S.M.1 - SVD does not overfit on the European network

In order to ensure that the creation of the RDPG on the European network does not lead to overfitting, we performed two numerical experiments.

First, we estimated the threshold that separates interactions from non-interactions based on a decreasing amount of species; this highlights that removing up to 50% of the total species in the network does not change the estimate of the threshold, suggesting that there is an important amount of information contained in the first 12 ranks of the network.

Second, we extracted $\mathcal{L}$ and $\mathcal{R}$, the left and right subspace of the entire network, at rank 12. Then, for every number n of interactions between 10 and $\text{links}(M) - 1$ (where M is the European metaweb), we define m as a network in which n interactions have been either randomly removed, randomly added, or both. We then define $\updownarrow$ and ∇ as the left and right subspaces coming from the rank-12 RDPG applied to this network, and compare the original network M to the one that was reconstructed after thresholding $\updownarrow\nabla$ by picking the cutoff that maximizes Youden's J measure.

This last experiment allows measuring the response of various measures of fit of the binary classifier to incomplete sampling. We are specifically interested in (i) the ability of RDPG to identify modified interactions, (ii) the ability of RDPG to function as a performant classifier in the presence of uncertainty in the original data, and (iii) the ability of RDPG to reconstruct biologically realistic data when interactions are withheld.

Threshold estimation is robust to species sub-sampling

In the initial experiment, we witheld an increasing number of species from the European metaweb, ranging from 20% for training and 80% for validation, to 90% for training and 10% for validation. Surprisingly, the estimation of the threshold, here presented as the mean and standard deviation of 50 folds for validation, is remarkably robust (and matches the value obtained using the entire network, as a dashed line). Specifically, even using 60% of species to estimate the threshold gave on average the same threshold as would be estimated based on the entire network; therefore, this establishes that the decision in the main text to use the entire European metaweb to set the threshold is correct.
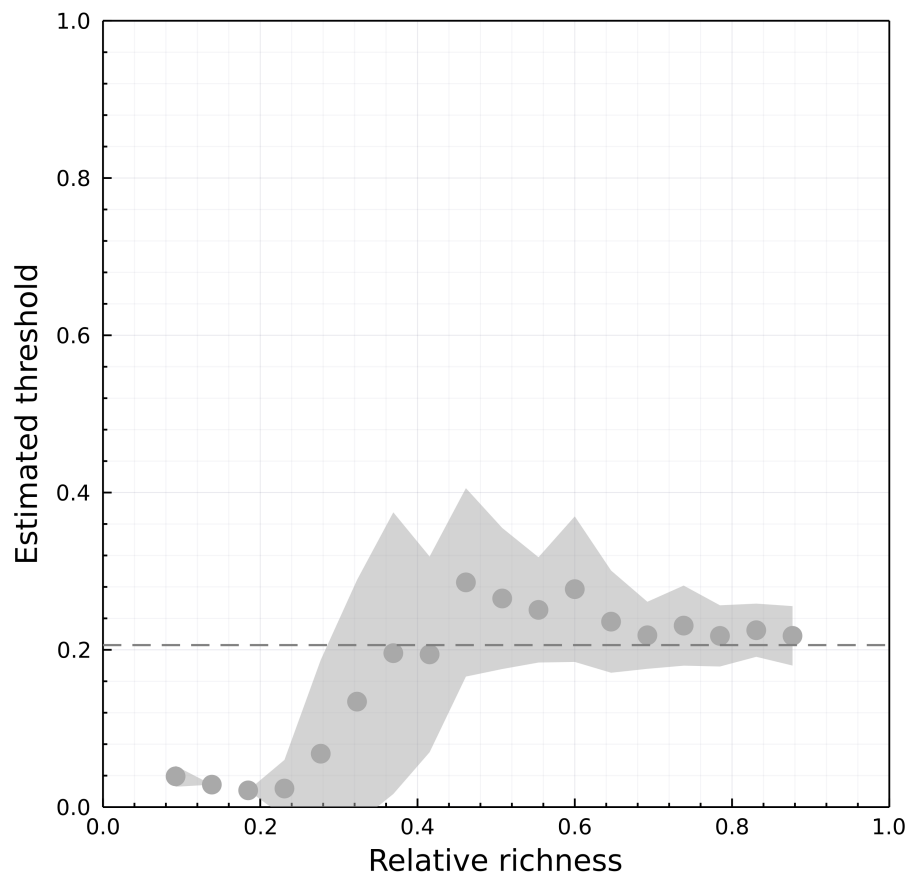
More strikingly, looking at the rates of true/false positive/negative, as illustrated below, it is clear that RDPG can be thresholded in a way that yields an almost perfect classifier:
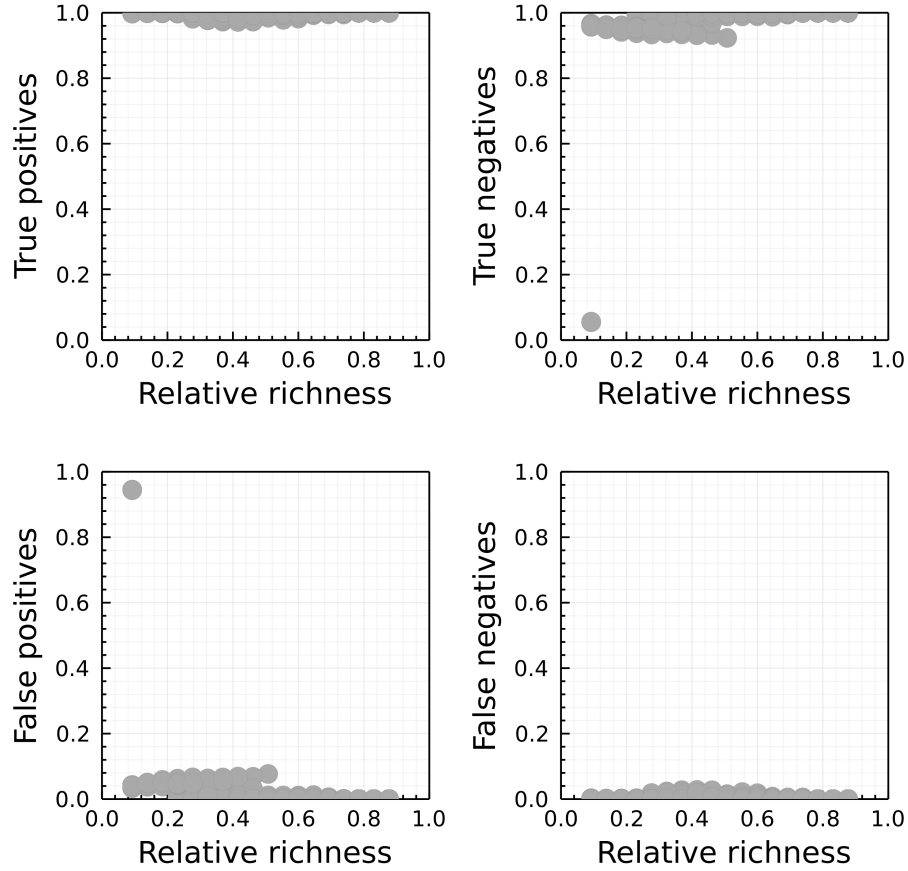
These results may be surprising, given that ecological models usually do not reach this degree of accuracy. That being said, we use the first 12 ranks of the network to approximate it, and this contains a lot of information; in short, the minute discrepancies between the predictions and the actual data can be attributed to leftover noise in the original dataset.

