

Working title: Ecological network control

¹ **E. Fernando Cagua¹, Kate L. Wootton^{1,2}, Daniel B. Stouffer¹**

² ¹ Centre for Integrative Ecology, School of Biological Sciences, University of Canterbury, Private Bag
³ 4800, Christchurch 8041, New Zealand

⁴ ² Current address: Department of Ecology, Swedish University of Agricultural Sciences, Box 7044, SE-750
⁵ 07 Uppsala, Sweden

⁶ **Author for correspondence:** Daniel Stouffer (daniel.stouffer@canterbury.ac.nz) - +64 3 364 2729 -
⁷ Centre for Integrative Ecology, School of Biological Sciences, University of Canterbury, Private Bag 4800,
⁸ Christchurch 8140, New Zealand

Introduction

A major goal in ecological research is to understand the role that species play in the biotic environment. Within community ecology, a complex-systems has led to the development of a variety analytical and simulation tools with which to understand the role of the role of species embedded in a network of interactions (Bascompte & Stouffer, 2009; Stouffer, Sales-Pardo, Sirer, & Bascompte, 2012). A particularly relevant dimension of any species' role is its ability to alter the abundance of other species in the community—since this impacts ecosystem function and biodiversity and might have knock-down effects on other processes and services. This ability is often referred to as “keystoness”. A significant proportion of the network tools used to estimate species roles in this context relies on the assumption that there is a link between a species centrality—a relative ranking of its importance that stems from social network research—and its ability to alter other species' abundance. At first glance, this assumption makes sense. At the core of representing ecological communities as networks sits the idea that the abundance of a species also depends on the abundance of its neighbours (the species it interacts with) and, by extension, the abundance of its neighbours' neighbours, and so on. More central species, by definition, are better connected and hence should be more likely to affect the abundance of other species in the network.

Though conceptually intuitive, the relationship between centrality and keystoness is largely phenomenological. On one hand, there is more than a dozen indices of centrality used in ecology. Choosing a particular metric can be a difficult task, especially because to date there are no clear methods to determine the suitability of an index to a particular application. Furthermore, recent studies have shown that definitions of keystoness based on centrality might not adequately predict the cascading extinctions keystone species are usually associated to. As a result, community ecology could benefit from an alternative, more mechanistically-grounded, approach to understand how species affect each other abundance. Species abundances are also influenced by the dynamic relationships between nodes, and the mechanisms of self-regulation, however, community and population dynamics can be modelled in innumerable ways, and empirical support for one versus another is often still ambiguous. The alternative approach should acknowledge ecosystem dynamics, but without being overly dependent on the particular choices of how they are characterised.

Among the various possibilities, *control theory* appears to be a strong candidate (Isbell & Loreau, 2013). Control theory is a widely studied branch of engineering used to determine and supervise the behaviour of dynamical systems (Motter, 2015). It is inherently designed to deal with system feedbacks and its application has recently been expanded to complex networks (Liu & Barabási, 2016). In correspondence with ecological questions, advances in this field have established a clear link between the structure of the network and the way nodes affect each other. Unlike centrality indices, however, this link is not based on a priori assumptions between network metrics and keystoness, instead, it is based on well-established

advances in both dynamical and complex-systems theory .

Despite the apparent overlap, the concepts and approaches of control theory are not yet commonly applied in an ecological context. Possibly, this is in part because research on the control of complex networks is generally more concerned with the state of the system as a whole. This is in contrast with the ecological role concept, where the focus is on the nodes that compose the network. At its fundamental level, control theory first determines whether a system is controllable or not; that is, it asks if a system can be driven to a desired state within a finite amount of time. Although the controllability of a network is a whole-system property, it has been recently shown that asking for the controllability of a complex-system is equivalent to finding a particular set of nodes. A set with which is possible to control the state of the whole network. Notably, this set of nodes is not always unique for a given network. This implies that an examination of the distinct sets provides a means to connect nodes with its *general* ability to modify the system to which they belong.

