

SCIENTIFIC REPORTS



OPEN

Correlations in the degeneracy of structurally controllable topologies for networks

Colin Campbell¹, Steven Aucott¹, Justin Ruths², Derek Ruths³, Katriona Shea⁴ & Réka Albert⁵

Received: 05 December 2016

Accepted: 08 March 2017

Published: 12 April 2017

Many dynamic systems display complex emergent phenomena. By directly controlling a subset of system components (nodes) via external intervention it is possible to indirectly control every other component in the system. When the system is linear or can be approximated sufficiently well by a linear model, methods exist to identify the number and connectivity of a minimum set of external inputs (constituting a so-called minimal control topology, or MCT). In general, many MCTs exist for a given network; here we characterize a broad ensemble of empirical networks in terms of the fraction of nodes and edges that are *always*, *sometimes*, or *never* a part of an MCT. We study the relationships between the measures, and apply the methodology to the T-LGL leukemia signaling network as a case study. We show that the properties introduced in this report can be used to predict key components of biological networks, with potentially broad applications to network medicine.

A system of interacting components can be represented by a network, where system components are represented by *nodes* (or *vertices*) and their interactions are represented by directed *edges* (or *links*) between nodes. The dynamical behavior of the system is constrained by the topology of the network, but, because an edge generally indicates the existence of a regulatory relationship without specifying its nature, network topology is necessarily an incomplete representation of the system's dynamical behavior. For instance, the dynamic relationship represented by a node *A* being regulated by two nodes *B* and *C* may be well-approximated by a truth table (e.g., the future state of *A* is given by *B* AND *C*, where the node state variables are binary, 0 or 1) or by an arbitrarily complex function of *B* and *C* (e.g., $dA/dt = B^C$, where the node state variables are real numbers).

Whatever the dynamical relationships, a common goal is to efficiently influence the behavior of the constituent complex system. For example, manipulating biological systems can potentially lead to the development of effective therapies for disease^{1–5}, ecological management may preserve vital ecosystem services^{6–10}, and technological systems such as the power grid must be carefully designed and managed to provide functional and efficient services to consumers^{11–15}. The field of network science aims to identify the extent to which the network structure alone can be used to study the dynamics, and more recently the control, of systems. In part because such analysis aims to identify model-agnostic properties and in part because it can provide informative first-order effects, it is common to model the evolution of systems by linear dynamics of the form

$$\frac{d}{dt}\mathbf{x} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{u}, \quad (1)$$

where \mathbf{x} is a vector of the nodes in the network, \mathbf{A} is the *adjacency matrix* that encodes the interactions in the network topology (including, for instance, diagonal terms corresponding to the free decay of a node state variable), and the term $\mathbf{B}\mathbf{u}$ encodes external, user-controlled signals fed into a subset of so-called directly-controlled nodes, i.e., the dynamics of these specific nodes are directly affected by these external signals. Many systems, of course, are inherently nonlinear and are at best approximately linear within a local region; assuming linear dynamics of the form of equation (1) in these cases is an extremely coarse approximation^{16–18}, thought note that linearizing a nonlinear system is in some cases a sufficient condition for local controllability¹⁹. Nonetheless, the study of linear dynamics is important not only for systems that *are* linear, but also as a comparative framework for nonlinear

¹Department of Physics, Washington College, Chestertown, MD 21620, USA. ²Department of Mechanical Engineering, University of Texas at Dallas, Richardson, TX 75080, USA. ³Department of Computer Science, McGill University, Montreal, Quebec H3A 2A7, Canada. ⁴Department of Biology, Pennsylvania State University, University Park, PA 16802, USA. ⁵Department of Physics, Pennsylvania State University, University Park, PA 16802, USA. Correspondence and requests for materials should be addressed to C.C. (email: ccampbell7@washcoll.edu)