RDPG recovers withheld interactions

RDPG is able to correct almost all *added* interactions, which is very strong evidence that the metaweb produced using it are not going to contain too much spurious interactions. When *removing* interactions, even when half are missing, RDPG was able to accurately reconstruct about 75 to 80% of them. Predictibly, the performance when both adding and removing interactions is in between the two scenarios.
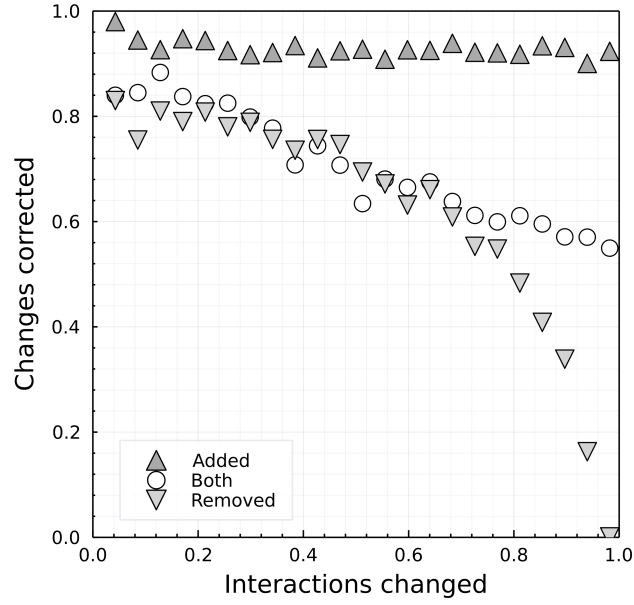
The stochasticity in the proportion of recovered interactions is larger when a small number of interactions are withheld, which makes sense as the *number* of interactions is far smaller (compared to the overall network size).

