

# Predicting metawebs: transfer of graph embeddings can help alleviate spatial data deficiencies

Tanya Strydom<sup>1,2,‡</sup>   Timothée Poisot<sup>1,2,‡</sup>

<sup>1</sup> Département de Sciences Biologiques, Université de Montréal, Montréal, Canada   <sup>2</sup> Quebec Centre for Biodiversity Science, Montréal, Canada

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

## Correspondance to:

Timothée Poisot — [timothee.poisot@umontreal.ca](mailto: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 Being able to infer *potential* interactions could be the catalyst for significant breakthroughs in our ability  
2 to start thinking about species interaction networks over large spatial scales (Hortal et al., 2015). In a  
3 recent overview of the field of ecological network prediction, Strydom, Catchen, et al. (2021) identified  
4 two challenges of interest to the prediction of interactions at large scales. First, there is a relative scarcity  
5 of relevant data in most places globally – paradoxically, this restricts our ability to infer interactions to  
6 locations where inference is perhaps the least required; second, accurate predictions often demand  
7 accurate predictors, and the lack of methods that can leverage small amount of data is a serious  
8 impediment to our global predictive ability. In most places, our most reliable biodiversity knowledge is  
9 that of a species pool: through the analysis of databases like GBIF or IUCN, it is possible to establish a list  
10 of species in a region of interest; but establishing the interactions between these species is difficult.

11 Following the definition of Dunne (2006), a metaweb is the ecological network analogue to the species  
12 pool; specifically, it inventories *potential* interactions between species for a spatially delimited area (and so  
13 captures the  $\gamma$  diversity of interactions). The metaweb is not a prediction of the network at a specific point  
14 within the spatial area it covers: it will have a different structure, notably by having a larger connectance  
15 (see *e.g.* Wood et al., 2015), from any of these local networks. These local networks (capturing the  $\alpha$   
16 diversity of interactions) are a subset of the metaweb’s species and interactions, and have been called  
17 “metaweb realizations” (Poisot et al., 2015). Differences between local networks and their metawebs are  
18 due to chance, species abundance and co-occurrence, local environmental conditions, and local  
19 distribution of functional traits, among others.

20 Because the metaweb represents the joint effect of functional, phylogenetic, and macroecological  
21 processes (Morales-Castilla et al., 2015), it holds valuable ecological information. Specifically, it is the  
22 “upper bounds” on what the composition of the local networks can be (see *e.g.* McLeod et al., 2021). These  
23 local networks may be reconstructed given appropriate knowledge of local species composition, providing  
24 information on structure of food webs at finer spatial scales. This has been done for example for  
25 tree-galler-parasitoid systems (Gravel et al., 2018), fish trophic interactions (Albouy et al., 2019), tetrapod  
26 trophic interactions (O’Connor et al., 2020), and crop-pest networks (Grünig et al., 2020). In this  
27 contribution, we highlight the *probabilistic* nature of metawebs, discuss how a family of machine learning  
28 tools (graph embeddings and transfer learning) can be used to overcome data limitations to metaweb  
29 inference, and highlight how the use of metawebs introduces important questions for the field of network  
30 ecology.

## **The metaweb is an inherently probabilistic object**

Dallas et al. (2017) suggested that most links in ecological networks are cryptic, *i.e.* uncommon or hard to observe. This argument echoes Jordano (2016): sampling 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, we expect to see an over-representation of low-probability interactions under a model that accurately predicts interaction probabilities. Yet the original metaweb definition, and indeed most past uses of metawebs, was based on the presence/absence of interactions. Moving towards *probabilistic* metawebs, by represent interactions as Bernoulli events (see *e.g.* Poisot et al., 2016), offers the opportunity to weigh these rare interactions appropriately. The inherent plasticity of interactions is important to capture: there have been documented instances of food webs undergoing rapid collapse/recovery cycles over short periods of time (*e.g.* Pedersen et al., 2017). These considerations emphasize why metaweb predictions should focus on quantitative (preferentially probabilistic) predictions; this should constrain the suite of appropriate models.

Yet it is important to recall that a metaweb is intended as a catalogue of all potential interactions, which is then filtered (Morales-Castilla et al., 2015). 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 will increase exponentially, and will do so for each undocumented interaction. Recent proposals suggest that machine learning algorithms, in these situations, can act as data generators (Hoffmann et al., 2019): high quality observational data can generate the core rules underpinning the network structure, and be supplemented with synthetic data coming from predictive models, increasing 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. A probabilistic metaweb represents an aggregation of informative priors on the interactions, an elusive information with the potential to boost our predictive ability (Bartomeus et al., 2016).

