

Predicting metawebs: graph embeddings can help alleviate spatial data deficiencies

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

¹ Département de Sciences Biologiques, Université de Montréal, Montréal, Canada ² Quebec Centre for Biodiversity Science, Montréal, Canada

[‡] These authors contributed equally to the work

Correspondance to:

Timothée Poisot — timothee.poisot@umontreal.ca

1. Metawebs, i.e. networks of potential interactions within a species pool, are a powerful abstraction to understand how large-scales species interaction networks are structured.
2. Because metawebs are typically expressed at large spatial and taxonomic scales, assembling them is a tedious and costly process; predictive methods can help circumvent the limitations in data deficiencies, by providing 'draft' metawebs.
3. One way to improve the predictive ability is to maximize the information used for prediction, by using graph embeddings rather than the list of species interactions. Graph embedding is an emerging field in machine learning that holds great potential for ecological problems.
4. In this perspective, we outline how the challenges associated with inferring metawebs line-up with the advantages of graph embeddings; furthermore, because metawebs are inherently spatial objects, we discuss how the choice of the species pool has consequences on the reconstructed network, but also embeds hypotheses about which human-made boundaries are ecologically meaningful.

1 Having a general solution for inferring *potential* interactions (despite the unavailability of interaction data)
2 could be the catalyst for significant breakthroughs in our ability to start thinking about species interaction
3 networks over large spatial scales (Hortal et al., 2015). In a recent overview of the field of ecological
4 network prediction, Strydom, Catchen, et al. (2021) identified two challenges of interest to the prediction
5 of interactions at large scales. First, there is a relative scarcity of relevant data in most places globally –
6 paradoxically, this restricts our ability to infer interactions to locations where inference is perhaps the least
7 required; second, accurate predictions often demand accurate predictors, and the lack of methods that can
8 leverage small amount of data is a serious impediment to our predictive ability globally.

9 Following the definition of Dunne (2006), a metaweb is a network analogue to the regional species pool;
10 specifically, it is an inventory of all *potential* interactions between species from a spatially delimited area
11 (and so captures the γ diversity of interactions). The metaweb is, therefore, *not* a prediction of the food
12 web at a specific locale within the spatial area it covers, and will have a different structure (notably by
13 having a larger connectance; see e.g. Wood et al., 2015). These local food webs (which captures the α
14 diversity of interactions) are a subset of the metaweb's species and interactions, and have been called
15 “metaweb realizations” (Poisot et al., 2015). Differences between local food web and their metaweb are
16 due to chance, species abundance and co-occurrence, local environmental conditions, and local
17 distribution of functional traits, among others.

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

The metaweb is an inherently probabilistic object

Yet, owing to the inherent plasticity of interactions, there have been documented instances of food webs undergoing rapid collapse/recovery cycles over short periods of time (Pedersen et al., 2017). The embedding of a network, in a sense, embeds its macro-evolutionary history, especially as RDPG captures ecological signal (Dalla Riva & Stouffer, 2016); at this point, it is important to recall that a metaweb is intended as a catalogue of all potential interactions, which should then be filtered (Morales-Castilla et al., 2015). In practice (and in this instance) the reconstructed metaweb will predict interactions that are plausible based on the species' evolutionary history, however some interactions would/would not be realized due to human impact.

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

[Figure 1 about here.]

57 Graph embedding offers promises for the inference of potential 58 interactions

59 Graph embedding fig. 1 is a varied family of machine learning techniques aiming to transform nodes and
60 edges into a vector space (Arsov & Mirceva, 2019), usually of a lower dimension, whilst maximally
61 retaining key properties of the graph (Yan et al., 2005). Ecological networks are an interesting candidate
62 for the widespread application of embeddings, as they tend to possess a shared structural backbone (Mora
63 et al., 2018), which hints at structural invariants that can be revealed at lower dimensions. Indeed,
64 previous work by Eklöf et al. (2013) suggests that food webs are inherently low-dimensional objects, and
65 can be adequately represented with less than ten dimensions. Simulation results by Botella et al. (2022)
66 suggest that there is no best method to identify architectural similarities between networks, and that
67 multiple approaches need to be tested and compared to the network descriptor of interest. This matches
68 with previous, more general results on graph embedding, which suggest that embedding algorithm choice
69 matters for the results (Goyal & Ferrara, 2018).

Table 1: Overview of some common graph embedding approaches, by time of publication, alongside examples of their use in the prediction of species interactions. Surprisingly, these methods have not yet been used routinely to predict species interactions; most of the examples we identified were either statistical associations, or analogues to joint species distribution models. ^a: statistical interactions; ^b: joint-SDM-like approach.

