

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

Tanya Strydom^{1,2,‡} Salomé Bouskila^{1,2,‡} Timothée Poisot^{1,2}

¹ Université de Montréal ² Québec Centre for Biodiversity Sciences

[‡] These authors contributed equally to the work

Correspondance to:

Timothée Poisot — timothee.poisot@umontreal.ca

Abstract: TODO

1 Introduction

2 There are two core challenges to our understanding of ecological networks in space, particu-
3 larly at macro-ecologically relevant scales (*e.g.* Trøjelsgaard and Olesen 2016). First, networks
4 within a location are difficult to sample properly (Jordano 2016a, 2016b), and this results in a
5 widespread “Eltonian shortfall” (Hortal et al. 2015). This first challenge (local incompleteness)
6 has been in large part addressed by the recent multiplication of methods aiming to predict inter-
7 actions within an existing network, a lot of which are reviewed in Strydom et al. (2021). Second,
8 recent analyses based on collected data (Poisot, Bergeron, et al. 2021) or metadata (Cameron et
9 al. 2019) highlight that ecological networks are currently studied in a biased subset of space and
10 bioclimates, which impedes our ability to generalize any local understanding of network struc-
11 ture. Having a general solution for the issue of metaweb inference (Morales-Castilla et al. 2015)
12 even in situations where minimal knowledge about existing interactions within the species pool
13 is known would catalyze significant breakthroughs in our ability to start thinking about species
14 interactions networks over large spatial scales.

15 Here, we present a general method for the transfer learning of network representations, relying
16 on similarities of species in a biologically relevant proxy space. This method allows to the
17 prediction of species interactions based on knowledge extracted from another species pool. Our
18 method is outlined in fig. 1; we provide an illustration based on learning an embedding of a
19 metaweb of trophic interactions for European mammals (Maiorano et al. 2020b, 2020a), and
20 based on phylogenetic relationships between mammals globally (Upham, Esselstyn, and Jetz
21 2019), transfer this representation to the pool of mammals in Canada.

22 [Figure 1 about here.]

23 There are a plurality of measures of species similarities that can be used for knowledge transfer
24 (see *e.g.* Morales-Castilla et al. 2015); nevertheless, phylogenetic proximity has a number of
25 desirable properties when working at large scales. Gerhold et al. (2015) make the point that
26 phylogenetic signal captures diversification of characters (large macro-evolutionary process),
27 but not necessarily community assembly (fine ecological process); Dormann et al. (2010) pre-
28 viously gave very similar conclusions. Interactions tend to conserve phylogenetic signal that

29 encompasses a wide range of ecological and evolutionary mechanisms (Mouquet et al. 2012;
30 Cavender-Bares et al. 2009), and most importantly, retain this signal even when it is not de-
31 tectable at the community scale (Poisot and Stouffer 2018; Hutchinson, Cagua, and Stouffer
32 2017). Finally, species interactions at macro-ecological scales seem to respond mostly to macro-
33 evolutionary processes (Price 2003); this is evidenced by the presence of conserved backbones in
34 food webs (Dalla Riva and Stouffer 2016), strong evolutionary signature on prey choice (Stouf-
35 fer et al. 2012), and strong phylogenetic signature in food web intervalty (Eklöf and Stouffer
36 2016). Taken together, these consideration suggest that phylogenies can be reliably be used to
37 transfer knowledge on species interactions.

38 **TK** Results/summary type paragraph

39 **Data used for the case study**

40 We use data on the European metaweb assembled by Maiorano et al. (2020b). We use the
41 definition of the metaweb first introduced by Dunne (2006), *i.e.* an inventory of all possible
42 interactions within species from a spatially delimited pool. The metaweb is not a prediction
43 of the food web at any specific locale within the frontiers of the species pool – in fact, these
44 local food webs are expected to have a subset of both the species and the interactions of their
45 metaweb (Poisot et al. 2012). This being said, as the metaweb represents the total of functional,
46 phylogenetic, and macroecological processes (Morales-Castilla et al. 2015), it is still worthy
47 of ecological attention. We induced the subgraph corresponding to all mammals by matching
48 species names in the original network first to the GBIF taxonomic backbone (GBIF Secretariat
49 2021), and retaining all those who matched to mammals; all nodes had valid matches to GBIF
50 at this step, and so this backbone is used for all name reconciliation steps as outlined below.

51 The European metaweb represents the knowledge we want to learn and transfer; the support for
52 transfer is here represented by the phylogenetic similarity of mammals. We used the mammalian
53 consensus supertree by Upham, Esselstyn, and Jetz (2019), for which all approx. 6000 names
54 have been similarly matched to their GBIF valid names. This step allows us to place each node
55 of the mammalian European metaweb in the phylogeny.

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

61 The European metaweb had XXX species, and the Canadian species pool had YYY - of these,
62 only 17 were shared, which highlights the need to rely on a knowledge transfer approach. In the
63 following sections, we describe the representational learning step applied to European data, the
64 transfer step through phylogenetic similarity, and the generation of a probabilistic metaweb for
65 the destination species pool.

66 Method description

67 The crux of the method is the transfer of knowledge of a known network, in order to predict
68 interactions between species from another location. In fig. 1, we give a high-level overview of
69 the approach; in the example around which this manuscript is built (leveraging detailed knowl-
70 edge about trophic interactions between Mammalia in Europe to predict the less known trophic
71 interactions between closely phylogenetically related Mammalia in Canada), we use a series of
72 specific steps for network embedding, trait inference, and network prediction and thresholding.
73 This being said, the methods at each step may be substituted by others when the data support it.

