Robustness of controllability and recoverability of complex networks

Sun, P.

DOI
10.4233/uuid:29dc55b9-f992-4246-a340-51e40d823451

Publication date
2022

Document Version
Final published version

Citation (APA)
Sun, P. (2022). *Robustness of controllability and recoverability of complex networks*. [Dissertation (TU Delft), Delft University of Technology]. https://doi.org/10.4233/uuid:29dc55b9-f992-4246-a340-51e40d823451

ROBUSTNESS OF CONTROLLABILITY AND RECOVERABILITY OF COMPLEX NETWORKS

ROBUSTNESS OF CONTROLLABILITY AND RECOVERABILITY OF COMPLEX NETWORKS

Dissertation

for the purpose of obtaining the degree of doctor
at Delft University of Technology,
by the authority of the Rector Magnificus, Prof.dr.ir. T.H.J.J. van der Hagen,
chair of the Board for Doctorates,
to be defended publicly on
Thursday 13 January 2022 at 12:30 o'clock

by

Peng SUN

Master of Engineering in Communication and Information Systems,
Shandong University, China,
born in Shandong, China.

This dissertation has been approved by the promotors.

Composition of the doctoral committee:

Rector Magnificus	chairperson
Prof. dr. ir. P. F. A. Van Mieghem	Delft University of Technology, promotor
Prof. dr. ir. R. E. Kooij	Delft University of Technology, promotor
Independent members:	
Prof. dr. M. E. Warnier	Delft University of Technology
Prof. dr. J. L. van den Berg	University of Twente
Prof. dr. J. L. Marzo	University of Girona, Spain
Dr. R. Bouffanais	University of Ottawa, Canada
Dr. J. L. A. Dubbeldam	Delft University of Technology
Prof. dr. ir. S. Hamdioui	Delft University of Technology, reserve member

Keywords: Complex networks, Network Resilience, Network Robustness, Network Controllability
Printed by: ProefschriftMaken

Front & Back: Peng Sun
Published by: Peng Sun
Email: sunpeng0626@hotmail.com

ISBN 978-94-6423-609-5

An electronic version of this dissertation is available at
http://repository.tudelft.nl/.

To my family

CONTENTS

SUMMARY

Network robustness describes a network's ability to provide and maintain an acceptable level of service in the face of failures and challenges to normal operation. Unfortunately, failures of networks, such as power outages in power systems, congestions in transportation networks, failures of routers on the Internet, happen frequently in our daily life and introduce a tremendous cascading effect on our society. We naturally expect that these networks have high robustness to maintain their performance in face of failures or attacks. As the first step, it is vital to investigate and analyze the robustness of networks so as to propose effective methods to improve network robustness.

The first part of the thesis mainly focuses on the robustness of network controllability in face of topological perturbations. In Chapter 2, we propose closed-form analytic approximations for the minimum number of driver nodes which denotes the controllability of the network. Inspired by the concept of critical links, we deduce and validate our approximations on both real-world and synthetic networks. We show that when the fraction of removed links is small, our approximations perform well. Besides, we also find that the critical link attack is the most effective among 4 considered attacks, as long as the fraction of removed links is smaller than the fraction of critical links. In Chapter 3, we focus on the controllability of swarm signalling networks with regular out-degree and bi-modal out-degree distribution. We deduce the generating functions in random failure process and then estimate the fraction of driver nodes with simulations. Results show that our estimations have high accuracy in predicting the fraction of driver nodes in case of random link failures. In order to further improve the accuracy of our proposed approximations in Chapter 4, we use a machine learning method to decrease the gap between our analytical approximations and the simulation results. We compare our approximations obtained by machine learning with existing analytical approximations and show that our approximations significantly outperform the existing closed-form analytical approximations in both synthetic and real-world networks. Apart from targeted attacks based upon the removal of critical links, we also propose analytical approximations for out-in degree-based attacks. In Chapter 5, we investigate the reachability-based robustness of controllability considering link-based random attack, targeted attack, as well as random attack under the protection of critical links. We validate our approximations using 200 real-world communication networks and some synthetic networks and find that our approximations perform well in most cases.

In the second part of the thesis, we work on the recoverability of networks. The recoverability of networks refers to the ability of a network to return to a desired performance level after suffering topological perturbations such as link failures. In Chapter 6, we propose a general topological approach and two recoverability indicators to measure the network recoverability for optical networks for two recovery scenarios. Furthermore, we employ the proposed approach to assess 20 real-world optical networks. Numerical results show that the network recoverability is coupled to the network topology, the robust-

ness metric and the recovery strategy. We also find that assortativity, which denotes the tendency of network nodes to connect preferentially to other nodes with similar degree, has the strongest correlation with both recoverability indicators. In Chapter 7, we adopted the framework of network recoverability and investigate the recoverability of network controllability for two recovery scenarios. We employ the proposed approach to assess swarm signalling networks with regular out-degree, and networks with bi-modal out-degree distributions. Besides, we also deduced the analytical results of the recoverability indicators by generating functions, which are close to the results based on simulations. In Chapter 8, we conclude this thesis and come up with some future work.

SAMENVATTING

De robuustheid van een netwerk beschrijft het vermogen van het netwerk om een acceptabel serviceniveau te bieden en te behouden in het geval van storingen en aanvallen op het netwerk. Helaas komen storingen van netwerken, zoals stroomuitval in energiesystemen, congestie in transportnetwerken en pakketverlies op internet, regelmatig voor in het dagelijks leven en hebben soms een aanzienlijke impact op onze samenleving. We verwachten natuurlijk dat deze netwerken een hoge robuustheid hebben, zodat het prestatieniveau behouden blijft bij storingen of aanvallen. Als eerste stap is het van vitaal belang om de robuustheid van netwerken te onderzoeken en om effectieve methoden voor te stellen om de netwerkrobuustheid te verbeteren.

Het eerste deel van het proefschrift richt zich voornamelijk op de robuustheid van netwerkbeheersbaarheid in het geval van topologische verstoringen. In Hoofdstuk 2 stellen we analytische benaderingen voor, in gesloten vorm, waarmee het minimum aantal z.g.n. driver nodes kan worden bepaald, nodig om het netwerk beheersbaar te maken. Geïnspireerd door het concept van kritieke verbindingen, deduceren en valideren we onze benaderingen op zowel echte als synthetische netwerken. We laten zien dat wanneer de fractie verwijderde links klein is, onze benaderingen zeer nauwkeurig zijn. Daarnaast tonen we ook aan dat de kritieke link-aanval het meest effectief is van 4 beschouwde aanvallen, zolang de fractie verwijderde links kleiner is dan de fractie kritieke links. In Hoofdstuk 3 richten we ons op de beheersbaarheid van zwerm-signaleringsnetwerken met reguliere uitgaande graad en bi-modale uitgaande graad distributie. We leiden genererende functies af in het geval van willekeurige uitval van verbindingen en schatten vervolgens de fractie van benodigde driver nodes m.b.v. simulaties. De resultaten tonen aan dat onze schattingen een hoge nauwkeurigheid hebben bij het voorspellen van de fractie van driver nodes, in het geval van willekeurige uitval van verbindingen. Om de nauwkeurigheid van onze voorgestelde benaderingen verder te verbeteren, gebruiken we in hoofdstuk 4 machine learning om de kloof tussen onze analytische benaderingen en de simulatieresultaten te verkleinen. We vergelijken onze benaderingen die zijn verkregen m.b.v. machine learning met bestaande analytische benaderingen en laten zien dat onze benaderingen aanzienlijk beter presteren dan de bestaande analytische benaderingen in gesloten vorm in zowel synthetische als echte netwerken. Naast gerichte aanvallen op basis van het verwijderen van kritieke links, stellen we ook een analytische benadering op, voor aanvallen gebaseerd op de uitgaande graad. In Hoofdstuk 5 onderzoeken we de op bereikbaarheid gebaseerde robuustheid van beheersbaarheid. We beschouwen hierbij willekeurige aanvallen op verbindingen, gerichte aanvallen en willekeurige aanvallen op verbindingen waarbij kritieke verbindingen worden beschermd. We valideren onze benaderingen met behulp van 200 communicatienetwerken en enkele synthetische netwerken en vinden dat onze benaderingen in de meeste gevallen nauwkeurig zijn.

In het tweede deel van het proefschrift richten we ons op de herstelbaarheid van netwerken. De herstelbaarheid van netwerken verwijst naar het vermogen van een netwerk

om terug te keren naar een gewenst prestatieniveau na topologische verstoringen zoals het verwijderen van verbindingen. In Hoofdstuk 6 stellen we een algemene topologische benadering voor, alsmede indicatoren, om de herstelbaarheid van netwerken voor optische netwerken voor twee herstelscenario's te kwantificeren. Verder gebruiken we de voorgestelde aanpak om de herstelbaarheid van 20 echte optische netwerken te bepalen. Numerieke resultaten tonen aan dat de herstelbaarheid van een netwerk is gekoppeld aan de netwerktopologie, de robuustheidsmetriek en de herstelstrategie. We vinden ook dat assortativiteit de sterkste correlatie heeft met beide indicatoren voor herstelbaarheid. In Hoofdstuk 7 hebben we het raamwerk van netwerkherstelbaarheid overgenomen en de herstelbaarheid van netwerkbeheersbaarheid onderzocht voor twee herstelscenario's. We gebruiken de voorgestelde aanpak om zwerm-signaleringsnetwerken met reguliere uitgaande graad en netwerken met bimodale uitgaande graad distributies te analyseren. Daarnaast hebben we ook de analytische resultaten van de herstelbaarheidsindicatoren afgeleid door gebruik te maken van genererende functies. De verkregen resultaten liggen dicht bij uitkomsten op basis van simulaties. In hoofdstuk 8 zetten we de resultaten van dit proefschrift op een rij. Verder doen we enkele suggesties voor toekomstig onderzoek.

1

INTRODUCTION

I have not failed. I've just found 10,000 ways that won't work.

