Effect of Network Topology on Performance of Multi-Agent Consensus in Adversarial Conditions

UNDERGRADUATE THESIS

Authored By
Neelaksh Singh

Prepared in partial fulfillment of the requirements of BITS F421T Thesis.

Supervisor
Dr. Vaibhav Katewa
Dept. of Electrical Communication Engineering
Indian Institute of Science, Bangalore

Co-Supervisor
Dr. Bijoy Krishna Mukherjee
Dept. of Electrical and Electronics Engineering
BITS Pilani - Pilani Campus

Birla Institute of Technology and Science Pilani - Pilani Campus
Department of Electrical and Electronics Engineering
December, 2021

THESIS

Effect of Network Topology on Performance of Multi-Agent Consensus in Adversarial Conditions

Undergraduate Thesis

Submitted in partial fulfillment of the requirements of
BITS F421T Thesis

By

Neelaksh Singh
ID No. 2018A3TS0337P

Under the supervision of:

Dr. Vaibhav Katewa
&
Dr. Bijoy K. Mukherjee

BIRLA INSTITUTE OF TECHNOLOGY AND SCIENCE PILANI, PILANI CAMPUS
December, 2024

Declaration of Authorship

I, Neelaksh Singh, declare that this Undergraduate Thesis titled, 'Effect of Network Topology on Performance of Multi-Agent Consensus in Adversarial Conditions' and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself

Signature

Date: December 14, 2024

Certificate

This is to certify that the thesis entitled, "*Effect of Network Topology on Performance of Multi-Agent Consensus in Adversarial Conditions*" and submitted by Neelaksh Singh ID No. 2018A3TS0337P in partial fulfillment of the requirements of BITS F421T Thesis embodies the work done by him under my supervision.

Supervisor
Dr. Vaibhav Katewa
Assistant Professor,
Indian Institute of Science, Bangalore
Date: December 14, 2024

Co-Supervisor
Dr. Bijoy K. Mukherjee
Assistant Professor,
BITS-Pilani, Pilani Campus
Date: December 14, 2024

"And once the storm is over, you won't remember how you made it through, how you managed to survive. You won't even be sure, whether the storm is really over. But one thing is certain. When you come out of the storm, you won't be the same person who walked in. That's what this storm's all about."

Haruki Murakami, Kafka on the Shore

BIRLA INSTITUTE OF TECHNOLOGY AND SCIENCE PILANI, PILANI CAMPUS

Abstract

Bachelor of Engineering

Effect of Network Topology on Performance of Multi-Agent Consensus in Adversarial Conditions

by Neelaksh SINGH

It has been proved in the literature that Cyber Physical Systems (CPS) are vulnerable to attacks. So much so that this has propelled CPS security as a major domain of interest in control systems research. Various practical attack models have been developed alongside algorithms which aim towards safeguarding the systems against the modelled attacks.

However, research in communication systems have shown that network topology plays an imperative role in creating secure systems, with some topologies being more vulnerable than others by design. Yet research in multi-agent control systems has focused majorly on the algorithmic development side of security. The effect of the multi-agent system's design, such as its communication network topology, on its overall resilience against adversarial attacks is yet to be explored. This thesis focuses on understanding the role of network topology on performance of multi-agent consensus in single-integrator systems under an attack. To this end, a heuristic function is derived to estimate security, i.e. performance under attack, which depends on topological connectivity properties. A general Lyapunov function for single-integrator consensus on arbitrary digraphs is also developed to assess convergence performance for all connectivity levels. The Lyapunov function is found to be valid for the set of digraphs of arbitrary connectivity given that their graph laplacian is a non-defective matrix. Since the set of defective matrices has very small dimensions compared to that of non-defective matrices, the analysis is applicable for a large class of digraphs. Such a generalized result is being presented for the first time in this thesis to the best of the author's knowledge.

Finally future directions in topology design based on rigorous simulation of the random DoS attack and exhibited patterns are presented. It is concluded experimentally that in the case of bounded adjacency matrix norm $\|A\|_F$, topologies with more but weaker links are, on an average, more secure than ones with lesser but stronger links. This suggests the existence of a large class of topologies which exhibit performance versus security trade-off patterns.

Acknowledgements

I will take this opportunity to thank my supervisor Prof. Vaibhav Katewa for introducing me to the world of multi-agent systems. I love robotics and have mostly done practical projects until now. Except, this time it was very theoretical. The freedom he allowed me on this project has taught me priceless lessons about research. Searching for the right direction to work in, formulating the problem, taking important decisions and setting my own goals and milestones which are all refined by his feedback; I felt myself grow as a researcher more than I ever had during the other projects.

I would also like to thank Prof. Bijoy Krishna Mukherjee, my supervisor at BITS Pilani - Pilani Campus for his never ending support. The entire process went as smooth as a polished pearl, which wouldn't be possible without his support.

Finally, special thanks to my family and friends who gave me clarity through times of brain fog, and taught me equanimity in adversity. The pandemic's difficulties led to various ups and downs. It was because of their constant support that this thesis sees completion today.

Contents

List of Figures

Dedicated to my family.

Chapter 1

Introduction

Humans always live in the fast lane. With every passing year, the demand for making processes like manufacturing, mobility, farming utilities, consumer electronics, and many more, faster than their current state increases. These demands are satisfied with cutting edge innovations in control systems and automation, new milestones are created and achieved, advancing humanity to the next level in automation technology. One such milestone led to the inception of Cyber Physical Systems (CPS), and the world witnessed the rise of applications like Smart-Grids, Avionics, Industrial Control, etc [1]. Researchers around the world are working on making them a tangible reality due to the plethora of benefits they bring with their adoption.

Unfortunately, there is a caveat. CPS rely heavily on networking and communication and are; therefore, vulnerable to attacks [2]. This has propelled security in CPS as a major research area in the last decade. This has been accompanied with the development of attack models followed by development of algorithms which are robust against them.

Towards attack development, authors in [3] developed various deterministic attack models and presented a practical taxonomy of such attacks. [4] demonstrated complicated byzantine attacks in multi-agent robotic systems. [5] presents an excellent overview of Denial of Service (DoS) attacks in CPS. Following the development of attack model, recent advances focus on resilient algorithm design and attack detection in CPS which has derived a lot of results from concepts in fault tolerant control systems theory. [6], [7] present holistic surveys in fault detection and attack detection. [8] is a seminal work in the domain of attack detection, where they present a mathematical framework for designing attack monitors for LTI nodes in a CPS. Research in attack detection was followed by development of attack resilient distributed control algorithms [9], [10]. Decentralized algorithm designs were popularized. Realizing that some attacks will deal a lot of damage despite protective algorithms, survivable CPS design paradigms were explored [11].

Despite making significant progress with respect to algorithm design as a layer of protection, a very significant aspect of CPS is still unexplored with respect to secure CPS design: the network topology. There is a significant dearth of literature in secure topology design. This inspired the author of this Thesis to explore the effect of network topology and its contribution to providing security against attacks. If the attack affects the links (link failure DoS), there must be a topology design which placates its impact. This has been demonstrated by research in communication topology design where researchers have derived metrics of network's vulnerability to attack, which inspired critical node detection in multi-agent networks [12]. For attack models which behave as contagions, research in financial networks provides an excellent basis [13] about how topological properties provide containment for different attack

strengths.

Therefore, it is imperative to identify topological properties which contribute to security. This is the main objective of this thesis. The final aim will be to derive important insights about the effect of topology on security by rigorous experiments analysis. To identify trends which may point towards the existence of a potential class of topologies which provide security against adversaries as their inherent property, to an extent. The adversary of choice is the random link disabling DoS attack, and the system is a multi-agent consensus system with single-integrator nodes. Existing results on topology design focus majorly on topology design for optimal consensus [14], [15]. Our objective is to find a heuristic to asses security as a function of security, which might propel research in optimal topology design for optimal consensus security, and to identify any potential trade-offs between nominal consensus performance and performance under attack which is hinted at through the intuitive example presented in the introduction to the next chapter.

This work is arranged among 4 chapters numbered from 2 to 5. Chapter 2 introduces multi-agent consensus and required preliminary concepts. The link disabling DoS attack's mathematical representation is developed which resembles matrix perturbations. Thus, important results from arbitrary matrix perturbations are presented towards the end. Chapter 3 establishes the final problem statement and assumptions about the attacker. Chapter 4 presents the novel general Lyapunov equation for arbitrary digraphs with non-defective graph laplacian, which is used to assess convergence performance under attack. Furthermore, Chapter 4 then builds upon perturbation theory of Sylvester Equation to derive the final security estimating heuristic for directed and undirected graphs. Finally, Chapter 5 presents the rigorous simulation results which were used as directional pointers for research throughout this Thesis and will present a lot of interesting problems to be explored in the domain of secure topology design in CPS for future research endeavours.

Chapter 2

Multi-Agent Consensus and Adversaries

Let's start off with an example which demonstrates the importance of analyzing performance in adversarial conditions and the importance of network topology in the same.

Imagine a group of drones which are inbound from source A to destination B. On the way, it is bound to encounter the adversary and somehow re-planning of paths is not possible, therefore, the system will for sure be subjected to attack. At the same time, the swarm is maybe carrying an important medicine which people at B need urgently. We immediately realize that a trade-off is induced between the vicinity to the attacker and the priority to reach the destination. Situations like these are not impossible, this scenario will be repeated over and over again during adversarial conflicts such as war.

The reader may point out that this is not the only trade-off possible; the author wholeheartedly agrees, because among the one discussed earlier there is a very important trade-off introduced by the topology between consensus performance and security. To explain using a simple scenario: a system of drones is well connected and they are exchanging data with each other over a weighted digraph. This system is subjected to a DoS attack which results in failure of some links while the system was executing a formation control algorithm with only a selected few receiving the global trajectory and the rest being held together through a formation consensus algorithm .Consider that a drone which was not receiving the direct global trajectory and was dependent on the ones which were had its links to the global receivers broken. This drone will most likely face a tracking error, therefore, there are two cases which are possible at this point:

1. The perturbed drone will send erroneous position data to its still connected neighbours which in as a result start generating erroneous control input themselves and then sending their erroneous data to their other neighbours. This makes a vicious cycle of error spreading, something similar to string instability in its structure.

2. The erroneous data will be transmitted to the neighbours but through links which have low weights (considering weights have a physical meaning such as link energy, more link energy means more energy required to disable them through DoS) and is eventually assimilated to zero. Here the system gets an error for some time but eventually restores itself when the DoS attack ends.

Clearly, the topology is playing an important role. The link weights have a direct effect on whether the performance and security of consensus algorithms clash with each other. The intuitive ideas presented here will help the reader to understand the various concepts such as potential performance v/s security trade-offs. The analysis presented in this thesis is a simplified but general one which will require some preliminary concepts from distributed

control systems theory and matrix perturbation theory. These concepts are presented in this chapter before the problem formulation and analysis is presented in detail.

2.1 Consensus in Multi-Agent Systems

Matrix perturbation theory deals with the study of changes in the spectrum of a matrix $A \in \mathbb{C}^{n\times n}$ is perturbed by another matrix B belonging to the same space as A. Eigenvalues define the natural and forced modes of any dynamical system and therefore are integral to the convergence analysis of the system. This section starts off by defining the multi-agent system, its dynamics and the consensus control law. This will be followed by defining the basic Denial-of-Service (DoS) attack finally concluding with the introduction of matrix perturbations in the consensus.

2.1.1 Consensus Problem on General DiGraph

Consider a multi-agent system defined by the *weighted digraph* $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{A})$ of order n where $\mathcal{V} = \{v_1, v_2, \ldots, v_n\}$ is the set of vertices, $\mathcal{I} = \{1, 2, \ldots, n\}$ is the node index set, $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ is the set of edges and $\mathcal{A} \in \mathbb{R}^{n\times n}_{\geq 0}$ is the adjacency matrix defined by $\mathcal{A} = [a_{ij}]$ with strictly non-negative weights $a_{ij} \in \mathbb{R}^{n\times n}_{\geq 0}$. An edge in $\mathcal{G}$ is denoted by $e_{ij} = (v_i, v_j),\ i, j \in \mathcal{I}$. Clearly, presence of an edge denotes a non-zero weight i.e. $e_{ij} \in \mathcal{E} \iff a_{ij} > 0$. The sets of in-neighbours $\mathcal{N}_i^{in}$ and out-neighbours $\mathcal{N}_i^{out}$ of node i are defined as follows:

$$\mathcal{N}_i^{in} = \{v_j \in \mathcal{V} \mid (v_j, v_i) \in \mathcal{E}\} \tag{2.1}$$

$$\mathcal{N}_i^{out} = \{v_j \in \mathcal{V} \mid (v_i, v_j) \in \mathcal{E}\} \tag{2.2}$$

Every node in the digraph represents a dynamic agent which follows the *agent dynamics* defined by:

$$\dot{x}_i = f_i(x_i, u_i), \quad i \in \mathcal{I} \tag{2.3}$$

Defining $x = [x_1, x_2, \ldots, x_n]^T$ as the system state and $u = [u_1, u_2, \ldots, u_n]^T$. Then, using the terminologies in [23] we define that a *dynamic graph* is a system where $\mathcal{G}_x = (\mathcal{G}, x)$ defines a *network* with *network state* $\mathcal{G}_x$ which evolves according to the *network dynamics* given by $\dot{x} = F(x, u)$. For the sake of simplicity we will consider a system with single integrator dynamics. The dynamics of agent i is then defined as:

$$\dot{x}_i = u_i \tag{2.4}$$

We consider the simple *consensus protocol* defined as:

$$u_i = \sum_{j=1}^{n} [a_{ij}(x_i - x_j)] \tag{2.5}$$

In steady state, the protocol 2.5 ensures $x_i \to x_j, \forall e_{ij} \in \mathcal{E}$, this leads to the result:

Theorem 2.1.1 (necessary and sufficient condition for consensus)**.** *The multi-agent system on topology $\mathcal{G}$ following agent dynamics 2.4, the protocol 2.5 achieves consensus at steady state if and only if $\mathcal{G}$ contains a spanning tree.*

The proof for theorem 1.1.1 is omitted for brevity. For the system state $x = [x_1, x_2, \ldots, x_n]^T$ the protocol 2.5 is simplified to the following compact form:

$$\dot{x} = -Lx \tag{2.6}$$

2.1.2 The Graph Laplacian L

The matrix L in (2.6) is known as the graph laplacian for the graph $\mathcal{G}$ and adjacency matrix A and is defined as:

$$\begin{aligned}
\mathcal{L}(A) &= L = D_{out} - A \\
D_{out} &= [d_{ij}] \\
d_{ij} &= \begin{cases} \sum\limits_{j \in \mathcal{N}_i^{out}} (a_{ij}), & if\ j = i \\ 0 & if\ j \neq i \end{cases} \\
L &= [l_{ij}] \\
l_{ij} &= \begin{cases} \sum\limits_{\substack{k=1 \\ k \neq i}}^{n} (a_{ik}) & j = i \\ -a_{ij} & j \neq i \end{cases}
\end{aligned} \tag{2.7}$$

The reader is referred to [23] for standard results on the graph laplacian. The important results which will be required in the upcoming sections are:

- Every eigenvalue of the graph laplacian has a positive real part, which can be proved using the Gersgorin disks theorem. The result that the eigenvalues can never be purely imaginary follow from the same proof.
- 0 is a simple eigenvalue of the laplacian of strongly connected digraphs. In this case it was proved in [23] that the rate of exponential convergence is bounded above by the second smallest nonzero eigenvalue of L called the *Feidler eigenvalue* [25]. For $Eig(L) = \{\lambda_1, \lambda_2, \ldots, \lambda_n\}$ such that $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$ the Feidler eigenvalue is λ_{m+1} is the algebraic multiplicity of 0 is m. A result related to this was proved in [23]

The case, however, is significantly different for digraphs which are not strongly connected. A substantially detailed analysis of the convergence of weakly connected digraphs is presented in [24]. To summarize briefly:

- The connected components $\mathcal{S}$ which don't receive data from any node outside ($v \in \mathcal{V} \setminus \mathcal{S}$) and only have outgoing edges are referred to as *authorities*. And the nodes which receive data from the authorities and share edges among them are called *intermediate nodes*.
- The nodes of a authorities will reach consensus with each other, however, different authorities may or may not settle at the same value, therefore, consensus between authorities is not necessary.
- The intermediate nodes will converge at an intermediate value to the authorities they are connected to. Usually a consensus is reached between the intermediate nodes and their steady state value can be expressed as a weighted average of the steady state values of the authorities.
- $rank(L) \leq n - m$ for m authorities and a digraph of order n.