74 Specifically, our approach can be summarized as follows: from the known network in Europe,
75 we use a truncated Singular Values Decomposition (t-SVD) to generate latent traits representing
76 a low-dimensional embedding of the network; these traits give an unbiased estimate of the node's
77 position in the latent feature spaces. Then, we map these latent traits onto a reference phylogeny
78 (other distance-based measures of species proximity that allow for the inference of features in
79 the latent space can be used, such as for example dissimilarity in functional traits). Based on
80 the reconstructed latent traits for species in the destination species pool, a Random Dot Product
81 Graph (RDPG) model predicts the interaction between species through a function of the nodes'
82 features. Thus, from latent traits and nodes position, we can infer interactions.

83 **Learning the origin network representation**

84 The first step in transfer learning is to learn the structure of the origin dataset. In order to do so,
85 we rely on an approach inspired from representational learning, where we learn a *representation*
86 of the metaweb, rather than a list of interactions. This approach is conceptually different from
87 other metaweb-scale predictions (*e.g.* ALBOUY, others?), in that the metaweb representation
88 is easily transferable. Specifically, we use Random Dot Product Graphs (RDPG; Young and
89 Scheinerman 2007) to create a number of latent variables that can be combined into an approx-
90 imation of the network adjacency matrix. RDPG results are known to have strong phylogenetic
91 signal, and to capture the evolutionary backbone of food webs (Dalla Riva and Stouffer 2016).

92 In addition, recent advances show that the latent variables produced this way can be used to
93 predict *de novo* network edges (Runghen, Stouffer, and Dalla Riva 2021).

94 The latent variables are created by performing a truncated Singular Value Decomposition (tSVD)
95 on the adjacency matrix. SVD is an appropriate embedding of ecological networks, which has
96 recently been shown to both capture their complex, emerging properties (Strydom, Dalla Riva,
97 and Poisot 2021), and to allow highly accurate prediction of the interactions within a single
98 network (Poisot, Ouellet, et al. 2021). Under tSVD, an adjacency matrix \mathbf{A} is decomposed into
99 three components, so that $\mathbf{A} = \mathbf{L}\Sigma\mathbf{R}^*$, for which Σ is an $r \times r$ diagonal matrix (where r is the
100 rank of matrix \mathbf{A}) containing only non-zero singular (σ) values, \mathbf{L} is an $m \times r$ matrix, and \mathbf{R} an
101 $n \times r$ matrix. In using a tSVD we retain only ‘informative’ (non-zero) σ values - which translates
102 to unique predation strategies within the matrix.

103 The latent variables used for the RDPG, called the left and right subspaces, are defined as $\mathcal{L} =$
104 $\mathbf{L}\sqrt{\Sigma}$, and $\mathcal{R} = \sqrt{\Sigma}\mathbf{R}$ – using the full rank of \mathbf{A} , $\mathcal{L}\mathcal{R}' = \mathbf{A}$, and using any smaller rank results in
105 $\mathcal{L}\mathcal{R}' \approx \mathbf{A}$. Using a rank of 1 for the tSVD provides a first-order approximation of the network,
106 etc.

107 The specific rank at which the SVD ought to be truncated is a difficult question. The purpose of
108 SVD is to remove the noise, expressed at high dimensions, and to focus on the signal, expressed
109 at low dimension. In datasets with a clear signal/noise demarcation, the scree plot of Σ can
110 show a sharp drop at the rank where noise starts. Alternatively, REF showed that the profile

111 likelihood (assuming a known distribution of Σ , or a normal distribution otherwise) peaks at the
112 rank maximizing the signal-to-noise ration. Because the European metaweb is almost entirely
113 known, the amount of noise is low; this is reflected in fig. 2 (left), where the scree plot shows
114 no important drop, and in fig. 2 (right) where the proportion of variance explained increases
115 smoothly at higher dimensions. For this reason, we default back to an arbitrary threshold that
116 explain 60% of the variance in the underlying data, corresponding to 12 dimensions.

117 [Figure 2 about here.]

118 Because RDPG relies on matrix multiplication, the higher dimensions essentially serve to make
119 specific interactions converge towards 0 or 1; therefore, for reasonably low ranks, there is no
120 guarantee that the values in the reconstructed network will be within the unit range. In order
121 to determine what constitutes an appropriate threshold, we performed the RDPG approach on
122 the European metaweb, and evaluated the threshold by treating this as a binary classification
123 problem, specifically assuming that both 0 and 1 in the European metaweb are all true. Given
124 the methodological given in REF, this seems like a reasonable assumption, although one that
125 does not hold for all metawebs. We used the thresholding approach presented in HAYSTACK,
126 and picked a cutoff that maximized Youden's J statistic (Youden 1950); the resulting cutoff was
127 0.22, and gave an accuracy above 0.99.

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

131 Transfer learning through phylogenetic relatedness

132 In order to transfer the knowledge from the European metaweb to the Canadian species pool,
133 we performed ancestral character estimation using a Brownian motion model, which is a con-
134 servative approach in the absence of strong hypotheses about the nature of phylogenetic signal
135 in the network decomposition (Litsios and Salamin 2012). We assumed that all traits (*i.e.* the
136 feature vectors for the left and right subspaces) were independent, which is a reasonable as-
137 sumption as every trait/dimension added to the tSVD has an *additive* effect to the one before it.

138 The Brownian motion algorithm returns the *average* value of the trait, and its upper and lower
139 bounds. Because we do not estimate other parameters of the traits distributions, we considered
140 that every species trait is represented as a uniform distribution between these bounds. Therefore,
141 the inferred left and right sub-spaces for the Canadian species pool ($\hat{\mathcal{L}}$ and $\hat{\mathcal{R}}$) have entries that
142 are distributions, representing the range of values for a given species at a given dimension.

143 These objects represent the transferred knowledge, which we can use for prediction of the Cana-
144 dian metaweb.

145 Probabilistic prediction of the destination network

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

150 [Figure 3 about here.]

