Ecological network control

Supplementary Information

Fernando Cagua, Kate L. Wootton, Daniel B. Stouffer

S1 Characteristics of the empirical networks

The networks studied had species richness ranging between 19 and 87 when considering only the largest component in each network). As shown by the network asymmetry *AS* (N. Blüthgen, Menzel, Hovestadt, Fiala, & Blüthgen, 2007), the networks had a low ratio of plants to pollinators overall. Furthermore, the networks had relatively low levels of nestedness (when measured using the quantitative version of the NODF index; Almeida-Neto & Ulrich, 2011). Details for each network can be found in the Table S1.

S2 Structural controllability

Structural controllability avoids the limitation of not knowing the exact values of the matrices *A* and *B*. Using structural controllability can be boiled down to two conditions: a system is controllable if there are no inaccessible nodes or dilations. A node is inaccessible if there are no directed paths between it and the input nodes. "Dilations are subgraphs in which a small subset of nodes attempts to rule a larger subset of nodes. In other words, there are more 'subordinates' than 'superiors'" (Liu & Barabási, 2016).

The goal of structural controllability is to use the information contained in **A** to generate a supportable estimate of **B**. This focus allows us to gain insight of the inherent controllability of a network, and the roles of the species that compose it, without being overly dependent on the particular choices of how the system dynamics are modelled or characterised. The trade-off of this approach is that, because of the assumption of linearity, structural controllability alone does not allow us to fully design the time-varying control signal *u(t)* that can drive the system from one particular equilibrium to another. Nevertheless, the lessons gained when assuming linearity—at both the network and the species level—are a prerequisite for eventually understanding nonlinear control (Liu & Barabási, 2016; Liu, Slotine, & Barabási, 2011).

Table S1: Properties of the analysed plant-pollinator communities. Here show the number of species (n_s), the number of plants (n_p), the number of pollinators (n_a), the network connectance (c), the network asymmetry (AS), and the network nestedness (NODF index). All properties correspond to the network's largest component. British networks were assembled by Lopezaraiza-Mikel et al. (2007), Spanish were networks assembled by Bartomeus et al. (2008).

site	invader	n_s	n_p	n_a	c	AS	NODF	location
1	—	35	9	26	0.17	-0.49	8.68	Cap de Creus, Spain
1	*Carpobrotus affine*	57	10	47	0.17	-0.65	13.27	Cap de Creus, Spain
2	—	40	10	30	0.16	-0.50	11.66	Cap de Creus, Spain
2	*Carpobrotus affine*	38	11	27	0.21	-0.42	15.04	Cap de Creus, Spain
3	—	31	7	24	0.19	-0.55	12.91	Cap de Creus, Spain
3	*Opuntia stricta*	33	8	25	0.18	-0.52	9.96	Cap de Creus, Spain
4	—	35	10	25	0.17	-0.43	12.43	Cap de Creus, Spain
4	*Carpobrotus affine*	57	14	43	0.14	-0.51	13.70	Cap de Creus, Spain
5	—	35	8	27	0.19	-0.54	11.91	Cap de Creus, Spain
5	*Opuntia stricta*	32	8	24	0.19	-0.50	10.96	Cap de Creus, Spain
6	—	30	9	21	0.17	-0.40	6.91	Cap de Creus, Spain
6	*Opuntia stricta*	37	9	28	0.17	-0.51	12.45	Cap de Creus, Spain
7	—	37	6	31	0.19	-0.68	18.33	Bristol, United Kingdom
7	*Impatients grandulifera*	57	8	49	0.20	-0.72	14.36	Bristol, United Kingdom
8	—	48	5	43	0.21	-0.79	5.87	Bristol, United Kingdom
8	*Impatients grandulifera*	87	15	72	0.11	-0.66	8.12	Bristol, United Kingdom
9	—	55	12	43	0.13	-0.56	13.01	Bristol, United Kingdom
9	*Impatients grandulifera*	86	11	75	0.13	-0.74	13.40	Bristol, United Kingdom
10	—	19	3	16	0.38	-0.68	9.99	Bristol, United Kingdom
10	*Impatients grandulifera*	54	6	48	0.21	-0.78	7.56	Bristol, United Kingdom

directed network　**alternative bipartite representation**　**matching in original network**

