

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 Hemming⁴ Benjamin Mercier^{3,2} Laura
J. Pollock⁵ Rogini Runghen⁷ Giulio V. Dalla Riva⁸ Timothée Poisot^{1,2}

¹ Département de Sciences Biologiques, Université de Montréal, Montréal, Canada ² Québec Centre for Biodiversity Sciences, Montréal, Canada ³ Université de Sherbrooke, Sherbrooke, Canada ⁴ Department of Forest Resources Management, University of British Columbia, Vancouver, Canada ⁵ Department of Biology, McGill University, Montréal, Canada ⁶ Department of Ecology & Evolutionary Biology, University of Toronto, Toronto, Canada ⁷ Centre for Integrative Ecology, School of Biological Sciences, University of Canterbury, Canterbury, New Zealand ⁸ 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

Abstract: 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 deficit of data of which species interact, and what we can expect the network structure of these interactions to be. 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 solution 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. We provide guidance on how this method can be adapted by substituting some approaches or predictors in order to make it more generally applicable to a broad family of ecological problems.

1 Introduction

2 There are two core challenges we are faced with in furthering our understanding of ecological
3 networks across space, particularly at macro-ecologically relevant scales (*e.g.* Trøjelsgaard and
4 Olesen 2016). First, networks within a location are difficult to sample properly (Jordano 2016a,
5 2016b), resulting in a widespread “Eltonian shortfall” (Hortal et al. 2015). This first challenge
6 (local incompleteness) has been, in large part, addressed by the recent multiplication of meth-
7 ods aiming to predict interactions *within* an *existing* network, a lot of which are reviewed in
8 Strydom et al. (2021). Second, recent analyses based on collected data (Poisot, Bergeron, et
9 al. 2021) or metadata (Cameron et al. 2019) highlight that ecological networks are currently
10 studied in a biased subset of space and bioclimates, which impedes our ability to generalize
11 any local understanding of network structure. Meaning that although the framework to address
12 incompleteness *within* a network exists there would still be regions that, due to a *lack* of local
13 interaction data, we are unable to infer potential species interactions. Having a general solution
14 for the issue of metaweb inference (Morales-Castilla et al. 2015) that, despite situations where
15 minimal knowledge about interactions within a species pool is known, is capable of producing
16 a plausible metaweb could be the catalyst for significant breakthroughs in our ability to start
17 thinking about species interactions networks over large spatial scales.

18 Here, we present a general method for the transfer learning of network representations, relying
19 on the similarities of species in a biologically/ecologically relevant proxy space (*e.g.* shared
20 morphology or ancestry). Transfer learning is a machine learning methodology that uses the
21 knowledge gained from solving one problem and applying it to a related (destination) problem
22 (Torrey and Shavlik 2010; Pan and Yang 2010). In this instance, we solve the problem of pre-
23 dicting trophic interactions between species, based on knowledge extracted from another species
24 pool for which interactions are known, using phylogenetic structure as a medium for transfer.
25 This allows us to construct a *probabilistic* metaweb for a community for which we have *no* prior
26 interaction data for the desired species pool. Our methodology is outlined in fig. 1, where we
27 provide an illustration based on learning an embedding of a metaweb of trophic interactions for
28 European mammals (known interactions; Maiorano et al. 2020b, 2020a), and based on phy-
29 logenetic relationships between mammals globally (Upham, Esselstyn, and Jetz 2019), infer a

30 metaweb for the Canadian mammalian species pool (interactions are treated as unknown in this
31 instance).

32 [Figure 1 about here.]

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

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

58 **Data used for the case study**

59 We use data on the European metaweb assembled by Maiorano et al. (2020b), following the
60 definition of the metaweb first introduced by Dunne (2006), *i.e.* an inventory of all possible
61 interactions within species from a spatially delimited pool. Notably the metaweb is not a pre-
62 diction of the food web at any specific locale within the frontiers of the species pool – in fact,
63 these local food webs are expected to have a subset of both the species and the interactions of
64 their metaweb (Poisot et al. 2012). This being said, as the metaweb represents the total of func-
65 tional, phylogenetic, and macroecological processes (Morales-Castilla et al. 2015), it is thus
66 still worthy of ecological attention. We induced the subgraph corresponding to all mammals by
67 matching species names in the original network first to the GBIF taxonomic backbone (GBIF
68 Secretariat 2021) and retaining all those who matched to mammals; all nodes had valid matches
69 to GBIF at this step, and so this backbone is used for all name reconciliation steps as outlined
70 below.

71 The European metaweb represents the knowledge we want to learn and transfer; the phyloge-
72 netic similarity of mammals here represents the support for transfer. We used the mammalian
73 consensus supertree by Upham, Esselstyn, and Jetz (2019), for which all approximatively 6000
74 names have been similarly matched to their GBIF valid names. This step allows us to place each
75 node of the mammalian European metaweb in the phylogeny.

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

81 After taxonomic cleaning and reconciliation as outlined in the following sections, the mam-
82 malian European metaweb had 260 species, and the Canadian species pool has 163; of these, 17
83 (about 4% of the total) are shared, and 89 species from Canada (54%) had at least one congeneric
84 species in Europe. The similarity for both species pool predictably increases with higher taxo-
85 nomic order, with 19% of shared genera, 47% of shared families, and 75% of shared orders; for

86 the last point, Canada and Europe each had a single unique order (*Didelphimorphia* for Canada,
87 *Erinaceomorpha* for Europe).

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

91 **Method description**

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

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

108 **Implementation and code availability**

109 The entire pipeline is implemented in *Julia* 1.6 (Bezanson et al. 2017) and is available under the
110 permissive MIT License at <https://osf.io/2zwqm/>. The taxonomic cleanup steps are done
111 using GBIF.jl (Dansereau and Poisot 2021). The network embedding and analysis is done

112 using EcologicalNetworks.jl (Banville, Vissault, and Poisot 2021; Poisot et al. 2019). The
113 phylogenetic simulations are done using PhyloNetworks.jl (Solís-Lemus, Bastide, and Ané
114 2017) and Phylo.jl (Reeve et al. 2016). A complete Project.toml file specifying the full tree
115 of dependencies is available alongside the code. This material also includes a fully annotated
116 copy of the entire code required to run this project (describing both the intent of the code and
117 discussing some technical implementation details), a vignette for every step of the process, and
118 a series of Jupyter notebooks with the text and code. The pipeline can be executed on a laptop
119 in a matter of minutes, and therefore does not require extensive computational power.

120 Step 1: Learning the origin network representation

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

131 The latent variables are created by performing a truncated Singular Value Decomposition (t-
132 SVD) on the adjacency matrix. SVD is an appropriate embedding of ecological networks, which
133 has recently been shown to both capture their complex, emerging properties (Strydom, Dalla
134 Riva, and Poisot 2021) and to allow highly accurate prediction of the interactions within a single
135 network (Poisot, Ouellet, et al. 2021). Under SVD, an adjacency matrix \mathbf{A} (where $\mathbf{A}_{m,n} \in \mathbb{B}$
136 where 1 indicates predation and 0 an absence thereof) is decomposed into three components
137 resulting in $\mathbf{A} = \mathbf{L}\Sigma\mathbf{R}$. Here, Σ is a $m \times n$ diagonal matrix and contains only singular (σ) values
138 along its diagonal, \mathbf{L} is a $m \times m$ unitary matrix, and \mathbf{R}' a $n \times n$ unitary matrix. Truncating the
139 SVD removes additional noise in the dataset by omitting non-zero and/or smaller σ values from

140 Σ using the rank of the matrix. Under a t-SVD $\mathbf{A}_{m,n}$ is decomposed so that Σ is a square $r \times r$
141 diagonal matrix (where r is the rank of \mathbf{A}) containing only non-zero σ values. Additionally, \mathbf{L}
142 is now a $m \times r$ semi unitary matrix and \mathbf{R}' a $n \times r$ semi-unitary matrix.

