- 490 Galiana, N., Barros, C., Braga, J., Ficetola, G. F., Maiorano, L., Thuiller, W., Montoya, J. M., & Lurgi, M.
491 (2021). The spatial scaling of food web structure across European biogeographical regions. *Ecography*,
492 n/a(n/a). <https://doi.org/10.1111/ecog.05229>
- 493 Galiana, N., Hawkins, B. A., & Montoya, J. M. (2019). The geographical variation of network structure is
494 scale dependent: Understanding the biotic specialization of hostparasitoid networks. *Ecography*, 42(6),
495 1175–1187. <https://doi.org/10.1111/ecog.03684>
- 496 Galiana, N., Lurgi, M., Claramunt-López, B., Fortin, M.-J., Leroux, S., Cazelles, K., Gravel, D., & Montoya,
497 J. M. (2018). The spatial scaling of species interaction networks. *Nature Ecology & Evolution*, 2(5),
498 782–790. <https://doi.org/10.1038/s41559-018-0517-3>
- 499 Garland, T., JR., Midford, P. E., & Ives, A. R. (1999). An Introduction to Phylogenetically Based Statistical
500 Methods, with a New Method for Confidence Intervals on Ancestral Values1. *American Zoologist*,
501 39(2), 374–388. <https://doi.org/10.1093/icb/39.2.374>
- 502 Garlaschelli, D., Hollander, F. den, & Roccaverde, A. (2018). Covariance structure behind breaking of
503 ensemble equivalence in random graphs. *Journal of Statistical Physics*, 173(3-4), 644–662.
504 <https://doi.org/10.1007/s10955-018-2114-x>
- 505 GBIF Secretariat. (2021). *GBIF Backbone Taxonomy*. <https://doi.org/10.15468/39omei>
- 506 Gerhold, P., Cahill, J. F., Winter, M., Bartish, I. V., & Prinzing, A. (2015). Phylogenetic patterns are not
507 proxies of community assembly mechanisms (they are far better). *Functional Ecology*, 29(5), 600–614.
508 <https://doi.org/10.1111/1365-2435.12425>
- 509 Gravel, D., Baiser, B., Dunne, J. A., Kopalke, J.-P., Martinez, N. D., Nyman, T., Poisot, T., Stouffer, D. B.,
510 Tylianakis, J. M., Wood, S. A., & Roslin, T. (2018). Bringing Elton and Grinnell together: A quantitative

- 511 framework to represent the biogeography of ecological interaction networks. *Ecography*, 0(0).
- 512 <https://doi.org/10.1111/ecog.04006>
- 513 Grünig, M., Mazzi, D., Calanca, P., Karger, D. N., & Pellissier, L. (2020). Crop and forest pest metawebs
514 shift towards increased linkage and suitability overlap under climate change. *Communications Biology*,
515 3(1, 1), 1–10. <https://doi.org/10.1038/s42003-020-0962-9>
- 516 Halevy, A., Norvig, P., & Pereira, F. (2009). The Unreasonable Effectiveness of Data. *IEEE Intelligent
517 Systems*, 24(2), 8–12. <https://doi.org/10.1109/MIS.2009.36>
- 518 Halko, N., Martinsson, P. G., & Tropp, J. A. (2011). Finding Structure with Randomness: Probabilistic
519 Algorithms for Constructing Approximate Matrix Decompositions. *SIAM Review*, 53(2), 217–288.
520 <https://doi.org/10.1137/090771806>
- 521 Herbert, F. (1965). *Dune* (1st ed.). Chilton Book Company.
- 522 Hoffmann, J., Bar-Sinai, Y., Lee, L. M., Andrejevic, J., Mishra, S., Rubinstein, S. M., & Rycroft, C. H. (2019).
523 Machine learning in a data-limited regime: Augmenting experiments with synthetic data uncovers
524 order in crumpled sheets. *Science Advances*, 5(4), eaau6792.
525 <https://doi.org/10.1126/sciadv.aau6792>
- 526 Holm, E. A. (2019). In defense of the black box. *Science*, 364(6435), 26–27.
527 <https://doi.org/10.1126/science.aax0162>
- 528 Hortal, J., de Bello, F., Diniz-Filho, J. A. F., Lewinsohn, T. M., Lobo, J. M., & Ladle, R. J. (2015). Seven
529 Shortfalls that Beset Large-Scale Knowledge of Biodiversity. *Annual Review of Ecology, Evolution, and
530 Systematics*, 46(1), 523–549. <https://doi.org/10.1146/annurev-ecolsys-112414-054400>
- 531 Hutchinson, M. C., Cagua, E. F., & Stouffer, D. B. (2017). Cophylogenetic signal is detectable in pollination
532 interactions across ecological scales. *Ecology*, n/a–n/a. <https://doi.org/10.1002/ecy.1955>
- 533 Jordano, P. (2016a). Chasing Ecological Interactions. *PLOS Biol*, 14(9), e1002559.
534 <https://doi.org/10.1371/journal.pbio.1002559>
- 535 Jordano, P. (2016b). Sampling networks of ecological interactions. *Functional Ecology*, 30(12), 1883–1893.
536 <https://doi.org/10.1111/1365-2435.12763>
- 537 Kawatsu, K., Ushio, M., van Veen, F. J. F., & Kondoh, M. (2021). Are networks of trophic interactions
538 sufficient for understanding the dynamics of multi-trophic communities? Analysis of a tri-trophic

