Reachability-based Robustness of Controllability in Complex Networks

Deven Parekh

Master of Science

School of Computer Science

McGill University

Montreal,Quebec

July, 2015

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of Master of Science

DEDICATION

This thesis is dedicated to my parents, who have always been proud of what I have accomplished so far.

ACKNOWLEDGEMENTS

I am sincerely thankful to my supervisor and mentor Prof. Derek Ruths and also to Prof. Justin Ruths, who have guided me in my research as well as my thesis.

ABSTRACT

In real world applications, control is always performed without perfect knowledge, perfect models, and often, under changing conditions. Such circumstances are particularly true of complex systems. As a result, application of control theory to complex systems requires the development and implementation of control policies that are robust to unexpected and potentially malicious changes to the underlying network. This thesis makes three important contributions along this direction. First, we introduce a new definition of robustness which captures realistic constraints imposed by many control problems. Second, we develop a novel algorithm for computing this robustness measure. Third, we conduct a thorough assessment of the control robustness of different synthetic networks to a wide array of attacks/network perturbations. We find that our robustness measure is behaviorally different from other robustness measurements in the literature and that the attacks considered highlight a number of ways in which network properties correlate with control robustness.

ABRÉGÉ

Dans les applications du monde réel, le contrôle est toujours effectuée sans une parfaite connaissance, des modèles parfaits, et souvent, dans des conditions changeantes. Ces circonstances sont particulièrement vrai des systèmes complexes. En conséquence, l'application de la théorie du contrôle de systèmes complexes nécessite l'élaboration et la mise en œuvre des politiques de contrôle qui sont robustes à des changements inattendus et potentiellement malveillants sur le réseau sous-jacent. Ce document fait trois contributions importantes le long de cette direction. Tout d'abord, nous introduisons une nouvelle définition de la robustesse qui capture des contraintes réalistes imposées par de nombreux problèmes de contrôle. Deuxièmement, nous développons un nouvel algorithme de calcul de cette mesure de robustesse. Troisièmement, nous procédons à une évaluation approfondie de la robustesse de contrôle des différents réseaux de synthèse pour un large éventail d'attaques / perturbations du réseau. Nous constatons que notre mesure de robustesse est comportemental différent des autres mesures de robustesse dans la littérature et que les attaques considérées mettre en évidence un certain nombre de façons dont les propriétés du réseau sont en corrélation avec le contrôle robustesse.

CONTRIBUTION OF AUTHORS

This thesis is based on the manuscript titled *Reachability-based Robustness of Network Controllability under Node and Edge Attacks* which was co-authored by my supervisor Prof. Derek Ruths and Prof. Justin Ruths [18]. The manuscript was mainly written by me with editing and guidance by the professors. Chapters 1 and 2 in this thesis are written by me. Chapters 3-5 in this thesis are based on the manuscript and the algorithms, implementations, and the results were generated by me with guidance of the professors.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1
Introduction

Networks are ubiquitous and they find applications in biological systems, transport systems, social interactions, food networks and many other natural as well as engineered systems. Complex networks are used to model and describe many of these systems that function around us everyday [1, 4, 15, 16]. Synthetic networks have been developed to model the structural characteristics of these natural networks, for purpose of studying their physical properties and developing networks of similar performance capabilities [3, 5, 9, 10]. Such networks have been analyzed for their tolerance against structural failures [2]. Apart from the studies of topological properties, recent work has highlighted the importance of understanding the extent to which complex network systems are controllable [14, 22].

A *controllable system* can be driven from any arbitrary state to any desired state in finite time through the application of external control inputs. For example, a network of power grid stations can be controlled externally through input signals from control units connected to some of the grid stations. The network is fully controllable (all the stations are controllable) if the state, such as voltage level, of each grid stations can be changed to any possible value in that state space. Like any property that is dependent on structure of complex networks, controllability is sensitive to perturbations or failures that occur to the networks. Understanding how network controllability changes in the context of node or link failures as well as

which network structure designs have control schemes that are most resilient to such failures is an essential part of making such theoretical formalisms practically useful. In this thesis, in order to understand and compare different synthetic networks for their resilience or capacity to maintain controllability under failures, we define and analyze a new measure of *robustness of the controllability*.

1.1 Problem Definition & Motivation

We strive to identify critical design characteristics of networks that provide increased controllability and robustness under component failures. Current research indicates a correlation between topological characteristics such as degree distribution and network controllability [14], but the role of underlying control structures is ignored. Effects of different types of node and link failures and attacks on robustness of network controllability have also been studied [17, 21], but with an indirect measure of robustness. In this study, we analyse the parameters such as our new measure of robustness, types of attacks, and structural properties of underlying control structures, to discover relationships between these characteristics and network controllability.

In current literature on network controllability, there appears to be a broader interest in statistics surrounding the number of controls necessary to control a complex network (e.g., [11, 14, 20, 22]). Certainly, it is practically useful to minimize the number of controls necessary to fully control a network. As opposed to research on network controllability, work on robustness in the context of control is nascent. Therefore, due to the broader interest, the recent studies on robustness of controllability have primarily measured robustness as the number of additional controls

required to maintain full controllability following a change (a failure or an attack) in the topology of the network [17, 21].

Such a definition of robustness effectively makes the assumption that new controls can be added to the network as components fail: e.g., such a definition of robustness assumes that one is always in a position to add more controls, only that we have a preference to add as few controls as possible. However, in practice, critical aspects of the system may be unknown, resources constrained, and regions flagged for direct control not easy to reach: thus, in many cases, controls cannot be simply added without great cost.

With this in mind, this thesis explores an alternative definition of robustness concerned with how the number of controllable (reachable) nodes changes due to a topological change (failure or attack). Such a definition offers a more direct way of interpreting robustness as it measures the loss of controllable nodes after a failure. On the other hand, the existing definition of robustness in literature measures an increase in number of controls needed after a failure to compensate for the loss of controllable nodes. In order to distinguish these notions of robustness, we refer to the existing definition as *control-based* and our proposed definition as *reachability-based robustness* (CR and RR, respectively). Different systems and conditions will determine which of these definitions will be appropriate to use - but certainly both capture practical constraints and objectives.

This thesis makes three core contributions. First, we formalize reachability-based robustness. This formalization involves the development of a complex algorithm which has been alluded to in literature, but (to our knowledge) has never

before been fleshed out or published [8, 19]. Second, using random network models, we establish how control-based and reachability-based definitions of robustness differ (and are similar) both over different types of networks and different types of attacks. Where attacks are concerned, we consider a much more comprehensive set than has been evaluated elsewhere in the literature. Specifically, we assess all standard node and edge attacks which depend on first-order degree properties. This extensive set of attacks constitutes our third contribution and reveals that there are significantly more nuanced factors determining the most effective attacks (and most robust configurations) than what is currently reported in the literature. Moreover, our study raises a number of questions about the relationship of robustness to network structure in general and to the nature of the control structures that govern the control of complex systems.

CHAPTER 2
Background and Prior Work

In this section, we outline characteristics of networks and the methods used to control them. This will serve as a basis and background for the analysis performed in this thesis.

2.1 Complex Networks

The use of networks to model systems, both simple and complex, has been prevalent for decades. Most commonly, these networks are constructed with *nodes* and *edges* according to graph theory. A node represents an entity, while an edge represents a connection between two nodes. This connection could be physically identifiable such as a power line in a power grid network, or it can be representational, such as human relationships in a social network. These edges can be either undirected or directed. The undirected edges imply that there is no directional constraint on flow of information between the nodes that share a particular edge. In a directed network, the connection only works in the direction marked by an arrow present at one end. It is very important to note that all of our study concerns with directed networks only. Additionally, edges can have weights associated with them. Weights can have different interpretation based on the different kinds of networks. In the case of an internet or router network for example, weight could represent bandwidth of a connection, while in an airport network, it could be the number of flights between two airports. In this thesis, the effect of edge weights is not studied in context of

controllability since we are interested only in structural controllability. The reason for ignoring weight is further explained in subsequent sections describing structural controllability.

Another important characteristic of networks is *degree* of a node which is the number of connections, or edges, connected to the node. This characteristic will be used extensively in this work, particularly in determining types of node and edge attacks used to define robustness of controllability. These basic properties of networks outlined before form the fundamental graph-theoretic representation of networks.

Network applications are extremely widespread and diverse in characteristics. Networks can be relatively small, for example the animal prey/predator food chain in a small community, or relatively large such as a network of human interactions. Table 2–1 shows some of the important real-world networks that are often used by the research community in the study of network controllability. In this thesis, we focus only on synthetic networks. However, it is important to note that our choice of synthetic networks and their parameters is such that they closely model many important types of real networks.

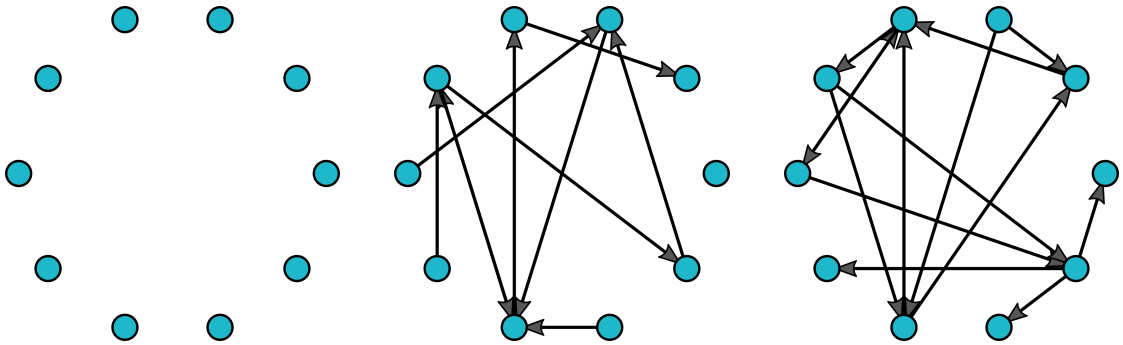
2.1.1 Synthetic Networks

Synthetic networks have been developed by researchers interested in modeling real networks to analyse their topological aspects as well for understanding mechanisms to design optimal networks. Many natural networks such as food chains, for example, have slowly evolved over years to adapt and survive under constantly varying environmental parameters. Understanding these networks provides valuable

Table 2–1: Examples of real-world networks

Networks	Description
Airport Networks	Nodes represent airports, and edges exist wherever there are flight(s) between the airports.
Amazon Co-purchase	The relationship between co-purchasing patterns on Amazon.com, indicating that the source of an edge is often copurchased with the target
C. Elegans Neural Network	A graph of Caenorhabtitis elegans (C. elegans) worm's neural network. Neurons are nodes and edges indicate existance of at least one synapse or gap junction between neurons
Corporate Ownership	Ownership relations among companies, where a directed link indicates that the source is an owner of the target
E-coli Transcription	A transcriptional regulation network for E. coli encoding interactions between transcription factors and operons
Food Web	Networks from various marine ecosystems, where the orientation of networks is such that directed edges point towards the flow of biomass, e.g., edges point from prey to predators.
Gnutella Networks	Nodes represent hosts in the Gnutella network topology and edges represent connections between the Gnutella hosts
Macaque Neural	Networks representing structural (axon projections) cortical connectivity in Macaque monkeys
Social networks	Online social networks, edges represent interactions between people
World Wide Web	Nodes represent webpages and edges are hyperlinks

insights into the principles of better network design. Commonly referred to as random graphs, the synthetic models used in this study are based on graph theory and probability theory. A random graph is generated when a given node or a set of nodes is added to a graph and then edges are added to connect them at random. Different formation mechanism yield a number of network properties and topologies, which characterize each type of the random graphs. While a number of such models exists, the following four types were chosen for this study for their unique characteristics and applications. These models and parameters for each of them were selected such that they represent a wide range of real world networks mentioned earlier.



(a) $p = 0$ (b) $p = 0.1$ (c) $p = 0.2$

Figure 2–1: An illustration of the Erdős-Rényi model using a 10 node network. By increasing probability p, the number of edges scales accordingly.