143 The specific rank at which the SVD ought to be truncated is a difficult question. The purpose of
144 SVD is to remove the noise (expressed at high dimensions) and to focus on the signal, (expressed
145 at low dimensions). In datasets with a clear signal/noise demarcation, a scree plot of Σ can
146 show a sharp drop at the rank where noise starts (Zhu and Ghodsi 2006). Because the European
147 metaweb is almost entirely known, the amount of noise is low; this is reflected in fig. 2 (left),
148 where the scree plot shows no important drop, and in fig. 2 (right) where the proportion of
149 variance explained increases smoothly at higher dimensions. For this reason, we default back to
150 an arbitrary threshold that explains 60% of the variance in the underlying data, corresponding
151 to 12 dimensions.

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

157 [Figure 2 about here.]

158 Because RDPG relies on matrix multiplication, the higher dimensions essentially serve to make
159 specific interactions converge towards 0 or 1; therefore, for reasonably low ranks, there is no
160 guarantee that the values in the reconstructed network will be within the unit range. In order
161 to determine what constitutes an appropriate threshold for probability, we performed the RDPG
162 approach on the European metaweb, and evaluated the probability threshold by treating this as a
163 binary classification problem, specifically assuming that both 0 and 1 in the European metaweb
164 are all true. Given the methodological details given in Maiorano et al. (2020b) and O'Connor
165 et al. (2020), this seems like a reasonable assumption, although one that does not hold for all
166 metawebs. We used the thresholding approach presented in Poisot, Ouellet, et al. (2021), and

¹⁶⁷ picked a cutoff that maximized Youden's J statistic (Youden (1950); a measure of the informedness (trust) of predictions); the resulting cutoff was 0.22, and gave an accuracy above 0.99.

¹⁶⁹ 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
¹⁷⁰ on phylogenetic similarity to do so.
¹⁷¹

¹⁷² **Steps 2 and 3: Transfer learning through phylogenetic relatedness**

¹⁷³ In order to transfer the knowledge from the European metaweb to the Canadian species pool, we
¹⁷⁴ performed ancestral character estimation using a Brownian motion model, which is a conserva-
¹⁷⁵ tive approach in the absence of strong hypotheses about the nature of phylogenetic signal in the
¹⁷⁶ network decomposition (Litsios and Salamin 2012). This uses the estimated feature vectors for
¹⁷⁷ the European mammals to create a state reconstruction for all species (conceptually something
¹⁷⁸ akin to a trait-based mammalian phylogeny using generality and vulnerability traits) and allows
¹⁷⁹ us to impute the missing (latent) trait data for the Canadian species that are not already in the Eu-
¹⁸⁰ ropean network; as we are focused on predicting contemporary interactions, we only retained the
¹⁸¹ values for the tips of the tree. We assumed that all traits (*i.e.* the feature vectors for the left and
¹⁸² right subspaces) were independent, which is a reasonable assumption as every trait/dimension
¹⁸³ added to the t-SVD has an *additive* effect to the one before it. Note that the Upham, Esselstyn,
¹⁸⁴ and Jetz (2019) tree itself has some uncertainty associated to inner nodes of the phylogeny. In
¹⁸⁵ this case study, we have decided to not propagate this uncertainty, as it would complexify the
¹⁸⁶ process. The Brownian motion algorithm returns the *average* value of the trait, and its upper and
¹⁸⁷ lower bounds. Because we do not estimate other parameters of the traits' distributions, we con-
¹⁸⁸sidered that every species trait is represented as a uniform distribution between these bounds;
¹⁸⁹ in a situation where the algorithm would return point values for all simulations, one could in
¹⁹⁰ theory either estimate the parameters of a distribution for each tip, or draw randomly from the
¹⁹¹ outputs. In all cases, the inferred left and right sub-spaces for the Canadian species pool ($\hat{\mathcal{L}}$ and
¹⁹² $\hat{\mathcal{R}}$) have entries that are distributions, representing the range of values for a given species at a
¹⁹³ given dimension.

¹⁹⁴ These objects represent the transferred knowledge, which we can use for prediction of the Cana-

195 dian metaweb.

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

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

201 [Figure 3 about here.]

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

211 Because the intervals around some trait values can be broad (in fact, probably broader than what
212 they would actually be, see *e.g.* Garland, Midford, and Ives 1999), we repeat the above process
213 2×10^5 times, which results in a probabilistic metaweb P , where the probability of an interaction
214 (here conveying our degree of trust that it exists given the inferred trait distributions) is given
215 by the number of times where it appears across all random draws N , divided by the number of
216 samples. An interaction with $P_{i,j} = 1$ means that these two species were predicted to interact in
217 all 2×10^5 random draws, etc..

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

219 Once the probabilistic metaweb for Canada has been produced, we followed a number of data
220 inflation steps to finalize it.

221 [Figure 4 about here.]

222 First, we extracted the subgraph corresponding to the 17 species shared between the European
223 and Canadian pools and replaced these interactions with a probability of 0 (non-interaction) or
224 1 (interaction). This represents a minute modification of the inferred network (about 0.8% of all
225 species pairs from the Canadian web), but ensures that we are directly re-using knowledge from
226 Europe.

227 Second, we looked for all species in the Canadian pool known to the Global Biotic Interactions
228 (GLOBI) database (Poelen, Simons, and Mungall 2014), and extracted their known interactions.
229 Because GLOBI aggregates observed interactions, it is not a *networks* data source, and there-
230 fore the only information we can reliably extract from it is that a species pair *was reported to*
231 *interact at least once*. This last statement should yet be taken with caution, as some sources in
232 GLOBI (*e.g.* Thessen and Parr 2014) are produced through text analysis, and therefore may not
233 document direct evidence of the interaction. Nevertheless, should the predictive model work,
234 we would expect that a majority of interactions known to GLOBI would also be predicted. After
235 performing this check, we set the probability of all interactions known to GLOBI (366 in total,
236 33 of which were not predicted by the model, for a success rate of 91%) to 1.

237 Finally, we downloaded the data from Strong and Leroux (2014), who mined various literature
238 sources to identify trophic interactions in Newfoundland. This dataset documented 25 inter-
239 actions between mammals, only two of which were not part of our (Canada-level) predictions,
240 resulting in a success rate of 92%. These two interactions were added to our predicted metaweb
241 with a probability of 1.

242 [Figure 5 about here.]

243 Because the confidence intervals on the inferred trait space are probably over-estimates, we de-
244 cided to apply a thresholding step to the interactions after the data inflation (fig. 5). Cirtwill

245 and Hambäck (2021) proposed a number of strategies to threshold probabilistic networks. Their
246 methods assume the underlying data to be tag-based sequencing, which represents interactions
247 as co-occurrences of predator and prey within the same tags; this is conceptually identical to
248 our Bernoulli-trial based reconstruction of a probabilistic network. We performed a full analy-
249 sis of the effect of various cutoffs, and as they either resulted in removing too few interactions,
250 or removing enough interactions that species started to be disconnected from the network, we
251 set this threshold for a probability equivalent to 0 to the largest possible value that still allowed
252 all species to have at least one interaction with a non-zero probability. The need for this slight
253 deviation from the Cirtwill and Hambäck (2021) method highlights the need for additional de-
254 velopment on network thresholding.

