DR GORDANA POPOVIC (Orcid ID : 0000-0002-1376-1058)

PROFESSOR DAVID IAIN WARTON (Orcid ID : 0000-0001-9441-6645)

DR ANGELA T MOLES (Orcid ID : 0000-0003-2041-7762)

Article type : Research Article

Handling editor: Dr David Murrell

Corresponding author mail-id: g.popovic@unsw.edu.au

Untangling direct species associations from indirect mediator species effects with graphical models

Gordana C. Popovic[*], David I. Warton [*], Fiona J. Thomson[†], Francis K. C. Hui[‡], Angela T. Moles[§]

[*]School of Mathematics and Statistics and the Evolution & the Ecology Research Centre, UNSW Sydney, NSW 2052, Australia
[†] Manaaki Whenua Landcare Research, Lincoln, 7640, New Zealand
[‡]Research School of Finance, Actuarial Studies & Statistics, Australia National University, Acton, ACT 2601, Australia
[§]School of Biological, Earth 0061nd Environmental Sciences & the Evolution & the Ecology Research Centre, UNSW Sydney, NSW 2052, Australia

Header: Direct species associations from graphical models

Word count: 6723 words

Summary

1. Ecologists often investigate co-occurrence patterns in multi-species data in order to gain insight into the ecological causes of observed co-occurrences. Apart from direct associations between the two species of interest, they may co-occur because of indirect effects, where both species respond to another variable, whether environmental or biotic (e.g. a mediator species).
2. A wide variety of methods are now available for modelling how environmental filtering drives species distributions. In contrast, methods for studying other causes of co-occurence are much more limited. “Graphical” methods, which can be used to study how mediator species impact co-occurrence patterns, have recently been proposed for use in ecology. However, available methods are limited to presence/absence data or methods assuming multivariate normality, which is problematic when analysing abundances.
3. We propose Gaussian copula graphical models (GCGMs) for studying the effect of mediator species on co-occurence patterns. GCGMs are a flexible type of graphical model which naturally accommodates all data types, *e.g.* binary (presence/absence), counts, as well as ordinal data and biomass, in a unified framework. Simulations demonstrate that GCGMs can be applied to a much broader range of data types than the methods currently used in ecology, and perform as well as or better than existing methods in many settings.
4. We apply GCGMs to counts of hunting spiders, in order to visualise associations between species. We also analyse abundance data of New Zealand native forest cover (on an ordinal scale) to show how GCGMs can be used analyse large and complex datasets. In these data, we were able to reproduce known species relationships as well as generate new ecological hypotheses about species associations.

INTRODUCTION

The study of how and when species co-occur is central to understanding large scale patterns in biodiversity, ecosystem services such as pollination and seed dispersal, and in helping scientists to predict how communities might re-assemble in response to climate change. There are three important drivers of co-occurrence: (1) Two species may cooccur because they both respond to the same (or similar) environmental variables. For example, they might both be more abundant in warmer climates (Hawkins et al., 2003). (2) The two species may both respond to the presence or abundance of another species (hereafter referred to as *mediator species*), for example, they may both be hunted by the same predator (Gilinsky, 1984), but have no direct association between them. (3) The two species might have a direct association, such as facilitation, seed dispersal, or pollination (Freestone, 2006).

Model–based methods for quantifying the effect of environmental variables on species' occurrence and abundance are relatively well-established. The impact of measured environmental variables can be investigated using multivariate extensions of generalised linear models (GLMs), like in Pollock et al. (2014) and Hui et al. (2013). The impact of unmeasured environmental variables can be inferred using generalised linear latent variable models for multivariate count and biomass data in ecology (Warton et al., 2015; Ovaskainen et al., 2017). Covariance matrices obtained after controlling for environmental variables using these methods can be used to deduce if species co-occur more or less frequently than predicted by the environment.

Simulation based methods for investigating species associations in ecology are most commonly based on null models (Gotelli & Ulrich, 2010; Strong Jr et al., 2014; D'Amen et al., 2017). These aim to test if matrices of species co-occurrence or abundances are consistent with a null hypothesis of no association (sometimes controlling for environmental gradients). They can also be used to investigate pairwise species association (Ulrich, 2008).

The above methods do not take into account whether species co-occur due to the presence or abundance of mediator species, that is, they cannot be used to differentiate between driver (2) and (3) of co-occurrence. The consequence of this can be far ranging, affecting estimates of species' ranges, co-distribution of species, and relationships with environmental variables (see Morales-Castilla et al., 2015, and references therein). To determine whether co-occurrence patterns between a pair of species are a result of relationships with mediator species, we must instead examine *conditional dependence* relationships between all the species by examining residual *precision* matrices (the inverse of the correlation matrix). Conditional dependence describes how pairs of species are related, after controlling for all the other species in the dataset.

Markov models for investigating conditional dependence in ecology for presence/absence data (Harris, 2016; Clark et al., 2018) have recently been developed. These add a crucial dimension to investigating interspecies associations, as they are able to tease apart drivers (2) and (3) of co-occurrence for the first time (see Figure 1 for a demonstration). Comparisons between Markov models and null models convincingly demonstrate that null models are indeed unable to account for indirect effect of mediator species, while Markov models can (Harris, 2016).

Conditional dependence relationships in count abundance data have been investigated in Morueta-Holme et al. (2016), by applying Gaussian graphical models to log transformed counts. Similarly to the Markov models above, this method was able to tease apart the different drivers of co-occurence, including indirect associations. Unfortunately, transforming count data can have undesirable effects on analyses (OHara & Kotze, 2010; Warton et al., 2016). This is especially the case for multi-species data (Warton & Hui, 2017) because counts are often small or zero, and most species are found at only a small number of sampling occasions. Aside from transforming the responses, there are currently no methods for investigating conditional dependence in ecology that are capable of modelling ordinal or biomass data, both of which are commonly collected in ecology. What is needed now are tools for modelling conditional dependence capable of handling a broader range of distribution types, including the previously studied presence/absence and counts, but also ordinal and biomass data, such that they can be directly applied to multi-species data without the need for data transformation.

In this paper we will demonstrate the use of Gaussian copula graphical models (GCGMs; Popovic et al., 2018) to uncover conditional relationships among species from abundance data, and hence untangle the impact of mediator species on the co-occurrence patterns between pairs of species. Gaussian copulas have exciting potential in ecology for the analysis of multivariate non-normal data (Anderson et al., 2019). Copula graphical models can naturally accommodate a wide variety of data types in a unified framework, including: binary variables (presence/absence); overdispersed counts; ordinal data; and biomass data. They do this by extending GLMs to accommodate multivariate data, essentially, by mapping GLM residuals onto the multivariate normal distribution. Using simulations, we show that GCGMs can be applied to a much broader range of data types than the methods currently used in ecology, and perform well in most settings. We then demonstrate this method on a count dataset of spider abundances, and a dataset of ordinal cover categories of plants in New Zealand native forests. The latter dataset is particularly challenging due to its size (1311 taxa recorded at 964 sites) as well as the ordinal nature of the response; neither feature can be accommodated in any graphical modelling method previously used in ecology.

MATERIALS AND METHODS

We start with a note on terminology. Throughout this manuscript we use the term *association* to refer to correlations between species occurrence or abundance. We further distinguish between *indirect associations*, which are observed correlations between two species without reference to other species (marginal), from *direct associations*, which are correlations that are still present after controlling for effects of other species (conditional). In the ecology literature, species associations as defined here are often referred to as species interactions (Harris, 2016; Clark et al., 2018; Ulrich, 2008). We distinguish between these concepts, as associations do not necessarily imply interactions (Dormann et al., 2018, see Discussion).

Graphical models and co-occurrence

The concept of conditional dependence will be familiar to most applied researchers who have fitted linear models, where the term is typically used to refer to *controlling for* variables using a covariate. To measure if there is, for example, an effect of temperature on leaf width, after controlling for altitude, we can simply put both temperature and altitude in a model predicting leaf width, and test for an effect of temperature. This is the same to asking whether leaf width is conditionally dependent on temperature given altitude.

A similar strategy can be employed to uncover conditional dependence relationships between species, by using the species of interest as the response, and all the other species as predictors. This node-wise method is in fact how Gaussian graphical models were first conceived and fitted (see: Banerjee et al., 2006; Meinshausen & Bühlmann, 2006). Modern methods for Gaussian graphical modelling use a LASSO-penalised likelihood approach (Friedman et al., 2008; Hastie et al., 2015) to estimate these conditional dependence relationships in Gaussian data simultaneously for all species, as they have better power and interpretability, but the intuition remains the same. Both estimation methods have been used in ecology, with Clark et al. (2018) implementing a node wise method which more easily scales to a large number of species, while Harris (2016) and Morueta-Holme et al. (2016) estimate the graph jointly for all species, thus allowing for greater accuracy.

Figure 1 demonstrates how graphical models can be used to disentangle whether correlations between species are due to indirect relationships via mediator species or direct relationships between the species of interest. Here we are interested in the relationship between two species A and B, which we observe to be negatively correlated (red, i and iv) in both scenarios given by the top and bottom rows. After controlling for a third species C (ii and v), we either find A and B are still related (top, ii), or that their relationship was completely explained by both species responding to the mediator species C (bottom, v). It is important to emphasise that the only way to distinguish between these scenarios is to look at conditional relationships as given by the middle and right columns of Figure 1. The "graph" in figures iii and vi, after which graphical models are named, is nothing but a convenient and visually appealing way to display these conditional dependence relationships. Conditional relationships are inferred from the matrix of partial correlations, which for Gaussian data is given by the inverse of the correlation matrix, known as the *precision* matrix.

The main distinction between graphical modelling methods and other methods commonly implemented in ecology to investigate species associations, like null models and latent variable models, is in the kind of relationship they model, calculate or estimate from the data. Previously, dependence relationships (or correlations) have been modelled, which confound direct and indirect associations. By contrast, modelling conditional dependence relationships can untangle direct and indirect associations. For presence-absence data this is illustrated in Harris (2016), who showed that null models can have false positive rates as high as 100% for associations between species which are not directly related, but are correlated due to associations with mediator species. While graphical models are very rarely used in ecology, they are increasingly used in many other disciplines, from modelling gene networks (Krämer et al., 2009) to brain connectivity (Huang et al., 2010) and traffic flows (Sun et al., 2012).

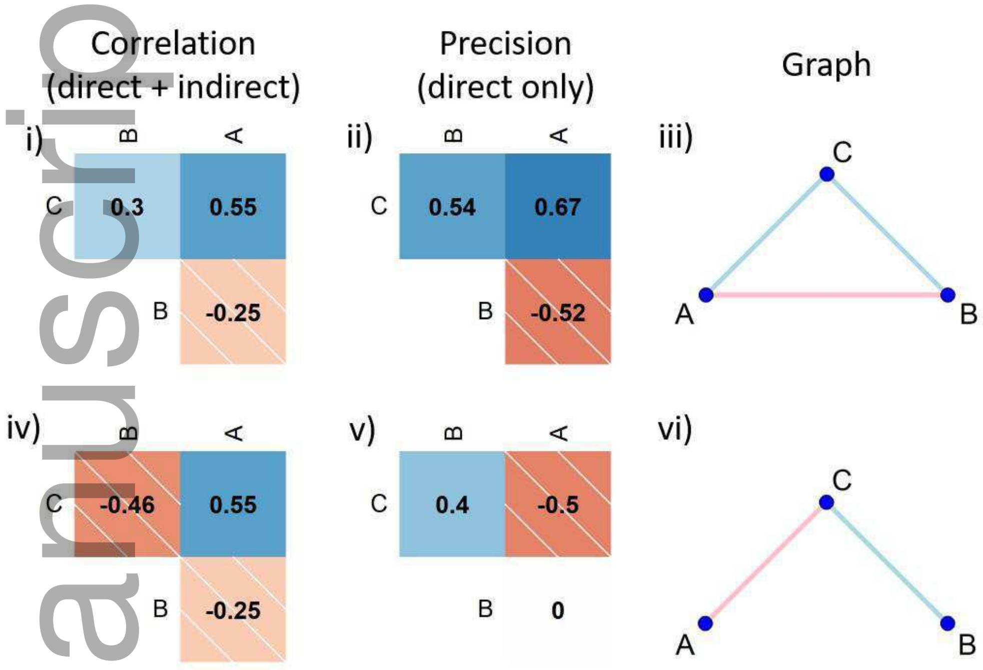


Figure 1: Left (i and iv): correlations between species A, B, and C (red dashed for negative and blue for positive), which include both direct associations and indirect mediator species effects. Middle and right: partial correlations between each pair of species, after controlling for the other species, which show direct associations only. The middle column (ii and v) plots partial correlations in a matrix, while the right column (iii and vi) plots this same information in a "graph". Correlations (left) between species A and B are the same ($\rho_{AB} = -0.25$) in both scenarios (top and bottom). To see if this relationship is explained by both A and B responding to C, a mediator species, we can examine the conditional relationships (ii and v). In the bottom row (iv – vi), the relationship between A and B is fully explained by them both responding to C, while in the top row (i – iii), A and B are related, even after controlling for C. We cannot untangle these without looking at conditional relationships in either the precision matrix (ii and v) or graph (iii and vi).

Gaussian copula graphical models

Multi-species data (also known as multivariate abundance data) consist of measurements of abundance of plants or animals (most commonly presence/absence, counts, cover or biomass) simultaneously collected for a large number of taxa, classified to species or another taxonomic level. Gaussian graphical models cannot be directly used to model multi-species data, as the responses are most often discrete or, in the case of biomass, semi-continuous. Gaussian copula graphical models provide a way to apply graphical models to any random variable, given a model for the statistical distribution of each species (its *marginal distribution*).