[Figure 1 about here.]

## 59 Graph embedding offers promises for the inference of potential 60 interactions

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

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. <sup>a</sup>: statistical interactions; <sup>b</sup>: joint-SDM-like approach.

Method	Embedding approach	Reference	Application
tSNE	nodes through statistical divergence	Hinton & Roweis (2002)	Cieslak et al. (2020) <sup>a</sup>
RDPG	graph through SVD	Young & Scheinerman (2007)	Poisot et al. (2021)
DeepWalk	graph walk	Perozzi et al. (2014)	Wardeh et al. (2021)
FastEmbed	graph through PCA/SVD analogue	Ramasamy & Madhoo (2015)	
LINE	nodes through statistical divergence	Tang et al. (2015)	

Method	Embedding approach	Reference	Application
SDNE	nodes through auto-encoding	D. Wang et al. (2016)	
node2vec	nodes embedding	Grover & Leskovec (2016)	
graph2vec	sub-graph embedding	Narayanan et al. (2017)	
DMSE	joint nodes embedding	D. Chen et al. (2017)	D. Chen et al. (2017) <sup>b</sup>
HARP	nodes through a meta-strategy	H. Chen et al. (2017)	
GraphKKE	graph embedding	Melnyk et al. (2020)	Melnyk et al. (2020) <sup>a</sup>
Joint methods	multiple graphs	S. Wang et al. (2021)	

74 The popularity of graph embedding techniques in machine learning is more prosaic than the search for  
 75 structural invariants: graphs are discrete objects, and machine learning techniques tend to handle  
 76 continuous data better. Bringing a sparse graph into a continuous, dense vector space (Xu, 2020) opens up  
 77 a broader variety of predictive algorithms, notably of the sort that are able to predict events as probabilities  
 78 (Murphy, 2022). Furthermore, the projection of the graph itself is a representation that can be learned;  
 79 Runghen et al. (2021), for example, used a neural network to learn the embedding of a network in which  
 80 not all interactions were known, based on nodes metadata. This example has many parallels in ecology  
 81 see 2, in which node metadata can be given by phylogeny or functional traits. Rather than directly  
 82 predicting biological rules (see *e.g.* Pichler et al., 2020 for an overview), which may be confounded by the  
 83 sparse nature of graph data, learning embeddings works in the low-dimensional space that maximizes  
 84 information about the network structure. This approach is further justified by the observation, for  
 85 example, that the macro-evolutionary history of a network is adequately represented by some graph  
 86 embeddings (RDGP; see Dalla Riva & Stouffer, 2016).

[Figure 2 about here.]

## **The metaweb embeds ecological hypotheses and practices**

The goal of metaweb inference is to provide information about the interactions between species at a large spatial scale. But as Herbert (1965) rightfully pointed out, “[y]ou can’t draw neat lines around planet-wide problems”; any inference of a metaweb at large scales must contend with several novel, 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, or by non-overlapping trait distributions in the known and unknown species. 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.

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. Metawebs can be constructed by assigning interactions in a list of species within geographic boundaries. The upside of this approach is that information at the country level is likely to be required for biodiversity assessments, as countries set goals at the national level (Buxton et al., 2021), and as quantitative instruments are designed to work at these scales (Turak et al., 2017); specific strategies are often enacted at smaller scales, nested within a specific country (Ray et al., 2021). But there is no guarantee that these boundaries are meaningful. In fact, we do not have a satisfying answer to the question of “where does a food web stop?”; the most promising 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 impossible in the absence of enough information about the network itself. This suggests that inferred metaweb should be further downscaled to allow *a posteriori* analyses.

The final family of problems relates less to ecological concepts and more to the praxis of ecological research. Operating under the context of national divisions, in large parts of the world, reflects nothing more than the legacy of settler colonialism. Indeed, the use of ecological data is not an apolitical act (Nost

117 & Goldstein, 2021), as data infrastructures tend to be designed to answer questions within national  
118 boundaries, and their use both draws upon and reinforces territorial statecraft; as per Machen & Nost  
119 (2021), this is particularly true when the output of “algorithmic thinking” (e.g. relying on machine  
120 learning to generate knowledge) can be re-used for governance (e.g. enacting conservation decisions at the  
121 national scale). We therefore recognize that predictive approaches deployed at the continental scale, no  
122 matter their intent, originate in the framework that contributed to the ongoing biodiversity crisis (Adam,  
123 2014), reinforced environmental injustice (Choudry, 2013; Domínguez & Luoma, 2020), and on Turtle  
124 Island especially, should be replaced by Indigenous principles of land management (Eichhorn et al., 2019;  
125 No’kmaq et al., 2021). As we see AI/ML being increasingly mobilized to generate knowledge that is  
126 lacking for conservation decisions (e.g. Lamba et al., 2019; Mosebo Fernandes et al., 2020), our discussion  
127 of these tools need to go beyond the technical, and into the governance consequences they can have.