- 539 insect food-web time-series. *Ecology Letters*, 24(3), 543–552. <https://doi.org/10.1111/ele.13672>
- 540 Kéfi, S., Berlow, E. L., Wieters, E. A., Navarrete, S. A., Petchey, O. L., Wood, S. A., Boit, A., Joppa, L. N.,
541 Lafferty, K. D., Williams, R. J., Martinez, N. D., Menge, B. A., Blanchette, C. A., Iles, A. C., & Brose, U.
542 (2012). More than a meal... integrating non-feeding interactions into food webs: More than a meal
543 *Ecology Letters*, 15(4), 291–300. <https://doi.org/10.1111/j.1461-0248.2011.01732.x>
- 544 Lamba, A., Cassey, P., Segaran, R. R., & Koh, L. P. (2019). Deep learning for environmental conservation.
545 *Current Biology*, 29(19), R977–R982. <https://doi.org/10.1016/j.cub.2019.08.016>
- 546 Litsios, G., & Salamin, N. (2012). Effects of Phylogenetic Signal on Ancestral State Reconstruction.
547 *Systematic Biology*, 61(3), 533–538. <https://doi.org/10.1093/sysbio/syr124>
- 548 Machen, R., & Nost, E. (2021). Thinking algorithmically: The making of hegemonic knowledge in climate
549 governance. *Transactions of the Institute of British Geographers*, 46(3), 555–569.
550 <https://doi.org/10.1111/tran.12441>
- 551 Maiorano, L., Montemaggioli, A., Ficetola, G. F., O'Connor, L., & Thuiller, W. (2020a). TETRA-EU 1.0: A
552 species-level trophic metaweb of European tetrapods. *Global Ecology and Biogeography*, 29(9),
553 1452–1457. <https://doi.org/10.1111/geb.13138>
- 554 Maiorano, L., Montemaggioli, A., Ficetola, G. F., O'Connor, L., & Thuiller, W. (2020b). *Data from:*
555 *Tetra-EU 1.0: A species-level trophic meta-web of European tetrapods* (Version 3, pp. 16596876 bytes)
556 [Data set]. Dryad. <https://doi.org/10.5061/DRYAD.JM63XSJ7B>
- 557 Marco, M. D., Venter, O., Possingham, H. P., & Watson, J. E. M. (2018). Changes in human footprint drive
558 changes in species extinction risk. *Nature Communications*, 9(1), 4621.
559 <https://doi.org/10.1038/s41467-018-07049-5>
- 560 McLeod, A., Leroux, S. J., Gravel, D., Chu, C., Cirtwill, A. R., Fortin, M.-J., Galiana, N., Poisot, T., & Wood,
561 S. A. (2021). Sampling and asymptotic network properties of spatial multi-trophic networks. *Oikos*,
562 n/a(n/a). <https://doi.org/10.1111/oik.08650>
- 563 Mora, B. B., Gravel, D., Gilarranz, L. J., Poisot, T., & Stouffer, D. B. (2018). Identifying a common backbone
564 of interactions underlying food webs from different ecosystems. *Nature Communications*, 9(1), 2603.
565 <https://doi.org/10.1038/s41467-018-05056-0>
- 566 Morales-Castilla, I., Matias, M. G., Gravel, D., & Araújo, M. B. (2015). Inferring biotic interactions from

- 567 proxies. *Trends in Ecology & Evolution*, 30(6), 347–356.
- 568 <https://doi.org/10.1016/j.tree.2015.03.014>
- 569 Mosebo Fernandes, A. C., Quintero Gonzalez, R., Lenihan-Clarke, M. A., Leslie Trotter, E. F., & Jokar
570 Arsanjani, J. (2020). Machine Learning for Conservation Planning in a Changing Climate.
571 *Sustainability*, 12(18, 18), 7657. <https://doi.org/10.3390/su12187657>
- 572 Mouquet, N., Devictor, V., Meynard, C. N., Munoz, F., Bersier, L.-F., Chave, J., Couteron, P., Dalecky, A.,
573 Fontaine, C., Gravel, D., Hardy, O. J., Jabot, F., Lavergne, S., Leibold, M., Mouillot, D., Münkemüller,
574 T., Pavoine, S., Prinzing, A., Rodrigues, A. S. L., ... Thuiller, W. (2012). Ecophylogenetics: Advances
575 and perspectives. *Biological Reviews*, 87(4), 769–785.
576 <https://doi.org/10.1111/j.1469-185X.2012.00224.x>
- 577 Nenzén, H. K., Montoya, D., & Varela, S. (2014). The Impact of 850,000 Years of Climate Changes on the
578 Structure and Dynamics of Mammal Food Webs. *PLOS ONE*, 9(9), e106651.
579 <https://doi.org/10.1371/journal.pone.0106651>
- 580 Neutel, A.-M., Heesterbeek, J. A. P., & de Ruiter, P. C. (2002). Stability in Real Food Webs: Weak Links in
581 Long Loops. *Science*, 296(5570), 1120–1123. <https://doi.org/10.1126/science.1068326>
- 582 No'kmaq, M., Marshall, A., Beazley, K. F., Hum, J., joudry, shalan, Papadopoulos, A., Pictou, S., Rabesca,
583 J., Young, L., & Zurba, M. (2021). “Awakening the sleeping giant”: Re-Indigenization principles for
584 transforming biodiversity conservation in Canada and beyond. *FACETS*, 6(1), 839–869.
- 585 Nost, E., & Goldstein, J. E. (2021). A political ecology of data. *Environment and Planning E: Nature and
586 Space*, 25148486211043503. <https://doi.org/10.1177/25148486211043503>
- 587 O'Connor, L. M. J., Pollock, L. J., Braga, J., Ficetola, G. F., Maiorano, L., Martinez-Almoyna, C.,
588 Montemaggioli, A., Ohlmann, M., & Thuiller, W. (2020). Unveiling the food webs of tetrapods across
589 Europe through the prism of the Eltonian niche. *Journal of Biogeography*, 47(1), 181–192.
590 <https://doi.org/10.1111/jbi.13773>
- 591 Pan, S. J., & Yang, Q. (2010). A Survey on Transfer Learning. *IEEE Transactions on Knowledge and Data
592 Engineering*, 22(10), 1345–1359. <https://doi.org/10.1109/TKDE.2009.191>
- 593 Park, J., & Newman, M. E. J. (2004). Statistical mechanics of networks. *Physical Review E*, 70(6), 066117.
594 <https://doi.org/10.1103/PhysRevE.70.066117>