The information stated above is sufficient to proceed with the discussion for the time being. Additional results will be stated and discussed as and when the need arises. Now that we have a grasp over the basic concepts in consensus and graph theory we will define the DoS attack model.

2.2 The Link Failure DoS Attack

DoS (Denial of Service) attacks are one of the most commonly studied attacks used for assessing the security of and designing attack resilient distributed algorithms [10], [11], [28]. [3] presents a model for all of the commonly encountered attacks in secure CPS (Cyber Physical Systems) literature. An important question then is, if a system is attacked, how can the system identify that it is under attack. Identification is an important first step to deploying a robust algorithm which may vary according to the attack model. The answer to the problem of attack detection and identification ([26]) is well researched one, owing to its similarities to the fault detection and identification, a problem studied in depth in the fault tolerant control literature [9], [10], [27]. Since DoS attacks are the simplest to model, the analysis of the trade-offs presented here will use DoS attacks for the sake of simplification.

As discussed in the introduction of this chapter, there are various scenarios where a well connected topology, although superior in performance, poses a greater threat when attacked, such cases mostly comprise of attacks which rely on the spread of malicious data through a host node. The higher the connectivity, the better is the host nodes' scope of virus spread. *The spread of malicious data might not be a deliberate effect of the attack itself but might be a consequence of the attacked node sending its perturbed data to its neighbors.* In fact, it can be argued that this is the case for almost every attack because perturbations are all disturbances that try to take the system away from the desired trajectory. In this sense, the author argues that *resilient algorithms aside, topology is imperative to safety.* This fact has not been analyzed in detail till date.

Therefore, to prove the point, we refer to the intuitive idea presented in the introduction to this chapter that link failures lead to a spread in perturbations. Following this logic, it makes sense to use the simplified model of link failure DoS attacks to analyze the effect of topology on performance in consideration with nominal performance.

2.2.1 Mathematical Representation of The Attack

Link failure means link removal. In the most simplified sense, a DoS attack on the edge e_{kl} connecting the nodes v_k and v_l is the removal of the edge to block any data exchange between v_k and v_l. Clearly, this means $a_{kl} = 0$. Consider the set of edges attacked $\mathcal{E}_{dos} = \{e_{k1}, e_{k2}, \ldots, e_{kn}\} \subseteq \mathcal{E}$ where the set of indices $\mathcal{I}_{dos} = \{k_1, k_2, \ldots, k_n\} \subseteq \mathcal{I} \times \mathcal{I}$. Then the attack matrix E is defined as follows:

$$E = E_{ij} = \begin{cases} a_{ij} & if (i,j) \in \mathcal{I}_{dos} \\ 0 & otherwise \end{cases} \tag{2.8}$$

In this case, denoting the quantities under attack with a tilde ($\tilde{}$), the attacked adjacency matrix $\tilde{A}$ is calculated as:

$$\tilde{A} = A - E \tag{2.9}$$

It is clear that E represents a digraph $\tilde{\mathcal{G}}$ which is a sub-graph of the digraph defined by A, i.e. $\mathcal{G}_{dos} \subseteq \mathcal{G}$. The resultant graph due to the attack ($\tilde{A}$) is $\tilde{\mathcal{G}} \subseteq \mathcal{G}$ with graph laplacian $\tilde{L}$. Then, denoting the graph laplacian of E as $L_E = L_{dos} = \mathcal{L}(E)$, the following result follows with a straightforward proof (omitted for brevity) :

$$\tilde{L} = \mathcal{L}(A) = L - L_E \tag{2.10}$$

Assumption 1. The attacker has limited attack resources which can be represented as the bounded attack matrix norm $\|E\| \leq \epsilon$. The norm can be any standard vector induced matrix norm defined as:

$$\|A\| = \frac{\|Ax\|}{\|x\|} \tag{2.11}$$

Note that, in specific cases one may want to use a *unitarily invariant norm.*

Definition 2.2.1 (Unitarily invariant norms)**.** Denoted as $|||A|||$, unitarily invariant norms are matrix norms which are invariant to multiplication with any two unitary matrices U and V such that $|||UAV||| = |||A|||$.

A famous example of unitarily invariant norms is the Frobenious norm defined by:

$$\|A\|_F = \sqrt{\sum_{i=1}^{n}\sum_{j=1}^{n}|a_{ij}|^2} \tag{2.12}$$

Unitarily invariant norms are usually symmetric gauge functions.

It is clear that these unitarily invariant norms are directly proportional to the absolute measure of nonzero elements contained within the matrix. Therefore, they are a strong representation of the total link strength within a weighted digraph irrespective of their connectivity.

Assumption 2. For the sake of clarity, it is assumed here that the weight of a link in a digraph represents its strength or the energy contained within the link. Higher the energy of a link and equivalent amount of high energy if required by the adversary to disable it through a link failure DoS attack. Therefore, the bound $\|E\| \leq \epsilon$ is an energy bound. A better representation is $\|E\|_{fr} \leq \zeta$ since it is a direct bound on the total attack energy available to the adversary. Such an attack will be referred to as ζ*-strong DoS attacks.*

Note that, the attack matrix E need not be a constant. The attack policy might be follow any route: optimal, random, heuristic, etc. They all boil down to link disabling, this means that the elements of the matrix E can be continuously time varying, in fact, they should be significantly continuous for most the schemes that take a finite non-zero time to completely disable an attack, such as the data overloading DoS which causes packet loss by overloading the link with data. In this case the bound on $\|E\|$ might not necessarily be an energy bound since overloading can use additional amounts of energy than link disabling, but the energy bound will still be directly related to the bound on $\|E\|_{fr}$, therefore, this definition of ζ-strong DoS attacks is a versatile one. Furthermore, to come back to the main point of discussion the attack matrix E can be time varying such that if the attack starts at some time say τ_s and ends at tau_a then $E(t)$ follows for some initial constant attack matrix E_c:

$$\begin{aligned} E(\tau_s) &= E_c \in \mathbb{R}^{n\times n}, \\ E(\tau_a) &= E \end{aligned} \tag{2.13}$$

where E is as defined in (1.6) <change this eqn ref>. With this we can model any attack policy as long as it can be expressed in the form of E(t) irrespective of whether it is linear or non-linear. Therefore, the final equation of the net laplacian matrix of the system becomes:

$$\tilde{L}(t) = L - E(t) \tag{2.14}$$

For the time intervals when the system is not under attack $E(t) = 0$. With the attack model established, the next step is to analyze the system under attack using the existing results in the literature for consensus algorithms with time varying topology.

2.3 Preliminary Concepts

Now we come to the main section of this thesis, where we will finally present the required concepts for analysing the performance of the system under attack. Throughout this section we will denote the set of eigenvalues of the graph laplacian as $Eig(L) = \{\lambda_1, \lambda_2, \ldots, \lambda_n\}$, such that $\lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n$. The left and right eigenvalues corresponding to λ_i are denoted as ω_i and v_i respectively. To avoid confusion, v_i will no longer be used to represent graph node number i. Consider the matrices $W = [\omega_1\ \omega_2\ \ldots\ \omega_n]$ and $V = [v_1\ v_2\ \ldots\ v_n]$.

Theorem 2.3.1 (Diagonalization Theorem and Matrix Exponential)**.** *For the left and right matrices W, V comprising of the left and right eigenvectors of the matrix $L \in \mathbb{C}^{n\times n}$, the following relation holds:*

$$L = W^T \Lambda V \tag{2.15}$$

where Λ is the Jordan canonical form of L. If λ_i is a simple eigenvalue of L, $\forall i \in \{1, 2, \ldots, n\} = \mathcal{I}$ then $\Lambda = diag\,(\lambda_1, \lambda_2, \ldots, \lambda_n)$. The exponential of the matrix L for distinct eigenvalues then follows:

$$e^L = W^T e^\Lambda V, \tag{2.16}$$

$$\textit{where, } e^\Lambda = \begin{bmatrix} e^{\lambda_1} & 0 & \ldots & 0 \\ 0 & e^{\lambda_2} & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & e^{\lambda_n} \end{bmatrix}$$

This result leads to the expression of steady state consensus value of a strongly connected digraph with distinct eigenvalues:

Theorem 2.3.2. *For a general strongly connected digraph $\mathcal{G}$ the final value at which the nodes converge is:*

$$\lim_{t\to\infty} x(t) = x_{ss} = \omega_1^T x(t_o) v_1 1_n \tag{2.17}$$

where $1_n = [1, 1, \ldots, 1]^T \in \mathbb{R}^{n\times n}$, and $x(t_o)$ represents the initial value, given that the control law starts at t_o.

Theorem 2.3.3 (dual basis theorem)**.** *For a matrix $L \in \mathbb{C}^{n\times n}$ with distinct eigenvalues, $span\,(v_1, v_2, \ldots, v_n) = span\,(\omega_1, \omega_2, \ldots, \omega_n) = \mathbb{C}^{n\times n}$, ($\mathbb{R}^{n\times n}$ if the eigenvalues and the eigenvectors are real (we have assumed that the graph weights are real numbers)) and the left and right eigenvectors form a dual basis, i.e.:*

$$\langle \omega_i, v_j \rangle = \omega_i^T v_j = \delta_{ij} \tag{2.18}$$

where δ_{ij} is the Kronecker-Delta function, given that $\|\omega_i\| = \|v_j\| = 1 \forall i, j \in \mathcal{I}$. In general, relaxing the condition of unit norm, the consequence of this theorem is that $W^T V$ is diagonal.

Using dual basis theorem we get the following reproduction formulae for any vector $u \in \mathbb{C}^{n\times n}$:

$$u = \sum_{i=1}^{n} \left[\frac{\omega_i^T u}{\omega_i^T v_i}\right] v_i \tag{2.19}$$

Therefore, from 2.6 we get:

$$x(t) = \sum_{i=1}^{n} \left[\frac{\omega_i^T x(t)}{\omega_i^T v_i} \right] v_i,$$
$$\dot{x}(t) = \sum_{i=1}^{n} \left[\frac{\omega_i^T (-Lx(t))}{\omega_i^T v_i} \right] v_i,$$
$$\dot{x}(t) = \sum_{i=1}^{n} \left[\frac{\omega_i^T (-\lambda_i x(t))}{\omega_i^T v_i} \right] v_i,$$

multiplying ω_i on both sides ,

$$\omega_i \dot{x}(t) = \lambda_i \omega_i^T x(t),$$
$$\text{let } y(t) = \omega_i^T x(t) \text{ then,}$$
$$\dot{y}(t) = -\lambda_i y(t) \implies y(t) = e^{-\lambda_i t} y(t_o),$$
$$\therefore \omega_i^T x(t) = e^{-\lambda_i t} \omega_i^T x(t_o),$$

Using 2.19 now yields the equation of $x(t)$:

$$x(t) = \sum_{i=1}^{n} e^{-\lambda_i t} \left[\frac{\omega_i^T x(t_o)}{\omega_i^T v_i} \right] v_i = e^{-Lt} x(t_o) \tag{2.20}$$

This leads us to two important conclusions:

1. If there is a disturbance or a perturbation in any of the components of $x(t)$, the lower magnitude eigenvalues will have a higher value of $e^{-\lambda_i t}$ and will, therefore, increase the perturbation.
2. Equation 2.20 is significantly invariant w.r.t vector translations and magnitudes of the eigenvectors due to the $\langle \omega_i, v_i \rangle$ term in the denominator.

Note the following theorem:

Theorem 2.3.4. *The eigenvalues of a matrix vary continuously with its elements. Note that vector induced norms and unitarily invariant norms and the eigenvalues are related as:*

$$\lambda_i \leq \|A\| \leq |||A||| \leq \|A\|_{tr} = tr(A), \ \forall i \in \mathcal{I} \tag{2.21}$$

At this point we have two options to quantify the extent to which performance degrades under attack:

1. It can be assumed in some cases that the initial topology, therefore, its resultant final steady state value is of strategic importance. Defined for protocol 2.5, the steady state 2.17 is topology dependent. In cases such as these an attack will cause a departure from the original non-attack steady state. Then, our focus will be to prove that the departure is proportional to the attack frequency, attack energy as well as initial connectivity. Its proportionality on initial connectivity is because the departures/perturbations from initial state originate at the attack affected nodes and spread faster with increased connectivity for energy bounded systems as we shall see later. From 2.32 it is clear that the eigenvalues directly affect this result. However, not all systems have a hard requirement on the initial value.
2. In almost all cases, achieving consensus is the only goal. Then, here the feasible heuristic to measure will be the impact an attack will have on consensus performance. This too

depends strongly on the eigenvalues of L, for the smaller they are in magnitude the weaker the convergence. Even in the first point, the smaller the eigenvalues for a given number of links (which usually means that the magnitude of the links is small, 2.3.4) more will the perturbation spread, since the attenuation of perturbation is dependent upon the exponential terms $e^{\lambda_i t}$.

From both the points above one thing is very clear, we need to study the variations in the eigenvalues. They dictate the performance of the system. Our objective therefore is to analyze the variation in eigenvalues when the system is attacked relative to when the system is not attacked. We have two options here as well:

1. Directly measure the eigenvalue variation between $\tilde{L}$ and L. This will allow us to analyze point 1.
2. Prove that the rate of convergence of the consensus protocol degrades under attack to analyze point 2.

The main objective is to find a heuristic as a function of the attack capabilities and system parameters defining the degradation of performance in both the cases above which can be used for topology design problems. Therefore, at this point, we turn to *Matrix Perturbation Theory* [31] for an all out analysis of the eigenvalue variation.

2.3.1 Optimal Matching Distance

The optimal matching distance, in general, is defined on all closed subsets L, M of some compact space $\mathbb{C}$, we will, however, define it in terms of the eigenvalue sets of matrices which follow the condition of compactness. Consider arbitrary matrices $A, B \in \mathbb{C}^{n\times n}$. Let $\alpha_1, \alpha_2, \ldots, \alpha_n \in spec(A) = Eig(A)$, $\beta_1, \beta_2, \ldots, \beta_n \in spec(B) = Eig(B)$ be the sets of eigenvalues of the matrices and let $\mathcal{I} = \{1, 2, \ldots, n\}$. Then the *optimal matching distance* $d(Eig(A), Eig(B))$ between the sets of eigenvalues is defined as:

$$d(Eig(A), Eig(B)) = \min_{\pi\in S_n} \max_{i\in\mathcal{I}} \left(|\alpha_i - \beta_{\pi(i)}|\right) \tag{2.22}$$

where S_n is the group of all possible permutations of n symbols in $\mathcal{I}$.

2.3.2 Hausdorff Distance

Consider closed subsets $L, M \subseteq \mathbb{C}$ for some compact space $\mathbb{C}$. Then the *one sided Hausdorff distance* $\nu(L, M)$ is defined as:

$$\nu(L, M) = \sup_{\lambda\in L} dist(\lambda, M) \tag{2.23}$$

where the distance between some point $\lambda \in \mathbb{C}$ and set $M \subseteq \mathbb{C}$ is defined as:

$$dist(\lambda, M) = \inf_{z\in M} dist(\lambda, z) \tag{2.24}$$

The distance between points $dist(\lambda, z)$, unless otherwise stated, is assumed to be euclidean. This of course, means that the space $\mathbb{C}$ is assumed to be a compact *Hilbert Space*. The two sided or *Bi-Directional Hausdorff Distance* or just Hausdorff Distance between the sets L and M, $h(L, M)$, is now defined as:

$$h(L, M) = \max(\nu(L, M), \nu(M, L)) \tag{2.25}$$

Therefore, the Hausdorff distance between the eigenvalue sets $Eig(A)$ and $Eig(B)$ as defined in the previous section is defined as:

$$h(Eig(A), Eig(B)) = \max(\nu(Eig(A), Eig(B)), \nu(Eig(B), Eig(A))) = \max_{\sigma \in S_n} \min_{i \in \mathcal{I}} \left(|\alpha_i - \beta_{\sigma(i)}|\right) \tag{2.26}$$

where S_n is as defined in the previous section.

Definition 2.3.1 (Hausdorff Distance)**.** If the Hausdorff Distance between two sets L and M is ϵ, this implies that ϵ is the smallest number such that an ϵ-nbd (ϵ-neighbourhood) of every point in L contains at least one point from M, and vice-versa. That is, $\forall\ \mu_i \in L,\ \exists\ \beta_j \in M$ such that $\beta_j \in nbd_\epsilon(\mu_i)$, and vice-versa for some $(i, j) \in \mathcal{I}$.