Here, we apply methods from control theory to an ecological problem and show how it can be used to gain insight into the role of species in an ecological network. Specifically, we outline the approach using a set of ten pairs of uninvaded and invaded plant-pollinator communities. We use invaded communities because there is strong empirical evidence showing that invasive species play an important role shaping the abundances of other species, and particularly in these ten networks (Bartomeus, Vilà, & Santamaría, 2008; Lopezaraiza-Mikel, Hayes, Whalley, & Memmott, 2007). This choice thus offers us an opportunity to explicitly contrast theoretical observations with empirical evidence. First, grounded in the difficulties usually involved with invasive-species eradication and ecosystem restoration (Woodford et al., 2016), we ask whether there are differences between the controllability of invaded and uninvaded networks. Then we expand existing methods control theory to effectively link the *controllability* of a network with the role of particular nodes. We ask—from a control-theoretic perspective—whether there are key differences between species in the role they play at driving the population of other members of the community and identify the ecological factors related to these differences. Finally, we compare the proposed approach to current methods based on species centrality and show how these methods are indeed valuable but only paint a limited picture in regards to the “keystoness” of a species.

Methods

We used ten paired pollination communities to apply the control theoretic approach. Each community pair was composed of a community invaded by a plant and a community free of the invasive species. Four pairs correspond to natural or semi-natural vegetation communities in the city of Bristol, UK

(Lopezaraiza-Mikel et al., 2007). These communities are comprised of 19–87 species (mean 55), and non-invaded plots were obtained by experimentally removing all the flowers of the invasive species *Impatiens grandulifera*. The other six pairs were obtained from lower diversity Mediterranean shrublands in Cap de Creus National Park, Spain (Bartomeus et al., 2008). These communities are comprised of 30–57 species (mean 38); in contrast to the above, uninvaded communities were obtained from plots that had not yet been colonised by either of the invasive species *Carpobrotus affine acinaciformis* or *Opuntia stricta*. The structure of all these communities was defined by the pollinator visitation frequency, which has been shown to be an appropriate surrogate for interspecific effects in pollination networks (Bascompte, Jordano, & Olesen, 2006; Vázquez, Morris, & Jordano, 2005). Full details about the empirical networks, the implications of using visitation frequency to quantify the strength of the associations, and the sensitivity of the control theoretic approach to undersampling of the interactions can be found in the supplementary information section S1, XX, and XX.

The first step in applying methods of control theory is to construct a directed network that is able to provide an indication of the extent to which species affect each other’s abundance. In some ecological networks, establishing the directionality can be relatively straightforward, for example when links represent biomass transfer or energy flow (Isbell & Loreau, 2013). In pollination networks, however, this directionality is less obvious as both species can, in theory, benefit from the interaction. We overcome that obstacle by noting that the extent to which species i affects species j relative to the extent to which j affects i can be summarised by the interaction asymmetry (Bascompte et al., 2006). This asymmetry is given by

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

where the dependence of plant i on pollinator j , d_{ij} , is the proportion of the visits from pollinator j compared to all pollinator visits to plant i . Previous research has shown that mutualistic interactions are often highly asymmetric in natural communities; in other words, if a plant species is largely dependent on a pollinator species, then that pollinator tends to depend rather weakly on the plant (or vice versa). We, therefore, create a directed link between species i and j when $d_{ij} - d_{ji} \geq 0$ to establish the most likely direction of control between a species pair (Figure 1a). Sometimes, however, there is no asymmetry in the dependences between species pairs ($d_{ij} = d_{ji}$), and we cannot infer the direction of control. As such, we deem both species to be equally likely to affect each other and create two directed links between them, one from i to j and another from j to i . By basing the direction of the links on the asymmetry of their dependence, we are able to generate a network that is consistent with the dynamics of the community while satisfying the requirements of structural controllability.

After constructing the directed network, we can now calculate the controllability of the different networks and investigate whether there are differences between invaded and uninvaded communities.

