

STRUCTURAL CONTROLLABILITY OF POLLINATION NETWORKS

E. Fernando Cagua, Kate L. Wootton, Johanna Voinopol-Sassu, and Daniel B. Stouffer

Center of Integrative Ecology, School of Biological Sciences, University of Canterbury

Introduction

Here we use plant pollinator communities to investigate the number of species that should be managed to control population dynamics of the whole community, the characteristics that determine whether a species should be managed or not and how invasive species fit.

Methods

Maximum matching

The number of nodes necessary to fully control a complex network can be calculated by counting the number of unmatched nodes in the network's maximum matching (Liu, Slotine, and Barabási 2011). In a directed network a matching is a subset of links in which no two links share a common starting node or a common ending node and node is matched if it is the ending node of one of the links in the matching (Figure 1a). A matching is maximum if the number of matched links (matching size) is maximal, and the sum of the weights of the matched links (matching weight) is the largest possible among all possible matchings (West 2001).

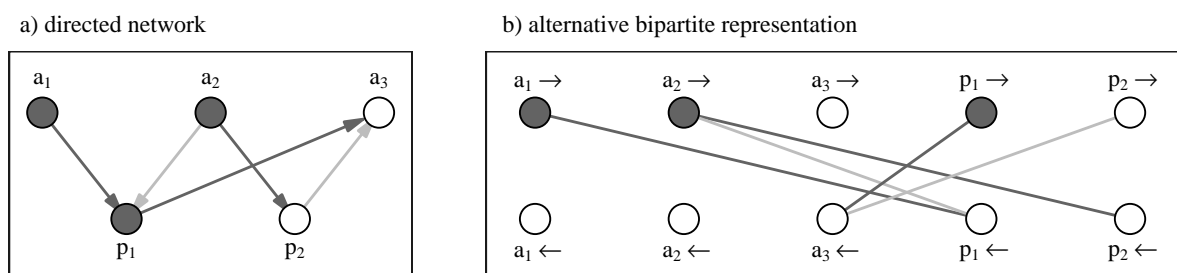


Figure 1: One of the possible maximum matchings of a network (dark grey links) defines a set matched (p_1 , p_2 , and a_3) and unmatched nodes (a_1 and a_2). The minimum number of species necessary to control this network equates to the number of unmatched nodes $n_d = 2$. Superior nodes (dark grey circles) drive the dynamics of other species and are therefore important for the controllability of the network because they either require external input or because they are part of the chain that transmit the external inputs to other species.

To find the maximum matching in a pollination network we used an alternative bipartite representation of the network (Figure 1b) in which the levels indicate the outgoing and incoming links to each node. Finding a matching in this alternative representation is equivalent to finding the set of links such that each node on the top level is matched to at most one node on the bottom level and vice versa (West 2001; Csardi and Nepusz 2006).

The link between network controllability and the maximum matching also has an intuitive basis. In control theory, a network is fully controllable when each node in the network has a "superior", or input that drives its dynamics. Matched nodes can be controlled within the network by the node pointing to it—its superior. However, unmatched nodes have no superior in the network and must be directly controlled by external inputs (Liu, Slotine, and Barabási 2011), what we refer to as management interventions.

Manageability

Systems that require a large number of inputs are more difficult to control. The *manageability* of a community is therefore determined by the minimum number of species on which we would need to apply management interventions if we were to gain full control of the abundances of all species in the ecological community (n_d , Equation 1).

$$m = 1 - \frac{n_d}{s} \quad (1)$$

We quantified the manageability of ten paired plant-pollinator communities. Each pair was composed by a community invaded by a plant and a community effectively free of the invasive species (Table S1). The networks were constructed from previously published visitation data collected from pollination communities in Bristol, UK (Lopezaraiza-Mikel et al. 2007) and Cap de Creus National Park, Spain (Bartomeus, Vilà, and Santamaría 2008). In the four British uninvaded communities the non-invaded plots were obtained by experimentally removing all the flowers of invasive species *Impatiens grandulifera*. In contrast, the Spanish uninvaded communities were obtained from plots that had not yet been colonised by the invasive species *Carpobrotus affine acinaciformis* or *Opuntia stricta*.