2.3.3 Preliminary Results from Matrix Perturbation Theory

Hadamard Inequality

Theorem 2.3.5 (Hadamard Inequality)**.** *Let X and Y be any two $n \times n$ matrices, let $\lambda \in spec(Y)$, then:*

$$|det(X - \lambda I)| \leq \|X - Y\| \left(\|X\| + \|Y\|\right)^{n-1} \tag{2.27}$$

Upper bounds on Hausdorff and Optimal Matching distance

Theorem 2.3.6. *Consider two $n \times n$ arbitrary matrices A and B, and let $Eig(A)$, $Eig(B)$ be the sets of their eigenvalues in any order, then the following relations hold true:*

$$h(Eig(A), Eig(B)) \leq n^{\frac{1}{n}} (2M)^{1-\frac{1}{n}} \|A - B\|, \tag{2.28}$$

$$d(Eig(A), Eig(B)) \leq c(n) n^{\frac{1}{n}} (2M)^{1-\frac{1}{n}} \|A - B\|, \tag{2.29}$$

$$\text{where } M = \max(\|A\|, \|B\|), \text{ and } c(n) = \begin{cases} n & , n \text{ is odd} \\ n-1 & , n \text{ is even} \end{cases}$$

Upper bounds on distance between eigenvalues

Theorem 2.3.7. *Let A and B be arbitrary $n \times n$ matrices then:*

$$d(Eig(A), Eig(B)) \leq 4 \left(\|A\| + \|B\|\right) \|A - B\|^{1-\frac{1}{n}} \tag{2.30}$$

and for some $a \in Eig(A)$ and $b \in Eig(B)$:

$$|b - a| \leq 4 \cdot 2^{-\frac{1}{n}} \left(\|A\| + \|B\|\right)^{1-\frac{1}{n}} \|A - B\|^{\frac{1}{n}} \tag{2.31}$$

2.3.4 Consensus Under Switching Topology

One important parameter is the deviation from the non-attack value. We know that if the trajectory were to remain unchanged then the system would have settled on the final value given by 2.17. However, the attack changes its topology. In fact, the problem now strongly resembles a *switched topology* consensus problem. It is shown in [29] that for switching according to the *dirac-delta* function ($\delta(t - t_{si})$, where t_{si} is the switching time for topology i, the expression of $x(t)$ becomes:

$$x(t) = e^{C_m(t - t_{s_m})} \left[\prod_{i=1}^{m-1} e^{C_i \Delta t_i}\right] x(t_o),\ t \in \left[t_{s_m}, t_{s_{(m-1)}}\right) \tag{2.32}$$

where C_i denotes the negative of the graph laplacian corresponding to the topology during the time interval $[t_{s_{(i-1)}}, t_{s_i})$.

Theorem 2.3.8 (consensus under switching topology)**.** *In the case of switching topology, if there exists a finite time duration ξ such that the union of the graphs over the time interval $[t, t+\xi]$ contains a spanning tree $\forall t \in \mathbb{R}_{\geq 0}$, then the multi-agent system achieves consensus asymptotically.*

A better, higher utility feasible heuristic to measure the reduction in performance of the attacked system is the reduction in the rate of convergence due to loss of links. To prove that a trade-off exists, we want to prove that a system with higher connectivity sometimes leads to higher reduction in convergence rates in the attacked system. This means that irrespective of the change in the consensus value due to attack, we will focus on the time taken to reach convergence, i.e. the attention is shifted towards the change in convergence rate due to attack.

Chapter 3

Problem Formulation

3.1 The DoS Attack

Assumption 3 (the attack scheme). The following assumptions specify the exact DoS attack model used in the following sections:

1. The attacker executes the DoS attacks at discrete time intervals in the form of DoS bursts lasting for a finite duration of time, each such interval is called an *attack interval*. The set of starting times of the attack intervals is denoted as $T_s = \{t_1^a, t_2^a, t_3^a, \ldots, t_m^a\}$ for m attack intervals. Consider the time interval $[t_k, t_{k+1})$ such that $t_k < t_k^a < t_{k+1}$, then the attack begins at t_k^a and ends at t_{k+1}, therefore, the duration of attack in attack interval k is $\tau_k = t_{k+1} - t_k^a$. Let the set of time indices be $\mathcal{I}_\tau = \{1, 2, \ldots, m\}$.

2. The attack model follows 2.14. It is assumed that the matrix $E(t)$ is non-zero and constant throughout $[t_k^a, t_{k+1})$, i.e. the *attacked links $\mathcal{E}_{dos}$ are disabled instantly upon attack.*

3. Furthermore, it is assumed that the system instantly restores itself to its original state during the *non-attack/safe intervals* $[t_k, t_k^a)$, i.e. $E(t)$ is equal to $0_{n\times n}$ for $\forall t \in [t_k, t_k^a)$. *Therefore, every topology switching is instantaneous.*

$$\tilde{L}(t) = \begin{cases} L & t \in [t_k, t_k^a) \\ L - E_k & t \in [t_k, t_k^a) \end{cases} \tag{3.1}$$

Assumption 4 (limited attack resources). The attacker is assumed to have limited resources over some given finite time period, therefore, it is assumed that the attack matrix E has a bounded norm for all intervals $\|E_k\| \leq \zeta \ \forall k \in \mathcal{I}_\tau$. i.e. the attacks are ζ-strong.

Assumption 5 (link selection scheme). The attacker randomly selects the links to disable one by one, but *the condition $\|E_k\| \leq \zeta$ is strict* in the sense that if the currently selected m links (say) don't violate the condition, but the selection of link number $m+1$ violates the condition, i.e. $\|E_k\| \geq \zeta$ if link number $m+1$ is added to $\|E_k\|$, then the attack will proceed with the previously selected m links. The links are *uniformly randomly selected* for addition in the attack matrix. This is because, it is assumed that weakened links can regain their weight once the switching stops since the communication link still exists thus a link weakening (weight reduction and not cancellation due to bounded attack) won't have any effect. *The attacker is assumed to use as much energy as possible while obeying the ϵ strength bounds. i.e. the attacker will continue to add links until it encounters the similar condition as described above, at which point, it will stop adding more links.*

3.2 The System under Attack

Assumption 6 (asymptotic convergence). It is assumed that the system achieves consensus asymptotically, since it follows from Assumption 3 that Theorem 2.3.8 holds true. This, of course, assumes that the switching between topologies doesn't produce an overall destabilizing effect, i.e. the switching between topologies at interval boundaries doesn't produce a destabilizing effect with stable modes at non-attack intervals.

Given , convergence is guaranteed. The *rate of convergence* dictates the overall performance of the system, since it is a measure of the speed with which consensus is achieved. It is proposed here that the rate of convergence is degraded when the system is attacked. Within any given time interval in which $\tilde{L}$ is constant, the eigenvalues of $\tilde{L}$ (and in some cases ω_i, for some $i \in \mathcal{I}$ as well) determine the convergence rate, therefore, closer the eigenvalues are to the RHP, slower is the convergence. The shift in eigenvalues must be a function of the attack resources and system parameters which will function as a heuristic to *estimate security*. If the attacked system $\tilde{L} = L - E_k$ and the normal system L contain a spanning tree, then eqn:bound on consensus convergence rate gives definite bounds on the convergence rate. The bounds will shift under attack, as the solution of the modified lyapunov equation eqn:modified lyapunov equation i.e. P changes to $\tilde{P}_k$ for fixed $Q = Q^T \in \mathbb{C}^{n\times n}$ and $\alpha \in \mathbb{R}_{>0}$ as $\tilde{L}$ changes from L to $L - E_k$, thus resulting in a shift of bounds. This shift too should be characterized by a function of the attack and system parameters and if the separation between the upper and lower bounds doesn't increase then it will imply guaranteed degradation of performance under attack.

The objective, therefore, is to find a function to estimate this shift.

Chapter 4

The Consensus Lyapunov Function

To measure the change in the convergence rate, we need to estimate the convergence rate first. We start again by asking the most important questions: Will the system still be stable after attack? Will it converge to a value even if it is not strongly connected? If yes, is there a way to prove this by the Lyapunov stability criteria? The chapter title gives a spoiler to the last question. It is clear that there certainly is a way to prove convergence for a general digraph regardless of its connectivity level, and that too, using the Lyapunov analysis.

To start answering the questions in order, the system will be stable after a attack, due to the protocol (2.5) and our discussion in section 2.1.2 about weakly connected digraphs. It is important to understand that the nodes will always settle to a value, its just that, unless a spanning tree exists they will never achieve consensus. From 3.2 and Theorem 2.3.8, it is clear that the system achieves consensus, but it is delayed due to the attacks. The delay is a consequence of link removal from matrix, which decreases $\|L\|_F$, and thus results in a net decrease of the positive real parts of the eigenvalues of L. Intuitively, this can be inferred from the fact that $tr(L) = \sum_{i=1}^{n} \lambda_i = \sum_{i=1}^{n} \sum_{j=1, j\neq i}^{n} a_{ij}$, where the rightmost side decreases as the links are removed from $\mathcal{G}$. There are cases where link removal might speed up the convergence, which shall be demonstrated in Chapter 5. However, such cases are practically rare and usually exist at the boundary of rank transitions due to attack. Therefore, under the state of attack, the system may or may not lose connectivity, and thus may stop achieving consensus.

Let's illustrate how attacks delay convergence with an example. Imagine that a strongly connected graph is subjected to the DoS attack described in section 2.2.1. Consider that this attack results in the graph becoming weakly connected. Assume that, the system converged to some final state during the attack, i.e. the nodes settle at a value and stop evolving for the rest of the attack interval. From the discussion on weakly connected digraph [24], the nodes won't converge to the same value. This implies that, when the attack interval ends and system network is restored, protocol (2.5) will start evolving the nodes again until they achieve agreement, an agreement they could have achieved earlier had it not been interrupted by connectivity loss due to the DoS attack. Of course, if the system achieves consensus during an attack interval it will converge once and for all. But, on an average, this mechanism of connectivity loss during attack intervals leads to delay in agreement. Furthermore, losing links without losing connectivity still slows down convergence due to the decrease in eigenvalues' real parts, as proved by the result (2.6).

Therefore, the system still converges to some value despite the attack, it just may to reach agreement. An agreement will certainly be achieved after a specific non-attack interval beyond some time τ_c. This leads to an important fact: a valid lyapunov function gives admissible

upper and lower bounds on the rate of convergence. This was used by the authors in [23] to prove that the convergence rate of protocol (2.5) for a digraph with a spanning tree is bounded above by the *Feidler Eigenvalue* λ_2 of the graph laplacian.

Clearly, since digraphs of arbitrary connectivity levels will be encountered while analyzing our case, it is important to attain an expression on the bound of the convergence rate of the multi-agent system which is valid irrespective of its connectivity levels, i.e. we must now seek a lyapunov function for general digraphs which not only answers the last question about proving convergence through Lyapunov analysis, but also provides convergence rate bounds.

4.1 Lyapunov function for general digraphs

Since the rate of convergence of the system is the rate of convergence of the lyapunov function to the final constant value in limit set, we present the lyapunov function for general connected digraphs as derived in [30].

Theorem 4.1.1 (lyapunov function for connected digraph)**.** *Consider a graph $\mathcal{G}$ containing a spanning tree and a multi-agent system connected by $\mathcal{G}$ whose augmented state vector is denoted as $x(t)$. Then for the laplacian L of $\mathcal{G}$, for any positive definite matrix $Q = Q^T \in \mathbb{R}^{n\times n}$, and for a positive real number $\alpha \in \mathbb{R}_{>0}$ there exists a positive definite matrix $P = P^T \in \mathbb{R}^{n\times n}$, such that the following modified lyapunov equation holds:*

$$PL + L^T P = Q - \alpha\left(P1_n\omega_1^T + \omega_1 1_n^T P\right) \tag{4.1}$$

Therefore, the following holds true for the system dynamics 2.6:

$$y(t) = x(t) - \left(\omega_n^T x(t)\right)1_n, \tag{4.2}$$

$$V(t) = y^T P y > 0, \tag{4.3}$$

$$\dot{V}(t) = -y^T Q y < 0 \tag{4.4}$$

Note that as $x(t) \to \lim_{t\to\infty} x(t), y(t) \to 0 \implies \lim_{t\to\infty} V(t) \to 0$. Thus, $V(t)$ is a valid lyapunov function for the digraph having a spanning tree. This relation gives us the generalized upper and lower bounds on the convergence rate of the system as:

$$-\frac{\lambda_{max}(Q)}{\lambda_{min}(P)}V(t) \le \dot{V}(t) \le -\frac{\lambda_{min}(Q)}{\lambda_{max}(P)}V(t) \tag{4.5}$$

However, the lyapunov function we have formulated so far is valid only for the case where a spanning tree exists. Therefore, the author hereby presents an extension to this concept to digraphs where $rank(L) = n - m$, which implies that the eigenvalue 0 has algebraic multiplicity m. Let Λ be the Jordan canonical form of L.

$$L = V\Lambda W^T \tag{4.6}$$

$$\Lambda = \left[\begin{array}{c|ccc} \Lambda_0 & 0 & \dots & 0 \\ \hline 0 & \Lambda_1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \Lambda_k \end{array}\right] = \left[\begin{array}{c|c} \Lambda_0 & \mathbf{0} \\ \hline \mathbf{0} & \Lambda_{n\setminus 0} \end{array}\right] \tag{4.7}$$

$$\Lambda_i = \begin{bmatrix} \lambda_i & 1 & 0 & \dots & 0 \\ 0 & \lambda_i & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_i & 1 \\ 0 & 0 & \dots & 0 & \lambda_i \end{bmatrix} \in \mathbb{C}^{g\times g} \text{ if algebraic multiplicity of } \lambda_i \text{ is equal to } g. \tag{4.8}$$

Clearly, Λ_0 is an $m \times m$ matrix while $\Lambda_{n\setminus 0}$ is $n-m \times n-m$. Lets define $\Lambda_0(\alpha)$ as the matrix containing $\lambda_0 = \alpha$ instead of $\lambda_0 = 0$, i.e.:

$$\Lambda_0(\alpha) = \begin{bmatrix} \alpha & 1 & 0 & \dots & 0 \\ 0 & \alpha & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \dots & \alpha & 1 \\ 0 & 0 & \dots & 0 & \alpha \end{bmatrix} \in \mathbb{R}^{m \times m} \tag{4.9}$$

$$\Lambda(\alpha) = \left[\begin{array}{c|c} \Lambda_0(\alpha) & \mathbf{0} \\ \hline \mathbf{0} & \Lambda_{n\setminus 0} \end{array}\right] \tag{4.10}$$

We now define the matrix $R(\alpha)$ for some $\alpha > 0 \in \mathbb{R}$ as:

$$R(\alpha) = V\Lambda(\alpha)W^T = \underbrace{\begin{bmatrix} v_1 & \dots & v_m \end{bmatrix}}_{V_0} \underbrace{\begin{bmatrix} v_{m+1} & \dots & v_n \end{bmatrix}}_{V_{n\setminus 0}} \left[\begin{array}{c|c} \Lambda_0(\alpha) & \mathbf{0} \\ \hline \mathbf{0} & \Lambda_{n\setminus 0} \end{array}\right] \left.\begin{bmatrix} \omega_1^T \\ \vdots \\ \omega_m^T \end{bmatrix}\right\} W_0^T \left.\begin{bmatrix} \omega_{m+1}^T \\ \vdots \\ \omega_n \end{bmatrix}\right\} W_{n\setminus 0}^T \tag{4.11}$$

$$\begin{aligned} R(\alpha) &= V_0\Lambda_0(\alpha)W_0^T + V_{n\setminus 0}\Lambda_{n\setminus 0}W_{n\setminus 0}^T, \\ L &= V_0\Lambda_0(0)W_0^T + V_{n\setminus 0}\Lambda_{n\setminus 0}W_{n\setminus 0}^T, \\ \therefore R(\alpha) - L &= V_0\left[\Lambda_0(\alpha) - \Lambda_0(0)\right]W_0^T, \\ \text{Now, } \Lambda_0(\alpha) - \Lambda_0(0) &= diag(\alpha, \alpha, \dots m \text{ times}) = D \text{ (say)}, \\ \therefore R(\alpha) - L &= V_0 D W_0^T = \alpha\sum_{i=1}^{m} v_i\omega_i^T = \alpha\sum_{i=1}^{m} v_i \otimes \omega_i, \\ R(\alpha) &= L + \alpha K, \text{ where } K = V_0W_0^T = \sum_{i=1}^{m} v_i \otimes \omega_i \text{ for } rank(L) = n-m \end{aligned} \tag{4.12}$$

Note that $R(\alpha)$ is Hurwitz for $\forall \alpha > 0$, therefore, (4.1) can be written as:

$$PL + L^TP = Q - \alpha\left(PK + K^TP\right) \tag{4.13}$$

Note that $K = V_0W_0^T$, V_0 comprises of the right eigenvectors while W_0 contains the dual basis for the eigenspace of eigenvalue 0. Assuming that the eigenvalue 0 is not defective and that the eigenvectors are normalized to unit magnitude, we get, $W_0^TV_0 = I_m = V_0^TW_0$. Then, $K^n = K, (K^T)^n = K^T \ \forall \ n \in \mathbb{N}$. Now, modifying the definition of y from (4.2) as:

$$y(t) = x - Kx \tag{4.14}$$

Tracing the steps of [30], the validity of (4.3) and (4.4) follows for y defined as in (4.14) by a straightforward proof utilizing properties of K.