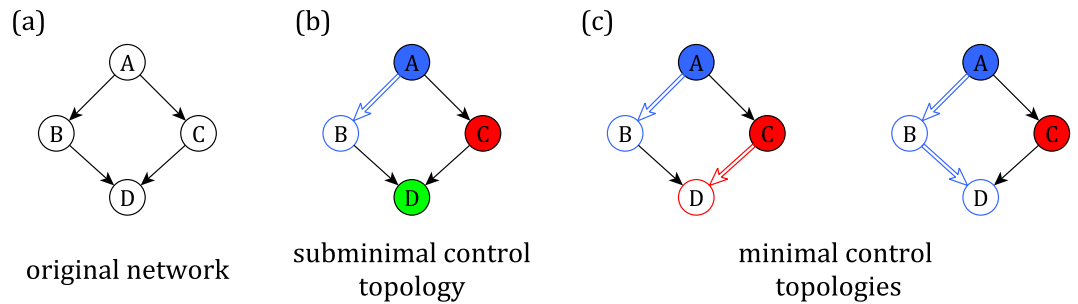


Figure 1. Control topologies in networks with linear, non-dissipative dynamics. (a) A simple directed network. We omit cycles from this example because in this framework they are inherently self-regulatory and their control follows immediately once the remainder of the network has been controlled²³. (b) In a control topology, every node is either directly controlled (colored nodes) or indirectly controlled (white nodes with colored outlines). Indirect control is achieved by placing nodes on a path originating at a directly controlled node (white edges with colored outlines). Importantly, in this framework every node can control at most one of its downstream neighbors and every pair of such paths are necessarily node-disjoint. (c) A control topology is minimal if it minimizes the number of controls. In this example node A must be directly controlled (it has no upstream nodes through which a control path may be routed) and either node B or node C must be directly controlled because node A can control at most one of its downstream neighbors.

dynamics: a system is nonlinear to the extent that it departs from linear expectations. The extent to which our results apply to nonlinear systems will depend on the form of the nonlinearity and the region of interest. While the networks we study behave according to a variety of models, we use their interaction networks to provide realistic structures for us to study the potential degeneracies of control properties.

In this case it is possible, with a sufficient number and placement of directly-controlled nodes and appropriate choices of time-varying control signals, to drive the system to any desired state in finite time^{20–23}. In principle one wishes to control the dynamics of a network with as few interventions as possible; minimizing the number of controls amounts to finding the *maximum matching* on the network (see Methods). Recent work has addressed the relationship between network topology and the properties of the directly-controlled nodes. For instance, the directly-controlled nodes are either source nodes or arise due to a *dilation*, where a node has more than one outgoing edge²³. By classifying dilations as *external* (if the outgoing edges point to sink nodes) or *internal* (otherwise), one can determine the fraction of controls in each of these categories (respectively denoted η_s , η_e , and η_i). These parameters constitute the *control profile* of a network. Empirical networks tend to be dominated by one of the three control profile parameters; diverse mechanisms exist by which synthetic networks may be generated with the same properties^{23–26}.

While the control profile offers insight into the types of directly-controlled nodes in terms of their topological location, it offers no direct insight into the *degeneracy* of control: there are generally many solutions to the maximum matching of a network. In other words, many combinations of nodes may be chosen for direct control. Furthermore, determining the maximum matching involves not only a set of directly-controlled nodes, but also a set of *matched edges*, which constitute *control signal paths* (see Fig. 1, Methods). It is therefore of interest to assess the importance of both nodes and edges in the set of all of the so-called *minimal control topologies* (MCTs). Specifically, here we wish to determine if a node is *always*, *sometimes*, or *never* directly controlled among all maximum matchings (i.e., across all MCTs). Similarly, we wish to determine if an edge is *always*, *sometimes*, or *never* on a control signal path.

Our contributions in this report are three-fold. First, we leverage the methodology of Jia *et al.*²¹ to characterize a new network-level statistic capturing the fraction of nodes that are always, sometimes, or never directly controlled (v_a , v_s , and v_n , respectively). We then develop a new method using a similar approach to assign a network-level statistic capturing the fraction of edges that always, sometimes, or never belong to the set of matched edges (fractions ε_a , ε_s , and ε_n , respectively). These new quantities echo the spirit of the control profile, which is a unique network-level statistic that captures the fraction of *controls* due to different functional structures. In contrast, these new quantities capture the fraction of *nodes and edges* that participate in the process of control signal dissemination through the network in varying ways. Thus the third contribution of this work is to provide insight into the connections between the functional classification of the control configuration (given by the control profile) and the degeneracy of the control configuration allowed by the network structure (captured by the statistics proposed in this report). Notably, we show that the existence of many internal dilations correlates with many edges *never* existing on a control path, suggesting that internal dilations tend to restrict the flexibility with which control signals propagate through a network. Finally, we apply the methodology of this report to the T-LGL leukemia signaling network and show broad agreement between the metrics introduced herein and existing experimental, computational, and analytical work that has identified nodes whose control play a pivotal role in the behavior of the network. While (as noted above) it is known that the study of structural controllability has some applications to systems characterized by nonlinear dynamics, the above-summarized results provide additional evidence suggesting that the methodology described in this report may be used in other biological networks to predict network components essential for control, with potentially broad applications to network medicine.

