Robustness of Network Controllability under Edge Removal

Justin Ruths and Derek Ruths

Abstract We introduce a quantitative measure of robustness of network controllability. Given a set of control nodes which drive the network, we investigate the effect of edge removal on the number of controllable nodes. We find that the mean degree of the network is a major factor in determining the robustness of random networks. Nonetheless, a comparison between real and random networks indicates a statistically significant difference which points to additional factors that influence the robustness of control of real complex networks.

1 Introduction

Mounting evidence confirms that network-oriented studies can uniquely characterize aspects of complex dynamics in wide ranging areas of biology, engineering, and social science: e.g., interaction of proteins in cellular pathways, diffusion of information within human communities, and optimal routing strategies on the internet [9, 5, 1]. However, as researchers seek techniques for engineering or influencing such complex systems, they are confronted by questions of controllability.

Real-world networks bring with them their own unique set of attributes and requirements not found in typical control systems. For example, in real complex networks (e.g., cellular biochemical pathways, social networks, and sprawling internet physical connectivity) external controls are not predefined and mechanisms for controls are not known, a priori. Furthermore, the structure of such systems is unlikely to be static over time and space: cascades of fundamental biochemical interactions differ among individuals, friendships may be made or lost, and internet connectivity changes as individual servers come online or fall offline. Thus, a schema for con-

Justin Ruths
Singapore University of Technology and Design, 20 Dover Drive, Singapore, e-mail: justinruths@sutd.edu.sg

Derek Ruths
McGill University, 3480 University Street, Montreal, Canada e-mail: derek.ruths@mcgill.ca

trolling such systems over long periods of time must be robust to these frequent and unpredictable structural changes. Because the addition and removal of edges and nodes in a network has no clear analog to modifications to a standard control system, this problem of robust network control presents a new challenge both to the network and control communities.

In this paper, we first define a quantitative measure of the robustness of a given control scheme for a given network. Second, we develop the necessary computational techniques to use our measure and network sampling to assess the statistical significance of robustness observed in a real-world network. This second result allows us to consider for the first time whether the controllability of real networks is more or less robust than might be expected at random.

2 Background

We consider a linear dynamic model of the directed network graph $G(A)$, which is composed of n nodes and L directed edges between nodes in the network. Systematically analyzing linear models is a key step in generalizing to broader classes of models, such as nonlinear dynamics. In addition, there are a large number of network phenomena that exhibit a good fit to a linear time-invariant model of the form [8, 13, 10], $\dot{x}(t) = Ax(t)$, where the state $x(t) \in \mathbb{R}^n$ is the value of all of the nodes at time t and $A \in \mathbb{R}^{n \times n}$ is the transpose of the adjacency matrix of the network, such that the value $A_{i,j}$ is the weight of a directed edge from node x_j to node x_i and zero if there is no such edge. In what follows, we study the controlled network $G(A,B)$, corresponding to adding m control nodes yielding the form, $\dot{x}(t) = Ax(t) + Bu(t)$ where the control $u(t) \in \mathbb{R}^m$ and $B \in \mathbb{R}^{n \times m}$ models the effect of the controls on the network. Determining the minimal B to make the graph $G(A,B)$ completely controllable has been a topic of particular interest [11, 3, 4]. A networked system is controllable if the controls are able to guide the system state, the value of the nodes, from an initial configuration $x(t_0) = x_0 \in \mathbb{R}^n$ to a final configuration $x(t_1) = x_1 \in \mathbb{R}^n$. The minimum number of control nodes which makes the $G(A,B)$ completely controllable (m_c) has been shown to correlate with the degree distribution, which is a measure of edges leaving nodes of the network [8].

Due to space constraints, we are unable to review the terms that arise in the following sections. These terms are well defined by a number of sources, however, they are likely new to many, even well-versed, network scientists [7, 11, 3, 4]. In particular, we direct you to review the concept of structural controllability and its relation to generic rank of the system $[A\ B]$; cacti structure of a network formed by paths and cycles of connected nodes; and maximum matching algorithm to find the largest set of nodes that can be uniquely paired amongst themselves using edges present in the network.