- 595 Pedersen, E. J., Thompson, P. L., Ball, R. A., Fortin, M.-J., Gouhier, T. C., Link, H., Moritz, C., Nenzen, H.,
596 Stanley, R. R. E., Tararu, Z. E., Gonzalez, A., Guichard, F., & Pepin, P. (2017). Signatures of the
597 collapse and incipient recovery of an overexploited marine ecosystem. *Royal Society Open Science*, 4(7),
598 170215. <https://doi.org/10.1098/rsos.170215>
- 599 Perretti, C. T., Munch, S. B., & Sugihara, G. (2013). Model-free forecasting outperforms the correct
600 mechanistic model for simulated and experimental data. *Proceedings of the National Academy of
601 Sciences*, 110(13), 5253–5257. <https://doi.org/10.1073/pnas.1216076110>
- 602 Pires, M. M., Koch, P. L., Fariña, R. A., de Aguiar, M. A. M., dos Reis, S. F., & Guimarães, P. R. (2015).
603 Pleistocene megafaunal interaction networks became more vulnerable after human arrival.
604 *Proceedings of the Royal Society B: Biological Sciences*, 282(1814), 20151367.
605 <https://doi.org/10.1098/rspb.2015.1367>
- 606 Poelen, J. H., Simons, J. D., & Mungall, C. J. (2014). Global biotic interactions: An open infrastructure to
607 share and analyze species-interaction datasets. *Ecological Informatics*, 24, 148–159.
608 <https://doi.org/10.1016/j.ecoinf.2014.08.005>
- 609 Poisot, T., Belisle, Z., Hoebeke, L., Stock, M., & Szefer, P. (2019). EcologicalNetworks.jl - analysing
610 ecological networks. *Ecography*. <https://doi.org/10.1111/ecog.04310>
- 611 Poisot, T., Bergeron, G., Cazelles, K., Dallas, T., Gravel, D., MacDonald, A., Mercier, B., Violet, C., &
612 Vissault, S. (2021). Global knowledge gaps in species interaction networks data. *Journal of
613 Biogeography*, n/a(n/a). <https://doi.org/10.1111/jbi.14127>
- 614 Poisot, T., Cirtwill, A. R., Cazelles, K., Gravel, D., Fortin, M.-J., & Stouffer, D. B. (2016). The structure of
615 probabilistic networks. *Methods in Ecology and Evolution*, 7(3), 303–312.
616 <https://doi.org/10.1111/2041-210X.12468>
- 617 Poisot, T., Ouellet, M.-A., Mollentze, N., Farrell, M. J., Becker, D. J., Albery, G. F., Gibb, R. J., Seifert, S. N.,
618 & Carlson, C. J. (2021, May 31). *Imputing the mammalian virome with linear filtering and singular
619 value decomposition*. <http://arxiv.org/abs/2105.14973>
- 620 Poisot, T., & Stouffer, D. B. (2018). Interactions retain the co-phylogenetic matching that communities lost.
621 *Oikos*, 127(2), 230–238. <https://doi.org/10.1111/oik.03788>
- 622 Poisot, T., Stouffer, D. B., & Gravel, D. (2015). Beyond species: Why ecological interaction networks vary