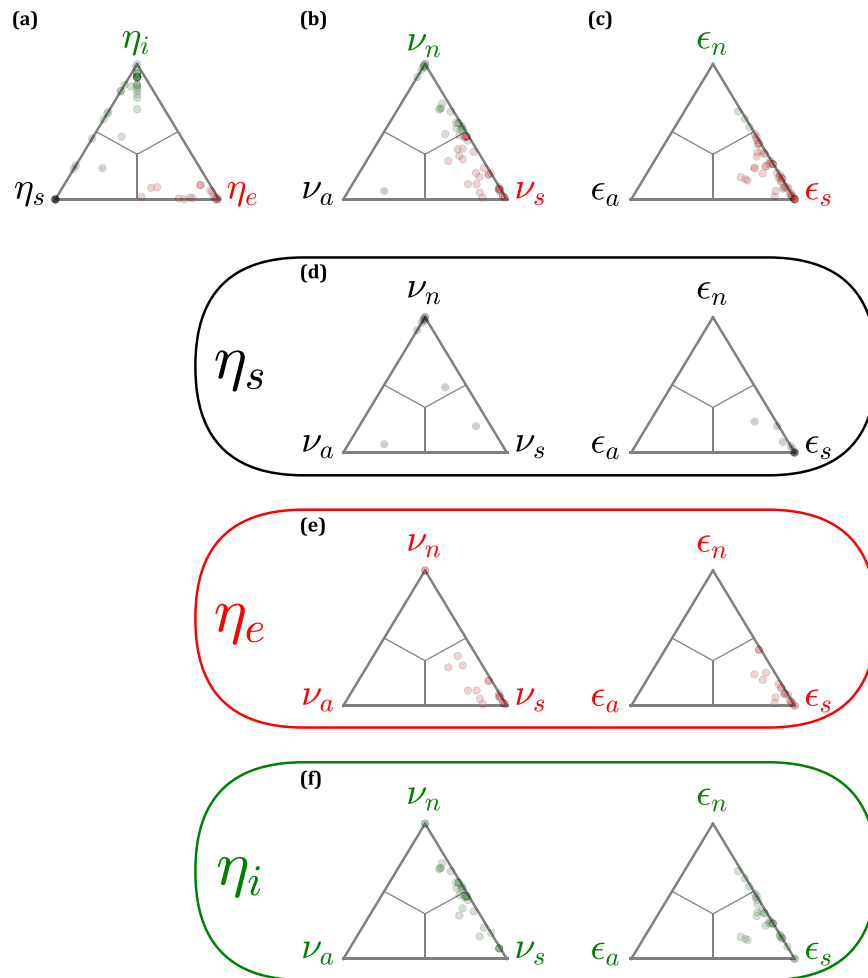


Figure 2. Distributions of empirical networks according to three sets of control measures. Each set includes three measures that sum to 1 for a given network. The distributions are shown on ternary plots, where a network at the center corresponds to a set of values (1/3, 1/3, 1/3) and a network at a corner corresponds to e.g. (1, 0, 0). Networks are represented with colored dots according to their maximal parameter. Interior lines indicate the regions where each parameter is largest. **(a)** The control profile of Ruths & Ruths²³. Directly-controlled nodes are either source nodes (η_s), arise due to internal dilations (η_i), or arise due to external dilations (η_e). **(b)** The fraction of nodes that are always (ν_a), sometimes (ν_s), or never (ν_n) directly controlled, when considering all control schemes that minimize the number of controls. **(c)** The fraction of edges that are always (ϵ_a), sometimes (ϵ_s), or never (ϵ_n) on a control signal path, when considering all control schemes that minimize the number of controls. **(d–f)** The degeneracy measures applied independently to the cases where η_s , η_e , and η_i are the dominant parameter in the control profile. Each plot is uniformly colored according to the corresponding dominant control profile parameter (as labeled on the left of the panel). The formatting of each plot otherwise follows (a–c).

Results

We analyze 58 empirical networks (see Table S1) and determine their distribution in parameter space for the control profile, node-based degeneracy, and edge-based degeneracy measures (Fig. 2a–c). As previously reported, the control profile of any one empirical network tends to be dominated by one of the control profile parameters²³. In contrast, the node-based degeneracy measures indicate that $\nu_a \ll 1$ for most networks, that is, few networks have a significant fraction of nodes that are always directly controlled (in agreement with the observation that most networks have relatively few directly-controlled nodes). This observation applies also to the edge-based degeneracy measures; $\epsilon_a \ll 1$. However, while networks are well-dispersed between ν_s and ν_n , most networks are skewed toward ϵ_s , meaning that while some networks have many nodes that are never directly controlled, few networks have many edges that are never on a control signal path.