3 Methods

We discuss two novel measures of the robustness of a control scheme given by B to changes in the network structure through edge removal. The first is a measure

that assigns a value to a network/control scheme pair based on how reachability responds to the removal of edges from the network. The second is a computational approach to estimating the statistical significance of a network/control scheme pair's robustness. This answers the question: how much more robust is the network/control pair than might be expected at random?

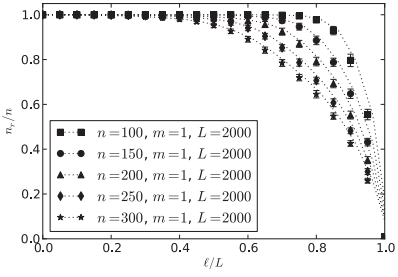

Fig. 1 The robustness profiles of Erdos-Renyi networks with the same number of original edges. The number of controllable nodes n_r of five ER networks are shown as a function of the number of edges, ℓ, removed. The displayed values are the averages of 50 separate percolation sequences, with corresponding error bars. The robustness measure, R, is given as the normalized area under this curve.

3.1 Measuring Robustness

Under fully structural controllable conditions, all nodes in the network are reachable by the specified control scheme. However, as edges are removed from the network (hereafter, *percolated*) some nodes will become unreachable. For a given sequence of edge percolations, the number of unreachable nodes will depend on the location of the controls in the network and the local and global structure of the network itself. Conceptually, given two different control schemes and a sequence of edge percolations, the more robust control scheme is the one that maintains a greater number of reachable nodes in the percolated network (in which the edges are missing).

In order to derive a quantitative measure of robustness, first consider that we can incrementally remove edges from a network with a specified control scheme and compute the number of nodes in the network that are still controllable. Such an iterative calculation will yield curves like those shown in Figure 1, which we call *robustness profiles*. In such a plot, the x-axis is the fraction of edges that have been percolated and the y-axis is the fraction of nodes that are reachable in the percolated network. The value of point ℓ/L corresponds to the fraction of nodes that are reachable after removing ℓ/L percent of the edges present in the original network. Curves will be monotonically decreasing since removing an edge cannot bring a new node under control.

Every combination of network (G), control scheme (B), and edge percolation sequence (ξ) has its own robustness profile, which we can summarize by integrating to find the area under the curve. We call this value the *integrated percolation (IP) robustness*,

$$R_{G,B,\xi} = \frac{1}{nL}\sum_{\ell=0}^{L} n_r(G-\xi[1:\ell],B) \tag{1}$$

where G is the network, n is the number of nodes in the network, L is the number of edges, B is the control scheme, $n_r(G',B')$ is the number of reachable nodes in the graph G' with controls B', and $\xi[1:\ell]$ is the first ℓ edges in the edge percolation sequence. $G-\xi[1:\ell]$ is graph G with the first ℓ edges in ξ removed.

Because the robustness profile always falls within a unit box (i.e., both axes of the plot are normalized to 1), the intuitive interpretation of $R_{G,B,\xi}$ is as the percent of the unit box that falls under the curve. This has the desired effect of returning larger values for more robust control schemes: if we fix the network and edge percolation sequence, a control scheme that retains more reachable nodes will have a curve with more of the area of the unit box under it.

Given the ability to compute the IP robustness for arbitrary edge percolation sequences, we can obtain an estimate of the robustness of a network-control scheme pair (hereafter, a *configuration*) by averaging $R_{G,B,\xi}$ over a large number of randomly generated edge percolation sequences, $\Xi = \{\xi_1, \xi_2, ...\}$,

$$R_{G,B} = \frac{1}{|\Xi|} \sum_{\xi \in \Xi} R_{G,B,\xi}. \tag{2}$$

We call this measure the *mean integrated percolation (MIP) robustness*. Analysis on a wide spectrum of networks have shown that, in practice (see Figures), the value of $R_{N,B,\xi}$ is quite consistent across different percolation sequences (i.e., the standard deviation for the statistic is diminishingly small). This consistency lends $R_{G,B}$ to being a reliable measure of robustness of a network-control scheme pair to arbitrary edge percolation processes. This will be further explored and validated in Section 4.

Hereafter, mentions of robustness refer to MIP robustness. Standard deviation will be shown in plots to indicate the variability in the estimate.