Erdős-Rényi Model. The Erdős-Rényi model is one of the simplest of the exponential random graph models. Under this model, a network of N nodes is generated using a single parameter, the probability p of existence of an edge e_{ij} from node n_i to n_j. Each edge's existence is independent of other edges. Therefore, given

N nodes, the expected number of edges [6] is $E\{L\} = pN(N-1)/2$. Figure 2–1 shows examples of ER graphs for different values of the probability p.

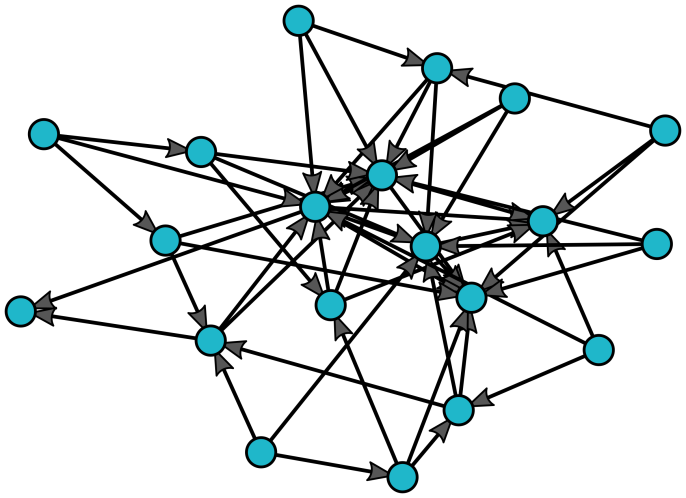


Figure 2–2: A Barabási-Albert network exhibiting preferential attachment characteristics.

Barabási-Albert Model. The Barabási-Albert (BA) model generates random scale-free networks using a preferential attachment mechanism. Scale-free networks are widely observed in natural and human-made systems, including the Internet, the world wide web, citation networks, and some social networks. Degree distribution of scale-free networks closely follow that of a power law i.e. the fraction of nodes with degree k, is of the form [3]

$$P(k) \sim k^{-\alpha}$$

In the case of BA networks, the power in equation above is taken as $\alpha = 3$. Formation mechanism of BA networks is as follows. Initially, there are m_0 connected nodes. Then each new node i is connected to m ($\leq m_0$) existing nodes with a

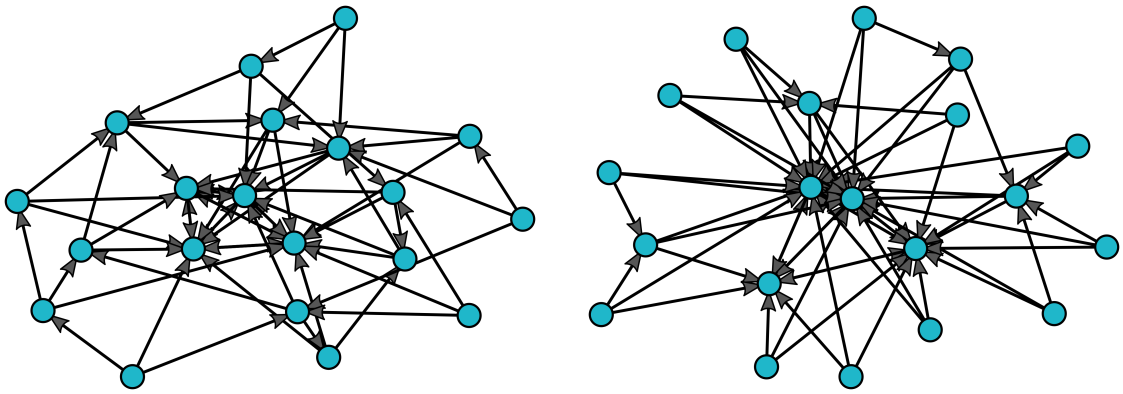
probability that is proportional to number of link that existing nodes have. The probability p_j that a new node connects to an existing node j is given by

$$p_j = \frac{k_j}{\sum_m k_m}$$

where k_j is degree of node j [1]. It can be seen that the new node prefers nodes with a higher degree, i.e. the nodes that are already more populated with connections. This signifies a concept called preferential attachment, which characterizes many scale-free networks. A sample Barabási-Albert network is shown in Figure 2–2. It can be clearly seen that very few nodes have much higher degree compared to the other nodes in the network.

Local Attachment Model. Proposed by Jackson and Rogers [10], the local attachment (LA) model is built from behavior typically seen in social networks. As the name suggests, the model generates a random networks where nodes find other nodes to form links with in one of the two ways: some are found uniformly at random, while others are found by searching locally through the current structure of the network (e.g., meeting friends of friends). In this model, nodes are added incrementally with m edges each. Of the m edges, r are connected randomly from a new node to existing nodes in network, while remaining $m - r$ edges are connected to neighbors of randomly chosen nodes. Clustering can be added to the network by increasing the fraction $m - r$. Thus, the model can be considered to have two parameters, number of links to add for each node (m) and clustering (c). From these parameters, the fraction of nodes to be added randomly can be obtained as $r = (1 - c)m$. This methodology ensures that the outward bound connections, or

out-degree, of each node is near uniform, but the in-degree will depend on how long the node has existed. Figure 2–3 shows two local attachment networks with different clustering values.

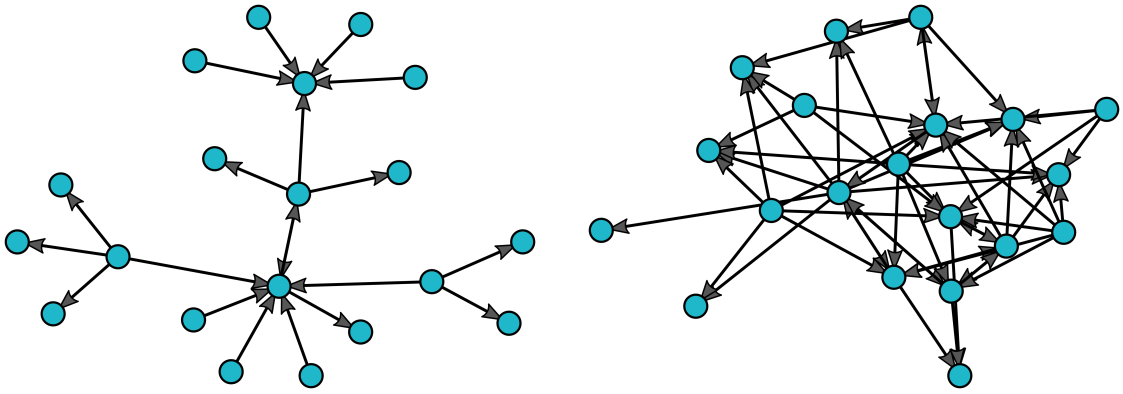


(a) Low clustering (b) High clustering

Figure 2–3: Two 15-node local-attachment networks with different clustering.

Duplication-Divergence Model. The Duplication-Divergence (DD) model was developed to describe the evolution of protein-protein interaction networks. The particular DD model used in this thesis is considered asymmetric, based on replica proteins that carry some but typically not all of the interaction links of the proteins from which they sprouted [9]. In this model, we start with an initial directed network having two nodes and two edges; an edge from node 0 to node 1 and vice versa. Then at each step, a random node n is duplicated to obtained a new node n' which is then connected to each neighbor of the node n with a probability s. If the new node n' does not obtained any connections, it is removed. This simple one-parameter network has been shown to approximate the degree distribution of realistic protein-protein

networks. Figure 2–4 shows two sample DD networks with different probabilities of duplication.



(a) Low probability of connecting to neighbours (b) High probability of connecting to neighbours

Figure 2–4: Two 15-node Duplication-Divergence networks with different probabilities of connections.

2.2 Controllability of Complex Networks

Consider a linear dynamical system represented by a directed graph $G(A)$, composed of N nodes and L edges, where A is the $N \times N$ adjacency matrix of the network for which the component a_{ij} is the edge weight from node x_i to node x_j. Dynamics of such a system can be described using a set of mathematical equations governing the state of each node in the system and any effect that one node may have on the other. Though many real world systems are complex, non-linear and non-homogeneous, thorough analysis of a linear model is a key step in generalizing to non-linear dynamics. In fact, many natural phenomenas fit well to a linear

time-invariant model of the form,

$$\frac{d}{dt}x(t) = Ax(t)$$

where $x(t) = [x_1(t), x_2(t), ..., x_N(t)]^T$ is a vector denoting the states of the N nodes at time t [14].

The above equation can describe dynamics of a network where nodes can change states only under influence of other connected nodes. In other words, external influences from outside of the network on states of the nodes are ignored. However, it is often desirable to influence networks externally for functional or performance related reasons. For example, in case of a power grid network, some grid stations might be required to be controlled or regulated directly for desired voltage levels. Similarly, networks usually require external influences in case of unexpected node or link failures. Sometimes we might desire to change states of a whole network by controlling few nodes only. For example, in case of social networks, where people are represented as nodes in the network and the connections (friendship, association, trust) between people as links, it would be desirable, for the purpose of advertisement for example, to be able to drive propagation of certain sentiment within the population by controlling a few highly influential individuals in the population. In case of airport networks modeling the flow of flight traffic, it is necessary to regulate as few control locations as possible for efficient performance.

In order to incorporate external control inputs in the model, we need to augment the dynamic system defined above. Then, such a controllable dynamical system (A, B) is one that can be driven from a given initial state to any desired final state

within a finite amount of time. Here, the $N \times m$ matrix B indicates the components or nodes of the system where m control input signals are applied. The control-augmented graph $G(A, B)$, formed by adding additional *control nodes*, one for each control inputs, to the original network $G(A)$, provides efficient methods for dealing with large-scale networks. The controlled network $G(A, B)$ can be described with the governing equation,

$$\frac{d}{dt}x(t) = Ax(t) + Bu(t) \tag{2.1}$$

where $u(t) = [u_1(t), u_2(t), ..., u_m(t)]^T$ is a vector of m control input signals at time t.

Let us consider a directed graph $G(A, B)$ shown in Figure 2–5, that represents a simple controlled system with the linear dynamics given by

$$\frac{d}{dt}\begin{bmatrix} x_1(t) \\ x_2(t) \\ x_3(t) \\ x_4(t) \\ x_5(t) \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ a_{31} & a_{32} & 0 & 0 & 0 \\ a_{41} & 0 & a_{43} & 0 & 0 \\ a_{51} & 0 & a_{53} & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \\ x_3(t) \\ x_4(t) \\ x_5(t) \end{bmatrix} + \begin{bmatrix} b_1 & 0 \\ 0 & b_2 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} u_1(t) \\ u_2(t) \end{bmatrix}$$

In this system, the matrix A represents the interaction strengths a_{ij} among the nodes in $G(A)$. The matrix B encodes the strengths (b_1, b_2) of control inputs ($u_1(t)$, $u_2(t)$) from control nodes X and Y (external to network $G(A)$) as well as locations of the nodes where the control inputs are applied. The control nodes X and Y are directly controlling nodes 2 and 1 respectively. Then, nodes 1 and 2 can further influence other nodes connected to them such that all nodes in the path $2 \to 3 \to 5$ are controlled by control node X and those in the path $1 \to 4$ are controlled by

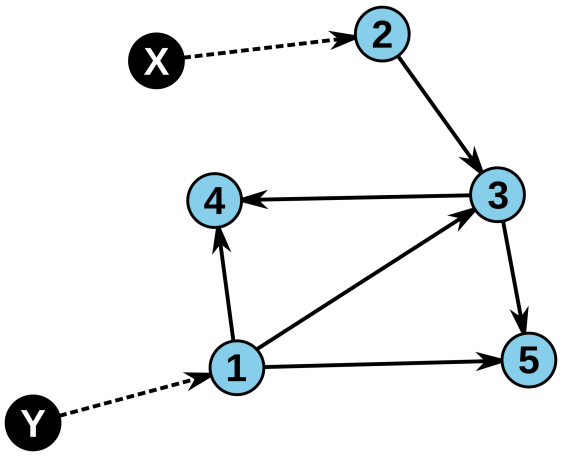


Figure 2–5: A simple network $G(A)$ with 5 nodes, when augmented by adding control nodes X and Y, is represented as $G(A, B)$.

control node Y. The nodes that can be controlled are called *controllable nodes*. Note that there are other possible choices of paths such as $X \rightarrow 2 \rightarrow 3 \rightarrow 4$ and $Y \rightarrow 1 \rightarrow 5$, through which control inputs can be applied such that total number of controllable nodes are same.