Thomas A. Edison

NETWORKS exist everywhere and have deeply integrated into our daily life. Examples around us and even inside of us include social networks (Facebook, Twitter, LinkedIn) [1], the Internet, transportation networks (airline, metro, train and bus networks) [2], telecommunication networks [3], neural networks [4], vascular networks [5] and so on.

The routine of our life and the society depends much on the normal operation of these networks. However, networks in real-life are frequently faced with failures or malicious attacks which degrade the performance of networks [6], such as the Denial-of-Service (DoS) attack on the Internet [7], cascading failures on the power grid [8], connection failures in telecommunication networks [9], etc. Naturally, we hope that these networks are robust enough to maintain their functionality under external perturbations, which leads to the research question, how can we characterize, measure and improve the robustness of networks?

Graph theory provides powerful approaches to investigate the robustness of networks. The history of the development of graph theory is presented briefly. In 1736, Leonhard Euler solved the Seven Bridges of Königsberg problem, which is regarded as the origin of graph theory. In 1959, Paul Erdős and Alfréd Rényi introduced probability theory into graph theory [10] and established the random graph theory. In 1998, Watts and Strogatz [11] discovered small-world phenomenon in numerous real-world networks and proposed a model to generate small-world networks. In 1999, Albert and Barabási [12] discovered the scale-free property in the Internet, whose degree distribution follows a power law. The Barabási–Albert model was proposed to generate scale-free networks, which adopts the preferential attachment mechanism in the network growth process.

1

Network robustness has been extensively investigated for decades from the perspective of graph theory. In 2000, Albert *et al.* [13] investigated the robustness of complex networks in face of node-based failures. The results show that scale-free networks are robust to random failures but extremely vulnerable to targeted attacks. The percolation model was employed to analytically measure the robustness of networks [14], [15] followed by a series of studies [16], [17]. In 2001, Newman *et al.* applied the generating functions [18] to the percolation model in random graphs with arbitrary degree distribution [19]. Wang *et al.* [20] derived the upper and lower bounds of the effective graph resistance under topological changes and illustrated a novel comparison method by considering the distance between the added or removed links. He *et al.* [21] proposed an approach on network modeling and robustness assessment for multimodal freight transport networks and found that the node criticality resembles a power-law distribution which implies a relatively robust state of the network against single random disruptions. In recent years, the research on robustness has switched to interdependent networks and focuses on analyzing the interconnection patterns between networks. Huang *et al.* [22] introduced a general technique which maps the targeted-attack problem in interdependent networks to the random-attack problem in a transformed pair of interdependent networks. Results showed that when the highly connected nodes are protected and have lower probability to fail in contrast to single scale-free networks. Parandehgheibi *et al.* [23] focused on the interdependency between the power grid and communication networks and developed heuristics to find a near-optimal solution to the minimum number of node failures needed to cause total blackout.

There is still lack of consensus in the definition of robustness and the approach for robustness assessment, which is due to the diversity of service and demand in real-world networks. For example, a short end-to-end delay is desired in telecommunication networks while a good connectivity is expected in transportation networks. In this thesis, we focus on the robustness of network controllablity as described in Section 1.1.

1.1. NETWORK CONTROLLABILITY

Controllability as one of the fundamental concepts in control theory, quantifies the ability to steer a dynamical system from an arbitrary initial state to an arbitrary terminal state in finite time [24]. In some networked systems, a set of proper input can control the state of the whole system. For example, in a bee colony, 5% of the bees are enough to guide the entire population towards a new beehive [25]. We consider linear, time-invariant dynamics on a directed network, which are described by:

$$\frac{dx(t)}{dt} = Ax(t) + Bu(t) \tag{1.1}$$

where the $N \times 1$ vector $x(t) = (x_1(t), x_2(t), ..., x_N(t))^T$ denotes the state of the system with N nodes at time t. The weighted matrix A is an $N \times N$ matrix which describes the network topology and the interaction strength between the components. The $N \times M$ matrix B is the input matrix which identifies the $M \leq N$ driver nodes controlled by outside input signals. The $M \times 1$ vector $u(t) = (u_1(t), u_2(t), ..., u_M(t))^T$ is the input signal vector. A driver node $j \in \{1, \ldots, M\}$ has an input signal u_j that is externally fed in $u_j(t)$. The fraction n_D

of driver nodes is normally chosen as the metric that measures the controllability of a network, a smaller the fraction n_D denotes a higher controlllability of a network.

The linear system defined by Eq.(1.1) is controllable, if and only if the $N \times NM$ controllability matrix:

$$C = (B, AB, A^2B, ..., A^{N-1}B) \tag{1.2}$$

has full rank, i.e., $rank(C) = N$. This criterion is called Kalman's controllability rank condition [26]. The rank of matrix C provides the dimension of the controllable subspace of the system. We need to choose the right input matrix B consisting of a minimum number of driver nodes to assure that the controllability matrix C has full rank. System (1.1) is said to be structurally controllable if it is possible to fix the non-zero parameters in A and B in such a way that the obtained system (A, B) satisfies Kalman's rank condition. We assume that the network described by A has no self-loops, i.e. all entries on the diagonal of A are zero.

The minimum number of driver nodes needed for structural controllability can be obtained through the "maximum matching" of the network. We define the source node of a directed link as the node from which the link originates and the target node as the node where the link terminates. A maximum matching of a directed network is a maximum set of links that do not share source or target nodes [27], as illustrated in Figure 1.1(a). Such links are named "matching links". Target nodes of matching links are matched nodes while the other nodes are unmatched nodes. For a given maximum matching, the unmatched nodes are the driver nodes needed for controlling the network.

In order to find the maximum number of matching links, so as to determine the minimum number of driver nodes N_D, a directed network G with N nodes and L links can be converted into a bipartite graph $B_{N,N}$ with $2N$ nodes and L links, as shown in Figure 1.1(b). A maximum matching in a bipartite graph can be obtained efficiently using the Hopcroft-Karp algorithm [28]. The Hopcroft-Karp algorithm guarantees to return the minimum number of driver nodes to completely control the network. The computational complexity of the Hopcroft-Karp algorithm to find all driver nodes is $O(\sqrt{N}L)$. In our simulations, we use the algorithm mentioned above to determine the number N_D of driver nodes and then get the fraction n_D of driver nodes which equals N_D/N.

1.2. RECOVERABILITY OF COMPLEX NETWORKS

As mentioned before, most research on network robustness focuses on the network's ability to withstand perturbations such as failures or attacks. However, the recovery process and methods are not considered in the degraded network after failures or attacks. In recent years, more and more research focuses on proposing recovery strategies and measuring their efficiency in restoring the performance of networks. Zhang *et al.* [29] introduced a resilience-based framework to optimise the scheduling of the post-disaster recovery strategies for road-bridge transportation networks. Two metrics were proposed for measuring rapidity and efficiency of the network recovery: total recovery time (TRT) and the skew of the recovery trajectory (SRT). Sun *et al.* [30] proposed two novel recovery strategies: Cold Backup Service Replacement Strategy (CBSRS) and Hot Backup Service Replacement Strategy (HBSRS). Results showed that the proposed strategies significantly improve the performance of service composition and effectively guarantee the

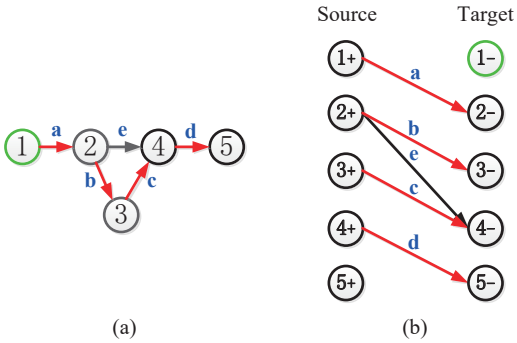


Figure 1.1: Driver nodes and critical links in a directed network G. (a) An example network G with $N = 5$ nodes and $L = 5$ directed links. Since link a, b, c and d are the maximum set of links that do not share source and target nodes, these links are matching link. Target nodes of these matching links, i.e., node 2, 3, 4 and 5, are matched nodes. Node 1 is an unmatched node. (b) The bipartite graph with $2N$ nodes and L links. Matching links are highlighted in red in the bipartite graph. Driver nodes are highlighted in green. To create the bipartite graph, each node V in the original network G will be translated into source node $V+$ and target node $V-$ in the bipartite graph. The first column of the bipartite graph are all possible source nodes, whereas the nodes in the second column are all possible target nodes. Links in the bipartite graph are determined by the directed links in the original network G. By using the Hopcroft-Karp algorithm, a maximum set of matching links can be found in the bipartite graph. None of the matching links share a common source or target node. Then, the target nodes of matching links are matched nodes. Other target nodes are unmatched nodes, which are also driver nodes.

availability and reliability of service composition in dynamic network. Almoghathawi *et al.* [31] studied the interdependent network restoration problem (INRP) and proposed a resilience-driven multi-objective optimization model which considers partial disruptions and recovery of the disrupted network components, and partial dependence between nodes in different networks. In [32] [33] [34], a comprehensive methodology for topology generation was proposed, and the analytical and experimental techniques used for evaluating the network attributes depend on the efficiency of the recovery strategy. The efficiency of a recovery strategy indicates how much network performance can be recovered in a given number of edges or nodes once they are reconnected. Di Muro *et al.* [35] proposed a recovery strategy to repair the nodes in the mutual boundary of functional clusters in two interdependent networks based on a critical probability of recovery above which the system is restored and below which it collapses.

Though a large amount of recovery strategies were proposed to restore the performance of networks, the work above did not consider the failure or the attack process when designing recovery strategies. In real life, we hope that networks are difficult to destroy but easy to recover which suggests that we should consider both the failure process and the recovery process to design a resilient network. Networks preferably have the ability to return to a desired performance level after suffering malicious attacks or random failures. We define such network capability as network recoverability and introduce the details in Section 6.

1.3. THESIS OBJECTIVES AND OUTLINE

This thesis is motivated by the increasing importance of robust networks in real life. We regard network robustness as a beneficial property for real-world networks which provide daily service to people. During random failures or malicious attacks, the performance of real-world networks degrades which affects our daily life.