Copula models are so named because they couple a multivariate distribution (e.g. multivariate Gaussian) with any set of marginal distributions. For abundance data in ecology, which are generally discrete, the most useful marginal distributions are the binomial for presence/absence, the Poisson, negative binomial or binomial for counts, a multinomial (often using a cumulative link model) for ordinal data and the Tweedie distribution for biomass data. Copulas allow us to combine these into one model with the correlation structure of a multivariate Gaussian, which in turn allows us to indirectly apply Gaussian graphical models with all these data types. Figure (2) demonstrates visually how copula models work. The term "Gaussian" in Gaussian copula graphical models refers only to the copula part of the model, the responses are modelled by the appropriate discrete distribution (*e.g.* negative binomial for counts). The Gaussian copula will be used here in preference to other copulas (*e.g.* vine copulas or archimedean copulas) as it allows the application of Gaussian graphical modelling algorithms in the estimation of the GCGMs (Popovic et al., 2018). Details of the model-fitting algorithm can be found in the Supplementary material, but briefly, it uses a Monte Carlo EM algorithm (Wei & Tanner, 1990), where repeated sets of Dunn-Smyth residuals (Dunn & Smyth, 1996) from the marginal model are used to estimate the copula likelihood, via importance sampling, with importance weights updated iteratively following each graphical lasso fit to (re)weighted data, until convergence (when the estimate of the precision matrix stabilises). Graphical lasso places a LASSO penalty on the elements of the precision matrix, which describes the conditional dependences. This penalty plays two roles. Firstly, it stabilises precision matrix estimates when the number of sites is not large compared to the number of species. Secondly, when this penalty is large, it shrinks some elements of the precision matrix to zero, thereby estimating which pairs of species are conditionally independent, *i.e.* have no direct association. Gaussian copulas and Gaussian graphical models are used together for ease and speed of estimation. An R package ecoCopula implementing GCGMs is available on github.

The GCGM model can easily incorporate any set of marginal distributions, including any GLMs, with and without covariates, and more complex models like generalised additive models, and cumulative link models for ordinal data. When using these models, we can apply standard tools for residual analysis and model selection to choose the appropriate marginal model for each species. Flexibility is the main advantage of GCGM over graphical modelling techniques previously used for multi-species data. The method implemented in Harris (2015) for example can only be used on presence/absence data, while the method of Morueta-Holme et al. (2016) requires multivariate normal data, or data that can be transformed to approximately satisfy multivariate normality.

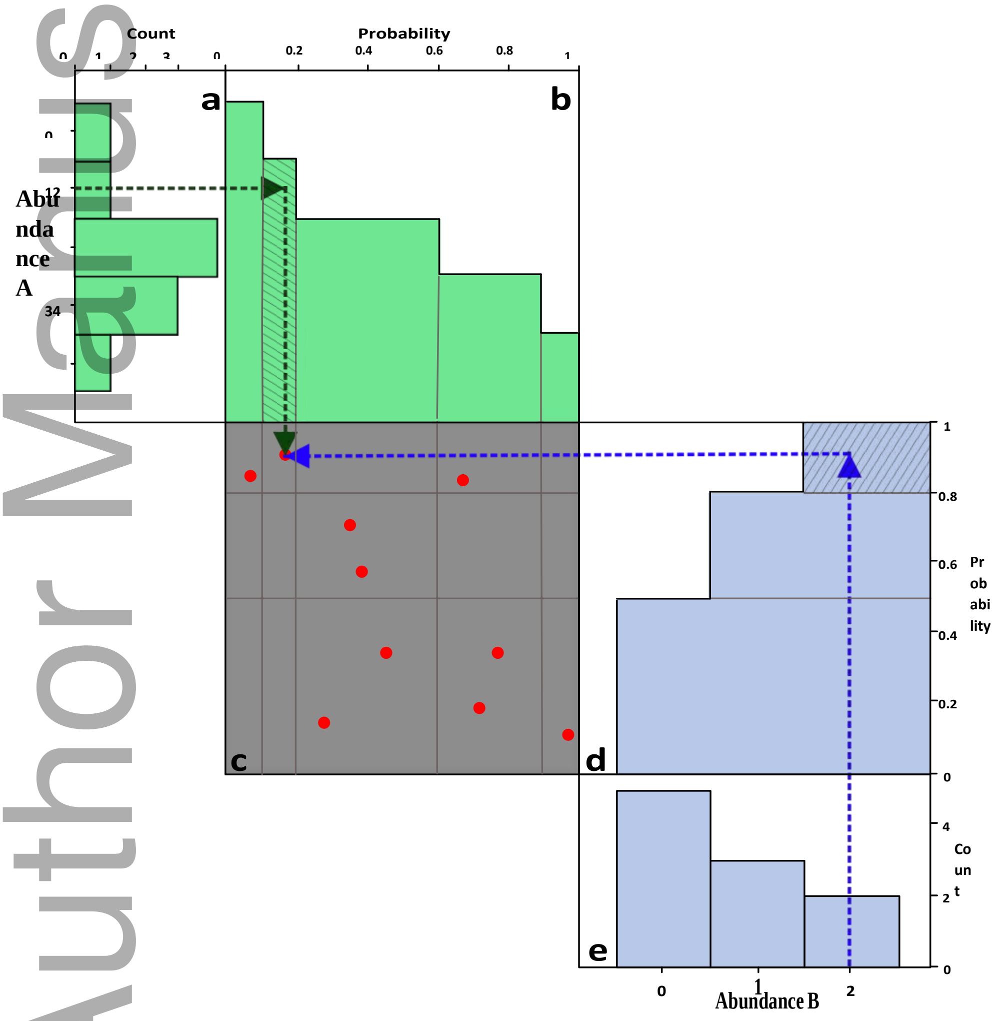


Figure 2: A visual illustration of copulas for discrete data, and their estimation. We have ten observations from two species, A and B, with abundances given by the histograms (panel a and e). For each species we calculate the cumulative distribution function (CDF; panel b and d). This can be accomplished based on the empirical, 'non-parametric' CDF, although generally in GCGMs, the CDF is calculated assuming a

parametric model, *e.g.* a Poisson regression model for counts. These CDFs give the estimated (marginal) distributions of each species separately. To model and estimate associations between species, we use these CDFs to jointly project each observation (site) onto the unit square (or unit (hyper)cube when there are more species). The dashed arrows represent the process of projection for one site/observation, at which we observe 1 individual of species A and 2 individuals of species B. Looking firstly at the species A, we project the observed count onto the region between the CDF value at our observed value (0.2) and the previous CDF value (0.1). We then generate a random uniform value inside this region. By doing the same for the observed count for species B, we thus calculate a randomised "residual" (red point). If we do this for all the observations, we obtain the pattern of red points in panel c, suggesting negative association between species A and B. To estimate the joint distribution between species A and B, we then assume a form (for example Gaussian copula) for the distribution in panel c (background shading), and use the points to estimate the correlation parameter ρ of this distribution (in this case $\rho = -0.46$).

Data

We use two datasets in our analysis and simulations. The first is a dataset consisting of counts of hunting spiders caught in pitfall traps, with 12 species found at 28 sites (van der Aart & Smeenk-Enserink, 1974). The primary aim of this study was to identify the main environmental factors associated with the distribution of the species studied. The hunting spider data is a well known ecological dataset popularised in ter Braak (1986). The data contain six covariates thought to be associated with spider abundance, namely: dry soil mass; percent cover of bare sand; percent cover of fallen leaves or twigs; percent cover of moss; percent cover of herb layer and reflection of the soil surface with a cloudless sky.

Manaa

Simulations

For each data type (presence/absence, counts, ordinal and biomass), we simulated data from competing models able to assign conditional relationships (direct associations) to such data, with 9 species and 50 sites used in all simulations. We simulated a heterogeneous environment by including a binary environmental covariate in simulations, and a homogeneous environment by excluding this covariate. All species association matrices were generated randomly. We then fitted a range of methods currently available in ecology to the simulated data, and assessed how well they were able to recover these relationships. Simulation code is available in the supplementary material (S2).

In order for associations to be interpretable, their sign and magnitude should be meaningful, such that if there is a larger negative association between species A and B compared to between species D and E, then we want this ordering to be reflected in the estimates of these associations. The Spearman's correlation coefficient (correlation of the order of associations) assesses the degree of correlation between the order of the two variables, and we will use it to gauge how well the methods perform in all the simulations below. We do not assess the actual estimates, as the different models are on very different scales, and we would not be able to straightforwardly compare across competing models. A Spearman's correlation of 1 here means the order of the estimated associations is identical to the order of the true associations used to generate the data. As most of the methods we compared do not incorporate shrinkage of the (partial) correlation parameters, we implemented GCGMs without any shrinkage in all simulations, though by default the ecoCopula package does choose a shrinkage parameter with BIC.

Binary co-occurrence (presence/absence) data

For binary data, we simulated from two models for conditional independence relationships between binary variables: the Markov model fitted in Harris (2015) and Clark et al. (2018), using the simulation method found in Appendix S2 of the former; and the GCGM with binary marginal distribution. We then fitted five methods to the data to see how well they were able to recover the direct species associations: GCGMs as implemented in the ecoCopula package; the rosalia package which implements the method of Harris (2015); the MRFcov package which implements the method of Clark et al. (2018); the gllvm package which implements the latent variables models with estimation as outlined in Hui et al. (2017); and the Fortran package Pairs which implements the approach of (Ulrich, 2008). The gllvm and Pairs packages do not estimate conditional dependence, so are unable to distinguish between direct and indirect associations.

Count data

For count data, we simulated from two methods that model conditional independence relationships between count variables: the Markov model proposed in Yang et al. (2015, hereafter yang), the GCGM assuming Poisson marginal distributions. We then compared the following five methods to assess how well they were able to recover direct species associations: GCGMs are implemented in the ecoCopula package; graphical lasso on log transformed counts as demonstrated in Morueta-Holme et al. (2016, hereafter ggmlog); the MRFcov package using the Poisson family argument, which transforms counts with a root mean square transformation before applying a Gaussian graphical model; the gllvm package; and the rosalia package with counts transformed to presence/absence.

Ordinal and biomass data

For both ordinal and biomass data, we simulated from the GCGM with cumulative link marginal models for ordinal data, and Tweedie marginal models for biomass data. We then compared the following two methods to assess how well they were able to recover direct species associations: GCGMs as implemented in the ecoCopula package; and the rosalia package, with the data transformed to presence/absence.

RESULTS

Simulation results

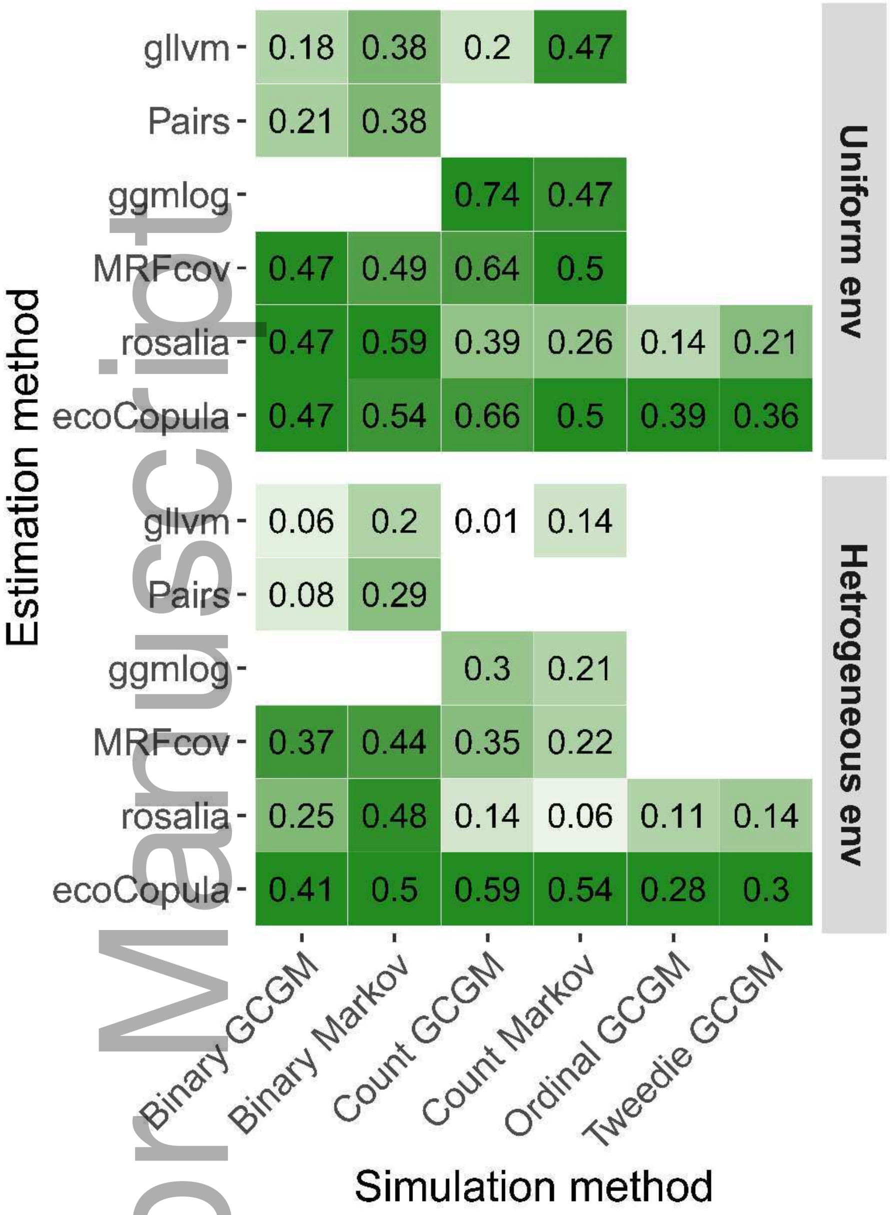


Figure 3: Average Spearman's correlation coefficient between estimated and true pairwise associations for each estimation method across 50 simulations of each combination of simulation method and environment (uniform and heterogeneous). Shading is standardised per scenario such that the highest correlation coefficient (best method) is always dark green. White cells without text indicate this combination of simulation and estimation methods was not implemented. While other estimation methods perform best in certain scenarios (e.g. rosalia for the binary Markov simulation model and ggmlog for the count GCGM simulation model, both in a uniform environment), ecoCopula consistently performs well, especially when the environment is heterogeneous, where it is the best performing estimation method in all scenarios. The methods which do not estimate direct associations (gllvm and Pairs) generally have the poorest performance.