- 623 through space and time. *Oikos*, 124(3), 243–251. <https://doi.org/10.1111/oik.01719>
- 624 Price, P. W. (2003). *Macroevolutionary theory on macroecological patterns*. Cambridge University Press.
- 625 Ray, J. C., Grimm, J., & Olive, A. (2021). The biodiversity crisis in Canada: Failures and challenges of
626 federal and sub-national strategic and legal frameworks. *FACETS*, 6, 1044–1068.
627 <https://doi.org/10.1139/facets-2020-0075>
- 628 Reeve, R., Leinster, T., Cobbold, C. A., Thompson, J., Brummitt, N., Mitchell, S. N., & Matthews, L. (2016,
629 December 8). *How to partition diversity*. <http://arxiv.org/abs/1404.6520>
- 630 Rosado, B. H. P., Dias, A., & de Mattos, E. (2013). Going Back to Basics: Importance of Ecophysiology
631 when Choosing Functional Traits for Studying Communities and Ecosystems. *Natureza &
632 Conservaç~ao Revista Brasileira de Conservaç~ao Da Natureza*, 11, 15–22.
633 <https://doi.org/10.4322/natcon.2013.002>
- 634 Runghen, R., Stouffer, D. B., & Dalla Riva, G. V. (2021). *Exploiting node metadata to predict interactions in
635 large networks using graph embedding and neural networks*.
636 <https://doi.org/10.1101/2021.06.10.447991>
- 637 Schoener, T. W. (1989). Food webs from the small to the large. *Ecology*, 70(6), 1559–1589.
- 638 Solís-Lemus, C., Bastide, P., & Ané, C. (2017). PhyloNetworks: A Package for Phylogenetic Networks.
639 *Molecular Biology and Evolution*, 34(12), 3292–3298. <https://doi.org/10.1093/molbev/msx235>
- 640 Stock, M. (2021). Pairwise learning for predicting pollination interactions based on traits and phylogeny.
641 *Ecological Modelling*, 14.
- 642 Stouffer, D. B., Sales-Pardo, M., Sirer, M. I., & Bascompte, J. (2012). Evolutionary Conservation of Species'
643 Roles in Food Webs. *Science*, 335(6075), 1489–1492. <https://doi.org/10.1126/science.1216556>
- 644 Strong, J. S., & Leroux, S. J. (2014). Impact of Non-Native Terrestrial Mammals on the Structure of the
645 Terrestrial Mammal Food Web of Newfoundland, Canada. *PLOS ONE*, 9(8), e106264.
646 <https://doi.org/10.1371/journal.pone.0106264>
- 647 Strydom, T., Catchen, M. D., Banville, F., Caron, D., Dansereau, G., Desjardins-Proulx, P., Forero-Muñoz,
648 N. R., Higino, G., Mercier, B., Gonzalez, A., Gravel, D., Pollock, L., & Poisot, T. (2021). A roadmap
649 towards predicting species interaction networks (across space and time). *Philosophical Transactions of*

- 650 *the Royal Society B: Biological Sciences*, 376(1837), 20210063.
651 <https://doi.org/10.1098/rstb.2021.0063>
- 652 Strydom, T., Dalla Riva, G. V., & Poisot, T. (2021). SVD Entropy Reveals the High Complexity of Ecological
653 Networks. *Frontiers in Ecology and Evolution*, 9. <https://doi.org/10.3389/fevo.2021.623141>
- 654 Thessen, A. E., & Parr, C. S. (2014). Knowledge extraction and semantic annotation of text from the
655 encyclopedia of life. *PloS One*, 9(3), e89550.
- 656 Torrey, L., & Shavlik, J. (2010). Transfer learning. In *Handbook of research on machine learning
657 applications and trends: Algorithms, methods, and techniques* (pp. 242–264). IGI global.
- 658 Trøjelsgaard, K., & Olesen, J. M. (2016). Ecological networks in motion: Micro- and macroscopic
659 variability across scales. *Functional Ecology*, 30(12), 1926–1935.
660 <https://doi.org/10.1111/1365-2435.12710>
- 661 Turak, E., Brazill-Boast, J., Cooney, T., Drielsma, M., DelaCruz, J., Dunkerley, G., Fernandez, M., Ferrier,
662 S., Gill, M., Jones, H., Koen, T., Leys, J., McGeoch, M., Mihoub, J.-B., Scanes, P., Schmeller, D., &
663 Williams, K. (2017). Using the essential biodiversity variables framework to measure biodiversity
664 change at national scale. *Biological Conservation*, 213, 264–271.
665 <https://doi.org/10.1016/j.biocon.2016.08.019>
- 666 Upham, N. S., Esselstyn, J. A., & Jetz, W. (2019). Inferring the mammal tree: Species-level sets of
667 phylogenies for questions in ecology, evolution, and conservation. *PLOS Biology*, 17(12), e3000494.
668 <https://doi.org/10.1371/journal.pbio.3000494>
- 669 van der Hoorn, P., Lippner, G., & Krioukov, D. (2018). Sparse Maximum-Entropy Random Graphs with a
670 Given Power-Law Degree Distribution. *Journal of Statistical Physics*, 173(3-4), 806–844.
671 <https://doi.org/10.1007/s10955-017-1887-7>
- 672 Vermaat, J. E., Dunne, J. A., & Gilbert, A. J. (2009). Major dimensions in food-web structure properties.
673 *Ecology*, 90(1), 278–282. <http://www.ncbi.nlm.nih.gov/pubmed/19294932>
- 674 Wood, S. A., Russell, R., Hanson, D., Williams, R. J., & Dunne, J. A. (2015). Effects of spatial scale of
675 sampling on food web structure. *Ecology and Evolution*, 5(17), 3769–3782.
676 <https://doi.org/10.1002/ece3.1640>
- 677 Xing, S., & Fayle, T. M. (2021). The rise of ecological network meta-analyses: Problems and prospects.