151 Specifically, we have adopted the following approach. For every entry in $\hat{\mathcal{L}}$ and $\hat{\mathcal{R}}$, we draw
152 a value from its distribution. This results in one instance of the possible left () and right ()
153 subspaces for the Canadian metaweb. These can be multiplied, to produce one matrix of real
154 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
155 predicted, it follows that the threshold ρ estimated for the European metaweb also applies. We
156 use this information to produce one random Canadian metaweb, $N = \hat{\mathcal{L}}\hat{\mathcal{R}}' \geq \rho$ fig. 3.

157 Because the intervals around some trait values can be broad (in fact, probably broader than what
158 they would actually be, see *e.g.* Garland, Midford, and Ives 1999), we repeat the above process
159 2×10^5 times, which results in a probabilistic metaweb P , where the probability of an interaction
160 (here conveying our degree of trust that it exists given the inferred trait distributions) is given
161 by the number of times where it appears across all random draws N , divided by the number of
162 samples. An interaction with $P_{i,j} = 1$ means that these two species were predicted to interact in
163 all 2×10^5 random draws, etc..

¹⁶⁴ **Data cleanup, discovery, validation, and thresholding**

¹⁶⁵ Once the probabilistic metaweb for Canada has been produced, we followed a number of data
¹⁶⁶ inflation steps to finalize it.

¹⁶⁷ [Figure 4 about here.]

¹⁶⁸ First, we extracted the subgraph corresponding to the 17 species shared between the European
¹⁶⁹ and Canadian pools, and replaced these interactions with a probability of 0 (non-interaction)
¹⁷⁰ or 1 (interaction). This represents a minute modification of the inferred network (about 0.8% of
¹⁷¹ all species pairs from the Canadian web), but ensures that are directly re-using knowledge from
¹⁷² Europe.

¹⁷³ Second, we looked for all species in the Canadian pool known to the Global Biotic Interactions
¹⁷⁴ (GLOBI) database **REF**, and extracted their known interactions. Because GLOBI aggregates
¹⁷⁵ observed interactions, it is not a *networks* data source, and therefore the only information we
¹⁷⁶ can reliably extract from it is that a species pair *was reported to interact at least once*. This last
¹⁷⁷ statement should yet be taken with caution, as some sources in GLOBI (*e.g.* **TK**) are produced
¹⁷⁸ through text-mining, and therefore do not document direct evidence of the interaction. Neverthe-
¹⁷⁹ less, should the predictive model work, we would expect that a majority of interactions known
¹⁸⁰ to GLOBI would also be predicted. After performing this check, we set the probability of all
¹⁸¹ interactions known to GLOBI (366 in total, 33 of which were not predicted by the model, for a
¹⁸² success rate of 91%) to 1.

¹⁸³ Finally, we downloaded the data from Strong and Leroux (2014), who mined various literature
¹⁸⁴ sources to identify trophic interactions in Newfoundland. This dataset had documented 25 in-
¹⁸⁵ teractions between mammals, only two of which were not part of our predictions, for a success
¹⁸⁶ rate of 92%. These two interactions were added to our predicted metaweb, with a probability of
¹⁸⁷ 1.

¹⁸⁸ [Figure 5 about here.]

¹⁸⁹ Because the confidence interval on the inferred trait space are probably an over-estimate, we
¹⁹⁰ decided to apply a thresholding step to the interactions after the data inflation fig. 5. Cirtwill

191 and Hambäck (2021) highlight a number of strategies to threshold probabilistic networks. Their
192 methods assume the underlying data to be tag-based sequencing, which represents interactions as
193 co-occurrences of predator and prey within the same tags; this is conceptually identical to our
194 Bernoulli-trial based reconstruction of a probabilistic network. We performed a full analysis
195 of the effect of various cutoffs, and as they either resulted in removing to few interactions, or
196 removing enough interactions that species started to be disconnected from the network, we set
197 the threshold for a probability equivalent to 0 to the largest possible value that still allow all
198 species to have at least one interaction with non-zero probability.

199 **Implementation and code availability**

200 The entire pipeline was implemented in *Julia* 1.6 (Bezanson et al. 2017), and is available un-
201 der the permissive MIT License at DOI. The taxonomic cleanup steps are done using GBIF.jl
202 (Dansereau and Poisot 2021). The network embedding and analysis is done using EcologicalNetworks.jl
203 (Banville, Vissault, and Poisot 2021; Poisot et al. 2019). The phylogenetic simulations are
204 done using PhyloNetworks.jl (Solís-Lemus, Bastide, and Ané 2017) and Phylo.jl (Reeve et
205 al. 2016). A complete Project.toml file specifying the full tree of dependencies is available
206 alongside the code, at <https://osf.io/2zwqm/>.

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

208 **TP** fig. 5 – the cutoffs using maximum curvature of central difference approximation of the
209 second order partial derivative result in respectively species being lost, or almost all links being
210 kept – we settled on the value that allowed all species to remained.

211 [Figure 6 about here.]

212 The t-SVD embedding is able to learn relevant ecological features for the network. fig. 6 shows
213 that the first rank correlates linearly with generality and vulnerability (Schoener 1989), *i.e.* the
214 number of preys and predators. Importantly, this implies that a rank 1 approximation repre-
215 sents the configuration model (Park and Newman 2004) for the metaweb, or (accounting for the

216 probabilistic nature of the degree) the soft configuration model (van der Hoorn, Lippner, and
217 Krioukov 2018), both of which are maximum entropy graph models (Garlaschelli, Hollander,
218 and Roccaverde 2018), with respectively sharp and soft local constraints. The configuration
219 model is widely used by ecologists in the context of null hypothesis significance testing of net-
220 work structure (*e.g.* Bascompte et al. 2003), and it is noteworthy that for this metaweb, the rele-
221 vant information was extracted at the first rank. Because the first rank corresponds to the leading
222 eigenvalue of the system, the results of fig. 6 have a straightforward interpretation: degree-based
223 processes are the most important in structuring the mammalian food web.