The aims of this dissertation are to model and analyse real-world networks and their robustness in terms of failures, malicious attacks. Thus, the main questions this thesis aims to answer are:

Chapter 2: Can we estimate the impact of random failures or targeted attacks on network controllability? Which type of network is more resilient to attacks? Which attack strategy is the most harmful?

Chapter 3: Can we propose analytical approximations for networks with specific degree distribution? If so, how is the performance of the approximations?

Chapter 4: Can we use machine learning methods to further improve our approximations for the number of driver nodes in random failures or targeted attack? What network metrics should be chosen as features? Are the approximations obtained by machine learning always better than other existing methods?

Chapter 5: What is the difference between the reachability-based controllability and the control-based controllability? How can we measure the reachability-based controllability for different attack strategies?

Chapter 6: How can we quantify the recoverability of optical networks? Which property of the optical network has strong correlation with its recoverability? Can we come up with a method to improve the recoverability of optical networks?

Chapter 7: How can we apply the framework for network recoverability to network controllability? Can we get analytical results to calculate the recoverability of networks in terms of controllability?

The dissertation is organized as follows. In Chapter 2, we propose closed-form analytic approximations for the minimum number of driver nodes needed to fully control networks, where links are removed according to both random and targeted attacks. We also do case studies for both real-world and synthetic networks. In Chapter 3, we deduce the fraction of driver nodes in random failure process for two types of swarm signalling networks. In Chapter 4, we apply machine learning models to further improve our approximations proposed in Chapter 2. We also compare our improved approximations with existing methods. In Chapter 5, we propose closed-form analytic approximations for the number of controllable nodes in sparse communication networks from the point of view of network controllability, considering link-based random attack, targeted attack, as well as random attack under the protection of critical links. We then compare our approximations with simulation results on communication networks. Chapter 6 proposes a general topological approach and recoverability indicators to measure the network recoverability for optical networks for two recovery scenarios: 1)only the links which are damaged in the failure process can be recovered and 2) links can be established between any pair of nodes that have no link between them after the failure process. We use the robustness envelopes of realizations and the histograms of two recoverability indicators to illustrate the impact of the random failure and recovery processes on the network performance. In Chapter 7, we adopted the framework of network recoverability and

investigate the recoverability of network controllability for two recovery scenarios. We also deduced the analytical results of the recoverability indicators by generating functions. Chapter 8 concludes this thesis.

1.4. PUBLICATION RELATED TO THIS THESIS

The following papers are completed by the author of this thesis while pursuing the Ph.D degree at Delft University of Technology.

1. P. Sun, R. E. Kooij, Z. He and P. Van Mieghem, *Quantifying the Robustness of Network Controllability*, 4th International Conference on System Reliability and Safety (ICSRS 2019), 20-22 November, Rome, Italy. **[Chapter 2]**

2. P. Sun, R. E. Kooij and R. Bouffanais, *Controllability of a class of Swarm Signalling Networks*, in preparation. **[Chapter 3]**

3. A. Dhiman, P. Sun and R. E. Kooij, *Using Machine Learning to Quantify the Robustness of Network Controllability*, Machine Learning for Networking - Third International Conference, MLN2020, Springer, p. 19-39. **[Chapter 4]**

4. P. Sun, R. E. Kooij and P. Van Mieghem, *Reachability-based Robustness of Controllability in Sparse Communication Networks*, IEEE Transactions on Network and Service Management (2021). **[Chapter 5]**

5. P. Sun, Z. He, R. E. Kooij and P. Van Mieghem, *Topological Approach to Measure the Recoverability of Optical Networks*, Optical Switching and Networking, 100617. **[Chapter 6]**

6. Z. He, P. Sun and P. Van Mieghem, *Topological approach to measure network recoverability*, **best paper award**, 11th International Workshop on Resilient Networks Design and Modeling (RNDM 2019), 14-16 October, Nicosia, Cyprus. **[Chapter 6]**

7. A. Chen, P. Sun and R. E. Kooij, *The recoverability of network controllability*, 5th International Conference on System Reliability and Safety (ICSRS 2021), 24-26 November, Palermo, Italy. **[Chapter 7]**

2

THE ROBUSTNESS OF NETWORK CONTROLLABILITY

In this chapter, we propose closed-form analytic approximations for the minimum number of driver nodes needed to fully control networks, where links are removed according to both random and targeted attacks. Our approximations rely on the concept of critical links. A link is called critical if its removal increases the required number of driver nodes. We validate our approximation on both real-world and synthetic networks. For random attacks, the approximation is always very good, as long as the fraction of removed links is smaller than the fraction of critical links. For some cases, the approximation is still accurate for larger fractions of removed links. The approximation for an attack, where first the critical links are removed, is also accurate, as long as the fraction of removed links is sufficiently small. Finally, we show that the critical link attack is the most effective among 4 considered attacks, as long as the fraction of removed links is smaller than the fraction of critical links.

2.1. INTRODUCTION

Our society nowadays depends critically on the proper functioning of a variety of infrastructures, such as the Internet, the power grid, water management networks and mobile communication networks. It is common practice to model such infrastructures as complex networks. Research over the last decades has led to a deep understanding of structural and robustness properties of complex networks [37], [38]. In recent years, the emphasis has shifted to understanding the controllability of such networks [39] [40] [41] [42]. Controllability is an essential property for the safe and reliable operation of real-life infrastructures. A system is said to be controllable if it can be driven from any initial state to any desired final state by external inputs in finite time [42]. Merging classical control theory with network science [43] introduced the notion of structural controllability. Let the $N \times N$ matrix A represent the network's wiring diagram, while the connection of M input signals

This chapter is based on the published paper [36].

to the network is described by the $N \times M$ input matrix B, where $M \leq N$. Then, the system characterized by (A, B) is said to be structurally controllable, if it is possible to fix the non-zero parameters in A and B in such a way that the obtained system (A, B) is controllable in the classical sense of satisfying Kalman's rank condition. Liu *et al.* [39] found a method that gives the minimum number of driver nodes, which are driven by external inputs, that are needed to achieve structural controllability of a directed network. As was pointed out by Cowan *et al.* [44], the results reported in Liu *et al.* [39] critically depend on the assumption that the network has no self-links, i.e. a node's internal state can only be changed upon interaction with a neighbor. In this chapter, we will also assume this condition. Ruths *et al.* [45] developed a theoretical framework for characterizing control profiles of networks. Yuan *et al.* [40] further proposed the concept of exact controllability based on the maximum multiplicity of all eigenvalues of the adjacency matrix A to find the driver nodes in networks. Jia *et al.* [41] classified each node into one of three categories, based on its likelihood of being included in a minimum set of driver nodes and discovered bimodal behaviour for the fraction of redundant nodes, when the average degree of the networks is high. Nepusz *et al.* [42] indicated that most real-world networks are more controllable than their randomized counterparts. Yan *et al.* [46] investigated the relation between the maximum energy needed for controllability and the number of driver nodes.

Real-world networks are often confronted with topological perturbations such as link-based random failures or targeted malicious attacks. For instance, in power grids, the breakdown of connections between different substations in some cases can be interpreted as random failures due to circuit aging or natural disasters. Malicious, and targeted attacks can seriously degrade the network performance [47]. In transportation networks, betweenness centrality-based targeted attacks can have a significant impact on normal operation [48].

Network robustness under topological perturbations has been widely investigated. The effective graph resistance [20], the viral conductance [49], the size of giant component [50], betweenness and eigenvector centrality are computed to measure the robustness of networks under topological perturbations. Wang *et al.* [51] investigated two interconnection topologies for interdependent networks and proposed the derivative of the largest mutually connected component as a new robust metric, which addresses the impact of a small fraction of failed nodes. Trajanovski *et al.* [52] studied the robustness envelope and concluded that centrality-based targeted attacks are sufficient for studying the worst-case behavior of real-world networks. Koç *et al.* [53] found that increasing the effective graph resistance of synthetic power systems results in decreased grid robustness against cascading failures by targeted attacks.

The robustness of the network controllability can be assessed by quantifying the increase in the minimum number of driver nodes N_D, under perturbation of the network topology. The impact of topological perturbations on the controllability of networks has been investigated extensively in recent years. Pu *et al.* [54] found that the degree-based node attack is more efficient than a random attack for degrading the controllability in directed random and scale-free networks. Nie *et al.* [55] found that the controllability of Erdős-Rényi random graphs with a moderate average degree is less robust, whereas a scale-free network with moderate power-law exponent shows a stronger ability to maintain its controllability, when these networks are under intentional link attack. Thomas *et al.* [56]

identified that the potency of a degree-based attack is directly related (on average) to the betweenness centrality of the edges being removed. Lu *et al.* [57] discovered that a betweenness-based strategy is quite efficient to harm the controllability of real-world networks. Mengiste *et al.* [58] introduced a new graph descriptor, 'the cardinality curve', to quantify the robustness of the control structure of a network to progressive link pruning.

The previous works on the robustness of network controllability listed above, have been mainly based upon simulations. In this chapter we quantify the robustness of network controllability by deriving analytical expressions, approximating the increase of the number of driver nodes, upon random and targeted link removals. Based upon methods from statistical physics, Liu *et al.* [39] already found analytical approximations for the number of driver nodes N_D, as a function of the nodes in- and out-degree distributions. However, the obtained expressions are an implicit set of equations, which are derived under the assumptions of $N \to \infty$ (thermodynamic limit) and sufficiently large average node degree.

We propose an analytical approximation to quantify the robustness of network controllability, based upon the concept of critical links, introduced in [39]. Links can be classified into three categories: critical, redundant, and ordinary [39]. A link is critical if its removal increases the number of driver nodes to remain in full control of the system. A link is redundant if it never belongs to a maximum matching. A link is ordinary if it is neither critical nor redundant. In this chapter, we want to derive analytical expressions for the increase in the minimum number of driver nodes, upon link removal. We will use the concept of critical links to construct such approximations, both for random link removals and targeted attacks. We show the performance of our approximations in both real-world and synthetic networks. Finally, we compare an attack based upon critical links, to attacks based upon topological properties, such as the out-in degree-based attack.