—— matched link　　—— unmatched link　　● matched node　　● unmatched node (driver)
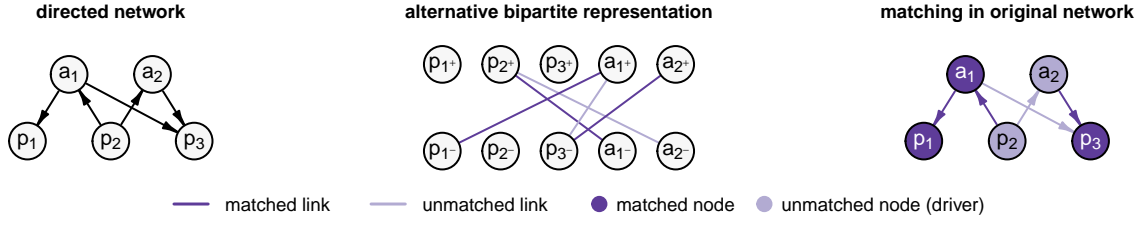
Figure S1: Finding a maximum matching in a complex network. (left) Directed network that indicate the direction of control between species. (center) Alternative bipartite representations of the directed networks. (right) The matchings in the bipartite representation mapped back to the original network.

S3   Maximum matching

S3.1 Finding a single maximum matching

Our approach to finding the minimum number of driver nodes relies on finding maximum matchings. We start with a directed network in which the direction of the link represents the direction of control (Figure S1 left panel). We then construct an alternative representation of the directed network in which each node of the directed network is represented by two nodes that indicate their outgoing and incoming links respectively (Figure S1 centre panel). Finding a maximum matching in this alternative representation is equivalent to finding the largest possible set of edges in which one node on the left-hand side is connected to at most one node on the right-hand side. To find the maximum matching we use the push-relabel algorithm implemented in `max_bipartite_matching` in the R package igraph 1.0.1 (Csardi & Nepusz, 2006). Once we have the matching (shown in the Figure S1 center panel) it is then easy to identify the roles of each node in this representation: nodes on the top-level that are connected to a matched link (dark purple) are superior while those connected to a matched link on the bottom-level are matched. This information can then be mapped back to the original representation to identify the control paths and the driver nodes in the network (Figure S1 right panel).

To further illustrate our methodology here, we also show the approach for the smallest of our empirical networks, the uninvaded network at site 10 (Table S1; Figure S2). This network is composed of 19 species of which three are non-invasive plants and the other 16 are pollinators. The one-to-one relationship between matched and superior nodes implies that in order to achieve full network controllability, most pollinators would be unmatched, and hence are classified as driver nodes that require external intervention. At the same time, both plants in the community, *Heracleum sphondylum* and *Rubus fructicosus*, and one of the pollinators, *Orthotylus/Lygocorus*, tend to be classified as superior nodes.
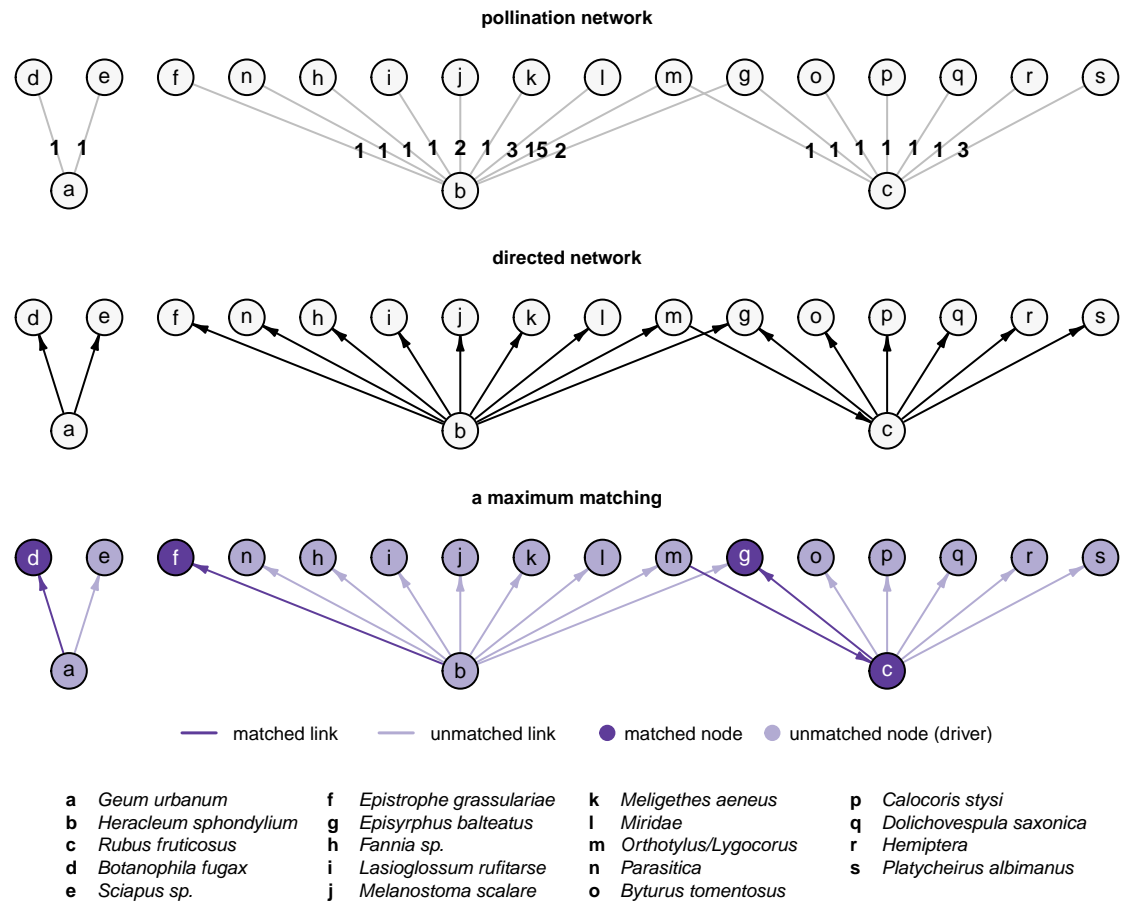
pollination network