A given system $G(A, B)$ is controllable (all N nodes in $G(A)$ are controllable) if and only if the controllability matrix given by

$$C = [B, AB, A^2B, ..., A^{N-1}B] \tag{2.2}$$

has full rank, i.e., $rank(C) = N$. This mathematical condition for controllability is called the Kalman's controllability rank condition [12].

There are some important questions about network controllability that current literature has attempted to answer: (1) Given a system $G(A)$ how to find a minimal B ($B \in \mathbb{R}^{N \times m}$ with smallest possible m) such that $G(A, B)$ is fully controllable (all

N nodes are controllable)? (2) Given a system $G(A, B)$ how many nodes are controllable? In practice, since resources available to deploy control inputs are limited, it is necessary to find minimum number of controls to achieve full controllability. Hence question (1) above is important and practically useful to solve. As we will see later in this thesis, answer to the question (1) is used as a basis for definition of robustness in the current literature. While an answer to the question (2), is the basis for robustness defined in this thesis.

2.2.1 Structural Controllability

The present thesis purposely ignores weights on the edges or links; in other words, the matrices A and B in equation 2.1 consists of either 0 or 1 instead of any real-valued weights. This is because in many dynamical systems to which this analysis is relevant, the strength or weight of a single connection or link in a complex network changes with time, and is rather expensive to measure or could possibly be unknown. The direction of influence however may be known, and therefore the focus of this work is on the generic properties of such networks, i.e., those that hold for *almost all* parameter values. Exact methods exist to find minimum set of controls to achieve full control of networks with arbitrary structures and link-weights [23]. However, as link weights for many real networks are either unknown or dynamic in nature, *structural controllability* introduced by Lin [13] is used by many researchers to deal with control of such systems. Furthermore, the Kalman's controllability rank mentioned before is computationally extensive to compute for large-scale networks. Therefore, tools from structural controllability described below make the analysis tractable.

Structural controllability deals with control of a network using only topological structure of the system, therefore weights of the edges can be ignored. It has been shown that if the system is structurally controllable, it is controllable for almost all choices of edge weights except for some pathological cases [14]. Let us consider some example networks as shown in Figure 2–6 to understand controllability and structural controllability.

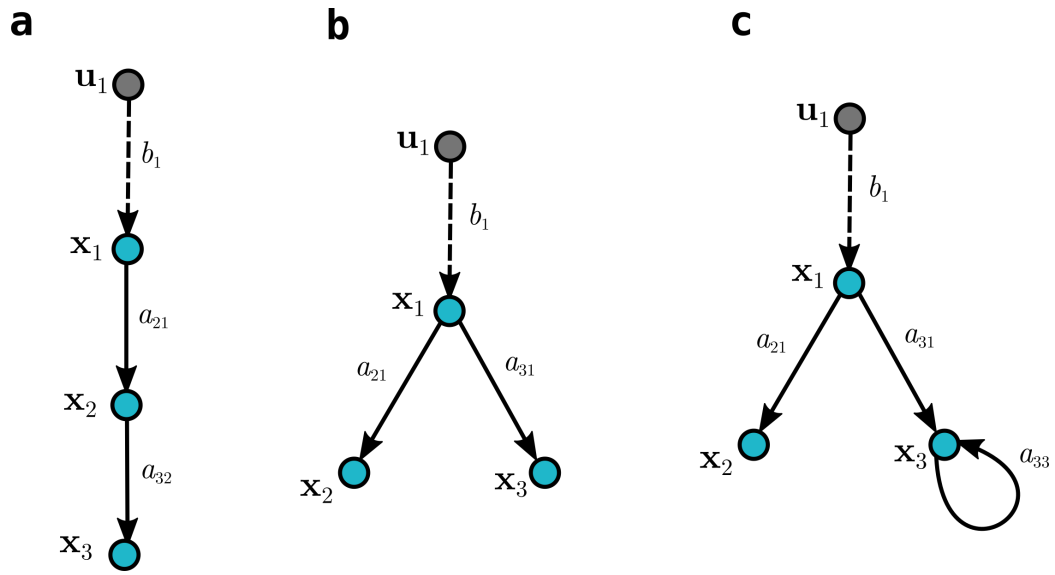


Figure 2–6: Controlling simple networks. (**a**) A directed path that can be completely controlled by controlling the starting node only. (**b**) A directed star can never be completely controlled by controlling the central hub (node X_1) only. (**c**) This example network, generated by adding a self-edge to the node X_3 of the star shown in **b**, can be completely controlled by controlling node X_1 only. (**d**) This network is completely controllable for almost all weights combinations except some pathological cases.

The linear dynamics of the network shown in Figure 2–6(a) can be written as

$$\frac{d}{dt}\begin{bmatrix} x_1(t) \\ x_2(t) \\ x_3(t) \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ a_{21} & 0 & 0 \\ 0 & a_{32} & 0 \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \\ x_3(t) \end{bmatrix} + \begin{bmatrix} b_1 \\ 0 \\ 0 \end{bmatrix} u(t)$$

The controllability matrix of the Kalman's rank condition is given by

$$\mathbf{C} = [\mathbf{B}, \mathbf{A}\cdot\mathbf{B}, \mathbf{A}^2\cdot\mathbf{B}] = b_1 \begin{bmatrix} 1 & 0 & 0 \\ 0 & a_{21} & 0 \\ 0 & 0 & a_{21}a_{32} \end{bmatrix}$$

Since $rank(\mathbf{C}) = 3 = N$, the system is controllable. Moreover, the system is always controllable if the weights a_{21}, a_{32} and b_1 are non-zero. Since the controllability is independent of the weights, it is called structural controllability.

The linear dynamics of the network shown in Figure 2–6(b) can be written as

$$\frac{d}{dt}\begin{bmatrix} x_1(t) \\ x_2(t) \\ x_3(t) \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ a_{21} & 0 & 0 \\ a_{31} & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \\ x_3(t) \end{bmatrix} + \begin{bmatrix} b_1 \\ 0 \\ 0 \end{bmatrix} u(t)$$

The controllability matrix is given by

$$\mathbf{C} = [\mathbf{B}, \mathbf{A}\cdot\mathbf{B}, \mathbf{A}^2\cdot\mathbf{B}] = b_1 \begin{bmatrix} 1 & 0 & 0 \\ 0 & a_{21} & 0 \\ 0 & a_{31} & 0 \end{bmatrix}$$

Since $rank(\mathbf{C}) = 2 < N$, the system is uncontrollable. In fact, this is independent of the weights; the system is uncontrollable no matter how their values are tuned.

The linear dynamics of the network shown in Figure 2–6(c) can be written as

$$\frac{d}{dt}\begin{bmatrix} x_1(t) \\ x_2(t) \\ x_3(t) \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ a_{21} & 0 & 0 \\ a_{31} & 0 & a_{33} \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \\ x_3(t) \end{bmatrix} + \begin{bmatrix} b_1 \\ 0 \\ 0 \end{bmatrix} u(t)$$

The controllability matrix is given by

$$\mathbf{C} = [\mathbf{B}, \mathbf{A}\cdot\mathbf{B}, \mathbf{A}^2\cdot\mathbf{B}] = b_1 \begin{bmatrix} 1 & 0 & 0 \\ 0 & a_{21} & 0 \\ 0 & a_{31} & a_{33}a_{31} \end{bmatrix}$$

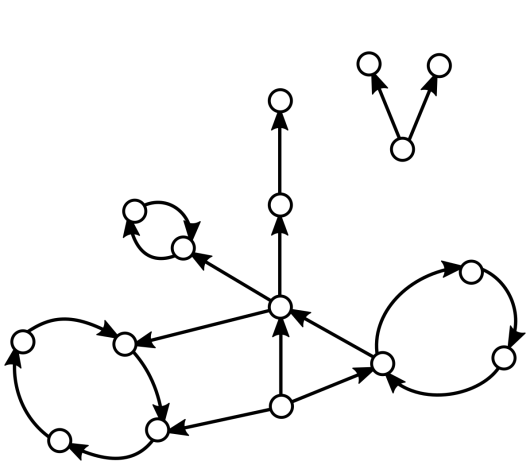
Since $rank(\mathbf{C}) = 3 = N$, the system is controllable. Moreover, the system is always controllable if the weights a_{21}, a_{32} and b_1 are non-zero. Thus, the system is structurally controllable. Note the difference between Figure 2–6(b) and 2–6(c). Presence of self-loop, or a cycle, changes the controllability of the system such that cycles can be effectively considered as self-controlling and do not require extra control inputs.

2.2.2 Control Structures

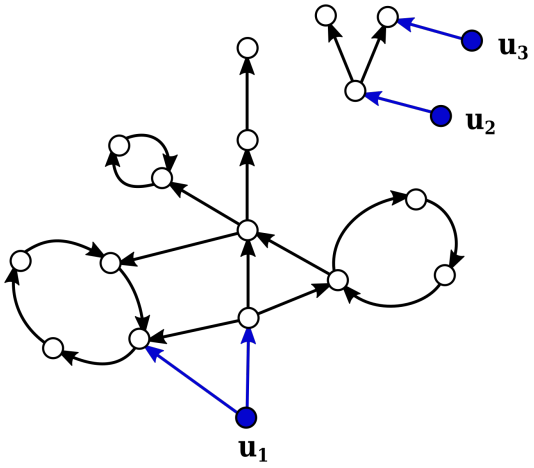
In his highly influential paper, Lin [13] formed a network representation of structural systems and defined key structures within the network that characterize controllability of the network. These controls structures are defined below.

- **Stem.** It is an elementary path $x_0, x_1, ..., x_t$ in network $G(A, B)$ from a node x_0 (also called *origin* of the stem) to node x_t (also called *terminus* of the stem) such that the origin x_0 is under direct control of an external control input. Each stem requires exactly one external control input (attach to origin of the stem) such that all nodes in the stem are controllable.
- **Cycle and Bud.** Cycle is a path in $G(A, B)$ such that its starting node and its ending node are the same i.e. the path forms a closed loop. Moreover, unlike a stem, a cycle might not be directly controlled by an external control input. As opposed to stem, cycles are self-regulatory and do not need extra control inputs; existing control inputs used for stems can also accommodate cycles. *Bud* is a cycle with an additional edge, the *distinguished edge*, that enters a node of the cycle from the a node of a stem. Thus, a bud is indirectly controlled via a stem. A non-bud cycle is one that is not attached to any stems; it is an independent cycle and it is directly attached to an external control input.
- **Cactus.** It is a subgraph composed of stems and cycles and is formed recursively as follows. A stem is a cactus. Given a stem S_0 and cycle $C_1, C_2, ...C_l$, the union $S_0 \cup C_1 \cup C_2 \cup ... \cup C_l$ is a cactus if for every $i(1 \leq i \leq l)$ the source node of the distinguished edge of C_i is not the terminus of S_0 and is the only node belonging to both C_i and $S_0 \cup C_1 \cup C_2 \cup ... \cup C_{i-1}$. A set of node-disjoint cacti is called a cacti.