Theorem 4.1.2 (lyapunov function for general digraph)**.** *Consider a graph $\mathcal{G}$ containing a spanning tree and a multi-agent system connected by $\mathcal{G}$ whose augmented state vector is denoted as $x(t)$. Then for the laplacian L of $\mathcal{G}$, such that $rank(L) = n-m$, for any positive*

definite matrix $Q = Q^T \in \mathbb{R}^{n\times n}$, and for a positive real number $\alpha \in \mathbb{R}_{>0}$ there exists a positive definite matrix $P = P^T \in \mathbb{R}^{n\times n}$, such that the following modified lyapunov equation holds:

$$PL + L^T P = Q - \alpha\left(PK + K^T P\right) \tag{4.15}$$

where $K = \sum_{i=1}^{n} v_i \otimes \omega_i$. Therefore, the following holds true for the system dynamics 2.6:

$$\begin{aligned} y(t) &= x(t) - Kx(t), \\ V(t) &= y^T P y > 0, \\ \dot{V}(t) &= -y^T Q y < 0 \end{aligned}$$

Solution set of weakly connected digraphs

The exponential bounded nature of convergence rate given by (4.5) holds true for weakly connected graphs as well. In fact, this analysis reveals the set which weakly connected systems converge to. Since Q is positive definite and $\dot{V}(t) = 0$ at steady state then $y(t \to \infty) = 0$. This means $(I - K)x_{ss} = 0$.

Then,

$$x_{ss} \in null(I - K), \text{ for weakly connected digraphs} \tag{4.16}$$

With all the preliminary concepts defined and in place, the final problem is formulated.

4.2 Perturbation Theory for Sylvester Equation: $XA + BX = C$

The problem is formulated as a matrix perturbation problem in L for $\tilde{L} = L - L_E$. This will, in turn, lead to perturbations in the matrix $R(\alpha)$ (and, therefore, K) defined in Theorem 4.1.2. Note that the equation (4.15) is equivalent to the continuous time Lyapunov equation $PR(\alpha) + R(\alpha)^T P = Q$, which is a special of the symmetric Sylvester equation. For constant Q and α, when the system is attacked the following changes occur: $L \to \tilde{L}$ and $K \to \tilde{K}$ using the usual notation to denote quantities under attack. This can be modeled as $\tilde{L} = L + \Delta L$, (of course, $\Delta L = -E$) and $\tilde{K} = K + \Delta K$. The perturbation in K are a result of the change in the eigenvectors of L. For simplicity, it is assumed that $rank(L)$ remains constant, then it is clear that ΔK is not independent of ΔL since the latter dictates the perturbation in the eigenvalues, so it is better to use $\tilde{R}(\alpha) = R(\alpha) + \Delta R$. A relationship exists between ΔR and L_E which will be a subjected to investigation in the upcoming sections. At this point, Theorem 8.3.1 from [32] is stated without proof since it is beyond the scope of this publication.

Theorem 4.2.1 (Perturbation Theorem for the Sylvester Equation)**.** *Let the Sylvester Equation $XA + BX = C$ have a unique solution X for $C \neq 0$. Let $\Delta A, \Delta B, \Delta C$, and ΔX be the perturbation in the matrices A, B, C, and X respectively. Let $\tilde{X}$ be the solution of the perturbed problem. That is, $\tilde{X}$ is a solution of:*

$$\tilde{X}\left(A + \Delta A\right) + \left(B + \Delta B\right)\tilde{X} = C + \Delta C \tag{4.17}$$

Let

$$\epsilon = \max\left\{\frac{\|\Delta A\|_F}{\alpha}, \frac{\|\Delta B\|_F}{\beta}, \frac{\|\Delta C\|_F}{\gamma}\right\} \tag{4.18}$$

where α, β, and γ are tolerances such that $\|\Delta A\|_F \leq \epsilon\alpha$, $\|\Delta B\|_F \leq \epsilon\beta$, and $\|\Delta C\|_F \leq \epsilon\gamma$.

Then,

$$\frac{\|\Delta X\|_F}{\|X\|_F} = \frac{\|\tilde{X} - X\|_F}{\|X\|_F} \leq \sqrt{3}\epsilon\delta \tag{4.19}$$

$$\textit{where } \delta = \left\|P^{-1}\right\|_2 \frac{(\alpha+\beta)\|X\|_F + \gamma}{\|X\|_F} \tag{4.20}$$

where P is defined as $P = (I_n \otimes B) + \left(A^T \otimes I_m\right)$.

Definition 4.2.1 (Matrix Separation)**.** The separation of two matrices A and B, denoted by $sep(A, B)$, is defined as:

$$sep(A, B) = \min_{X \neq 0} \frac{\|AX - XB\|_F}{\|X\|_F} \tag{4.21}$$

Thus, using the definition of the induced norms, it can be proved that:

$$\left\|P^{-1}\right\|_2 = \frac{1}{\sigma_{min}(P)} = \frac{1}{sep(B, -A)} \tag{4.22}$$

Using the separation function, the inequality (4.19) can be written as:

$$\frac{\|\Delta X\|_F}{\|X\|_F} < \sqrt{3}\epsilon \frac{1}{sep(B, -A)} \frac{(\alpha+\beta)\|X\|_F + \gamma}{\|X\|_F} \tag{4.23}$$

4.2.1 Application to the Attacked System

The lyapunov equation during the attack interval can be modeled as follows:

$$\begin{aligned} &\tilde{P}\tilde{R}(\alpha) + (\tilde{R}(\alpha))^T\tilde{P} = Q, \\ &(P + \Delta P)(R(\alpha) + \Delta R) + ((R(\alpha))^T + \Delta R^T)(P + \Delta P) = Q \\ &\Delta Q = 0, \ (Q \text{ remains constant}) \end{aligned} \tag{4.24}$$

It is known that $\|E\| \leq \zeta$, therefore, $\|L_E\| \leq \eta$ for some $\eta > 0$. Then $\Delta R = \Delta L + \Delta K = -L_E + \Delta K$. ΔK in turn depends on the outer products of vectors from a subset of normalized eigenvector pairs of L, therefore $\|\Delta K\|$ must be bounded. This implies that there exists some $\kappa > 0 \in \mathbb{R}$, such that, $\|\Delta R\|_F \leq \kappa \|R(\alpha)\|$. Then from equation (4.18), $\epsilon = \kappa$. Note that $\alpha = \beta$, and $\gamma = 0$ (as $\Delta Q = \mathbf{0}$). Applying Theorem 4.2.1 then gives:

$$\frac{\|\Delta P\|_F}{\|P\|_F} < \frac{2\sqrt{3}\kappa^2}{sep((R(\alpha))^T, -R(\alpha))} \tag{4.25}$$

Let λ_P and $\tilde{\lambda}_P$ be any two eigenvalues of P and $\tilde{P}$, respectively. Using (2.31) from Theorem 2.3.7 and from (2.21) we get:

$$\begin{aligned} |\tilde{\lambda}_P - \lambda_P| &\leq 4 \cdot 2^{-\frac{1}{n}} \left(\|P\| + \|\tilde{P}\|\right)^{1-\frac{1}{n}} \|\tilde{P} - P\|^{\frac{1}{n}} \\ &< 4 \cdot 2^{-\frac{1}{n}} \left(\|P\|_F + \|\tilde{P}\|_F\right)^{1-\frac{1}{n}} \|\tilde{P} - P\|_F^{\frac{1}{n}}, \quad (n \geq 1) \\ &< 4 \cdot 2^{-\frac{1}{n}} \left(\|P\|_F + \|P\|_F + \|\Delta P\|_F\right)^{1-\frac{1}{n}} \|\Delta P\|_F^{\frac{1}{n}} \\ &\text{using (4.25)} \\ |\tilde{\lambda}_P - \lambda_P| &< 4 \cdot 2^{2-\frac{1}{n}} \|P\|_F \left(1 + \frac{\sqrt{3}\kappa^2}{sep\left((R(\alpha))^T, -R(\alpha)\right)}\right)^{1-\frac{1}{n}} \left(\frac{\sqrt{3}\kappa^2}{sep\left((R(\alpha))^T, -R(\alpha)\right)}\right)^{\frac{1}{n}} \end{aligned} \tag{4.26}$$

From here onward it becomes important to investigate the trend followed by $\tilde{\lambda}_P$. The suspected decrease in convergence rate will result if $\tilde{\lambda}_P$ increases with the attack, especially the maximum and the minimum eigenvalues of $\tilde{P}$. In that case the window of bounds confining the actual convergence rate will shift in the direction of a decreasing trend, indicating a degradation of performance. If this holds true, then equation (4.26) indicates that the extent of performance degradation depends upon the separation term $sep((R(\alpha))^T, -R(\alpha))$. The nature of this separation term is discussed in [34].

The Matrix Separation Term: $sep(A, B)$

(4.21) is re-written in the form $sep(R^T, -R) = \min_{\|X\|_F=1} \left\|R^T X + XR\right\|_F$. For simplicity, lets start by analyzing the case of undirected graphs first. Note that for this section, R and $R(\alpha)$ are used interchangeably.

4.2.2 The case of undirected graphs

In the case of undirected graphs, the notion of weak connectivity no longer exists, the graph can either be connected or can be disconnected. Since the link removal is now symmetric, the matrices L, $\tilde{L}$ and E are Hermitian. The outer product of the eigenvectors i.e. $v_i \otimes v_i = v_i v_i^T$ is symmetric, implying that $R(\alpha)$ is hermitian as well.

It was shown in [35] that for normal matrices A and B:

$$sep(A, B) = \min_{i,j} |\lambda_i(A) - \lambda_j(B)| \tag{4.27}$$

The spectrum of $(R(\alpha))^T$ and (α) is symmetrically distributed about the imaginary axis, then for eigenvalues of L, $\lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n$:

$$sep(-R^T, R) = \begin{cases} 2\alpha & \alpha < \lambda_2 \\ 2\lambda_2 & \alpha \geq \lambda_2 \end{cases} \tag{4.28}$$

Clearly, for $\alpha < \lambda_2$ the bound can be arbitrarily large, therefore, for a conservative bound we should consider $\alpha \geq \lambda_2$. Then from 4.26, with similar assumptions, we get:

$$|\tilde{\lambda}_P - \lambda_P| < 4 \cdot 2^{2-\frac{1}{n}} \|P\|_F \left(1 + \frac{\sqrt{3}\kappa^2}{\lambda_2}\right)^{1-\frac{1}{n}} \left(\frac{\sqrt{3}\kappa^2}{\lambda_2}\right)^{\frac{1}{n}} \tag{4.29}$$

Equations (4.26) and (4.29) are the heuristics we were seeking all along. Similar results can be obtained for the Hausdorff Distance (section 2.3.2) and the Optimal Matching distance (section 2.3.1) of the eigenvalues, using the results from Theorem 2.3.6. Of special interest is the Hausdorff distance result since it establishes an upper bound on the minimum shift of eigenvalues from equation (2.28). For this bound to be a valid heuristic it should closely follow the actually profile. We analyze this aspect numerically in the next chapter.

4.3 Discussion and Conclusion

When working with these levels of generality where here is absolutely no assumption about the digraphs, the only level of simplification is provided by the particular DoS attack model chosen. Even then, it still behaves as a very general matrix perturbation problem. This leads us to the question: Will this level of generalization help answer the questions we have been asking ourselves all along with certainty? The answer is both yes and no. No, it won't give

us direct deterministic results, but we can get the range in which those results lie, and that range is determined by the bounds developed in this chapter. It is important to note that our choice of α and Q matrix in the Lyapunov equation affect these bounds. It is proposed here that *there must be an optimal choice of* α *and the matrix* Q *such that the bounds* (4.26) *and* (4.29) *are the tightest possible for a given digraph* $\mathcal{G}$. Of course, for a given α and Q, the choice of the digraph $\mathcal{G}$ affects the credibility of these bounds as we shall see with the example of undirected graphs in the next chapter. Some bounds on the separation term are also presented in [36].

So, we may have just worked with bounds until now without arriving at concrete analytical results, but, this certainly directs us towards the right direction.

Directional clarity is what this Thesis seeks to provide for driving future research in the domain. So that researchers, in the future, can get an idea about what may and what may not lead to the answers they seek.

To present the domain of topological effects on consensus performance from the broadest perspective possible and highlight the spots one should narrow it down to through the refreshing unprecedented perspective of spectral analysis and matrix perturbations. This leads us to the final chapter of this thesis which documents the experimental analysis used to validate or reject various hypotheses throughout this research.

Chapter 5

Simulations and Experimental Analysis

This chapter is all about experimental validation of the hypotheses made earlier in the text and forming new hypotheses based on their results and then validate some of them through further experimental validation. This chapter forms the final stage of this thesis having refuted previously held beliefs and opening newer doors for research by pointing in the right direction.

We will start by rigorously simulating the spectral variation of a randomly generated digraph as its links are removed. Then we will check the variation in the solution to the modified lyapunov equation (4.15) P. Finally, we will perform a monte-carlo like simulation of norm bounded digraphs being attacked by randomly generated norm bounded attacks. The graph topology will be varied by changing the number of non-zero links to check how network connectivity affects post-attack consensus performance.

Please note that validation in throughout this chapter doesn't mean actual proof, because an actual proof of any concept will be accompanied by concrete theoretical results. The reader is hereby requested keep in mind that the usage of the word validation will mostly refer to suggestion, and the two will be used interchangeably throughout the course of this chapter. Wherever a concrete verification is presented, the reason is explicitly stated.

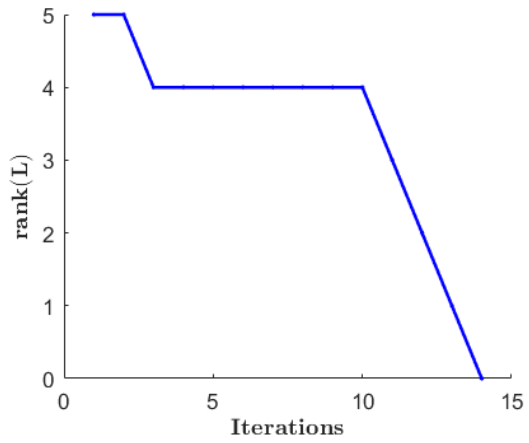
5.1 Spectral Trends with Link Disabling

5.1.1 Spectral Variation of Graph Laplacian

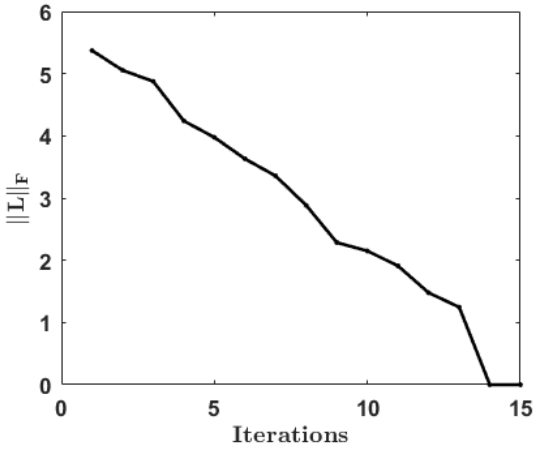
A digraph was randomly generated with 6 nodes and link weights $a_{ij} \in [0.5, 1] \cup \{0\}$. Each weight is uniformly selected from the range $[0, 1]$, and is allotted the value of zero if it is less than 0.50. This gives each edge a 50% chance of being included in the initial digraph. A link, i.e. non-zero weight element of A, is selected at random (uniformly randomly among all non-zero weights) and is then removed. This all links are removed one by one and the spectral properties are recorded at each stage.

