Senior Editor Comments to Authors

I have now received the reviewers' reports and a recommendation from the Associate Editor who handled the review process of your resubmission. Copies of their reports are included below. Based on their evaluations, I regret to inform you that we are unable to publish your paper in Methods in Ecology and Evolution in its current form.

However, we would be willing to consider another resubmission, which takes into consideration the new feedback you have received. You will see that several comments really ask for a structural rework towards the core of the "Perspective" format of MEE, which is about the stimulation of scientific debate and the offer of conceptual advances. While the ms is giving some of that, it does not convince readers and reviewers even after 2 reviews.

I think that reframing how it is presented would go a long way to appeal to readers and current & future users of graph embedding. It is not usual for us to offer the possibility to resubmit after a first "reject and resubmit' decision when the reviews are of that extent and we did so because of there is a potential for the ms to have all these comments (mentioned below) addressed. A new version including all these (numerous) comments would still need much work but would really allow the ms to be a useful roadmap towards predicting metaweb (even if that means going above the word limit).

In addition, there are several comments from the previous round that were not addressed but still created issues with the second revision; dealing with them and using the additional perspective of this new revision will help move the ms forward.

Associate Editor Comments to Authors

(There are no comments.)

Reviewer 3 Comments to Authors

General comments

The manuscript is a resubmission of a previous version that I did not review nor see. I understand that the manuscript is proposed for publication under the "Perspective" type in MEE. Besides the current version, I had access to the first reports and the authors' answers.

I think that the contents of this Perspective manuscript do not - as required by the journal guidelines - "stimulate [much] scientific debate". Neither does it significantly "offer conceptual advances or opinions or identify gaps". I'll now try to motivate this opinion.

In their answers, the authors “point out that the submitted version of this present manuscript includes” 3 different elements. It is unclear to me whether they consider this to be the complete list of the manuscript contributions.

Anyway, let me discuss the first of these elements. It is “an overview of embedding techniques and their application to species interaction networks”. I do not agree with this first claim. In my opinion, Table 1 that lists a large number of graph embedding techniques is not sufficient to be considered an overview.

The authors barely address the difference between nodes embedding and graph embedding: while I can clearly see the second column of the table, the text does not contain any sentence that could help a reader not familiar with these embedding techniques to learn about this major difference. The text is even misleading when (line 75) the authors say that “Their [the graph embedding techniques] main goal is to learn a low dimensional vector representations for the nodes of the graph (embeddings)”. An overview of embedding techniques should go beyond a list of references and come with (at least some lines of) an introduction to the methods and their main differences.

Going back to the manuscript contributions, the end of the paper’s introduction (from lines 37 to 40), sets a slightly different list: “In this contribution, we highlight the power in viewing (and constructing) metawebs as probabilistic objects in the context of rare interactions, 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.” From this, I understand that the 3 elements pointed out by the authors in their answer do not constitute the core of their contribution, which is rather described by these previous lines. This leads me to my next remark.

The previous reports raised the concern that the current manuscript “adds little to the previous published paper”. The authors answered that “there is not a single instance where areas of overlap are clearly identified”. I do identify specific places that support the reports concern:

The contents of the paper as described at the end of the introduction match the ideas underlying the previous published work;

Section “Graph embedding offers promises for the inference of potential interactions” is not only “an overview of embedding techniques” but rather it is biased towards “their application to species interaction networks”, which is the topic of the previous published paper. Graph embedding techniques are promising for ecological networks, obviously not only in the context of inferring a metaweb. This leads us back to the main concern with this paragraph: that the authors do not even have a sentence about the major difference between node embeddings and whole graphs embeddings shows that they are not interested in these techniques for ecological networks in general.

Figure 1 is titled “Overview of the embedding process”, which describes only part A and B so half of the scheme. The remainder is concerned with the method from the previously published paper.

The third element of the list pointed out by the authors in their answer is “a discussion of the remaining technical and methodological challenges associated with this approach”. Here I understood that “this” refers to “prediction through embeddings” and the title of the last section (“The metaweb embeds both ecological hypotheses and practices”) is slightly misleading as its contents are rather biased towards inference of a metaweb (after its embedding).