The analysed networks were quantified using visitation frequency, which has been shown to be an appropriate surrogate for inter-specific effects in pollination networks (Vázquez, Morris, and Jordano 2005; Bascompte, Jordano, and Oelsen 2006). Nevertheless we evaluated the effect of using pollinator efficiency or importance as measures of species interactions (Ne'Eman et al. 2010; Ballantyne, Baldock, and Willmer 2015) and found consistent results (see Supplementary Information).

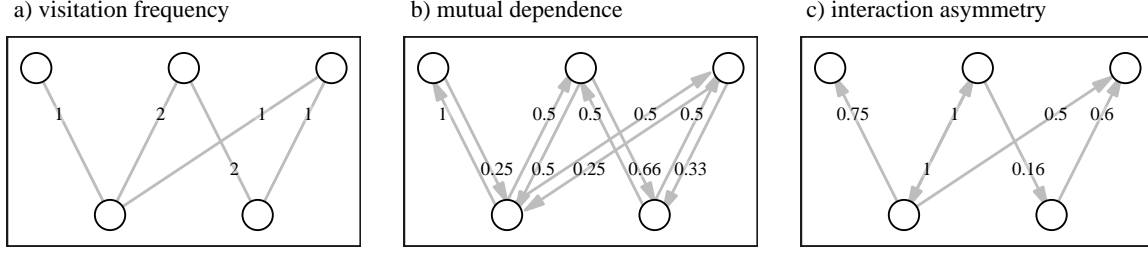


Figure 2: Mutualistic networks are usually described quantitatively by the relative number of visits between each plant and animal species (a). The mutual dependencies between interacting species (b) depend directly on the relative visitation frequency. The dependency of plant i on pollinator j , d_{ij} , is the proportion of the visits coming from pollinator j compared to all pollinator visits to plant i . Likewise, the dependency of pollinator j on plant i , d_{ji} , is the ratio of the visits by pollinator j to plant i and all visits of pollinator j . Nevertheless, mutual dependencies in pollination networks tend to be highly assymetric. The relative differences of dependencies (c) then is a good indicator

Interspecific effects

Our maximum matching algorithm requires a directed network in which a link between species i and j pointing to species j indicates that the abundance of species j can be affected by the abundance of species i . The extent to which a pair of interacting species affect each other—the mutualism strengtth—can be quantified by the magnitude of the mutual dependence (Bascompte, Jordano, and Oelsen 2006). Therefore, we initially assumed that a pair of interacting species can affect each other abundances. The effect on each other being inverselly proportional to the magnitude of the dependencies. This assumption renders a bipartite network in which interacting pairs are connected by two directed links (Figure 2b).

Nevertheless, mutual dependencies have been shown to be highly assymetric in natural communities (Bascompte, Jordano, and Oelsen 2006). In other words, if a plant species is largely dependent on a pollinator species, then the pollinator tends to depend only very weakly on the plant. This suggest that it may be possible to relax our initial assumption and simplify the network such that interacting species are connected by one directed link when mutual dependencies are assymetric. The direction of this link being consistent with the dominant direction of control in the bidirectional case (Figure 2c). The extent to which a species i affects species j relative to the extent to which species j affects species i can then be summarised by the interaction asymmetry (Equation 2).

$$a(i,j) = \frac{|d_{ij} - d_{ji}|}{\max(d_{ij}, d_{ji})} \quad (2)$$

We quantified the manageability of the analysed pollination communities by finding the maximum matching of the corresponding pollination network following the method previously explained. If a network had more than one component (two species are in different components if there exist no path between them) we analysed the largest one. We then performed a paired Spearman correlation test to compare the manageability obtained when including bidirectional effects between interacting species, and when simplifying the

network to include only the dominant direction of the asymmetry.