We are also interested in the relationships between these measures. In Fig. 2(d–f) we show the distributions of degeneracy measures separately for networks where each of the control profile measures is largest (e.g., $\eta_s > \eta_e, \eta_i$). We find, for instance, that η_s -dominated networks tend to be ν_n - and ϵ_s -dominated. In other words, networks where many of the directly-controlled nodes are source nodes tend to have many nodes that are never directly controlled, and many edges that can be (but are not necessarily) on control signal paths. In contrast, η_e -dominated

| | η_s | η_i | η_e | ν_a | ν_s | ν_n | ε_a | ε_s | ε_n | N | E | $\langle k \rangle$ | $\langle C \rangle$ | τ |
|-----------------|----------|----------|----------|---------|---------|---------|-----------------|-----------------|-----------------|-------|-------|---------------------|---------------------|--------|
| η_s | 1.00 | -0.23 | -0.39 | 0.80 | -0.56 | 0.39 | 0.16 | -0.03 | -0.10 | -0.35 | -0.45 | -0.06 | 0.29 | 0.26 |
| η_i | | 1.00 | -0.24 | -0.14 | 0.06 | 0.05 | 0.06 | -0.65 | 0.74 | -0.34 | -0.31 | 0.14 | 0.19 | 0.18 |
| η_e | | | 1.00 | -0.04 | 0.68 | -0.66 | 0.30 | -0.03 | 0.06 | 0.43 | 0.27 | -0.44 | -0.64 | -0.61 |
| ν_a | | | | 1.00 | -0.23 | 0.01 | 0.42 | -0.19 | 0.07 | -0.22 | -0.44 | -0.38 | -0.01 | -0.06 |
| ν_s | | | | | 1.00 | -0.92 | 0.04 | 0.11 | -0.01 | 0.48 | 0.41 | -0.31 | -0.65 | -0.63 |
| ν_n | | | | | | 1.00 | -0.13 | -0.17 | 0.10 | -0.56 | -0.45 | 0.43 | 0.72 | 0.73 |
| ε_a | | | | | | | 1.00 | -0.49 | 0.29 | 0.14 | -0.26 | -0.79 | -0.52 | -0.58 |
| ε_s | | | | | | | | 1.00 | -0.94 | 0.21 | 0.37 | 0.11 | 0.01 | 0.03 |
| ε_n | | | | | | | | | 1.00 | -0.29 | -0.32 | 0.07 | 0.12 | 0.09 |

Table 1. Spearman correlation coefficients between control parameters and basic network measures. The table shows the total number of nodes and edges (N and E , respectively), the average degree $\langle k \rangle$, average clustering coefficient $\langle C \rangle$, and network transitivity τ . Table entries are colored according to their values (shades of blue for positive values and shades of red for negative); coefficients with a magnitude below 0.2 are written in light gray text. Black lines bracket intra-measure correlations (e.g. among η_s , η_e , and η_i).

networks tend to be ν_s - and ε_s -dominated, meaning that networks with an abundance of sink nodes tend to have significant flexibility both in choice of directly-controlled nodes and control signal paths. Finally, η_i -dominated networks tend to be ε_s -dominated (though less so than in the case of either η_s -dominance or η_e -dominance) and distributed between ν_s and ν_n . Thus, networks with many internal dilations may or may not have flexibility in terms of choice of directly-controlled nodes, and tend to have at least a moderate degree of flexibility in control signal paths.

We perform a pairwise Spearman correlation analysis between each of the 9 control-related parameters considered here and with basic network properties (Table 1). Unsurprisingly, there is a strong negative correlation between ε_a and each of the average node degree, average clustering coefficient, and network transitivity (-0.79 , -0.52 , and -0.58 , respectively): more connections per node and/or an increased frequency of closed triads afford greater flexibility in assigning control signal paths. Interestingly, these same properties are negatively correlated with ν_s and positively correlated with ν_n : a richer local structure constrains nodal participation (as directly controlled nodes) in MCTs. Furthermore, the number of nodes and edges exhibit weak negative correlation with ν_n and ε_n , indicating that larger networks are more likely to access more nodes and edges in at least some MCTs. Measures within a set tend to be negatively correlated with one another, with the notable exception of ε_a and ε_n (0.29). Some powerful trends exist between sets, as well, as suggested by Fig. 2: η_s is correlated with ν_a (0.8), η_i is correlated with ε_n (0.74), and η_e is correlated with ν_s (0.68).