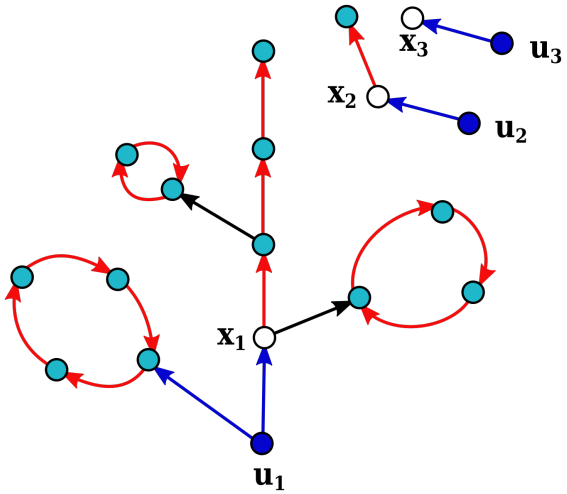
Figure 2–7 gives an overview of cacti structure in a sample network. Original network denoted by $G(A)$ is augmented with control nodes u_1,u_2 and u_3 to form a controlled network $G(A, B)$, where matrix B encodes nodes to which the control



(a) Network **G(A)**



(b) Controlled Network **G(A,B)**



(c) Cacti Structure

Figure 2–7: An illustration of Cacti Structure in a network

inputs are applied (shown with blue arrows in Figure 2–7(b)). Since the three control nodes can fully control the network, there is a cacti which spans the network as shown in Figure 2–7(c). The cacti is composed of three stems starting with nodes x_1,x_2 and x_3 (a stem can have just one isolated node). There are also three cycles in the cacti, one of which is an independent (non-bud) cycle directly controlled by u_1 while the other two are buds that are indirectly controlled by u_1. Since there are three stems, the number of controls required to fully control the network is also 3.

Nodes in $G(A, B)$ can be uncontrollable in two scenarios, *inaccessibility* and *dilations*. A node is inaccessible if there is no directed path reaching the node from any of the control nodes. Inaccessible nodes are nodes that are simply not reachable from the control nodes, hence it is not possible to exert a controlling influence over them. A dilation exists in the graph $G(A, B)$ if a subset S of nodes in $G(A)$, can be found such that the number of nodes in the inbound neighborhood set of S, given by $|T(S)|$, is smaller than the number of nodes in S, given by $|S|$. The inbound neighborhood set $|T(S)|$ is the set of nodes with directed edges going into S. Simply put, dilations imply a situation in the network whereby there is not a sufficient number of inputs to control all nodes in S. There are at most $|T(S)|$ control inputs leading into S and $|T(S)| < |S|$. The existence of cacti that span the graph $G(A, B)$ relies upon the control-augmented graph having no inaccessible nodes and no dilations.

A fundamental result from structural controllability thereom by Lin [13] states that the following three statements are equivalent:

1. A linear control system (A, B) is structurally controllable.
2. (a) The directed graph $G(A, B)$ contains no inaccessible nodes.

(b) The directed graph $G(A, B)$ contains no dilations.

3. $G(A, B)$ is spanned by cacti.

Since only stems require control inputs and cycles don't, if one can find a cacti spanning the network $G(A, B)$ that minimizes number of stems, one can find minimum number of controls required to fully control the network. Using statement 2 mentioned above, a method to find such a cacti is described in the next section.

2.3 Finding Control Structures

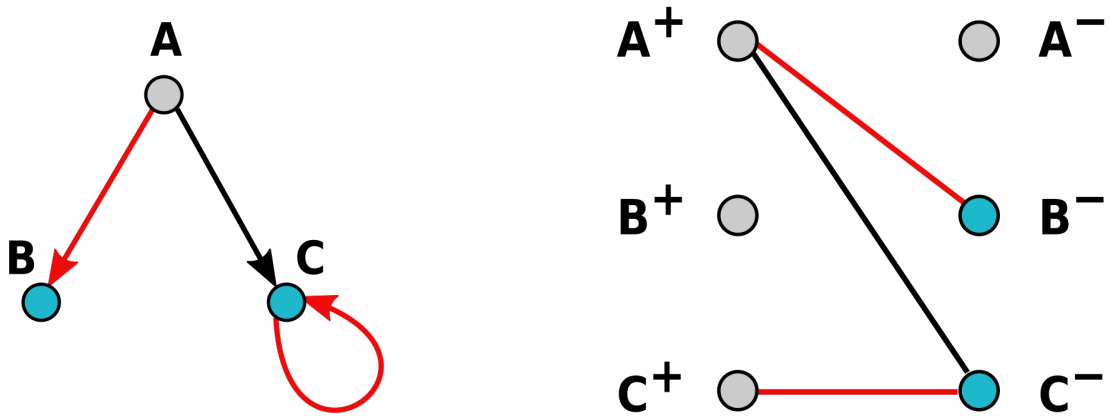
Here we describe a method to find the control structures in a given network in order to obtain minimum number of control inputs and their locations to fully control the network.

As we saw in the last section, an important task in structural controllability of networks is to construct a minimal B ($B \in \mathbb{R}^{N\times m}$ with smallest possible m), such that there are no inaccessible nodes and there are no dilations. A solution to the task can be found in the maximum matching algorithm from graph theory that can find a set of disjoint and simple paths and cycles which maximally cover nodes of a given network $G(A)$ [7, 14]. Edges in the maximum matching obtained from the algorithm can be joined to form stems and cycles. The *unmatched nodes* from the algorithm are the nodes without any matched inbound edges. They identify the location of either inaccessible nodes or dilations in $G(A)$. By connecting *control nodes*, which represent the external control inputs, to these unmatched nodes, we can create a spanning cacti for $G(A, B)$ without inaccessible nodes or dilations. A cycle which is not a bud (no distinguished edge exists from a node in a stem to a node in the cycle) can be converted into a bud by adding a distinguished edge from a control node to a

node in the cycle. The matrix B then encodes the new edges from the control nodes to the unmatched nodes and the new distinguished edges from the control nodes to the non-bud cycles. This method, due to Liu et al. [14], is described in detail below with an example of the maximum matching algorithm.

2.3.1 Maximum Unweighted Matching in Directed Graphs

Maximum matching of a network is the largest set of nodes that can be uniquely paired amongst themselves using edges within the network. Algorithms exist for obtaining the maximum matching of bipartite graphs [7]. However, most networks are not bipartite graphs and therefore, an extra step must be taken to convert a directed graph to an undirected bipartite graph.



(a) (b)

Figure 2–8: An illustration of finding minimum number of controls inputs using maximum matching. (a) A directed network $G(A)$. (b) Undirected Bipartite network $G_B(A)$ showing a maximum matching.

Figure 2–8(a) depicts the original network $G(A)$. Figure 2–8(b) shows $G_B(A)$, the bipartite representation of $G(A)$. In $G_B(A)$, each node from $G(A)$ has been translated into two nodes, each belonging to one of the two groups: the *positive* $(+)$ group representing the out-bound edges of $G(A)$ and the *negative* $(-)$ group representing the inbound edges. For example, the edge $A \rightarrow B$ in $G(A)$ is translated to the edge $A^+ \rightarrow B^-$ in $G_B(A)$. Now, the Hopcroft-Karp algorithm can be applied to $G_B(A)$ to get a maximum matching in $G_B(A)$, which is shown as the red-colored edges in Figure 2–8(b). This matching in $G_B(A)$ are translated back to a corresponding matching in $G(A)$ which is also shown as the red-colored edges. *Matched nodes* in $G(A)$ have an in-bound edge (shown in red) which is included in the matching. Thus, nodes B and C, shown in blue, are matched. *Unmatched nodes*, shown in grey, don't have an in-bound edge which is in the matching. Therefore, the unmatched nodes must be controlled directly by external control inputs. In Figure 2–8(a), the unmatched node A must be directly controlled. All matched nodes in $G(A)$ can be joined together using edges in the matching to form stems and cycles. The unmatched nodes are the origins or starting nodes of these stems. In Figure 2–8(a), $A \rightarrow B$ is a stem, while $C \rightarrow C$ is a cycle.

The maximum matching ensures that as many nodes as possible are matched; the number of unmatched nodes is minimized. Therefore, the minimum number of control inputs and their locations are given by the unmatched nodes. In many cases, a maximum matching is not unique; different sets of edges can be chosen to produce different maximum matchings of the same size. However, the number of unmatched nodes, and hence the number of control inputs, remains the same.

2.4 Robustness

Networks, and the underlying control structures, are susceptible to failure and attack, and thereby must be designed to survive these unavoidable events. Their robustness, or ability to function, after links or nodes have been removed or incapacitated is a key characteristic and measurement of the quality of their design. We first look at previous research that attempts to define and understand robustness of networks with regards to their topological structure. Then we discuss previous works on robustness of controllability (or control structure) of networks.

2.4.1 Robustness of Networks

The robustness of synthetic networks, in particular the exponential random models (such as Erdős-Rényi) and scale-free models (such as Barabási-Albert), have been studied in depth. Most of these studies view robustness of a network as its resilience to any change in important network properties under node or edge failure or removal.

Albert et al. [2] studied behavior of the scale-free and the exponential networks under different kinds of node failures. One of the network properties they used to study resilience or robustness of the networks was *diameter*, which they defined as "the average length of the shortest paths between any two nodes in the network". Figure 2–9 is one of their results which shows changes in the diameter d of the networks as a fraction of nodes are removed from the networks in steps. They considered two kinds of node removal or percolation: random (when nodes are removed randomly, simulating a random failure) and degree-based (when nodes with highest

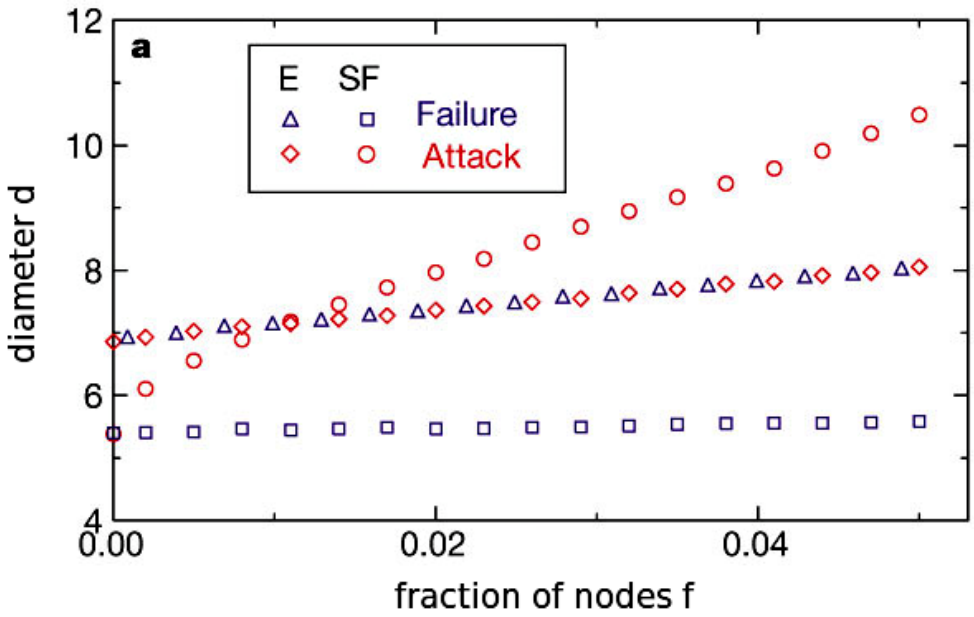


Figure 2–9: Comparison between the exponential (E) and scale-free (SF) network models, each containing $N = 10,000$ nodes and $20,000$ links. The blue symbols correspond to the diameter of the exponential (triangles) and the scale-free (squares) networks when a fraction f of the nodes are removed randomly (error tolerance). Red symbols show the response of the exponential (diamonds) and the scale-free (circles) networks to attacks, when the most connected nodes are removed. [2]

degree are removed first, simulating a targeted attack). They found that, under random node failure, the scale-free network exhibited better durability or robustness, through lower recorded diameter on average, than the exponential networks. This can be largely attributed to the significant differences in topological structure of the two networks. The exponential network exhibits this behavior because of its relative homogeneity; since all the nodes have approximately the same number of links, they contribute equally to the interconnectedness of the network and therefore, each

node removal inflicts the same relative amount of damage. Meanwhile, the scale-free network is heterogeneous; there exist only a few nodes with a high degree of connectivity and a large fraction of nodes with a smaller degree. Therefore, under random failures, loss of a low-degree node is more likely than that of a high-degree node. However, contrary to random failures, under high-degree attacks on nodes (when nodes with highest degree are removed first), the scale-free networks exhibit a significant increase in diameter, while the exponential networks perform almost the same as in case of random failures. Again, this can be attributed to the network homogeneity.