This chapter is organized as follows. In Section 2.2 and 2.3, we propose analytic approximations for the minimum number of driver nodes N_D when the network is under random attacks and targeted attacks, respectively. In Section 2.4, we compare the robustness of controllability under four different attack methods. Section 2.5 concludes the chapter.

2.2. NUMBER OF DRIVER NODES UNDER RANDOM ATTACKS

In this section, we assume that links are removed from the network uniformly at random. We derive an analytical approximation for the minimum number of driver nodes N_D for random attacks and show the performance of the approximation for real-world and synthetic networks.

2.2.1. THE FRACTION l OF REMOVED LINKS IS LESS THAN THE FRACTION OF CRITICAL LINKS l_c

For a network with N nodes and L links, denote the minimum number of driver nodes by N_{D0}. The number of critical links L_C can be determined by applying the Hopcraft-Karp algorithm L times, by considering all L networks that are obtained by removing exactly one link from the original network. If we denote the number of removed links by m, then the fraction of removed links $l = \frac{m}{L}$, while the fraction of critical links l_c satisfies $l_c = \frac{L_c}{L}$.

We consider the case $l \leq l_c$, i.e. m links are removed uniformly at random, under the condition that the number of removed links $m \leq L_c$. Now assume that of these m links i links are critical ($i \leq m$) and, hence, $m - i$ links are non-critical. We assume that the set of critical links is nearly unchanged when the fraction of removed links is small. Invoking the fact that after removing a critical link, the minimum number of driver nodes N_D increases by one [39], thus, when i critical links are iteratively removed one by one, the minimum number of driver nodes N_D increases by one in each iteration. For the $m - i$ removed non-critical links, the minimum number of driver nodes N_D remains the same. We show in the Appendix that this leads to the following approximation $n_{D,rand}$ for the normalized minimum number of driver nodes:

$$n_{D,rand} = \frac{N_{D0} + lL_c}{N} \tag{2.1}$$

VALIDATION FOR REAL-WORLD NETWORKS

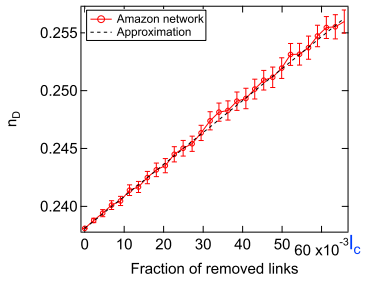
We evaluate the performance of the approximation $n_{D,rand}$ in (2.1) for 8 real-world networks. Table 2.1 presents the properties of the 8 real-world networks: the number of nodes (N), the number of links (L), the initial minimum number of driver nodes (N_{D0}) and the number of critical links (L_c).

Table 2.1: Properties of the 8 real-world networks

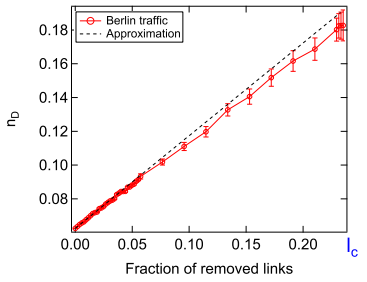
Networks	N	L	N_{D0}	L_c
Amazon network [59]	105	441	25	29
Berlin traffic network [60]	224	523	14	123
IEEE118 power grid [61]	118	179	38	36
Illinois students network [62]	70	366	3	8
Hagy Chesapeake Bay ecosystem [63]	37	215	9	4
INSNA social network [64]	60	94	35	5
s838 [39]	512	819	119	179
TRN-Yeast-2 [39]	688	1079	565	23

Figure 2.1 shows the comparison between our approximation Eq. (2.1) and simulation results in the considered real-world networks. For each figure, the right-most point at the horizontal axis denotes the fraction of critical links l_c. We use 10000 realizations and obtain mean values for the fraction of minimum number of driver nodes n_D, together with the 95%− confidence interval, for each fraction l. Visual inspection of Figure 2.1 confirms that our approximation (2.1) is close to the simulation results for the 8 real-world networks, when the fraction of removed links l satisfies $l \leq l_c$.

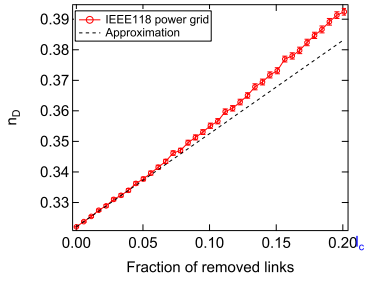
To further quantify the accuracy of the approximation $n_{D,rand}$, Table 2.2 gives two performance indicators. K different values of the fraction of removed links, i.e., $c_1, c_2, ..., c_K$, are evenly determined in the interval $[0, l_c]$. Let $n_D^*(c_i)$ and $n_D(c_i)$ denote the mean simulated n_d and the approximation (2.1) at the fraction of removed links $l = c_i$, respectively. The performance indicator γ denotes the fraction of the interval $[0, l_c]$ for which the absolute value of the relative error between the approximation and the mean simulated



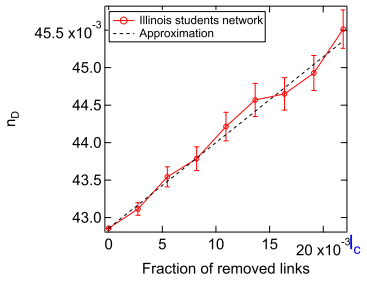
(a) Amazon



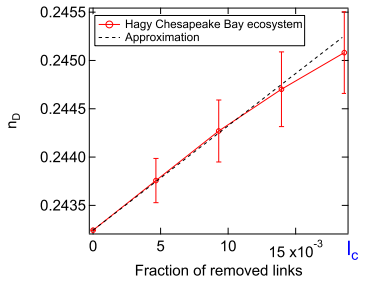
(b) Berlin traffic



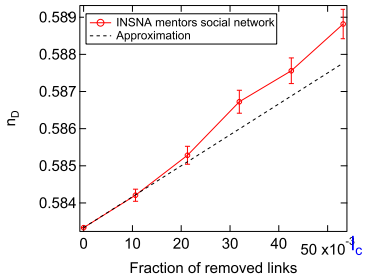
(c) IEEE118 power grid



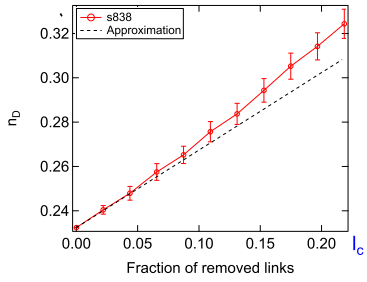
(d) Illinois students



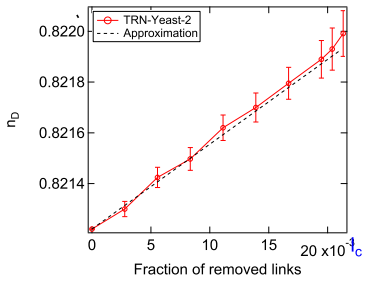
(e) Hagy Chesapeake Bay



(f) INSNA



(g) s838



(h) TRN-Yeast-2

Figure 2.1: Performance of the approximation (2.1) for the normalized minimum number of driver nodes n_D as a function of the fraction of removed links l in real-world networks under random attacks. The results for each fraction l are based on 10000 simulations.

value, does not exceed 5%.

$$\gamma = \frac{\sum_{i=0}^{K} \mathbf{1}_{\left|\frac{n_D^*(c_i)-n_D(c_i)}{n_D^*(c_i)}\right| \leq 5\%}}{K}$$

Finally, r denotes the absolute value of the relative error between the approximation and the mean value obtained through simulation, at $l = l_c$. Table 2.2 shows for all real-

Table 2.2: Performance indicators for the approximation $N_{D,rand}$ for the 8 real-world networks; $l \leq l_c$

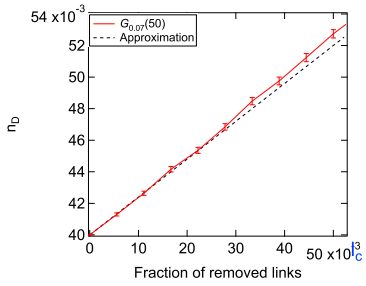
Networks	γ	r
Amazon	100%	0.11%
Berlin traffic	100%	4.82%
IEEE118 power grid	100%	2.31%
Illinois students	100%	0.35%
Hagy Chesapeake Bay	100%	0.07%
INSNA	100%	0.20%
s838	100%	4.80%
TRN-Yeast-2	100%	0.01%

world networks that the approximation (2.1) for $n_{D,rand}$ performs very well for $l \leq l_c$. For 5 out of the 8 considered networks, the absolute value of the relative error at $l = l_c$ is less than 0.5%.

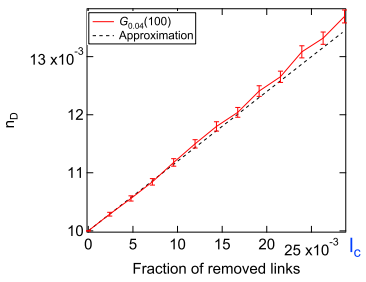
SYNTHETIC NETWORKS

Next we test our approximation Eq.(2.1) on two types of synthetic networks. When generating the directed Erdős-Rényi random network $G_p(N)$ with N nodes, the probability that every node has an outbound link to the other nodes is p. We generate the scale-free network $BA(N, M_0, M)$ by using the Barabási-Albert (BA) model, where N is the number of nodes, M is the number of out-going links for each new node added to the current network. We assume that initially the network consists of a complete digraph on M_0 nodes, where M_0 equals M. In the initial complete digraph, every pair of distinct nodes is connected by a pair of unique links (one in each direction). New nodes are added to the network one at a time. Each new node is connected to M existing nodes with a probability that is proportional to the number of links that the existing nodes already have. Figure 2.2 shows that both for Erdős-Rényi and Barabási-Albert (BA) networks, our analytic approximation (2.1) for $n_{D,rand}$ fits well with simulation results, when the fraction of removed links l is less than the fraction of critical links l_c. For the results depicted in Figure 2.2, Table 2.3 reports the performance indicators γ and r introduced in the previous subsection. Table 2.3 shows that also for the considered synthetic networks, the approximation $n_{D,rand}$ performs very well for $l \leq l_c$.