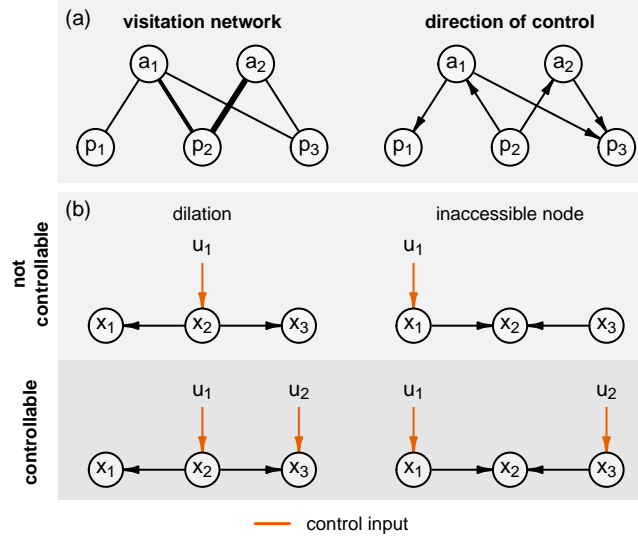


Figure 1: Direction of control and controllability conditions. (a) To establish the direction of control we start with a weighted visitation network (on the left). In this network, the width of the links corresponds to the frequency of visitation between animals a_i and plants p_i , with wider link indicating more visits. Plant p_1 is visited exclusively by a_1 but p_1 represents only a small fraction of the floral resources of a_1 . Therefore the population of p_1 is more likely to be affected by a_1 than viceversa. We represent this with a directed link from a_1 to p_1 (right). The direction of control between all other species pairs can be similarly determined by inspecting the difference between their relative dependences. (b) Once we have established the directions of control we can determine whether the network is controllable or not. A system defined by a directed network (with state nodes x_i ; species populations in an ecological context) and external control inputs (nodes u_i , orange links) is structurally controllable if it satisfies two conditions: it has no dilations (expansions in the network) and no inaccessible nodes. The system on the top left is not controllable because there is a dilation in which node x_2 is being used to control two nodes simultaneously, in other words there are less superiors (x_2) than subordinates (x_1 and x_3). The network in the top right is not controllable because node x_3 is inaccessible for the only input node u_1 in the system. Both systems can be made controllable by adding an extra input node (top).

Controllability

A system is said to be controllable if it is possible to steer it from an initial to an arbitrary final state within finite time (Kalman, 1963). A simple dynamic system can be described by $\frac{dx}{dt} = Ax + Bu(t)$, where the change of its state over time ($\frac{dx}{dt}$) depends on its current state x (for example, the species' abundances), an external time-varying input $u(t)$ (the control signal), and two matrices A and B , which encode information about the network structure and how species respond to external inputs, respectively. In classic control theory, determining whether this system is controllable can easily be achieved by checking that its controllability matrix $R = [B \ AB \ A^2B \ \dots \ A^{n-1}B]$ has full rank. In complex systems, however, employing this rank condition, or numerical approximations of it, is unfeasible because it is hard to fully parameterize A and B (either because the weight of the links changes over time or because they are difficult to measure). Here we use an approach based on the structural controllability theorem (Lin, 1974), which assumes that we are confident that the elements of A and B have either non-zero or zero values (there is an interaction or not), but that we are less sure about the precise magnitude of the non-zero values. Using this structural approach we can find out the controllability of a system under almost all non-zero parameter realisations. Structural controllability also has a graphical interpretation that allows us to translate the control problem into a topological one: a network is controllable if there are no inaccessible nodes—nodes without incoming links—or dilations—*expansions* of the network (Figure 1b; Supplementary Information section S2).

We are often able to estimate A in ecological networks, as it represents the interactions between species. Part of the control problem resides on estimating a supportable estimation of B , which represents the links between external inputs and species. Naively, any ecological community (and any system for that matter) could be controlled if we control the state of every species independently. Such approach is typically unpractical. We are instead interested in finding a minimum driver node-set (effectively finding B) with which to make the system controllable. However, a brute force search of this minimum driver node set is computationally prohibitive for most networks as it involves the evaluation of 2^N different controllability matrices. We employ a recently developed approach that shows that the control problem of finding the minimum driver node set can be mapped into a graph-theoretic problem: maximum matching (Liu & Barabási, 2016; Liu, Slotine, & Barabási, 2011).