3.2 Assessing Real Network Robustness

In the preceding subsections, we were interested in simply being able to compute and compare the robustness of different network-control configurations. Such techniques are not sufficient, however, to quantify the extent to which real networks are designed for robust control. The notion of "designed robustness", whether through engineering or evolution, suggests that such networks would be expected to have more robustness than might be observed by chance. In this subsection we formalize this notion and provide a method for estimating the probability of observing a given network's MIP robustness by chance.

A standard practice in network science for establishing the statistical significance of a given network feature is to compare the feature in the network of interest to the same feature in a set of synthetic networks drawn from a null model (e.g., [6, 9]). In this instance, we will compare the robustness of the network of interest, G_0, to the robustness of a set of randomized networks, $\mathbf{Z} = \{G_1, G_2, ...\}$. In order to facilitate comparison, the random networks must match the number of nodes and edges in the original network ($n_i = n_0$ and $L_i = L_0$ for $i \geq 1$). Furthermore, we preserve the degree distribution of the original network in the random networks: this is done because, as we will see, mean degree is strongly associated with narrow ranges of robustness. Thus, the degree distribution of the network alone can explain much of its robustness. By holding degree distribution constant, our results estimate the extent to which more higher-order degree features and local motifs present within the real networks lend it to robust control.

To perform an actual comparison, a large number of random networks satisfying the above constraints are generated. The average MIP robustness is computed for the null model for a range of control scheme sizes: $R_{\mathbf{Z}}(m) = \frac{1}{|\mathbf{Z}|}\sum_{G\in\mathbf{Z}} R_{G,B_m}$. The MIP robustness for the original network is also computed for a range of control scheme sizes: $R_0(m) = R_{G_0,B_m}$. For a given value of m, $R_{\mathbf{Z}}(m)$ and $R_0(m)$ can be directly compared. Furthermore, since both measures are means, the overlap in their standard deviations can be used to assess the statistical significance of the original network's MIP robustness score.

4 Results & Discussion

In this section, we apply our definitions and methods to understanding two specific questions. First, we quantify the robustness of a class of synthetic networks and explore the connection between degree distribution and MIP robustness. Second, we determine the extent to which real networks are more robust than might be expected by chance, which is a first step towards understanding whether robust control is a feature designed into some complex systems.

4.1 Random Networks

Random network models have the capacity to generate large numbers of networks which share certain properties, but are random in every other way. For this reason, such network models have been fruitfully used to study the impact of specific network properties on phenomena of interest. Here we employ them for similar purposes: to understand the extent to which network properties influence MIP robustness. There are many different random network models; in the present paper we focus on Erdos-Renyi (ER) networks which are generated by cycling through all pairs of nodes and establishing an edge between these nodes with probability p [2]. The number of edges in an ER network is, on average, $L = n^2 p$.

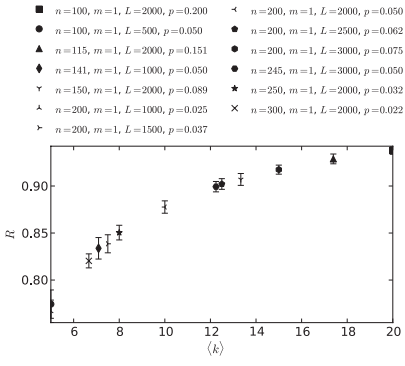

Fig. 2 The MIP robustness of Erdos-Renyi networks. We compute the robustness profiles and the corresponding MIP robustness measures, $R_{G,B}$, for a variety of ER networks. We plot the robustness measure versus the mean degree $\langle k\rangle = L/n$.

Figure 1 depicts the effect of edge removal on the number of controllable nodes. The generated ER networks have the same number of edges ($L = 2000$) with varying number of total nodes ($n \in \{100, 150, 200, 250, 300\}$). In addition the number of controls is kept constant, $m = 1$, which is crucial for a fair comparison. The robustness profiles in Figure 1 are the result of averaging 50 percolation sequences for each network, as described by Equation 2. If we start with a control scheme (B) that

makes $G(A,B)$ fully controllable, the robustness profile will start with $n_r(0) = n$ and end at $n_r(L) = m$. Although the curves in Figure 1 are derived for a single realization of each ER network, we have observed that any ER network with the same n and L parameters will result in effectively indistinguishable robustness profiles, which we omit here for brevity.