255 Results and discussion of the case study

256 In fig. 5, we examine the effect of varying the cutoff on $P(i \rightarrow j)$ on the number of links, species,
257 and connectance. Determining a cutoff using the maximum curvature, or central difference
258 approximation of the second order partial derivative, as suggested by *e.g.* Cirtwill and Hambäck
259 (2021), results in respectively species being lost, or almost all links being kept. We therefore
260 settled on the value that allowed all species to remain with at least one interaction. This result,
261 in and of itself, suggests that additional methodological developments for the thresholding of
262 probabilistic networks are required.

263 [Figure 6 about here.]

264 The t-SVD embedding is able to learn relevant ecological features for the network. fig. 6 shows
265 that the first rank correlates linearly with generality and vulnerability (Schoener 1989), *i.e.* the
266 number of preys and predators. Importantly, this implies that a rank 1 approximation represents
267 the configuration model for the metaweb, *i.e.* a set of random networks generated from a given
268 degree sequence (Park and Newman 2004). Accounting for the probabilistic nature of the de-
269 grees, the rank 1 approximation also represents the *soft* configuration model (van der Hoorn,
270 Lippner, and Krioukov 2018). Both models are maximum entropy graph models (Garlaschelli,

271 Hollander, and Roccaverde 2018), with sharp (all network realizations satisfy the specified de-
272 gree sequence) and soft (network realizations satisfy the degree sequence on average) local con-
273 straints, respectively. The (soft) configuration model is an unbiased random graph model widely
274 used by ecologists in the context of null hypothesis significance testing of network structure
275 (*e.g.* Bascompte et al. 2003) and can provide informative priors for Bayesian inference of net-
276 work structure (*e.g.* J.-G. Young, Cantwell, and Newman 2021). It is noteworthy that for this
277 metaweb, the relevant information was extracted at the first rank. Because the first rank corre-
278 sponds to the leading singular value of the system, the results of fig. 6 have a straightforward
279 interpretation: degree-based processes are the most important in structuring the mammalian
280 food web.

281 Discussion

282 One important aspect in which Europe and Canada differ (despite their comparable bioclimatic
283 conditions) is the legacy of human impacts, which have been much longer in Europe. Nenzén,
284 Montoya, and Varela (2014) showed that even at small scales (the Iberian peninsula), mammal
285 food webs retain the signal of both climate change and human activity, even when this human
286 activity was orders of magnitude less important than it is now. Similarly, Yeakel et al. (2014)
287 showed that changes in human occupation over several centuries can lead to food web collapse.
288 Megafauna in particular seems to be very sensitive to human arrival (Pires et al. 2015). In short,
289 there is well-substantiated support for the idea that human footprint affects more than the risk of
290 species extinction (Marco et al. 2018), and can lead to changes in interaction structure. Yet, ow-
291 ing to the inherent plasticity of interactions, there have been documented instances of food webs
292 undergoing rapid collapse/recovery cycles over short periods of time (Pedersen et al. 2017).
293 The embedding of a network, in a sense, embeds its macro-evolutionary history, especially as
294 RDGP captures ecological signal (Dalla Riva and Stouffer 2016); at this point, it is important to
295 recall that a metaweb is intended as a catalogue of all possible interactions, which should then
296 be filtered (Morales-Castilla et al. 2015). In practice (and in this instance) the reconstructed
297 metaweb will predict interactions that are plausible based on the species' evolutionary history,
298 however some interactions would not be realized due to human impact.

299 Cirtwill et al. (2019) previously made the point that network inference techniques based on
300 Bayesian approaches would perform far better in the presence of an interaction-level informa-
301 tive prior; the desirable properties of such a prior would be that it is expressed as a probability,
302 preferably representing a Bernoulli event, the value of which would be representative of rele-
303 vant biological processes. We argue that the probability returned at the very last step of our
304 framework may serve as this informative prior; indeed, the output of our analysis can be used in
305 subsequent steps, also possibly involving expert elicitation to validate some of the most strongly
306 recommended interactions. One important *caveat* to keep in mind when working with interac-
307 tion inference is that interactions can never really be true negatives (in the current state of our
308 methodological framework and data collection limitations); this renders the task of validating
309 a model through the usual application of binary classification statistics very difficult (although
310 see Strydom et al. 2021 for a discussion of alternative suggestions).

311 As Herbert (1965) rightfully pointed out, “[y]ou can’t draw neat lines around planet-wide prob-
312 lems”; in this regard, our approach must contend with two interesting problems. The first is the
313 limit of the metaweb to embed and transfer. If the initial metaweb is too narrow in scope, notably
314 from a taxonomic point of view, the chances of finding another area with enough related species
315 to make a reliable inference decrease. This is notably true if the metaweb is assembled in an area
316 with mostly endemic species. Conversely, the metaweb should be reliably filled, which assumes
317 that the S^2 interactions in a pool of S species have been examined, either through literature
318 surveys or expert elicitation. The second problem is to determine which area should be used to
319 infer the new metaweb in, as this determines the species pool that must be used. In our applica-
320 tion, we focused on the mammals of Canada. The upside of this approach is that information at
321 the country level is likely to be required by policy makers and stakeholders for their biodiversity
322 assessment, as each country tends to set goals at the national level (Buxton et al. 2021) for which
323 quantitative instruments are designed (Turak et al. 2017), with specific strategies often enacted
324 at smaller scales (Ray, Grimm, and Olive 2021). Yet these national divisions, in large parts
325 of the world, reflect nothing except for the legacy of settler colonialism, and operating under
326 them must be done under the clear realization that they contributed to the ongoing biodiversity
327 crisis (Adam 2014), can reinforce environmental injustice (Choudry 2013; Domínguez and Lu-

328 oma 2020), and on Turtle Island especially, will probably end up being replaced by Indigenous
329 principles of land management (Eichhorn, Baker, and Griffiths 2019; No’kmaq et al. 2021).

330 **Acknowledgements:** We acknowledge that this study was conducted on land within the tra-
331 ditional unceded territory of the Saint Lawrence Iroquoian, Anishinabewaki, Mohawk, Huron-
332 Wendat, and Omàmiwininiwak nations. TP, TS, DC, and LP received funding from the Canadian
333 Institut for Ecology & Evolution. FB is funded by the Institut de Valorisation des Données. TS,
334 SB, and TP are funded by a donation from the Courtois Foundation. CB was awarded a Mi-
335 tacs Elevate Fellowship no. IT12391, in partnership with fRI Research, and also acknowledges
336 funding from Alberta Innovates and the Forest Resources Improvement Association of Alberta.
337 M-JF acknowledges funding from NSERC Discovery Grant and NSERC CRC RR is funded by
338 New Zealand’s Biological Heritage Ngā Koiora Tuku Iho National Science Challenge, admin-
339 istered by New Zealand Ministry of Business, Innovation, and Employment. BM is funded by
340 the NSERC Alexander Graham Bell Canada Graduate Scholarship and the FRQNT master’s
341 scholarshipTP acknowledges financial support from NSERC through the Discovery Grants and
342 Discovery Accelerator Supplement programs.