Maximum-matching is a widely studied topic in graph theory and is commonly used in multiple applications, ranging from dating apps and wireless communications to organ transplant allocation and peer to peer file sharing. A matching in an unweighted directed graph is defined as a set of links that do not share a common start or end nodes; the largest possible matching is called maximum matching. For example, in a network composed by jobs and job applicants, a matching is any pairing between applicants and positions that satisfies one basic constraint: an applicant can be assigned to at most one position and vice-versa.

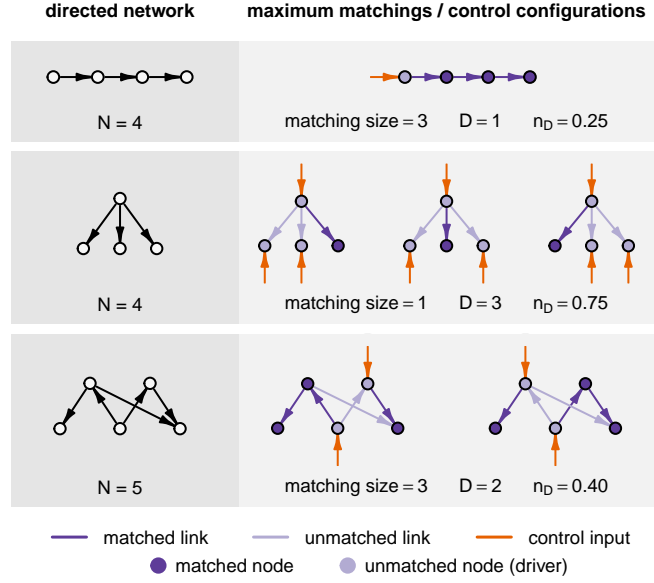


Figure 2: Maximum matchings and control configurations. In directed networks, a maximum matching is the largest possible set of links that do not share start or end nodes (dark purple). Maximum matchings are not necessarily unique, instead, each of them is related to a possible minimum driver node set in the network (the nodes to which an external control input, in orange, should be applied in order to ensure controllability). The size of the minimum driver node set D corresponds exactly to the number of unmatched nodes (the number of nodes in the network N minus the matching size). To account for network size, we use the size of the minimum driver node set relative to the total number of nodes $n_D = D/N$ as a measure of the extent to which the network structure can be harnessed to control the system.

Consequently, a maximum matching is an optimal pairing, one that maximises the number of applicants with jobs and the number of positions filled. At first glance, however, is hard to identify how exactly matchings are related to structural controllability. The key is to note that the conditions of structural controllability imply that, just like the one to one relationship between jobs and applicants, there needs to be a one to one relationship between *superior* and *subordinate* nodes (Figure 1b, bottom left). We use the maximum matching algorithm to find an optimal pairing of superior and subordinate nodes in a way consistent with the controllability conditions (supplementary information section S3.1). The result is that a matching can actually be decomposed into a set of paths that reveal how a control signal can flow across the links in a network to reach every node that composes it. And so, as recently shown (Liu et al., 2011), the minimum driver node set, those to which an external control input should be applied to make the system controllable, corresponds exactly to the *unmatched* nodes in the network (Figure 2).