The above study uses a method of understanding and measuring robustness of a network, in which changes in topological structure of the network are observed under node or link failures. Such a method can also be applied to understand robustness of different kinds of network properties such as controllability. The same concept is used in the current literature on robustness of controllability as well as this thesis.

2.4.2 Robustness of Controllability

In this thesis, "robustness of controllability" is viewed as resilience of a network with regards to its controllability, or the ability of a network to maintain controllability under node or link failures. Understanding the robustness under this view or interpretation, also appears to be the goal of other researchers discussed below, when they define and formulate models of the robustness. However, our novel definition and model of the robustness are more directly related to the above goal, and hence, are one significant step further in strengthening our understanding of the robustness.

As we have already seen, often it is desirable to find minimum number of controls N_c such that a network is fully controllable. When a network undergoes node or link failure, it may no longer maintain full controllability with the same number of controls N_c. Therefore, under failures, a network loses controllability and may require more controls to maintain full controllability. To our knowledge, all the existing studies of robustness of controllability have measured the increase in the minimum number of controls required as a proxy for the reduction in controllability due to a failure. We refer to this indirect approach of measuring robustness as *control-based robustness* (CR).

Liu et al. [14] showed that sparse and heterogeneous networks are difficult to control and provided a method to observe robustness under edge failures based on change in N_c upon edge removal. Pu et al. [21] investigated the behavior of controllability of various networks under random, targeted, and cascading failures of nodes. As shown in Figure 2–10, they found that under degree-based failures or attack, both Erdős-Rényi (ER) and scale-free (SF) networks need more controls on average to maintain full controllability than they need under random failures. Hence, they conclude that for both the networks, degree-based attacks are more effective (damaging) than random attacks, i.e. both the networks are less robust against degree-based attacks than random attacks. Furthermore, they observed that a larger number of edges and greater network homogeneity increases the robustness of network controllability. Finally, Nie et al. [17] analyzed robustness of control under random and targeted cascading failures. They report that ER networks with smaller average degrees are more robust against a highest-load cascading attack while SF networks with

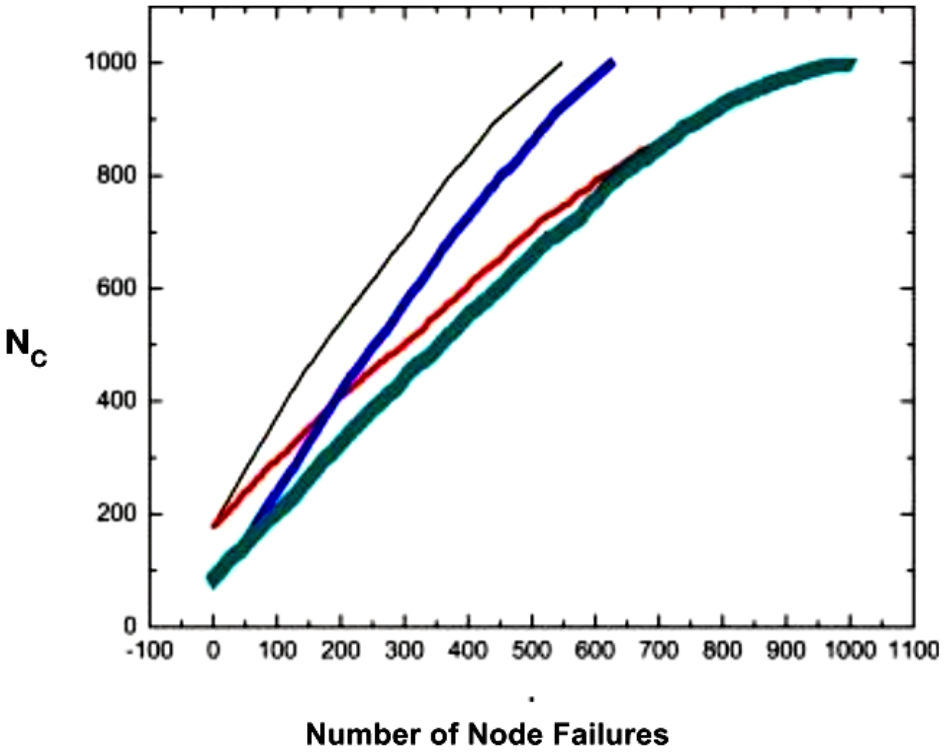


Figure 2–10: Minimum number of controls required for full control N_c vs number of node failures. Scale-free network is shown in black (degree-based failure) and red (random failure). ER network is shown in blue (degree-based failure) and dark-green (random failure). [21]

smaller power-law exponents are more vulnerable than those with large exponents. Furthermore, random attacks are shown to be more effective than targeted attacks for less heterogeneous networks under moderate edge removal rates.

Such a control-based robustness highlighted in the above studies, is measured in terms of change in minimum controls required, and hence has some practical limitations. In real world systems, resources to assign controls are limited and it is not always possible to bring system under full control after a failure. In light of

such practical issues, it is desirable to have a measure of robustness which provides a better perspective on how controllability of a network changes under failures when the controls assigned to the network cannot be increased. In the subsequent sections, we define this new measure of robustness, which we call reachability-based robustness, as well as we perform analysis on various synthetic networks using a wide array of node and edge attacks to demonstrate similarities as well as differences between the reachability- and control-based robustness measures.

CHAPTER 3
Methods and Data

In this section, we define reachability-based robustness, the algorithmic means by which it is calculated, and the network models as well as the attacks which will be used to empirically assess the attributes of the both control- and reachability-based robustness.

3.1 Reachability-based Robustness

Our goal is to investigate a robustness measure that focuses on how much of the network *remains under control* in the presence of an attack. Unlike control-based robustness, in our measure, no new controls are added after a node or link failure. The original set of controls designed for the original network remain in place (except those whose nodes were removed by the attack, if any). Our measure of robustness asks how many nodes are *still* under control after the perturbation to the network topology. In this way, we can expect the number of nodes under control (or the *controllable nodes*) to decrease with increasing node or link failures. To assess the robustness of a network over a series of failures affecting 0 to 50% of the network (nodes or edges, depending on the attack), we consider the average or mean of the number of controllable nodes for each step of the series.

Notice, however, that our formulation thus far requires a particular assignment of controls to nodes in the network (called the *control configuration*) before a series of failures occurs. In order to obtain a robustness measure for a network, we sample

multiple control configurations for the network and report the robustness score as the average of values obtained for each configuration. In this study, we used a sample of 10. While a small number, this yielded results with extremely low variance. As an interesting direction for future work, this suggests that different minimal control configurations have highly similar robustness for a wide array of events.

Returning to the issue of computing our robustness value for a network, however, we still have a problem. Specifically, how do we compute the number of controllable nodes in an arbitrary network controlled by an arbitrary control configuration (the need for handling an "arbitrary" network stems from the fact that a perturbation could affect a network in any number of ways)? In order to do this, we require a simple and efficient algorithm for finding the cacti control structure given a fixed set of controls. While this problem is discussed as *finding the generic dimension of controllable subspace* in literature [8, 19], quite remarkably, a clear algorithmic approach appears to be missing. This algorithm is the first of our contributions in this thesis.

To understand the problem, consider the example in Figure 3–1(a), which shows a network G with two controls X and Y attached to driver nodes A and B. The number of minimum controls N_c is 2 while number of controllable nodes N_r is 8. Figure 3–1(b) demonstrates an increase in N_c after node C is removed from G. After removal of C, in order to fully control the network, a new control Z needs to be attached to the unmatched node D. Thus, N_c increases to 3. On the other hand, Figure 3–1(c) shows the decrease in controllable nodes N_r given the same set of

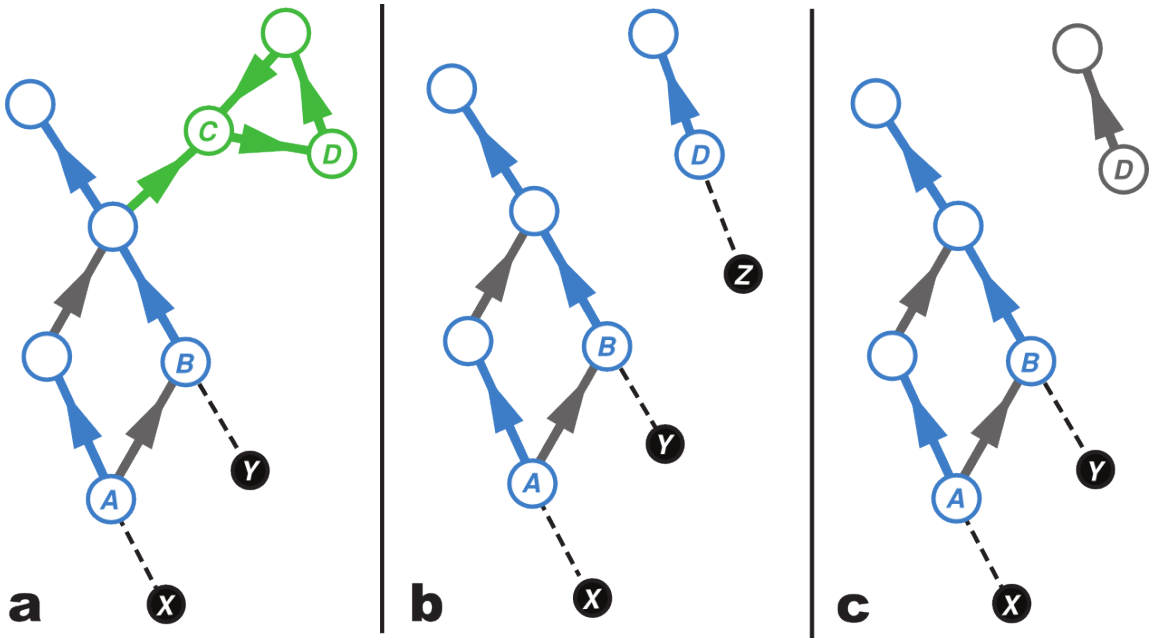


Figure 3–1: **(a)** An example network G with driver nodes A and B. Stems are highlighted in blue, cycles in green **(b)** N_c increases by 1 as node C is removed and cycle is broken into a stem which requires a new control Z **(c)** N_r decreases by 3 as node C is removed and stem starting with D becomes uncontrollable

controls (X and Y) as before percolation. The value of N_r reduces to 5 since the stem starting with node D is no longer reachable using controls X and Y.

In order to calculate N_r given controls X and Y, Hosoe's theorem can be used [8]. Let us recall the linear system given in Equation 2.1 as a graph $G(A, B)$. The generic dimension of the controllability matrix C (Equation 2.2) is defined as

$$rank(C) = \max_{G_{sub} \in G} |E(G_{sub})| \tag{3.1}$$

where G_{sub} is the set of all stem/cycle disjoint subgraphs of the $G(A, B)$ that are reachable from controls B and $|E(G_{sub})|$ is the number of edges in the subgraph G_{sub}.