Next , it is interesting to note that the threshold "adapts" to the amount of missing information - the dashed line corresponds to the threshold we used in the manuscript. Adding interactions specifically did not result in an increase in the threshold, further suggesting that RDPG is extremely good at removing spurious interactions.
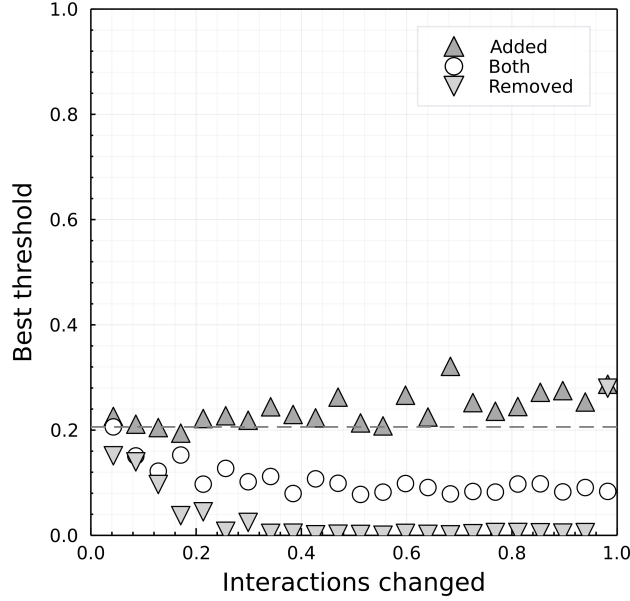
The important consequence of this result is that training the RDPG on a sub-sample of the network (*i.e.* one missing interactions) would result in a lower threshold, thereby potentially creating more false positives when applied to novel data; this further justifies our decision to use the entire evidence to estimate the threshold.

RDPG yields an accurate classifier

More important than the recovery of removed interaction is the fact that the classifier should have a good global performance. One measure to assess this is the area under the receiving operator characteristic curve, or ROC-AUC. By this measure, the RDPG remains an excellent classifier even if 50% of interactions are withheld, and no matter what the amount of changes are made by adding or both adding and removing interactions.
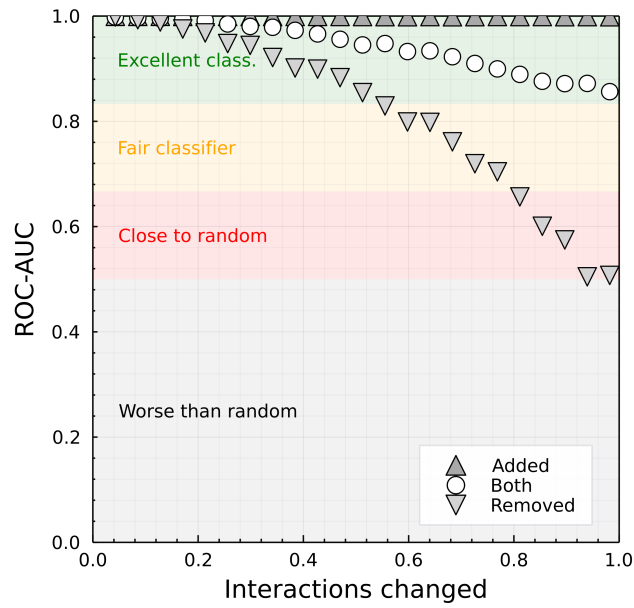
The overall agreement between a classifier and the actual data can be measured by Cohen's κ, which gives a similar result.
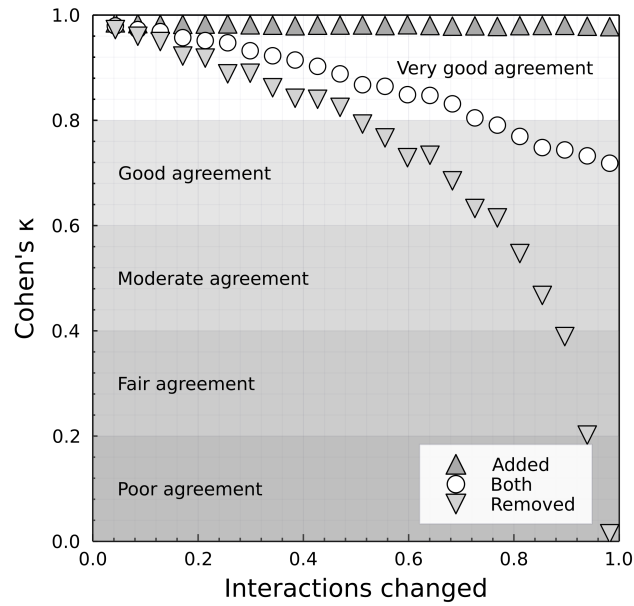
These two diagnostic figures reveal that, although we used a probably exhaustive list of interactions to do the initial RDPG, there are chances that the approach would work on less complete datasets.