Differences between invaded and uninvaded networks

Our first main objective is to investigate whether the controllability of a community is associated with invasion status or not. Finding out exactly, how difficult is to control a network depends strongly on the particularities of the desired control trajectory as well as the dynamical relationship between nodes. However, we are interested in understanding the controllability of a network in a more general sense, such

that it can be applied even when the precise control scenario is known only incompletely. We instead employ an indicator that follows from our approach: the size of the minimum driver node set. This simple metric provides a general indication of how difficult controlling a network might be, as systems that require a large number of external inputs to be fully controlled are intuitively more difficult or costly to manage. Specifically, following the structural controllability literature, we use the size of the minimum driver node set relative to the total number of nodes $n_D = \frac{D}{N}$ as a measure of the extent to which the network structure can be harnessed to control the community. In an ecological context, external inputs can be thought of as management interventions that modify the abundance of a particular species. For instance, achieving full control in a “network” in which species do not interact at all is relatively difficult as we would require an intervention for every single species. Conversely, the structure of a linear trophic chain can be harnessed to achieve full control using just one intervention targeted the top species; the control signal would then cascade through the trophic levels and reach other species in the community. After finding the minimum driver node set in each of our networks we want to test whether invasion status or other predictors are correlated to this metric of controllability. We, therefore, use a set of generalised linear models (with binomial error structure) to investigate the effect of invasion status while also including covariates related to species richness, since one might naively expect to see a negative relationship between richness and controllability (Menge, 1995). These covariates included the total number of species, plant richness, pollinators richness, the ratio of plant to pollinator richness, the link density (connectance), and the study site (as a two-level factor). In addition, we also explored whether real networks differ in their architecture from random ones in a concerted way (supplementary information section S4).

Species roles

While calculating the size of the minimum drive node set can measure the controllability of an ecological community, it does not provide information about the role that particular species play. Our second key question is related to how species differ in their ability to drive the population dynamics of the community. Ecologically, these differences are relevant because resources and data are limited, and therefore full control is unfeasible. To answer this question, we make two kinds of distinctions between species from a control perspective.

First, we harness the fact that maximum matchings in a network are often not unique, and each maximum matching indicates unique paths that can potentially be used to control the network. Specifically, we use the frequency with which a species is part of the multiple minimum driver node sets, its control capacity ϕ , as an estimation of its relative importance in driving the state of the community as a whole (Jia & Barabási, 2013). For example, a species with a control capacity $\phi_i = 1$ is a one that requires external

input in every single control configuration. This is, if we were to drive the community to an arbitrary alternate state, we would need to apply a management intervention to this species under any optimal control strategy.

Intuitively, to calculate a species control capacity ϕ we need to enumerate all possible maximum matchings (we detail the algorithm to find all maximum matchings in the Supplementary Information section S3.2). However, enumerating all maximum matchings is extremely expensive from a computational perspective—a network with a couple dozen nodes could be controlled with several hundred million unique matchings. To solve this problem we instead employ a recently developed algorithm that reveals the control correlations between the nodes in the graph (Zhang, Lv, & Pu, 2016, Supplementary Information section S3.3). Using this algorithm we are able to obtain every minimum driver node set with information obtained from a single maximum matching. The end result is that we are now able to calculate the control capacity ϕ_i of every node in linear time—rather than polynomial as the algorithm for matching enumeration requires.

A species control capacity ϕ allows us to identify species that are critical to change or maintain the state of the community. However, in ecological settings, changing the state of the community as a whole is often not required or undesired. Consequently, we make a second distinction between nodes based on whether they form part of possible control paths in the network and therefore are likely to influence the abundance of another species. We identify those species by checking whether they are classified as a superior node, that is they are at the beginning of a matched link. While unmatched nodes correspond to those that should be influenced with an external control input, superior nodes correspond to those that influence other species internally. Conveniently, with the exception of some pathological cases, superior nodes are such across every control configuration and therefore knowing one maximum matching provides enough information on whether a species is likely to play an important role at influencing the abundance of other species or not.