Algorithm 1 Algorithm to find Cacti for fixed controls

Require: Network G, Control node set C (G includes nodes in C and edges to driver nodes)

Ensure: Cacti representing control structure

$G' \leftarrow G - x$, x not reachable from C // using Depth First Search
Create a Bipartite Graph G_B with:
    $\triangleright$ $2|V(G')|$ nodes, a pair of +ve and -ve nodes for each node in G'
for all edge (u, v) in G' **do**
    Add an edge (u^+, v^-) to G_B with weight 1
end for
for all control node c in C **do**
    $d \leftarrow Neighbor(c)$ // d is driver node
    Add an edge (c^+, d^-) to G_B with weight 1
    for all node x in G' such that x is not in C **do**
        Add an edge (x^+, c^-) to G_B with weight 0
    end for
end for
for all node u in G' **do**
    Add an edge (u^+, u^-) to G_B with weight 0 // self loop
end for
// Add large enough weight to make it a perfect matching
Add weight W to all edges in G_B such that $W > \sum weight(e), \forall e \in Edges(G_B)$
Perform weighted maximum matching algorithm on G_B to get a matching M
Map edges in M back to edges in G and join them to form stems and cycles of cacti.
Number of controllable nodes is number of matched nodes in cacti // control nodes in C are not matched

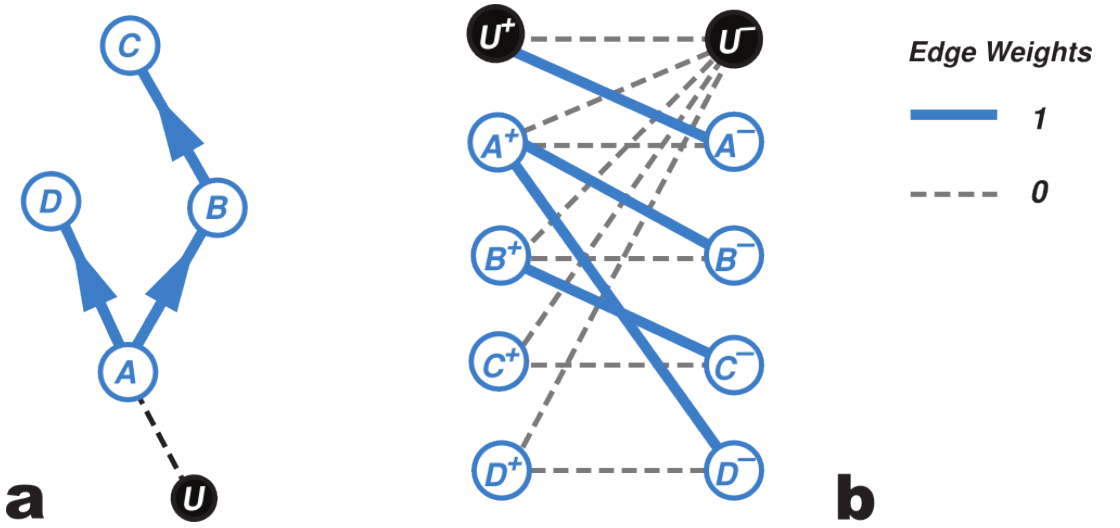


Figure 3–2: **(a)** An example network G with $N = 5$ nodes with external control node U attached to driver node A **(b)** Bipartite graph with $2N$ nodes, a pair of +ve and -ve nodes for each node in G.

Though the proof of the theorem is well presented [8], a algorithm to calculate the rank is missing. On the other hand, Poljak [19] gives a graph-theoretic proof of the theorem, which describes a method to calculate the rank in Equation 3.1 as well as to construct the cacti structure in the graph $G(A, B)$. However, the solution presented therein requires solving an integer linear program, which is computationally more expensive than our method and requires sophisticated linear program solvers. Therefore, we convert the solution into one that involves finding perfect maximum-weighted matching in a bipartite graph created from $G(A, B)$. The solution is explained further in Algorithm 1 and Figure 3–2). Using Fibonacci Heap in the implementation, we have the time complexity of the algorithm to be $\mathcal{O}(NL + mN^2 + 4N^2 log 2N)$ where N and L are number of nodes and edges in $G(A, B)$ respectively and m is the

number of controls. It can be seen that for sparse graphs, the performance is better than $\mathcal{O}(N^3)$.

3.2 Network Models and Data

Our present study is focused on synthetic network models that have been described before. Because the formation mechanisms of these synthetic models are known to us, they offer an opportunity to establish connections between the robustness of controllability in networks that have features which are frequently observed in nature. One of the distinguishing features of this thesis over the current literature on the robustness of network controllability is our broader survey of synthetic network models. Previous work has established that while these models share some common topological statistics (e.g., scale-free degree distribution) they show significant differences in their control properties [22].

We focus on directed synthetic networks using the Erdős-Rényi (ER), Barabási-Albert (BA), local attachment (LA), and duplication-divergence (DD) models of network generation. In this study, all networks have $N = 1000$ nodes; average degrees $k = 2, 4, 6, 8, 10, 12$ were considered except for duplication divergence networks, in which the parameter of the model does not directly involve degrees of the nodes and hence the formation mechanisms yields networks with highly varied average degree. Worth noting is that other choices of parameters did not substantively affect any of the results reported here. Further, these average degree values represent already a highly conservative approximation of the average degrees seen in real world networks. In all results presented, each realization of network type and parameter

values is generated with 10 different instances to provide a notion of expected (average) behavior. Even with this rather small level of averaging, for $N = 1000$ the error bars are very small in most cases, underscoring that our results are stable and this approach is sufficient to observe the expected control properties of these networks. The network models are described below.

Erdős-Rényi (ER) Using random connection model described in [5], the random networks were generated until the number of edges (E) was within an acceptable tolerance: $|E - kN| < 0.001kN$. The homogeneity of ER networks typically leads to networks that have very few controls with relatively long stems and cycles.

Barabási-Albert (BA) These networks were generated using the preferential attachment model presented in [3]. BA networks are inherently acyclic, which limits the range of the effect that a control can have in the network. Therefore, BA networks are characterized by a large number of controls (typically due to source nodes) and short stems and cycles.

Local Attachment (LA) The networks were created using local attachment model [10], which has been decribed in the previous section. In this study, we chose clustering values as $c = 0, 0.25, 0.5, 0.75$. LA networks are also acyclic and tend to exhibit similar control characteristics to that of BA networks.

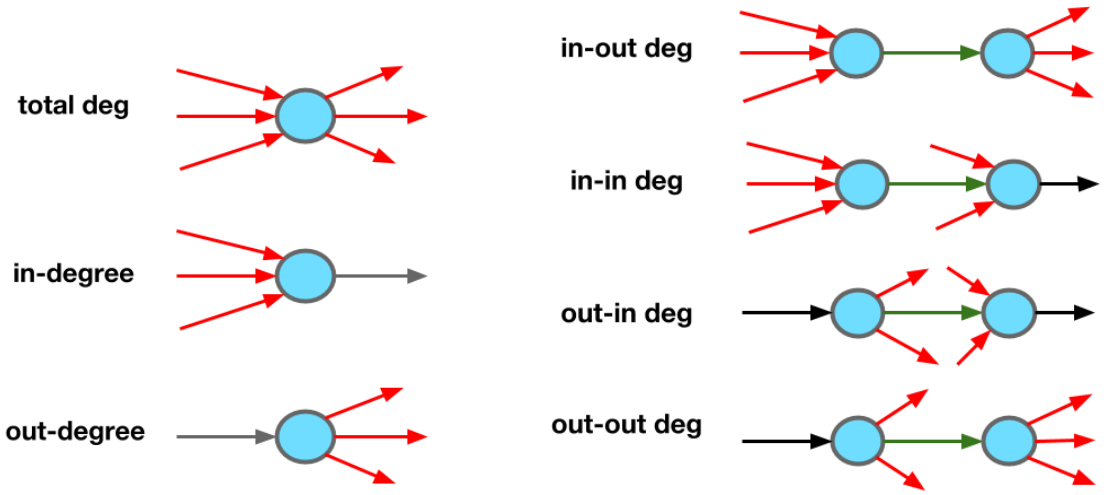
Duplication Divergence (DD) In duplication-divergence model, a node is duplicated and its edges are kept with a probability s [9]. Values for probability of duplication $s = 0.1, 0.3, 0.5, 0.7, 0.9$ were considered. DD networks have the most diverse control profiles of the synthetic networks surveyed here.

3.3 Types of Attacks

Another way we distinguish our work from other studies of robustness of network controllability is through the inclusion of a variety of node and edge attack types. While attacker models could potentially consider any level of knowledge about the system, we focus on the more realistic scenario where attackers have access to the most fundamental structural information given by the degree of the nodes in the network. Thus, as with other studies, we consider attacks that leverage degree information to identify critical nodes. Pósfai et al. [20] explored the role that various relative degree relationships determine properties of network controllability. With this as guidance, we consider such relative degree relationships in our study of robustness of network controllability.

For targeted node attacks, we select the node to be attacked based on its in-degree, out-degree, or total degree (Figure 3–3(a)). A node may be important, in the context of control, with high in-degree because it is a node through which many potential paths may pass or with high out-degree because it has the potential to propagate the influence of a control to many neighbors. Through this investigation we begin to shed light on the relative importance of these factors.

With regard to edge attacks, we consider the degree information of the source (s) and target (t) nodes of the edge. We can, therefore, consider all four combinations of looking at in- and out-degree information for both of these nodes: in-in, in-out, out-in, out-out, and total degree relationships. These combinations explore the extent to which the edge is important due to being a funnel (in-in), being a source (out-out), being a bridge (in-out), or other such functions. We considered five edge attacks



(a) Node attacks. The node is selected for removal based its degree (shown as red edges).

(b) Edge attacks. Green edge is selected for removal based on sum of degrees of its nodes (shown as red edges).

Figure 3–3: Types of attacks.

(Figure 3–3(b)) which are functions of the degree of s and t. Edges were selected in descending order of the score returned by a given function.

Finally, because we are interested in the contrast between random failures and targeted attacks, we also evaluate a random node and edge percolation (*attack*).

CHAPTER 4
Results

In this section we summarize how we generated results that show effects of node and edge attacks on reachability- and control-based robustness for different networks. We explain the plots involved and highlight key observations that can be made from plots.

While the node- and edge-based attacks differ to some extent, the approach to assessing robustness (both numerically and visually) was consistent across network models and attack types. As seen in Figures 4–1 and 4–2, we take the fixed set of control inputs to be a minimum control set required to control the original (unperturbed) network. Because the minimum controls guarantee complete controllability, $n_r = N_r/N = 1$ before the percolation process begins. With each step, 5% of nodes/edges are removed up to a total of 50% percolation. As opposed to Pu et al. [21], we do not keep as node in network after it is removed (or fully disconnected as is the case when edges are removed), so that the change in N_r reflects also the change in network size. Because the number and location of the controls cannot change, when the nodes they directly connect to are removed, that control will lose it's connection to the network and no longer be able to contribute to controlling the networks. Figures 4–1 and 4–2 show the change in the fraction of controllable nodes ($n_r = N_r/N$) for different networks with respect to node/edge percolation with different attacks.