Method	Embedding approach	Reference	Species interactions
			application
RDPG	graphs through SVD	Young & Scheinerman (2007)	Poisot et al. (2021)
tSNE	nodes through statistical divergence	Hinton & Roweis (2002)	Cieslak et al. (2020) ^a
DeepWalk	graph walk	Perozzi et al. (2014)	Wardeh et al. (2021)
FastEmbed	graph through PCA/SVD analogue	Ramasamy & Madhow (2015)	

Method	Embedding approach	Reference	Species interactions
			application
LINE	nodes through statistical divergence	Tang et al. (2015)	
SDNE	nodes through auto-encoding	D. Wang et al. (2016)	
node2vec	node embedding	Grover & Leskovec (2016)	
graph2vec	sub-graph embedding	Narayanan et al. (2017)	
DMSE	joint node embedding	D. Chen et al. (2017)	D. Chen et al. (2017) ^b
HARP	nodes through a meta-strategy	H. Chen et al. (2017)	
GraphKKE	graph embedding	Melnyk et al. (2020)	Melnyk et al. (2020) ^a
Joint methods	multiple graphs	S. Wang et al. (2021)	

70 But the popularity of graph embedding techniques in machine learning is rather more intuitive than the
71 search for structural invariants: while graphs are discrete objects, machine learning techniques tend to
72 handle continuous data better. Therefore, bringing a sparse graph into a continuous, dense vector space
73 (Xu, 2020) opens up a broader variety of predictive algorithms.

74 **TK** Transfer + embedding graf

75 Rather than directly predicting biological rules (see *e.g.* Pichler et al., 2020 for an overview), which may be
76 confounded by the sparse nature of graph data, learning embeddings works in the low-dimensional space
77 that maximizes information about the network structure.

78 [Figure 2 about here.]

The metaweb embeds strong ecological hypotheses

As Herbert (1965) rightfully pointed out, “[y]ou can’t draw neat lines around planet-wide problems”; in this regard, any inference of a metaweb at large scales must contend with several interesting and interwoven families of problems.

The first is the spatial and taxonomic limit of the metaweb to embed and transfer. If the initial metaweb is too narrow in scope, notably from a taxonomic point of view, the chances of finding another area with enough related species (through phylogenetic relatedness or similarity of functional traits) to make a reliable inference decreases; this would likely be indicated by large confidence intervals during estimation of the values in the low-rank space, but the lack of well documented metawebs is currently preventing the development of more concrete guidelines. The question of phylogenetic relatedness and dispersal is notably true if the metaweb is assembled in an area with mostly endemic species, and as with every predictive algorithm, there is room for the application of our best ecological judgement. Conversely, the metaweb should be reliably filled, which assumes that the S^2 interactions in a pool of S species have been examined, either through literature surveys or expert elicitation.

TK Supp. Mat. 1 provides some guidance as to the type of sampling effort that should be prioritized. While RDPG was able to maintain very high predictive power when interactions were missing, the addition of false positive interactions was immediately detected; this suggests that it may be appropriate to err on the side of “too many” interactions when constructing the initial metaweb to be transferred.

The second series of problems are related to determining which area should be used to infer the new metaweb in, as this determines the species pool that must be used.

TK In our application, we focused on the mammals of Canada. The upside of this approach is that information at the country level is likely to be required by policy makers and stakeholders for their biodiversity assessment, as each country tends to set goals at the national level (Buxton et al., 2021) for which quantitative instruments are designed (Turak et al., 2017), with specific strategies often enacted at smaller scales (Ray et al., 2021). And yet, we do not really have a satisfying answer to the question of “where does a food web stop?”; the current most satisfying solutions involve examining the spatial consistency of network area relationships (Fortin et al., 2021; see e.g. Galiana et al., 2018, 2019, 2021), which is of course impossible in the absence of enough information about the network itself. This suggests that an *a posteriori* refinement of the results may be required, based on a downscaling of the metaweb.

108 The final family of problems relates less to the availability of data or quantitative tools, and more to the
109 praxis of spatial ecology. Operating under the context of national divisions, in large parts of the world,
110 reflects nothing more than the legacy of settler colonialism. Indeed, the use of ecological data is not an
111 apolitical act (Nost & Goldstein, 2021), as data infrastructures tend to be designed to answer questions
112 within national boundaries, and their use both draws upon and reinforces territorial statecraft; as per
113 Machen & Nost (2021), this is particularly true when the output of “algorithmic thinking” (*e.g.* relying on
114 machine learning to generate knowledge) can be re-used for governance (*e.g.* enacting conservation
115 decisions at the national scale). We therefore recognize that methods such as we propose operate under
116 the framework that contributed to the ongoing biodiversity crisis (Adam, 2014), reinforced environmental
117 injustice (Choudry, 2013; Domínguez & Luoma, 2020), and on Turtle Island especially, should be replaced
118 by Indigenous principles of land management (Eichhorn et al., 2019; No’kmaq et al., 2021). As we see
119 AI/ML being increasingly mobilized to generate knowledge that is lacking for conservation decisions (*e.g.*
120 Lamba et al., 2019; Mosebo Fernandes et al., 2020), our discussion of these tools need to go beyond the
121 technical, and into the governance consequences they can have.