To validate the utility of these measures, we consider the dynamic model of survival signaling network relevant to T-LGL leukemia^{2,27}. In this disease a fraction of white blood cells activated in response to a stimulus escape the process of activation induced cell death, survive, and after a while start attacking healthy cells. The dynamics of this network are defined by Boolean functions, from which a topological network can be extracted such that $A \rightarrow B$ if node A exists in the update function for node B . In the T-LGL network the node representing apoptosis (i.e., programmed cell death) is of particular interest. Its OFF state, together with the deregulation (abnormally high or low activity) of a subset of nodes, indicates the abnormal, leukemic state. Conversely, if in a leukemic cell the state of apoptosis cell changes from OFF to ON, the cell is committed to the process of cell death. Existing work has identified the minimal set of nodes whose sustained expression can lead to the leukemic state. This set consists of three source nodes: the initial stimulus, as well as the external molecules platelet-derived growth factor (PDGF) and interleukin (IL) 15, both of which were experimentally observed to be over-abundant in the blood of T-LGL leukemia patients²⁷. Prior work has also identified nodes whose direct control can lead to apoptosis of leukemic cells, despite the continued presence of these source nodes. Control of any one of 18 nodes (of 57 total) leads to apoptosis according to at least two of the following three types of evidence: experimental verification²⁷, simulation of Boolean dynamics², and analysis of the topology of the network once it has been expanded to topologically encode the Boolean rules²⁸.

The extent to which we expect the present metrics to agree with prior work is mitigated to some extent by the scope of the methodologies: in most prior work the quantity of interest is the state of a *single* node (apoptosis), whereas structural controllability seeks to achieve a desired state for *every* node in the network. Furthermore, the methodology used here assumes dynamics that obey equation (1), which is quite different from the Boolean framework used in the prior work being considered here. Therefore, a conservative expectation is that the two methodologies do not contradict one another. Specifically, assuming that the target state reflects induced apoptosis of a leukemic cell, we expect that the three source nodes necessary for the leukemic state are always directly controlled and the 18 apoptosis-inducing nodes should, at minimum, sometimes be directly controlled and/or be connected to an edge that is always on a control path.

We verify that this is the case: all three source nodes are always directly controlled, and 15 of the 18 key nodes have at least one incoming or outgoing “always” edge (indicating that they take part in a critical signaling pathway in terms of control) and/or are sometimes directly controlled (indicating that in some cases system

control may require direct control of these nodes). Furthermore, the remaining three key nodes are connected to at least 4 “sometimes” edges (indicating flexibility in the manner in which control signals are routed through these nodes). Indeed, despite the different methodological frameworks, the agreement is rather strong: all 18 nodes are connected to more “always” and/or more “sometimes” edges than expected by random chance, and only 2 are connected to more “never” edges than expected by random chance (for more details, see the Supplementary Information).

Discussion

Effectively influencing the behavior of complex interacting systems is a broad, multi-disciplinary goal. Accordingly, there is significant interest in discovering general techniques by which the dynamics of systems from different domains (e.g., technological and biological) may be guided by external intervention. We here consider systems that obey the linear dynamics of equation (1), where it has been shown that complete control is possible by feeding external control signals into a subset of the system components^{20–23}. These directly-controlled nodes are chosen such that every other node in the network is reached via non-overlapping paths originating at the directly-controlled nodes (see Fig. 1). The directly-controlled nodes and these control paths together constitute a *control topology*; a minimal *control topology* (MCT) is one in which the number of inputs is minimized.

In this report we consider the degeneracy of minimal control (i.e. the extent to which different minimal control topologies exist for a given network) in linear systems. Specifically, we characterize every system component (node) and interaction (edge) as being *always*, *sometimes*, or *never* on a MCT (the fraction of all nodes in these categories are respectively represented by the parameters ν_a , ν_s , and ν_n for nodes and ε_a , ε_s , and ε_n for edges). We study a broad selection of empirical networks and find that they are generally distributed between ν_s and ν_n while ν_a tends to be small. While we can unambiguously state that nodes are always directly controlled only if they are source nodes²³, in all but the simplest networks more precise statements require analysis of the maximum-matching problem and/or perturbing a MCT via a breadth-first search (see Methods). However, the flexibility of control in this framework is reflected by the typically high values of ε_s , suggesting that there are generally many ways for control signals to propagate through a network, even if there is relatively little flexibility in the choice of nodes to be directly controlled.