For the authors (line 172) “The first open research problem is the taxonomic and spatial limit of the metaweb to embed and transfer” and (line 189) “The second series of problems relate to determining which area should be used to infer the new metaweb”. These two points could have formed a perspective in the previous published paper. The last part (from line 203, “praxis of ecological research”) opens to more general considerations but again contains specific remarks related to their previous work: “Applying any embedding to biased data does not debias them” (line 206); “the need to appraise and correct biases that are unwittingly propagated to algorithms when embedded” (line 215).

The second element of the list pointed out by the authors in their answer is “a discussion of the properties of metawebs that make them amenable to prediction through embeddings”. I did not clearly identify which part of the manuscript corresponds to that element. I can only suppose that this refers to the paragraph that “a metaweb is an inherently probabilistic object”, but as far as I understand, the paragraph does not make a clear link between these properties and the amenability to prediction through embedding.

While I am convinced that “a metaweb is an inherently probabilistic object” and found interesting the part of the manuscript between lines 42 to 57, I did not understand how this is combined with the second half of this section (namely that “high quality observational data” can be combined “with synthetic data coming from predictive models” to “increase the volume of information available for inference”). More precisely, that ‘[the metaweb] fixes an upper bound on which interactions can exist’ is not clearly improved by a probabilistic version of this metaweb.

The manuscript points the need for the construction of metawebs at large spatial and taxonomic scales. The authors are not specific about what “large” is exactly. It would be interesting to be more specific on that or provide some examples. Is this a world-wide scale? A continental scale? Any scale for which aggregation of local data is necessary? Anything else? Line 192 (and below) appears the mention of “country level”; However countries are too heterogeneous in their sizes to answer my point. Also, the term ‘continental scale’ appears on line 217 but in a specific sentence and I am not convinced that this is exactly what the authors have in mind when mentioning “large” scales.

The abstract contains the sentence (point 4) “[we] discuss how the choice of the species pool has consequences on the reconstructed network”. This is indeed an interesting question. But I did not see anything in the text that could refer to this.

Minor comments

line 10: “accurate predictors are important for accurate predictions”. Indeed, but what is your point?

line 13: replace GBIF and UICN by the full names.

line 73: “Graph (or Network) embedding (fig. 1)”. You should modify the reference to “fig1. A, B” because the rest of the figure is not part of the embedding process. By the way, the caption of figure 1 “Overview of the embedding process” is also misleading as (again) this title only describes half of the scheme.

I understand that the paragraph on GNN (from line 88) was added to answer a referee concern, but it is disproportionate: you use as much space not to speak about GNN as to speak about (ML) graph embeddings.

I understand that the illustration of metaweb embedding was added to answer one of the referees of the first round. Nonetheless, I do not see the added value of it.

On line 147, the authors claim to see “an inflection point around 25 dimensions”. I do not see any inflection, but I understand this is a reasonable compromise.

line 219: “Particularly on Turtle Island and other territories”. I did not understand why in a very general paragraph you refer “particularly” to this specific example. Maybe “for example” would be more suited.

Reviewer 4 Comments to Authors

General comments

The paper summarized the key challenges of inferring metawebs based on graph embedding approaches. It also highlighted the significant advantages of using graph embedding and transfer learning techniques for species interaction network prediction and other ecological problem applications. The paper provided a very important research direction of applying advanced graph embedding and transfer learning to tackle diverse inference tasks for species interaction networks.

Two main questions about this paper are listed below.

1) Fig.1 is a good diagram that shows the whole pipeline with the input graph adjacency matrix, output graph embedding and combined with transfer learning technique. I would also suggest the authors to include some experimental results based on graph embedding and transfer learning for specie interaction inference with real dataset.

2) Please update the article information for the paper (Xu, M.. Understanding graph embedding methods and their applications. SIAM Review, 2021) in the reference section.

Reviewer 5 Comments to Authors

General comments

This manuscript consists in a revised version (re-submission) of a perspective paper dedicated to the potential contribution of graph embedding to metaweb prediction.

The authors provided a pedagogical illustration on a host-parasites system, aiming at predicting (in a probabilistic way) the links of this bipartite network using Random Dot Product Graph embedding. This illustration implements key elements of Fig. 1 and might invite the reader to use or develop the embedding framework on various datasets. The figure showing the decrease of the loss with the rank of the embedding is particularly welcome since it shows to what extent network structure can be reasonably summarised in few dimensions using a specific embedding.