To examine whether any species-level structural properties could predict a species control capacity ϕ and the likelihood of being a superior node *sigma*, we calculated these properties for each of our networks. In the networks that contained reciprocal links (because there was not asymmetry in the dependences of a species pair), we averaged a species' ϕ and *sigma* across every possible “non-reciprocal” realisation of the network (more details can be found in the Supplementary Information section S3.4). *TODO: Still need to decide how to do this and how to compare traditional centrality metrics to the new control based ones. Will probably base the comparison on stability/feasibility metrics or/and on coextinction of species groups.*

References

- 215 Bartomeus, I., Vilà, M., & Santamaría, L. (2008). Contrasting effects of invasive plants in plantpollinator
216 networks. *Oecologia*, 155(4), 761–770. doi:10.1007/s00442-007-0946-1
- 217 Bascompte, J., & Stouffer, D. B. (2009). The assembly and disassembly of ecological networks. *Philosophical*
218 *Transactions of the Royal Society B: Biological Sciences*, 364(1524), 1781–1787. doi:10.1098/rstb.2008.0226
- 219 Bascompte, J., Jordano, P., & Olesen, J. M. (2006). Asymmetric Coevolutionary Networks Facilitate
220 Biodiversity Maintenance. *Science*, 312(5772), 431–433. doi:10.1126/science.1123412
- 221 Isbell, F., & Loreau, M. (2013). Human impacts on minimum subsets of species critical for maintaining
222 ecosystem structure. *Basic and Applied Ecology*, 14(8), 623–629. doi:10.1016/j.baae.2013.09.001
- 223 Jia, T., & Barabási, A.-L. (2013). Control Capacity and A Random Sampling Method in Exploring
224 Controllability of Complex Networks. *Scientific Reports*, 3(1). doi:10.1038/srep02354
- 225 Kalman, R. E. (1963). Mathematical Description of Linear Dynamical Systems. *Journal of the Society for*
226 *Industrial and Applied Mathematics Series A Control*, 1(2), 152–192. doi:10.1137/0301010
- 227 Lin, C. T. (1974). Structural Controllability. *IEEE Transactions on Automatic Control*, 19(3), 201–208.
228 doi:10.1109/TAC.1974.1100557
- 229 Liu, Y.-Y., & Barabási, A.-L. (2016). Control principles of complex systems. *Reviews of Modern Physics*,
230 88(3). doi:10.1103/RevModPhys.88.035006
- 231 Liu, Y.-Y., Slotine, J.-J., & Barabási, A.-L. (2011). Controllability of complex networks. *Nature*, 473(7346),
232 167–173. doi:10.1038/nature10011
- 233 Lopezaraiza-Mikel, M. E., Hayes, R. B., Whalley, M. R., & Memmott, J. (2007). The impact of an alien
234 plant on a native plantpollinator network: An experimental approach. *Ecology Letters*, 10(7), 539–550.
235 doi:10.1111/j.1461-0248.2007.01055.x
- 236 Menge, B. A. (1995). Indirect Effects in Marine Rocky Intertidal Interaction Webs: Patterns and
237 Importance. *Ecological Monographs*, 65(1), 21–74. doi:10.2307/2937158
- 238 Motter, A. E. (2015). Networkcontology. *Chaos*, 25, 097621. doi:10.1063/1.4931570
- 239 Stouffer, D. B., Sales-Pardo, M., Sirer, M. I., & Bascompte, J. (2012). Evolutionary Conservation of
240 Species' Roles in Food Webs. *Science*, 335(6075), 1489–1492. doi:10.1126/science.1216556
- 241 Vázquez, D. P., Morris, W. F., & Jordano, P. (2005). Interaction frequency as a surrogate for the total
242 effect of animal mutualists on plants: Total effect of animal mutualists on plants. *Ecology Letters*, 8(10),

243 1088–1094. doi:10.1111/j.1461-0248.2005.00810.x

244 Woodford, D. J., Richardson, D. M., MacIsaac, H. J., Mandrak, N. E., van Wilgen, B. W., Wilson, J. R.
245 U., & Weyl, O. L. F. (2016). Confronting the wicked problem of managing biological invasions. *NeoBiota*,
246 *31*, 63–86. doi:10.3897/neobiota.31.10038

247 Zhang, X., Lv, T., & Pu, Y. (2016). Input graph: The hidden geometry in controlling complex networks.
248 *Scientific Reports*, *6*(1). doi:10.1038/srep38209