224 Conclusions

225 A cautionary tale regarding validation. Interaction can never be a true negative. Cannot use a
226 empirical subset of a bigger (predicated) network for validation -> because *e.g.* dietary shifts
227 across range.
228 Cirtwill et al. (2019) – we need an informative interaction-level prior on probability: this method
229 might be it

230 References

- 231 Banville, Francis, Steve Vissault, and Timothée Poisot. 2021. “Mangal.jl and EcologicalNet-
232 works.jl: Two Complementary Packages for Analyzing Ecological Networks in Julia.” *Jour-*
233 *nal of Open Source Software* 6 (61): 2721. <https://doi.org/10.21105/joss.02721>.
- 234 Bascompte, J., P. Jordano, C. J. Melian, and J. M. Olesen. 2003. “The Nested Assembly of Plant-
235 Animal Mutualistic Networks.” *Proceedings of the National Academy of Sciences* 100 (16):
236 9383–87. <https://doi.org/10.1073/pnas.1633576100>.
- 237 Bezanson, J., A. Edelman, S. Karpinski, and V. Shah. 2017. “Julia: A Fresh Approach to Numer-
238 ical Computing.” *SIAM Review* 59 (1): 65–98. <https://doi.org/10.1137/141000671>.
- 239 Cameron, Erin K., Maja K. Sundqvist, Sally A. Keith, Paul J. CaraDonna, Erik A. Mousing,
240 Karin A. Nilsson, Daniel B. Metcalfe, and Aimée T. Classen. 2019. “Uneven Global

- 241 Distribution of Food Web Studies Under Climate Change.” *Ecosphere* 10 (3): e02645.
- 242 <https://doi.org/10.1002/ecs2.2645>.
- 243 Cavender-Bares, Jeannine, Kenneth H. Kozak, Paul V. A. Fine, and Steven W. Kembel. 2009.
- 244 “The Merging of Community Ecology and Phylogenetic Biology.” *Ecology Letters* 12 (7):
- 245 693–715. <https://doi.org/10.1111/j.1461-0248.2009.01314.x>.
- 246 Cirtwill, Alyssa R., Anna Eklf, Tomas Roslin, Kate Wootton, and Dominique Gravel. 2019. “A
- 247 Quantitative Framework for Investigating the Reliability of Empirical Network Construc-
- 248 tion.” *Methods in Ecology and Evolution* 0 (ja). <https://doi.org/10.1111/2041-210X.13180>.
- 249
- 250 Cirtwill, Alyssa R., and Peter Hambäck. 2021. “Building Food Networks from Molecular Data:
- 251 Bayesian or Fixed-Number Thresholds for Including Links.” *Basic and Applied Ecology* 50
- 252 (February): 67–76. <https://doi.org/10.1016/j.baae.2020.11.007>.
- 253 Dalla Riva, Giulio V, and Daniel B. Stouffer. 2016. “Exploring the Evolutionary Signature of
- 254 Food Webs’ Backbones Using Functional Traits.” *Oikos* 125 (4): 446–56. <https://doi.org/10.1111/oik.02305>.
- 255
- 256 Dansereau, Gabriel, and Timothée Poisot. 2021. “SimpleSDMLayers.jl and GBIF.jl: A Frame-
- 257 work for Species Distribution Modeling in Julia.” *Journal of Open Source Software* 6 (57):
- 258 2872. <https://doi.org/10.21105/joss.02872>.
- 259
- 260 Dormann, Carsten F., Bernd Gruber, Marten Winter, and Dirk Herrmann. 2010. “Evolution
- 261 of Climate Niches in European Mammals?” *Biology Letters* 6 (2): 229–32. <https://doi.org/10.1098/rsbl.2009.0688>.
- 262
- 263 Dunne, Jennifer A. 2006. “The Network Structure of Food Webs.” In *Ecological Networks:*
- 264 *Linking Structure and Dynamics*, edited by Jennifer A Dunne and Mercedes Pascual, 27–86.
- 265 Oxford University Press.
- 266
- 267 Eklöf, Anna, and Daniel B. Stouffer. 2016. “The Phylogenetic Component of Food Web Struc-
- ture and Intervallity.” *Theoretical Ecology* 9 (1): 107–15. <https://doi.org/10.1007/s12080-015-0273-9>.