122 **Acknowledgements:** We acknowledge that this study was conducted on land within the traditional
123 unceded territory of the Saint Lawrence Iroquoian, Anishinabewaki, Mohawk, Huron-Wendat, and
124 Omàmiwininiwak nations. TP, TS, DC, and LP received funding from the Canadian Institute for Ecology &
125 Evolution. FB is funded by the Institute for Data Valorization (IVADO). TS, SB, and TP are funded by a
126 donation from the Courtois Foundation. CB was awarded a Mitacs Elevate Fellowship no. IT12391, in
127 partnership with fRI Research, and also acknowledges funding from Alberta Innovates and the Forest
128 Resources Improvement Association of Alberta. M-JF acknowledges funding from NSERC Discovery
129 Grant and NSERC CRC. RR is funded by New Zealand’s Biological Heritage Ngā Koiora Tuku Iho
130 National Science Challenge, administered by New Zealand Ministry of Business, Innovation, and
131 Employment. BM is funded by the NSERC Alexander Graham Bell Canada Graduate Scholarship and the
132 FRQNT master’s scholarship. LP acknowledges funding from NSERC Discovery Grant (NSERC
133 RGPIN-2019-05771). TP acknowledges financial support from NSERC through the Discovery Grants and
134 Discovery Accelerator Supplement programs.

References

- Adam, R. (2014). *Elephant treaties: The Colonial legacy of the biodiversity crisis*. UPNE.
- Albouy, C., Archambault, P., Appeltans, W., Araújo, M. B., Beauchesne, D., Cazelles, K., Cirtwill, A. R., Fortin, M.-J., Galiana, N., Leroux, S. J., Pellissier, L., Poisot, T., Stouffer, D. B., Wood, S. A., & Gravel, D. (2019). The marine fish food web is globally connected. *Nature Ecology & Evolution*, 3(8, 8), 1153–1161. <https://doi.org/10.1038/s41559-019-0950-y>
- Arsov, N., & Mirceva, G. (2019, November 26). *Network Embedding: An Overview*. <http://arxiv.org/abs/1911.11726>
- Botella, C., Dray, S., Matias, C., Miele, V., & Thuiller, W. (2022). An appraisal of graph embeddings for comparing trophic network architectures. *Methods in Ecology and Evolution*, 13(1), 203–216. <https://doi.org/10.1111/2041-210X.13738>
- Buxton, R. T., Bennett, J. R., Reid, A. J., Shulman, C., Cooke, S. J., Francis, C. M., Nyboer, E. A., Pritchard, G., Binley, A. D., Avery-Gomm, S., Ban, N. C., Beazley, K. F., Bennett, E., Blight, L. K., Bortolotti, L. E., Camfield, A. F., Gadallah, F., Jacob, A. L., Naujokaitis-Lewis, I., ... Smith, P. A. (2021). Key information needs to move from knowledge to action for biodiversity conservation in Canada. *Biological Conservation*, 256, 108983. <https://doi.org/10.1016/j.biocon.2021.108983>
- Chen, D., Xue, Y., Fink, D., Chen, S., & Gomes, C. P. (2017). *Deep Multi-species Embedding*. 3639–3646. <https://www.ijcai.org/proceedings/2017/509>
- Chen, H., Perozzi, B., Hu, Y., & Skiena, S. (2017, November 16). *HARP: Hierarchical Representation Learning for Networks*. <http://arxiv.org/abs/1706.07845>
- Choudry, A. (2013). Saving biodiversity, for whom and for what? Conservation NGOs, complicity, colonialism and conquest in an era of capitalist globalization. In *NGOization: Complicity, contradictions and prospects* (pp. 24–44). Bloomsbury Publishing.
- Cieslak, M. C., Castelfranco, A. M., Roncalli, V., Lenz, P. H., & Hartline, D. K. (2020). T-Distributed Stochastic Neighbor Embedding (t-SNE): A tool for eco-physiological transcriptomic analysis. *Marine Genomics*, 51, 100723. <https://doi.org/10.1016/j.margen.2019.100723>
- Csermely, P. (2004). Strong links are important, but weak links stabilize them. *Trends in Biochemical Sciences*, 29(7), 331–334. <https://doi.org/10.1016/j.tibs.2004.05.004>

163 Dalla Riva, G. V., & Stouffer, D. B. (2016). Exploring the evolutionary signature of food webs' backbones
 164 using functional traits. *Oikos*, 125(4), 446–456. <https://doi.org/10.1111/oik.02305>

