S1 Supplementary Methods

S1.1 Model code and parameterisation

We make the following observations in addition to comments in the code. Firstly, identifiable interactions (beta_ij in the code) are defined as a vector, which must then be matched to their correct position in the interaction matrix. This is the role of the icol and irow vectors defined in the data block. Our Github repository also contains the data_prep.R file, which will show how to create these vectors from appropriate input data. Secondly, the joint_model.stan code is written to use a negative binomial function (lines 86 and 88, and again 102 and 104) with an exponential inverse link (lines 56 and 68), the latter is equivalent to a log link function. These lines, along with the dispersion parameter (lines 29 and 77, and wherever the negative binomial function is used) must all be modified if a different family and link function are to be used. Thirdly, we impose the following constraints to improve convergence, avoid over-parameterisation and maintain identifiability of our parameters r_i and e_j (Huber et al. 2004; Kidziński et al. 2021; Niku et al. 2021). We define both the effect and the response parameters as unit vectors, which means we only require T-1 degrees of freedom (where T is the total number of neighbour elements) to estimate all effect values and S-1 degrees of freedom (where S is the total number of focal elements) to estimate all response values. This loss of a degree of freedom arises from the fact that the matrix whose i, j-elements are all products $r_i \times e_j$ is of rank 1. Though both latent variables r_i and e_j may be susceptible to sign switching when running the model on multiple chains, estimates for identifiable and unidentifiable interactions should not be affected. All r_i and e_j's are multiplied by a a unique, positive constrained $weight$ parameter which controls the average strength of the resulting interactions.

S1.2 Assessing parameter identifiability

We assess which interactions parameters are identifiable in the data_prep.R file (lines 30-44) and store this information in the matrix Q. To do so, we first create a model matrix X_i for each focal species i. This matrix has as many rows as performance measurements for focal species i and $t + 1$ columns, where t is the total number of neighbouring species across the full dataset. Its first column is a vector of 1's (representing the intercept) and all remaining columns denote the neighbour abundances recorded

for each observation. We transform this matrix into its row-reduced echelon form R_i using the rref() function from the pracma package (Borchers 2022), and we then left multiply R_i by its transpose to produce a $t + 1 \times t + 1$ symmetric matrix Z_i. We can use this matrix Z_i to determine which neighbour abundances are linearly independent of all other neighbour abundances as follows. We ignore the first row of Z_i because it corresponds to the intercept which we always want to include in our model. For every other row $k > 1$, its corresponding predictor is linearly independent if and only if it consists entirely of 0's except for its k'th element. If this is true, the interaction parameter corresponding to that row (neighbour abundance) is thus identifiable and we assign it a value of 1 in the i'th row and $k - 1$'th column of Q. Alternatively, the corresponding element in Q is given a value of 0. Examining the Z_i matrix of each focal species i therefore allows us to construct the full Q matrix of inferrable pairwise interaction parameters. Given that Z_i is symmetric, note that the checks for linear independence could equivalently be done using the columns of Z_i.

In a similar way, we also evaluate whether predictors (neighbours) are linearly independent across the entire dataset in order to be able to correctly distinguish their respective effect parameters e_j for the RIM. The same procedure as above is performed across the entire model matrix X (without subsetting for each focal species) and neighbours in the entire matrix Z are evaluated for linear independence. This is done early on in the master.R file (lines 27-37) as a check before transforming the data into the format required by STAN. Given enough data, a lack of linear independence across the entire dataset is very unlikely; indeed, it would only arise if multiple neighbours' densities were perfectly correlated across all neighbourhoods.

S1.3 Evaluating model convergence for the case study data

The NDD-only model, the RI-only model and the joint model were all run on the case study data with 4 chains and 7000 iterations, of which the first 5000 were discarded. The NDDM converged well with $\hat{R} < 1.01$ for all parameters. Convergence for the RIM-only model was less ideal with few γ and B parameters reaching $\hat{R}$ values below 1.01, with a median $\hat{R}$ of 1.35 for values in γ and of 1.13 for elements of B. Convergence for the joint model was improved over the RIM, especially for the interaction parameters, with a median $\hat{R}$ of 1.24 for values in γ and of 1.005 for elements of B. This was achieved by setting