- 678 *Global Ecology and Conservation*, 30, e01805. <https://doi.org/10.1016/j.gecco.2021.e01805>
- 679 Yeakel, J. D., Pires, M. M., Rudolf, L., Dominy, N. J., Koch, P. L., Guimarães, P. R., & Gross, T. (2014).
680 Collapse of an ecological network in Ancient Egypt. *PNAS*, 111(40), 14472–14477.
681 <https://doi.org/10.1073/pnas.1408471111>
- 682 Youden, W. J. (1950). Index for rating diagnostic tests. *Cancer*, 3(1), 32–35.
683 [https://doi.org/10.1002/1097-0142\(1950\)3:1%3C32::AID-CNCR2820030106%3E3.0.CO;2-3](https://doi.org/10.1002/1097-0142(1950)3:1%3C32::AID-CNCR2820030106%3E3.0.CO;2-3)
- 684 Young, J.-G., Cantwell, G. T., & Newman, M. E. J. (2021). Bayesian inference of network structure from
685 unreliable data. *Journal of Complex Networks*, 8(6). <https://doi.org/10.1093/comnet/cnaa046>
- 686 Young, S. J., & Scheinerman, E. R. (2007). Random Dot Product Graph Models for Social Networks. In A.
687 Bonato & F. R. K. Chung (Eds.), *Algorithms and Models for the Web-Graph* (pp. 138–149). Springer.
688 https://doi.org/10.1007/978-3-540-77004-6_11
- 689 Zhu, M., & Ghodsi, A. (2006). Automatic dimensionality selection from the scree plot via the use of profile
690 likelihood. *Computational Statistics & Data Analysis*, 51(2), 918–930.
691 <https://doi.org/10.1016/j.csda.2005.09.010>

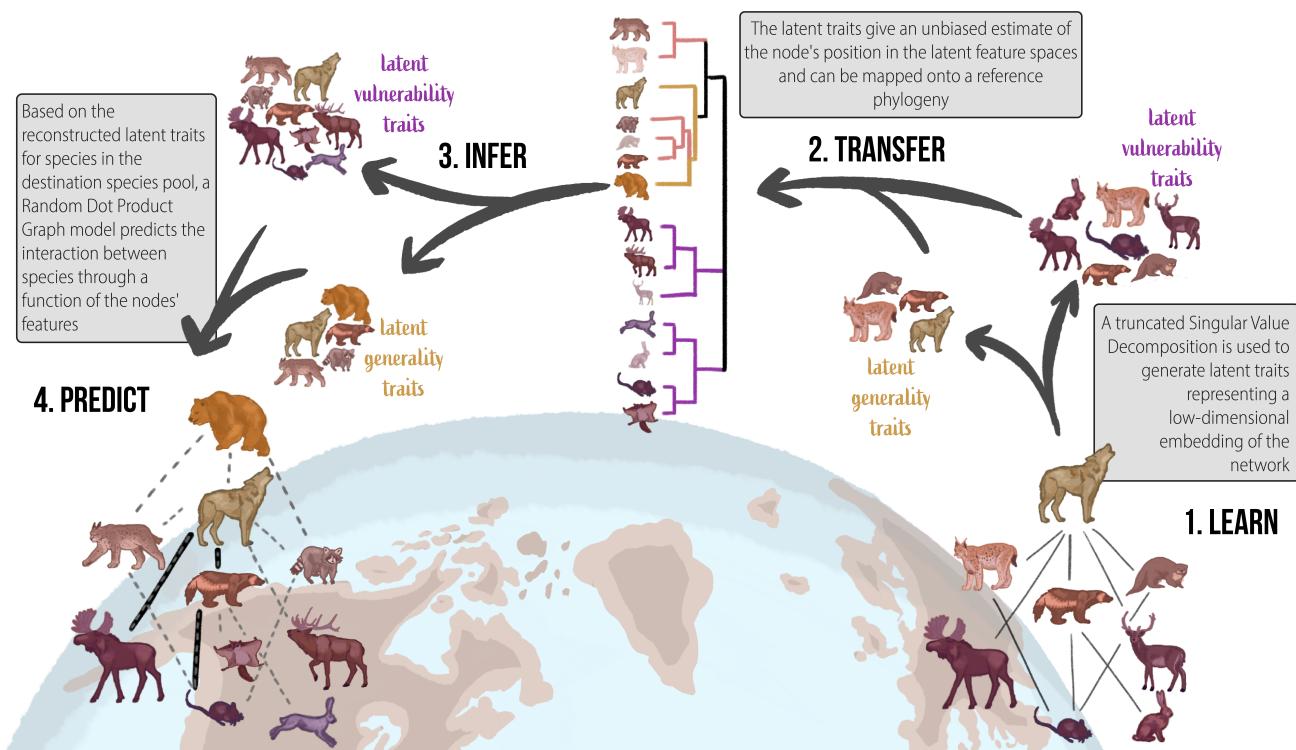


Figure 1: Overview of the phylogenetic transfer learning (and prediction) of species interaction networks. Starting from an initial, known, network, we learn its representation through a graph embedding step (here, a truncated Singular Value Decomposition; Step 1), yielding a series of latent traits (latent vulnerability traits are more representative of species at the lower trophic-level and latent generality traits are more representative of 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).