- 268 Garland, THEODORE, JR., PETER E. Midford, and ANTHONY R. Ives. 1999. “An In-
269 troduction to Phylogenetically Based Statistical Methods, with a New Method for Con-
270 fidence Intervals on Ancestral Values1.” *American Zoologist* 39 (2): 374–88. <https://doi.org/10.1093/icb/39.2.374>.
- 272 Garlaschelli, Diego, Frank den Hollander, and Andrea Roccaverde. 2018. “Covariance Struc-
273 ture Behind Breaking of Ensemble Equivalence in Random Graphs.” *Journal of Statistical*
274 *Physics* 173 (3-4): 644–62. <https://doi.org/10.1007/s10955-018-2114-x>.
- 275 GBIF Secretariat. 2021. “GBIF Backbone Taxonomy.” <https://doi.org/10.15468/39omei>.
- 276 Gerhold, Pille, James F. Cahill, Marten Winter, Igor V. Bartish, and Andreas Prinzing. 2015.
277 “Phylogenetic Patterns Are Not Proxies of Community Assembly Mechanisms (they Are Far
278 Better).” *Functional Ecology* 29 (5): 600–614. <https://doi.org/10.1111/1365-2435.12425>.
- 280 Hoorn, Pim van der, Gabor Lippner, and Dmitri Krioukov. 2018. “Sparse Maximum-Entropy
281 Random Graphs with a Given Power-Law Degree Distribution.” *Journal of Statistical Physics*
282 173 (3-4): 806–44. <https://doi.org/10.1007/s10955-017-1887-7>.
- 283 Hortal, Joaquín, Francesco de Bello, José Alexandre F. Diniz-Filho, Thomas M. Lewinsohn,
284 Jorge M. Lobo, and Richard J. Ladle. 2015. “Seven Shortfalls That Beset Large-Scale
285 Knowledge of Biodiversity.” *Annual Review of Ecology, Evolution, and Systematics* 46 (1):
286 523–49. <https://doi.org/10.1146/annurev-ecolsys-112414-054400>.
- 287 Hutchinson, Matthew C., Edgar Fernando Cagua, and Daniel B. Stouffer. 2017. “Cophyloge-
288 netic Signal Is Detectable in Pollination Interactions Across Ecological Scales.” *Ecology*,
289 n/a–. <https://doi.org/10.1002/ecy.1955>.
- 290 Jordano, Pedro. 2016a. “Chasing Ecological Interactions.” *PLOS Biology* 14 (9): e1002559.
291 <https://doi.org/10.1371/journal.pbio.1002559>.
- 292 ———. 2016b. “Sampling Networks of Ecological Interactions.” *Functional Ecology*, Septem-
293 ber. <https://doi.org/10.1111/1365-2435.12763>.
- 294 Litsios, Glenn, and Nicolas Salamin. 2012. “Effects of Phylogenetic Signal on Ancestral State
295 Reconstruction.” *Systematic Biology* 61 (3): 533–38. <https://doi.org/10.1093/sysbio/>

296 syr124.

297 Maiorano, Luigi, Alessandro Montemaggiori, Gentile Francesco Ficetola, Louise O'Connor,
298 and Wilfried Thuiller. 2020a. "Data from: Tetra-EU 1.0: A Species-Level Trophic Meta-
299 Web of European Tetrapods." Dryad. <https://doi.org/10.5061/DRYAD.JM63XSJ7B>.

300 ———. 2020b. "TETRA-EU 1.0: A Species-Level Trophic Metaweb of European Tetrapods."
301 Edited by Allen Hurlbert. *Global Ecology and Biogeography* 29 (9): 1452–57. <https://doi.org/10.1111/geb.13138>.

303 Morales-Castilla, Ignacio, Miguel G. Matias, Dominique Gravel, and Miguel B. Araújo. 2015.
304 "Inferring Biotic Interactions from Proxies." *Trends in Ecology & Evolution* 30 (6): 347–56.
305 <https://doi.org/10.1016/j.tree.2015.03.014>.

306 Mouquet, Nicolas, Vincent Devictor, Christine N. Meynard, Francois Munoz, Louis-Félix Bersier,
307 Jérôme Chave, Pierre Couteron, et al. 2012. "Ecophylogenetics: Advances and Perspec-
308 tives." *Biological Reviews* 87 (4): 769–85. <https://doi.org/10.1111/j.1469-185X.2012.00224.x>.

310 Park, Juyong, and M. E. J. Newman. 2004. "Statistical Mechanics of Networks." *Physical
311 Review E* 70 (6): 066117. <https://doi.org/10.1103/PhysRevE.70.066117>.

312 Poisot, Timothée, Zacharie Belisle, Laura Hoebelke, Michiel Stock, and Piotr Szefer. 2019.
313 "EcologicalNetworks.jl - Analysing Ecological Networks." *Ecography*. <https://doi.org/10.1111/ecog.04310>.

315 Poisot, Timothée, Gabriel Bergeron, Kevin Cazelles, Tad Dallas, Dominique Gravel, Andrew
316 MacDonald, Benjamin Mercier, Clément Violet, and Steve Vissault. 2021. "Global Knowl-
317 edge Gaps in Species Interaction Networks Data." *Journal of Biogeography* n/a (n/a). <https://doi.org/10.1111/jbi.14127>.

319 Poisot, Timothée, Elsa Canard, David Mouillot, Nicolas Mouquet, and Dominique Gravel. 2012.
320 "The Dissimilarity of Species Interaction Networks." *Ecology Letters* 15 (12): 1353–61.
321 <https://doi.org/10.1111/ele.12002>.

322 Poisot, Timothée, Alyssa R. Cirtwill, Kévin Cazelles, Dominique Gravel, Marie-Josée Fortin,
323 and Daniel B. Stouffer. 2016. "The Structure of Probabilistic Networks." Edited by Jana