We consider these measures against the fraction of controls that are source nodes, sink nodes, and internal nodes (η_s , η_e and η_i , respectively), quantities which are fixed for a given network²³. The fact that η_s is positively correlated with ν_a follows from their definitions: η_s is the fraction of *directly-controlled nodes* that are source nodes, and ν_a is the fraction of *all nodes* that are source nodes (see Methods). The fact that η_i is positively correlated with ε_n indicates the existence of some rigidity in control signal paths cases where most of the directly-controlled nodes are neither sources nor sinks. In contrast, the correlation between η_e and ν_s suggests flexibility when most of the control nodes are sink nodes. In other words, there is flexibility in choosing *which* sink nodes are directly controlled and which are not, likely in part because there are multiple paths from source nodes to different sink nodes. It is also interesting to note that the correlations between node-based degeneracy measures and edge-based degeneracy measures tends to be weak (with the exception of ν_a and ε_a , the correlation magnitudes are uniformly below 0.4), indicating a relative disconnect between the node-based and edge-based degeneracy measures considered in this report.

While an interesting topic from a strictly theoretical standpoint, characterizing control degeneracy also has significant practical implications. In a biological system, for instance, a particular group of signaling molecules may be implicated in many theoretically viable control strategies. This, in turn, could incentivize the development of (e.g., pharmacological) techniques to influence the molecules in question. Indeed, we have shown broad agreement between the techniques developed here and existing work concerning the dynamics of the T-LGL leukemia signaling network, and the techniques described herein could be used to identify potential candidates for regulatory control in other biological networks. In an ecological system, the abundance of a particular group (or groups) of species, for example invasive or endangered species, may be controlled to initiate a cascade of changes in the abundances of other species^{6–8,29–32}. Species implicated in many viable control strategies under an appropriate modeling framework may, therefore, be prime candidates for direct manipulation to effectively manage ecological communities. Regardless of context, in cases where the nature of any nonlinearity is unknown, the methods developed here may provide insight into which components are essential for control.

We observe in this study that there exist meaningful correlations between the degeneracy of the control topology (directly controlled nodes and matched edges) and the functional divisions offered by the control profile. While aggregated statistics of network controllability have offered fruitful insights in the past, moving forward – to understand more precisely how network topology is related to network control – will require knowledge about all the possible control paths that can be used to control a network. Ultimately we aim to provide a clear mapping between the structure of the network and the ability we have to control such a system. Here we have provided a new dimension such that we can use the types of degeneracy exhibited by the control topology along with the dominance of certain types of control structures (given by the control profile) to triangulate more informed inferences on the network structures that are most important for network control.

A tempting avenue for future work is the development of procedures that identify the *fraction* of control paths that contain a given node or edge. This information would allow the categorical analysis considered in this report to be complemented by analysis on a continuum: nodes in V_a are in 100% of all control paths, nodes in V_n are in 0%, and nodes in V_s are somewhere in between (and similarly for edges). Studying the properties that drive nodes and edges to have comparatively high or low participation in control paths promises to enhance our understanding of the relationship between the structure and controllability of complex systems. In addition, we note that the diverse selection of empirical networks evaluated in this study offers insight into network structures that are independent of context. While this follows related work and avoids sample bias that arises when considering

traditional generative models^{23,26,33}, taking a similar approach as this study, but focused upon a particular empirical context (e.g., cellular signaling networks, ecological networks) may offer network-specific insight.

Methods

Control Topology. We define a control topology as a set of directly-controlled nodes, N_d , and the corresponding control signal paths that yield indirect control over every other node in the network^{20,23}. A control topology is minimal if it additionally minimizes $|N_d|$. Prior work has generally assessed the properties of a single minimal control topology (MCT) for a given network; a MCT is often obtained via the Hopcroft-Karp algorithm^{20,23,34,35}. In Fig. 1 we show several control topologies for a simple network.

Node-based assessment of MCT degeneracy. Because many MCTs generally exist for all but the simplest networks, we wish to characterize the nodes in a network according to the frequency with which they are directly controlled in a MCT. Specifically, a node is *always*, *sometimes*, or *never* directly controlled in a MCT; we denote the set of nodes in these categories as V_a , V_s , and V_n , respectively. Similarly, we denote the size of each set, normalized by the total number of nodes in the network, as ν_a , ν_s , and ν_n .

To categorize the nodes in this way, we adopt the method proposed by Jia *et al.*²¹. Suppose a single MCT has been determined, and consider first the set of directly-controlled nodes N_d . Clearly every node $n \in N_d$ is a member of either V_a or V_s . Making this distinction is trivial in light of the fact that the set of source nodes is identical to V_a ²¹. It follows immediately that the directly-controlled nodes in the MCT that are *not* source nodes are members of V_s .

