February 28, 2025

Grant Foster
Coker Life Sciences Building
715 Sumter Street
Columbia, SC 29208, USA
Email: fostergt@email.sc.edu

Dr. Pedro Peres-Neto
Editor-in-Chief
Oikos

Dear Dr. Peres-Neto,

We appreciate the time and thoughtful comments of all three reviewers of this manuscript, and believe their feedback has been helpful for improving the overall quality and clarity of our work. At their core, we believe that some of the concerns raised by reviewer 3 come from the standpoint of choosing a single "best" approach to predict links. We argue that this is an important perspective, but believe it has been more thoroughly explored by other papers than ours (including by Nunes-Martinez and Pires [2024](1) using the same dataset). We hope that our comparative approach instead better addresses how three alternative different sources of predictive information: phylogeny, species traits, and latent network features, differ in their predictive capacity and relative importance when predicting different sorts of ecological links. The suggested methods of phylogenetic imputation based on latent features after Strydhom (2022)(2), while powerful for making predictions, confounds latent features with a presumed phylogenetic signal, which prevents us from answering our central research question.

The reviewer also raised methodological concerns, particularly regarding the potential overestimation of model performance due to inflated AUC values caused by an overrepresentation of 0's (non-interactions) in the dataset. We used a relatively strict class-balancing scheme of enforcing a 3:1 ratio in our training data in the main text to control for this impact of sparseness. In response to these concerns, we further repeat our analysis with alternative training prevalence ratios of 1:1 and 1:10 and directly compare these to our main text results in new supplemental figures S5, S6, and S7. We find that while altering the training set prevalence values (within a reasonable range) does alter the absolute predicted suitability values as expected (S5), there is no consistent effect on overall model performance (S6) or most importantly variable importance (S7) across any of our models.

Finally, the reviewer raised concerns about potential information leakage between the training and test sets in models that utilize latent features. As information from the entire interaction set is used to create our latent features, even when we remove large amounts of records from the training set, they may still ostensible leave a potential amount of "signal" in the remaining training values. The method of using SVD-derived features for prediction is shared by a number of other studies, including machine learning approaches looking to predict interactions outside the realm of ecology. These include predicting drug-disease interactions (3), non-coding RNA sequences and disease (4), and potential interactions within gene networks (5). Despite this precedent, we were similarly concerned about the potential for information leakage during the design of our analysis, and it ultimately led to the two-part analysis present in our main text. In our main text initially compare potential model performance between models trained on test and training sets, before then repeating the analysis with the entire dataset to compare variable importance across models. We believe the congruence of our results across these sections, both in terms of model performance and variable importance (S7), speaks to the validity of our model training approach.

We appreciate the editor for allowing us to respond to these concerns, as well as believe the reviews provided have made the manuscript stronger as a whole. Below we more fully respond to these concerns line by line. Comments provided by the reviewer are highlighted in blue. Additionally, we have provided an updated version of our manuscript and supplemental material below that incorporates the comments made by this and other reviewers.

Reviewer 3's Comments to the Author:
In this manuscript the authors test whether plant-frugivore interactions can be predicted from phylogenetic structure, species traits and latent structure. The main result is that latent structural features are important to predict the interactions. The main issue here is that there is some circularity in the analysis. Latent structure is obtained from the decomposition of the interaction matrix. The decomposition of the matrix generates axes that summarize the information in the matrix. So, it is expected that those axes will perform well when predicting the interaction matrix. This technique has been used recently to estimate interactions, but combined with other steps, such as

Also, because the matrix is sparse other models seem to perform well because they basically have to predict the zeroes. Any model that generates a sparse matrix will have high accuracy. Looking at table 2 we see for instance that the trait only model has a considerably lower rate of true positives (almost five times lower than other models) and high rate of false negatives. I know this dataset well and the main problem with it is that most species are highly undersampled. You cannot predict matrix structure based on traits or phylogeny when sampling is such an important factor generating interaction imbalances between species. The phylogenetic structure the authors find is most likely a mix between true phylogenetic signal in interaction patterns and phylogenetic signal in sampling. One way to address that is to use only well sampled species to train the model and then estimate interactions for under-sampled species (as in Nunes-Martinez and Pires 2024). Moreover, although the authors justify their study based on the need to predict interactions, they offer no guidance on how their analysis could be used to estimate interactions.