- 324 Vamosi. *Methods in Ecology and Evolution* 7 (3): 303–12. <https://doi.org/10.1111/2041-210X.12468>.
- 326 Poisot, Timothée, Marie-Andrée Ouellet, Nardus Mollentze, Maxwell J. Farrell, Daniel J. Becker,
327 Gregory F. Albery, Rory J. Gibb, Stephanie N. Seifert, and Colin J. Carlson. 2021. “Im-
328 putting the Mammalian Virome with Linear Filtering and Singular Value Decomposition.”
329 *arXiv:2105.14973 [q-Bio]*, May. <http://arxiv.org/abs/2105.14973>.
- 330 Poisot, Timothée, and Daniel B. Stouffer. 2018. “Interactions Retain the Co-Phylogenetic
331 Matching That Communities Lost.” *Oikos* 127 (2): 230–38. <https://doi.org/10.1111/oik.03788>.
- 333 Price, Peter W. 2003. *Macroevolutionary Theory on Macroecological Patterns*. Cambridge
334 University Press.
- 335 Reeve, Richard, Tom Leinster, Christina A. Cobbold, Jill Thompson, Neil Brummitt, Sonia N.
336 Mitchell, and Louise Matthews. 2016. “How to Partition Diversity.” *arXiv:1404.6520 [q-
337 Bio]*, December. <http://arxiv.org/abs/1404.6520>.
- 338 Runghen, Rogini, Daniel B Stouffer, and Giulio V Dalla Riva. 2021. “Exploiting Node Metadata
339 to Predict Interactions in Large Networks Using Graph Embedding and Neural Networks,”
340 June. <https://doi.org/10.1101/2021.06.10.447991>.
- 341 Schoener, Thomas W. 1989. “Food Webs from the Small to the Large.” *Ecology* 70 (6): 1559–
342 89.
- 343 Solís-Lemus, Claudia, Paul Bastide, and Cécile Ané. 2017. “PhyloNetworks: A Package for
344 Phylogenetic Networks.” *Molecular Biology and Evolution* 34 (12): 3292–98. <https://doi.org/10.1093/molbev/msx235>.
- 346 Stouffer, D. B., M. Sales-Pardo, M. I. Sirer, and J. Bascompte. 2012. “Evolutionary Conserva-
347 tion of Species’ Roles in Food Webs.” *Science* 335 (6075): 1489–92. <https://doi.org/10.1126/science.1216556>.
- 349 Strong, Justin S., and Shawn J. Leroux. 2014. “Impact of Non-Native Terrestrial Mammals on
350 the Structure of the Terrestrial Mammal Food Web of Newfoundland, Canada.” *PLOS ONE*
351 9 (8): e106264. <https://doi.org/10.1371/journal.pone.0106264>.

- 352 Strydom, Tanya, Michael David Catchen, Francis Banville, Dominique Caron, Gabriel Dansereau,
353 Philippe Desjardins-Proulx, Norma Forero, et al. 2021. “A Roadmap Toward Predicting
354 Species Interaction Networks (Across Space and Time).” Preprint. EcoEvoRxiv. <https://doi.org/10.32942/osf.io/eu7k3>.
- 356 Strydom, Tanya, Giulio V. Dalla Riva, and Timothée Poisot. 2021. “SVD Entropy Reveals the
357 High Complexity of Ecological Networks.” *Frontiers in Ecology and Evolution* 9. <https://doi.org/10.3389/fevo.2021.623141>.
- 359 Trøjelsgaard, Kristian, and Jens M. Olesen. 2016. “Ecological Networks in Motion: Micro- and
360 Macroscopic Variability Across Scales.” *Functional Ecology* 30 (12): 1926–35. <https://doi.org/10.1111/1365-2435.12710>.
- 362 Upham, Nathan S., Jacob A. Esselstyn, and Walter Jetz. 2019. “Inferring the Mammal Tree:
363 Species-Level Sets of Phylogenies for Questions in Ecology, Evolution, and Conservation.”
364 *PLOS Biology* 17 (12): e3000494. <https://doi.org/10.1371/journal.pbio.3000494>.
- 366 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).
- 367 Young, Stephen J., and Edward R. Scheinerman. 2007. “Random Dot Product Graph Models for
368 Social Networks.” In *Algorithms and Models for the Web-Graph*, edited by Anthony Bonato
369 and Fan R. K. Chung, 138–49. Lecture Notes in Computer Science. Berlin, Heidelberg:
370 Springer. https://doi.org/10.1007/978-3-540-77004-6_11.
- 371 Zhu, Mu, and Ali Ghodsi. 2006. “Automatic Dimensionality Selection from the Scree Plot via
372 the Use of Profile Likelihood.” *Computational Statistics & Data Analysis* 51 (2): 918–30.
373 <https://doi.org/10.1016/j.csda.2005.09.010>.

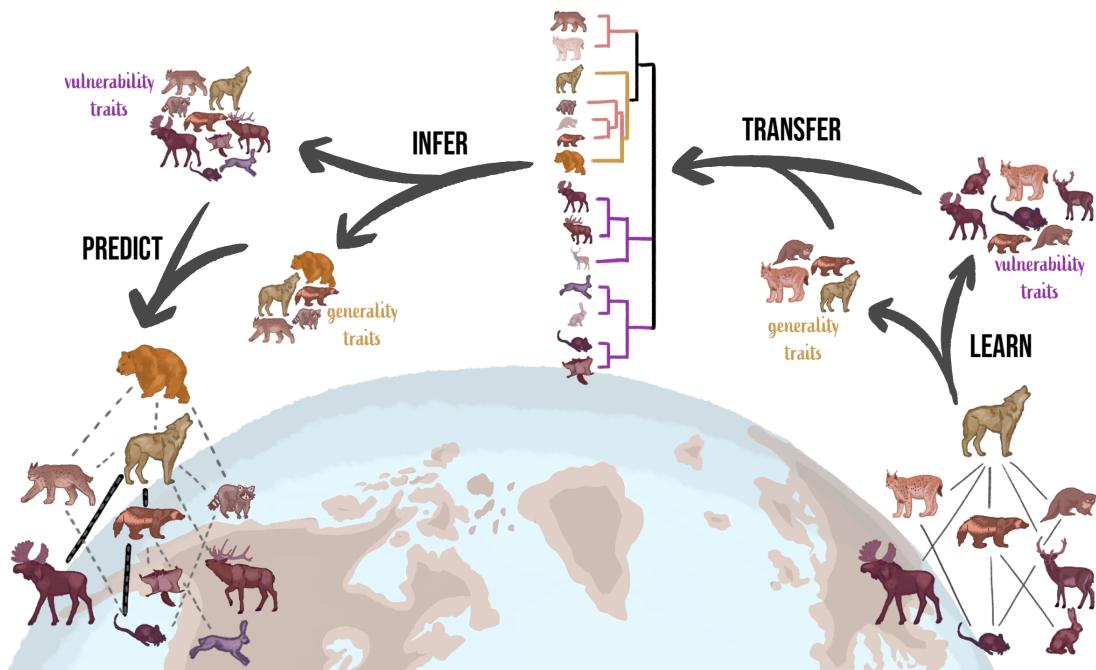


Figure 1: Overview of the phylogenetic transfer learning of species interactions networks. Starting from an initial network, we learn its representation through a graph embedding step (here, a truncated Singular Values Decomposition), yielding a series of latent traits; second, for the destination species pool, we perform ancestral character estimation using a phylogeny (here, using a Brownian model for the latent traits); we then sample from the reconstructed distribution of latent traits 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.