343 References

- 344 Adam, Rachelle. 2014. *Elephant Treaties: The Colonial Legacy of the Biodiversity Crisis*.
345 UPNE.
- 346 Albouy, Camille, Philippe Archambault, Ward Appeltans, Miguel B. Araújo, David Beauchesne,
347 Kevin Cazelles, Alyssa R. Cirtwill, et al. 2019. “The Marine Fish Food Web Is Globally
348 Connected.” *Nature Ecology & Evolution* 3 (8): 1153–61. <https://doi.org/10.1038/s41559-019-0950-y>.
- 349
- 350 Banville, Francis, Steve Vissault, and Timothée Poisot. 2021. “Mangal.jl and EcologicalNet-
351 works.jl: Two Complementary Packages for Analyzing Ecological Networks in Julia.” *Jour-
352 nal of Open Source Software* 6 (61): 2721. <https://doi.org/10.21105/joss.02721>.
- 353 Bascompte, J., P. Jordano, C. J. Melian, and J. M. Olesen. 2003. “The Nested Assembly of Plant-
354 Animal Mutualistic Networks.” *Proceedings of the National Academy of Sciences* 100 (16):

- 355 9383–87. <https://doi.org/10.1073/pnas.1633576100>.
- 356 Beckerman, Andrew P., Owen L. Petchey, and Philip H. Warren. 2006. “Foraging Biology
357 Predicts Food Web Complexity.” *Proceedings of the National Academy of Sciences* 103
358 (37): 13745–49. <https://doi.org/10.1073/pnas.0603039103>.
- 359 Bezanson, J., A. Edelman, S. Karpinski, and V. Shah. 2017. “Julia: A Fresh Approach to Numer-
360 ical Computing.” *SIAM Review* 59 (1): 65–98. <https://doi.org/10.1137/141000671>.
- 361 Boeckaerts, Dimitri, Michiel Stock, Bjorn Criel, Hans Gerstmans, Bernard De Baets, and Yves
362 Briers. 2021. “Predicting Bacteriophage Hosts Based on Sequences of Annotated Receptor-
363 Binding Proteins.” *Scientific Reports* 11 (1): 1467. <https://doi.org/10.1038/s41598-021-81063-4>.
- 364 Braga, Mariana P., Niklas Janz, Sören Nylin, Fredrik Ronquist, and Michael J. Landis. 2021.
365 “Phylogenetic Reconstruction of Ancestral Ecological Networks Through Time for Pierid
366 Butterflies and Their Host Plants.” *Ecology Letters* n/a (n/a). <https://doi.org/10.1111/ele.13842>.
- 368 Buxton, Rachel T., Joseph R. Bennett, Andrea J. Reid, Charles Shulman, Steven J. Cooke,
369 Charles M. Francis, Elizabeth A. Nyboer, et al. 2021. “Key Information Needs to Move
370 from Knowledge to Action for Biodiversity Conservation in Canada.” *Biological Conserva-*
371 *tion* 256: 108983. <https://doi.org/10.1016/j.biocon.2021.108983>.
- 372 Cameron, Erin K., Maja K. Sundqvist, Sally A. Keith, Paul J. CaraDonna, Erik A. Mousing,
373 Karin A. Nilsson, Daniel B. Metcalfe, and Aimée T. Classen. 2019. “Uneven Global
374 Distribution of Food Web Studies Under Climate Change.” *Ecosphere* 10 (3): e02645.
375 <https://doi.org/10.1002/ecs2.2645>.
- 376 Cavender-Bares, Jeannine, Kenneth H. Kozak, Paul V. A. Fine, and Steven W. Kembel. 2009.
377 “The Merging of Community Ecology and Phylogenetic Biology.” *Ecology Letters* 12 (7):
378 693–715. <https://doi.org/10.1111/j.1461-0248.2009.01314.x>.
- 379 Choudry, Aziz. 2013. “Saving Biodiversity, for Whom and for What? Conservation NGOs,
380 Complicity, Colonialism and Conquest in an Era of Capitalist Globalization.” In *NGOiza-
381 tion: Complicity, Contradictions and Prospects*, 24–44. Bloomsbury Publishing.
- 382 Cirtwill, Alyssa R., Anna Eklf, Tomas Roslin, Kate Wootton, and Dominique Gravel. 2019. “A

- 383 Quantitative Framework for Investigating the Reliability of Empirical Network Construc-
384 tion.” *Methods in Ecology and Evolution* 0 (ja). <https://doi.org/10.1111/2041-210X.13180>.
- 386 Cirtwill, Alyssa R., and Peter Hambäck. 2021. “Building Food Networks from Molecular Data:
387 Bayesian or Fixed-Number Thresholds for Including Links.” *Basic and Applied Ecology* 50:
388 67–76. <https://doi.org/10.1016/j.baae.2020.11.007>.
- 389 Dalla Riva, Giulio V., and Daniel B. Stouffer. 2016. “Exploring the Evolutionary Signature of
390 Food Webs’ Backbones Using Functional Traits.” *Oikos* 125 (4): 446–56. <https://doi.org/10.1111/oik.02305>.
- 392 Dansereau, Gabriel, and Timothée Poisot. 2021. “SimpleSDMLayers.jl and GBIF.jl: A Frame-
393 work for Species Distribution Modeling in Julia.” *Journal of Open Source Software* 6 (57):
394 2872. <https://doi.org/10.21105/joss.02872>.
- 395 Domínguez, Lara, and Colin Luoma. 2020. “Decolonising Conservation Policy: How Colo-
396 nial Land and Conservation Ideologies Persist and Perpetuate Indigenous Injustices at the
397 Expense of the Environment.” *Land* 9 (3): 65. <https://doi.org/10.3390/land9030065>.
- 398 Dormann, Carsten F., Bernd Gruber, Marten Winter, and Dirk Herrmann. 2010. “Evolution
399 of Climate Niches in European Mammals?” *Biology Letters* 6 (2): 229–32. <https://doi.org/10.1098/rsbl.2009.0688>.
- 401 Dunne, Jennifer A. 2006. “The Network Structure of Food Webs.” In *Ecological Networks:
402 Linking Structure and Dynamics*, edited by Jennifer A Dunne and Mercedes Pascual, 27–86.
403 Oxford University Press.
- 404 Eichhorn, Markus P., Kate Baker, and Mark Griffiths. 2019. “Steps Towards Decolonising
405 Biogeography.” *Frontiers of Biogeography* 12 (1): 1–7. <https://doi.org/10.21425/F5FBG44795>.
- 407 Eklöf, Anna, and Daniel B. Stouffer. 2016. “The Phylogenetic Component of Food Web Struc-
408 ture and Intervallity.” *Theoretical Ecology* 9 (1): 107–15. <https://doi.org/10.1007/s12080-015-0273-9>.

