Predicting metawebs: transfer of 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`

Last revision: *January 13, 2022*

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 ecoloigcal problems.

4. In this perspective, we outline how the challenges associated with infering 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.

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

Following the definition of Dunne (2006), a metaweb is the ecological network analogue to the species pool; specifically, it iventories *potential* interactions between species from a spatially delimited area (and so captures the γ diversity of interactions). The metaweb is not a prediction of the network at a specific point within the spatial area it covers: it will have a different structure, notably by having a larger connectance (see *e.g.* Wood et al., 2015), from any of these local networks. These local networks (capturing the α diversity of interactions) are a subset of the metaweb's species and interactions, and have been called "metaweb realizations" (Poisot et al., 2015). Differences between local networks and their metawebs are due to chance, species abundance and co-occurrence, local environmental conditions, and local distribution of functional traits, among others.

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

Graph embedding offers promises for the inference of potential interactions

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

Method	Embedding approach	Reference	Application
tSNE	nodes through statistical divergence	Hinton & Roweis (2002)	Cieslak et al. (2020) [a]
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 & Madhow (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) [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)	

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

[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.

The final family of problems relates less to the availability of data or quantitative tools, and more to the praxis of spatial ecology. 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 & Goldstein, 2021), as data infrastructures tend to be designed to answer questions within national boundaries, and their use both draws upon and reinforces territorial statecraft; as per Machen & Nost (2021), this is particularly true when the output of "algorithmic thinking" (*e.g.* relying on machine learning to generate knowledge) can be re-used for governance (*e.g.* enacting conservation decisions at the national scale). We therefore recognize that methods such as we propose operate under the framework that contributed to the ongoing biodiversity crisis (Adam, 2014), reinforced environmental injustice (Choudry, 2013; Domínguez & Luoma, 2020), and on Turtle Island especially, should be replaced by Indigenous principles of land management (Eichhorn et al., 2019; No'kmaq et al., 2021). As we see AI/ML being increasingly mobilized to generate knowledge that is lacking for conservation decisions (*e.g.* Lamba et al., 2019; Mosebo Fernandes et al., 2020), our discussion of these tools need to go beyond the technical, and into the governance consequences they can have.

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

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

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

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

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

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

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

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

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

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

Galiana, N., Hawkins, B. A., & Montoya, J. M. (2019). The geographical variation of network structure is scale dependent: Understanding the biotic specialization of host–parasitoid networks. *Ecography*,

42(6), 1175–1187. https://doi.org/10.1111/ecog.03684

Galiana, N., Lurgi, M., Claramunt-López, B., Fortin, M.-J., Leroux, S., Cazelles, K., Gravel, D., & Montoya,

J. M. (2018). The spatial scaling of species interaction networks. *Nature Ecology & Evolution*, *2*(5),

782–790. https://doi.org/10.1038/s41559-018-0517-3

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

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

Gravel, D., Baiser, B., Dunne, J. A., Kopelke, J.-P., Martinez, N. D., Nyman, T., Poisot, T., Stouffer, D. B.,

Tylianakis, J. M., Wood, S. A., & Roslin, T. (2018). Bringing Elton and Grinnell together: A quantitative

framework to represent the biogeography of ecological interaction networks. *Ecography*, *0*(0).

https://doi.org/10.1111/ecog.04006

Grover, A., & Leskovec, J. (2016). Node2vec: Scalable Feature Learning for Networks. *Proceedings of the*

22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 855–864.

https://doi.org/10.1145/2939672.2939754

Grünig, M., Mazzi, D., Calanca, P., Karger, D. N., & Pellissier, L. (2020). Crop and forest pest metawebs

shift towards increased linkage and suitability overlap under climate change. *Communications Biology*,

3(1, 1), 1–10. https://doi.org/10.1038/s42003-020-0962-9

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

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

Hoffmann, J., Bar-Sinai, Y., Lee, L. M., Andrejevic, J., Mishra, S., Rubinstein, S. M., & Rycroft, C. H. (2019).

Machine learning in a data-limited regime: Augmenting experiments with synthetic data uncovers

order in crumpled sheets. *Science Advances*, *5*(4), eaau6792.

https://doi.org/10.1126/sciadv.aau6792

Hortal, J., de Bello, F., Diniz-Filho, J. A. F., Lewinsohn, T. M., Lobo, J. M., & Ladle, R. J. (2015). Seven

Shortfalls that Beset Large-Scale Knowledge of Biodiversity. *Annual Review of Ecology, Evolution, and*

Systematics, *46*(1), 523–549. https://doi.org/10.1146/annurev-ecolsys-112414-054400

Jordano, P. (2016). Sampling networks of ecological interactions. *Functional Ecology*, *30*(12), 1883–1893.