Figure 1 shows that changing the number of nodes changes the curvature of the robustness profile, with lower numbers of nodes yielding a higher MIP robustness measure, as given by Equations 1 and 2. This result agrees with our intuition based on cacti-based control scheme construction: if the network is more dense, there are more edges for each node. In this case, it is more likely that m stems (node-disjoint paths) will be able to cover a larger part of the network. Moreover, more edges per node increase the chance that an alternative path exists around a location where an edge was removed.

With this insight we now aim to characterize the property (or properties) that help to determine the robustness of a controlled network. We now broaden the parameters used to generate the ER networks. In addition to those shown in Figure 1 we include several ER networks with the same number of nodes ($n = 200$) and also several with the same density ($p = 0.05$). Again we compute the robustness profiles and the corresponding MIP robustness measures for 50 percolation sequences. These were plotted against several choices of network parameters, such as n, L, and p. However, none had as strong a correlation as the mean degree $\langle k \rangle$, which is shown in Figure 2. This strong correlation supports the idea that there may be a causal relationship between degree-based network properties and the network's MIP robustness. More generally, the presence of a strong correlation with mean degree confirms that aspects of network structure can, indeed, influence the robustness of controllability, suggesting that other features besides mean degree may also confer some amount of robustness.

4.2 Real Networks

Although random networks can reveal certain underlying concepts of control robustness, our ultimate goal is to quantify, characterize, and engineer robustness in real-world systems. Fundamental to this investigation is determining whether real networks are more or less robust than might be expected by chance. For example, do certain systems have robust controllability as a design criterion whereas others sacrifice this in favor of other attributes such as performance, adaptability, and efficiency? We study the difference in MIP robustness between several real networks and their cooresponding degree-preserving random networks, including a food web, email correspondence (the East Anglia email dataset), and protein networks [14, 12]. We chose these networks because the concept of controllability is directly relevant to each: controls can disrupt or correct disruptions within the food chain of an ecosystem, influence the spread of information among a population, and be used to alter the trajectory of biochemical systems.

The comparison of robustness between two arbitrarily sized and controlled networks is challenging. In order to make such a comparison possible between a real

and a random network, we maintained the degree distribution of the network as well as its number of nodes and edges. We further consider the MIP robustness differences for a range of numbers of controls (m). In Figure 3 we show the MIP robustness for both real and shuffled networks, constraining the control set to a size m, where m ranges between 1 to n. The dashed and dotted vertical lines correspond to the number of controls, needed to fully control the real (m_c^{real}) and shuffled (m_c^{shuff}) networks, respectively. For all $m < m_c^{\text{real}}$, we observe that the shuffled networks (denoted by $\times$) are consistently more robust than the original real network (denoted by ■). However, after m_c^{real}, the relative robustness is reversed and the real networks are more robust. In a statistical t-test of the MIP robustness values, this switch is confirmed with a confidence beyond 0.05 for all tested networks (note that as $m/n \to 1$, the robustness difference disappears since both networks approach fully controllable even under percolation). This confidence is upheld prior to the dashed line and several nodes past it until the measures begin to coincide at the far right. In the upper and lower subplots, corresponding to the food web and yeast protein networks, the trends have a statistical significance beyond 0.0005.

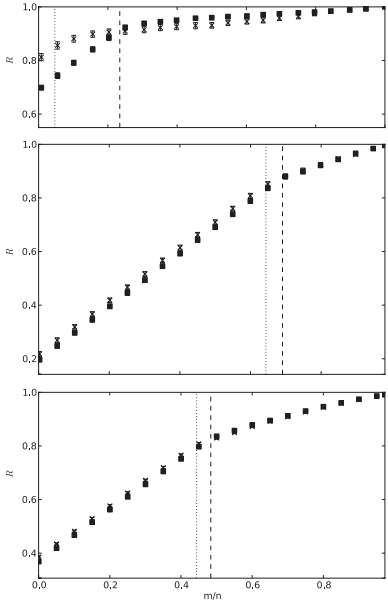

Fig. 3 The MIP robustness of real networks. In order from top to bottom, we compute robustness profiles and the corresponding MIP robustness measures (■) for a food web, email correspondence, and yeast protein networks. We compare these results with degree-preserving shuffled networks ($\times$). The dashed and dotted vertical lines correspond to the number of controls, m_c, needed to fully control the real and shuffled networks, respectively. A t-test confidence level beyond 0.05 confirms the definitive switch from shuffled to real networks being more controllable before and after the dashed line, respectively. For the food web and yeast plots this confidence is beyond 0.0005.