It remains only to consider the nodes $n \notin N_d$. Clearly every such node is a member of either V_s or V_n . To determine the membership of one such node n_i , we force it to be directly controlled: if $|N_d|$ increases as a result, then it immediately follows that no MCT directly controls node n_i and therefore $n_i \in V_n$. Otherwise, $n_i \in V_s$. We repeat this procedure for all nodes $n \notin N_d$. It is possible to force a node to be directly controlled by perturbing the original MCT with an algorithmic complexity $O(EN)$ (see SI).

Edge-based control classification. Here we are interested in similarly classifying *edges* as *always*, *sometimes*, or *never* existing on the path of a control signal. We respectively define the sets of nodes in these categories as E_a , E_s , and E_n , and the normalized sizes of these sets as ε_a , ε_s , and ε_n . As in the case of the node-based analysis, we begin by applying the Hopcroft-Karp algorithm to the network in question to determine one MCT. From this MCT we obtain a set of edges on control paths, L_c (e.g., edge A->B in Fig. 1a).

Clearly edges $l \in L_c$ are members of E_a or E_s . Similarly, edges $l \notin L_c$ are members of E_s or E_n . In the first case, removing one such edge l_{ji} (denoting an edge from node j to node i) and re-evaluating the number of directly-controlled nodes via the Hopcroft-Karp algorithm serves to categorize the node: if $|N_d|$ increases, $l_{ji} \in E_a$; otherwise $l_{ji} \in E_s$. In the second case we may wish to force an edge $l_{ji} \notin L_c$ to be on a control path and similarly re-apply the Hopcroft-Karp algorithm; however, no simple modification to the network guarantees that the Hopcroft-Karp algorithm will force $l_{ji} \in L_c$.

We therefore develop alternatives for both of the above cases; the approach has a complexity of $O(E^2)$ (see SI).

References

- Campbell, C. & Albert, R. Stabilization of perturbed Boolean network attractors through compensatory interactions. *BMC Syst. Biol.* **8**, 53 (2014).
- Saadatpour, A. *et al.* Dynamical and Structural Analysis of a T Cell Survival Network Identifies Novel Candidate Therapeutic Targets for Large Granular Lymphocyte Leukemia. *PLoS Comput Biol.* **7**, e1002267 (2011).
- Kreeger, P. K. & Lauffenburger, D. A. Cancer systems biology: a network modeling perspective. *Carcinogenesis* **31**, 2–8 (2010).
- Steinway, S. N. *et al.* Network Modeling of TGF Signaling in Hepatocellular Carcinoma Epithelial-to-Mesenchymal Transition Reveals Joint Sonic Hedgehog and Wnt Pathway Activation. *Cancer Res.* **74**, 5963–5977 (2014).
- Huang, S., Ernberg, I. & Kauffman, S. Cancer attractors: A systems view of tumors from a gene network dynamics and developmental perspective. *Semin. Cell Dev. Biol.* **20**, 869–876 (2009).
- LaBar, T., Campbell, C., Yang, S., Albert, R. & Shea, K. Restoration of plant–pollinator interaction networks via species translocation. *Theor. Ecol.* **7**, 209–220 (2014).
- LaBar, T., Campbell, C., Yang, S., Albert, R. & Shea, K. Global versus local extinction in a network model of plant–pollinator communities. *Theor. Ecol.* **6**, 495–503 (2013).
- Levin, S. A. Towards a science of ecological management. *Conserv. Ecol.* **3**, 6 (1999).
- Attwill, P. M. The disturbance of forest ecosystems: the ecological basis for conservative management. *For. Ecol. Manag.* **63**, 247–300 (1994).
- Liebold, A. M. & Tobin, P. C. Population Ecology of Insect Invasions and Their Management*. *Annu. Rev. Entomol.* **53**, 387–408 (2008).
- Dobson, I. Complex networks: Synchrony and your morning coffee. *Nat. Phys.* **9**, 133–134 (2013).
- Hines, P., Blumsack, S., Cotilla Sanchez, E. & Barrows, C. The Topological and Electrical Structure of Power Grids. In *2010 43rd Hawaii International Conference on System Sciences (HICSS)*, doi: 10.1109/HICSS.2010.398 1–10 (2010).
- Motter, A. E., Myers, S. A., Anghel, M. & Nishikawa, T. Spontaneous synchrony in power-grid networks. *Nat. Phys.* **9**, 191–197 (2013).
- Pagani, G. A. & Aiello, M. The Power Grid as a complex network: A survey. *Physica A* **392**, 2688–2700 (2013).
- Kinney, R., Crucitti, P., Albert, R. & Latora, V. Modeling cascading failures in the North American power grid. *Eur. Phys. J. B* **46**, 101–107 (2005).
- Motter, A. E. Network control. *Chaos Interdiscip. J. Nonlinear Sci.* **25**, 97621 (2015).
- Cornelius, S. P., Kath, W. L. & Motter, A. E. Realistic control of network dynamics. *Nat. Commun.* **4** (2013).
- Gates, A. J. & Rocha, L. M. Control of complex networks requires both structure and dynamics. *Sci. Rep.* **6**, 24456 (2016).
- Liu, Y.-Y. & Barabási, A.-L. Control principles of complex systems. *Rev. Mod. Phys.* **88** (2016).
- Liu, Y.-Y., Slotine, J.-J. & Barabási, A.-L. Controllability of complex networks. *Nature* **473**, 167–173 (2011).
- Jia, T. *et al.* Emergence of bimodality in controlling complex networks. *Nat. Commun.* **4** (2013).
- Pósfai, M., Liu, Y.-Y., Slotine, J.-J. & Barabási, A.-L. Effect of correlations on network controllability. *Sci. Rep.* **3** (2013).
- Ruths, J. & Ruths, D. Control Profiles of Complex Networks. *Science* **343**, 1373–1376 (2014).
- Ruths, J. & Ruths, D. Response to Comment on ‘Control profiles of complex networks’. *Science* **346**, 561–561 (2014).