Naively one might expect to see a negative relationship between diversity and manageability (Menge 1995). We therefore used a set of generalised linear models to investigate the relationship between community manageability and (i) the total number of species, (ii) plant richness, (iii) pollinators richness, and (iv) the ratio of plants to pollinators richness. Because both invaded and British communities tended to be more diverse, we also assessed whether observed differences in manageability were due to these two factors. Competing models were assessed by comparing the Akaike Information Criterion corrected for small sample sizes (AICc).

To quantify the extent to which manageability is characteristic of the network structure, we compared the manageability of empirical networks to networks generated by two null models of visitation. One set of random visitation networks was generated by maintaining the species degree (the number of interactions per species) while varying the number of visits per species. A second set of random network maintained the species strength (total sum of visits) while varying the degree of species (Oksanen et al. 2016). We then calculated the mutual dependences and interaction asymmetries of the random networks and calculated the p-value of the manageability of each empirical network and the corresponding manageability of 999 randomisations.

failed to come up with a better wording

We also quantified the extent to which the relative manageability of a pollination community is given by the dominant direction of control, we maintained the structure of network but randomised the direction of the interaction asymmetries. This is, we first calculated the asymmetries for each community and then shuffled the direction of the link between species. Similar as in previous null models, we calculated the p-value of the empirical manageability and that of the randomisations.

Driver species

The manageability of a community m provides an indication of the magnitude of n_d but not the identity of the species that can compose this set. In fact, maximum matchings are often not unique in ecological networks and different sets of species of size n_d could be used to control the network (Liu, Slotine, and Barabási 2011). Nevertheless, species vary in terms of the frequency with which they are classified as superior nodes (Figure 1). This frequency provides an indication of their relative importance to drive the state of the community. We therefore call *driver species* those that are identified as needing directed interventions in all possible matchings, and consequently likely to have a disproportionate effect on the abundances of other species in the community.

elaborate concept

Unlike unweighted networks, where a maximum matching is defined by its size (the number of matched edges), in weighted networks, a maximum matching is one whose edges have the largest possible total weight among all possible matchings. This means that in weighted networks there is a number of matchings that

have maximal size, but are not maximum matchings because its total weight is not maximal too. However the weight of species dependencies is unexact, a reflection of the fact that interaction strengths and sampling methods themselves are stochastic to some extent.

We therefore identified the driver species in each of the networks by computing all possible matchings in the network that have the same size and have a weight above a cut-off of 0.5 times the weight of the maximum matching. To do so, we first generated the network's line graph—an alternative representation in which nodes are the interactions of the original network and links represent common ending points. We then calculated the complement of the line graph—a graph in which nodes are adjacent if and only if they are not adjacent in the line graph. All possible maximum matchings can be found by enumerating all cliques of size $1 - n_d$ in the complement graph (West 2001; Csardi and Nepusz 2006). Because of the number of matchings depends on the cut-off and the weighting of the links, we evaluated the effects of i) using a different cut-offs value between 0 and 1 and ii) weighting the links by the number of visits.

Similar as when we compared the manageability obtained when assuming bidirectional links or single links determined by the asymmetry (Figure 2), we compared the relative importance of species under these two assumptions by calculating a paired Spearman correlation test between species' relative importance. Nevertheless because we are more interested on the agreement between highly ranked species we also calculated a weighted correlation coefficient r_{w2} which gives more weight to higher ranks (??), and the Jaccard similarity of the driver species identified by both methods.

We then tackled the question whether some species level structural properties can predict the relative importance of driver species. Here, we used a Generalised Linear Mixed-Effects model (??) to evaluate the effect on the species' relative importance of measures of centrality (degree, betweenness, and overlap), measures related to network robustness (contribution to nestedness), and measures of strength of association and dependency (visitation and dependency strength) (Dormann, Gruber, and Fründ 2008). The community was included as a random effect. Candidate models were compared using AIC and the relative importance of the explanatory variables was evaluated using the sum of Akaike weights over candidate models that accounted for 95% of the evidence (Burnham and Anderson 2003; Bates et al. 2014; Barto 2016).

All analysis were performed in R 3.2.2 (R Core Team 2015).