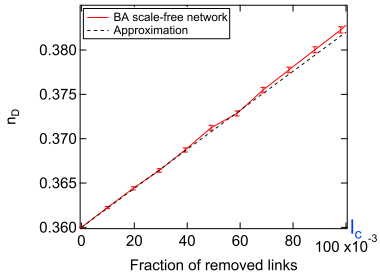
The overall conclusion of this subsection is that our approximation $n_{D,rand}$ in Eq.(2.1) gives a very good estimation for the minimum number of driver nodes, if the fraction of randomly removed links l is smaller than, or equal to, the fraction of critical links l_c.



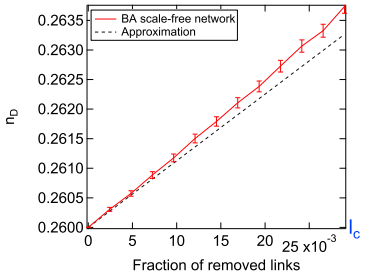
(a) ER network $G_{0.07}(50)$



(b) ER network $G_{0.04}(100)$



(c) N=200, $E[D]$=4



(d) N=500, $E[D]$=8

Figure 2.2: The normalized minimum number of driver nodes n_D as a function of the fraction of removed links l in synthetic networks under random attacks. In each sub-figure, we generate 100 corresponding synthetic networks and calculate the average fraction of critical links l_c and the average value of n_D for each fraction of removed links. For each network, the value of n_D for each fraction l is based on 10000 simulations.

Table 2.3: Performance indicators for the approximation $n_{D,rand}$ for the 4 synthetic networks; $l \leq l_c$

Networks	γ	r
ER: $G_{0.07}(50)$	100%	2.08%
ER: $G_{0.04}(100)$	100%	1.80%
BA: N=200, $E[D]$=4	100%	0.29%
BA: N=500, $E[D]$=8	100%	0.09%

2.2.2. THE FRACTION l OF REMOVED LINKS IS LARGER THAN THE FRACTION OF CRITICAL LINKS l_c

Because in most cases l_c is quite small, we also estimate the normalized minimum number of driver nodes n_D when the fraction l of removed links is larger than the fraction l_c of critical links. Therefore, for $l \geq l_c$, we propose a simple closed-form approximation for n_D:

$$n_D = al^2 + bl + c \tag{2.2}$$

where the parameters a, b and c will be determined by some boundary conditions. For the first two boundary conditions we assume that, for $l = l_c$, Eq.(2.2) has the same value and the same derivative as Eq.(2.1). This leads to the equations $N_{D0} + l_c L_c = N(al_c^2 + bl_c + c)$ and $L_c = N(2al_c + b)$, respectively. Finally, if we remove all links, i.e. $l = 1$, all nodes need

to be controlled. This gives the boundary condition $1 = a + b + c$. Solving for a, b and c and combining with the approximation Eq.(2.1), we obtain the following approximation for n_D for all values of l:

$$n_{D,rand} = \begin{cases} \frac{N_{D0}+lL_c}{N} & l \le l_c \\ \frac{al^2+bl+c}{N} & l \ge l_c \end{cases} \tag{2.3}$$

with, $a = \frac{N-N_{D0}-L_c}{N(l_c-1)^2}$, $b = L_cN - 2al_c$, and $c = 1 - L_cN + a(2l_c - 1)$. Eq.(2.3) respresents a closed-form approximation for n_D, which only depends on N, L, N_{D0} and L_C. The computational complexity of the approximation is $O(\sqrt{N}L^2)$, which is needed for the computation of L_C.

We compare the approximation (2.3) with simulation results for the 8 real-world networks and two types of synthetic networks. Figure 2.3 shows that for moderate values of the fraction of removed links, the approximation exhibits a very good fit for the real-world networks. This is quantified in Table 2.4 where we show two performance indicators: r which denotes the relative error at $l = 0.2$ and l^*, which represents the smallest value of l, where the relative error between the approximation and the simulated mean exceeds 5%.

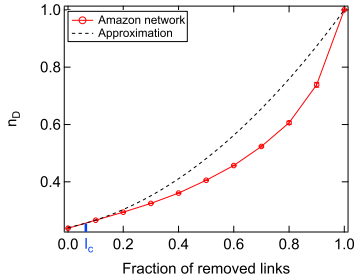
Table 2.4: Performance indicators for the approximation $n_{D,rand}$ for the 8 real-world networks

Networks	r	l^*
Amazon	3.12%	0.32
Berlin traffic	3.15%	0.24
IEEE118 power grid	2.31%	0.29
Illinois students	30.20%	0.12
Hagy Chesapeake Bay	5.22%	0.19
INSNA	1.50%	0.68
s838	4.15%	0.23
TRN-Yeast-2	0.39%	0.72

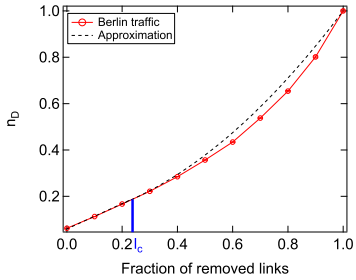
Figure 2.3 illustrates that the approximation both under- and overestimates the value of n_D. Table 2.4 shows that the approximation is the most accurate for the INSNA social network and the TRN-Yeast-2 network, while the least accurate for Illinois students network and Hagy Chesapeake Bay ecosystem. According to Table 2.4, for 6 out of the 8 real-world networks, for random link removals up to 20%, the absolute value of the relative error of the approximation (2.3) does not exceed 5%. For the worst performing network, Hagy Chesapeake Bay, 12% of the links can be removed before the absolute relative error exceeds 5%.

Finally, Figure 2.4 shows that the comparison for Erdős-Rényi and Barabási-Albert networks, leads to the same conclusions as above. The performance indicators r and l^* for the 4 synthetic networks are given in Table 2.5.

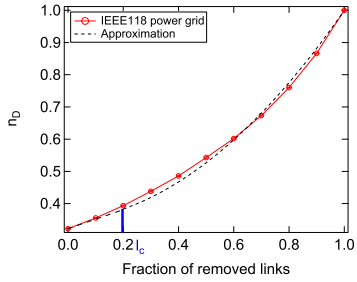
The overall conclusion of this subsection is that our approximation $n_{D,rand}$ in Eq. (2.3), in most cases, also gives a good estimation for the minimum number of driver nodes, if the fraction of randomly removed links l is larger than the fraction of critical links l_c, but still sufficiently small.



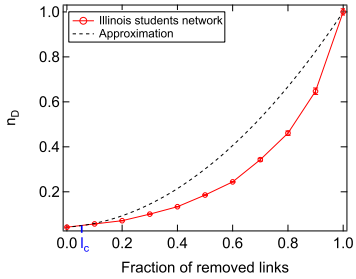
(a) Amazon



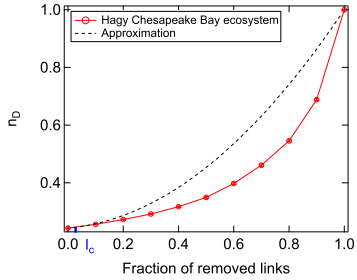
(b) Berlin traffic



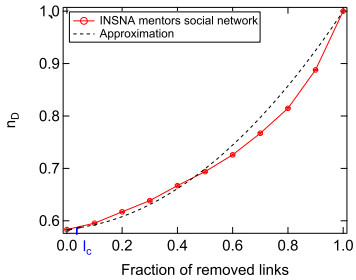
(c) IEEE118 power grid



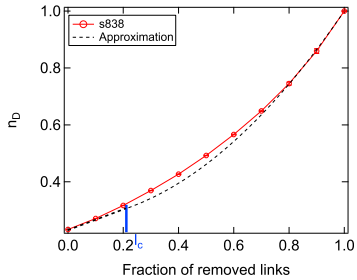
(d) Illinois students



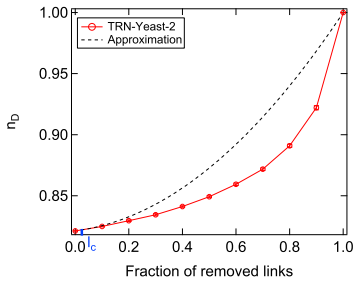
(e) Hagy Chesapeake Bay



(f) INSNA

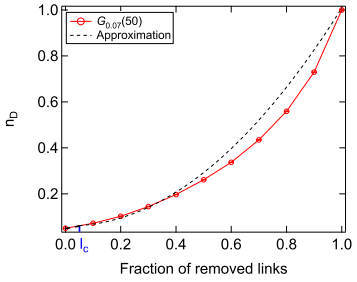


(g) s838

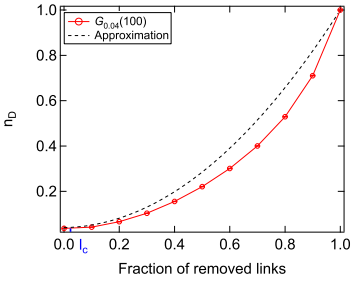


(h) TRN-Yeast-2

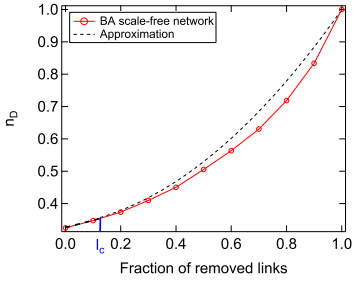
Figure 2.3: The normalized minimum number of driver nodes n_D as a function of the fraction of removed links l in real-world networks under random attacks. In each plot, the dashed line shows the simulation results and the solid line shows our approximation. The simulation results for each fraction l are based on 10000 simulations.