https://doi.org/10.1111/1365-2435.12763

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

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

McLeod, A., Leroux, S. J., Gravel, D., Chu, C., Cirtwill, A. R., Fortin, M.-J., Galiana, N., Poisot, T., & Wood, S. A. (2021). Sampling and asymptotic network properties of spatial multi-trophic networks. *Oikos*, *n/a*(n/a). https://doi.org/10.1111/oik.08650

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

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

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

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

Murphy, K. P. (2022). *Probabilistic machine learning: An introduction.* MIT Press. probml.ai

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

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

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

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

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

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

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

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

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

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

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

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

Wardeh, M., Baylis, M., & Blagrove, M. S. C. (2021). Predicting mammalian hosts in which novel

coronaviruses can be generated. *Nature Communications*, *12*(1, 1), 780.

https://doi.org/10.1038/s41467-021-21034-5

Wood, S. A., Russell, R., Hanson, D., Williams, R. J., & Dunne, J. A. (2015). Effects of spatial scale of

sampling on food web structure. *Ecology and Evolution*, *5*(17), 3769–3782.

https://doi.org/10.1002/ece3.1640

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

http://arxiv.org/abs/2012.08019

Yan, S., Xu, D., Zhang, B., & Zhang, H.-J. (2005). Graph embedding: A general framework for

dimensionality reduction. *2005 IEEE Computer Society Conference on Computer Vision and Pattern*

Recognition (CVPR'05), *2*, 830–837 vol. 2. https://doi.org/10.1109/CVPR.2005.170

Young, S. J., & Scheinerman, E. R. (2007). Random Dot Product Graph Models for Social Networks. In A.

Bonato & F. R. K. Chung (Eds.), *Algorithms and Models for the Web-Graph* (pp. 138–149). Springer.

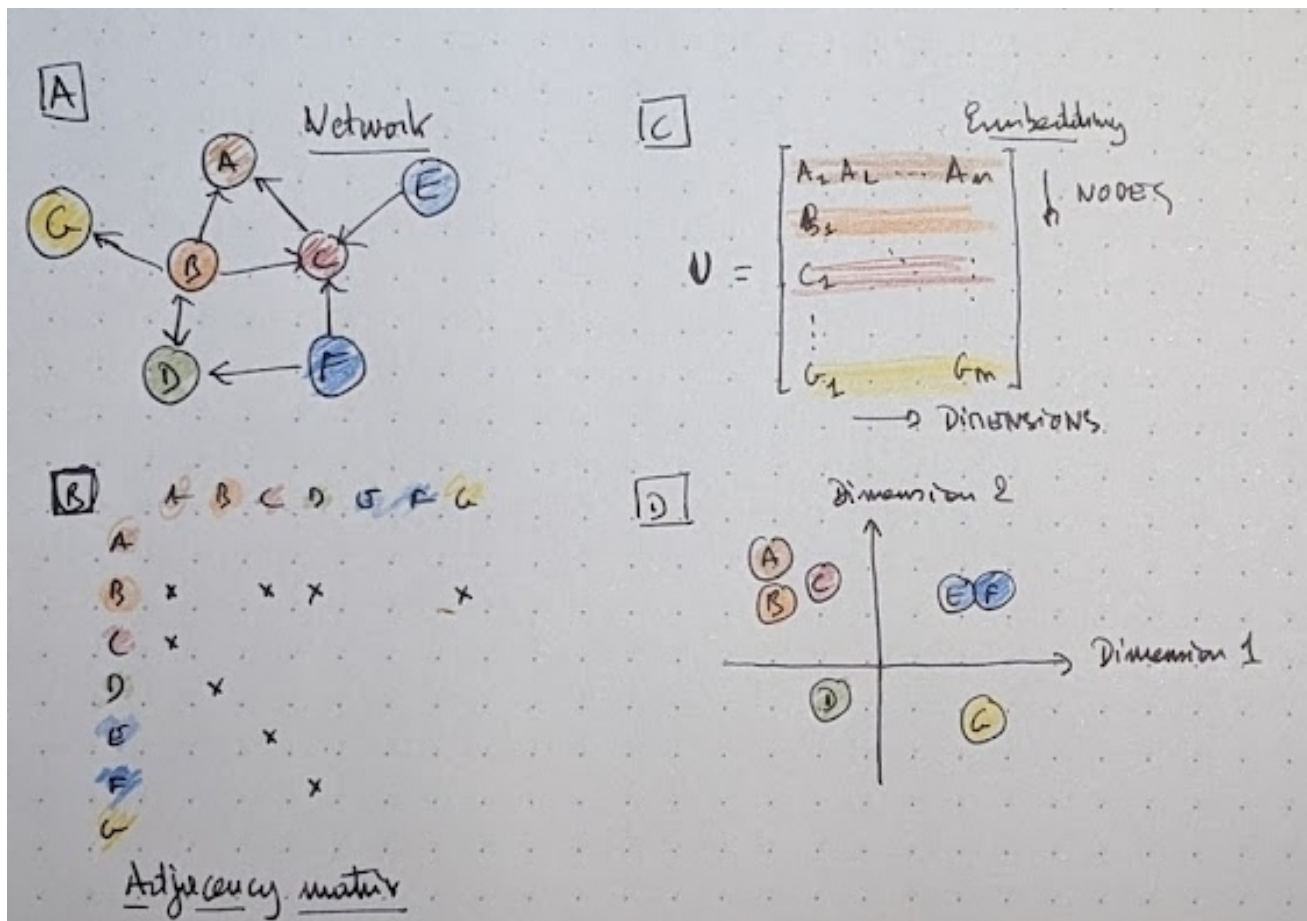https://doi.org/10.1007/978-3-540-77004-6_11