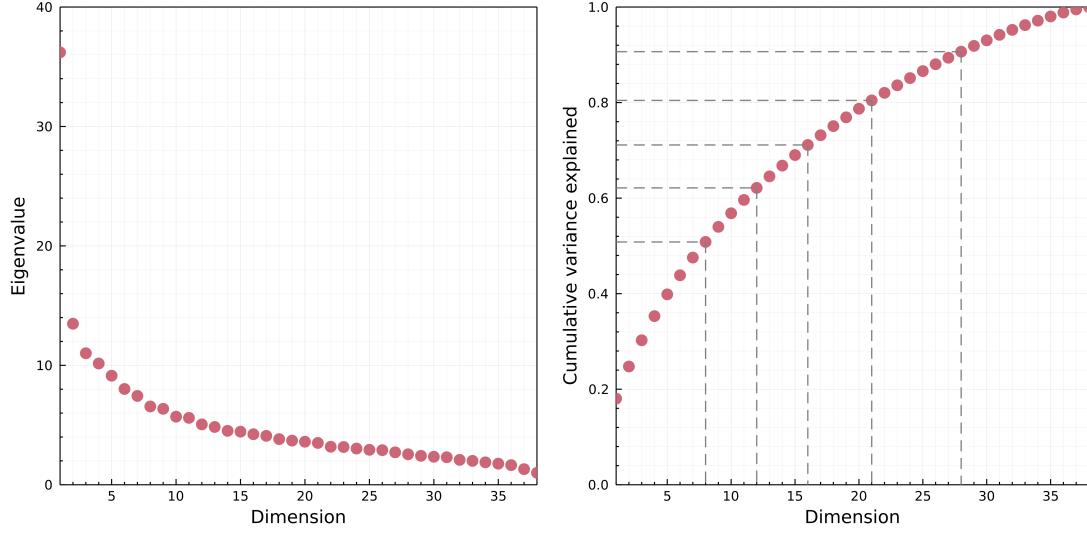


Figure 2: Left: representation of the screeplot of the eigenvalues from the tSVD on the European metaweb. The screeplot shows no obvious drop in the eigenvalues that may be leveraged to automatically detect a minimal dimension for embedding, after *e.g.* Zhu and Ghodsi (2006). Right: cumulative fraction of variance explained by each dimensions up to the rank of the European metaweb. The grey lines represents cutoff at 40, 60, and 80% 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.

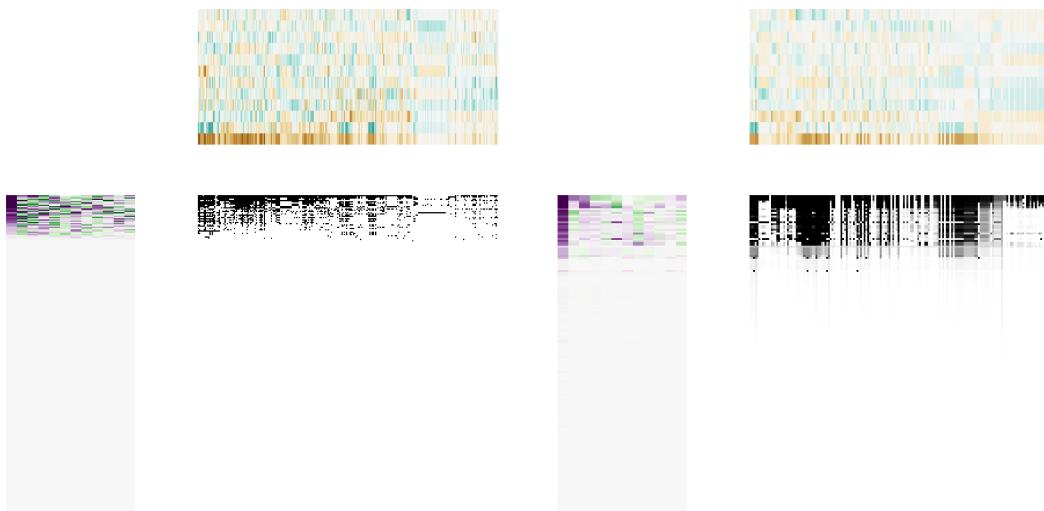


Figure 3: Visual representation of the left (green/purple) and right (green/brown) subspaces, alongside the adjacency matrix of the food web they encode (greyscale). The European metaweb is on the left, and the imputed Canadian metaweb (before data inflation) on the right. This figure illustrates how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