The following conclusions are drawn about the spectral properties of the graph laplacian:

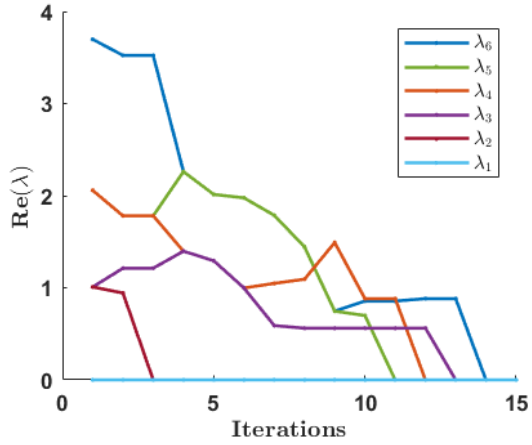
- Fig. 5.1b verifies that the Frobenious norm of the graph laplacian decreases monotonically as links are removed since the Frobenious norm is directly proportional to the square of the weight of each link which has a very straightforward proof from (2.12).
- Figs. 5.1c-5.1f clearly suggest that all the eigenvalues, especially their real parts, which dictate the convergence rate, decrease as the links are subsequently removed.



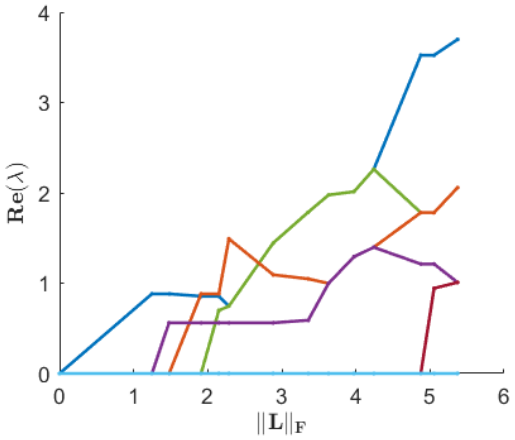
(a) Variation of $rank(L)$ with link removal.



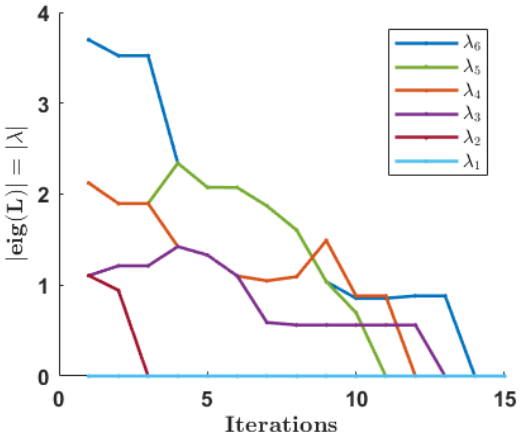
(b) Variation in $\|L\|_F$ with link removal.



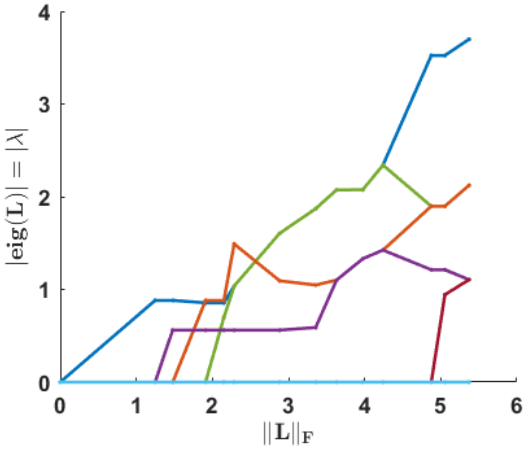
(c) Variation in $Re(eig(L))$ with link removal.



(d) Variation in $Re(eig(L))$ with $\|L\|_F$.



(e) Variation in $|eig(L)|$ with link removal.



(f) Variation in $|eig(L)|$ with $\|L\|_F$.

Figure 5.1: Variation of spectrum of the graph laplacian as network edges are removed one-by-one similarly to what happens during the link-failure DoS attack. One can consider that with each subsequent iteration, the attack strength increases. The graph was randomly generated and all the links weights are uniform random numbers.

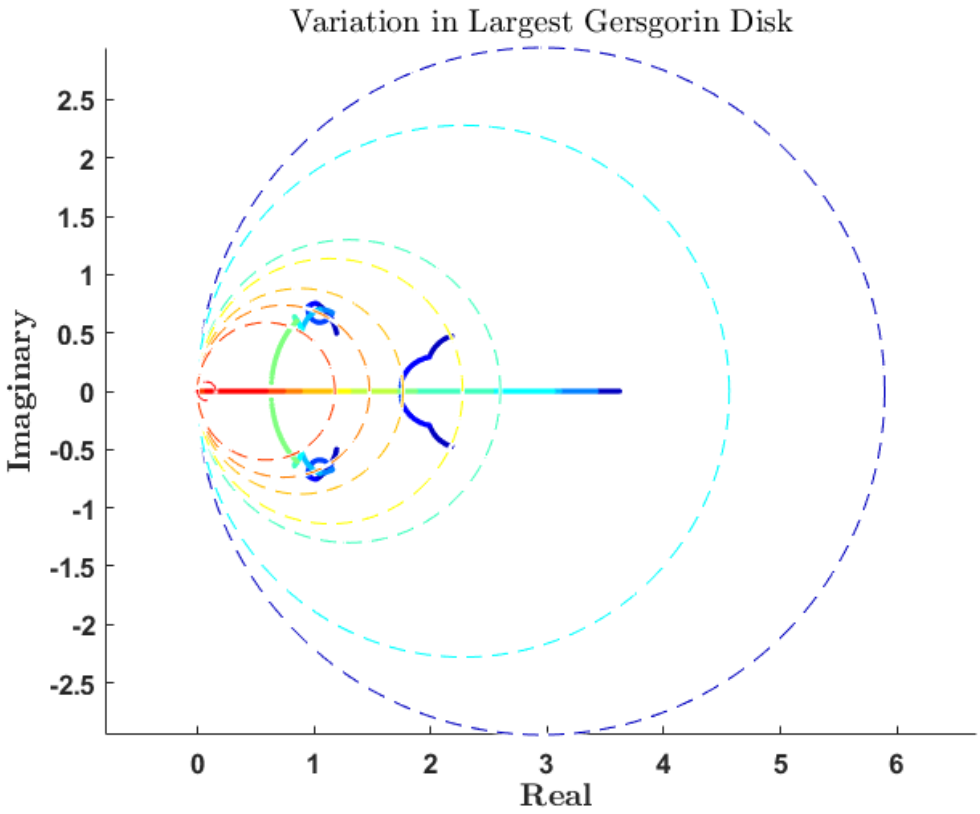


Figure 5.2: Variation in the Largest Gersgorin Disk as the links are removed. The transition of the eigenvalues towards the origin of the complex plane is clearly visible as the disk shrinks with subsequent link removal. The color gradients in the eigenvalue locations of L corresponding to their respective disks of the same colour.

The decrease of the eigenvalues is not necessarily monotonic. But it is guaranteed that they will eventually reach zero. This is a consequence of the Gersgorin disks theorem.

Theorem 5.1.1 (Gersgorin Disks Theorem)**.** *The eigenvalues of a square matrix* $M = [m_{ij}]$ *is contained within the union of the Gersgorin disks, which are circles in complex plane defined by the complex function* $|z - m_{ii}| = \sum_{j=1, j\neq i}^{n} m_{ij}$.

$$spec(M) \subseteq \bigcup_{i=1}^{n} \left\{ z \mid |z - m_{ii}| = \sum_{j=1 j\neq i}^{n} m_{ij} \right\} \tag{5.1}$$

A straightforward proof follows from the application of Theorem 5.1.1 to the graph laplacian L (the reader is referred to [23]). The largest Gersgorin disk of L contains all the other Gersgorin disks. For L, $l_{ii} = \sum_{j=1, j\neq i}^{n} l_{ij}$ which is equal to the radius of the Gersgorin disks. This means that as links are removed the largest converges to become a point on the origin while still lying entirely on the right half of the complex plane. Therefore, since all the eigenvalues of L are constrained to stay within the Gersgorin disk, they all approach the origin and are guaranteed to become zero when all the links are removed as illustrated in Fig. 5.2. The decrease is not monotonic, but on a average, a performance degrade is guaranteed even if connectivity stays the same which is suggested by the decreasing trends in the eigenvalues corresponding to the horizontal segments of 5.1a (between link disabling iteration number 5 to 10 for example).

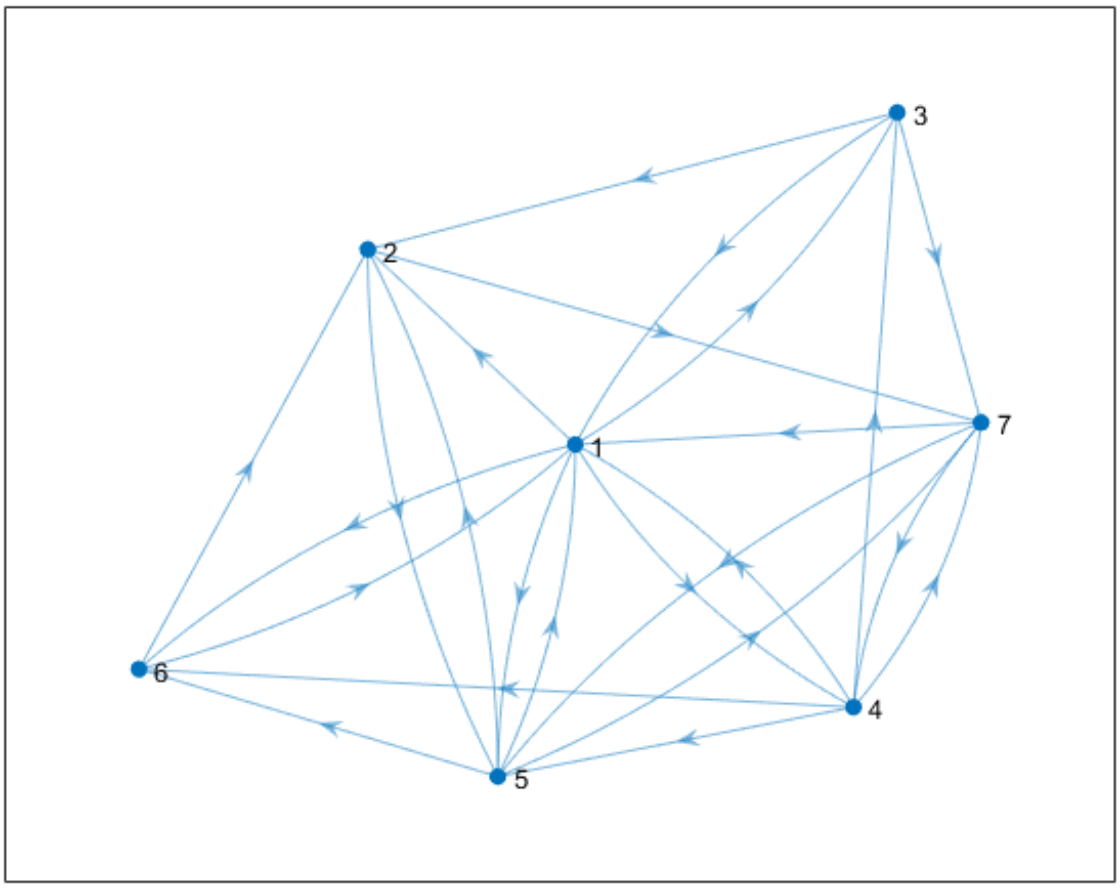


Figure 5.3: Initial digraph $\mathcal{G}$ before link disabling.

5.1.2 Variation in P and Convergence Rate Bounds

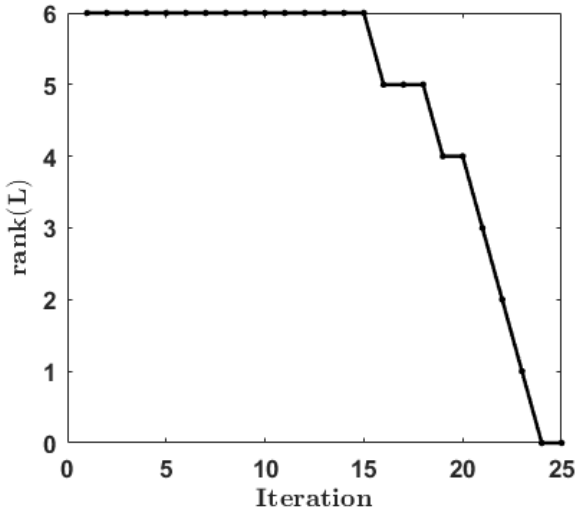
In the previous section, we discussed the variation of spectral properties of the graph laplacian L. In this section, following the discussion from section 4.3, the convergence rate bounds given by the solution to the modified lyapunov equation (4.15), i.e. P are simulated.

The digraph is generated randomly using a the same method as the previous section. $\alpha = 0.2, Q = diag(2, 2, 2, \ldots n\ times)$ are selected. The initial digraph G before any link disabling is performed is shown in Fig. 5.3. Links are disabled using the same scheme as the one used in the previous section.

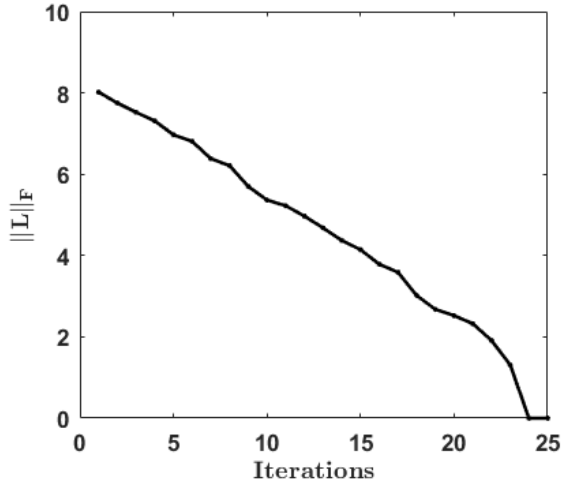
Fig. 5.4a and 5.4b reiterates the results from the last section. Fig. 5.4c is a consequence of the fact that, when all the links are disabled, $L = 0 = K$; therefore, $R(\alpha) = diag(\alpha, \alpha, \ldots n\ times)$. Then, $P = 0.5 \times \alpha \times Q$. Hence, as all the links are disabled, the eigenvalues of P all tend to become equal and are given by $\lambda_{P,i} = 0.5 \times \alpha \times q_{ii}$, which simplifies to 5 for all eigenvalues in this case.

The results in Fig. 5.5 suggests that $\|R(\alpha)\|_F$ varies linearly with $\|L\|_F$ for larger values of $\|L\|_F$, i.e. for graphs with higher connectivity. A specific trend in $\|K\|_F$ wasn't expected and the graph is probably random. Any pattern, if existent, couldn't be revealed in this simulation and may need additional analytical and experimental results. The properties of K are important since they can change the nature of the solutions of the lyapunov equation and are therefore open for future work. Due to the seemingly arbitrary variation of $\|K\|_F$, a specific pattern cannot be established for $\|P\|_F$ as well. Although the final value is known due to the analysis in the previous paragraph.

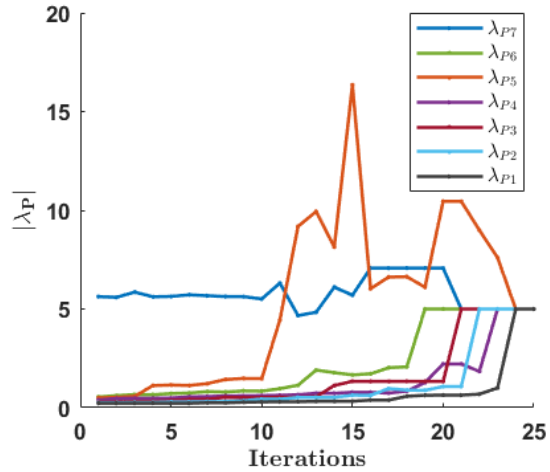
Finally, Figs. 5.6b and 5.6c illustrates the variation of the convergence rate bounds. This



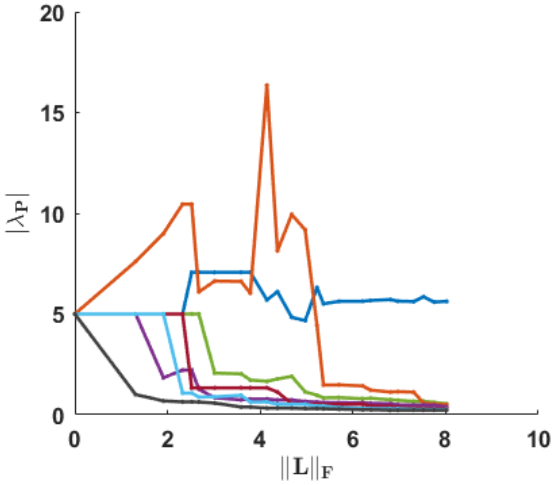
(a) Variation of $rank(L)$ with link removal.