directed network

a maximum matching

matched link — unmatched link — ● matched node ● unmatched node (driver)

a	*Geum urbanum*	**f**	*Epistrophe grassulariae*	**k**	*Meligethes aeneus*	**p**	*Calocoris stysi*
b	*Heracleum sphondylium*	**g**	*Episyrphus balteatus*	**l**	*Miridae*	**q**	*Dolichovespula saxonica*
c	*Rubus fruticosus*	**h**	*Fannia sp.*	**m**	*Orthotylus/Lygocorus*	**r**	*Hemiptera*
d	*Botanophila fugax*	**i**	*Lasioglossum rufitarse*	**n**	*Parasitica*	**s**	*Platycheirus albimanus*
e	*Sciapus sp.*	**j**	*Melanostoma scalare*	**o**	*Byturus tomentosus*		

Figure S2: Illustration of the procedure with an empirical community. The visitation network (top), the number of visits between species pairs is shown on each link. The directed network in which the direction of control is determined based on the mutual dependences (middle). One of the possible maximum matchings of this network (bottom).
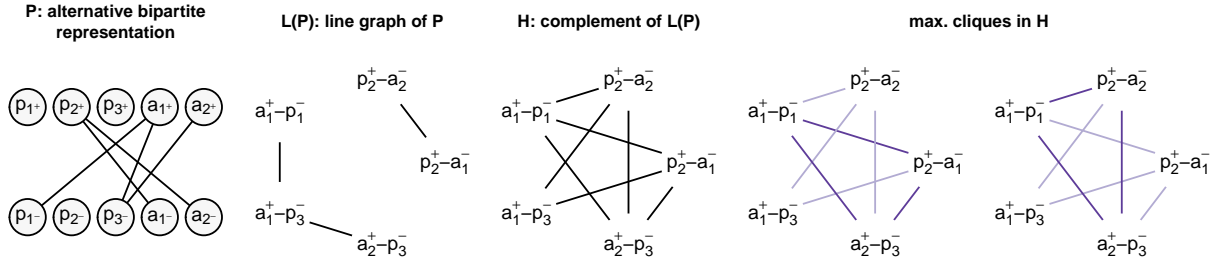
Figure S3: Finding all possible maximum matchings. From left to right: alternative bipartite representation of the directed network in Figure S1a. Line graph of the network. The complement of the network. The two maximal cliques are shown in dark purple.