165 Dallas, T., Park, A. W., & Drake, J. M. (2017). Predicting cryptic links in host-parasite networks. *PLOS*
 166 *Computational Biology*, 13(5), e1005557. <https://doi.org/10.1371/journal.pcbi.1005557>

167 Domínguez, L., & Luoma, C. (2020). Decolonising Conservation Policy: How Colonial Land and
 168 Conservation Ideologies Persist and Perpetuate Indigenous Injustices at the Expense of the
 169 Environment. *Land*, 9(3, 3), 65. <https://doi.org/10.3390/land9030065>

170 Dunne, J. A. (2006). The Network Structure of Food Webs. In J. A. Dunne & M. Pascual (Eds.), *Ecological*
 171 *networks: Linking structure and dynamics* (pp. 27–86). Oxford University Press.

172 Eichhorn, M. P., Baker, K., & Griffiths, M. (2019). Steps towards decolonising biogeography. *Frontiers of*
 173 *Biogeography*, 12(1), 1–7. <https://doi.org/10.21425/F5FBG44795>

174 Eklöf, A., Jacob, U., Kopp, J., Bosch, J., Castro-Urgal, R., Chacoff, N. P., Dalsgaard, B., de Sassi, C., Galetti,
 175 M., Guimarães, P. R., Lomáscolo, S. B., Martín González, A. M., Pizo, M. A., Rader, R., Rodrigo, A.,
 176 Tylianakis, J. M., Vázquez, D. P., & Allesina, S. (2013). The dimensionality of ecological networks.
 177 *Ecology Letters*, 16(5), 577–583. <https://doi.org/10.1111/ele.12081>

178 Fortin, M.-J., Dale, M. R. T., & Brimacombe, C. (2021). Network ecology in dynamic landscapes.
 179 *Proceedings of the Royal Society B: Biological Sciences*, 288(1949), rspb.2020.1889, 20201889.
 180 <https://doi.org/10.1098/rspb.2020.1889>

181 Galiana, N., Barros, C., Braga, J., Ficetola, G. F., Maiorano, L., Thuiller, W., Montoya, J. M., & Lurgi, M.
 182 (2021). The spatial scaling of food web structure across European biogeographical regions. *Ecography*,
 183 *n/a(n/a)*. <https://doi.org/10.1111/ecog.05229>

184 Galiana, N., Hawkins, B. A., & Montoya, J. M. (2019). The geographical variation of network structure is
 185 scale dependent: Understanding the biotic specialization of host–parasitoid networks. *Ecography*,
 186 42(6), 1175–1187. <https://doi.org/10.1111/ecog.03684>

187 Galiana, N., Lurgi, M., Claramunt-López, B., Fortin, M.-J., Leroux, S., Cazelles, K., Gravel, D., & Montoya,
 188 J. M. (2018). The spatial scaling of species interaction networks. *Nature Ecology & Evolution*, 2(5),
 189 782–790. <https://doi.org/10.1038/s41559-018-0517-3>

190 Goyal, P., & Ferrara, E. (2018). Graph embedding techniques, applications, and performance: A survey.

191 *Knowledge-Based Systems*, 151, 78–94. <https://doi.org/10.1016/j.knosys.2018.03.022>

192 Gravel, D., Baiser, B., Dunne, J. A., Kopelke, J.-P., Martinez, N. D., Nyman, T., Poisot, T., Stouffer, D. B.,
 193 Tylanakis, J. M., Wood, S. A., & Roslin, T. (2018). Bringing Elton and Grinnell together: A quantitative
 194 framework to represent the biogeography of ecological interaction networks. *Ecography*, 0(0).
 195 <https://doi.org/10.1111/ecog.04006>

196 Grover, A., & Leskovec, J. (2016). Node2vec: Scalable Feature Learning for Networks. *Proceedings of the*
 197 *22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 855–864.
 198 <https://doi.org/10.1145/2939672.2939754>

199 Grünig, M., Mazzi, D., Calanca, P., Karger, D. N., & Pellissier, L. (2020). Crop and forest pest metawebs
 200 shift towards increased linkage and suitability overlap under climate change. *Communications Biology*,
 201 3(1, 1), 1–10. <https://doi.org/10.1038/s42003-020-0962-9>

202 Herbert, F. (1965). *Dune* (1st ed.). Chilton Book Company.

203 Hinton, G., & Roweis, S. T. (2002). Stochastic neighbor embedding. *NIPS*, 15, 833–840.

204 Hoffmann, J., Bar-Sinai, Y., Lee, L. M., Andrejevic, J., Mishra, S., Rubinstein, S. M., & Rycroft, C. H. (2019).
 205 Machine learning in a data-limited regime: Augmenting experiments with synthetic data uncovers
 206 order in crumpled sheets. *Science Advances*, 5(4), eaau6792.
 207 <https://doi.org/10.1126/sciadv.aau6792>