For binary co-occurrence (presence/absence) data, the methods that do not differentiate between direct and indirect associations (gllvm and Pairs) performed poorly at recovering direct species relationships relative the other three methods, which do model indirect associations (Figure 3). This result was consistent with that of Harris (2016).

For count data, as with the binary data, gllvm, which does not differentiate between direct and indirect associations, performed worse at recovering direct species relationships relative to other methods (Figure 3). The Markov model as estimated by the roselia R package, which requires the data to be transformed to presence/absence but does model direct associations, performed poorly relative to the other methods (with the exception of gllvm). This was not surprising as truncating data to presence/absence potentially loses a lot of information, especially for abundant species. For all simulation methods in a uniform environment, the three estimation methods which modelled abundance and direct associations (ecoCopula, ggmlog, and MRFcov) produced similar results. However, in the presence of (known) environmental heterogeneity, ecoCopula notably outperformed the other methods.

For both ordinal and biomass data, the Markov model as estimated by the roselia R package, which requires the data to be transformed to presence/absence, performed poorly relative to the ecoCopula, which models the abundance data directly.

Overall the ecoCopula package performs very well in the simulations. This is consistent with the observation that methods which estimate both direct associations, and model the data directly without transformations, are best able to estimate direct species associations. In addition, the ecoCopula package can control for measured environmental gradients, which the other methods do not, and so performs particularly well relative to other methods in the presence of environmental heterogeneity.

Data analysis results

GCGM analysis of counts

The first dataset we analyse are counts of hunting spiders (described in Methods). For these data we used negative binomial marginal distributions, as these can account for the overdispersion often observed in abundance count data. We can include predictors in these marginal distributions, and so will consider partial correlation between species both before and after accounting for observed environmental variables.

After controlling for the effect of other species (Figure 4, b and e), partial correlations between species tend to be smaller than in the correlation matrices (a and d), with many pairs of species being estimated to be conditionally independent. For these species, their correlation is explained by relationships with other species in the data. In addition, environment can explain some of the (partial) correlation between species, with 60% (22/36) less non-zero partial correlations between species after controlling for environment.

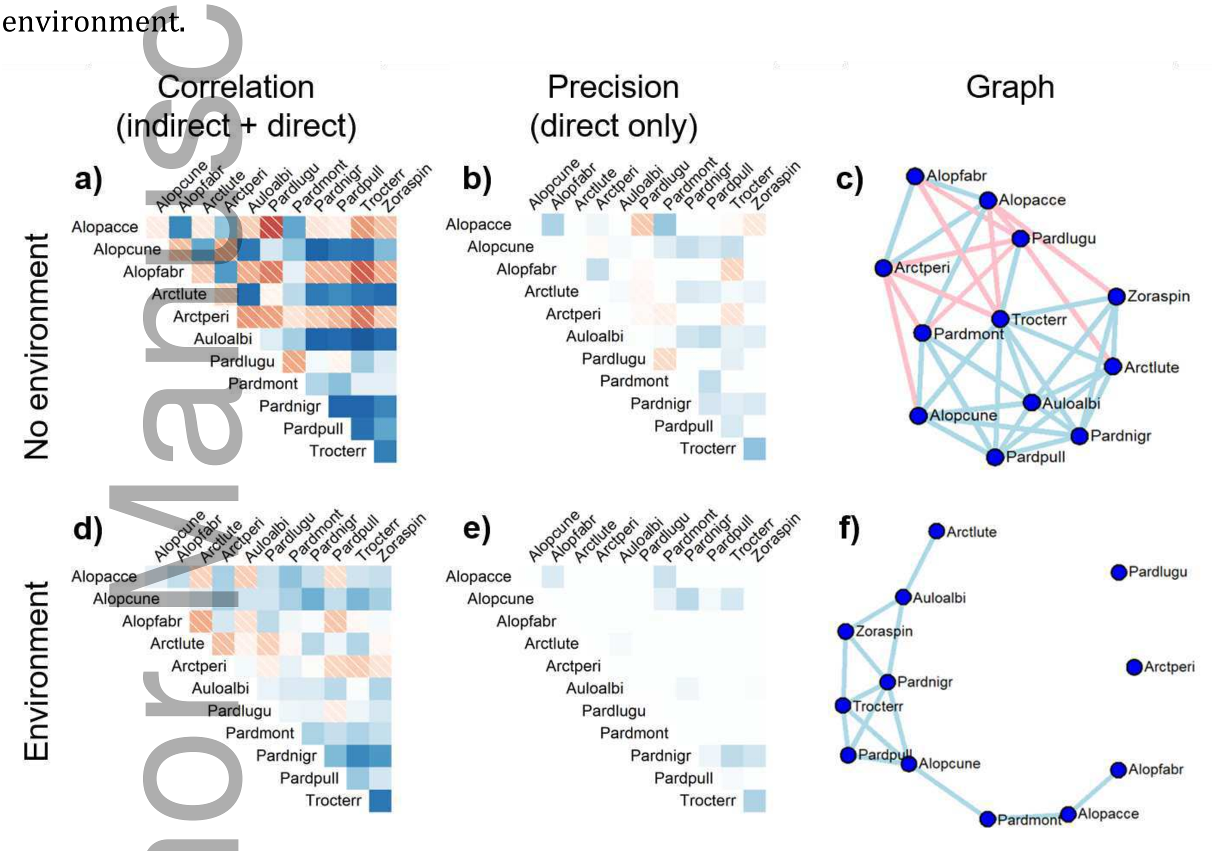


Figure 4: Raw and partial correlation between spider species. Correlations are positive (blue) if species co-occur more often than expected by their prevalence, negative (red, dashed) if they co-occur less often, and zero (white) if they co-occur at the rate expected by their prevalence. Correlations are plotted for the raw abundances (a – c) and after controlling for measured environmental variables (d – f). The left column (a and d) shows raw correlations. Middle and right columns both display partial correlations, after additionally controlling for mediator species, with "graphs" on the right (c and f), which are plots of the network of conditional dependence. Names are abbreviations based on the first four letters of the genus, then the first for letters of the species names (see Supplementary Material 1 Appendix B for details).

Interpreting the graph of spiders, we observe that *Pardosa lugubris* (Pardlugu) has negative associations with both *Alopecosa accentuata* (Alopacce) and *Pardosa monticola* (Pardmont), who are positively associated, before controlling for covariates (Figure 4, a – c). However, these negative associations are absent after controlling for covariates (Figure 4, d – f). Looking at Figure 5, we can see that Pardlugu has increased abundance in the presence of bare soil, while both Alopacce and Pardmont have the opposite response. These differences in environmental response seem to be the main reason for the perceived negative associations seen in Figure 4 (a – e), rather than direct negative associations between the species.

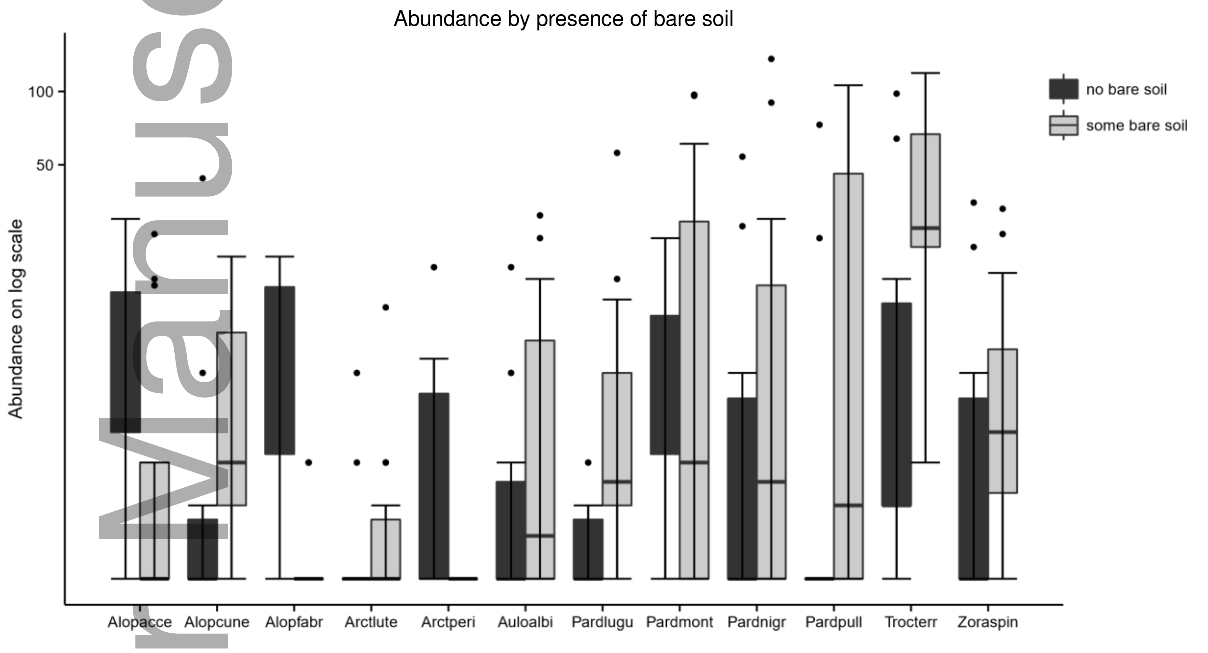


Figure 5: Spider abundance plotted against the presence of any bare soil. Notice that many of the species abundances differ according to the presence of any bare soil. For example, *Pardosa lugubris* (Pardlugu) has increased abundance in the presence of bare soil, while both *Alopecosa accentuata* (Alopacce) *and Pardosa monticola* (Pardmont) have decreased abundance.

GCGM analysis of large ordinal dataset

We applied GCGMs to the New Zealand forest cover data (see Methods for description), a complex dataset that could not have been analysed using pre-existing methods. In New Zealand, podocarp-broadleaf forest and beech forests tend not to co-occur, but are not obviously separated by geography or climate (Wardle, 2002, p.672). We therefore predicted that these two main forest types would form separate groupings in our analysis.

In order to investigate associations between species in New Zealand native forests, we fitted a GCGM with cumulative link (Agresti, 2010) marginal distributions to these data. We fit models with different values of the shrinkage parameter, with large values giving very sparse graphs with few direct associations, and small values giving dense graphs, but all including slope and altitude as covariates, we then chose the value of the shrinkage parameter with BIC.

Due to the large number of species, the graph of all correlated species (Figure 6) is fairly complex, nevertheless we can see some important patterns. For the 1311 New Zealand forest species in the data, there are 858,705 ($1311 \times (1311 - 1)/2$) possible pairwise direct associations between them. We estimated the vast majority (857,968, $> 99\%$) of species pairs to have no direct association. There are an estimated 656 ($< 1\%$) positive direct associations (blue lines), with the two ends of this gradient of positive association linked by a small number (22, $< 1\%$) of direct negative associations. This gradient is seen as a *U*-shaped group of blue positive associations on Figure 6. The dominant species at one end of the gradient tend to be associated with silver beech forest (*Lophozonia menziesii* (NOTMEN), *Raukaua simplex* (RAUSIM), *Myrsine divaricata* (MYRDIV), *Blechnum procerum* (BLEPRO), *Coprosma foetidissima* (COPFOE), and *Notogrammitis billardierei* (GRABIL)), while the other end of the gradient is dominated by species associated with the early-mid stages of regeneration of disturbed lowland forest (*Cyathea dealbata* (CYADEA), *Melicytus ramiflorus* (MELRAM), *Knightia excelsa* (KNIEXC) and *Uncinia uncinata* (UNCUNC)). This is partially consistent with our initial expectation that there would be a separation between the two main forest types of New Zealand.

Some of the relationships we found were surprising, and can be used to generate hypotheses for further investigation. In Figure 6, the separation between the two main forest types of New Zealand was not supported for the three species in *Fuscospora*, the other genus of southern beech present in New Zealand (*Fuscospora solandri* (NOTSOL), *F. cliffortioides* (NOTCLI), *F. fusca* (NOTFUS)). The three *Fuscospora* species fell in different parts of the graph rather than clustering together or with *Lophozonia menziesii*, and were not associated with many other species (either negatively or positively).