`adapt_delta = 0.99` and `maximum_treedepth = 20`; we did not increase these values beyond as this was unlikely to resolve all convergence difficulties. In addition to the lack of other warnings and the good convergence of these models on simulated data, it stands to reason that the lack of convergence observed here is at least partially due to a difficult geometry of the posteriors. Both the r and e parameters are constrained due to their parameterisation as unit vectors; this means that if a neighbour j only engages in few interactions, uncertainty about its e_j can create uncertainty for the values of e for other neighbours. Similarly, a shift in either r_i or in the values of e_j for all observed neighbours of i can produce similar values in B. If one value of r_i gets shifted during sampling, this will necessarily shift all other r values, which may impact e_j and then in turn all other e values, which can then also affect γ_i. The degree to which this correlation is problematic for $\hat{R}$ will depend intimately on the data at hand.

This lack of convergence leads to multi-modal posteriors for those parameters with higher $\hat{R}$, as different chains get "stuck" in slightly different local optima. For our case study data, we observed that these local optima were very close and overlapping in both parameter values and in the likelihood which is being maximised (Fig. S1). Indeed, the Monte Carlo standard error for all parameters always remained lower than the posterior standard deviation, for both the RIM-only and joint model. This means that all parameters were estimated to a higher accuracy than their standard error, and further simulation was unlikely to substantially modify posterior inferences. Furthermore, posterior intervals for these parameters were similar across all models and strongly overlapping. Fig. S2 shows the posterior intervals of all γ parameters across all three models to illustrate this behaviour (showing this for the B parameters would be unwieldy due to the high number of pairwise interactions in our case study). Comparison between all three model estimates was only possible because, whilst the NDDM sets to 0 all interactions where species were observed to co-occur but the interactions remained unindentifiable, those accounted for less than 5% of all interactions in the dataset.