208 Hortal, J., de Bello, F., Diniz-Filho, J. A. F., Lewinsohn, T. M., Lobo, J. M., & Ladle, R. J. (2015). Seven
 209 Shortfalls that Beset Large-Scale Knowledge of Biodiversity. *Annual Review of Ecology, Evolution, and*
 210 *Systematics*, 46(1), 523–549. <https://doi.org/10.1146/annurev-ecolsys-112414-054400>

211 Jordano, P. (2016). Sampling networks of ecological interactions. *Functional Ecology*, 30(12), 1883–1893.
 212 <https://doi.org/10.1111/1365-2435.12763>

213 Lamba, A., Cassey, P., Segaran, R. R., & Koh, L. P. (2019). Deep learning for environmental conservation.
 214 *Current Biology*, 29(19), R977–R982. <https://doi.org/10.1016/j.cub.2019.08.016>

215 Machen, R., & Nost, E. (2021). Thinking algorithmically: The making of hegemonic knowledge in climate
 216 governance. *Transactions of the Institute of British Geographers*, 46(3), 555–569.
 217 <https://doi.org/10.1111/tran.12441>

218 McLeod, A., Leroux, S. J., Gravel, D., Chu, C., Cirtwill, A. R., Fortin, M.-J., Galiana, N., Poisot, T., & Wood,

219 S. A. (2021). Sampling and asymptotic network properties of spatial multi-trophic networks. *Oikos*,
220 *n/a(n/a)*. <https://doi.org/10.1111/oik.08650>

221 Melnyk, K., Klus, S., Montavon, G., & Conrad, T. O. F. (2020). GraphKKE: Graph Kernel Koopman
222 embedding for human microbiome analysis. *Applied Network Science*, 5(1), 96.
223 <https://doi.org/10.1007/s41109-020-00339-2>

224 Mora, B. B., Gravel, D., Gilarranz, L. J., Poisot, T., & Stouffer, D. B. (2018). Identifying a common backbone
225 of interactions underlying food webs from different ecosystems. *Nature Communications*, 9(1), 2603.
226 <https://doi.org/10.1038/s41467-018-05056-0>

227 Morales-Castilla, I., Matias, M. G., Gravel, D., & Araújo, M. B. (2015). Inferring biotic interactions from
228 proxies. *Trends in Ecology & Evolution*, 30(6), 347–356.
229 <https://doi.org/10.1016/j.tree.2015.03.014>

230 Mosebo Fernandes, A. C., Quintero Gonzalez, R., Lenihan-Clarke, M. A., Leslie Trotter, E. F., & Jokar
231 Arsanjani, J. (2020). Machine Learning for Conservation Planning in a Changing Climate.
232 *Sustainability*, 12(18, 18), 7657. <https://doi.org/10.3390/su12187657>

233 Narayanan, A., Chandramohan, M., Venkatesan, R., Chen, L., Liu, Y., & Jaiswal, S. (2017, July 17).
234 *Graph2vec: Learning Distributed Representations of Graphs*. <http://arxiv.org/abs/1707.05005>

235 Neutel, A.-M., Heesterbeek, J. A. P., & de Ruiter, P. C. (2002). Stability in Real Food Webs: Weak Links in
236 Long Loops. *Science*, 296(5570), 1120–1123. <https://doi.org/10.1126/science.1068326>

237 No'kmaq, M., Marshall, A., Beazley, K. F., Hum, J., Joudry, shalan, Papadopoulos, A., Pictou, S., Rabesca,
238 J., Young, L., & Zurba, M. (2021). “Awakening the sleeping giant”: Re-Indigenization principles for
239 transforming biodiversity conservation in Canada and beyond. *FACETS*, 6(1), 839–869.

240 Nost, E., & Goldstein, J. E. (2021). A political ecology of data. *Environment and Planning E: Nature and*
241 *Space*, 25148486211043503. <https://doi.org/10.1177/25148486211043503>

242 O'Connor, L. M. J., Pollock, L. J., Braga, J., Ficetola, G. F., Maiorano, L., Martinez-Almoyna, C.,
243 Montemaggiore, A., Ohlmann, M., & Thuiller, W. (2020). Unveiling the food webs of tetrapods across
244 Europe through the prism of the Eltonian niche. *Journal of Biogeography*, 47(1), 181–192.
245 <https://doi.org/10.1111/jbi.13773>

246 Pedersen, E. J., Thompson, P. L., Ball, R. A., Fortin, M.-J., Gouhier, T. C., Link, H., Moritz, C., Nenzen, H.,

Stanley, R. R. E., Taranu, Z. E., Gonzalez, A., Guichard, F., & Pepin, P. (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>

Perozzi, B., Al-Rfou, R., & Skiena, S. (2014). DeepWalk: Online learning of social representations. *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 701–710. <https://doi.org/10.1145/2623330.2623732>