- 410 Garland, THEODORE, JR., PETER E. Midford, and ANTHONY R. Ives. 1999. “An In-
411 troduction to Phylogenetically Based Statistical Methods, with a New Method for Con-
412 fidence Intervals on Ancestral Values1.” *American Zoologist* 39 (2): 374–88. <https://doi.org/10.1093/icb/39.2.374>.
- 413
- 414 Garlaschelli, Diego, Frank den Hollander, and Andrea Rocaverde. 2018. “Covariance Struc-
415 ture Behind Breaking of Ensemble Equivalence in Random Graphs.” *Journal of Statistical*
416 *Physics* 173 (3-4): 644–62. <https://doi.org/10.1007/s10955-018-2114-x>.
- 417 GBIF Secretariat. 2021. “GBIF Backbone Taxonomy.” <https://doi.org/10.15468/39omei>.
- 418 Gerhold, Pille, James F. Cahill, Marten Winter, Igor V. Bartish, and Andreas Prinzing. 2015.
419 “Phylogenetic Patterns Are Not Proxies of Community Assembly Mechanisms (they Are Far
420 Better).” *Functional Ecology* 29 (5): 600–614. <https://doi.org/10.1111/1365-2435.12425>.
- 421
- 422 Halko, N., P. G. Martinsson, and J. A. Tropp. 2011. “Finding Structure with Randomness: Prob-
423 abilistic Algorithms for Constructing Approximate Matrix Decompositions.” *SIAM Review*
424 53 (2): 217–88. <https://doi.org/10.1137/090771806>.
- 425 Herbert, Frank. 1965. *Dune*. First. Philadelphia: Chilton Book Company.
- 426 Hoorn, Pim van der, Gabor Lippner, and Dmitri Krioukov. 2018. “Sparse Maximum-Entropy
427 Random Graphs with a Given Power-Law Degree Distribution.” *Journal of Statistical Physics*
428 173 (3-4): 806–44. <https://doi.org/10.1007/s10955-017-1887-7>.
- 429 Hortal, Joaquín, Francesco de Bello, José Alexandre F. Diniz-Filho, Thomas M. Lewinsohn,
430 Jorge M. Lobo, and Richard J. Ladle. 2015. “Seven Shortfalls That Beset Large-Scale
431 Knowledge of Biodiversity.” *Annual Review of Ecology, Evolution, and Systematics* 46 (1):
432 523–49. <https://doi.org/10.1146/annurev-ecolsys-112414-054400>.
- 433 Hutchinson, Matthew C., Edgar Fernando Cagua, and Daniel B. Stouffer. 2017. “Cophyloge-
434 netic Signal Is Detectable in Pollination Interactions Across Ecological Scales.” *Ecology*,
435 n/a-. <https://doi.org/10.1002/ecy.1955>.
- 436 Jordano, Pedro. 2016a. “Chasing Ecological Interactions.” *PLOS Biol* 14 (9): e1002559.
437 <https://doi.org/10.1371/journal.pbio.1002559>.

- 438 ———. 2016b. “Sampling Networks of Ecological Interactions.” Edited by Daniel Stouffer.
439 *Functional Ecology* 30 (12): 1883–93. <https://doi.org/10.1111/1365-2435.12763>.
- 440 Litsios, Glenn, and Nicolas Salamin. 2012. “Effects of Phylogenetic Signal on Ancestral State
441 Reconstruction.” *Systematic Biology* 61 (3): 533–38. <https://doi.org/10.1093/sysbio/syr124>.
- 443 Maiorano, Luigi, Alessandro Montemaggioli, Gentile Francesco Ficetola, Louise O’Connor,
444 and Wilfried Thuiller. 2020a. “Data from: Tetra-EU 1.0: A Species-Level Trophic Meta-
445 Web of European Tetrapods.” Dryad. <https://doi.org/10.5061/DRYAD.JM63XSJ7B>.
- 446 ———. 2020b. “TETRA-EU 1.0: A Species-Level Trophic Metaweb of European Tetrapods.”
447 Edited by Allen Hurlbert. *Global Ecology and Biogeography* 29 (9): 1452–57. <https://doi.org/10.1111/geb.13138>.
- 449 Marco, Moreno Di, Oscar Venter, Hugh P. Possingham, and James E. M. Watson. 2018. “Changes
450 in Human Footprint Drive Changes in Species Extinction Risk.” *Nature Communications* 9
451 (1): 4621. <https://doi.org/10.1038/s41467-018-07049-5>.
- 452 Morales-Castilla, Ignacio, Miguel G. Matias, Dominique Gravel, and Miguel B. Araújo. 2015.
453 “Inferring Biotic Interactions from Proxies.” *Trends in Ecology & Evolution* 30 (6): 347–56.
454 <https://doi.org/10.1016/j.tree.2015.03.014>.
- 455 Mouquet, Nicolas, Vincent Devictor, Christine N. Meynard, Francois Munoz, Louis-Félix Bersier,
456 Jérôme Chave, Pierre Couteron, et al. 2012. “Ecophylogenetics: Advances and Perspec-
457 tives.” *Biological Reviews* 87 (4): 769–85. <https://doi.org/10.1111/j.1469-185X.2012.00224.x>.
- 459 Nenzén, Hedvig K., Daniel Montoya, and Sara Varela. 2014. “The Impact of 850,000 Years of
460 Climate Changes on the Structure and Dynamics of Mammal Food Webs.” *PLOS ONE* 9
461 (9): e106651. <https://doi.org/10.1371/journal.pone.0106651>.
- 462 No’kmaq, M’sit, Albert Marshall, Karen F. Beazley, Jessica Hum, shalan joudry, Anastasia Pa-
463 padopoulos, Sherry Pictou, Janet Rabesca, Lisa Young, and Melanie Zurba. 2021. “‘Awak-
464 ening the Sleeping Giant’: Re-Indigenization Principles for Transforming Biodiversity Con-
465 servation in Canada and Beyond.” *FACETS* 6 (1): 839–69.

- 466 O'Connor, Louise M. J., Laura J. Pollock, João Braga, Gentile Francesco Ficetola, Luigi Maiorano, Camille Martinez-Almoyna, Alessandro Montemaggiori, Marc Ohlmann, and Wilfried Thuiller. 2020. "Unveiling the Food Webs of Tetrapods Across Europe Through the Prism of the Eltonian Niche." *Journal of Biogeography* 47 (1): 181–92. <https://doi.org/10.1111/jbi.13773>.
- 471 Pan, Sinno Jialin, and Qiang Yang. 2010. "A Survey on Transfer Learning." *IEEE Transactions on Knowledge and Data Engineering* 22 (10): 1345–59. <https://doi.org/10.1109/TKDE.2009.191>.
- 474 Park, Juyong, and M. E. J. Newman. 2004. "Statistical Mechanics of Networks." *Physical Review E* 70 (6): 066117. <https://doi.org/10.1103/PhysRevE.70.066117>.
- 476 Pedersen, Eric J., Patrick L. Thompson, R. Aaron Ball, Marie-Josée Fortin, Tarik C. Gouhier, Heike Link, Charlotte Moritz, et al. 2017. "Signatures of the Collapse and Incipient Recovery of an Overexploited Marine Ecosystem." *Royal Society Open Science* 4 (7): 170215. <https://doi.org/10.1098/rsos.170215>.
- 480 Pires, Mathias M., Paul L. Koch, Richard A. Fariña, Marcus A. M. de Aguiar, Sérgio F. dos Reis, and Paulo R. Guimarães. 2015. "Pleistocene Megafaunal Interaction Networks Became More Vulnerable After Human Arrival." *Proceedings of the Royal Society B: Biological Sciences* 282 (1814): 20151367. <https://doi.org/10.1098/rspb.2015.1367>.
- 484 Poelen, Jorrit H., James D. Simons, and Chris J. Mungall. 2014. "Global Biotic Interactions: An Open Infrastructure to Share and Analyze Species-Interaction Datasets." *Ecological Informatics* 24: 148–59. <https://doi.org/10.1016/j.ecoinf.2014.08.005>.
- 487 Poisot, Timothée, Zacharie Belisle, Laura Hoebelke, Michiel Stock, and Piotr Szefer. 2019. "EcologicalNetworks.jl - Analysing Ecological Networks." *Ecography*. <https://doi.org/10.1111/ecog.04310>.
- 490 Poisot, Timothée, Gabriel Bergeron, Kevin Cazelles, Tad Dallas, Dominique Gravel, Andrew MacDonald, Benjamin Mercier, Clément Violet, and Steve Vissault. 2021. "Global Knowledge Gaps in Species Interaction Networks Data." *Journal of Biogeography* n/a (n/a). <https://doi.org/10.1111/jbi.14127>.