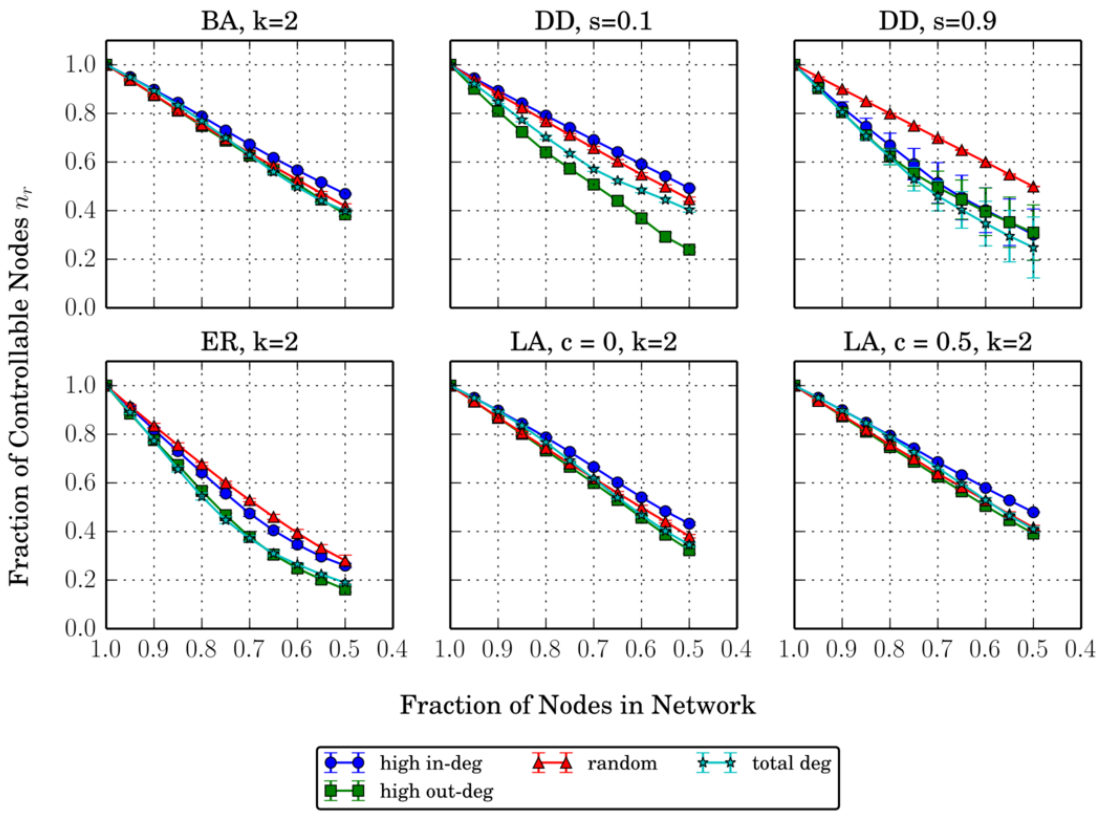


Figure 4–1: The robustness of control structures in the four network models to degree node-based attacks. Only the results for one parameter choice are shown — these are representative of all other parameter choices.

4.1 Connected Components and Stem Lengths

In order to understand the behavior of attacks on robustness we analyzed how the number of strongly-connected components N_{SCC} and average stem length vary under particular edge attacks (see Figure 4–3). Since BA and LA networks are acyclic, N_{SCC} remains constant at 1000 hence not shown in the figures. While for ER networks, we can see that the effectiveness of an attack is correlated with the an increase in N_{SCC}. Also there is a strong correlation between change in average stem

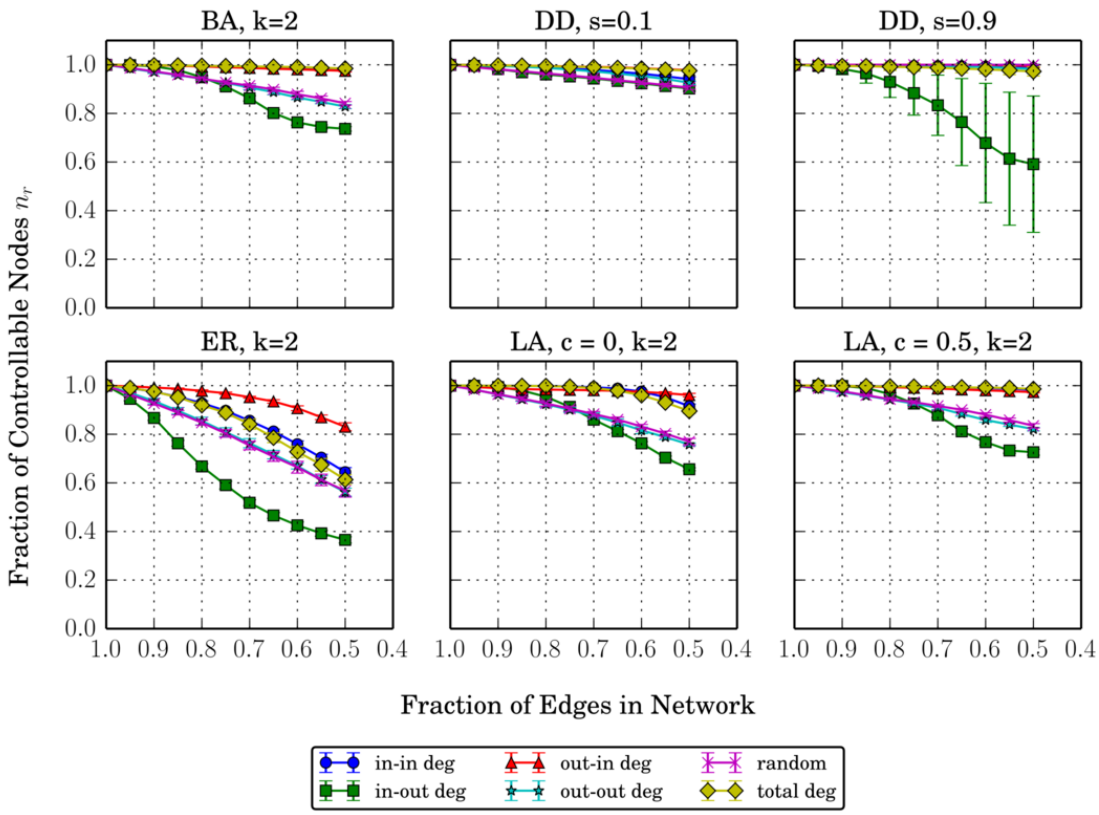


Figure 4–2: The robustness of control structures in the four network models to degree edge-based attacks. Only the results for one parameter choice are shown — these are representative of all other parameter choices.

length shown in Figure 4–3(b) and controllability plots in Figure 4–2. For example, the in-out attack which tends to be the most effective also tends to create shorter stems on average after percolation.

4.2 Initial Observations

Node attacks. There are few important observations that can be inferred from the plots in Figure 4–1. Unlike [21], we find that degree-based attacks are not always more effective (more damaging) than random attack. For example, in the case of BA

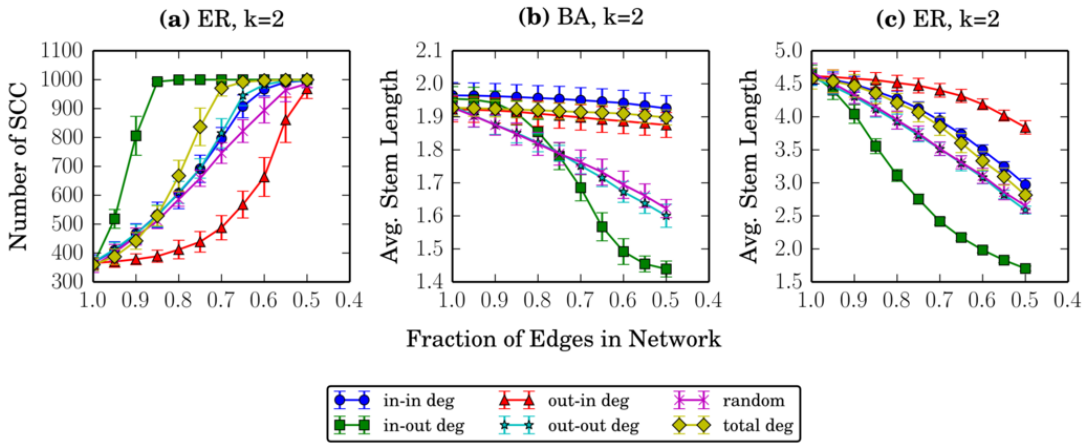


Figure 4–3: Variation in number of strongly connected components and average stem length under edge attacks

and LA networks, random attack does nearly the same as the most effective, high out-degree, attack. In the case of DD and ER networks random attack is the least effective. We also find significant variations among different types of degree-based attacks. High out-degree attacks stands out as being the most effective in most of the networks, while high in-degree and total degree attacks show considerable difference in effectiveness across network types. These differences underscore the importance of evaluating various metrics for robustness of network controllability.

Edge attacks. As seen in Figure 4–2, the in-out degree attack initially starts out being less effective than a random attack but after a few steps, it rapidly degrades controllability. This effect can be best explained by observing the change in the number of strongly connected components (SCC) as well as the change in average stem/cycle lengths (described next). It is observed that in-out degree attack rapidly creates a larger number of SCCs than other attacks and also produces stems/cycles of very short lengths. Also noteworthy is that out-in degree and total degree are

almost always least effective, while random attack maintains average effectiveness. This further confirms our proposition that degree-based attacks exhibit varying effectiveness relative to random attack for different networks and suggesting the role of more nuanced network features in the phenomena.

CHAPTER 5
Discussion

In this section we discuss important properties of reachability-based robustness such as its relation to control-based robustness. We also analyse the robustness measure as a function of network and attack types.

5.1 Robustness definitions have different behavior

Figures 5–1(a) and (b) show how the number of controllable nodes, N_r (or fraction $n_r = N_r/N$), decreases and the number of minimum controls, N_c, increases for BA and ER networks under edge percolation. In order to verify whether changes in N_r tell us a different story about robustness of networks than changes in N_c, in Figure 5–1(a) and (b) we also plotted $\Delta N_r(t) - \Delta N_c(t)$ for each percolation step t, where $\Delta N_r(t) = N_r(0) - N_r(t)$ and $\Delta N_c(t) = N_c(t) - N_r(0)$. The intuition here is to capture increase in N_c after percolation and compare it with decrease in N_r. We can see that in case of BA network, difference between the decrease in N_r and the increase in N_c for some attacks is almost negligible. While for more effective attacks such as in-out degree attack, there is a greater decrease in N_r for a given increase in N_c. This gap between N_r and N_c widens greatly in case of ER network. This strongly suggests that N_r (reachability) based measures reveal different aspects of network vulnerability to particular node- and edge-based attacks compared with measures using N_c. In fact, as opposed to results by Pu et al. [21], we find that ER networks are

less robust than BA networks. Therefore, our definition of robustness indeed presents a behaviorally different and significant point of view for understanding robustness.

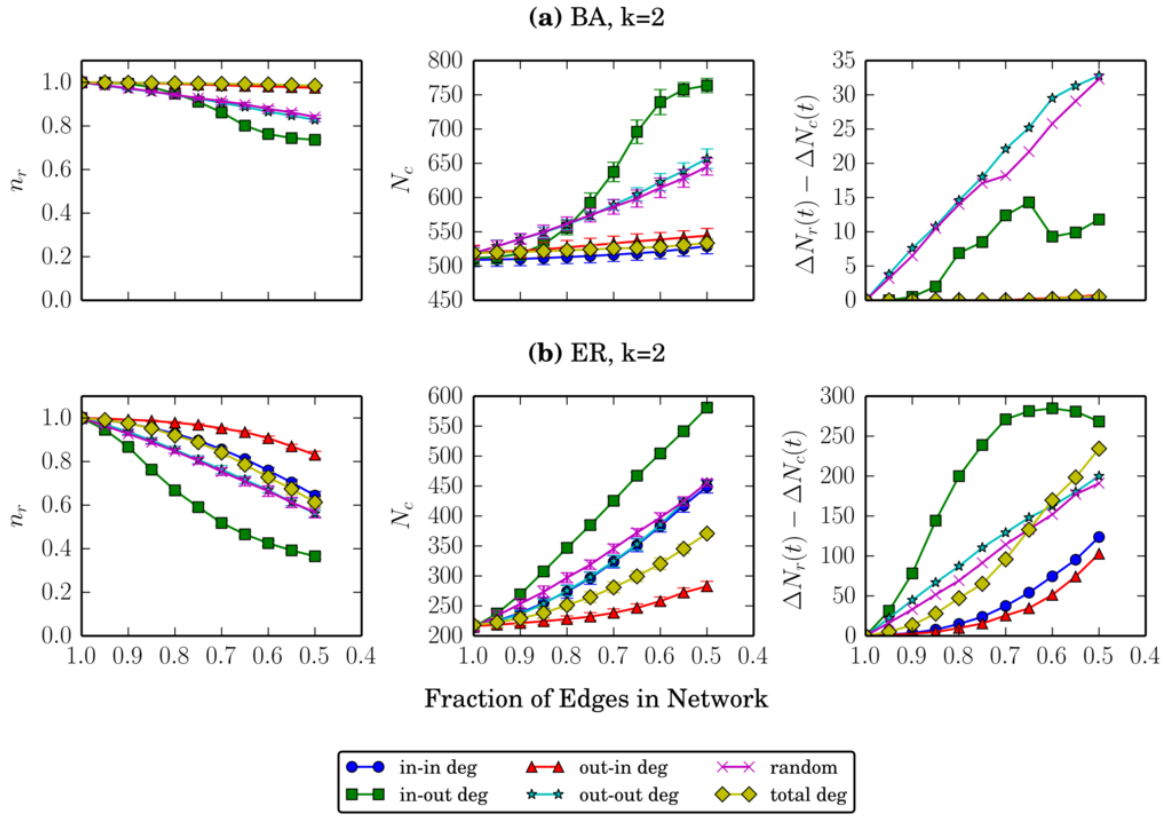


Figure 5–1: Comparison showing Reachability- and Control-based robustness measures for ER an BA networks. $n_r = N_r/N$ is fraction of controllable nodes and N_c is minimum number of controls required for full controllability