Pichler, M., Boreux, V., Klein, A.-M., Schleuning, M., & Hartig, F. (2020). Machine learning algorithms to infer trait-matching and predict species interactions in ecological networks. *Methods in Ecology and Evolution*, 11(2), 281–293. <https://doi.org/10.1111/2041-210X.13329>

Poisot, T., Cirtwill, A. R., Cazelles, K., Gravel, D., Fortin, M.-J., & Stouffer, D. B. (2016). The structure of probabilistic networks. *Methods in Ecology and Evolution*, 7(3), 303–312. <https://doi.org/10.1111/2041-210X.12468>

Poisot, T., Ouellet, M.-A., Mollentze, N., Farrell, M. J., Becker, D. J., Albery, G. F., Gibb, R. J., Seifert, S. N., & Carlson, C. J. (2021, May 31). *Imputing the mammalian virome with linear filtering and singular value decomposition*. <http://arxiv.org/abs/2105.14973>

Poisot, T., Stouffer, D. B., & Gravel, D. (2015). Beyond species: Why ecological interaction networks vary through space and time. *Oikos*, 124(3), 243–251. <https://doi.org/10.1111/oik.01719>

Ramasamy, D., & Madhow, U. (2015). Compressive spectral embedding: Sidestepping the SVD. In C. Cortes, N. Lawrence, D. Lee, M. Sugiyama, & R. Garnett (Eds.), *Advances in neural information processing systems* (Vol. 28). Curran Associates, Inc. <https://proceedings.neurips.cc/paper/2015/file/4f6ffe13a5d75b2d6a3923922b3922e5-Paper.pdf>

Ray, J. C., Grimm, J., & Olive, A. (2021). The biodiversity crisis in Canada: Failures and challenges of federal and sub-national strategic and legal frameworks. *FACETS*, 6, 1044–1068. <https://doi.org/10.1139/facets-2020-0075>

Strydom, T., Bouskila, S., Banville, F., Barros, C., Caron, D., Farrell, M. J., Fortin, M.-J., Hemming, V., Mercier, B., Pollock, L., Runghen, R., Riva, G. V. D., & Poisot, T. (2021). *Food web reconstruction through phylogenetic transfer of low-rank network representation*. <https://doi.org/10.32942/osf.io/y7sdz>

Strydom, T., Catchen, M. D., Banville, F., Caron, D., Dansereau, G., Desjardins-Proulx, P., Forero-Muñoz,

275 N. R., Higino, G., Mercier, B., Gonzalez, A., Gravel, D., Pollock, L., & Poisot, T. (2021). A roadmap
 276 towards predicting species interaction networks (across space and time). *Philosophical Transactions of*
 277 *the Royal Society B: Biological Sciences*, 376(1837), 20210063.
 278 <https://doi.org/10.1098/rstb.2021.0063>

279 Tang, J., Qu, M., Wang, M., Zhang, M., Yan, J., & Mei, Q. (2015). LINE: Large-scale Information Network
 280 Embedding. *Proceedings of the 24th International Conference on World Wide Web*, 1067–1077.
 281 <https://doi.org/10.1145/2736277.2741093>

282 Turak, E., Brazill-Boast, J., Cooney, T., Drielsma, M., DelaCruz, J., Dunkerley, G., Fernandez, M., Ferrier,
 283 S., Gill, M., Jones, H., Koen, T., Leys, J., McGeoch, M., Mihoub, J.-B., Scanes, P., Schmeller, D., &
 284 Williams, K. (2017). Using the essential biodiversity variables framework to measure biodiversity
 285 change at national scale. *Biological Conservation*, 213, 264–271.
 286 <https://doi.org/10.1016/j.biocon.2016.08.019>

287 Wang, D., Cui, P., & Zhu, W. (2016). Structural Deep Network Embedding. *Proceedings of the 22nd ACM*
 288 *SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1225–1234.
 289 <https://doi.org/10.1145/2939672.2939753>

290 Wang, S., Arroyo, J., Vogelstein, J. T., & Priebe, C. E. (2021). Joint Embedding of Graphs. *IEEE*
 291 *Transactions on Pattern Analysis and Machine Intelligence*, 43(4), 1324–1336.
 292 <https://doi.org/10.1109/TPAMI.2019.2948619>

293 Wardeh, M., Baylis, M., & Blagrove, M. S. C. (2021). Predicting mammalian hosts in which novel
 294 coronaviruses can be generated. *Nature Communications*, 12(1, 1), 780.
 295 <https://doi.org/10.1038/s41467-021-21034-5>

296 Wood, S. A., Russell, R., Hanson, D., Williams, R. J., & Dunne, J. A. (2015). Effects of spatial scale of
 297 sampling on food web structure. *Ecology and Evolution*, 5(17), 3769–3782.
 298 <https://doi.org/10.1002/ece3.1640>