- 494 Poisot, Timothée, Elsa Canard, David Mouillot, Nicolas Mouquet, and Dominique Gravel. 2012.
495 “The Dissimilarity of Species Interaction Networks.” *Ecology Letters* 15 (12): 1353–61.
496 <https://doi.org/10.1111/ele.12002>.
- 497 Poisot, Timothée, Alyssa R. Cirtwill, Kévin Cazelles, Dominique Gravel, Marie-Josée Fortin,
498 and Daniel B. Stouffer. 2016. “The Structure of Probabilistic Networks.” Edited by Jana
499 Vamosi. *Methods in Ecology and Evolution* 7 (3): 303–12. <https://doi.org/10.1111/2041-210X.12468>.
- 500 Poisot, Timothée, Marie-Andrée Ouellet, Nardus Mollentze, Maxwell J. Farrell, Daniel J. Becker,
501 Gregory F. Albery, Rory J. Gibb, Stephanie N. Seifert, and Colin J. Carlson. 2021. “Im-
502 putting the Mammalian Virome with Linear Filtering and Singular Value Decomposition.”
503 *arXiv:2105.14973 [q-Bio]*. <http://arxiv.org/abs/2105.14973>.
- 504 Poisot, Timothée, and Daniel B. Stouffer. 2018. “Interactions Retain the Co-Phylogenetic
505 Matching That Communities Lost.” *Oikos* 127 (2): 230–38. <https://doi.org/10.1111/oik.03788>.
- 506 Price, Peter W. 2003. *Macroevolutionary Theory on Macroecological Patterns*. Cambridge
507 University Press.
- 508 Ray, Justina C., Jaime Grimm, and Andrea Olive. 2021. “The Biodiversity Crisis in Canada:
509 Failures and Challenges of Federal and Sub-National Strategic and Legal Frameworks.”
510 *FACETS* 6: 1044–68. <https://doi.org/10.1139/facets-2020-0075>.
- 511 Reeve, Richard, Tom Leinster, Christina A. Cobbold, Jill Thompson, Neil Brummitt, Sonia N.
512 Mitchell, and Louise Matthews. 2016. “How to Partition Diversity.” *arXiv:1404.6520 [q-
513 Bio]*. <http://arxiv.org/abs/1404.6520>.
- 514 Runghen, Rogini, Daniel B Stouffer, and Giulio V Dalla Riva. 2021. “Exploiting Node Metadata
515 to Predict Interactions in Large Networks Using Graph Embedding and Neural Networks.”
516 <https://doi.org/10.1101/2021.06.10.447991>.
- 517 Schoener, Thomas W. 1989. “Food Webs from the Small to the Large.” *Ecology* 70 (6): 1559–
518 89.
- 519 Solís-Lemus, Claudia, Paul Bastide, and Cécile Ané. 2017. “PhyloNetworks: A Package for
520

- 522 Phylogenetic Networks.” *Molecular Biology and Evolution* 34 (12): 3292–98. <https://doi.org/10.1093/molbev/msx235>.
- 523
- 524 Stock, Michiel. 2021. “Pairwise Learning for Predicting Pollination Interactions Based on Traits
525 and Phylogeny.” *Ecological Modelling*, 14.
- 526 Stouffer, D. B., M. Sales-Pardo, M. I. Sirer, and J. Bascompte. 2012. “Evolutionary Conserva-
527 tion of Species’ Roles in Food Webs.” *Science* 335 (6075): 1489–92. <https://doi.org/10.1126/science.1216556>.
- 528
- 529 Strong, Justin S., and Shawn J. Leroux. 2014. “Impact of Non-Native Terrestrial Mammals on
530 the Structure of the Terrestrial Mammal Food Web of Newfoundland, Canada.” *PLOS ONE*
531 9 (8): e106264. <https://doi.org/10.1371/journal.pone.0106264>.
- 532
- 533 Strydom, Tanya, Michael David Catchen, Francis Banville, Dominique Caron, Gabriel Dansereau,
534 Philippe Desjardins-Proulx, Norma Forero, et al. 2021. “A Roadmap Toward Predicting
535 Species Interaction Networks (Across Space and Time).” Preprint. EcoEvoRxiv. <https://doi.org/10.32942/osf.io/eu7k3>.
- 536
- 537 Strydom, Tanya, Giulio V. Dalla Riva, and Timothée Poisot. 2021. “SVD Entropy Reveals the
538 High Complexity of Ecological Networks.” *Frontiers in Ecology and Evolution* 9. <https://doi.org/10.3389/fevo.2021.623141>.
- 539
- 540 Thessen, Anne E., and Cynthia Sims Parr. 2014. “Knowledge Extraction and Semantic Anno-
541 tation of Text from the Encyclopedia of Life.” *PloS One* 9 (3): e89550.
- 542
- 543 Torrey, Lisa, and Jude Shavlik. 2010. “Transfer Learning.” In *Handbook of Research on Ma-*
chine Learning Applications and Trends: Algorithms, Methods, and Techniques, 242–64.
IGI global.
- 544
- 545 Trøjelsgaard, Kristian, and Jens M. Olesen. 2016. “Ecological Networks in Motion: Micro- and
Macroscopic Variability Across Scales.” *Functional Ecology* 30 (12): 1926–35. <https://doi.org/10.1111/1365-2435.12710>.
- 546
- 547 Turak, Eren, James Brazill-Boast, Tim Cooney, Michael Drielsma, Jocelyn DelaCruz, Gillian
548 Dunkerley, Miguel Fernandez, et al. 2017. “Using the Essential Biodiversity Variables
549 Framework to Measure Biodiversity Change at National Scale.” *Biological Conservation*,

- 550 SI:Measures of biodiversity, 213: 264–71. <https://doi.org/10.1016/j.biocon.2016.08.019>.
- 552 Upham, Nathan S., Jacob A. EsSELSTYN, and Walter Jetz. 2019. “Inferring the Mammal Tree:
553 Species-Level Sets of Phylogenies for Questions in Ecology, Evolution, and Conservation.”
554 *PLOS Biology* 17 (12): e3000494. <https://doi.org/10.1371/journal.pbio.3000494>.
- 555 Yeakel, Justin D., Mathias M. Pires, Lars Rudolf, Nathaniel J. Dominy, Paul L. Koch, Paulo R.
556 Guimarães, and Thilo Gross. 2014. “Collapse of an Ecological Network in Ancient Egypt.”
557 *PNAS* 111 (40): 14472–77. <https://doi.org/10.1073/pnas.1408471111>.
- 558 Youden, W. J. 1950. “Index for Rating Diagnostic Tests.” *Cancer* 3 (1): 32–35. [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).
- 560 Young, Jean-Gabriel, George T Cantwell, and M E J Newman. 2021. “Bayesian Inference of
561 Network Structure from Unreliable Data.” *Journal of Complex Networks* 8 (6). <https://doi.org/10.1093/comnet/cnaa046>.
- 563 Young, Stephen J., and Edward R. Scheinerman. 2007. “Random Dot Product Graph Models for
564 Social Networks.” In *Algorithms and Models for the Web-Graph*, edited by Anthony Bonato
565 and Fan R. K. Chung, 138–49. Lecture Notes in Computer Science. Berlin, Heidelberg:
566 Springer. https://doi.org/10.1007/978-3-540-77004-6_11.
- 567 Zhu, Mu, and Ali Ghodsi. 2006. “Automatic Dimensionality Selection from the Scree Plot via
568 the Use of Profile Likelihood.” *Computational Statistics & Data Analysis* 51 (2): 918–30.
569 <https://doi.org/10.1016/j.csda.2005.09.010>.