However, I got a bit surprised not to find the ecological interpretation of this embedding in terms of response and effects traits. I think the authors should better try to link ecological theory in general and ecological hypothesis associated with machine learning methods throughout the manuscript. I understand that the manuscript is centered on methods to predict metawebs with somehow incomplete sampling or knowledge. Consequently, as many research papers on applied machine learning in ecology, ecological hypothesis and theory are a bit behind the scene.

I think that a perspective paper should clarify possible ecological interpretation of machine learning methods. If the manuscript is clear and enlightening on the probabilistic metaweb approach, ecological hypothesis associated to link predictions are much more obscure. The following points should be somehow addressed in the manuscript to get additional perspectives:

What are the interpretations of Random Dot Product Graph embeddings in terms of latent traits ?

What are the hypothesis behind link prediction using other information (traits, phylogeny as in Strydom et al 2022) ?

What structures are considered in the different embeddings ? Table 1 mentions several embedding techniques and separates node from graph embedding. However, even two node embeddings algorithm can have different interpretations. For example, tsne is based on neighbor (local structure) whereas node2vec relies on random walks, so paths in the network (global structure). Both could be interpreted in ecological terms. Predicting metawebs with these two methods do not hold the same hypothesis on species interactions. How does it affect potential applications ? I think Table 1 must be enriched more (maybe add a figure or split it) in order to be a roadmap and not simply a catalog. It should mix and bridge embeddings, ecological interpretations and even some illustrations to guide the reader in this machine learning jungle. Such clarification should be also present in the manuscript

To what extent co-occurence data should be considered in embedding approaches ? Co-occurence is considered for statistical association networks but probably not in the same manner than the deep learning model of (Strydom et al 2021). So, to what extent co-occurence data can be used to predict interaction ? Table 1 mentions statistical association methods and JSDM. The authors should be more clear in Table 1 and in the manuscript on the link/differences between association methods (that predicts associations from co-occurence) and link prediction using embeddings. The term statistical association is present in Table 1 but not in the text, it can be quite confusing for the reader.

I think this manuscript could be considered for publication in Methods in Ecology and Evolution but it must offer broader perspectives than Strydom et al. 2022. To do so, the authors should try to address somehow the previous listed points.

Minor comments

L20: Local networks capture alpha-diversity of interactions but also beta diversity since interactions from the metaweb can be absent from local networks species absence.

L24: Yes, I agree that Saravia et al. 2021 shows that local network structure (represented by network metrics blind to species identity) does not differ from the one expected from a null model (Trophic Theory of Island Food webs). However, the manuscript focuses on alpha-diversity metrics. In terms of species and interaction composition, they can still differ even if they have more or less the same structure. It think this point should be clarified. Indeed, otherwise, it somehow states that we do not need to focus on local composition since local networks have the same structure as the metaweb.

L63: Does it "generate" or uncover the core rules associated to species interactions ? I do not really understand the point of "generating rules". For me, statistical methods try, using abstract representations, to uncover/formulate biological rules.

L79: Yes, embedding methods can exhibit structural invariants in ecological networks. However, what characteristics of the networks should be considered in the model (links,paths,motifs…)?

L83: If the choice of the embedding matters for the result, why not providing to the reader some sort of roadmap of these embeddings techniques for different applications ?

L85: For the moment, Table 1 is more a catalog than a roadmap.

L86: Ok, here comes different network representations (latent traits or random walks). How do we interpret these methods in ecological terms ? Which one is the most suitable in the various potential applications.

L99: Bohan et al. 2017 mentions several associations methods to build networks. If I am correct, these associations methods are not mentioned in the manuscript. The authors must clarify this point.

L115: The authors should clarify the ecological hypothesis associated to such approach and the potential limits of predicting interactions using node metadata.

L134: I think this illustration is relevant but it must be enriched in order, for example, to compare different embedding or incorporating meta-data. To what extent will different embedding techniques provide different networks ? Will the loss with the rank be similar if you use link (as in the RDGP model) or path based embedding ? Moreover, Stochastic Block Model can also but used to perform link predictions (Gaucher et al. 2019; Link Prediction in the Stochastic Block Model with Outliers. stat.) and is related to RDGP, I think it deserves a mention here.

L136: What are the ecological hypothesis associated to RDGP model in terms of response and effect traits ?