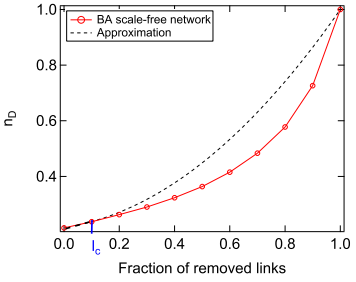
(a) ER network $G_{0.07}(50)$



(b) ER network $G_{0.04}(100)$



(c) N=200, $E[D]$=4



(d) N=500, $E[D]$=8

Figure 2.4: The normalized minimum number of driver nodes n_D as a function of the fraction of removed links l in synthetic networks under random attacks. The results for each fraction l are based on 10000 simulations.

Table 2.5: Performance indicators for the approximation (2.1) for the 4 synthetic networks

Networks	r	l^*
ER: $G_{0.07}(50)$	2.32%	0.47
ER: $G_{0.04}(100)$	23.56%	0.08
BA: N=200, $E[D]$=4	1.47%	0.57
BA: N=500, $E[D]$=8	3.25%	0.28

2.3. DRIVER NODES UNDER TARGETED ATTACKS

In this section, we quantify the impact of targeted link attacks on the minimum number of driver nodes. We assume that the attacker knows the critical links, which will be attacked first. We consider two scenarios. In the first scenario, the attacker removes critical links uniformly at random. We call this a random critical link attack. For the second scenario, we rank the critical links according to some network property. Inspired by the degree-based attack methods adopted in [56], we will rank the critical links in ascending order of their out-in degree $\delta_{i,j}$, which is defined as the sum of the out-degree of its source node d_i^{out} and the in-degree of its target node d_j^{in}, i.e., $\delta_{i,j} = d_i^{\text{out}} + d_j^{\text{in}}$. We refer to the second case as a targeted critical link attack. For both scenarios, we first remove critical links in the original networks. After all critical links are removed, the other links are removed uniformly at random. Attacks based upon critical links removal were also suggested by Mengiste et al. [58], however, only simulations results were reported.

2.3.1. THE FRACTION l OF REMOVED LINKS IS LESS THAN THE FRACTION OF CRITICAL LINKS l_c

Again, we will derive an approximation for the minimum number of driver nodes. We assume that, as long as the number of removed links $m \leq L_c$, the removal of each link increases the minimum number of driver nodes N_D by one. Consequently, when the number of removed links is smaller than L_c (the fraction of removed links l is smaller than l_c), the approximation for the minimum number of driver nodes N_D increases linearly with the fraction of removed links l. When the number of removed links equals the number of critical links L_c, the minimum number of driver nodes N_D equals $N_{D0} + L_c$. Thus, when the fraction l of removed links is no more than the fraction l_c of critical links, we obtain the following approximation for n_D:

$$n_{D,crit} = \frac{N_{D0} + lL}{N} \tag{2.4}$$

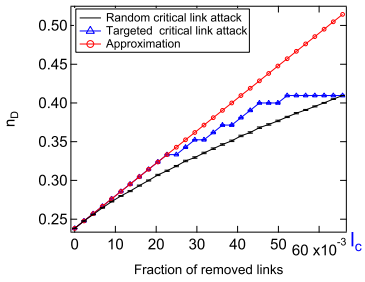
We evaluate the performance of (2.4) in our 8 real-world networks. Figure 2.5 shows that the targeted critical link attack is slightly more efficient than the random critical link in increasing the minimum number of driver nodes. Considering the small difference between the two scenarios, in the remainder of the chapter, we will only consider random critical link attack, and simply refer to it as critical link attack. For all cases the approximation (2.4) is a good fit for sufficiently small l, while in some cases this holds for all $l \leq l_c$. We also observe that the approximation (2.4) provides a worst-case estimate for the number of needed driver nodes. Comparing with the critical link attack, we quantify the performance of the approximation (2.4) in Table 2.6. We use γ, the fraction of the interval $[0, l_c]$ where the absolute value of the relative error does not exceed 5%, and the absolute value of the relative error r at $l = l_c$, as the performance indicators.

Table 2.6: Performance indicators for the approximation $n_{D,crit}$ for the 8 real-world networks; $l \leq l_c$

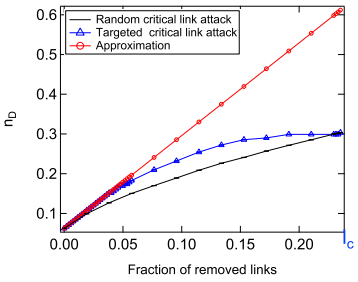
Networks	γ	r
Amazon	100%	0.68%
Berlin traffic	6.38%	79.60%
IEEE118 power grid	78.58%	7.25%
Illinois students	33.33%	22.22%
Hagy Chesapeake Bay	100%	0%
INSNA	100%	0%
s838	70%	9.88%
TRN-Yeast-2	100%	0.68%

While for 4 of the 8 considered real-world networks the approximation (2.4) for $n_{D,crit}$ is very good, the approximation is reasonable for two networks (IEEE118 power grid and s838) and rather poor for the remaining two (Berlin traffic and Illinois students). However, approximation (2.4) always seems to overestimate the normalized minimum number of driver nodes n_D and, hence, approximation (2.4) can be considered a worst-case approximation.

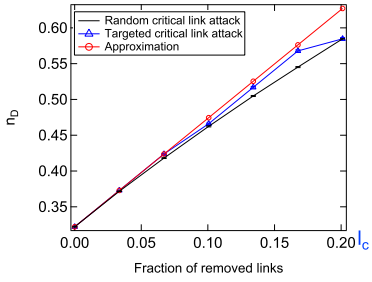
Next we evaluate the performance of (2.4) in synthetic networks. Figure 2.6 shows that



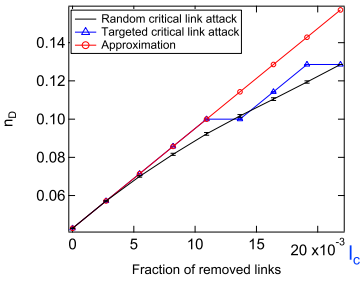
(a) Amazon



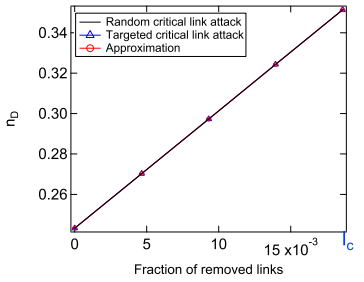
(b) Berlin traffic



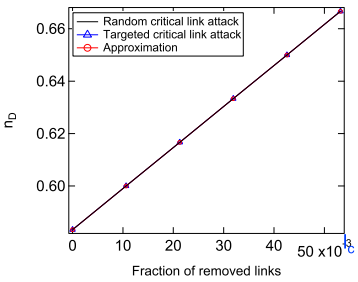
(c) IEEE118 power grid



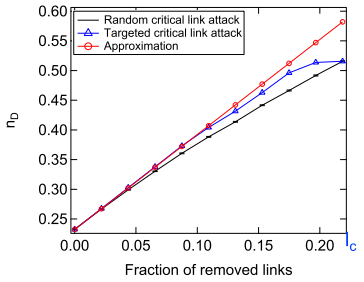
(d) Illinois students



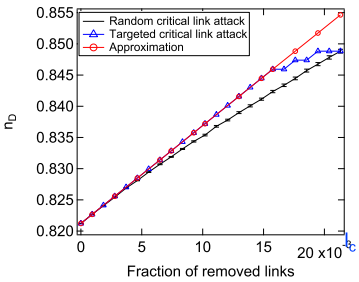
(e) Hagy Chesapeake Bay



(f) INSNA



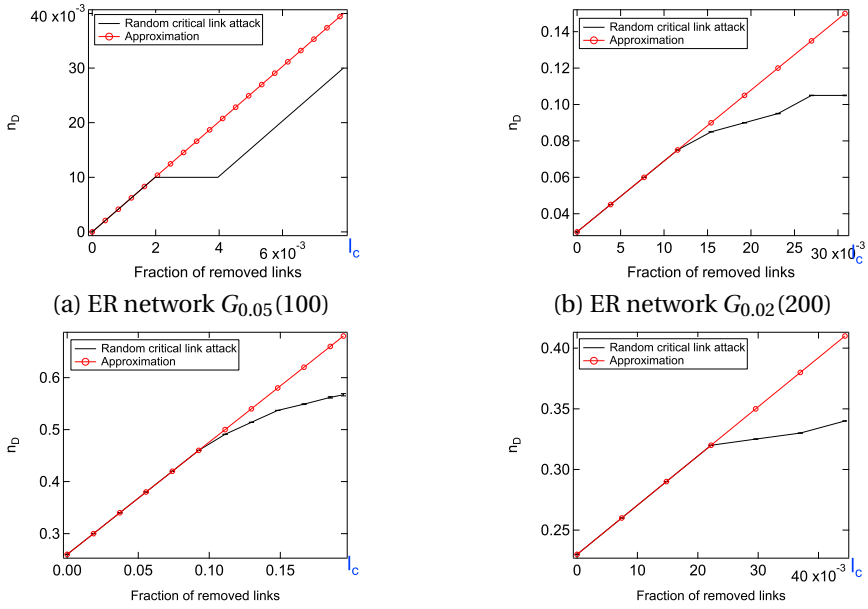
(g) s838



(h) TRN-Yeast-2

Figure 2.5: Performance of the approximation for the normalized minimum number of driver nodes n_D as a function of the fraction of removed links l in real-world networks under targeted attacks. The results for each fraction l are based on 10000 simulations.

our approximation Eq.(2.4) fits well with the simulation results in the first few removal steps. Qualitatively we observe the same behaviour as in Figure 2.5.



(a) ER network $G_{0.05}(100)$ (b) ER network $G_{0.02}(200)$

(c) BA network with $E[D]$=4 (d) BA network with $E[D]$=8

Figure 2.6: Performance of the approximation for the normalized minimum number of driver nodes n_D as a function of the fraction of removed links l in synthetic networks under targeted attacks.