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

## 141 References

- 142 Adam, R. (2014). *Elephant treaties: The Colonial legacy of the biodiversity crisis*. UPNE.
- 143 Albouy, C., Archambault, P., Appeltans, W., Araújo, M. B., Beauchesne, D., Cazelles, K., Cirtwill, A. R.,  
144 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>

Bartomeus, I., Gravel, D., Tylianakis, J. M., Aizen, M. A., Dickie, I. A., & Bernard-Verdier, M. (2016). A common framework for identifying linkage rules across different types of interactions. *Functional Ecology*, 30(12), 1894–1903. <http://onlinelibrary.wiley.com/doi/10.1111/1365-2435.12666/full>

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>

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

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

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

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

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

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

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

191 Galiana, N., Barros, C., Braga, J., Ficetola, G. F., Maiorano, L., Thuiller, W., Montoya, J. M., & Lurgi, M.  
 192 (2021). The spatial scaling of food web structure across European biogeographical regions. *Ecography*,  
 193 *n/a*(n/a). <https://doi.org/10.1111/ecog.05229>

194 Galiana, N., Hawkins, B. A., & Montoya, J. M. (2019). The geographical variation of network structure is  
 195 scale dependent: Understanding the biotic specialization of host–parasitoid networks. *Ecography*,  
 196 42(6), 1175–1187. <https://doi.org/10.1111/ecog.03684>

197 Galiana, N., Lurgi, M., Claramunt-López, B., Fortin, M.-J., Leroux, S., Cazelles, K., Gravel, D., & Montoya,  
 198 J. M. (2018). The spatial scaling of species interaction networks. *Nature Ecology & Evolution*, 2(5),  
 199 782–790. <https://doi.org/10.1038/s41559-018-0517-3>

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

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

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

206     Grover, A., & Leskovec, J. (2016). Node2vec: Scalable Feature Learning for Networks. *Proceedings of the*  
207     *22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 855–864.  
208     <https://doi.org/10.1145/2939672.2939754>

209     Grünig, M., Mazzi, D., Calanca, P., Karger, D. N., & Pellissier, L. (2020). Crop and forest pest metawebs  
210     shift towards increased linkage and suitability overlap under climate change. *Communications Biology*,  
211     3(1, 1), 1–10. <https://doi.org/10.1038/s42003-020-0962-9>

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

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

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

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

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

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

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

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

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

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

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

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

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

243 Murphy, K. P. (2022). *Probabilistic machine learning: An introduction*. MIT Press. [probml.ai](https://probml.ai)

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

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

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

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

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

257 Pedersen, E. J., Thompson, P. L., Ball, R. A., Fortin, M.-J., Gouhier, T. C., Link, H., Moritz, C., Nenzen, H.,  
 258 Stanley, R. R. E., Taranu, Z. E., Gonzalez, A., Guichard, F., & Pepin, P. (2017). Signatures of the  
 259 collapse and incipient recovery of an overexploited marine ecosystem. *Royal Society Open Science*, 4(7),  
 260 170215. <https://doi.org/10.1098/rsos.170215>

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

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

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

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

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

275 Ramasamy, D., & Madhow, U. (2015). Compressive spectral embedding: Sidestepping the SVD. In C.  
 276 Cortes, N. Lawrence, D. Lee, M. Sugiyama, & R. Garnett (Eds.), *Advances in neural information*  
 277 *processing systems* (Vol. 28). Curran Associates, Inc. <https://proceedings.neurips.cc/paper/2015/file/4f6ffe13a5d75b2d6a3923922b3922e5-Paper.pdf>  
 278

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

282 Runghen, R., Stouffer, D. B., & Dalla Riva, G. V. (2021). *Exploiting node metadata to predict interactions in*  
 283 *large networks using graph embedding and neural networks*.  
 284 <https://doi.org/10.1101/2021.06.10.447991>