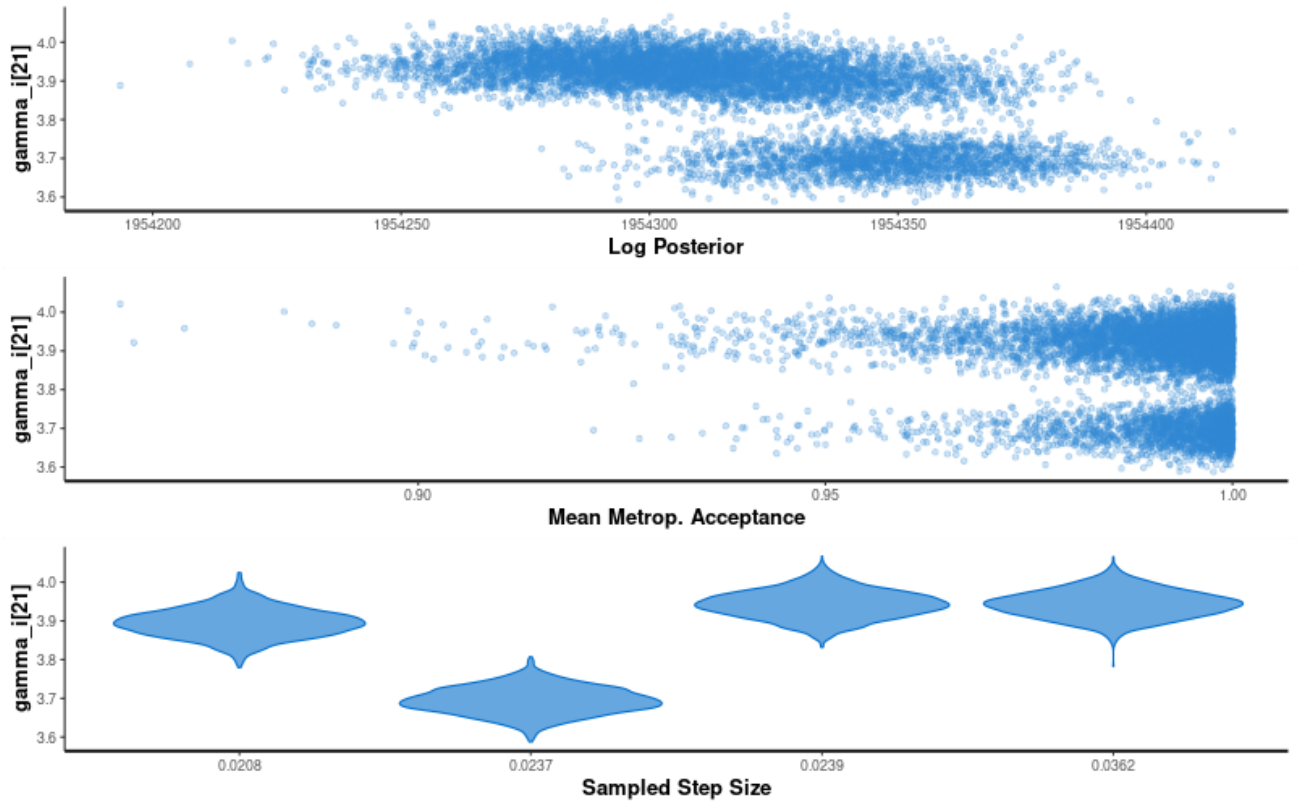

Figure S1: Diagnostic plot for the least convergent γ_i parameter ($\hat{R} = 3.21$) in the joint model. For all graphs, values for that γ_i are given on the y-axis. The first graph shows a scatterplot of γ_i against the accumulated log-posterior on the x-axis, whilst the second graph is a scatterplot of γ_i against the average Metropolic acceptance rate. Each point corresponds to one sample. The bottom graph is a violin plot of the distribution of γ_i at each of the sampled step sizes (one per chain). This figure illustrates that despite the chains not converging on the exact same values for γ_i, estimates are very close between chains and the associated log posteriors overlap. No divergent transitions were present. The figure was plotted using the `stan_par()` function from the rstan package.