Results

Manageability

We found that if we were to control the dynamics of the whole community, we would need to control between 42 and 88% (mean 63%) of the species in the community if we were to harness bidirectional links between species, and between 55 and 92% (mean 76%) if we only take into account the strongest dependence. This

justify

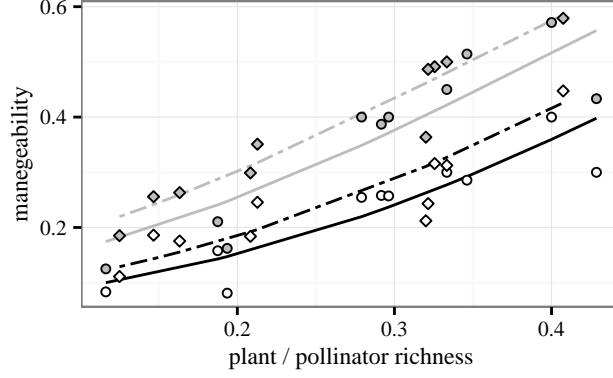


Figure 3: The relative manageability of a plant-pollinator community is largely influenced by the ratio of plant to pollinator species richness and to a smaller extent to the status of invasion. Invaded communities (circles, dashed lines) having larger levels of manageability. White and grey points correspond to manageabilities calculated using mutual dependences and interaction asymmetry respectively.

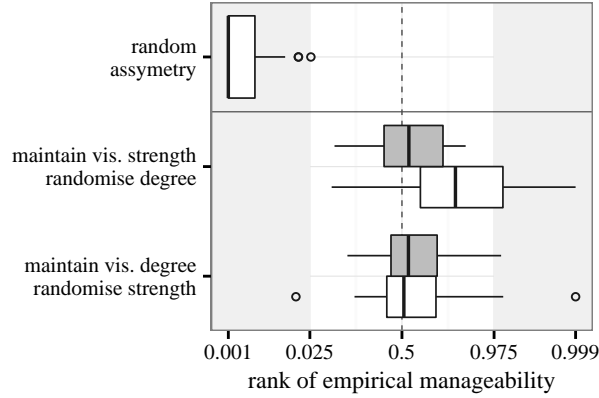


Figure 4: The manageability of empirical networks was compared to network randomisations using a normalised rank. The shaded areas in light grey (< 0.025 and > 0.975) suggest a significant difference between the empirical network and its randomisations. The empirical manageability is much smaller than those of randomisations in which the direction of asymmetries has been randomised. In contrast the manageability of network in which the visitation strength or degree was maintained was not different to the manageability of the empirical networks. Dark grey and white boxplots represent manageabilities obtained using mutual dependencies and asymmetries respectively.

corresponds to manageabilities between 0.12 to 0.58 and 0.08 to 0.45 respectively (Table S1). Despite different absolute values of manageability between these assumptions, the relative manageability of different communities was maintained (Spearman rank correlation 0.93, $p = 1.7 \times 10^{-9}$). In addition, the ratio of plant to pollinator species richness is positively correlated with manageability ($Pr(> \chi^2) = 8.5 \times 10^{-22}$). Similarly, invaded communities tend to be more difficult to manage than uninvaded ones ($Pr(> \chi^2) = 0.036$). In contrast, measures of biodiversity (like plant or pollinator richness) or the location of the study did not have a significant effect on the manageability (Figure 3, Table S3).

The manageability of empirical networks varied little when compared to network randomisations that maintained the species strength or species degree. This result is consistent regardless of whether we consider mutual dependences or the interaction asymmetry. Contrastingly, empirical networks have smaller manageabilities than network randomisations in which the direction of the dependencies is altered but the structure is maintained.

factor	est.	No. mod.	imp.	C.I.
(Intercept)	3.20	3	1.00	2.03
dependency strength	37.25	3	1.00	14.82
contribution to nestedness	1.42	3	1.00	1.01
visitation strength	1.38	3	1.00	1.14
eigen-centrality	-0.07	1	0.29	0.39
guild (pollinator)	0.14	1	0.22	0.94

Table 1: The strength of dependency, contribution to nestedness and visitation strength had a significant effect on all the models that accounted for 95% of the evidence based on AICc. Coefficient estimates were averaged following Buckland, Burnham, and Augustin (1997); confidence intervals were calculated following Lukacs, Burnham, and Anderson (2010).