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

288 Strydom, T., Catchen, M. D., Banville, F., Caron, D., Dansereau, G., Desjardins-Proulx, P., Forero-Muñoz,  
 289 N. R., Higino, G., Mercier, B., Gonzalez, A., Gravel, D., Pollock, L., & Poisot, T. (2021). A roadmap  
 290 towards predicting species interaction networks (across space and time). *Philosophical Transactions of*  
 291 *the Royal Society B: Biological Sciences*, 376(1837), 20210063.  
 292 <https://doi.org/10.1098/rstb.2021.0063>

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

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

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

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

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

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

313 Xu, M. (2020, December 14). *Understanding graph embedding methods and their applications*.

314 <http://arxiv.org/abs/2012.08019>

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

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

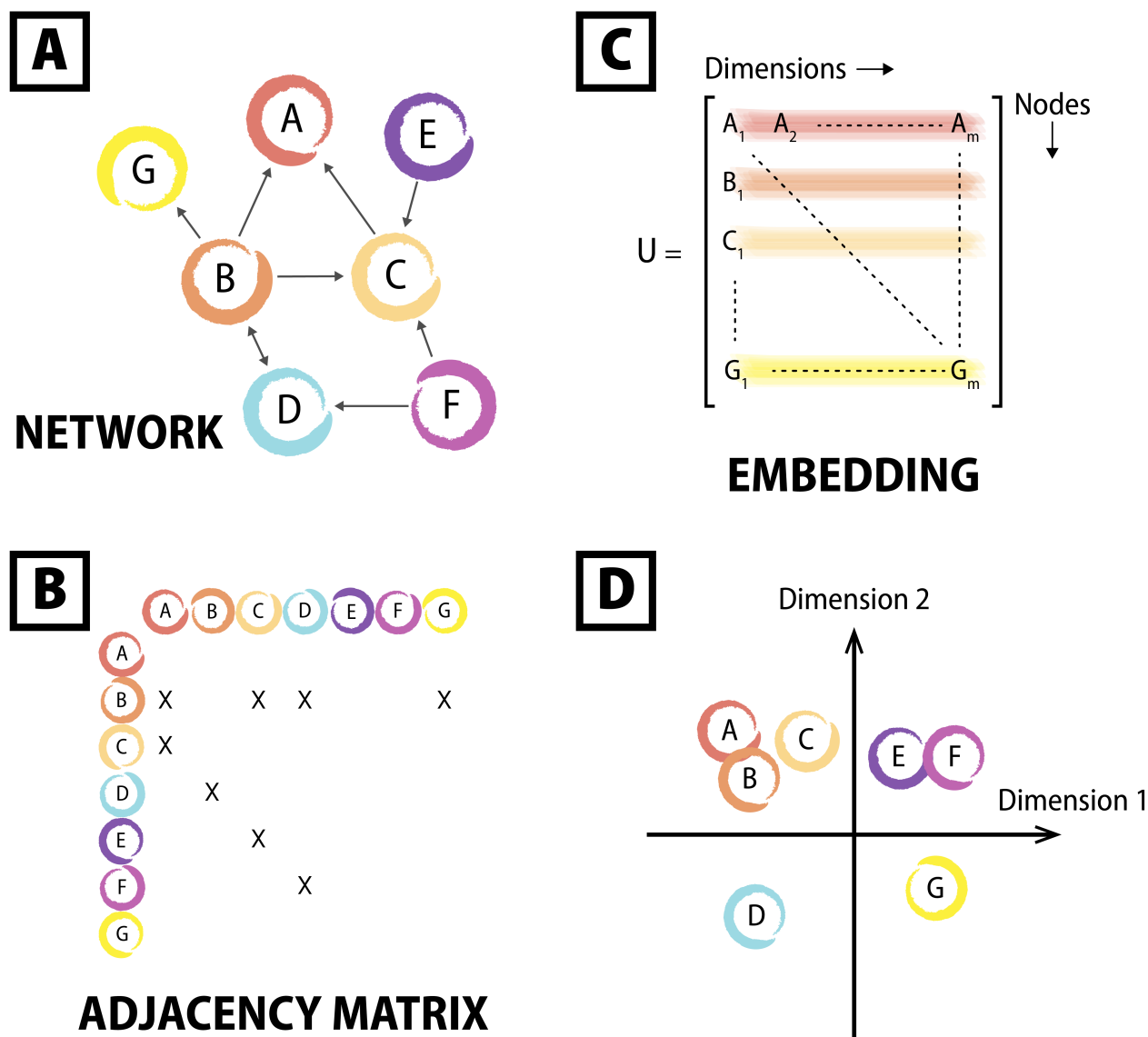


Figure 1: Overview of the embedding process. A network (A), possibly represented as its adjacency matrix (B), is converted into a lower-dimensional object (C) where nodes, subgraphs, or edges have specific values (see tbl. 1). For the purposes of prediction, this low-dimensional object encodes feature 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.