Comment 1: Ln 96. and seed dispersal interactions
Response: We have incorporated this text and its associated citation to Nunes-Martinez and Pires (2024) on this line.

Comment 2: Ln 114. Because the Atlantic frugivory database includes species from the entire Atlantic Forest, many of the plant-frugivore interactions are forbidden simply because species do not co-occur. If that is not taken into account in the model, it is unlikely to perform well. How did you deal with that in the trait based model? Could you incorporate sampling in the trait model?
Response: We agree that both 1) the presence of interactions forbidden by current species distributions, as well as 2) the potential effects of differential sampling are important features to consider when working with this dataset. We attempt to respond to the first point below, and further discuss the second point in our answer to comment 5.

For the purposes of prioritization of targeting sampling efforts of potential interactions, we argue withholding direct information on the role of geographic is important, especially when considering interactions that may occur in future climatic or biotic contexts. Two species that ostensibly have the potential to interact may fail to do so due to a variety of factors, including separation across space or time. However, as species geographic ranges or phenological periods shift, interaction networks may potentially rewire to include interactions that were previously forbidden. If we can uncover factors associated with the potential for interaction, we can better make predictions about how future changes may potentially drive changes in network structure, a goal we think is particularly salient in the context of global change. In the context of geographical constraints, truly "forbidden" links are only forbidden under current climatic and biotic contexts. If we were maximizing our ability to predict links according to current climatic conditions and range distributions alone, we agree that including information about contemporary distributions would likely increase model performance. However, in pursuing the question of how these different information types compare and overlap in predictive capacity broadly, we argue that a metaweb approach of predicting the potential broadest set of possible interactions is more informative. Any given local network will be a subset of this metaweb, with the realized set of interactions being determined by local assembly processes. This general approach is shared by a number of other studies predicting potential interactions across large spatial scales (6; 2; 7).

Comment 3: Ln189. High AUC can be obtained just because the matrix is sparse.
Response: It is true that in high AUC values can be inflated by class-imbalanced data, especially when researchers fail to address these class-imbalances in the training dataset for a give model. In our main-text presentation of our results, we attempt to control for this factor by dropping 0's from each training set until we achieve a 25% prevalence of positive associations, compared to the 3.7% observed in the empirical network (see lines 166-169 for details). More broadly however, we hope to emphasize the relative performance of our models in comparison to one another over their absolute predictive power, as well as how the same variables change in importance across models. To investigate the sensitivity of these features to our choice of training prevalence ratios, we repeated the first section of our analysis with alternative ratios of either 1:1 or 1:10. We have added 3 corresponding figures to the supplement of our paper exploring the results of these analysis.

Figures S5 represents pairwise comparisons of the predicted suitability of all links in models trained with either 1:3 or 1:1 training prevalence ratios. We see that increasing the prevalence of 1's in the dataset does increase the absolute value of predicted suitabilities (especially those with high uncertainty), but has very little impact on the relative ranking of potential links (Spearman's $\rho = 0.964$ across all predictions). Figure S6 shows the distribution of overall AUC values across the 7 models presented in our main text when trained on tests sets with alternative prevalence ratios, showing no significant impact on overall model performance. Finally, we also include figure S7, which shows pairwise relative variable importance across models trained on either a 1:3 or 1:1 prevalence ratio. The majority of points on this plot fall very close to a 1:1 line, indicating very little effect of choice of training ratio on variable importance. There is particularly strong agreement between the most important variables for each model,

as shown in the top-right quadrant of each subplot, further indicating no serious difference in variable importance across training ratios. All in all, we hope these results help assuage reviewer concerns that the performance of our models are being artificially inflated due to sparsity.