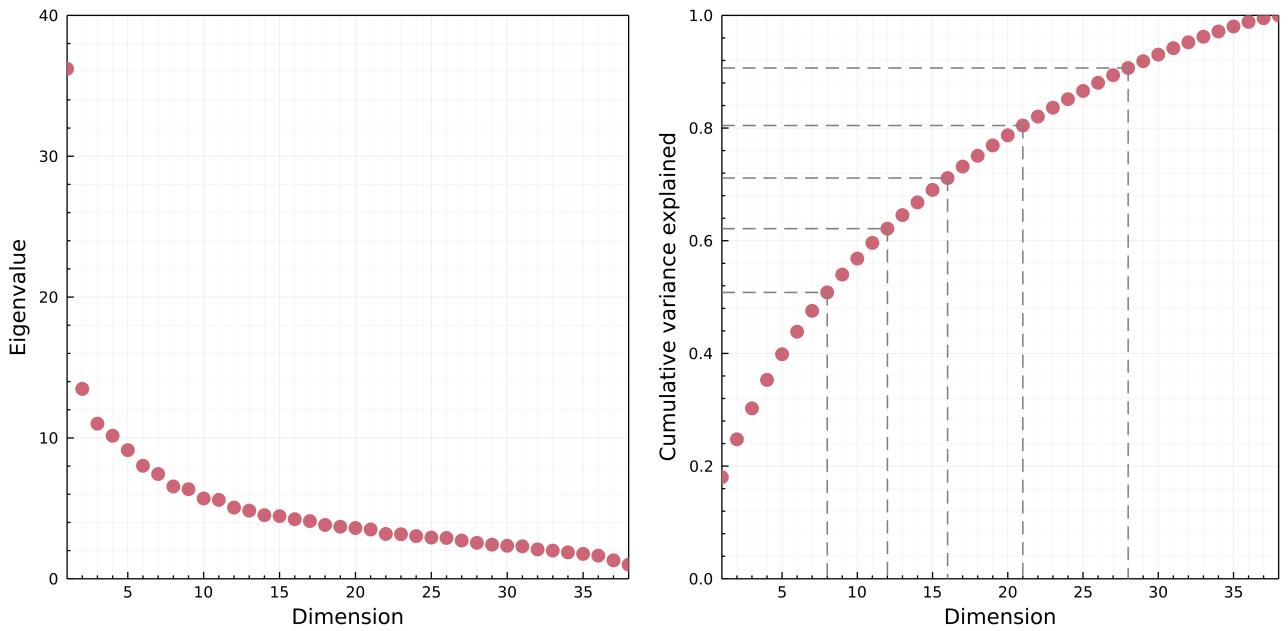


Figure 2: Left: representation of the scree plot of the singular values from the t-SVD on the European metaweb. The scree plot shows no obvious drop in the singular values that may be leveraged to automatically detect a minimal dimension for embedding, after e.g. Zhu & Ghodsi (2006). Right: cumulative 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.

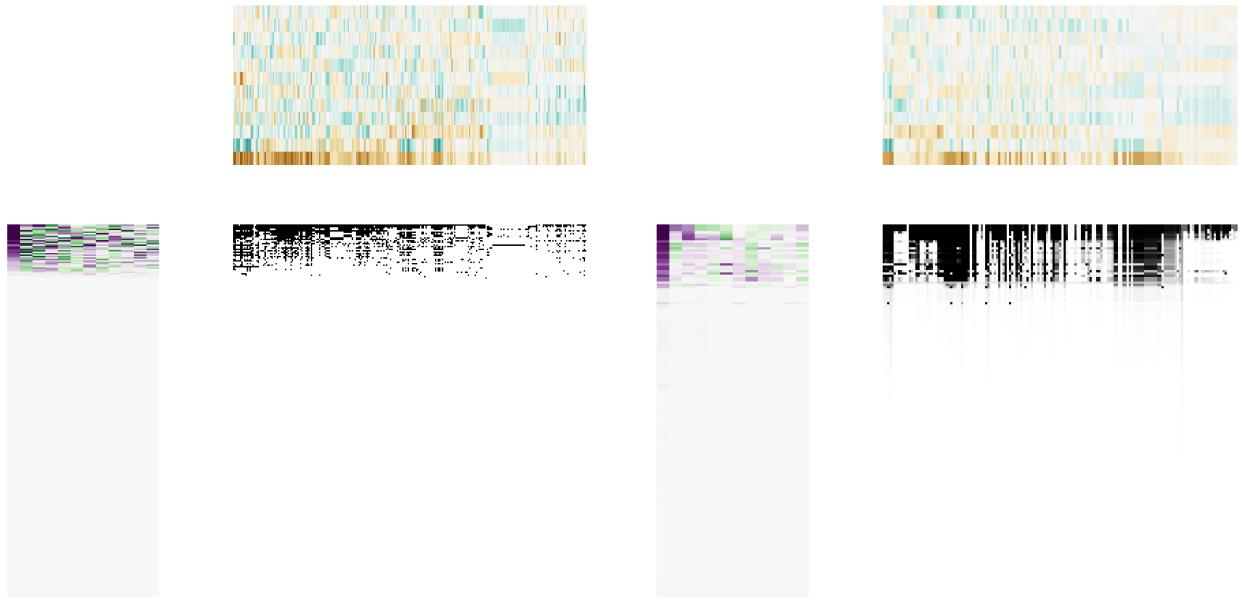


Figure 3: Visual representation of the left (green/purple) and right (green/brown) subspaces, alongside the adjacency matrix of the food web they encode (greyscale). The European metaweb is on the left, and the imputed Canadian metaweb (before data inflation) on the right. This figure illustrates how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