299 Xu, M. (2020, December 14). *Understanding graph embedding methods and their applications*.
 300 <http://arxiv.org/abs/2012.08019>

301 Yan, S., Xu, D., Zhang, B., & Zhang, H.-J. (2005). Graph embedding: A general framework for
 302 dimensionality reduction. *2005 IEEE Computer Society Conference on Computer Vision and Pattern*
 303 *Recognition (CVPR'05)*, 2, 830–837 vol. 2. <https://doi.org/10.1109/CVPR.2005.170>

304 Young, S. J., & Scheinerman, E. R. (2007). Random Dot Product Graph Models for Social Networks. In A.
305 Bonato & F. R. K. Chung (Eds.), *Algorithms and Models for the Web-Graph* (pp. 138–149). Springer.
306 https://doi.org/10.1007/978-3-540-77004-6_11

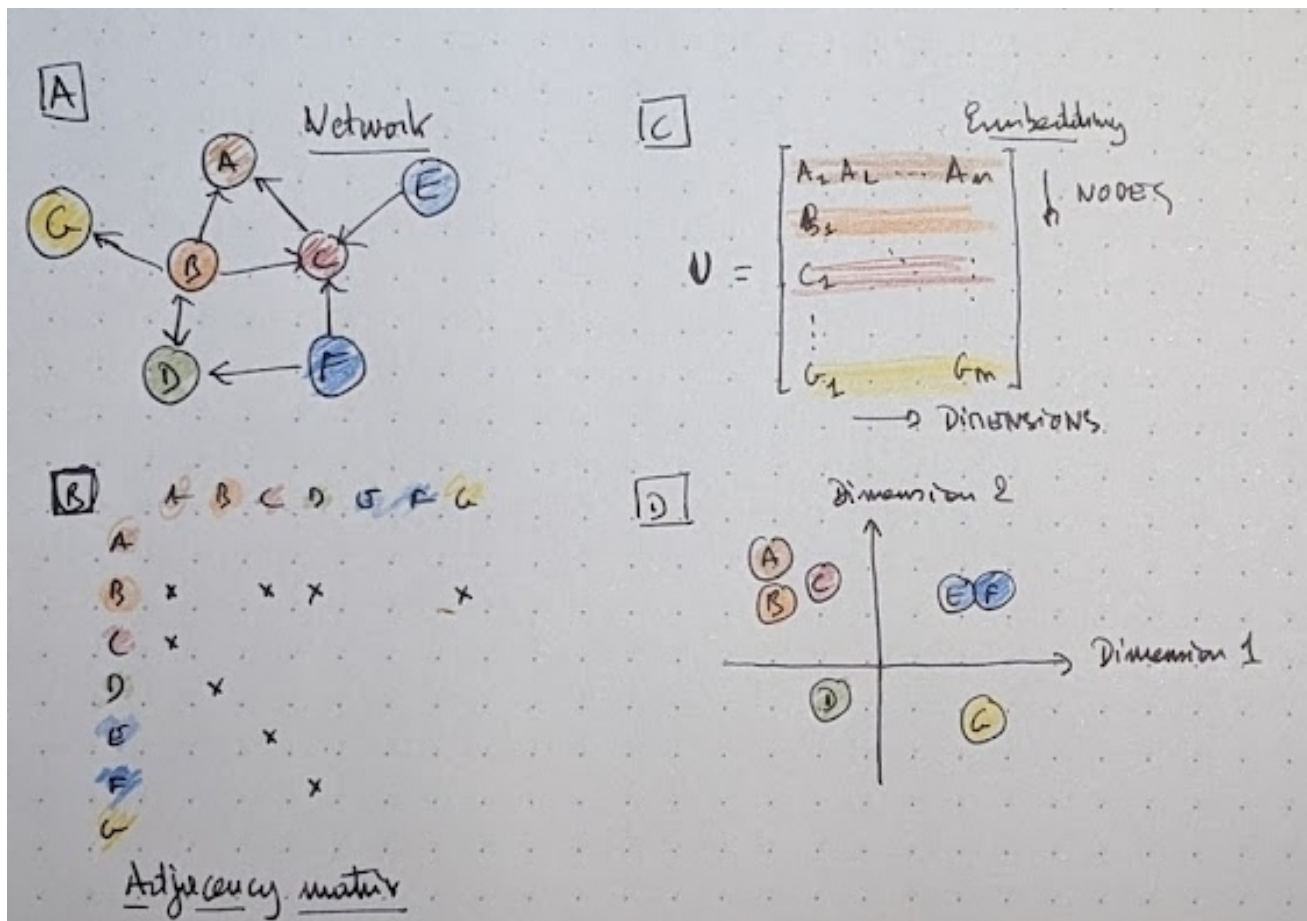


Figure 1: Overview of the embedding process. A network (A), possibly represented as its adjacency matrix (B), is converted into a lower-dimensional object (C) where nodes, subgraphs, or edges have specific values (see tbl. 1). For the purposes of prediction, this low-dimensional object encodes feature vectors for e.g. the nodes. Embedding also allows to visualize the structure in the data differently (D), much like with a principal component analysis.