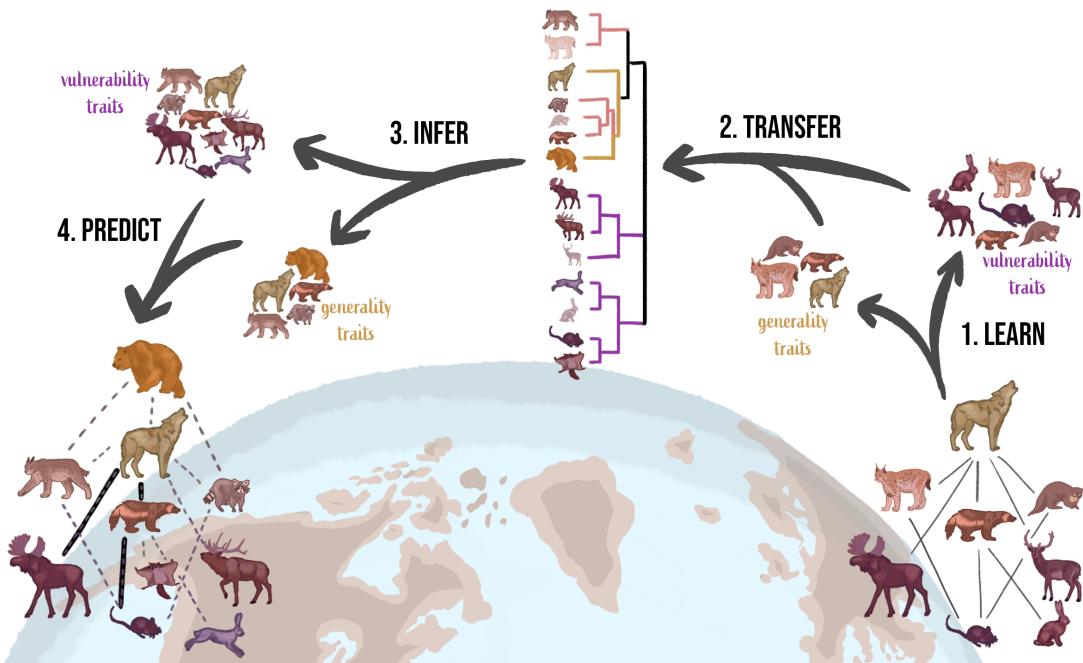


Figure 1: Overview of the phylogenetic transfer learning (and prediction) of species interactions networks. Starting from an initial, known, network, we learn its representation through a graph embedding step (here, a truncated Singular Value Decomposition; Step 1), yielding a series of latent traits (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).

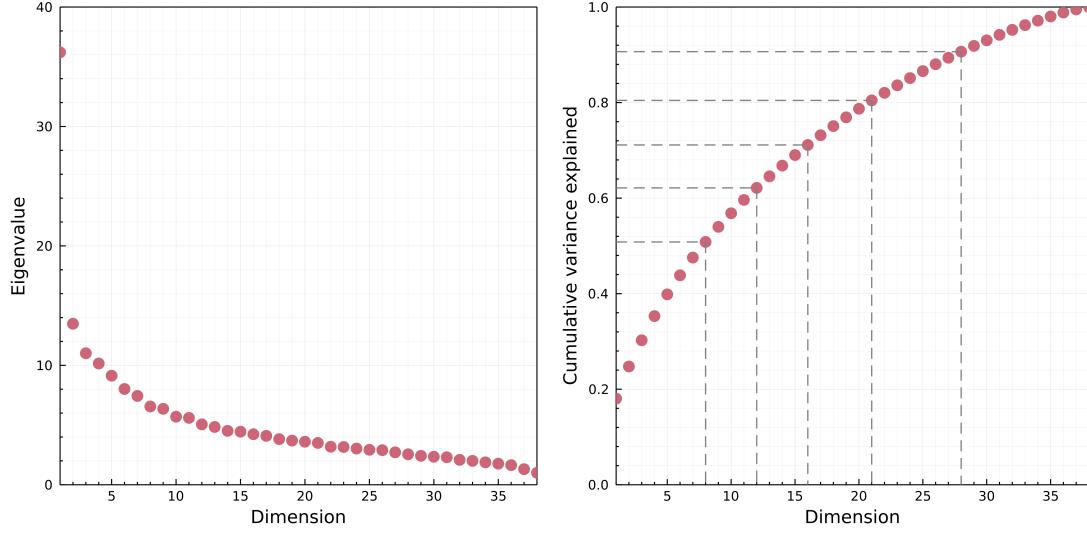


Figure 2: Left: representation of the screeplot of the singular values from the t-SVD on the European metaweb. The screeplot shows no obvious drop in the singular values that may be leveraged to automatically detect a minimal dimension for embedding, after *e.g.* Zhu and Ghodsi (2006). Right: cumulative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

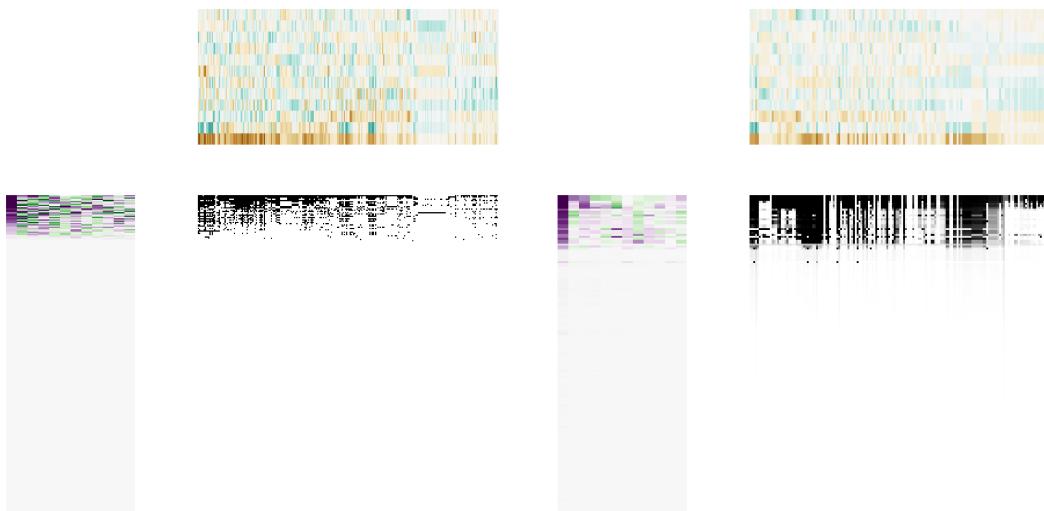


Figure 3: Visual representation of the left (green/purple) and right (green/brown) subspaces, alongside the adjacency matrix of the food web they encode (greyscale). The European metaweb is on the left, and the imputed Canadian metaweb (before data inflation) on the right. This figure illustrates how 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.