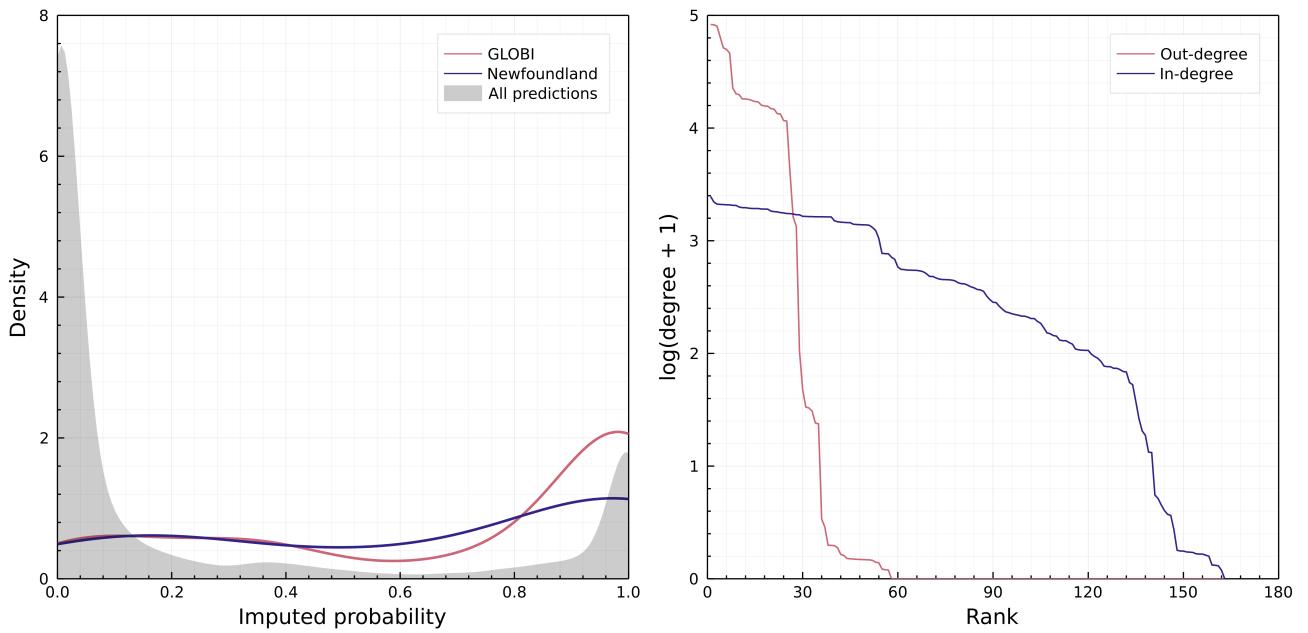


Figure 4: Left, comparison of the probabilities of interactions assigned by the model to all interactions (grey curve), the subset of interactions found in GLOBI (red), and in the Strong & Leroux (2014) Newfoundland dataset (blue). The model recovers more interactions with a low probability compared 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.