25. Campbell, C., Shea, K. & Albert, R. Comment on ‘Control profiles of complex networks’. *Science* **346**, 561–561 (2014).
26. Campbell, C., Ruths, J., Ruths, D., Shea, K. & Albert, R. Topological constraints on network control profiles. *Sci. Rep.* **5**, 18693 (2015).
27. Zhang, R. *et al.* Network model of survival signaling in large granular lymphocyte leukemia. *Proc. Natl. Acad. Sci.* **105**, 16308–16313 (2008).
28. Zañudo, J. G. T. & Albert, R. Cell Fate Reprogramming by Control of Intracellular Network Dynamics. *PLoS Comput. Biol.* **11**, e1004193 (2015).
29. Campbell, C., Yang, S., Albert, R. & Shea, K. Plant-pollinator community network response to species invasion depends on both invader and community characteristics. *Oikos*, doi: 10.1111/oik.02039 406–413 (2014).
30. Campbell, C., Yang, S., Shea, K. & Albert, R. Topology of plant-pollinator networks that are vulnerable to collapse from species extinction. *Phys. Rev. E* **86**, 21924 (2012).
31. Barrows, C. W. *et al.* A framework for monitoring multiple-species conservation plans. *J. Wildl. Manag.* **69**, 1333–1345 (2005).
32. Knops, J. M. H. *et al.* Effects of plant species richness on invasion dynamics, disease outbreaks, insect abundances and diversity. *Ecol. Lett.* **2**, 286–293 (1999).
33. Pequito, S., Preciado, V. M., Barabási, A.-L. & Pappas, G. J. Trade-offs between driving nodes and time-to-control in complex networks. *Sci. Rep.* **7**, 39978 (2017).
34. Hopcroft, J. E. & Karp, R. M. An $n^{3/2}$ Algorithm for Maximum Matchings in Bipartite Graphs. *SIAM J. Comput.* **2**, 225–231 (1973).
35. Commault, C., Dion, J.-M. & van der Woude, J. W. Characterization of generic properties of linear structured systems for efficient computations. *Kybernetika* **38**, 503–520 (2002).

Acknowledgements

The authors gratefully acknowledge Shaun Ramsey for useful discussions during the preparation of this report. This work was supported by NSF Grant DMS-1313115.

Author Contributions

C.C. and S.A. conceived of the study and performed analysis. All authors analyzed results. C.C. wrote the first draft of the manuscript; J.R., D.R., K.S. and R.A. contributed to revisions.

Additional Information

Supplementary information accompanies this paper at <http://www.nature.com/srep>

Competing Interests: The authors declare no competing financial interests.

How to cite this article: Campbell, C. *et al.* Correlations in the degeneracy of structurally controllable topologies for networks. *Sci. Rep.* **7**, 46251; doi: 10.1038/srep46251 (2017).

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

© The Author(s) 2017