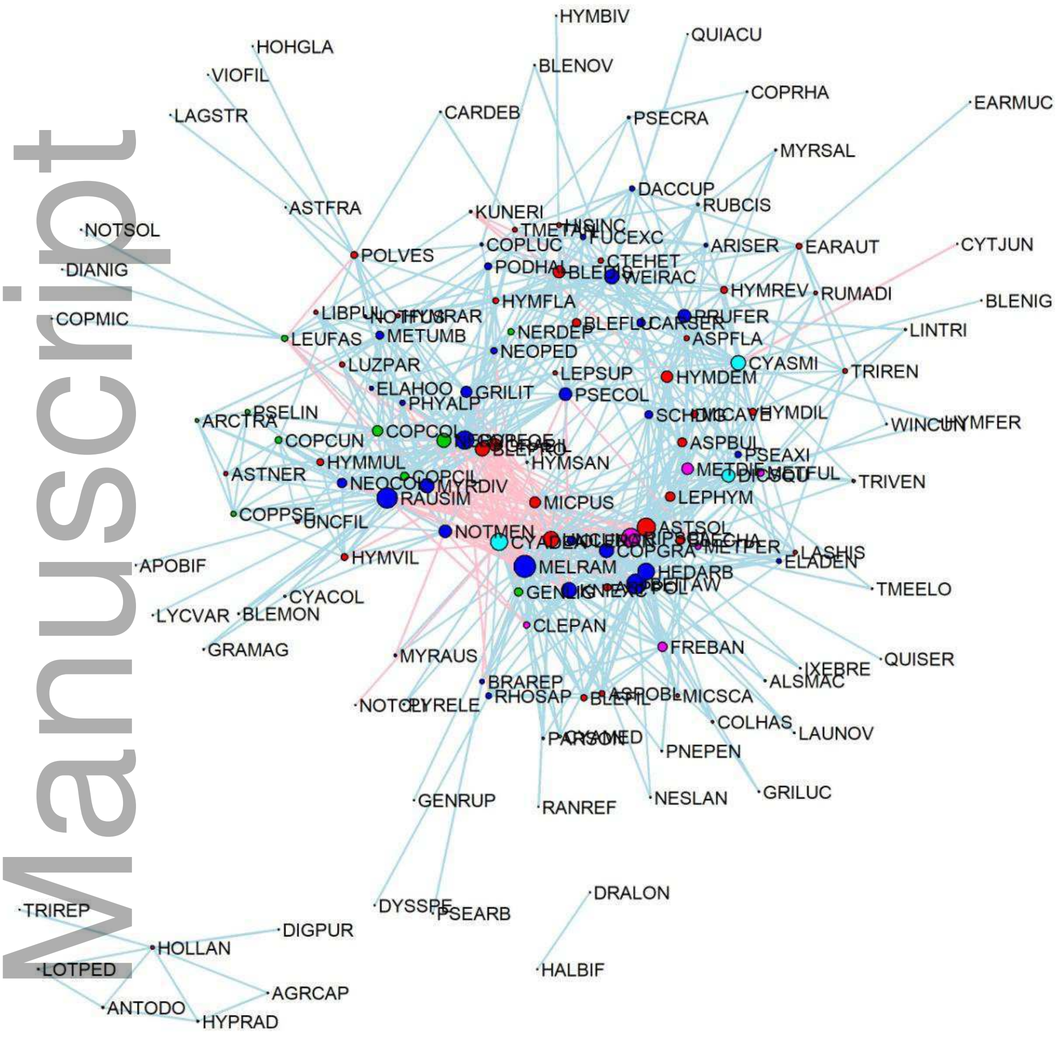


Figure 6: Graph for all species after controlling for covariates (slope and altitude) and mediating species. There are a range of positive (blue) and negative (red) partial correlations between species. The colour of the dot corresponds to growth form; herbs (red), shrubs (green), trees (blue), vines (magenta) and tree ferns (cyan). To simplify the display for all plots, we excluded species that were conditionally independent of all other species. We employ the Fruchterman Reingold algorithm (Fruchterman & Reingold, 1991) to position nodes on the graphs in two dimensions. Some species codes reflect outdated taxonomy (e.g. *Nothofagus menziesii* has been renamed *Lophozonia menziesii*, but retains the NOTMEN label in the dataset). Species at one end of the spectrum (NOTMEN, RAUSIM, MYRDIV, BLEPRO, COPFOE, and GRABIL) are associated with silver beech forest, while the other end of the gradient is dominated by species associated with the early-mid stages of regeneration of disturbed lowland forest (CYADEA, MELRAM, KNIEXC, and UNCUNC).

For more informative plots, we can “zoom in” on certain subsets of species, to better examine relationships between individual species. These subset graphs can display partial correlation between the species after accounting for covariates and all mediator species (including those not in the subset displayed).

We are able to distinguish between exotic herb species (*Trifolium repens* (TRIREP), *Digitalis purpurea* (DIGPUR), *Holcus lanatus* (HOLLAN), *Agrostis capillaris* (AGRCAP), *Lotus pedunculatus* (LOTPED), *Anthoxanthum odoratum* (ANTODO), *Hypochaeris radicata* (HYPRAD) and *Libertia pulchella*(LIBPUL)), which are mostly positively correlated with one another, and negatively correlated with the native species (green text) by simply looking at the plot of raw correlations (Figure 7a). This pattern is consistent with the fact that these exotic species tend to have poor performance in low light conditions (e.g. Devkota et al., 1997), and thus differ from many of the native herbaceous species (a group which includes a range of understorey ferns) in being uncommon in forest understoreys (pers. obs. A. Moles).

In addition to discerning the same group of exotic herbaceous species (purple) on the graph (Figure 7c) we can additionally see that there is no evidence that these exotic species are associated or interacting with natives, as they are estimated to be conditionally independent of them (Figure 7b), a point which was not clear from the negative correlations of Figure 7a. We can also see that one of the exotic species, *Libertia pulchella*(LIBPUL) is more closely associated with the native species. This may indicate that while the other exotic species are growing separate to native species, possibly in disturbed habitats, the *Libertia pulchella* is co-occurring with native species.

Curiously, we found that *Hymenophyllum sanguinolentum* (HYMSAN) and *Hymenophyllum villosum* (HYMVIL) have a negative partial correlation, even though they can cooccur (Brownsey & Perrie, 2014). These two species are often confused with each other. This negative relationship may therefore be an artefact of misclassification rather than a true negative relationship. Field ecologists may identify one or the other of these species, and assume everything similar in a plot is the same fern. The result would be that these species be recorded to co-occur less often than expected by chance.

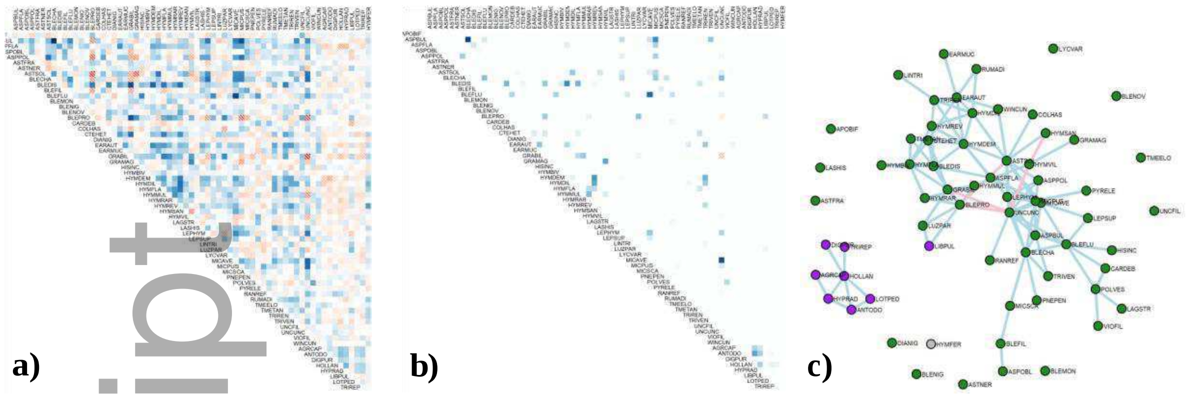


Figure 7: Herbs: partial correlations (b and c) and correlations (a). The graph clearly shows a group of exotic (purple) herbaceous (*Trillium repens* (TRIREP), *Digitalis purpurea* (DIGPUR), *Holcus lanatus* (HOLLAN), *Agrostis capillaris* (AGRCAP), *Lotus pedunculatus* (LOTPED), *Anthoxanthum odoratum* (ANTODO) and *Hypochaeris radicata* (HYPRAD)) that are positively associated with one another, and not associated with any native (green) species. While all the plots reveal that the exotic species are distinct from the native species (rightmost eight species on plots a and b), partial correlations (b and c) additionally show that one exotic species *Libertia pulchella* (LIBPUL) is more closely related to the native species, than the other exotic species.

We observe negative partial correlations between *Fuscospora cliffortioides* (NOTCLI) and other species when not controlling for any covariates (Figure 8a), which are lost after controlling for altitude and slope (Figure 8b). This is consistent with out understanding that *Fuscospora cliffortioides* occurs in montane and subalpine forest that tend to occur at altitudes between 400 m - 1380 m above sea level, whereas *Prumnopitys ferruginea* (PRUFER), *Weinmannia racemosa* (WEIRAC) and *Raukaua simplex* (RAUSIM) occur in lower altitude forest types, that range from sea-level up to 700 m in the south island and up to 1100 m in the north island (Wiser et al., 2011).

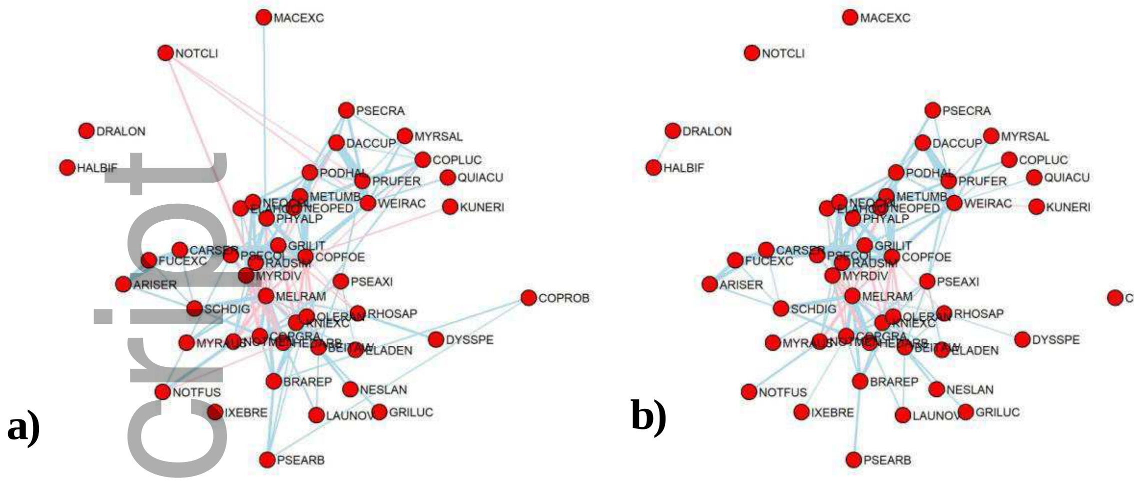


Figure 8: Partial correlation graph of trees before (a) and after (b) controlling for covariates. Looking at these graphs can suggest which correlations are explained by the covariates modelled. Negative partial correlations between *Fuscospora cliffortioides* (NOTCLI) and other species are not present after controlling for covariates (slope and altitude), presumably because *Fuscospora cliffortioides* tends to occur at higher altitudes.

DISCUSSION

We have demonstrated a new method for exploring whether co-occurrence patterns between species may be explained by indirect mediator species relationships, by response to environmental variables, or by neither. GCGMs can be used to analyse a wide variety of data types commonly observed in multi-species data in ecology, and can accommodate datasets with large numbers of sampling units and/or species in a computationally efficient manner. Simulations showed that GCGMs perform as well or better than competing methods in most scenarios, especially when the data are ordinal or biomass, or when there is known environmental heterogeneity. They are able to reproduce known relationships between species, and can generate hypotheses based on surprising partial correlation patterns, as demonstrated in our analysis of the New Zealand native forest cover data.

Ecologists have long been studying direct and indirect species interactions using manipulative experiments (e.g. Strauss, 1991; Dill et al., 2003) where for a small number of species, presences and abundances are manipulated to measure the effect on other species. Such manipulative experiments can tease apart direct and indirect species interactions. Studies using observational co-occurrence or co-abundance data have sometimes also been interpreted as species interactions, including in Harris (2015) and Morueta-Holme et al. (2016). However, some care has to be taken with this interpretation (Dormann et al., 2018). Species which appear to interact may both be responding to *unmeasured* environmental variables. In addition, when species appear not to be interacting (partial correlation is zero), this should not be interpreted as evidence of no interaction without some consideration of power and effect size, because perhaps we do not have sufficient data to estimate these interactions (*e.g.* rare species had fewer connections in Figure 6, probably for this reason).

While being generally more flexible than other methods, GCGMs do not have all the functionality of existing methods. In particular, as noted in Clark et al. (2018), the sign and strength of direct associations can change with environmental gradients (*e.g.* He et al., 2013). Some methods, including the Mrfcov package we evaluate in this paper, are able to model such changing associations. GCGMs could be extended to model these by using a similar node-wise algorithm applied to the Dunn-Smyth residuals with an iterative weighting scheme, and this is an avenue of future research.

With these caveats in mind, we recommend GCGMs as a method for visualising multivariate data, to be used in combination with other visualisation methods like ordination. While species interactions cannot be directly inferred from observational data, GCGMs can be used as an exploratory tool to generate hypotheses about relationships between species, to inform further research.

Gaussian copulas, as used here, have exciting potential in ecology for the analysis of multivariate non-normal data (Anderson et al., 2019). For example, the algorithm used to construct GCGMs here is quite general, and can be used to fit any desired covariance model on iteratively reweighted data (Popovic et al., 2018). Thus copulas can be readily used for ordination, using an iteratively reweighted factor analysis, as a fast, large-sample alternative to the hierarchical methods currently used for model-based ordination (Warton et al., 2015). Copulas can also be used as a simulation model (as in Warton et al., 2017), and have potential as a tool for likelihood-based inference about multivariate data.

Undirected networks, such as those simulated and estimated by GCGMs and Markov models (e.g. Figure 8) also hold potential for further research in ecology. They can, for example, be analysed with graph centrality measures like degree (the number of edges connecting the node/species) and betweenness (the degree to which the path between other nodes/species must pass though the node in question) among many others (Freeman, 1978). It may be possible to use these to determine the importance of species in the ecosystems modelled We look forward to seeing further progress using a copula approach and graph theory for multivariate analysis in ecology.

ACKNOWLEDGEMENTS

FKCH is supported by an ANU cross-disciplinary research grant. DIW was supported by an Australian Research Council Future Fellowship (FT120100501). GCP was supported by the Australia Postgraduate Award and ARC Discovery Project scheme (DP180103543). ATM is supported by an Australia Research Council Discovery Grant (DP180100836). FJT is supported from the Marsden Fast-Start Fund and the Royal Society of New Zealand.

We acknowledge the use of data drawn from the Natural Forest plot data collected between January 2002 and March 2007 by the LUCAS programme for the Ministry for the Environment, New Zealand.

In addition, we thank two anonymous reviewers for critically reading the manuscript and suggesting substantial improvements.

AUTHOR CONTRIBUTIONS

GCP carried out the data analysis and lead writing. DIW and FKCH provided ideas and input, and contributed critically to drafts. ATM and FJT assisted with interpretation of results from and ecological perspective and contributed to the writing for ecological aspects.

