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 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.

Having a general solution for inferring *potential* interactions (despite the unavailability of interaction data) could be the catalyst for significant breakthroughs in our ability to start thinking about species interaction networks over large spatial scales. In a recent overview of the field of ecological network prediction, Strydom 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 predictive ability globally.

Following the definition of Dunne (2006), a metaweb is a network analogue to the regional species pool; specifically, it is an inventory of all *potential* interactions between species from a spatially delimited area (and so captures the γ diversity of interactions). The metaweb is, therefore, *not* a prediction of the food web at a specific locale within the spatial area it covers, and will have a different structure (notably by having a larger connectance; see *e.g.* Wood et al., 2015). These local food webs (which captures 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 food web and their metaweb 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, in turn, can 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). Whereas the original metaweb definition, and indeed most past uses of metawebs, was based on the presence/absence of interactions, we focus on *probabilistic* metawebs where interactions are represented as the chance of success of a Bernoulli trial (see *e.g.* Poisot et al., 2016); therefore, not only does our method recommend 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 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.

Graph embedding offers promises the inference of potential interactions

The metaweb embeds hypotheses about which spatial boundaries are meaningful

As Herbert (1965) rightfully pointed out, "[y]ou can't draw neat lines around planet-wide problems"; in this regard, our approach (and indeed, any inference of a metaweb at large scales) must contend with several interesting and interwoven families of problems. The first is the 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 to make a reliable inference decreases; this would likely be indicated by large confidence intervals during ancestral character estimation, 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. 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.

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 (see *e.g.* Galiana et al., 2018, 2019, 2021; **Fortin2021NetEco?**), 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

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

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.

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

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

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

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.

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

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

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

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

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., 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

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., 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

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

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