5.2 Robustness dependence on network types

Figure 5–2 highlights the extent to which different network types will have different responses to various attacks. The fact that the robustness signatures differ indicates that clustering, degree homogeneity, the presence of cycles, and other

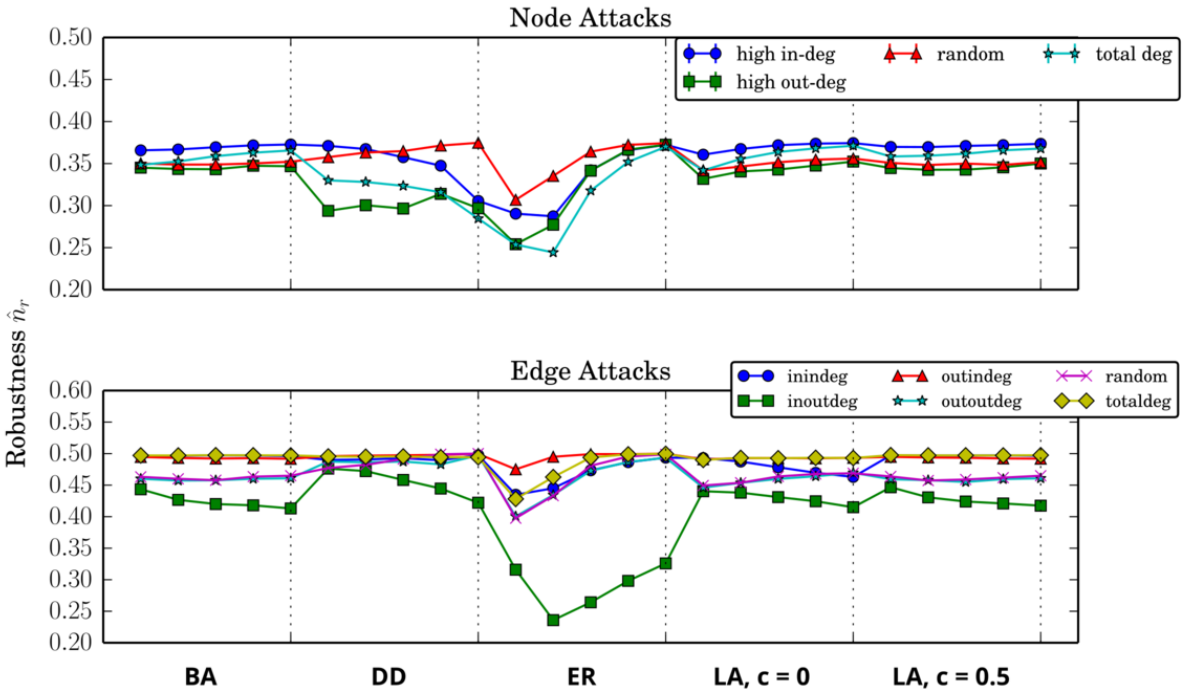


Figure 5–2: Reachability-based robustness for all networks grouped by type. Five data points, each for an average degree k=2,4,6,8,10 (or probability s=0.1, 0.3, 0.5, 0.7, 0.9 in case of DD) is shown for each network type.

higher-order network features may impact a particular network's vulnerability to attack. While an investigation into the nature of these signature differences is beyond the scope of this thesis, we can highlight a number of intriguing trends which deserve attention in future work.

- *Clustering does not affect robustness to node attacks.* Notice that BA networks and LA networks with high clustering tend to exhibit similar behavior for all

node attacks. LA networks have both a scale-free degree distribution and clustering - thus the similarity in behavior suggests that the addition of clustering does not affect the overall vulnerability to node attacks.

- *Degree distribution.* Even though both LA with zero clustering and ER are both random models, the former creates acyclic networks and exhibits greater resistance to attacks than ER which is more homogeneous than LA and consists of cycles.
- *Average degree sometimes can play a role.* Our analysis indicates that, under node attacks, $\hat{N}_r$ does not show much variability across different average degrees except for the case of ER networks. In the case of ER networks, robustness increases with average degree (likely due to presence of more edges). The effect of clustering in LA networks is also negligible for node attacks. $\hat{N}_r$ shows greater variation across attack types in DD and ER networks than in LA and BA networks.

  Similarly for edge attacks, there is only slight change in $\hat{N}_r$ with increasing average degree, except for the case of ER networks. It is surprising to find that the in-out degree attack in fact is more effective with increase in average degree. In-out degree attack is systematically the most effective attack for all networks, as well as across different average degrees. It is also interesting to note that duplication divergence (DD) networks are significantly more robust than other network types against edge attacks. This suggests that biological protein networks might have evolved to favor stability of protein-protein interactions.

5.3 Robustness dependence on attack types

While this point has already been indirectly explored above, it is worth highlighting that Figure 5–2 (as well as the other results figures) demonstrate the variable effect a given attack can have. Interestingly, in some instances, node attacks have more-or-less equivalent robustness scores; whereas in others, the same attacks can have very different and highly variable robustness scores. This perspective suggests that, in addition to understanding how particular network structures achieve differential degrees of robustness; another fruitful approach might consider the means by which different attacks achieve similar (or different) robustness levels across a wide array of networks.

CHAPTER 6
Conclusion

The robustness of control structures will be an important consideration when applying theory governing the control of complex systems. In this thesis we have proposed a new measure of robustness which we consider to capture many of the constraints that arise when controlling real-world systems. We find that this measure functionally differs from existing measures in the literature and that, when subjected to a variety of node- and edge-based attacks, yields trends that suggest ways in which network properties relate to the robustness of control structures. We show that our measure gives not just a practical but a necessary view of robustness. A novel and efficient method to evaluate the robustness measure is presented in this thesis. We strongly believe that the important results we obtained, using a wide array of network models and attack types in our study, will help in robust network design.

Our method and results are directly useful in at least two following ways. First, when resources available to control a network are limited and a full controllability can't be achieved, it is often desirable to select an allocation of controls from a set of feasible allocations such that controllability is maximized. Our method to calculate controllability given an a set of controls can be used effectively in this case. Second, our results can help to choose a synthetic network model for any practical system by comparing different models for their robustness against particular attacks that the system is likely to encounter. Furthermore, the specific model parameters, such as

average degree for example, of such networks can also be tuned for robustness based on patterns observed in the results.

One of our future goals is to continue our study and include real networks in our analysis of robustness. Our promising preliminary results suggest an interesting pattern of robustness among real networks. Our next future goal is to study a slightly different measure of robustness where control inputs are allowed to move to different nodes but their number is kept fixed. In practical scenarios, we may find that after a node or link failure, if control inputs are allowed to be reassigned, controllability can be increased. We are currently working on methods to calculate controllability given a fixed number of movable control inputs. We hope to find interesting results to compare against our existing measure of robustness.

References

[1] Réka Albert and Albert-László Barabási. Statistical mechanics of complex networks. *Reviews of modern physics*, 74(1):47, 2002.

[2] Réka Albert, Hawoong Jeong, and Albert-László Barabási. Error and attack tolerance of complex networks. *Nature*, 406(6794):378–382, 2000.

[3] Albert-László Barabási and Réka Albert. Emergence of scaling in random networks. *science*, 286(5439):509–512, 1999.

[4] Stephen P Borgatti, Ajay Mehra, Daniel J Brass, and Giuseppe Labianca. Network analysis in the social sciences. *science*, 323(5916):892–895, 2009.

[5] P. Erdős and A. Rényi. On Random Graphs, I. *Publicationes Mathematicae*, 6: 290–297, 1959.

[6] P. Erdős and A Rényi. On the evolution of random graphs. In *Publication of the Mathematical Institute of the Hungarian Academy of Sciences*, pages 17–61, 1960.

[7] John E Hopcroft and Richard M Karp. An n^5/2 algorithm for maximum matchings in bipartite graphs. *SIAM Journal on computing*, 2(4):225–231, 1973.

[8] Shigeyuki Hosoe. Determination of generic dimensions of controllable subspaces and its application. *Automatic Control, IEEE Transactions on*, 25(6):1192–1196, 1980.

[9] Iaroslav Ispolatov, PL Krapivsky, and A Yuryev. Duplication-divergence model of protein interaction network. *Physical review E*, 71(6):061911, 2005.

[10] Matthew O Jackson and Brian W Rogers. Meeting strangers and friends of friends: How random are social networks? *The American economic review*, pages 890–915, 2007.

[11] Tao Jia, Yang-Yu Liu, Endre Csóka, Márton Pósfai, Jean-Jacques Slotine, and Albert-László Barabási. Emergence of bimodality in controlling complex networks. *Nature communications*, 4, 2013.

[12] Rudolf Emil Kalman. Mathematical description of linear dynamical systems. *Journal of the Society for Industrial & Applied Mathematics, Series A: Control*, 1(2):152–192, 1963.

[13] Ching-Tai Lin. Structural controllability. *Automatic Control, IEEE Transactions on*, 19(3):201–208, 1974.

[14] Yang-Yu Liu, Jean-Jacques Slotine, and Albert-László Barabási. Controllability of complex networks. *Nature*, 473(7346):167–173, 2011.

[15] Ron Milo, Shai Shen-Orr, Shalev Itzkovitz, Nadav Kashtan, Dmitri Chklovskii, and Uri Alon. Network motifs: simple building blocks of complex networks. *Science*, 298(5594):824–827, 2002.

[16] Mark Newman, Albert-László Barabási, and Duncan J Watts. *The structure and dynamics of networks*. Princeton University Press, 2006.

[17] Sen Nie, Xuwen Wang, Haifeng Zhang, Qilang Li, and Binghong Wang. Robustness of controllability for networks based on edge-attack. *PloS one*, 9(2): e89066, 2014.

[18] Deven Parekh, Derek Ruths, and Justin Ruths. Reachability-based robustness of network controllability under node and edge attacks. In *Signal-Image Technology and Internet-Based Systems (SITIS), 2014 Tenth International Conference on*, pages 424–431. IEEE, 2014.

[19] Svatopluk Poljak. On the generic dimension of controllable subspaces. *Automatic Control, IEEE Transactions on*, 35(3):367–369, 1990.

[20] Márton Pósfai, Yang-Yu Liu, Jean-Jacques Slotine, and Albert-László Barabási. Effect of correlations on network controllability. *Scientific reports*, 3, 2013.

[21] Cun-Lai Pu, Wen-Jiang Pei, and Andrew Michaelson. Robustness analysis of network controllability. *Physica A: Statistical Mechanics and its Applications*, 391(18):4420–4425, 2012.

[22] Justin Ruths and Derek Ruths. Control profiles of complex networks. *Science*, 343(6177):1373–1376, 2014.

[23] Zhengzhong Yuan, Chen Zhao, Zengru Di, Wen-Xu Wang, and Ying-Cheng Lai. Exact controllability of complex networks. *Nature communications*, 4, 2013.