S3.2 Finding all possible maximum matchings

The algorithm implemented in `max_bipartite_matching`, however, is only able to find **one** of possibly many maximum matchings in a network. Though one maximum matching is enough to calculate n_D and hence to provide an indication of the manageability of a community, it is not sufficient to estimate the role of individual species. To do that, we need to calculate all possible maximum matchings (or, equivalently, all maximal cardinality matchings in weighted networks like ours). To do this, we again start from the alternative bipartite representation in Figure S1b and assign an identity to each of the links in the network. We will call this bipartite representation *P*. We then construct the line graph of the alternative bipartite representation *L(P)* (Figure S3). Each node in *L(P)* represents a link in *P* and these are connected to each other if and only if they share a common node in *P*. We then calculate *H*, the complement graph of *L(P)* and identify all of its maximal cliques (Figure S3). Here some extra definitions are necessary. First, *H* is a graph with the same nodes as *L(P)* but that has a link between two nodes if and only if there is not a link in *L(P)*. Second, a clique is a subset of nodes such that all pairs of them are linked. Lastly, a maximal clique is a clique such that there are no cliques composed of more nodes (Gutin, 2013). In this example, there are two maximal cliques: the one formed by 1, 3 and 5, and the one formed by 2, 3 and 5. The final step is then to map these cliques onto the original network to obtain all possible maximal cardinality matchings as shown in Figure 2 in the main text.

S3.3 Control adjacency and the input graph

The algorithm proposed by Zhang, Lv, & Pu (2016) enable us to find all minimum driver node sets without calculating all possible maximum matchings. The algorithm is based on the construction of the input graph of the directed network. Components in the input graph reveal the correlations between nodes from a structural control perspective (Figure S4). The input graph is constructed based on the *control adjacency* of nodes in a maximum matching. Two nodes x_i and x_j are said to be control adjacent if there is a node x_k connecting x_i and x_j with an unmatched link $x_k - x_i$ and a matched link $x_k - x_j$. For example a_1 is control adjacent to a_2 in the top panel of Figure S4. If a node x_i is part of a minimum
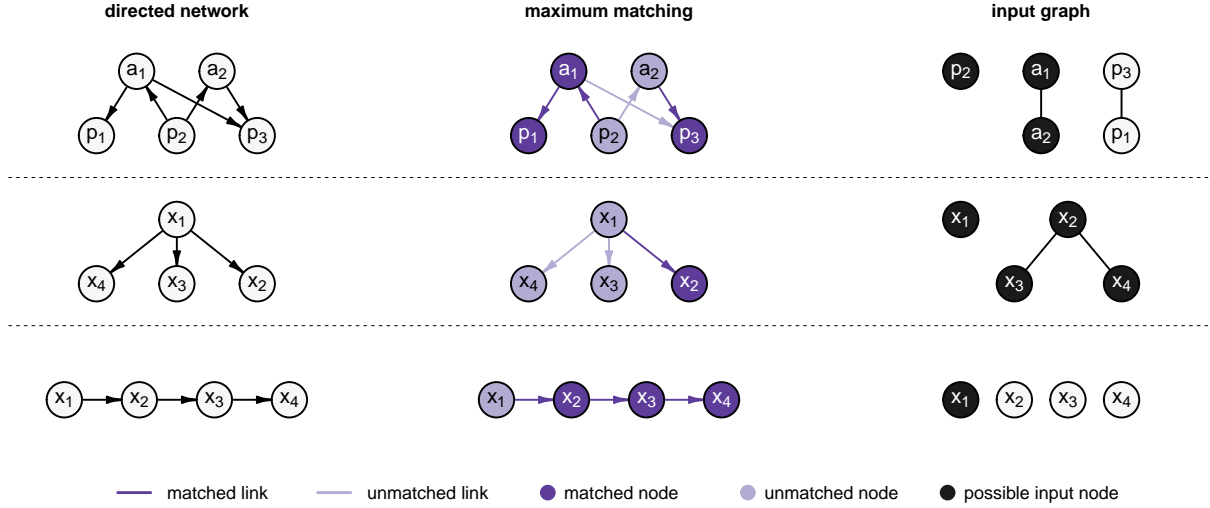
Figure S4: Input graph. Two nodes x_i and x_j are said to be control adjacent if there is a node x_k connecting x_i and x_j with an unmatched link $x_k - x_i$ and a matched link $x_k - x_j$. For example, in the top panel, a_1 is control adjacent to a_2 (via p_2), p_1 is control adjacent to p_3 (via a_1) and p_2 has no control adjacent nodes. As a_1 and p_2 belong to the minimum driver node set under the control configuration found with the maximum matching shown, then every node in the same component as these two nodes is also a possible driver node. By combining the nodes from the components with minimum driver node sets is possible to enumerate the different control configurations. For example the network in the top panel has two possible minimum driver node sets $[p_2, a_1]$ and $[p_2, a_2]$ (Figure 2). The network in the middle panel has three minimum driver node sets $[x_1, x_2]$, $[x_1, x_3]$, and $[x_1, x_4]$. Finally the network in the bottom pannel has only one $[x_1]$.