DATA ACCESSIBILITY

Hunting spider data are available in the mvabund package in R, which can be downloaded from the Comprehensive R Archive Network at https://CRAN.R-project.org/package=mvabund.. New Zealand native forest data used for these analyses are available at Zenodo: https://doi.org/10.5281/zenodo.3256709. Analysis code is provided in supplementary material, and in the ecoCopulaR package on github (gordy2x/ecoCopula, DOI: 10.5281/zenodo.3257118)).

REFERENCES

Agresti, A. (2010). *Analysis of ordinal categorical data*. New York: John Wiley & Sons, New York.

Anderson, M. J., de Valpine, P., Punnett, A., & Miller, A. E. (2019). A pathway for multivariate analysis of ecological communities using copulas. *Ecology and Evolution*.

Banerjee, O., Ghaoui, L. E., d'Aspremont, A., & Natsoulis, G. (2006). Convex optimization techniques for fitting sparse Gaussian graphical models. In *Proceedings of the 23rd international conference on Machine learning*, (pp. 89–96).

Brownsey, P. J., & Perrie, L. R. (2014). Flora of New Zealand - Ferns and Lycophytes. In *Breitwieser, I. Heenan, Wilton, A.D. Flora of New Zealand*. Manaaki Whenua Press, Lincoln.

Clark, N. J., Wells, K., & Lindberg, O. (2018). Unravelling changing interspecific interactions across environmental gradients using markov random fields. *Ecology*, *99*(6), 1277–1283.

D'Amen, M., Mod, H. K., Gotelli, N. J., & Guisan, A. (2017). Disentangling biotic interactions, environmental filters, and dispersal limitation as drivers of species cooccurrence. *Ecography*, *41*(8), 1233–1244.

Devkota, N. R., Kemp, P. D., & Hodgson, J. (1997). Screening pasture species for shade tolerance. *27*, 119–128.

Dill, L. M., Heithaus, M. R., & Walters, C. J. (2003). Behaviorally mediated indirect interactions in marine communities and their conservation implications. *Ecology*, *84*(5), 1151–1157.

Dormann, C. F., Bobrowski, M., Dehling, D. M., Harris, D. J., Hartig, F., Lischke, H., Moretti, M. D., Pagel, J., Pinkert, S., Schleuning, M., Schmidt, S. I., Sheppard, C. S., Steinbauer, M. J., Zeuss, D., & Kraan, C. (2018). Biotic interactions in species distribution modelling: 10 questions to guide interpretation and avoid false conclusions. *Global Ecology and Biogeography*, *27*(9), 1004–1016.

Dunn, P. K., & Smyth, G. K. (1996). Randomized quantile residuals. *Journal of Computational and Graphical Statistics*, *5*(3), 236–244.

Freeman, L. C. (1978). Centrality in social networks conceptual clarification. *Social Networks*, *1*(3), 215–239.

Freestone, A. L. (2006). Facilitation drives local abundance and regional distribution of a rare plant in a harsh environment. *Ecology*, *87*(11), 2728–2735.

Friedman, J., Hastie, T., & Tibshirani, R. (2008). Sparse inverse covariance estimation with the graphical LASSO. *Biostatistics*, *9*(3), 432–441.

Fruchterman, T. M. J., & Reingold, E. M. (1991). Graph drawing by force-directed placement. *Software: Practice and Experience*, *21*(11), 1129–1164.

Gilinsky, E. (1984). The role of fish predation and spatial heterogeneity in determining benthic community structure. *Ecology*, *65*(2), 455–468.

Gotelli, N. J., & Ulrich, W. (2010). The empirical Bayes approach as a tool to identify non-random species associations. *Oecologia*, *162*(2), 463–477.

Harris, D. J. (2015). Generating realistic assemblages with a joint species distribution model. *Methods in Ecology and Evolution*, *6*(4), 465–473.

Harris, D. J. (2016). Inferring species interactions from co-occurrence data with Markov networks. *Ecology*, *97*(12), 3308–3314.

Hastie, T., Tibshirani, R., & Wainwright, M. (2015). *Statistical learning with sparsity: the Lasso and generalizations*. CRC Press, Florida.

Hawkins, B. A., Field, R., Cornell, H. V., Currie, D. J., Gugan, J.-F., Kaufman, D. M., Kerr, J. T., Mittelbach, G. G., Oberdorff, T., O'Brien, E. M., Porter, E. E., & Turner, J. R. G. (2003). Energy, water, and broad-scale geographic patterns of species richness. *Ecology*, *84*(12), 3105–3117.

He, Q., Bertness, M. D., & Altieri, A. H. (2013). Global shifts towards positive species interactions with increasing environmental stress. *Ecology Letters*, *16*(5), 695–706.

Huang, S., Li, J., Sun, L., Ye, J., Fleisher, A., Wu, T., Chen, K., & Reiman, E. (2010). Learning brain connectivity of Alzheimer's disease by sparse inverse covariance estimation. *NeuroImage*, *50*(3), 935 – 949.

Hui, F. K. C., Warton, D. I., Foster, S. D., & Dunstan, P. K. (2013). To mix or not to mix: comparing the predictive performance of mixture models vs. separate species distribution models. *Ecology*, *94*(9), 1913–1919.

Hui, F. K. C., Warton, D. I., Ormerod, J. T., Haapaniemi, V., & Taskinen, S. (2017). Variational approximations for generalized linear latent variable models. *Journal of Computational and Graphical Statistics*, *26*(1), 35–43.

Krämer, N., Schäfer, J., & Boulesteix, A.-L. (2009). Regularized estimation of largescale gene association networks using graphical Gaussian models. *BMC Bioinformatics*, *10*(1), 384.

Manaaki Whenua Landcare Research (2018). *National Vegetation Survey databank*. URL https://www.landcareresearch.co.nz

Meinshausen, N., & Bühlmann, P. (2006). High-dimensional graphs and variable selection with the Lasso. *The Annals of Statistics*, *34*(3), 1436–1462.

Morales-Castilla, I., Matias, M. G., Gravel, D., & Arajo, M. B. (2015). Inferring biotic interactions from proxies. *Trends in Ecology and Evolution*, *30*(6), 347 – 356.

Morueta-Holme, N., Blonder, B., Sandel, B., McGill, B. J., Peet, R. K., Ott, J. E., Violle, C., Enquist, B. J., Jrgensen, P. M., & Svenning, J.-C. (2016). A network approach for inferring species associations from co-occurrence data. *Ecography*, *39*(12), 1139–1150.

OHara, R. B., & Kotze, D. J. (2010). Do not log-transform count data. *Methods in Ecology and Evolution*, *1*(2), 118–122.

Ovaskainen, O., Tikhonov, G., Norberg, A., Guillaume Blanchet, F., Duan, L., Dunson, D., Roslin, T., & Abrego, N. (2017). How to make more out of community data? a conceptual framework and its implementation as models and software. *Ecology Letters*, *20*(5), 561–576.

Pollock, L. J., Tingley, R., Morris, W. K., Golding, N., O'Hara, R. B., Parris, K. M., Vesk, P. A., & McCarthy, M. A. (2014). Understanding co-occurrence by modelling species simultaneously with a joint species distribution model (JSDM). *Methods in Ecology and Evolution*, *5*(5), 397–406.

Popovic, G. C., Hui, F. K. C., & Warton, D. I. (2018). A general algorithm for covariance modeling of discrete data. *Journal of Multivariate Analysis*, *165*, 86 – 100.

Strauss, S. Y. (1991). Indirect effects in community ecology: Their definition, study and importance. *Trends in Ecology & Evolution*, *6*(7), 206 – 210.

Strong Jr, D. R., Simberloff, D., Abele, L. G., & Thistle, A. B. (2014). *Ecological communities: conceptual issues and the evidence*. Princeton University Press, Princeton.

Sun, S., Huang, R., & Gao, Y. (2012). Network-scale traffic modeling and forecasting with graphical lasso and neural networks. *Journal of Transportation Engineering*, *138*(11), 1358–1367.

ter Braak, C. J. (1986). Canonical correspondence analysis: A new eigenvector technique for multivariate direct gradient analysis. *Ecology*, *67*(5), 1167–1179.

Ulrich, W. (2008). Pairs – a FORTRAN program for studying pair-wise species associations in ecological matrices. *URL www.keib.umk.pl/pairs*.

van der Aart, P., & Smeenk-Enserink, N. (1974). Correlations between distributions of hunting spiders (Lycosidae, Ctenidae) and environmental characteristics in a dune area.
Netherlands Journal of Zoology, *25*(1), 1–45.

Wardle, P. (2002). *Vegetation of New Zealand*. The Blackburn Press, New Jersey, USA.

Warton, D. I., Blanchet, F. G., OHara, R. B., Ovaskainen, O., Taskinen, S., Walker, S. C., & Hui, F. K. C. (2015). So many variables: Joint modeling in community ecology. *Trends in Ecology and Evolution*, *30*(12), 766–779.

Warton, D. I., & Hui, F. K. C. (2017). The central role of mean-variance relationships in the analysis of multivariate abundance data: a response to Roberts (2017). *Methods in Ecology and Evolution*, *8*(11), 1408–1414.

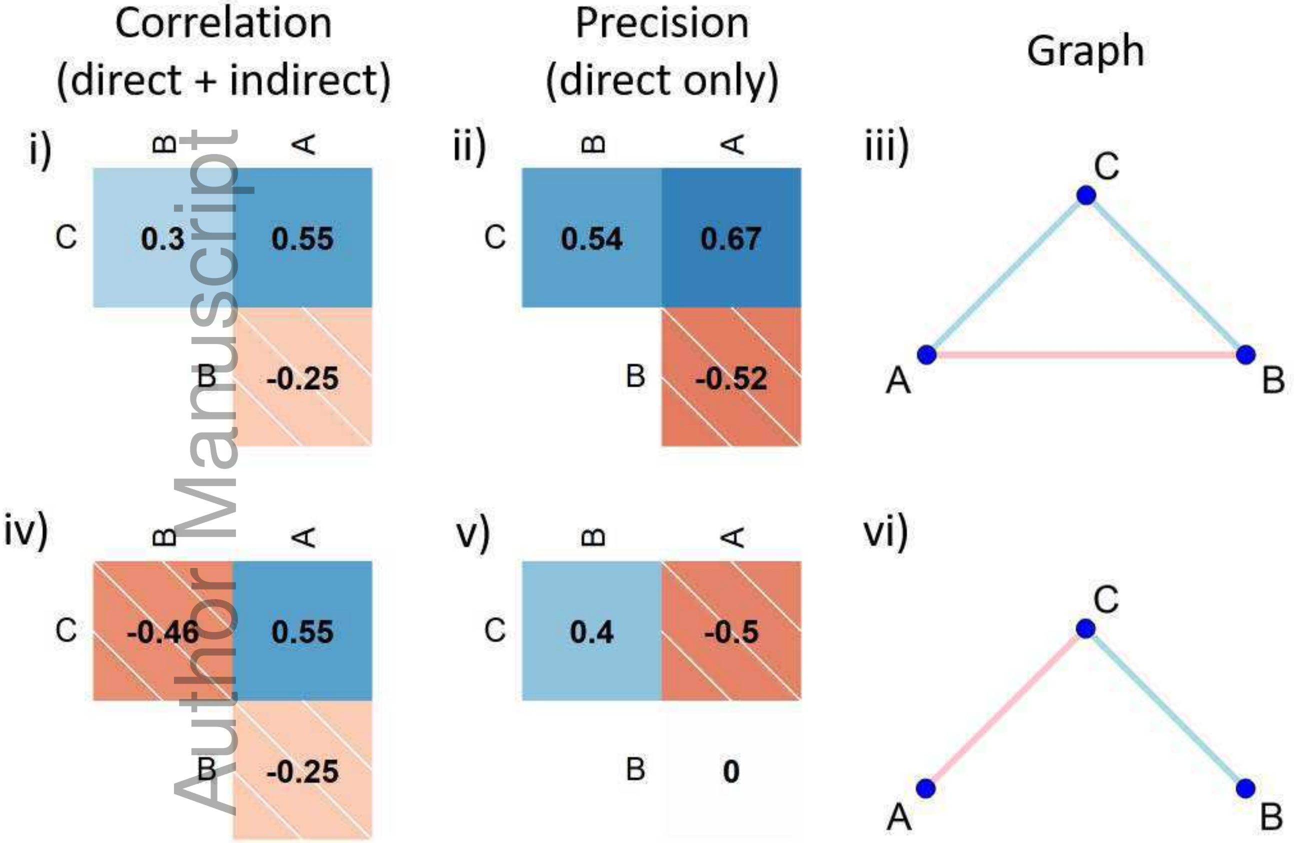
Warton, D. I., Lyons, M., Stoklosa, J., & Ives, A. R. (2016). Three points to consider when choosing a LM or GLM test for count data. *Methods in Ecology and Evolution*, *7*(8), 882–890.