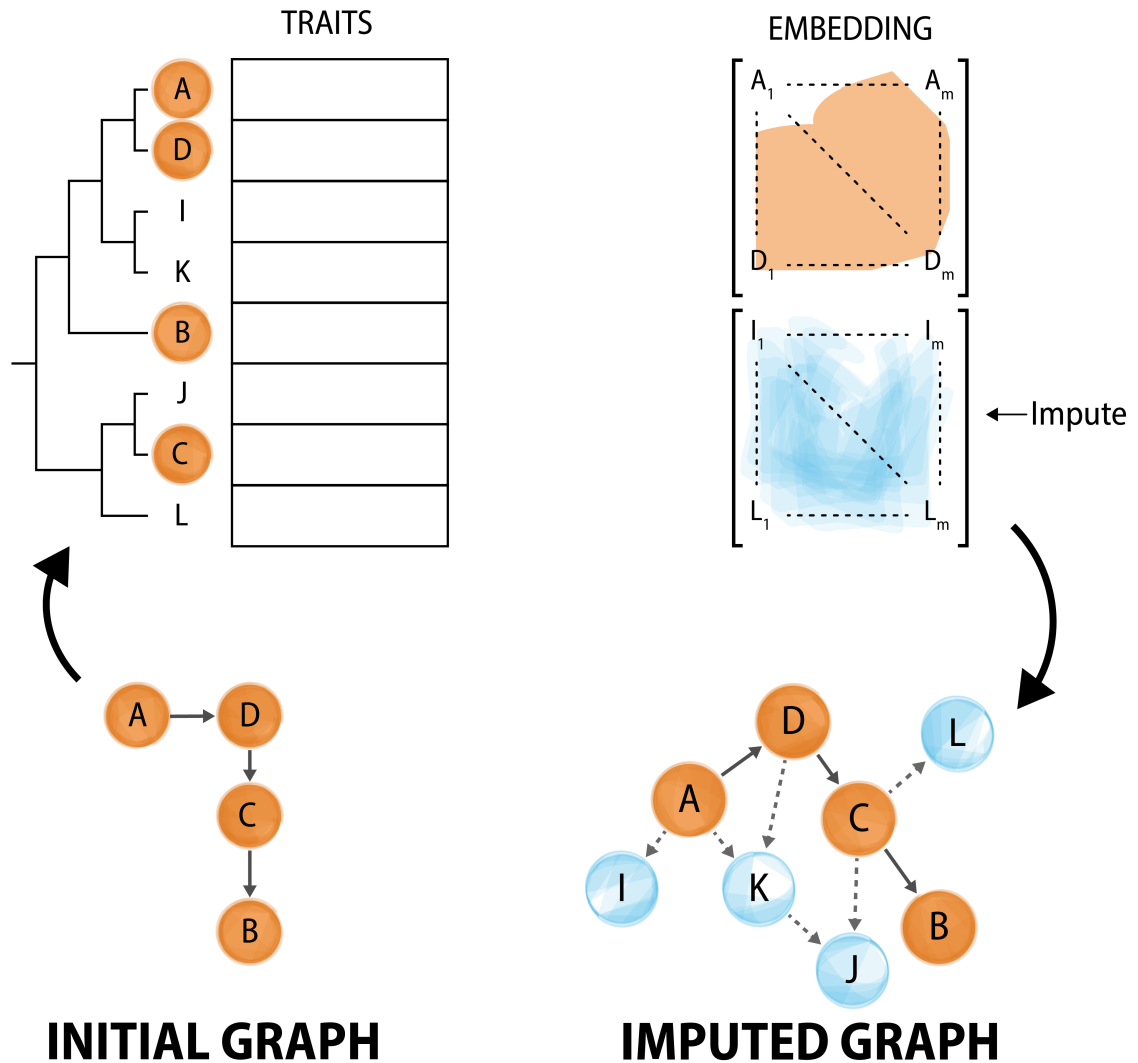


Figure 2: From a low-dimensional feature vector (see fig. 1), it is possible to develop predictive approaches. Nodes in an ecological network are species, for which we can leverage phylogenetic relatedness (*e.g.* Strydom, Bouskila, et al., 2021) or functional traits to fill the values of additional 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.