driver node set then every node in the same component of x_i in the input graph is a possible driver node. Every node within the same component in the input graph has the same control capacity ϕ and the control capacity ϕ_i of node x_i is inversely proportional to the number of nodes in the same component of x_i.

S3.4 Networks with reciprocal links

Our procedures to find minimum driver node sets is only adequate when there are no reciprocal (bidirectional) links in the network. The proportion of reciprocal links in relatively low. In fact the proportion of species pairs that are reciprocally connected in our empirical networks is just 2.97%. However, reciprocal links are not a rare occurrence as 16 out of 20 empirical networks had at least one. To calculate our controllability metrics at both the network (the relative size of the minimum driver node set, n_D) and species level (control capacity, ϕ, and the likelihood of being a superior node, σ) we first need to generate all possible versions of the of the network that include only non-reciprocal links.

Based on the available information, all of these network is as likely to be one that best represent the superior/matched relationships in the network. As such, we average the three metrics (n_D, ϕ, and σ) across all the networks without reciprocal links without weighting by the number of possible control configurations.

directed network with reciprocal links

D = 1.75
n_D = 0.44
No. control config. = 5

non–reciprocal network realisations

D = 2
n_D = 0.5
No. control config. = 1

D = 2
n_D = 0.5
No. control config. = 1

D = 2
n_D = 0.5
No. control config. = 2

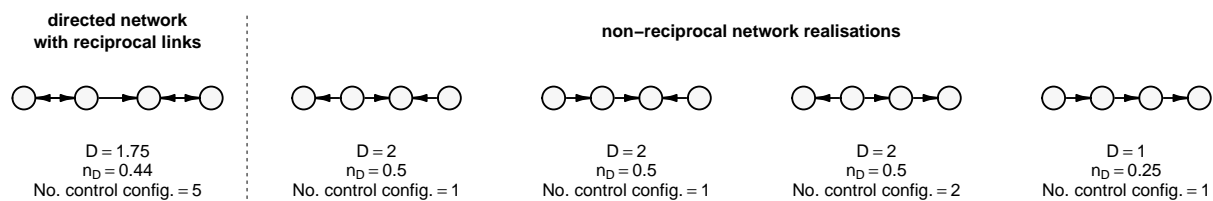D = 1
n_D = 0.25
No. control config. = 1

Figure S5: In order to calculate the controllability metrics for a network with reciprocal links (left), we first decompose the network into several versions that contain no reciprocal links (right). As each of these networks represent a different equiprobable scenario, the controllability metrics for the network with reciprocal links are the unweighted mean of these metrics across the non-reciprocal networks.

S4 Empirical vs. random expectation of controllability

References

Almeida-Neto, M., & Ulrich, W. (2011). A straightforward computational approach for measuring nestedness using quantitative matrices. *Environmental Modelling & Software*, *26*(2), 173–178. doi:10.1016/j.envsoft.2010.08.003

Blüthgen, N., Menzel, F., Hovestadt, T., Fiala, B., & Blüthgen, N. (2007). Specialization, Constraints, and Conflicting Interests in Mutualistic Networks. *Current Biology*, *17*(4), 341–346. doi:10.1016/j.cub.2006.12.039

Csardi, G., & Nepusz, T. (2006). The igraph software package for complex network research. *InterJournal, Complex Systems*, *1695*(5), 1–9.

Gutin, G. (2013). Independence and Cliques. In P. Zhang (Ed.), *Handbook of Graph Theory* (Second Edi, pp. 475–489). Chapman and Hall/CRC. doi:10.1201/b16132-31

Liu, Y.-Y., & Barabási, A.-L. (2016). Control principles of complex systems. *Reviews of Modern Physics*, *88*(3). doi:10.1103/RevModPhys.88.035006

Liu, Y.-Y., Slotine, J.-J., & Barabási, A.-L. (2011). Controllability of complex networks. *Nature*, *473*(7346), 167–173. doi:10.1038/nature10011

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