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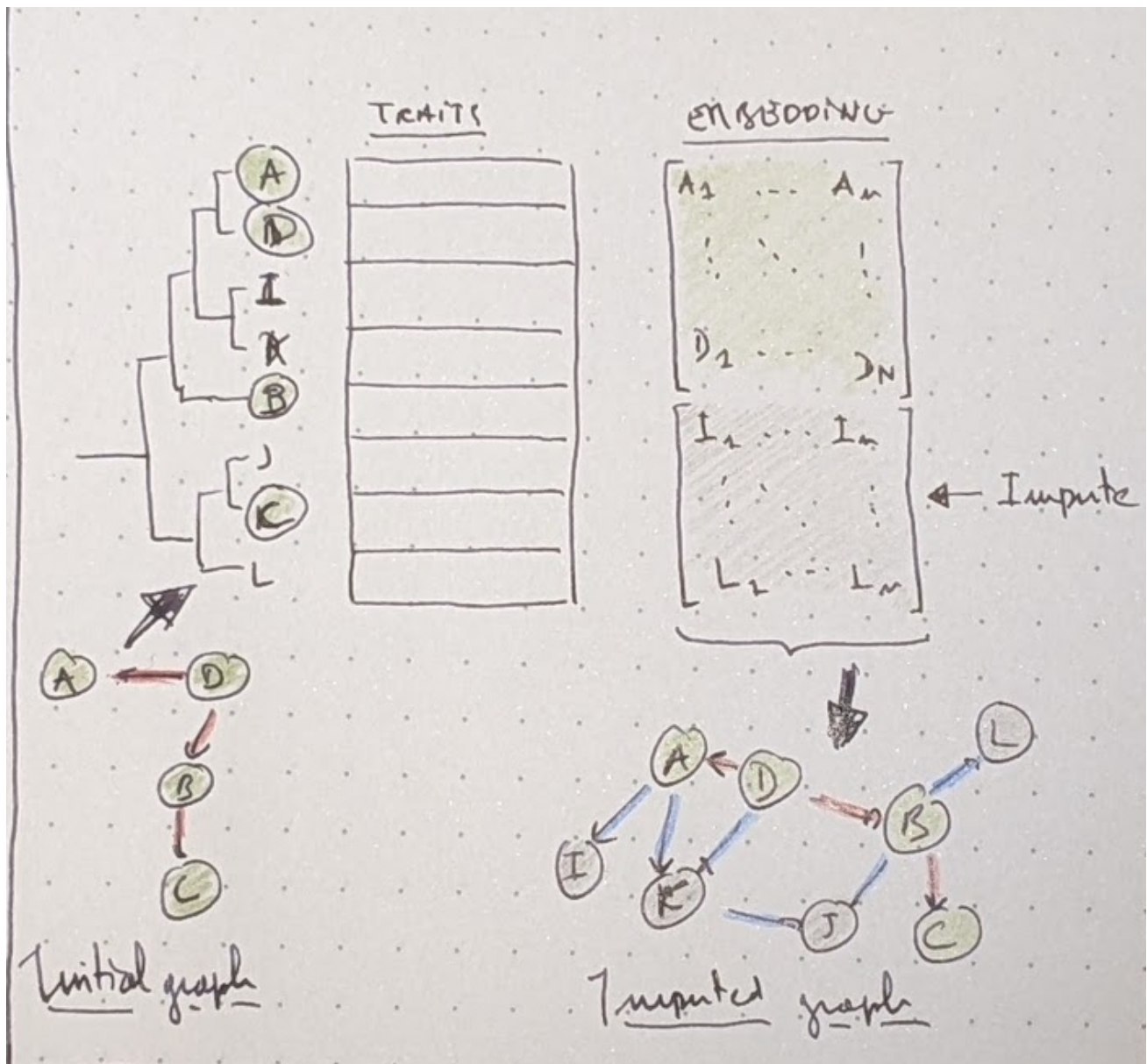
Figure 2: From a low-dimensional feature vector (see fig. 1), it is possible to develop predictive approaches. Nodes in an ecological network are species, for which we can leverage phylogenetic relatedness (*e.g.* Strydom, Bouskila, et al., 2021) or functional traits to fill the values of additional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