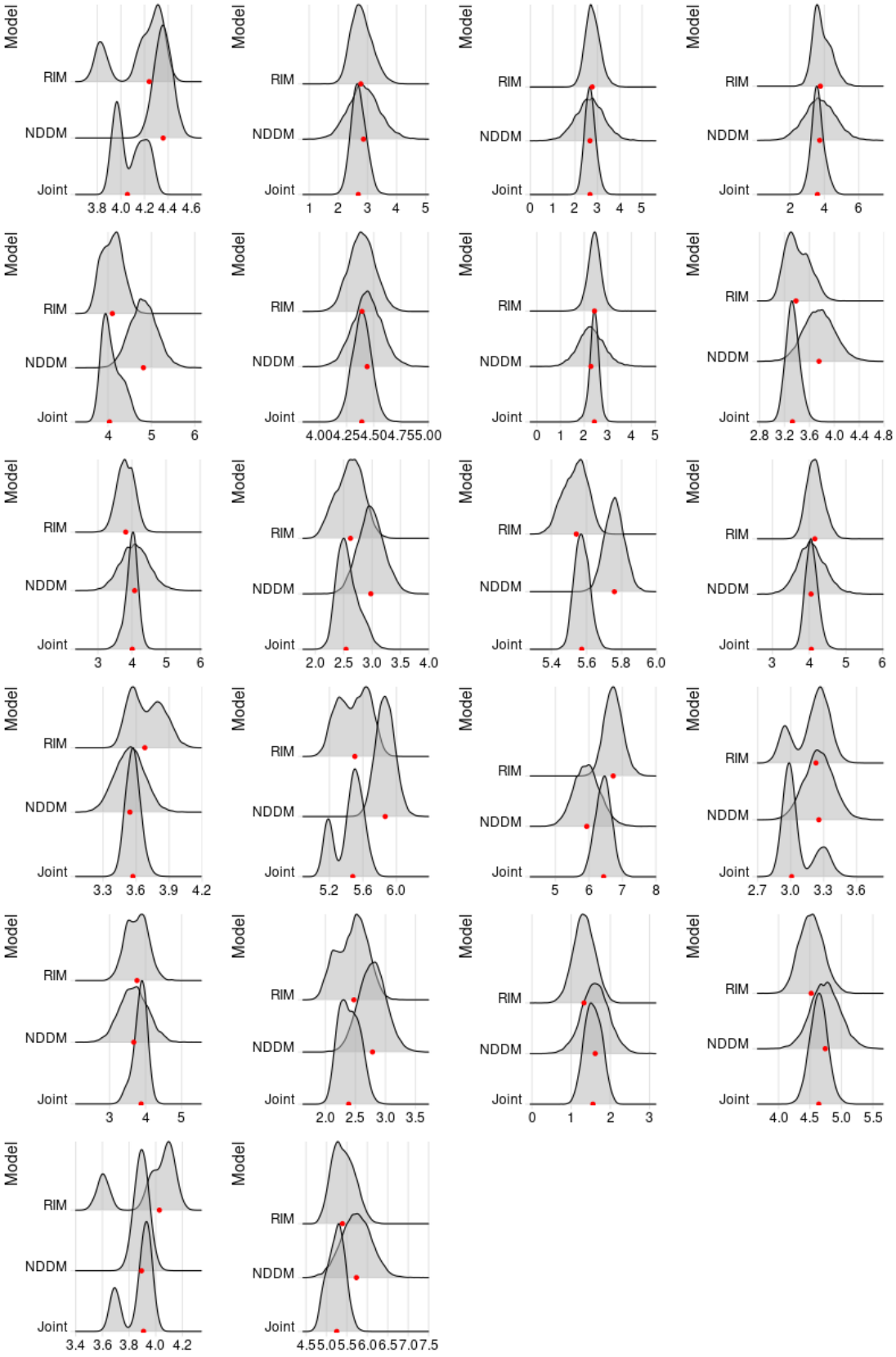

Figure S2: See caption on next page.

Figure S2: (Caption for previous page.) Posterior interval distributions for all 22 γ parameters as returned by the RIM-only, NDDM-only and joint model. For all graphs, posterior values are given on the x-axis and correspond to $\log(\gamma)$. The red dot indicates the median of each distribution. The γ_i with the highest $\hat{R}$ in the joint model ($\hat{R} = 3.21$) is shown in the bottom-most left corner. Even when posteriors are multi-modal, note that posterior distributions constrast considerably to these parameters' assumed prior distribution.

S1.4 Case study Data

S1.4.1 Community data

We applied this framework to annual wildflower community dataset from Western Australia. This system is a diverse and well-studied community of annual plants which germinate, grow, set seed and die within approximately 4 months every year. Individual fecundity data were collected in 2016, when 100 50 x 50 cm plots established in the understory of West Perenjori Reserve (29^{o}28'01.3"S 116^{o}12'21.6"E) were monitored over the length of the full field season. The resulting dataset includes between 29 to over 1000 counts of individual plant seed production from 22 different focal species (with a median of 108 observations per species), in addition to the identity and densities of all neighbouring individuals within the interaction neighbourhood of each focal plant. Interaction neighbourhoods varied in radius from 3 to 5 cm depending on the size of the focal species (Martyn 2020). Total neighbourhood diversity was 71 wildflower species, 19 of which were recorded fewer than 10 times across the whole dataset. The species-specific effects of this latter group of species on focals were deemed negligible due to their extremely low density, they were thus grouped into an 'other' category and their effects on focals averaged. This resulted in 53 potential neighbour identities. Plots were randomly positioned across the reserve to capture natural environmental variation in the system. Major factors known to impact plant composition and within species abundances are soil P concentrations, shade, and the presence of woody debris (Dwyer et al. 2015). Natural levels of soil P are very low in this system with elevated P concentrations occurring predominantly on the edge of the reserve near agricultural fields that use P fertiliser. All study plots

were placed far from reserve edges to avoid artificially elevated levels of P, which are easy to detect due to an association between high P levels and extensive exotic grass abundances, which are largely absent from the core reserve where this study was conducted. Thus, environmental variation captured by this dataset reflects natural variation in tree cover, soil nutrients and woody debris. To account for confounding effects between plot location and plant density, half of all plots were thinned (a quarter to 60% abundance and a quarter to 30%). Thinning did not target any particular species.

In study systems which allow it, a proportion of interaction neighbourhoods can instead be thinned prior to the experiment to randomly remove neighbouring individuals and provide observations for low-density estimates of interactions. Though this steps is not strictly necessary, thinning certain neighbourhoods can also reduce potential confounding effects between the environment and interactions and thus provide more accurate estimates of interaction effects. Environmental data known to affect performance can also be recorded and included in the model (Bimler et al. 2018) to minimise those confounding effects.

We required species demographic rates (seed survival and germination) in order to scale model interaction estimates into interaction strengths. Species demographic rates for 16 of our focal species were estimated from a database of field experiments carried out between 2016 and 2019 where seedbags were placed in the field to estimate germination rates, and ungerminated seeds were evaluated in the lab for survivability. The remaining species were assigned mean demographic rates from these experiments. Further details on the methods used for collecting those seed rates are available in section S1.4.2.