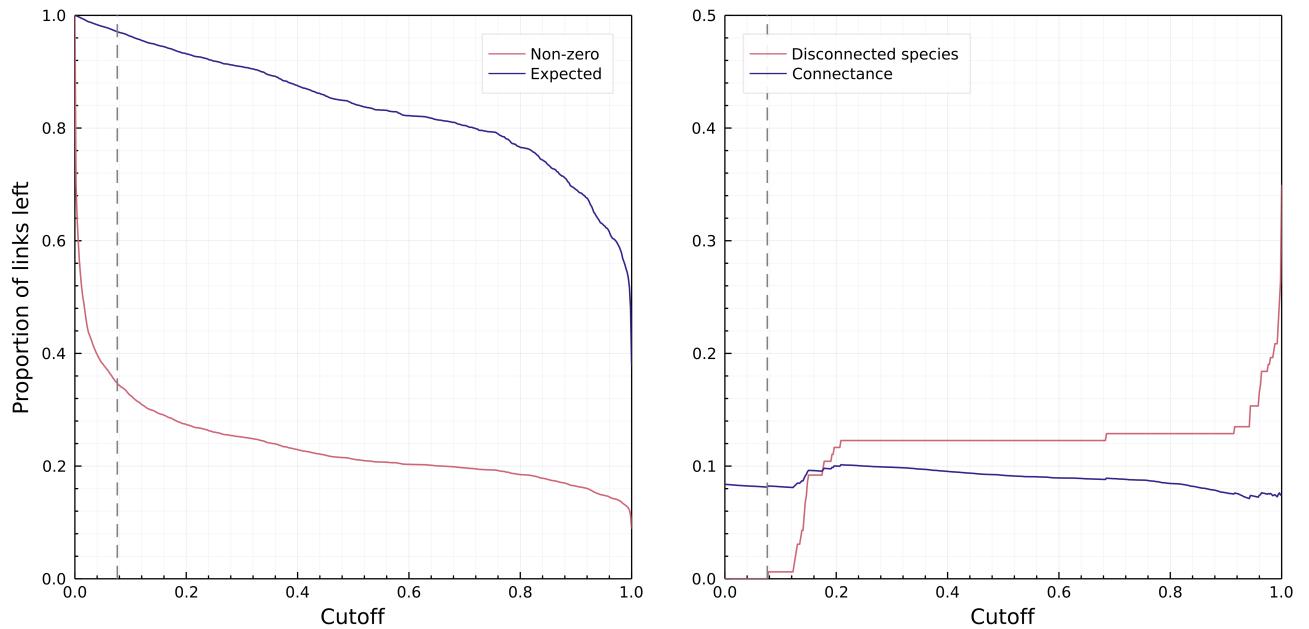


Figure 5: Left: effect of varying the cutoff for probabilities to be considered non-zero on the number of unique links and on \hat{L} , the probabilistic estimate of the number of links assuming that all interactions are independent. Right: effect of varying the cutoff on the number of disconnected species, 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.

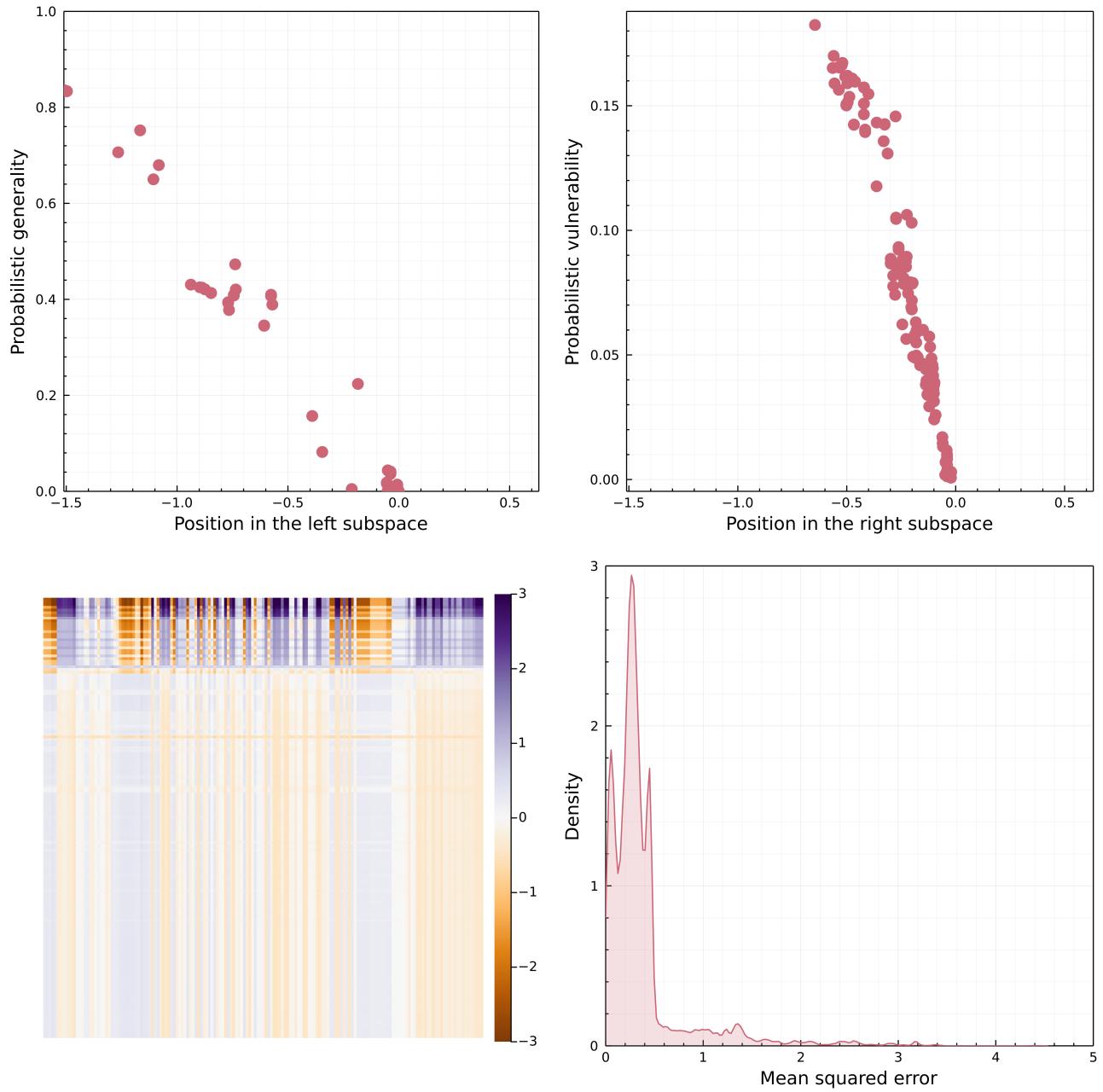


Figure 6: Top: biological significance of the first dimension. Left: there is a linear relationship between the values on the first dimension of the left subspace and the generality, *i.e.* the relative number of preys, *sensu* Schoener (1989). Species with a value of 0 in this subspace are at the bottom-most 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.