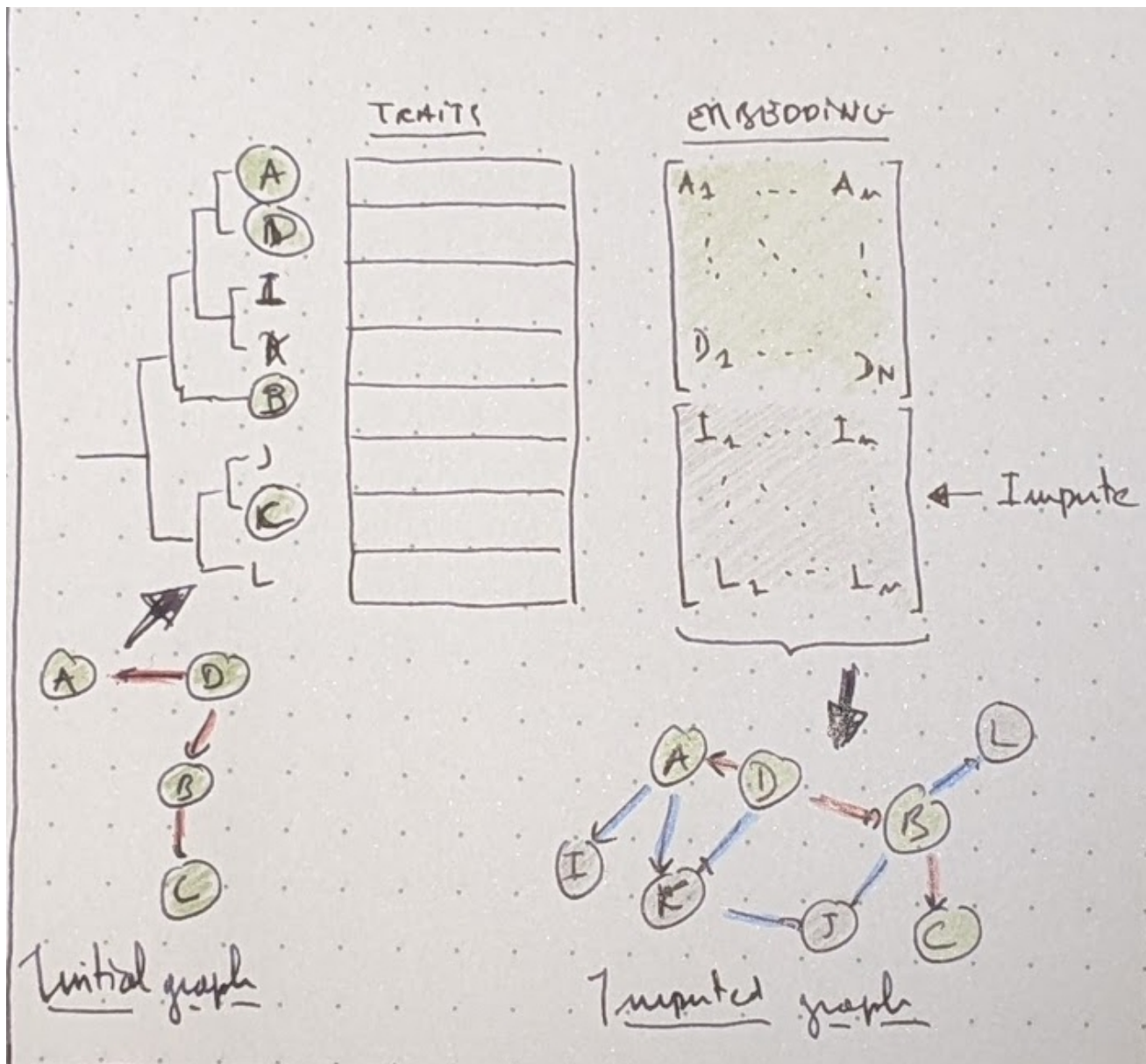


Figure 2: From a low-dimensional figure vector (see fig. 1), it is possible to develop predictive approaches. Nodes in an ecological network are species, for which we can leverage phylogenetic relatedness (e.g. Strydom, Bouskila, et al., 2021) or functional traits to fill the values of additional species we would like to project in this space (here, I, J, K, and L) from the embedding of known species (here, A, B, C, and D). Because embeddings can be projected back to a graph, this allows to reconstruct a network with these new species. This approach constitutes an instance of transfer learning.