Comment 4: 237. latent traits can be used to predict interaction patterns if you use a protocol to obtain latent traits from an input matrix and then estimate the latent traits for species with limited information as done by Strydon et al. (2021) and Nunes-Martinez and Pires (2024). You only have information on latent traits if there is an interaction matrix, different from phylogenetic information and traits, which are independent from the interactions.
Response: With the goal of maximizing prediction performance, we agree that approaches that leverage multiple information types (such as methods that phylogenetically impute latent features as utilized in Strydon et al. (2021) and Nunes-Martinez and Pires (2024)) are powerful approaches. However, we argue that our purpose here is to compare the information and subsequent predictive performance of three alternative sources of information that could affect species interactions: traits, phylogeny, and latent network features. Aforementioned combinatorial approaches are useful ways to deal with the differences in sampling across species, by assuming a phylogenetic signal to the latent features that may govern interaction probabilities, and is an insightful way to improve overall prediction. However, the phylogenetic conservation of the signal is a serious assumption, and furthermore the approach makes it impossible to separate the effects of phylogenetic and latent features on interaction probability. We submit this paper with the goal of comparing these different information types in their ability to predict interactions, both jointly and in concert, and as such believe utilizing an approach after Strydon et al. 2021 is somewhat antithetical to our aims.

Comment 5: Ln 273. if all interactions are correctly predicted this assumes that all interactions (and "non-interactions") are well known, i.e., sampling is assumed to be perfect, and we know this is not true. Therefore, the models are most likely missing a lot of interactions, and you are overestimating their true performance.
Response: To the second point, we acknowledge that incomplete sampling is a serious limitation this dataset, and indeed nearly all datasets looking to empirically observe species interactions. Species with low abundance or detectability are by their nature difficult to sample on their own, and their interactions with similarly rare or low abundance partners are multiplicatively more difficult to sample. We employ cross-validation techniques that evaluate model performance on holdout test sets, a consistent approach used in many preceding link prediction studies (8; 9; 10; 11). However, when applying our methods to the entire network, we are in many ways most interested in where our models seem to be "wrong" according to our known set of interactions. Many of the interactions with high predicted suitability but that do not occur in our dataset may actually exist in wild communities. We hope that link prediction methods like the ones we use here can help fill in some of these gaps through guiding future sampling efforts.

References

[1] Nunes Martinez A, Mistretta Pires M. 2024 Estimated missing interactions change the structure and alter species roles in one of the world's largest seed-dispersal networks. *Oikos* p. e10521.

[2] Strydom T, Bouskila S, Banville F, Barros C, Caron D, Farrell MJ, Fortin MJ, Hemming V, Mercier B, Pollock LJ et al.. 2022 Food web reconstruction through phylogenetic transfer of low-rank network representation. *Methods in Ecology and Evolution* **13**, 2838–2849.

[3] Wu G, Liu J, Yue X. 2019 Prediction of drug-disease associations based on ensemble meta paths and singular value decomposition. *BMC bioinformatics* **20**, 1–13.

[4] Zeng M, Lu C, Zhang F, Li Y, Wu FX, Li Y, Li M. 2020 SDLDA: lncRNA-disease association prediction based on singular value decomposition and deep learning. *Methods* **179**, 73–80.

[5] Yeung MS, Tegnér J, Collins JJ. 2002 Reverse engineering gene networks using singular value decomposition and robust regression. *Proceedings of the National Academy of Sciences* **99**, 6163–6168.

[6] Hao X, Holyoak M, Zhang Z, Yan C. 2025 Global Projection of Terrestrial Vertebrate Food Webs Under Future Climate and Land-Use Changes. *Global Change Biology* **31**, e70061.

[7] Dansereau G, Barros C, Poisot T. 2024 Spatially explicit predictions of food web structure from regional-level data. *Philosophical Transactions B* **379**, 20230166.

[8] Strydom T, Catchen MD, Banville F, Caron D, Dansereau G, Desjardins-Proulx P, Forero-Muñoz NR, Higino G, Mercier B, Gonzalez A et al.. 2021 A roadmap towards predicting species interaction networks (across space and time). *Philosophical Transactions of the Royal Society B* **376**, 20210063.

[9] Dallas T, Park AW, Drake JM. 2017 Predicting cryptic links in host-parasite networks. *PLoS computational biology* **13**, e1005557.

[10] Stock M, Piot N, Vanbesien S, Meys J, Smagghe G, De Baets B. 2021 Pairwise learning for predicting pollination interactions based on traits and phylogeny. *Ecological Modelling* **451**, 109508.

[11] Kendall LK, Rader R, Gagic V, Cariveau DP, Albrecht M, Baldock KC, Freitas BM, Hall M, Holzschuh A, Molina FP et al.. 2019 Pollinator size and its consequences: Robust estimates of body size in pollinating insects. *Ecology and Evolution* **9**, 1702–1714.