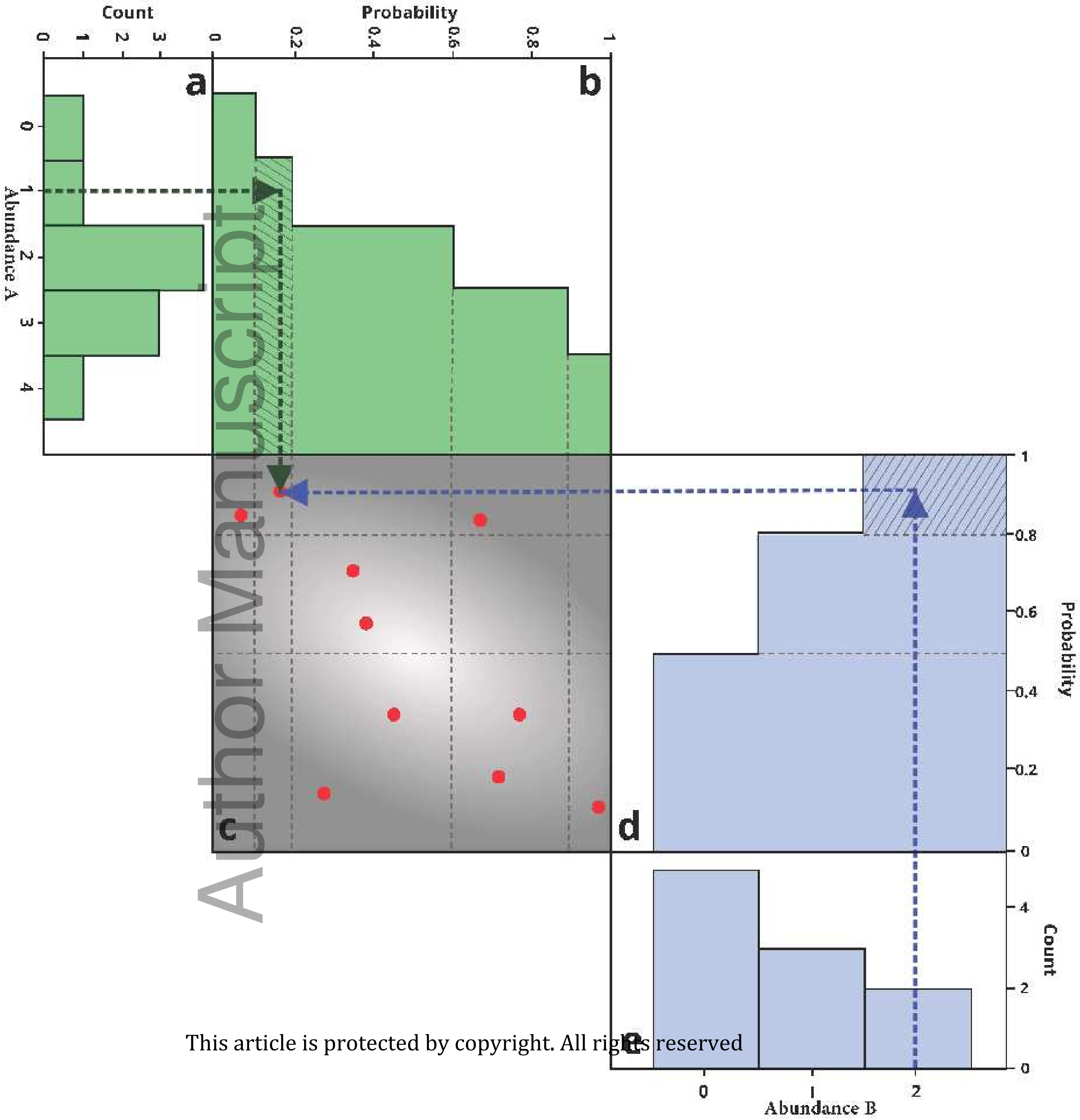
Warton, D. I., Thibaut, L., & Wang, Y. A. (2017). The PIT-trap – a "model-free" bootstrap procedure for inference about regression models with discrete, multivariate responses. *PloS one*, *12*(7), e0181790.

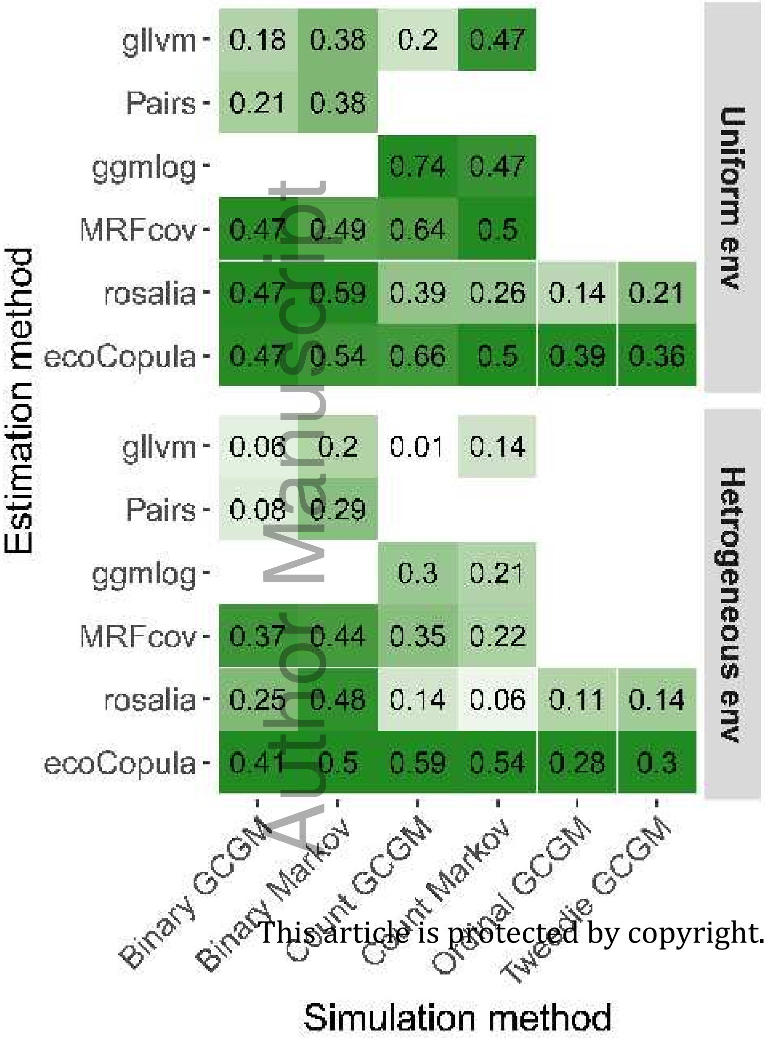
Wei, G. C., & Tanner, M. A. (1990). A Monte Carlo implementation of the EM algorithm and the poor man's data augmentation algorithms. *Journal of the American Statistical Association*, *85*(411), 699–704.

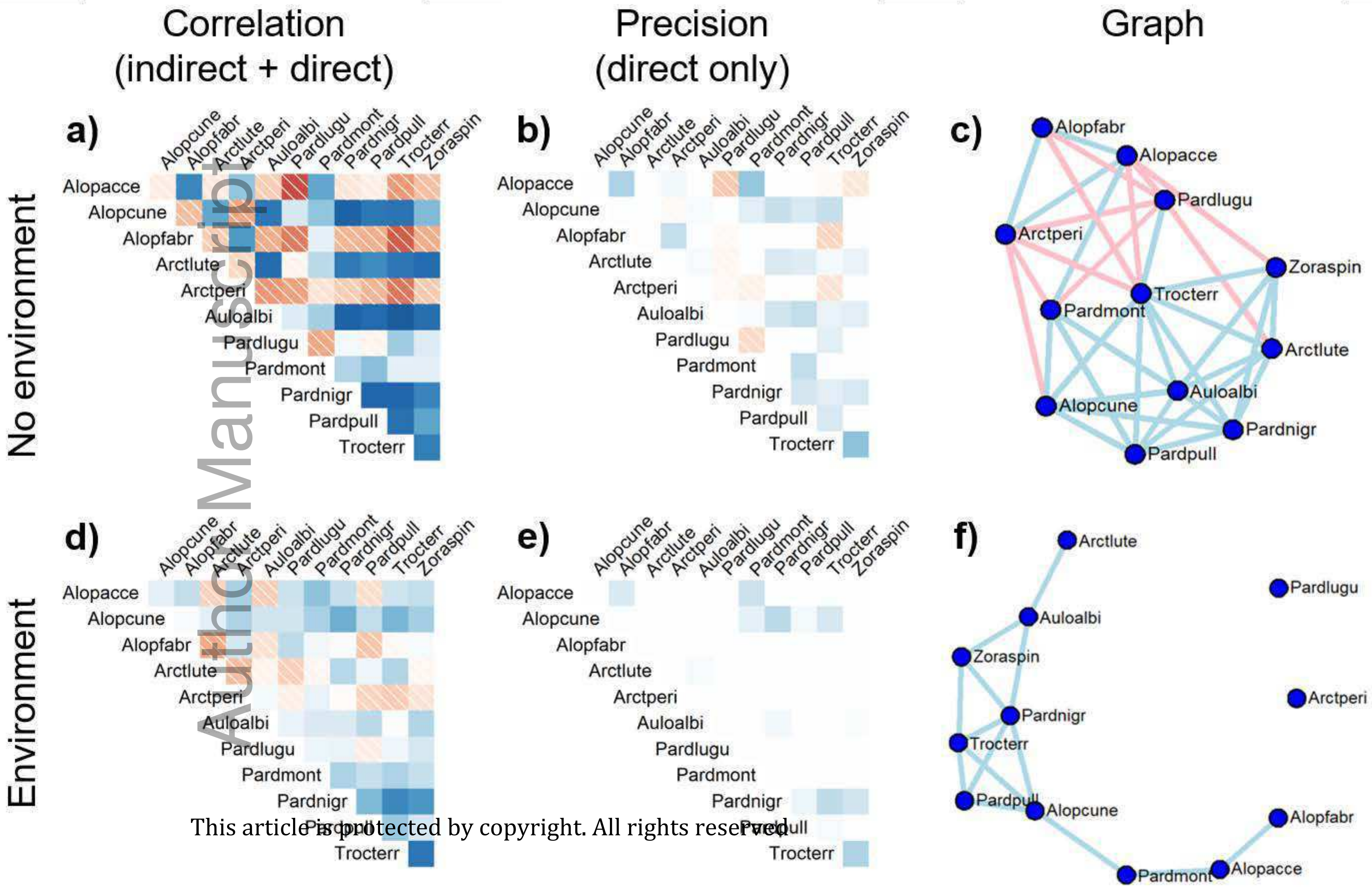
Wiser, S. K., Hurst, J. M., Wright, E. F., & Allen, R. B. (2011). New Zealand's forest and shrubland communities: a quantitative classification based on a nationally representative plot network. *Applied Vegetation Science*, *14*(4), 506–523.

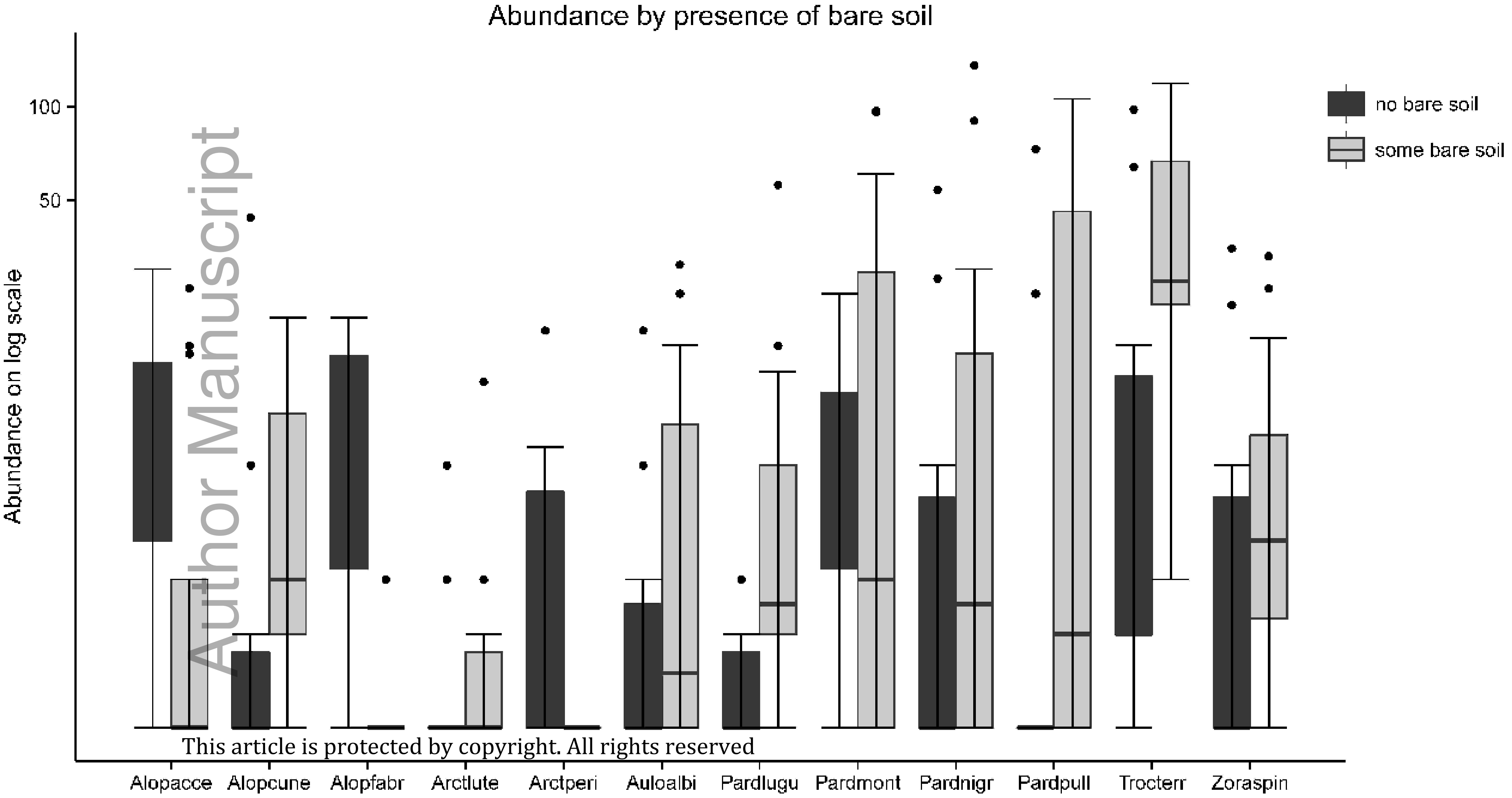
Yang, E., Ravikumar, P., Allen, G. I., & Liu, Z. (2015). Graphical models via univariate exponential family distributions. *The Journal of Machine Learning Research*, *16*(1), 3813–3847.