Driver species

Species vary widely on the degree of their importance for network manageability. Our method to calculate the relative importance of species relies on calculating all possible maximum matchings. The number of matchings increases exponentially with the number of links and nodes in a network. When we use mutual dependencies, there are two links between each interacting species, and we were only able to find the importance of different species in the eleven smallest networks. For the remaining nine networks, the number of maximum matchings is much larger than 4×10^{10} and the computation of the species relative importance became unfeasible.

Nevertheless, we were able to find the relative importance of species in the single link version of the networks where links were weighted by the asymmetry. For the networks in which we were able to calculate the importance using mutual dependencies, we found a strong correlation between both approaches (Spearman's rank correlation ranged between $\rho = [0.6, 0.78]$, median $\tilde{\rho} = 0.69$). Indeed, when using the weighted rank correlation coefficient r_{w2} , we found that the agreement was stronger for highly ranked species than for lowly ranked species ($r_{w2} = 0.72$ versus $r_{w2} = 0.59$). Moreover both approaches identified a highly similar set of driver species (Jaccard similarity $J = [0.5, 1]$, $\tilde{J} = 0.67$).

Although plants showed significantly higher levels of importance for network control than pollinators (Mann-Whitney-Wilcoxon test, $p < 0.001$), this difference can be largely explained by the strength of dependency of the species (the sum of the dependencies for a given species). Indeed, when averaging over the set of GLMMs that accounted at least 95% of the evidence based on AICc, we found that the strength of dependency is the single most important factor determining the relative importance of species for network control (Table 1, Figure 5). Contrastingly, the number of visits a species receives (or performs) and its contribution to nestedness have only a marginal effect on its relative importance for network control. In addition, in the ten communities that were invaded, invasive species were invariably classified as a driver species.

consider moving the table to the supplementary

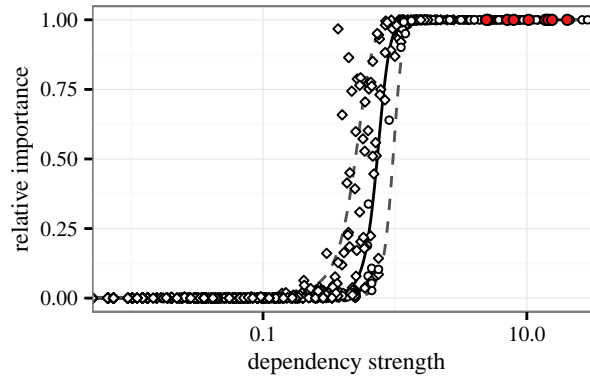


Figure 5: Dependency strength is the single most important factor explaining the importance of species for network control. Visitation levels and contribution to nestedness had a significant albeit small effect on the importance (dashed lines correspond to \pm one standard deviation on these factors). Invasive species (red) were invariably classified as driver species (those having the maximum importance for network control).

Acknowledgements

The authors thank Dr. Takeuki Uno for the insight provided to find the set of all maximum matching algorithms, and Bernat Bramon, Matthew Hutchinson and Marilia Gaiarsa for feedback in early stages of the project. EFC acknowledges the support from the University of Canterbury Doctoral Scholarship, the University of Canterbury Meadow Mushrooms Postgraduate Scholarship, a travel grant from the European Space Agency and a Rutherford Discovery Fellowship (to DBS). DBS acknowledges the support of a Rutherford Discovery Scholarship, administered by the Royal Society of New Zealand.