S1.4.2 Seed germination and survival data

Seed demographic rates were collected from a set of field experiments conducted by T. Martyn, M. Raymundo and I. Towers at Perenjori reserve between 2015 and 2019. Experiments differed both in the methods and in which focal species were included in ways which are detailed further below, such that each focal species had a different number of replicates across all experiments. Given how much seed rates estimates have been found to vary within species and according to a range of both individual and environmental factors, we chose to average results from these multiple experiments for each focal

species in order to provide a point estimate which captures a wide range of conditions under which seeds may grow. For those species which did not have any field estimates of seed rates (*Austrostipa elegantissima*, *Erodium sp.*, *Petrorhagia dubia*) or seed survival rate (*Gilberta tenuifolia*), no replication (*Waitzia acuminata*) or an unrealistically low estimate of germination rate (*Goodenia pusilliflora*), we substituted the community mean instead.

For each experiment, mature seeds were collected at the end of the growing season (September - October) from multiple populations of each focal species located throughout the reserve. Immature or damaged seeds were not included, and collected seed was homogenised for each focal species to eliminate bias associated with local adaptation within populations. Germination rate was estimated by planting seeds in the field along gradients of soil phosphorus, woody canopy cover and herbaceous vegetation density and either directly counting the number of seeds which had germinated or comparing recruitment rates to unplanted plots after a sufficient amount of time had elapsed. Seeds were planted during late-September to mid-October, mimicking natural seed dispersal timing for wildflowers in the area. Seed survival rates were estimated using either the remaining seeds or a separate batch of seeds and assessing viability of the seeds using tetrazolium staining.

T. Martyn experiment: For 19 focal species (the full species list excluding *A. elegantissima*, *P. dubia* and *G. tenuifolia*), germination bags containing 20 seeds each were planted in the field in 2016 across multiple areas of Perenjori Reserve. Out of the 30 bags, 19 were collected in 2017 and the remaining were collected in 2018. Due to a severe drought in 2017, half of the bags collected that year were watered during the field season and prior to collection. Germination bags were then brought back to the Mayfield Lab facilities at the University of Queensland, Brisbane, and seeds extracted. Seeds were examined for signs of germination in the field (broken or empty seed coat) and those remaining were placed in germination trays and a germination chamber to mimic light and temperature conditions conducive to germination. Trays were watered with Gibberellic acid once to twice a week and seedlings were recorded and removed until no more seedlings emerged. Remaining, ungerminated seeds were then assessed as dead (mouldy) or potentially viable. The remaining potentially viable seeds were assessed for viability using tetrazolium staining, contributing to our estimates of seed survival rates. For this procedure, embryos in each seed

were exposed by either removing the seed coat or by creating a thin cut along the seed coat. The exposed embryos were then placed on a six-well germination plate and 2 ml of 0.25% Tetrazolium solution was added to each well to stain the embryos, before covering them and storing them at 25°C overnight. To check for staining, embryos were dissected under a dissecting microscope. Viable seeds showed a dark pink embryo while non-viable seeds did not stain or were stained in a splotchy way.

M. Raymundo experiment: This experiment was carried out on the focal species *H. glutinosum*, *T. cyanopetala*, *T. ornata* and *V. rosea* from 2015 to 2017. However, a severe drought in 2017 made the second round of data collection impossible and thus we only include results for 2016 here. Ten plots were established measuring 0.5 m x 0.5 m at each of three sites in Perenjori Reserve for a total of 30 plots. Each plot was divided into 25 0.1 m x 0.1 m subplots and focal species were randomly assigned a subplot in each plot. Thirty seeds of each focal species were planted in the designated subplot in late September 2015 and a plastic ring 10 cm in diameter and 1 cm high was placed in each subplot where seeds were added to limit seed movement among subplots. Another five subplots were assigned plastic rings to serve as controls for the effect of the rings on non-experimental communities. The remaining 15 subplots served as controls where no seeds or rings were added allowing for recruitment from either natural dispersal or from the seed bank. Blocks were placed in such a way as to span shaded and open areas, bare ground and dense herbaceous vegetation, and areas with native dominated and exotic dominated assemblages. Before implementing the experiment in 2015, plots were surveyed to record the number and identity of all adult plants in each subplot. Due to the randomisation of seed addition into subplots, some subplots had focal species already in them. As all focal species were common to this reserve, it is also likely that seeds for all species were in the seed banks in at least some subplots. There was no way to determine this in advance, though when adult individuals of a focal species were present in a subplot prior to the implementation of our experiment, we expected that some seedlings in the following year would be from the seedbank as well as our planted seeds and looked for evidence of this (more than 30 individuals) in data from 2016. We therefore compared average densities of successful focal recruits and those which emerged in situ between sown, control, and ringed subplots to assess seed limitation and germination rate. To measure seed survival rates, thirty seeds of each focal species were also assessed for viability using tetrazolium staining using the same procedure as for the T. Martyn experiment.