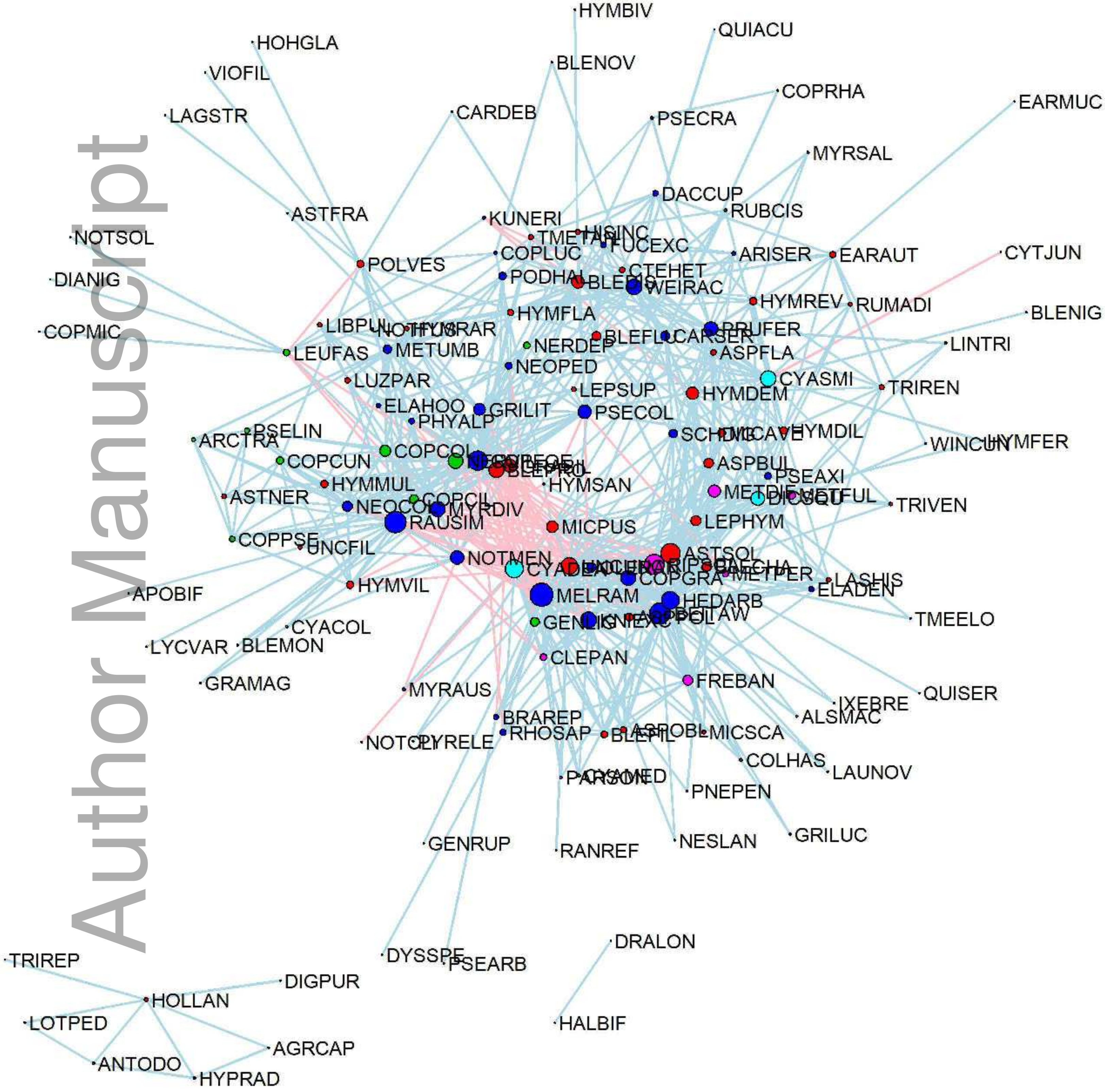

Correlation
(direct + indirect)
Precision
(direct only)
Graph
i)
B
A
C
0.3
0.55
B
-0.25
ii)
B
A
C
0.54
0.67
B
-0.52
iii)
C
A
B
iv)
B
A
C
-0.46
0.55
B
-0.25
v)
B
A
C
0.4
-0.5
B
0
vi)
C
A
B

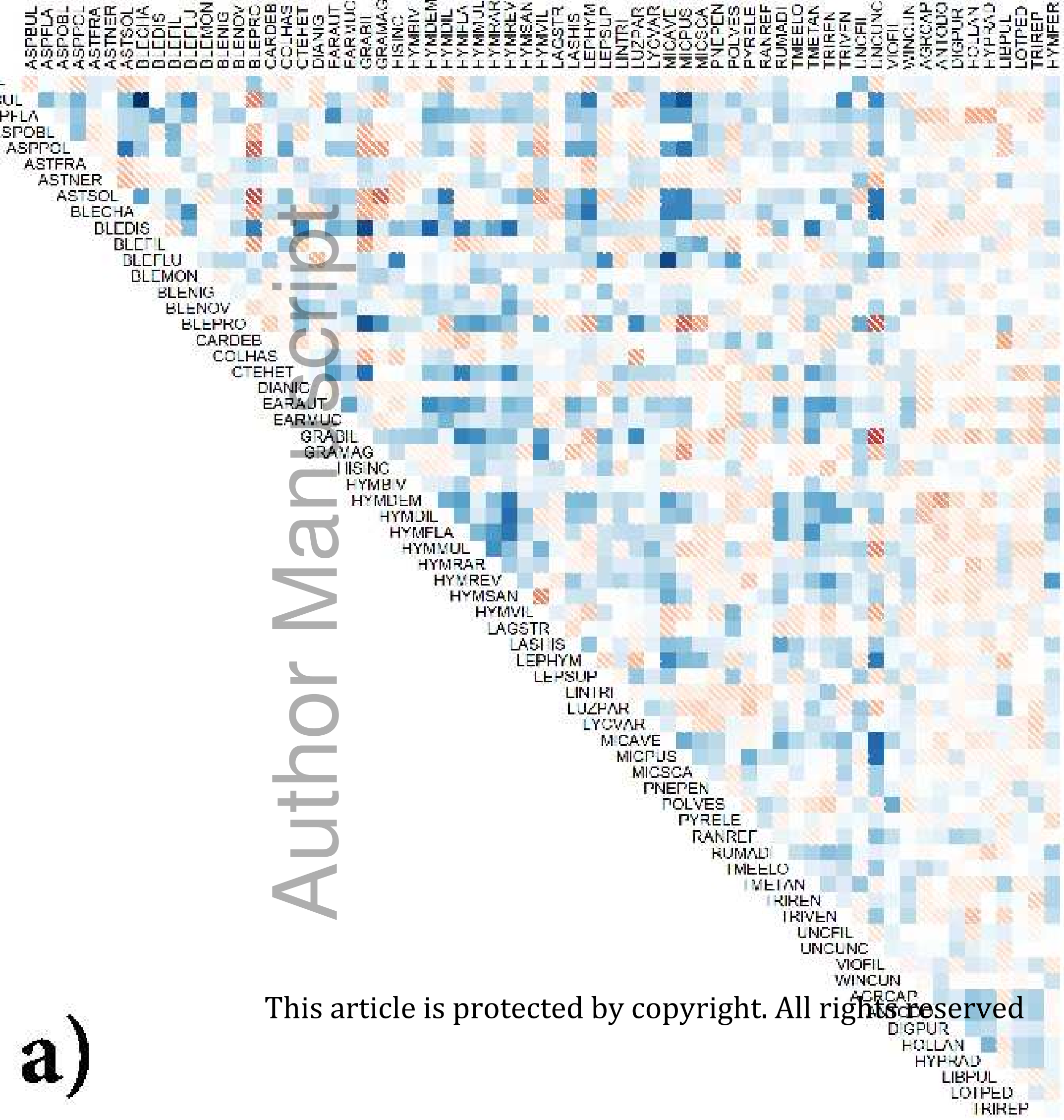

Count
0
1
2
3
Probability
0
0.2
0.4
0.6
0.8
1
a
b
0
1
2
3
4
Abundance A
c
d
1
0.8
0.6
0.4
0.2
0
Probability
4
2
0
Count
e
0
1
2
Abundance B

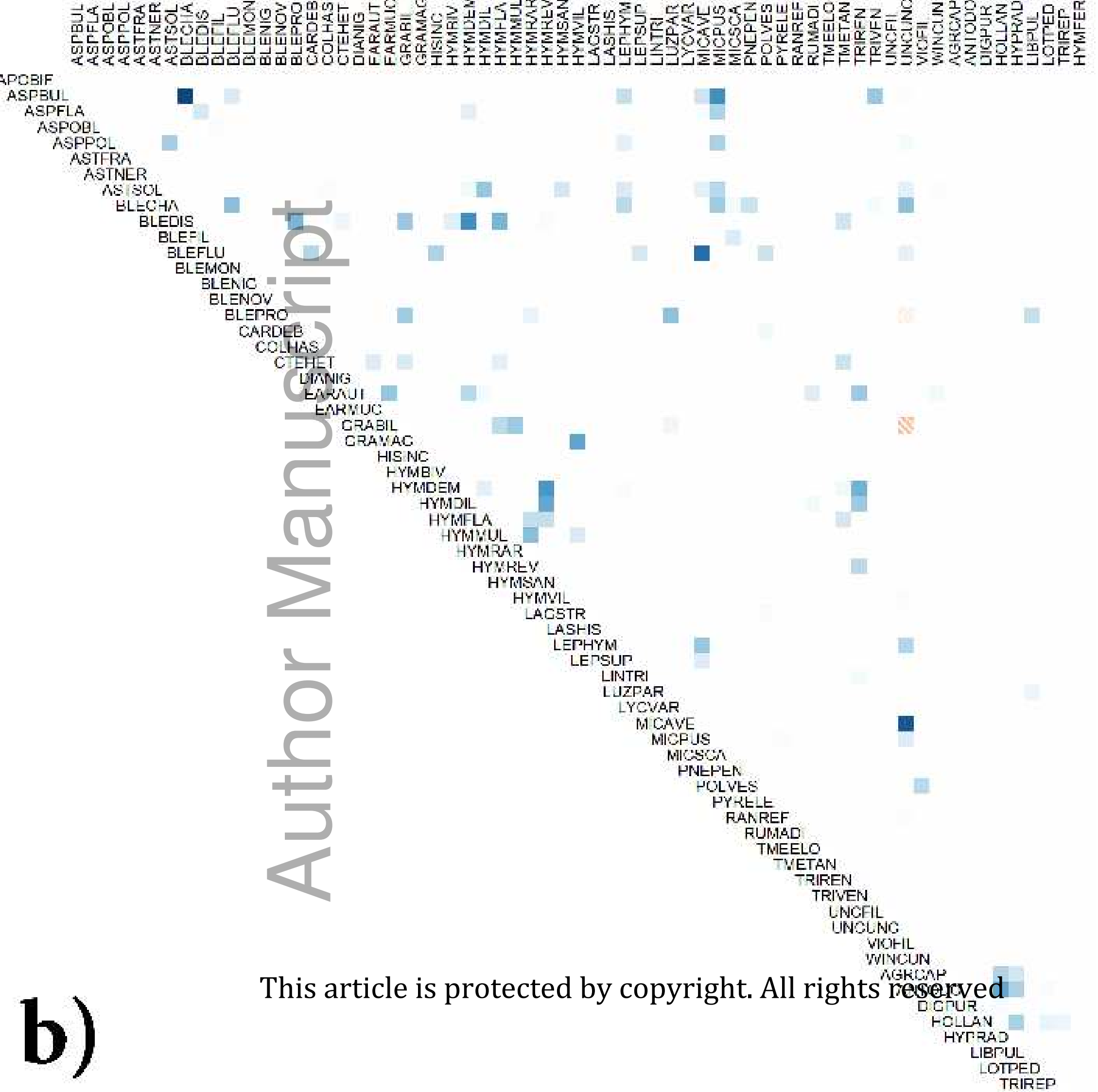

Uniform env
gllvm
0.18
0.38
0.2
0.47
Pairs
0.21
0.38
ggmlog
0.74
0.47
MRFcov
0.47
0.49
0.64
0.5
rosalia
0.47
0.59
0.39
0.26
0.14
0.21
ecoCopula
0.47
0.54
0.66
0.5
0.39
0.36
Hetrogeneous env
gllvm
0.06
0.2
0.01
0.14
Pairs
0.08
0.29
ggmlog
0.3
0.21
MRFcov
0.37
0.44
0.35
0.22
rosalia
0.25
0.48
0.14
0.06
0.11
0.14
ecoCopula
0.41
0.5
0.59
0.54
0.28
0.3
Estimation method
Binary GCGM
Binary Markov
Count GCGM
Count Markov
Ordinal GCGM
Tweedie GCGM
Simulation method


Correlation
(indirect + direct)
Precision
(direct only)
Graph
No environment
Environment
a)
b)
c)
d)
e)
f)
Alopcune
Alopfabr
Arctlute
Arctperi
Auloalbi
Pardlugu
Pardmont
Pardnigr
Pardpull
Trocterr
Zoraspin
Alopacce