(b) Variation in $\|L\|_F$ with link removal.



(c) Variation in λ_P with link removal.



(d) Variation in λ_P with $\|L\|_F$.

Figure 5.4: Variation of solution's spectrum of the modified Lyapunov Equation (4.15), i.e. P as the links are disabled.

is a clear performance degradation, which the matrix P was expected to reflected following the suggestions of the previous section. $\zeta_U = -\frac{\lambda_{q,min}}{\lambda_{p,max}}$ and $\zeta_L = -\frac{\lambda_{q,max}}{\lambda_{p,min}}$ are referred to as the upper and the lower bounds on convergence rates due to their positions. It is interesting to observe that, in the case of random link removal, every link removal leads a monotonic increase in the lower convergence rate bound. The window between the upper and the lower bounds shrinks and ζ_L becomes less negative. ζ_U on the other hand exhibits a slight decrease, but it is almost negligible. Therefore, it is safe to conclude that on an average, the convergence rate of the system will decrease at some point. As ζ_L decreases beyond its current value, it will be forced to decrease with ζ_L for some iterations.

5.2 DoS Attacks on an Undirected Graph at Different Connectivity Levels

In the last section of this thesis we reach perhaps the most important and interesting section of all we have encountered so far. It is a culmination of all the analyses so far. This section simulated randomly generated link disabling DoS attacks on the system where all the changes in all the parameters, before and after the attack are recorded.

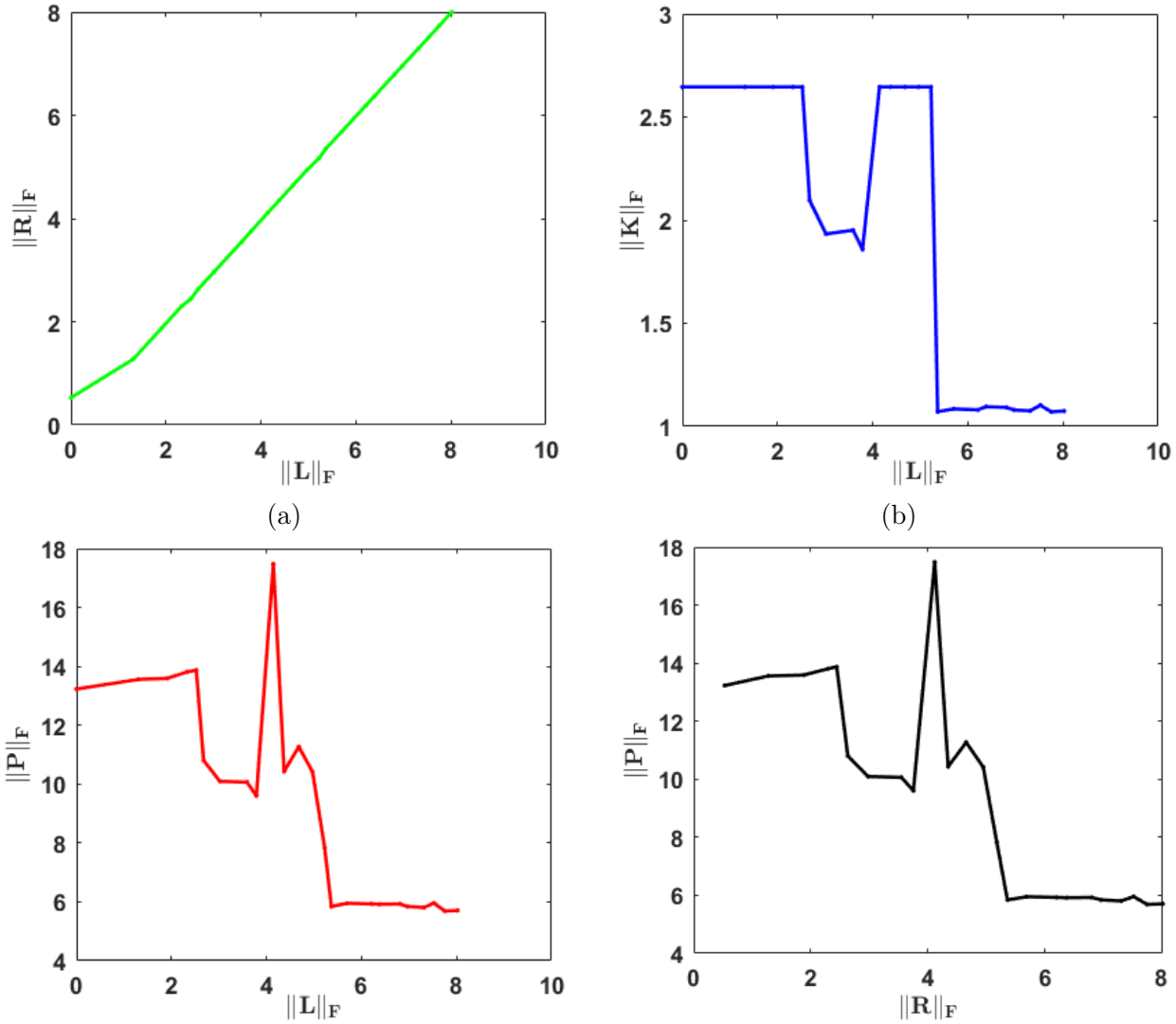
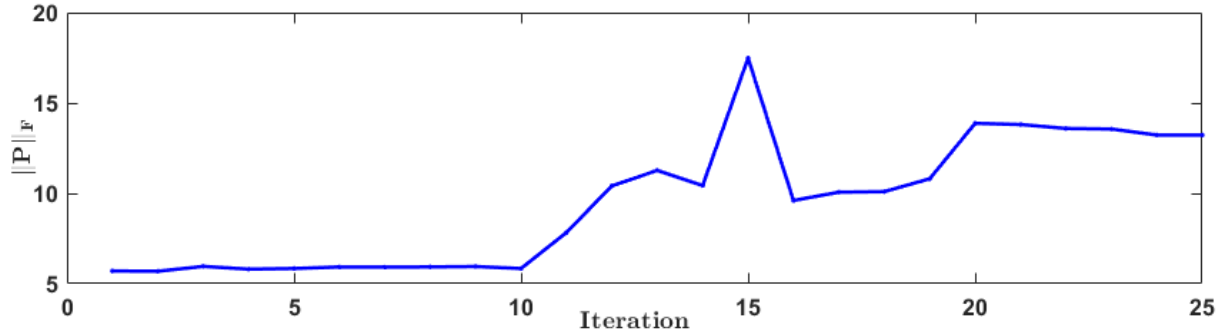
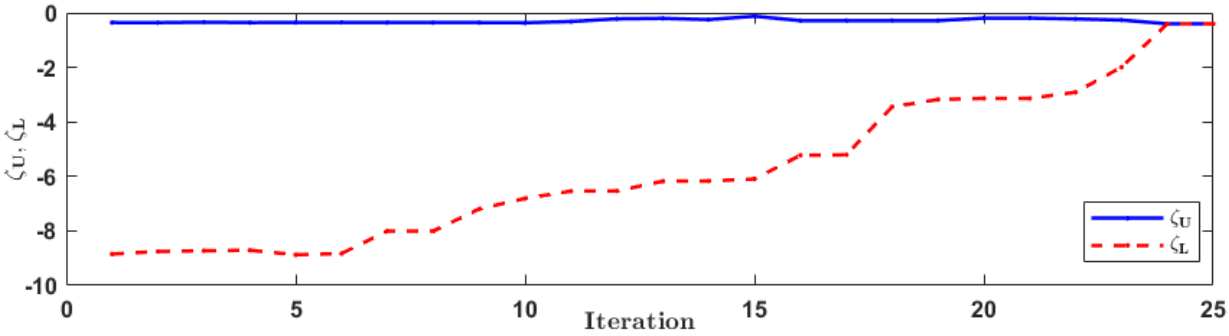


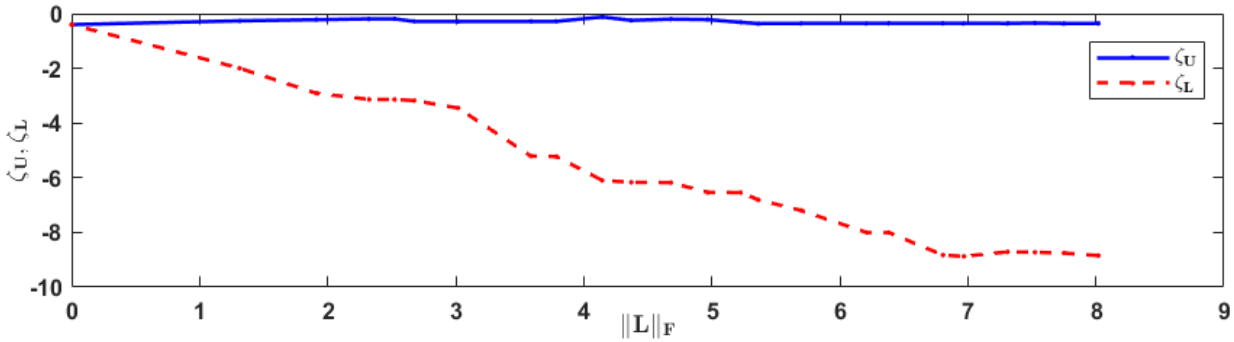
Figure 5.5: Variation of the modified Lyapunov equation matrix norms $\|R\|_F$, $\|K\|_F$, and the solution norm $\|P\|_F$ with $\|L\|_F$.



(a) Variation in $\|P\|_F$ with link removal.



(b) Variation in ζ_U, ζ_L with link disabling



(c) Variation in ζ_U, ζ_L with $\|L\|_F$

Figure 5.6: Variation of the solution norm $\|P\|_F$ with link disabling and the variation in convergence rate bounds ζ_U, ζ_L as further links are disabled, as well as with $\|L\|_F$ to illustrate the performance degradation despite no rank loss, as seen in Fig. 5.4a

.

5.2.1 The Simulation Scheme

The number of nodes in the undirected graph $\mathcal{G}$ is fixed at 7. The norm of the adjacency matrix is fixed to a particular value, $\|A\|_F = 20$ for this case. The number of non-zero links N is varied from 30% to 100% of the maximum possible amount of non-zero links without self-connections. For a given number of non-zero links, a random matrix is A is generated which contains N links and has the Frobenious norm exactly equal to the specified value. A ϵ-strong random DoS attack matrix E is generated using the scheme described in Assumption 5. The value of ϵ was set to 0.7 for this simulation, i.e. the maximum attack strength was chosen to be 70%.

The system is subjected to the DoS attack E. This process is repeated some m times for each value of N. m was chosen to be 10000 for this simulation. The parametric quantities recorded at each step are then averaged out to give the mean statistic variation with N. α was taken to be equal to 2, less than λ_2, so that the bound tightness can be analyzed as a consequences of the other properties alone. $Q = diag(2, 2, \ldots n\ times)$.

5.2.2 Spectral Properties of L and P

Figs. 5.7f, 5.7e, and 5.10a represent the spectral variation in L and the rank changes in L.

- It can be seen in Fig. 5.7f that with an increase in N, the eigenvalues of L tend to come closer to each other, which is accompanied by a reduction in performance degradation as discussed in the previous section. This clustering behaviour is explained by the theory of spectral clustering [37]. As number of link disablings required to disintegrate the graph into connected components (i.e. decrease the rank by 1) increases, the spectral-gap between the eigenvalues of L decreases. Therefore, if we have more non-zero links, the edge cut of the graph becomes bigger, thus decreasing the spectral-gap. It is hypothesised here that, the security improvement is a consequence of the fact that the no. of links added on an overall after being attacked as N increases must be positive which leads to the improvement in security despite the links being weaker.

- The previous point then leads to another hypothesis. The links are weak, but the $x_i - x_j$ term in protocol (2.5) can be arbitrarily large which is why it is possible that weaker links might not be causing a loss. In practical situations though, the link weights usually lead to a decrease in communication capacity, therefore, the exchange of values $(x_i - x_j)$ can be modeled by a saturation function at the simplest, whose saturation limits in turn must be directly proportional to the link weight such that weaker links lead to lesser information exchange. Will weaker but more links still be better than lesser but stronger links? This is an exciting problem which will be explored in future research.

- Note that the adding more but links resulting in a net increase in links post-attack is also suggested by the trend in the post-attack rank reduction (thus connectivity reduction). Figs. 5.10a and 5.7e show that the rank loss decreases as N increases. This also raises another question: What is the attack was so strong that it would have led to an overall decrease in links after attack? Will such strong attacks be feasible for the resource limited attacker? These are questions left to be answered in a future study.

- It is very interesting to see the increasing trend in $\|L\|_F$ with N, which is approximately linear near the middle. This might be a consequence of the inherent properties of norm bounded generation of digraphs.

Even more interesting than the observations about L, are the observations of the properties

of P:

- Fig. 5.8c reveals a very interesting $f(x) = c/x, c \in \mathbb{R}$ like curve. Such a seemingly strong resemblance to an elementary function suggests that there can be a way to analytically calculate the variation of $\|P\|_F$ with N, which, if successful, will certainly lead to stronger bounds for random DoS attacks. If such a result exists, it must depend upon how the links are selected i.e. the probability distribution of link selection, the $\|A\|_F$, α and Q. This is another direction to explore for future research.
- The eigenvalue trend in Fig. 5.8a resembles the clustering behavior of L's eigenvalues with one of the eigenvalues fixed at the 0.5. This behavior couldn't be explained by the end of the research term for this Thesis, and is therefore another topic for a future discussion. Note that, a decrease in the eigenvalues of P represents an increase in security since they will make the convergence bounds more negative.
- Fig. 5.9a seems to suggest that the shift in the eigenvalues of P decreases with N, which, on an average, suggests increased security since the shift is almost always in the degrading direction as suggested by section 5.1.2.

5.2.3 Change in Convergence Rate Bounds

The trends in Figs. 5.7a - Figs. 5.7d show that: With an increase in the number of non-zero links, the degradation of performance decreases. Note that, due to the nature of the attack scheme, for a given attack strength ϵ, the number of link failures by DoS attack should increase with N since more and more weights can be added to the matrix E as a consequence of Assumption 5. This increase in attack strength is illustrated by the increasing trend of $\|E\|_F$ in Fig. 5.10b. The same figure also confirms that the norm bounds are obeyed by the adjacency and attack matrices at each step.

So far, the previous sub-section on the properties of L and P have strongly suggested that an increase in the number of non-zero links increases security. *That is, having a denser topology with weaker links is more effective against norm bounded random link disabling DoS attacks as opposed to sparse topologies with stronger links.* If there is an element of truth to this statement, then Figs. 5.7a suggest a very interesting case: Observe that ζ_U is slightly bending towards the positive side insinuating the existence of *cases where the nominal consensus performance (i.e. without the presence of an adversary) decreases for denser topologies with weaker links.* The large sample size suggests that such cases are substantial, they exist and are out there. The class of such digraphs is seemingly not small and; therefore, is worth analyzing. *This class of digraphs may exhibit potential performance v/s security tradeoffs at the simulation's attack strength and weaker attacks since their nominal performance will tend to decrease while the performance under adversarial conditions increases with N.* This is at least true for the attack strength used in the simulation, the results might not be the same for stronger attacks and is something a researcher analyzing this problem should take in account, if possible.

5.2.4 The Heuristic Upper Bound

Fig. 5.9 suggests that the heuristic gets closer to the actual shift as N increases. However, below a threshold connectivity value, approx 40% in this case, the heuristic becomes arbitrarily large and no longer provides practical bounds. The ranges shown in the figure start at approx. 50%. Note that the difference between the heuristic bound and the actual is less than 10 for the most part. This suggests that the heuristic bound is quite accurate for undirected graphs with a high level of connectivity. The case of the undirected and the directed graph

is differentiated by the matrix separation term $sep(R^T(\alpha), -R(\alpha))$. Therefore, the case of the directed and the undirected graph won't differ for the most part. This is true for almost all of the figures which showed a very similar trend to that of the undirected graph. The search for stronger bounds will continue. The next step in this direction will be to find a numerically stable way of calculating the matrix separation term and use it to check the heuristic's admissibility in the general digraph case.