I. Towers experiment: This experiment was carried out on the focal species *A. calendula*, *G. berardiana*, *H. glutinosum*, *H. glabra*,*P. aroides*, *P. debilis*, *P. canescens*, *T. cyanopetala*, *T. ornata*, *V. rosea* and *W. acuminata* in 2018 and 2019. Pairs of free-draining germination trays were deployed across a gradient of canopy cover in mid-October of both years, filled with soil, which had either been collected from the field and heat-sterilised to render pre-existing seeds nonviable (2019), or simply collected from the roadside (2018). Each germination tray consisted of 24 cells, with two cells randomly assigned to each focal species. In each cell, 15 seeds of the designated focal species were broadly distributed and lightly misted with water to facilitate seed-soil contact and minimise removal by wind. Trays placed in 2018 used seeds collected at the end of the 2017 growing season and dry after-ripened at 60°C for a month before being stored in cool, dry conditions at the University of Queensland. Seeds planted in 2019 were collected at the end of the 2018 growing season and were placed directly from the field into the germination trays. To re-establish microbial communities for those trays where the soil had been heat-treated, seeds were lightly covered with a small amount of untreated soil collected from the site in which they were buried. Untreated soil was collected from directly underneath coarse woody debris in patches where it was present as prior research in this system has shown that the effect of coarse woody debris on plant performance is partially attributable to debris-specific soil microbial communities (A. Pastore unpublished data). Some of the trays received the additions of leaf litter, but the results of this treatment were not included for the seed rates used in this study. Seed germination rate was measured by counting the number of seedlings which emerged in the field, but seed survival rate was not calculated in this experiment.

S1.5 A model for annual plant population dynamics

The joint model framework returns species-specific estimates of intrinsic fitness (γ_i), as well as as a focal species $\times$ neighbour species composite matrix B of identifiable ($\beta_{i,j}$) and unidentifiable ($r_i e_j$) interaction estimates which quantify the effects of one neighbour j on the intrinsic fitness of a focal species i. Though useful as they are, these estimates can lead to a wider range of potential applications when integrated into models of population dynamics. For example, we might be more interested in the effects of neighbours on the density or growth rate of a focal species rather than on its proxy for lifetime

reproductive success. Importantly, it is necessary to specify a model describing population dynamics in order to draw conclusions about the effects of interactions and network structure on the maintenance of community diversity and stability.

We defined the following model for annual plants with a seed bank (Levine and HilleRisLambers 2009; Mayfield and Stouffer 2017; Bimler et al. 2018) which describes the rate of change in a focal species' *i* abundance of seeds in a seed bank from one year to the next:

$$\frac{N_{i,t+1}}{N_{i,t}} = (1 - g_i)\, s_i + g_i F_{i,t} \tag{S1}$$

where $F_{i,t}$ measures the number of viable seeds produced per germinated individual whilst g_i and s_i are the seed germination and seed survival rate, respectively. In a simplified case where the focal species *i* interacts with only one other species *j*, our use of a log link function implies that $F_{i,t}$ in this model of population dynamics is given by:

$$F_{i,t} = \lambda_i e^{-\alpha_{ii} g_i N_{i,t} - \alpha_{ij} g_j N_{j,t}} \tag{S2}$$

where λ_i corresponds to seed number in the absence of interaction effects, and α_{ii} and α_{ij} are the interaction strengths between species i and its intraspecific and interspecific neighbours respectively. Here it is α_{ij} and α_{ii} which are equivalent to $B_{i,j}$ in Eq. 3 of the main text.