RDPG recreates ecologically realistic networks

In this section, we present the relationship between the empirical measure of the network structure (dashed line) and the reconstructed estimate based on RDPG after the optimal threshold has been applied. We focus on connectance (for its broad relevant to food web structure) first:
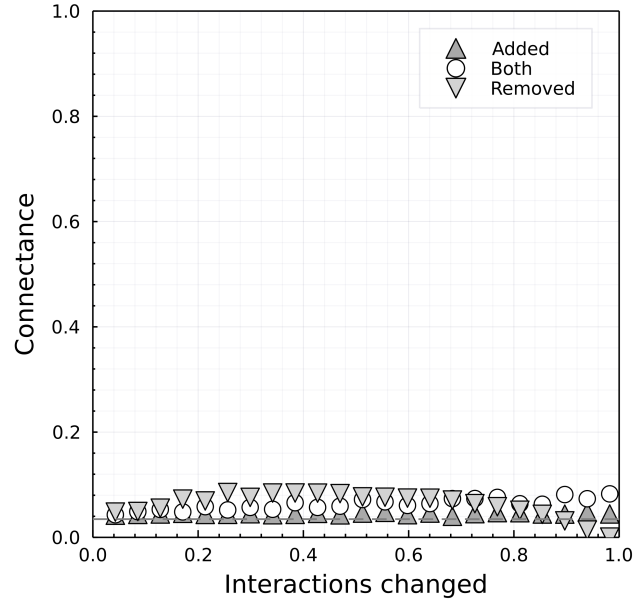
Connectance increases slightly when initial information is incomplete, but saturates at a value of around 0.12 – this is still within the bounds of connectances expected for food webs.

Next, we look at the ratio between direct competition ($a \rightarrow (b, c)$) and apparent competition ($(a, b) \rightarrow c$) motifs, as motifs are known to be conserved blocks in food webs:
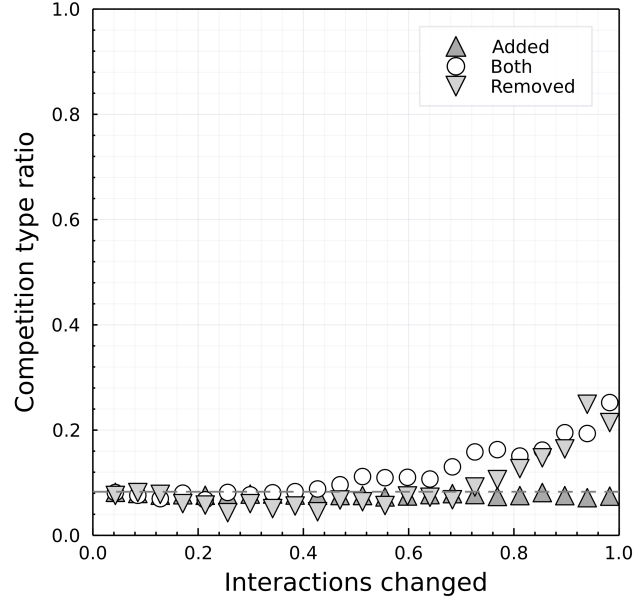
This ratio remains close to the real one up until 75% of initial interactions are modified.

Consequences

Based on these results, applying RDPG on the entire European network is reasonable, especially since (i) the threshold is insensitive to the number of withheld species, and (ii) removing interactions would artificially lower the threshold. Interestingly, the RDPG remains an excellent binary classifier even in the face of strong data modifications, which suggests that our framework can be used even in the absence of a complete metaweb. Even more importantly, the addition of wrong interactions to the original dataset was never an issue for the RDPG classifier, which was almost always able to remove them.