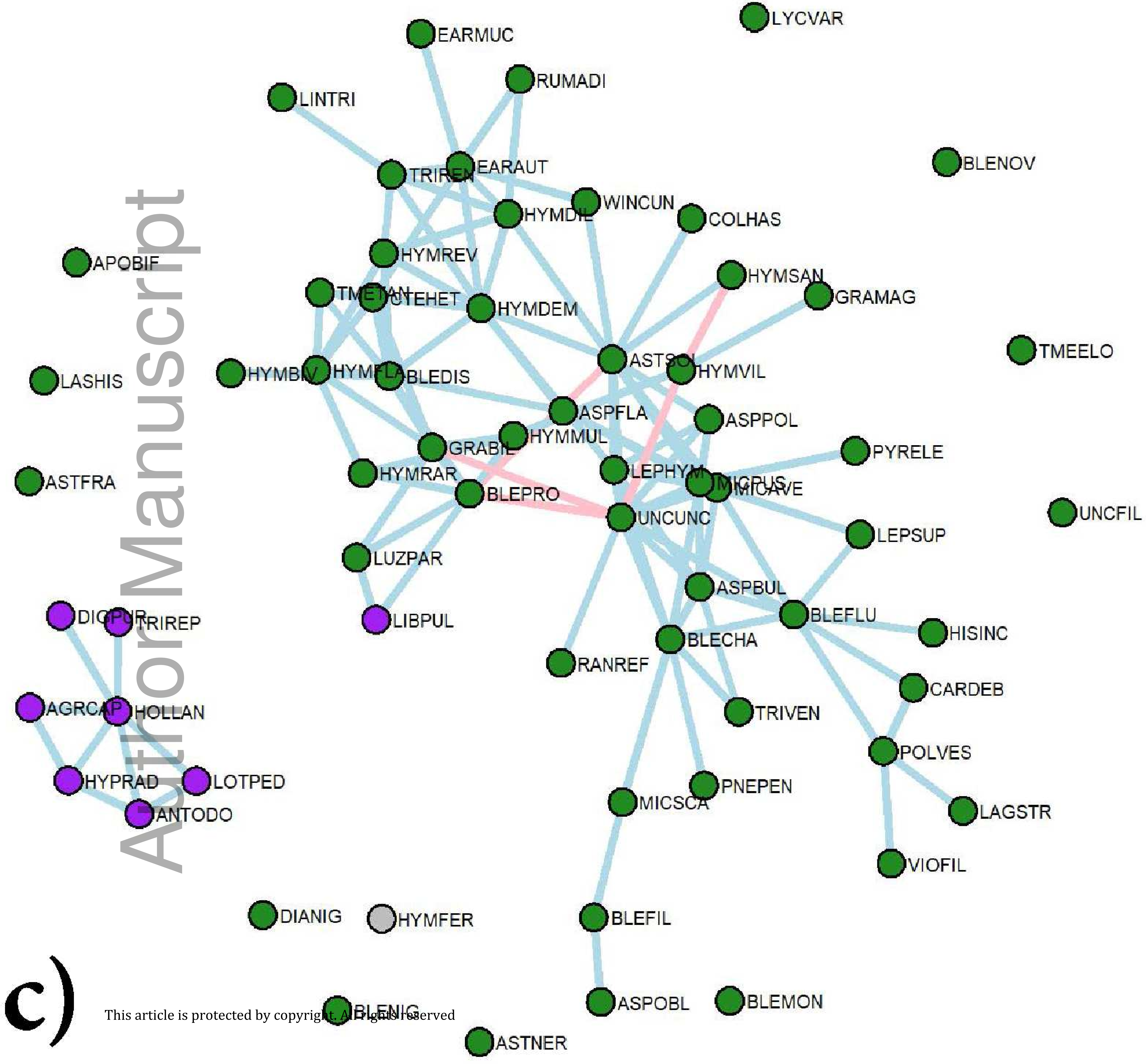

Abundance by presence of bare soil
Abundance on log scale
100
50
no bare soil
some bare soil
Alopacce
Alopcune
Alopfabr
Arctlute
Arctperi
Auloalbi
Pardlugu
Pardmont
Pardnigr
Pardpull
Trocterr
Zoraspin

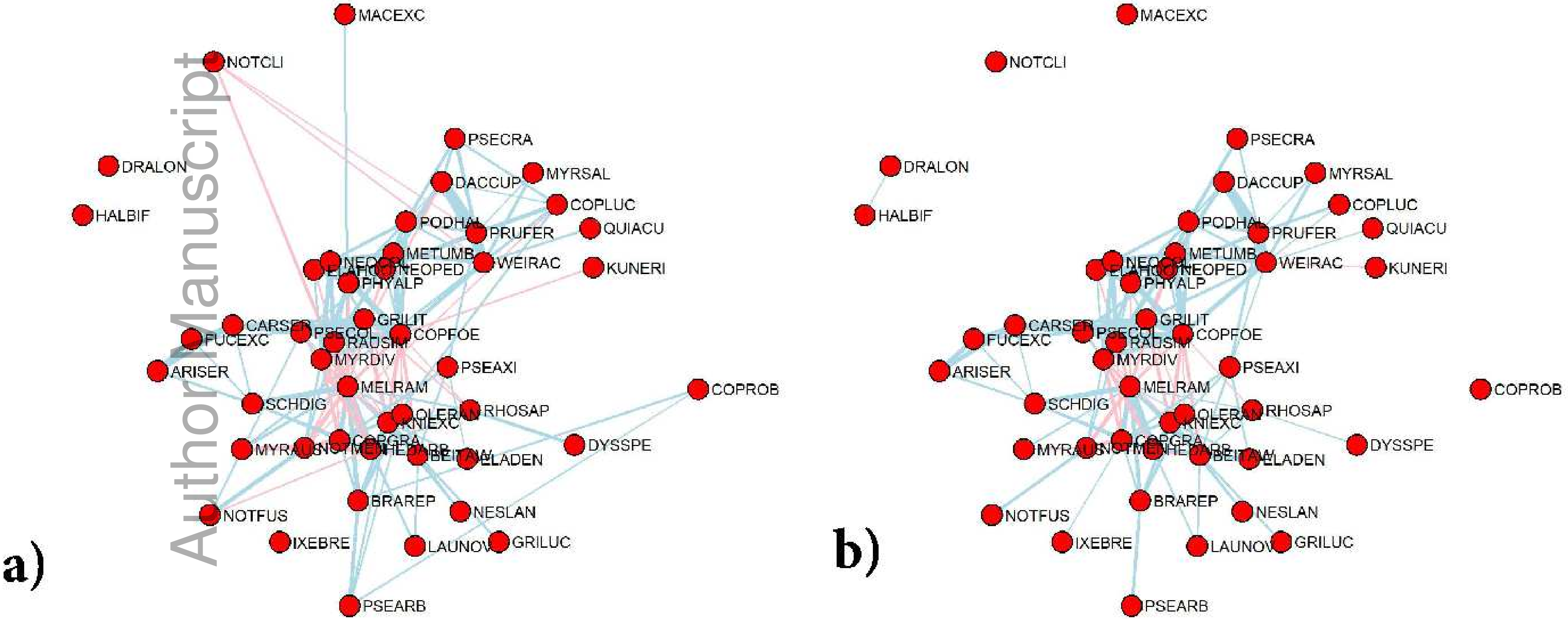

HYMBIV
QUIACU
HOHGLA
BLENOV
VIOFIL
COPRHA
EARMUC
LAGSTR
CARDEB
PSECRA
MYRSAL
DACCUP
ASTFRA
RUBCIS
KUNERI
NOTSOL
TMETAN
HISINC
FUCEXC
COPLUC
ARISER
EARAUT
CYTJUN
POLVES
DIANIG
CTEHET
PODHAL
BLEDIS
WEIRAC
HYMREV
RUMADI
HYMFLA
BLENIG
LIBPUL
NOTHUS
HYMRAR
COPMIC
BLEFLU
CARSER
PRUFER
LINTRI
NERDEP
METUMB
LEUFAS
ASPFLA
NEOPED
CYASMI
LUZPAR
TRIREN
LEPSUP
HYMDEM
ELAHOO
GRILIT
PSECOL
PHYALP
PSELIN
MICAVE
HYMDIL
ARCTRA
WINCUN
HYMFER
COPCOL
ASPBUL
COPCUN
BLEPRO
HYMSAN
PSEAXI
HYMMUL
METFUL
ASTNER
COPCIL
TRIVEN
NEOCOL
MYRDIV
RAUSIM
LEPHYM
MICPUS
COPPSE
UNCFIL
ASTSOL
NOTMEN
COPGRA
METPER
LASHIS
ELADEN
HYMVIL
APOBIF
MELRAM
HEDARB
TMEELO
CYACOL
LYCVAR
BLEMON
CLEPAN
GRAMAG
FREBAN
MYRAUS
IXEBRE
QUISER
ALSMAC
BRAREP
RHOSAP
BLEFIL
ASPOBL
MICSCA
PYRELE
COLHAS
LAUNOV
CYAMED
PNEPEN
GRILUC
GENRUP
NESLAN
RANREF
DRALON
TRIREP
DYSSPE
PSEARB
DIGPUR
HOLLAN
LOTPED
HALBIF
AGRCAP
ANTODO
HYPRAD


ASPBUL
ASPFLA
ASPOBL
ASPPOL
ASTFRA
ASTNER
ASTSOL
BLECHA
BLEDIS
BLEFIL
BLEFLU
BLEMON
BLENIG
BLENOV
BLEPRO
CARDEB
COLHAS
CTEHET
DIANIC
EARAUT
EARVUC
GRABIL
GRAMAG
HISINC
HYMBIV
HYMDEM
HYMDIL
HYMFLA
HYMMUL
HYMRAR
HYMREV
HYMSAN
HYMVIL
LAGSTR
LASHIS
LEPHYM
LEPSUP
LINTRI
LUZPAR
LYCVAR
MICAVE
MICPUS
MICSCA
PNEPEN
POLVES
PYRELE
RANREF
RUMADI
TMEELO
TMETAN
TRIREN
TRIVEN
UNCFIL
UNCUNC
VIOFIL
WINCUN
AGRCAP
ANTODO
DIGPUR
HOLLAN
HYPRAD
LIBPUL
LOTPED
TRIREP
HYMFER
a)


ASPBUL
ASPFLA
ASPOBL
ASPPOL
ASTFRA
ASTNER
ASTSOL
BLECHA
BLEDIS
BLEFIL
BLEFLU
BLEMON
BLENIG
BLENOV
BLEPRO
CARDEB
COLHAS
CTEHET
DIANIG
FARAUT
FARMUC
GRABIL
GRAMAC
HISINC
HYMBIV
HYMDEM
HYMDIL
HYMFLA
HYMMUL
HYMRAR
HYMREV
HYMSAN
HYMVIL
LACSTR
LASHIS
LEPHYM
LEPSUP
LINTRI
LUZPAR
LYCVAR
MICAVE
MICPUS
MICSCA
PNEPEN
POLVES
PYRELE
RANREF
RUMADI
TMEELO
TMETAN
TRIREN
TRIVEN
UNCFIL
UNCUNC
VIOFIL
WINCUN
AGRCAP
ANTODO
DIGPUR
HOLLAN
HYPRAD
LIBPUL
LOTPED
TRIREP
HYMFER
APOBIF
ASPBUL
ASPFLA
ASPOBL
ASPPOL
ASTFRA
ASTNER
ASTSOL
BLECHA
BLEDIS
BLEFIL
BLEFLU
BLEMON
BLENIG
BLENOV
BLEPRO
CARDEB
COLHAS
CTEHET
DIANIG
FARAUT
FARMUC
GRABIL
GRAMAC
HISINC
HYMBIV
HYMDEM
HYMDIL
HYMFLA
HYMMUL
HYMRAR
HYMREV
HYMSAN
HYMVIL
LACSTR
LASHIS
LEPHYM
LEPSUP
LINTRI
LUZPAR
LYCVAR
MICAVE
MICPUS
MICSCA
PNEPEN
POLVES
PYRELE
RANREF
RUMADI
TMEELO
TMETAN
TRIREN
TRIVEN
UNCFIL
UNCUNC
VIOFIL
WINCUN
AGRCAP
ANTODO
DIGPUR
HOLLAN
HYPRAD
LIBPUL
LOTPED
TRIREP

b)


LYCVAR
EARMUC
RUMADI
LINTRI
BLENOV
TRIREN
EARAUT
WINCUN
HYMDIL
COLHAS
APOBIF
HYMREV
HYMSAN
GRAMAG
TMETAN
CTEHET
HYMDEM
TMEELO
LASHIS
HYMBIV
HYMFLA
BLEDIS
ASTSOL
HYMVIL
ASPFLA
ASPPOL
HYMMUL
GRABIL
PYRELE
ASTFRA
HYMRAR
LEPHYM
MICPUS
MICAVE
BLEPRO
UNCUNC
UNCFIL
LEPSUP
LUZPAR
ASPBUL
DIGPUR
TRIREP
LIBPUL
BLEFLU
HISINC
BLECHA
RANREF
CARDEB
AGRCAP
HOLLAN
TRIVEN
POLVES
HYPRAD
LOTPED
PNEPEN
MICSCA
LAGSTR
ANTODO
VIOFIL
DIANIG
HYMFER
BLEFIL
c)
BLENIG
ASPOBL
BLEMON
ASTNER


a)
MACEXC
NOTCLI
DRALON
HALBIF
PSECRA
MYRSAL
DACCUP
COPLUC
PODHAL
PRUFER
QUIACU
METUMB
NEOCOL
ELAHOO
NEOPED
WEIRAC
KUNERI
PHYALP
GRILIT
CARSER
PSECOL
RAUSIM
COPFOE
FUCEXC
MYRDIV
ARISER
PSEAXI
MELRAM
COPROB
SCHDIG
OLERAN
RHOSAP
KNIEXC
DYSSPE
COPGRA
MYRAUS
NOTMEN
HEDARB
BEITAW
ELADEN
BRAREP
NESLAN
NOTFUS
IXEBRE
LAUNOV
GRILUC
PSEARB
b)
MACEXC
NOTCLI
DRALON
HALBIF
PSECRA
MYRSAL
DACCUP
COPLUC
PODHAL
PRUFER
QUIACU
METUMB
NEOCOL
ELAHOO
NEOPED
WEIRAC
KUNERI
PHYALP
GRILIT
CARSER
PSECOL
RAUSIM
COPFOE
FUCEXC
MYRDIV
ARISER
PSEAXI
MELRAM
COPROB
SCHDIG
OLERAN
RHOSAP
KNIEXC
DYSSPE
COPGRA
MYRAUS
NOTMEN
HEDARB
BEITAW
ELADEN
BRAREP
NESLAN
NOTFUS
IXEBRE
LAUNOV
GRILUC
PSEARB