We determine the scaled interaction strengths α'''s by including λ_i, g_i and s_i in such a way that these variables are cancelled out when the α'''s are substituted for the α's in our annual plant population model (Godoy and Levine 2014; Bimler et al. 2018).

$$\alpha''_{ij} = \frac{g_j \alpha_{ij}}{ln(\eta_i)} \tag{S3}$$

with $\eta_i = \frac{\lambda_i g_i}{\theta_i}$ and $\theta_i = 1 - (1 - g_i)(s_i)$. Note that our model evaluates the rate of change of seeds in the seed bank, and this is reflected in the scaling terms used to compare interaction strengths between focal

species. Substituting α'''s for α's in Eq. S1 gives us:

$$\frac{N_{i,t+1}}{N_{i,t}} = (1 - \theta_i) + \theta_i \eta_i e^{-ln(\eta_i)(\alpha''_{ii} N_{i,t} + \alpha''_{ij} N_{j,t})} \tag{S4}$$

where we can see that the α'''s are directly proportional to the density of neighbours. Relating this population model to the joint model framework, we recover the following:

$$B_{i,j} = \alpha_{ij} \tag{S5}$$

$$\gamma_i = ln(\lambda_i) \tag{S6}$$

$$\eta_i = \frac{e^{\gamma_i} g_i}{\theta_i} \tag{S7}$$

$$ln(\eta_i) = \gamma_i \left(ln(g_i) - ln(\theta_i)\right) \tag{S8}$$

$$\alpha''_{i,j} = \frac{g_j B_{i,j}}{ln(\eta_i)} = \frac{g_j B_{i,j}}{\gamma_i \left(ln(g_i) - ln(\theta_i)\right)} \tag{S9}$$

As we show here, the exact form of the rescaled interactions as well as intrinsic fitness can therefore vary depending on the specific population dynamic model applied and may include other demographic rates which reflect species-level differences in growth and mortality. In our case study, we were only able to scale interactions between pairs of focal species because we required demographic rates for both species involved in the interaction (Eq. S9) and rates for non-focal species were not available. Because intrinsic fitness is estimated by the model framework and not directly observed, we used the mean of the γ_i posterior distribution returned by our model in our scaling of the interaction coefficients.

References

Bimler, M. D. et al. (2018). "Accurate predictions of coexistence in natural systems require the inclusion of facilitative interactions and environmental dependency". In: *Journal of Ecology* 106.5, pp. 1839–1852. DOI: 10.1111/1365-2745.13030.

Borchers, H. W. (2022). *pracma: Practical Numerical Math Functions. R package version 2.3.8*. URL: https://cran.r-project.org/package=pracma.

Dwyer, J. M. et al. (2015). "Climate moderates release from nutrient limitation in natural annual plant communities". In: *Global Ecology and Biogeography* 24.5, pp. 549–561. DOI: 10.1111/geb.12277.

Godoy, O. and Levine, J. M. (2014). "Phenology effects on invasion success: Insights from coupling field experiments to coexistence theory". In: *Ecology* 95.3, pp. 726–736. DOI: 10.1890/13-1157.1.

Huber, P., Ronchetti, E., and Victoria-Feser, M.-P. (2004). "Estimation of Generalized Linear Latent Variable Models". In: *Journal of the Royal Statistical Society . Series B (Statistical Methodology)* 66.4, pp. 893–908.

Kidziński, Ł. et al. (2021). "Generalized Matrix Factorization". In: *arXiv preprint*.

Levine, J. M. and HilleRisLambers, J. (2009). "The importance of niches for the maintenance of species diversity." In: *Nature* 461.7261, pp. 254–257. DOI: 10.1038/nature08251.

Martyn, T. E. (2020). "Understanding the role of direct and indirect interactions in mediating local plant diversity". PhD Thesis. University of Queensland, p. 246. DOI: https://doi.org/10.14264/uql.2020.824.

Mayfield, M. M. and Stouffer, D. B. (2017). "Higher-order interactions capture unexplained complexity in diverse communities". In: *Nature Ecology & Evolution* 1, pp. 1–7. DOI: 10.1038/s41559-016-0062.

Niku, J. et al. (2021). "Analyzing environmental-trait interactions in ecological communities with fourth-corner latent variable models". In: *Environmetrics* 32.April, e2683. DOI: 10.1002/env.2683.