It can be inferred from equation (4.26), that the separation term has a strong influence on the bound validity. The separation term's can cause the bound to misbehave for smaller norms of $R(\alpha)$. $\|R(\alpha)\|_F$ linearly increases with $\|L\|_F$, as we shall see in the next chapter. We will also see that $\|L\|_F$ increases as the number of non-zero links increases for constant $\|A\|_F$ and, of course, for a given number of non-zero links $\|L\|_F$ increases with $\|A\|_F$. An estimate of this separation term is given by authors in [36] and important insights on its behavior are given in [34].

5.3 Conclusion

This brings this thesis to an end. It was a breathtaking pursuit to find the needle in the ocean of haystacks. While we still haven't found the needle we have narrowed down the search to a few haystacks, which is still a substantial improvement given the amount of generalization that was used throughout this text. While the author tried to cover as many aspects as possible, there a lot of things which are still needed to be considered to further narrow the search down. A majority of them are mentioned are throughout the text. The reader is encouraged to formulate their own hypotheses, simulate, draw inference, and pursue the direction thus obtained.

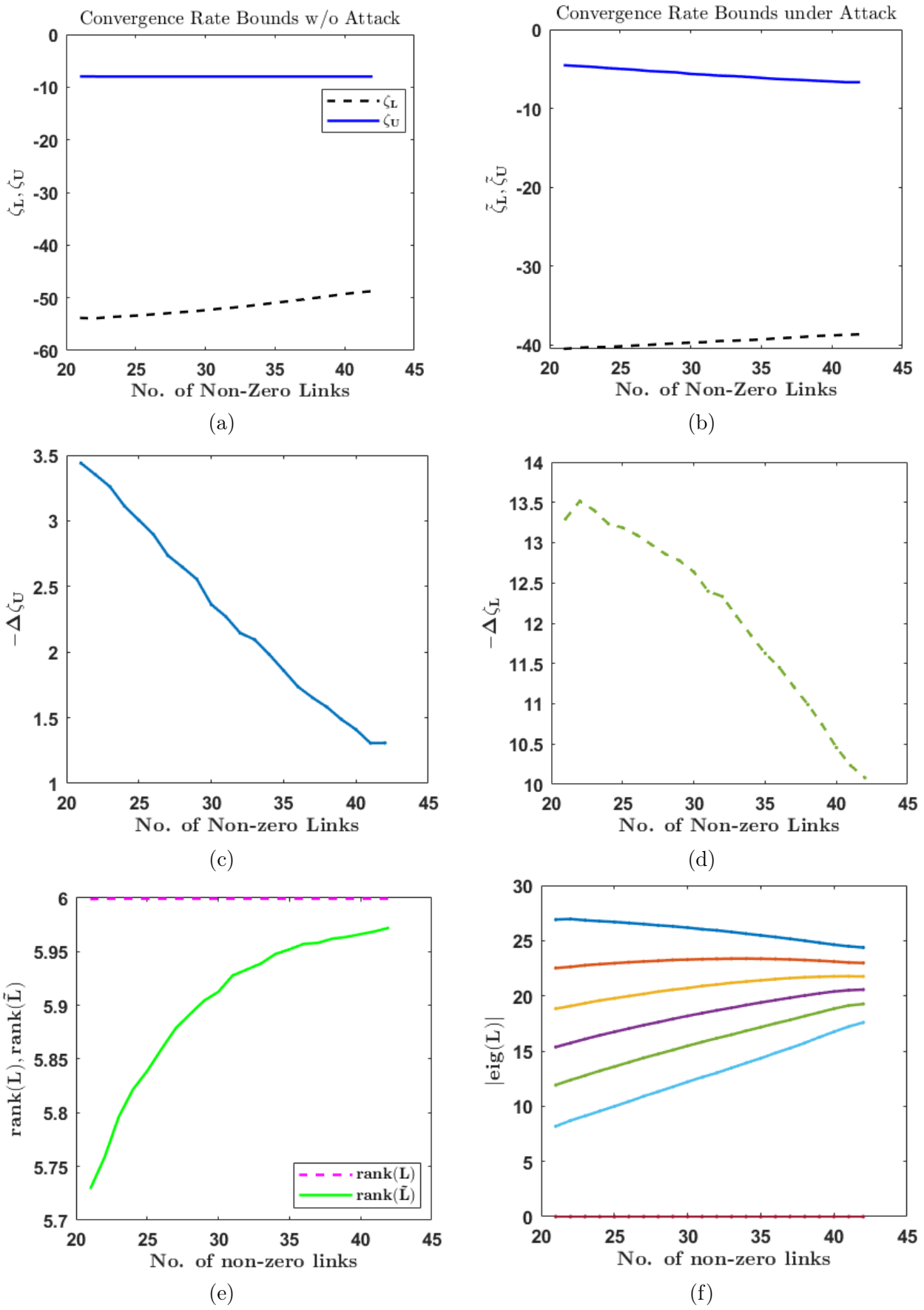
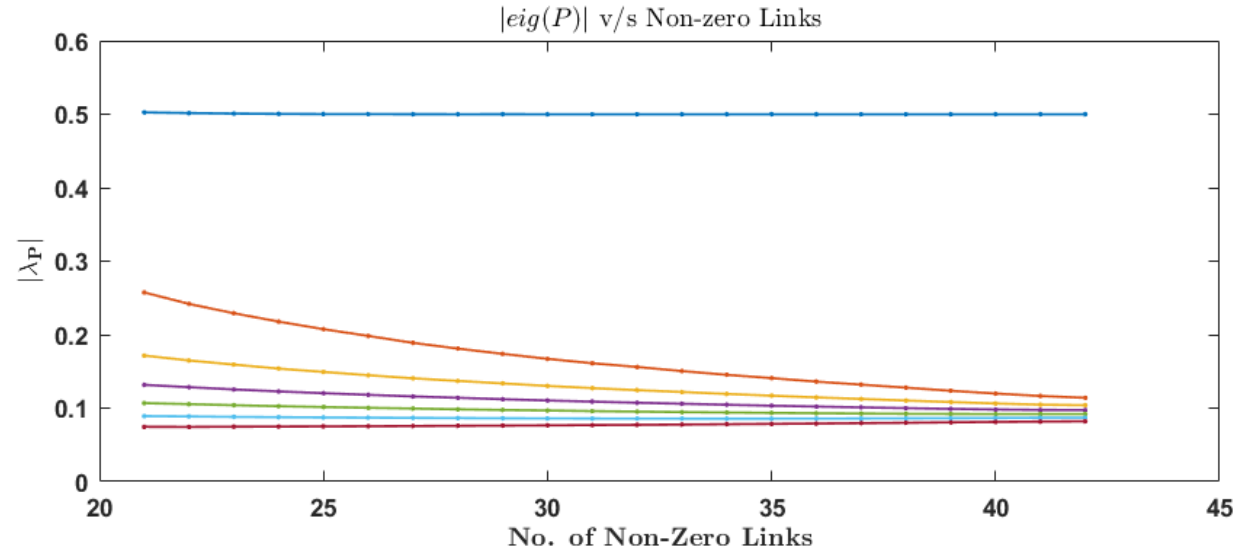
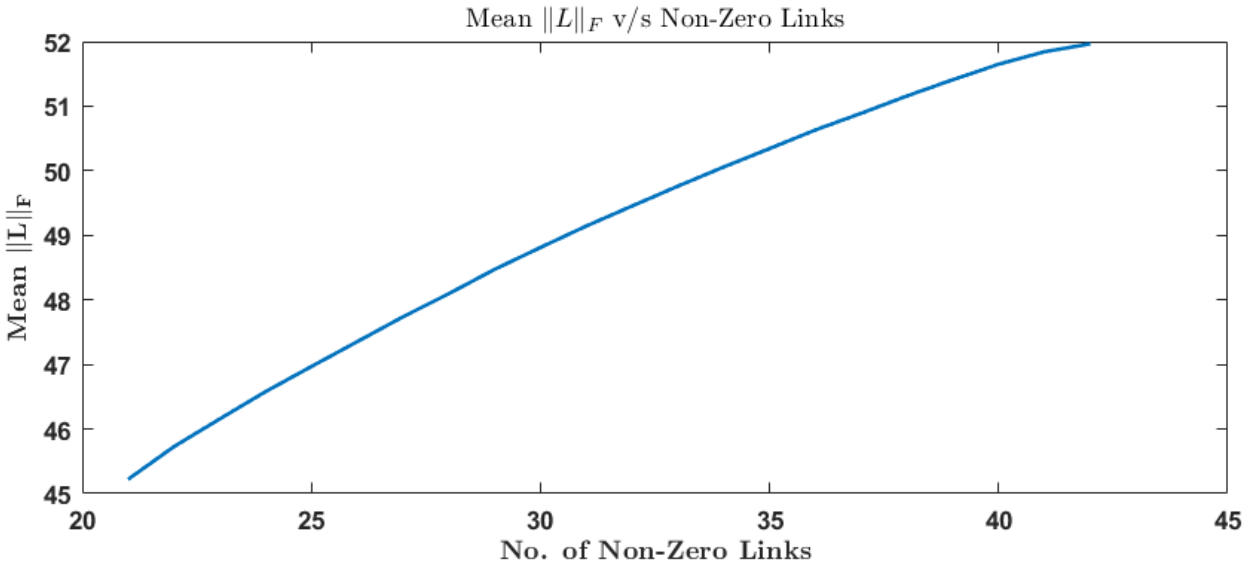


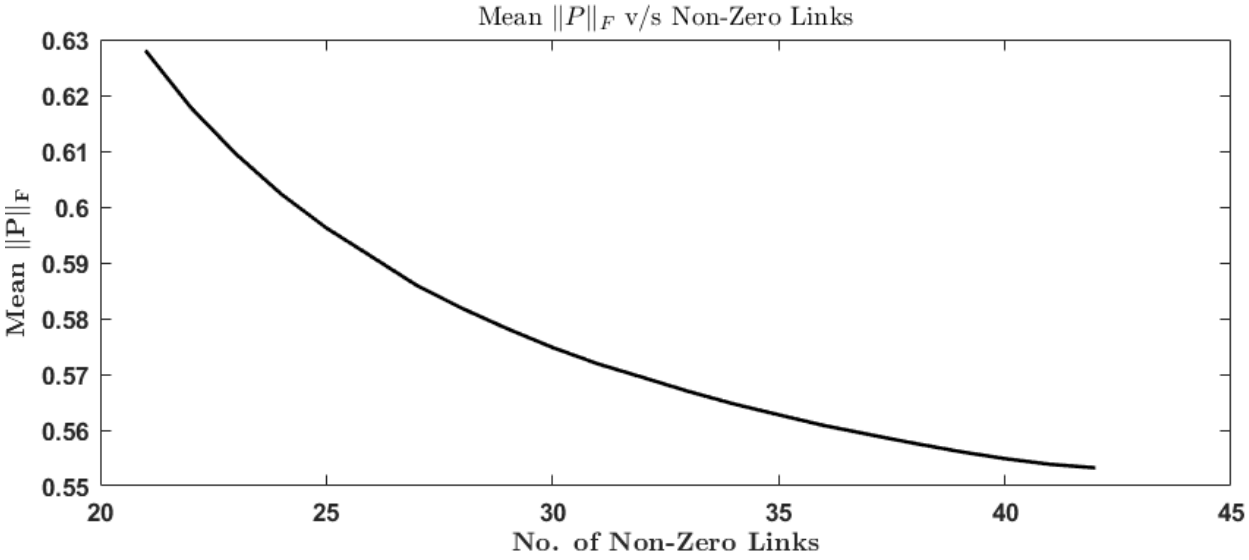
Figure 5.7: First four: Variation of the convergence rate bounds ζ_U, ζ_L, and their degradation $-\Delta\zeta_U, -\Delta\zeta_L$ (lesser value means lesser degradation) with no. of non-zero links at constant $\|A\|_F$ for an undirected graph. Bottom Two: Variation in the spectrum and rank of the graph laplacian as the no. of non-zero links increases. Note: $-\Delta\zeta_U = \tilde{\zeta}_U - \zeta_U$, and similarly for $\Delta\zeta_L$.



(a)

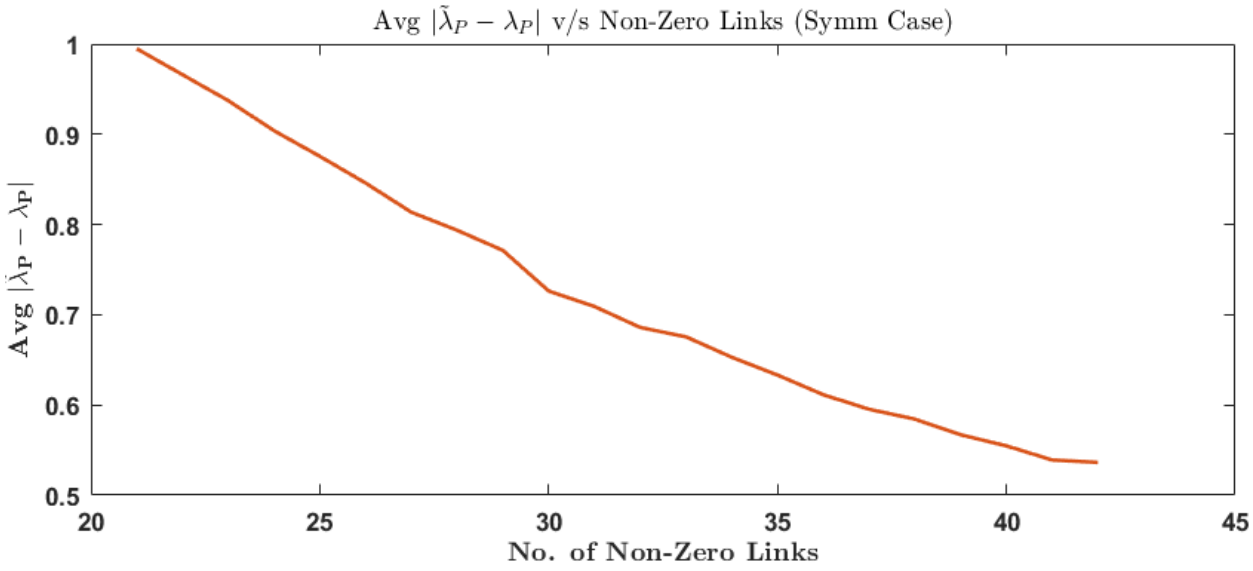


(b)

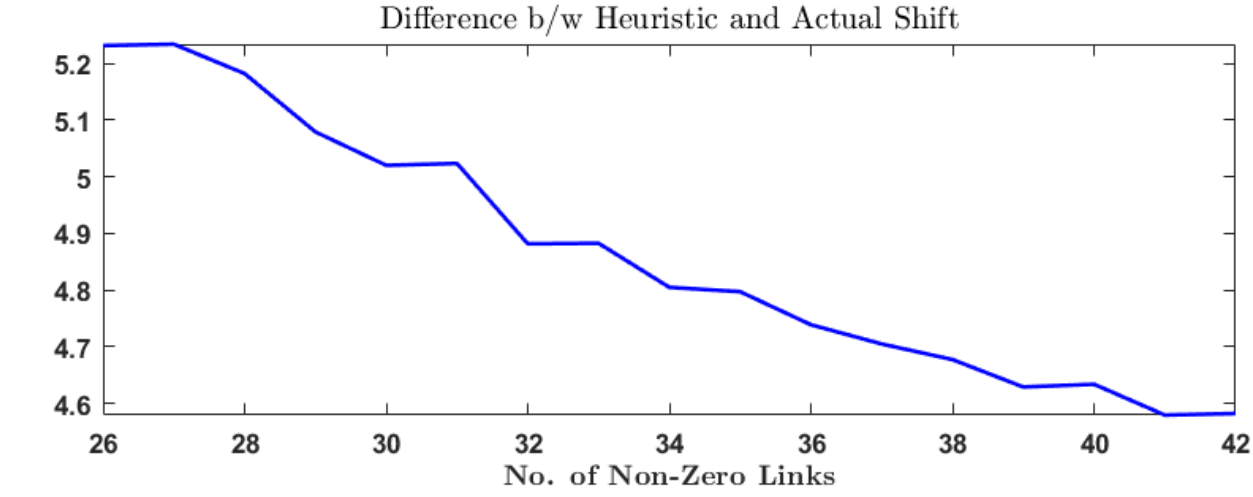


(c)

Figure 5.8: Spectral variation in P as the number of non-zero links are increases along-with the variation in $\|P\|_F$ and $\|L\|_F$ for undirected digraphs.