These observations provide an intriguing starting point for deeper investigation into the basis and origins of robust control in real networks. The fact that the switching behavior is both significant and conserved across networks drawn from such different real-world systems suggests that certain kinds of control robustness may be preferred by quite diverse natural systems. For example, the fact that robustness is enriched only for $m \geq m_c^{\text{real}}$, may imply that real systems are trading off robustness for other desirable attributes such as efficiency or adaptability. The fact that these real networks are more robust than networks drawn from degree-preserving random

models suggests that more subtle and, potentially, local structures are involved in implementing robust control.

5 Conclusion & Future Work

Where complex systems are useful or relevant to human pursuits, there will be an interest in the ability control and influence them. Because many natural systems experience frequent and unpredictable structural changes, we seek controls that are robust to changes in the structure of the controlled system.

In this paper, we have approached the issue of characterizing and comparing the robustness of networks. We have outlined a methodology for investigating the robustness of networks under edge removal, which augments the current methods in structured systems and graph theory literature.

There are a number of promising directions for future work in this area. In order to compare two real-world networks, we must develop a formal way of comparing robustness between control configurations of different sizes. Additionally, in order to understand the ways in which control robustness can be constructed, a careful assessment is needed of the contributions that different network properties (e.g., clustering, centrality, motifs) make to the robust controllability of a network.

References

1. L. A. S. A. e. a. David Lazer, Alex Pentland. Life in the network: the coming age of computational social science.
2. P. Erdos and A. Renyi. On random graphs. *Publicationes Mathematicae*, 6:290–297, 1959.
3. K. Glober and L. M. Silverman. Characterization of structural controllability. *IEEE Transactions on Automatic Control*, AC-21(4):534–537, 1976.
4. S. Hosoe and K. Matsumoto. On the irreducibility condition in the structural controllability theorem. *IEEE Transactions on Automatic Control*, AC-24(6):963–966, 1979.
5. H. Kitano. Systems biology: A brief overview. *Science*, 295(5560):1662–1664, 2002.
6. F. Li, T. Long, Y. Lu, Q. Ouyang, and C. Tang. The yeast cell-cycle network is robustly designed. *Proceedings of the National Academy of Sciences of USA*, 101(14):4781–4786.
7. C.-T. Lin. Structural Controllability. *IEEE Transactions on Automatic Control*, AC-19(3):201–208, 1974.
8. Y.-Y. Liu, J.-J. Slotine, and A.-L. Barabasi. Controllability of complex networks. *Nature*, 473:167–173, 2011.
9. R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon. Network motifs: Simple building blocks of complex networks. *Science*, 298(5594):824–827, 2002.
10. A. R. Rahmani, M. Ji, M. Mesbahi, and M. B. Egerstedt. Controllability of multi-agent systems from a graph-theoretic perspective. *SIAM Journal on Control and Optimization*, 48(1):162–186, 2009.
11. R. W. Shields and J. B. Pearson. Structural controllability of multi-input linear systems. *IEEE Transactions on Automatic Control*, AC-21(2):203–212, 1976.
12. S. Sun, L. Ling, N. Zhang, G. Li, and R. Chen. Topological structure analysis of the protein-protein interaction network in budding yeast. *Nucleic Acids Research*, 31(9):2443–2450, 2003.
13. H. G. Tanner. On the controllability of nearest neighbor interconnections. *CDC 43rd IEEE Conference on*, 3:2467–2472, 2004.
14. R. E. Ulanowicz, C. Bondavalli, and M. Egnotovich. Network analysis of trophic dynamics in south florida ecosystem, fy 97: The florida bay ecosystem. Technical Report UMCES-CBL 98-123, Chesapeake Biological Laboratory, 1998.