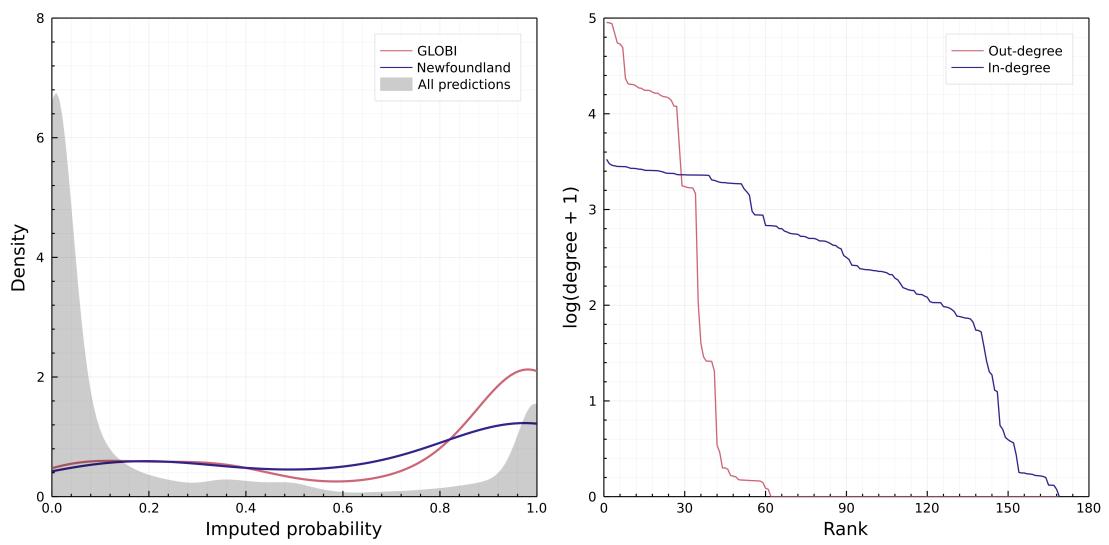


Figure 4: GLOBI comparison REF TK/TP

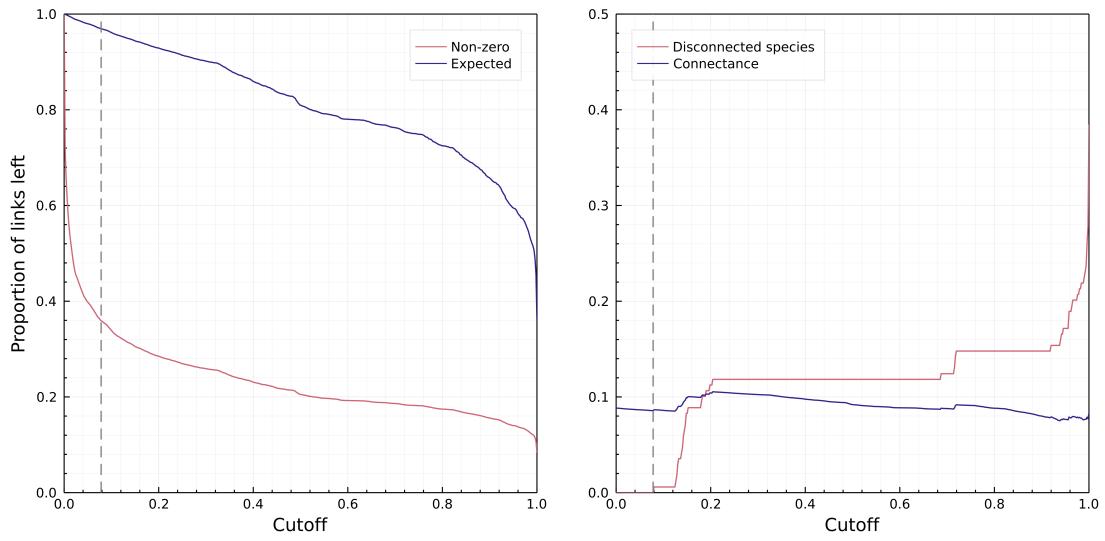


Figure 5: Left: effect of varying the cutoff for probabilities to be considered non-zero on the number of unique links and on \hat{L} , the probabilistic estimate of the number of links assuming that all interactions are independant. Right: effect of varying the cutoff on the number of disconnected species,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

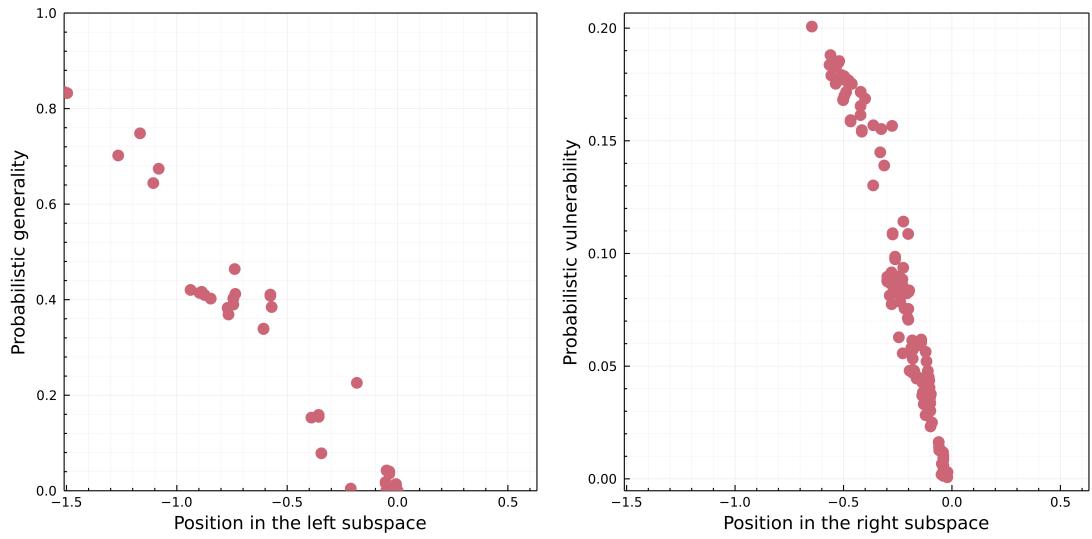


Figure 6: Left: there is a linear relationship between the values on the first dimension of the left subspace and the generality, *i.e.* the relative number of preys, *sensu* Schoener (1989). Species with a value of 0 in this subspace are at the bottom-most trophic level. Right: there is, similarly,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