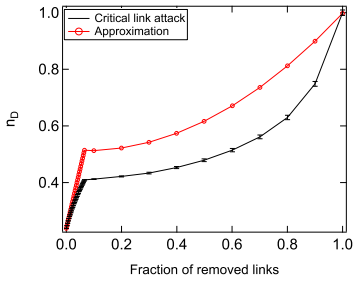
2.3.2. THE FRACTION l OF REMOVED LINKS IS LARGER THAN THE FRACTION OF CRITICAL LINKS l_c

We now construct an approximation when the number of removed links is larger than L_c (the fraction of removed links l is larger than l_c), in a similar way as in the previous section. Again assuming that for $l \geq l_c$ it holds that n_D is quadratic in l, we obtain $N_D = dl^2 + el + f$. Boundary conditions are now obtained from the assumptions that the parabola passes through $(1,1)$ and $(l_c, N_{D0} + L_C N)$ and has a zero derivative at the latter point. This leads to the following approximation for n_D for all values of l:

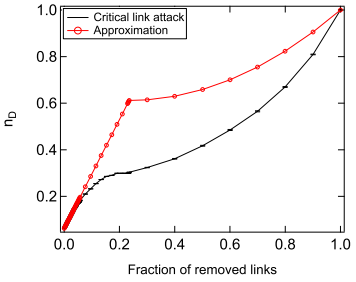
$$n_{D,crit} = \begin{cases} \frac{N_{D0}+lL}{N} & l \leq l_c \\ \frac{dl^2+el+f}{N} & l \geq l_c \end{cases} \tag{2.5}$$

with, $d = \frac{N-N_{D0}-l_cL}{N(l_c-1)^2}$, $e = -2dl_c$, and $f = 1 + d(2l_c - 1)$.

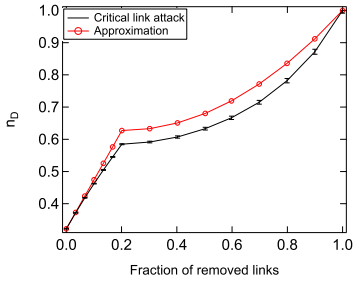
From Figure 2.7 and Figure 2.8, we can find the approximation $n_{D,crit}$ fits well with simulation results when the fraction of removed links is sufficiently small. When the fraction of removed links is getting larger, the difference between our approximation and simulation results is relatively large. However, in all cases the approximation seems to serve as a worst-case estimate for the number of required driver nodes. This implies that approximation (2.5) can have value in risk assessment studies.



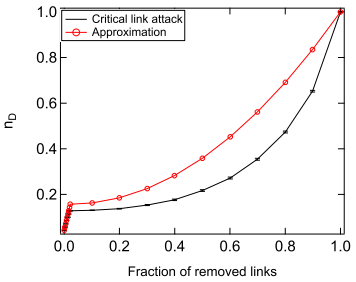
(a) Amazon



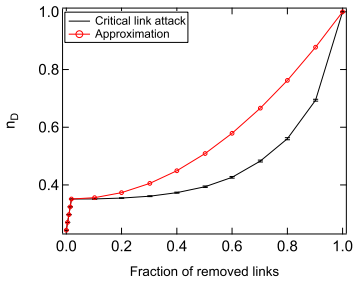
(b) Berlin traffic



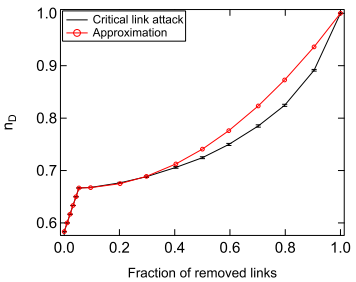
(c) IEEE118 power grid



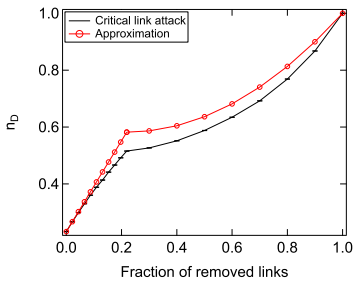
(d) Illinois students



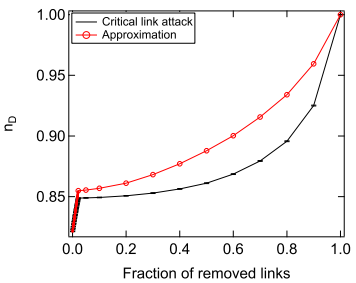
(e) Hagy Chesapeake Bay



(f) INSNA

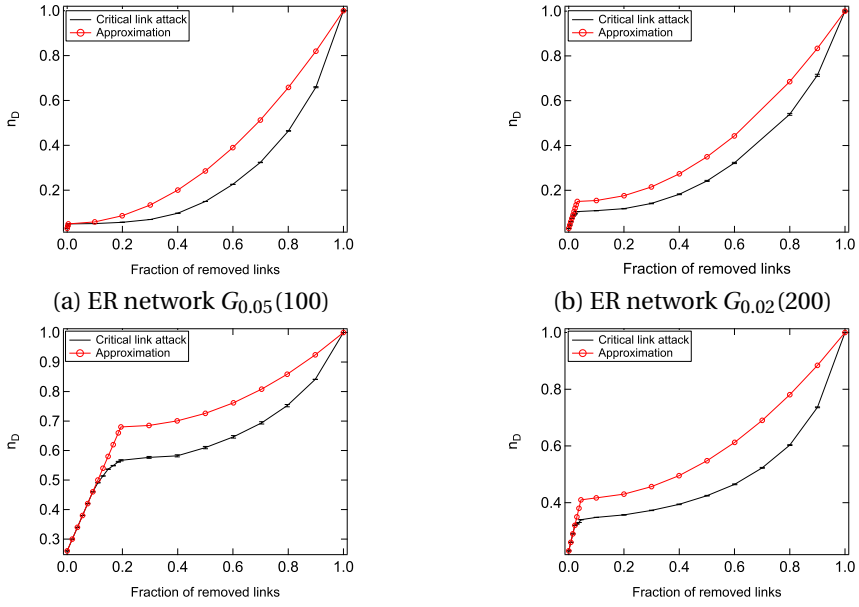


(g) s838



(h) TRN-Yeast-2

Figure 2.7: Performance of the approximation for the normalized minimum number of driver nodes n_D as a function of the fraction of removed links l in real-world networks under targeted attacks.



(a) ER network $G_{0.05}(100)$ (b) ER network $G_{0.02}(200)$

(c) BA network with $E[D]$=4 (d) BA network with $E[D]$=8

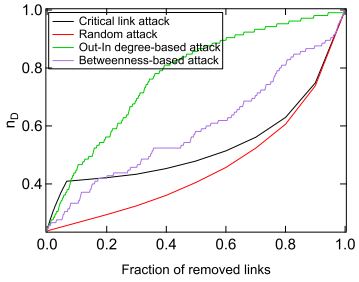
Figure 2.8: Performance of the approximation for the normalized minimum number of driver nodes n_D as a function of the fraction of removed links l in synthetic networks under targeted attacks.

2.4. COMPARISON OF n_D UNDER DIFFERENT ATTACK STRATEGIES

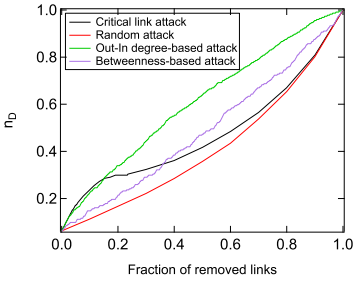
In this section, we compare the minimum number of driver nodes for link removals under four attack strategies: (a) critical link attack (targeted attack), (b) out-in degree-based attack, (c) betweenness-based attack and (d) random attack. In the out-in degree-based attack, we remove links one by one in the ascending order of the out-in degree using the recalculated out-in degree distribution at every removal step. In the betweenness-based attack, we remove links one by one in the descending order of the betweenness using the recalculated betweenness distribution at every removal step.

Figure 2.9 and Figure 2.10 show that, for most values of l, the out-in degree-based attack is the most harmful attack strategy. In other words, the out-in degree-based attack strategy is more efficient than other attack strategies in increasing the minimum number of driver nodes N_D, and, thus, degrading the controllability of the networks. However, if the fraction of removed links is small ($l \leq l_c$), the critical link attack is more effective than the out-in degree based attack. The most obvious case where this happens is for the TRN-Yeast-2 network, see Figure 2.9(h). When the fraction of removed links becomes larger, the critical link attack becomes less effective than the out-in degree-based attack. For large values of l, the targeted attack approaches the random attack. The random attack is the least effective attack strategy.

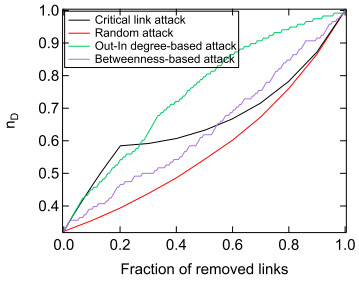
From results in Figure 2.9 and Figure 2.10, we can deduce that the links with a small out-in degree have a strong tendency to be critical links, whose removal increases the