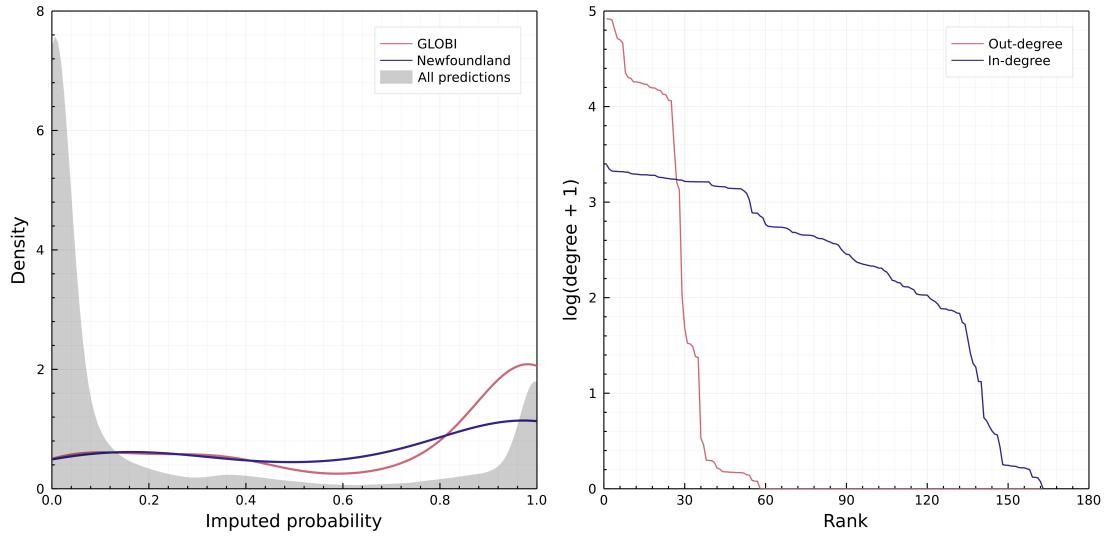


Figure 4: Left, comparison of the probabilities of interactions assigned by the model to all interactions (grey curve), the subset of interactions found in GLOBI (red), and in the Strong and Leroux (2014) Newfoundland dataset (blue). The model recovers more interaction with a low probability compared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

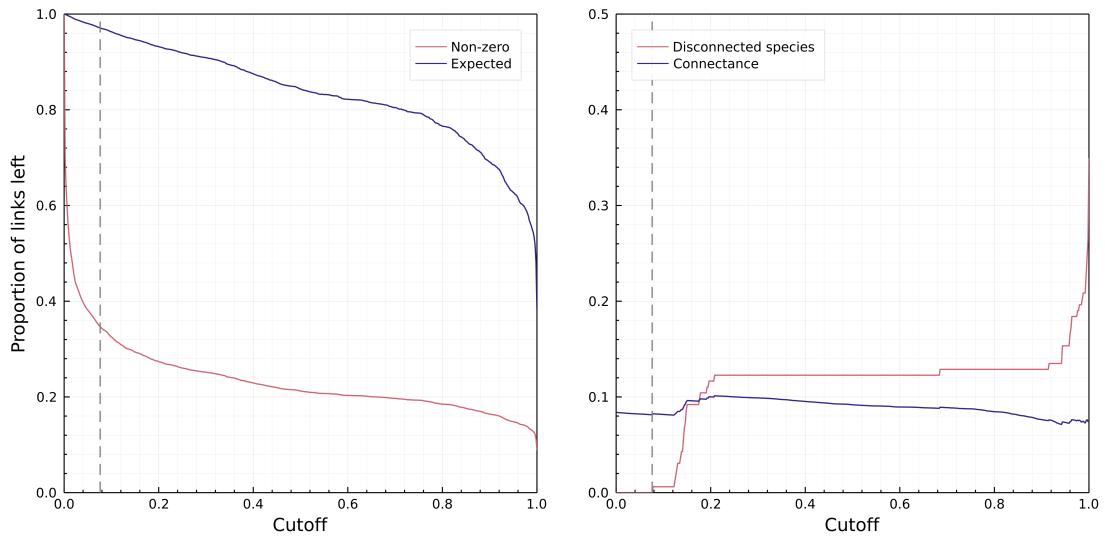


Figure 5: Left: effect of varying the cutoff for probabilities to be considered non-zero on the number of unique links and on \hat{L} , the probabilistic estimate of the number of links assuming that all interactions are independent. Right: effect of varying the cutoff on the number of disconnected species,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

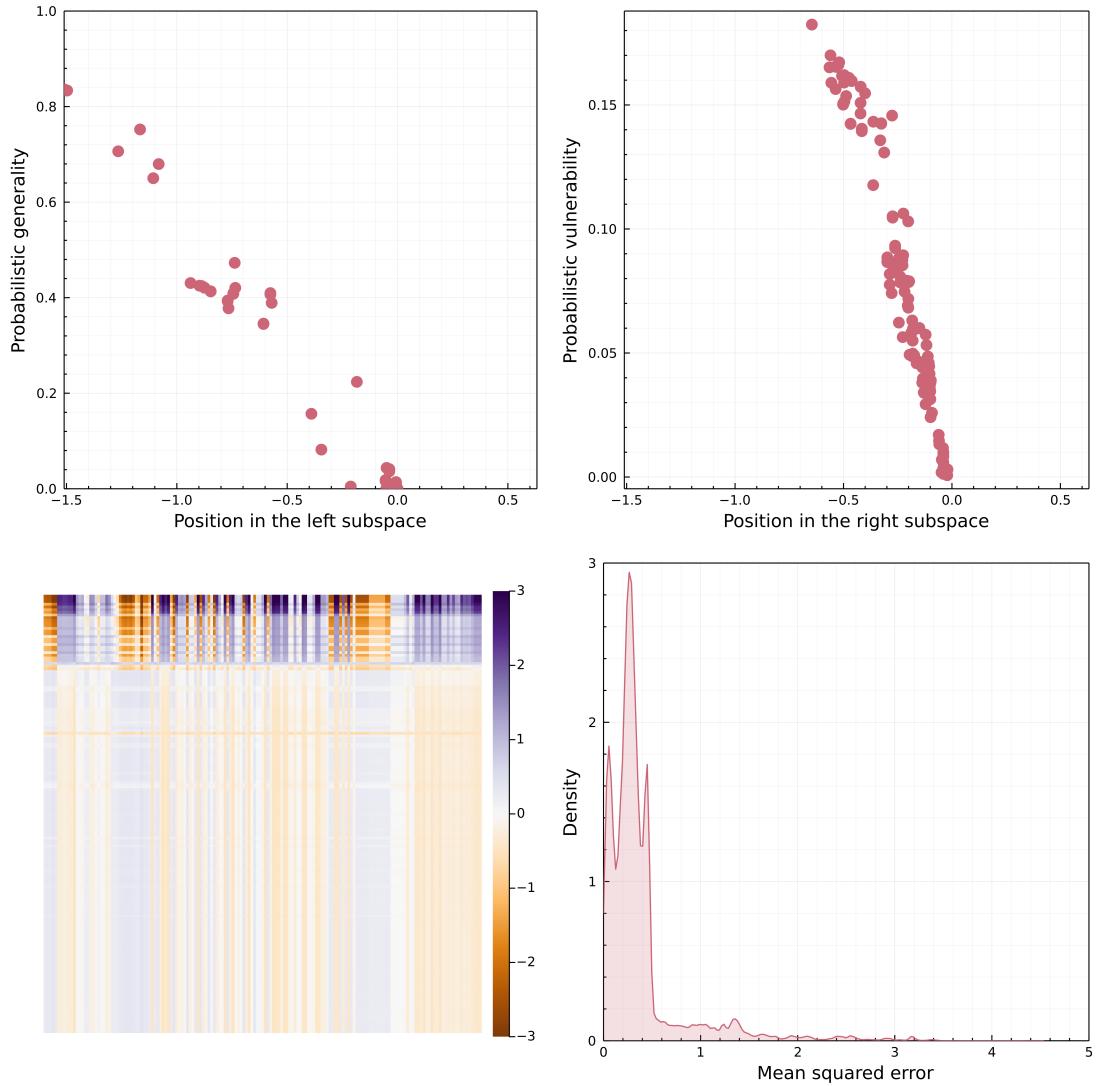


Figure 6: Top: biological significance of the first dimension. Left: there is a linear relationship between the values on the first dimension of the left subspace and the generality, *i.e.* the relative number of preys, *sensu* Schoener (1989). Species with a value of 0 in this subspace are at the bottom-most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