References

- Ballantyne, Gavin, Katherine C R Baldock, and Pat G Willmer (2015). “Constructing more informative plant-pollinator networks: visitation and pollen deposition networks in a heathland plant community”. In: *Proceedings of the Royal Society B* 282.1814, p. 20151130. DOI: [10.1098/rspb.2015.1130](https://doi.org/10.1098/rspb.2015.1130).
- Bartomeus, Ignasi, Montserrat Vilà, and Luís Santamaría (2008). “Contrasting effects of invasive plants in plant-pollinator networks”. In: *Oecologia* 155.4, pp. 761–770. DOI: [10.1007/s00442-007-0946-1](https://doi.org/10.1007/s00442-007-0946-1).
- Barto, Kamil (2016). *MuMIn: Multi-Model Inference*.
- Bascompte, Jordi, Pedro Jordano, and Jens Oelsen (2006). “Asymmetric Coevolutionary Networks Facilitate Biodiversity Maintenance”. In: *Science* 312.5772, pp. 431–433. DOI: [10.1126/science.1123412](https://doi.org/10.1126/science.1123412).
- Bates, Douglas et al. (2014). *lme4: Linear mixed-effects models using Eigen and S4*.
- Buckland, S T, K P Burnham, and N H Augustin (1997). “Model Selection: An Integral Part of Inference”. In: *Biometrics* 53.2, pp. 603–618. DOI: [10.2307/2533961](https://doi.org/10.2307/2533961).
- Burnham, Kenneth P. and David R. Anderson (2003). *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. Second. New York: Springer, p. 488.

187 Csardi, Gabor and Tamas Nepusz (2006). “The igraph software package for complex network research”. In: *Inter-*
188 *Journal Complex Sy*, p. 1695.

189 Dormann, Carsten F, Bernd Gruber, and Jochen Fründ (2008). “Introducing the bipartite Package: Analysing Eco-
190 logical Networks”. In: *R News* 8.October, pp. 8–11. DOI: [10.1159/000265935](https://doi.org/10.1159/000265935).

191 Liu, Yang-Yu, Jean-Jacques Slotine, and Albert-László Barabási (2011). “Controllability of complex networks.” In:
192 *Nature* 473.7346, pp. 167–173. DOI: [10.1038/nature10011](https://doi.org/10.1038/nature10011).

193 Lopezaraiza-Mikel, Martha E. et al. (2007). “The impact of an alien plant on a native plant-pollinator network: An
194 experimental approach”. In: *Ecology Letters* 10.7, pp. 539–550. DOI: [10.1111/j.1461-0248.2007.01055.x](https://doi.org/10.1111/j.1461-0248.2007.01055.x).

195 Lukacs, Paul M., Kenneth P. Burnham, and David R. Anderson (2010). “Model selection bias and Freedman’s
196 paradox”. In: *Annals of the Institute of Statistical Mathematics* 62.1, pp. 117–125. DOI: [10.1007/s10463-009-](https://doi.org/10.1007/s10463-009-0234-4)
197 [0234-4](https://doi.org/10.1007/s10463-009-0234-4).

198 Menge, Bruce A. (1995). “Indirect effects in marine rocky intertidal interaction webs: patterns and importance”. In:
199 *Ecological Monographs* 65.1, p. 21. DOI: [10.2307/2937158](https://doi.org/10.2307/2937158).

200 Ne’Eman, Gidi et al. (2010). “A framework for comparing pollinator performance: Effectiveness and efficiency”. In:
201 *Biological Reviews* 85.3, pp. 435–451. DOI: [10.1111/j.1469-185X.2009.00108.x](https://doi.org/10.1111/j.1469-185X.2009.00108.x).

202 Oksanen, Jari et al. (2016). *vegan: Community Ecology Package*.

203 R Core Team (2015). *R: A Language and Environment for Statistical Computing*. Vienna.

204 Vázquez, Diego P., William F. Morris, and Pedro Jordano (2005). “Interaction frequency as a surrogate for the
205 total effect of animal mutualists on plants”. In: *Ecology Letters* 8.10, pp. 1088–1094. DOI: [10.1111/j.1461-](https://doi.org/10.1111/j.1461-0248.2005.00810.x)
206 [0248.2005.00810.x](https://doi.org/10.1111/j.1461-0248.2005.00810.x).

207 West, DB (2001). *Introduction to graph theory*. Second. Upper Saddle River: Prentice Hall.