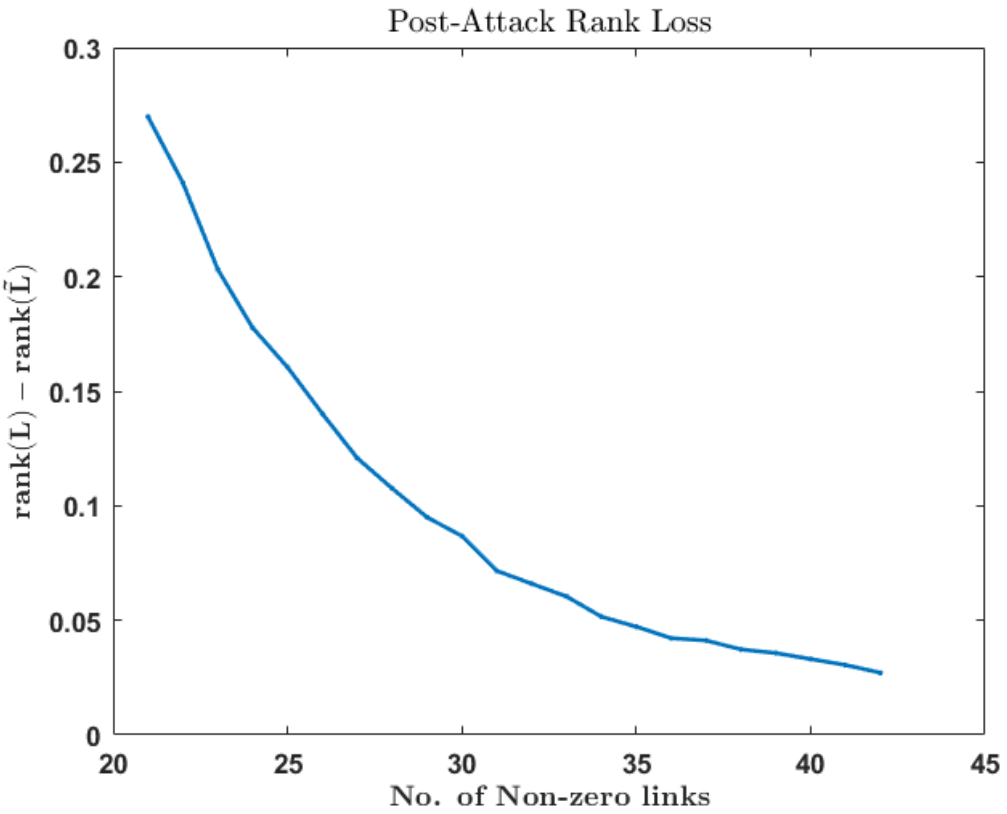


(a)

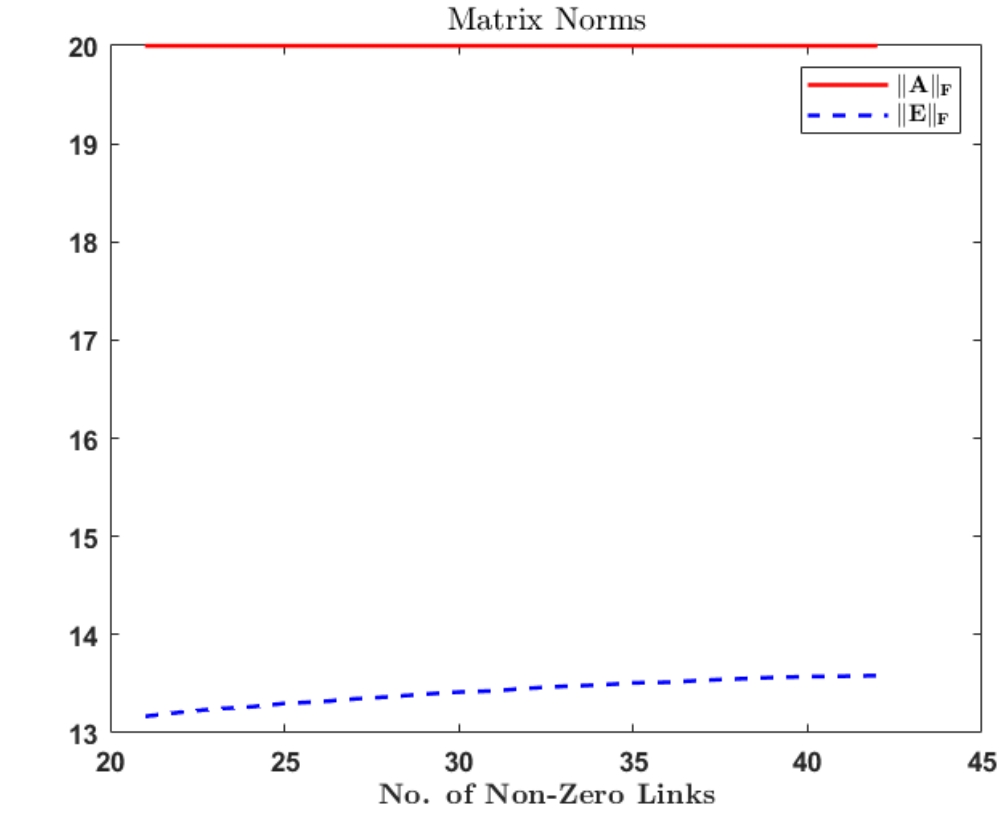


(b)

Figure 5.9: Change in the *average maximum eigenvalue shift* as the no. of non-zero links increases, and the change in heuristic closeness to the actual value for the undirected digraph case defined by Equation (4.29).



(a)



(b)

Figure 5.10: First four: Variation of the convergence rate bounds ζ_U, ζ_L, and their degradation $-\Delta\zeta_U, -\Delta\zeta_L$ (lesser value means lesser degradation) with no. of non-zero links at constant $\|A\|_F$ for an undirected graph. Bottom Two: Variation in the spectrum and rank of the graph laplacian as the no. of non-zero links increases.

Bibliography

[1] S. K. Khaitan and J. D. McCalley, "Design techniques and applications of cyberphysical systems: A survey," *IEEE Systems Journal*, vol. 9, no. 2, pp. 350–365, 2015. DOI: 10.1109/JSYST.2014.2322503.

[2] A. Humayed, J. Lin, F. Li, and B. Luo, "Cyber-physical systems security—a survey," *IEEE Internet of Things Journal*, vol. 4, no. 6, pp. 1802–1831, 2017. DOI: 10.1109/JIOT.2017.2703172.

[3] A. Teixeira, D. Pérez, H. Sandberg, and K. H. Johansson, "Attack models and scenarios for networked control systems," in *Proceedings of the 1st International Conference on High Confidence Networked Systems*, ser. HiCoNS '12, Beijing, China: Association for Computing Machinery, 2012, pp. 55–64, ISBN: 9781450312639. DOI: 10.1145/2185505.2185515. [Online]. Available: https://doi.org/10.1145/2185505.2185515.

[4] G. Deng, Y. Zhou, Y. Xu, T. Zhang, and Y. Liu, "An investigation of byzantine threats in multi-robot systems," ACM, Oct. 2021, pp. 17–32, ISBN: 9781450390583. DOI: 10.1145/3471621.3471867. [Online]. Available: https://dl.acm.org/doi/10.1145/3471621.3471867.

[5] A. Cetinkaya, H. Ishii, and T. Hayakawa, "An overview on denial-of-service attacks in control systems: Attack models and security analyses," *Entropy*, vol. 21, no. 2, 2019, ISSN: 1099-4300. DOI: 10.3390/e21020210. [Online]. Available: https://www.mdpi.com/1099-4300/21/2/210.

[6] I. Hwang, S. Kim, Y. Kim, and C. E. Seah, "A survey of fault detection, isolation, and reconfiguration methods," *IEEE Transactions on Control Systems Technology*, vol. 18, pp. 636–653, 3 May 2010, ISSN: 10636536. DOI: 10.1109/TCST.2009.2026285.

[7] D. Ding, Q. L. Han, Y. Xiang, X. Ge, and X. M. Zhang, "A survey on security control and attack detection for industrial cyber-physical systems," *Neurocomputing*, vol. 275, pp. 1674–1683, Jan. 2018, ISSN: 18728286. DOI: 10.1016/j.neucom.2017.10.009.

[8] F. Pasqualetti, F. Dorfler, and F. Bullo, "Attack detection and identification in cyber-physical systems," *IEEE Transactions on Automatic Control*, vol. 58, pp. 2715–2729, 11 2013, ISSN: 00189286. DOI: 10.1109/TAC.2013.2266831.

[9] I. Hwang, S. Kim, Y. Kim, and C. E. Seah, "A survey of fault detection, isolation, and reconfiguration methods," *IEEE Transactions on Control Systems Technology*, vol. 18, no. 3, pp. 636–653, 2010. DOI: 10.1109/TCST.2009.2026285.

[10] D. Ding, Q.-L. Han, Y. Xiang, X. Ge, and X.-M. Zhang, "A survey on security control and attack detection for industrial cyber-physical systems," *Neurocomputing*, vol. 275, pp. 1674–1683, 2018, ISSN: 0925-2312. DOI: https://doi.org/10.1016/j.neucom.2017.10.009. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0925231217316351.

[11] A. A. Cardenas, S. Amin, and S. Sastry, "Secure control: Towards survivable cyber-physical systems," in *2008 The 28th International Conference on Distributed Computing Systems Workshops*, 2008, pp. 495–500. DOI: 10.1109/ICDCS.Workshops.2008.40.

[12] M. Di Summa, A. Grosso, and M. Locatelli, "Branch and cut algorithms for detecting critical nodes in undirected graphs," *Computational Optimization and Applications*, vol. 53, no. 3, pp. 649–680, Dec. 2012. DOI: 10.1007/s10589-012-9458-y. [Online]. Available: https://doi.org/10.1007/s10589-012-9458-y.

[13] D. Acemoglu, A. Ozdaglar, and A. Tahbaz-Salehi, "Systemic risk and stability in financial networks," *American Economic Review*, vol. 105, no. 2, pp. 564–608, Feb. 2015. DOI: 10.1257/aer.20130456. [Online]. Available: https://www.aeaweb.org/articles?id=10.1257/aer.20130456.

[14] M. Rafiee and A. M. Bayen, "Optimal network topology design in multi-agent systems for efficient average consensus," Institute of Electrical and Electronics Engineers Inc., 2010, pp. 3877–3883, ISBN: 9781424477456. DOI: 10.1109/CDC.2010.5717719.

[15] Z. Liu, H. Zhang, P. Smith, and Q. Hui, "Optimizing weighted graph topology for robust network information dissemination," 2012, pp. 3329–3334. DOI: 10.1109/CDC.2012.6426594.

[16] K. K. Oh, M. C. Park, and H. S. Ahn, "A survey of multi-agent formation control," *Automatica*, vol. 53, pp. 424–440, Mar. 2015, ISSN: 00051098. DOI: 10.1016/j.automatica.2014.10.022.

[17] B. Genge, I. Kiss, and P. Haller, "A system dynamics approach for assessing the impact of cyber attacks on critical infrastructures," *International Journal of Critical Infrastructure Protection*, vol. 10, pp. 3–17, Sep. 2015, ISSN: 18745482. DOI: 10.1016/j.ijcip.2015.04.001.

[18] M. Zhu and S. Martínez, "Attack-resilient distributed formation control via online adaptation," 2011, pp. 6624–6629, ISBN: 9781612848006. DOI: 10.1109/CDC.2011.6161327.

[19] S. Nikoletseas, G. Prasinos, P. Spirakis, and C. Zaroliagis, "Attack propagation in networks," vol. 36, Sep. 2003, pp. 553–574. DOI: 10.1007/s00224-003-1087-5.

[20] C. Z. Bai, F. Pasqualetti, and V. Gupta, "Data-injection attacks in stochastic control systems: Detectability and performance tradeoffs," *Automatica*, vol. 82, pp. 251–260, Aug. 2017, ISSN: 00051098. DOI: 10.1016/j.automatica.2017.04.047.

[21] M. M. Zavlanos, M. B. Egerstedt, and G. J. Pappas, "Graph-theoretic connectivity control of mobile robot networks," vol. 99, Institute of Electrical and Electronics Engineers Inc., 2011, pp. 1525–1540. DOI: 10.1109/JPROC.2011.2157884.

[22] H. Zhang, Y. Shu, P. Cheng, and J. Chen, "Privacy and performance trade-off in cyber-physical systems," *IEEE Network*, vol. 30, pp. 62–66, 2 Mar. 2016, ISSN: 08908044. DOI: 10.1109/MNET.2016.7437026.

[23] R. Olfati-Saber and R. Murray, "Consensus problems in networks of agents with switching topology and time-delays," *IEEE Transactions on Automatic Control*, vol. 49, no. 9, pp. 1520–1533, 2004. DOI: 10.1109/TAC.2004.834113.

[24] S. Di Cairano, A. Pasini, A. Bemporad, and R. Murray, "Convergence properties of dynamic agents consensus networks with broken links," in *2008 American Control Conference*, 2008, pp. 1362–1367. DOI: 10.1109/ACC.2008.4586682.

[25] C. W. Wu, "Algebraic connectivity of directed graphs," *Linear and Multilinear Algebra*, vol. 53, no. 3, pp. 203–223, 2005. DOI: 10.1080/03081080500054810. eprint: https://doi.org/10.1080/03081080500054810. [Online]. Available: https://doi.org/10.1080/03081080500054810.

[26] F. Pasqualetti, F. Dörfler, and F. Bullo, "Attack detection and identification in cyber-physical systems," *IEEE Transactions on Automatic Control*, vol. 58, no. 11, pp. 2715–2729, 2013. DOI: 10.1109/TAC.2013.2266831.

[27] "On fault identification," *Model-based Fault Diagnosis Techniques*, pp. 441–461, DOI: 10.1007/978-3-540-76304-8_14.

[28] M. Zhu and S. Martínez, "Attack-resilient distributed formation control via online adaptation," in *2011 50th IEEE Conference on Decision and Control and European Control Conference*, 2011, pp. 6624–6629. DOI: 10.1109/CDC.2011.6161327.

[29] W. Ren and R. Beard, "Consensus seeking in multiagent systems under dynamically changing interaction topologies," *IEEE Transactions on Automatic Control*, vol. 50, no. 5, pp. 655–661, 2005. DOI: 10.1109/TAC.2005.846556.

[30] E. Panteley, A. Loría, and S. Sukumar, "Strict lyapunov functions for consensus under directed connected graphs," in *2020 European Control Conference (ECC)*, 2020, pp. 935–940. DOI: 10.23919/ECC51009.2020.9143719.

[31] R. Bhatia, *Perturbation Bounds for Matrix Eigenvalues*. Society for Industrial and Applied Mathematics, 2007. DOI: 10.1137/1.9780898719079. eprint: https://epubs.siam.org/doi/pdf/10.1137/1.9780898719079. [Online]. Available: https://epubs.siam.org/doi/abs/10.1137/1.9780898719079.

[32] D. Boley and B. N. Datta, "Numerical methods for linear control systems," 1994.

[33] G. Golub, S. Nash, and C. Van Loan, "A hessenberg-schur method for the problem ax + xb= c," *IEEE Transactions on Automatic Control*, vol. 24, no. 6, pp. 909–913, 1979. DOI: 10.1109/TAC.1979.1102170.

[34] J. M. Varah, "On the separation of two matrices," *SIAM Journal on Numerical Analysis*, vol. 16, no. 2, pp. 216–222, 1979, ISSN: 00361429. [Online]. Available: http://www.jstor.org/stable/2156829.

[35] G. W. Stewart, "Error and perturbation bounds for subspaces associated with certain eigenvalue problems," *SIAM Review*, vol. 15, no. 4, pp. 727–764, 1973, ISSN: 00361445. [Online]. Available: http://www.jstor.org/stable/2028728.

[36] B. Kågström and P. Poromaa, "Lapack-style algorithms and software for solving the generalized sylvester equation and estimating the separation between regular matrix pairs," vol. 22, no. 1, pp. 78–103, Mar. 1996, ISSN: 0098-3500. DOI: 10.1145/225545.225552. [Online]. Available: https://doi.org/10.1145/225545.225552.

[37] A. Ng, M. Jordan, and Y. Weiss, "On spectral clustering: Analysis and an algorithm," in *Advances in Neural Information Processing Systems*, T. Dietterich, S. Becker, and Z. Ghahramani, Eds., vol. 14, MIT Press, 2002. [Online]. Available: https://proceedings.neurips.cc/paper/2001/file/801272ee79cfde7fa5960571fee36b9b-Paper.pdf.

Fin