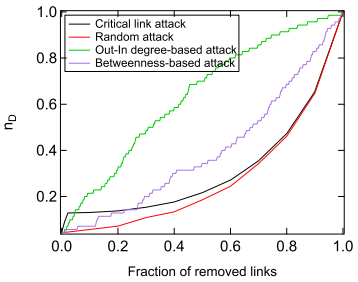
(a) Amazon



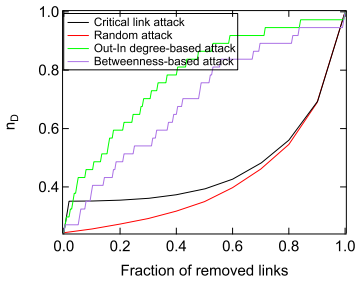
(b) Berlin traffic



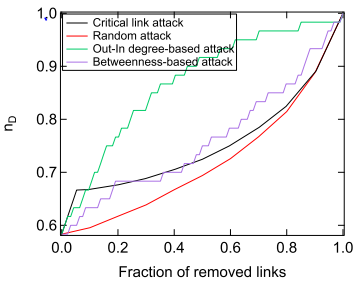
(c) IEEE118 power grid



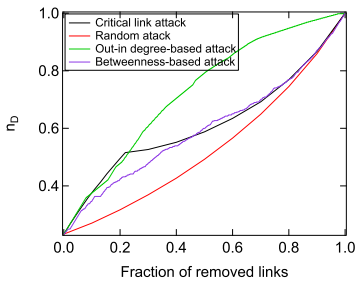
(d) Illinois students



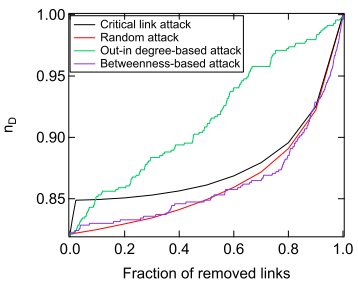
(e) Hagy Chesapeake Bay



(f) INSNA



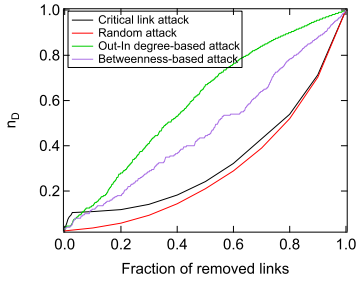
(g) s838



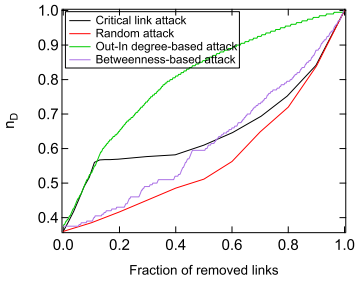
(h) TRN-Yeast-2

Figure 2.9: Performance of different attack strategies in real-world networks.

minimum number of driver nodes N_D more efficiently. The maximum matching, which



(a) ER network $G_{0.02}(200)$



(b) BA network with $E[D]$=4

Figure 2.10: Performance of different attack strategies in synthetic networks.

is used to determine N_D, can explain this phenomenon. As shown in Figure 1.1, link a and link d have a small out-in degree which equals 2. The number of matching links will decrease by 1 after removing either link a or link d. Consequently, the number of unmatched (driver) nodes will increase by one. Thus, link a and link d are critical links. Link e has a larger out-in degree which equals 4. The number of matching links is unchanged after removing link e. Link e is not a critical link. As a result, the link with a larger out-in degree is less likely to be a critical link since after removing this link, other links which share the same source or target node with this link, can also be alternative matching links.

2.5. CONCLUSION

In this study, we derived analytical closed-form approximations for the minimum number of driver nodes N_D needed to control networks, as a function of the fraction of removed links, both for random and targeted attacks. Our approximations rely on the notion of critical links. As targeted attack we consider the case, where first critical links are removed. Both for random and targeted attacks, our approximation is linear in the fraction of removed links l, as long as this fraction is smaller than the fraction of critical links. For fractions of removed links larger than the fraction of critical links, our approximation is quadratic in l. We validated our approximation through simulations on real-world and synthetic networks. For random attacks, the approximation is always very good, as long as the fraction of removed links is smaller than the fraction of critical links. For some cases, the approximation is still accurate for larger fractions of removed links. The approximation for attacks targeting the critical links is also accurate, as long as the fraction of removed links is sufficiently small. The approximation for the targeted attack always serves as a worst-case estimate. Finally, we showed that the critical link attack is the most effective among 4 considered attacks, as long as the fraction of removed links is smaller than the fraction of critical links.

3

CONTROLLABILITY OF A CLASS OF SWARM SIGNALLING NETWORKS

In this chapter, we propose closed-form analytical expressions to determine to the minimum number of driver nodes that is needed to control a specific class of networks. We consider swarm signalling networks with regular out-degree distribution where a fraction p of the links is unavailable. We further apply our method to networks with bi-modal out-degree distributions. Our approximations are validated through intensive simulations, and have high accuracy when compared with simulation results for both types of out-degree distribution.

3.1. INTRODUCTION

Network controllability is an essential property for the safe and reliable operation of real-world infrastructureshas and has been a hot reseach topic in recent years [39] [40] [41] [42]. A system is said to be controllable if it can be driven from any initial state to any desired final state by external inputs in finite time [24]. Merging classical control theory with network science [43] introduced the notion of structural controllability. Let the $N \times N$ matrix A represent the wiring diagram of a network with N nodes, while the connection of M input signals to the network is described by the $N \times M$ input matrix B, where $M \leq N$. Then, the system characterized by (A, B) is structurally controllable, if it is possible to find the non-zero parameters in A and B in such a way that the obtained system (A, B) is controllable in the classical sense of satisfying Kalman's rank condition.

Liu *et al.* [39] seminally use maximum matching to get the minimum number N_D of driver nodes, which are driven by external inputs, that are needed to achieve structural controllability of a directed network. However, the results reported in Liu *et al.* [39] critically depend on the assumption that the network has no self-links, i.e. a node's internal state can only be changed upon interaction with a neighbor [44]. Yuan *et al.* [40] further proposed the concept of exact controllability based on the maximum multiplicity of all eigenvalues of the adjacency matrix A to find the driver nodes in networks. Ruths *et*

al. [45] developed a theoretical framework for characterizing control profiles of networks. Jia *et al.* [41] classified each node into one of three categories, based on its likelihood of being included in a minimum set of driver nodes and discovered bimodal behaviour for the fraction of redundant nodes, when the average degree of the networks is high. Yan *et al.* [46] investigated the relation between the maximum energy needed for controllability and the number of driver nodes. Nepusz *et al.* [42] indicated that most real-world networks are more controllable than their randomized counterparts. Zhang *et al.* [65] studied the change of network controllability in growing networks and found the lower bound of the maximum number of nodes that can be added to a network while keeping the number of driver nodes unchanged.

The robustness of network controllability under perturbation of the network topology has been investigated extensively. Lu *et al.* [57] discovered that a betweenness-based strategy is quite efficient to harm the controllability of real-world networks. Lou *et al.* [66] present a search for the network configuration with optimal robustness of controllability against random node-removal attacks. Wang *et al.* [67] proposed a dynamic cascading failure model and investigated the controllability robustness in real logistics networks. Nie *et al.* [55] found that the controllability of Erdős-Rényi random networks with a moderate average degree is less robust, whereas a scale-free network with moderate power-law exponent shows a stronger ability to maintain its controllability, when these networks are under intentional link attack. Sun *et al.* [36] proposed closed-form analytic approximations for the minimum number of driver nodes needed to fully control networks, where links are removed according to both random and targeted attacks. Kamareji *et al.* [68] discussed the resilience and controllability of dynamic collective behaviours for a class of Swarm Signalling Networks (SSNs). The SSNs are modelled as directed graphs where the nodes have k-regular out-degree and Poisson in-degree with average k. Following the seminal paper by Liu *et al.* [39], an implicit equation is derived, whose solution leads to the minimum number of driver nodes to control the whole swarm [68]. However, upon validation of the formula given in [68] through simulation, we found obvious difference between the analytical results and simulation results.

The aim of this chapter is threefold. First, we correct the assumption when calculating the minimum fraction of driver nodes given in [68] and back this up with simulations. Second, we generalize the results by considering SSNs where a fraction p of the links is removed at random. Also for this case we are capable of deriving an implicit equation, whose solution leads to the minimum number of driver nodes. Finally, we relax the condition that the out-degree is regular: we look at bi-modal out-degree distributions, where the out-degree is k_1 for a fraction α of the nodes and k_2 for a fraction $1-\alpha$ of the nodes. Also for this case we consider scenarios with unavailable links.

3.2. GENERATING FUNCTIONS

As introduced in Section 1.1, the Hopcroft-Karp algorithm is applied to find the minimum number of driver nodes in a network. The Hopcroft-Karp algorithm works efficiently when the network is small and sparse. However, when the network is large and dense, the Hopcroft-Karp algorithm is no longer efficient in finding the number of driver nodes. A general expression for the minimum number N_D of driver nodes by using generating functions [37] is applicable, which is also provided in [39], as long as the closed-form

degree distribution of the network is known. In the rest of the chapter, we use the general expression to estimate the minimum number N_D of driver nodes in the SSNs with regular out-degree distribution and then deduce the general formula by considering the scenario when a fraction p of the links is unavailable. We then relax the condition that the out-degree is regular and look into networks with bi-modal out-degree distributions.

In a network, let x denote the probability that a link is in state X. For example, X can denote the existence of a link. We assume that the states of links are independent from each other. Then, the probability that all the links of a node with degree k are in state X is x^k. Averaging this probability by the degree distribution of the network, we then obtain the probability that all the links of a randomly chosen node are in state X. According to the definition of the generating function [69], this probability can be written as

$$G(x) = \sum_{k=0}^{\infty} p_k x^k, \tag{3.1}$$

where p_k is the probability that a randomly chosen node in the network has degree k. Let $x = 1$, then we obtain $G(1) = \sum_{k=0}^{\infty} p_k = 1$. Besides, the average degree $< k >$ of the network can be expressed as:

$$< k >= G^{'}(1) = \sum_{k=0}^{\infty} k p_k. \tag{3.2}$$

Assuming the degree of the node reached by following a randomly chosen link is k, the probability that all the other links of this node are in state X is x^{k-1}. The distribution of the degrees of the nodes reached by following a randomly chosen link is called excess degree distribution q_k, which depends on the degree distribution p_k; the larger p_k is, the larger q_k is. Furthermore, following a link, it is easier to reach a node with larger k. Hence, we have

$$q_k \propto k p_k. \tag{3.3}$$

The normalized distribution q_k satisfies

$$q_k = \frac{k p_k}{\sum_{k=0}^{\infty} k p_k} = \frac{k p_k}{< k >}. \tag{3.4}$$

Thus, the probability that all the other links of a node reached by following a randomly chosen link are in state X is

$$H(x) = \sum_{k=1}^{\infty} q_k x^{k-1} = \sum_{k=1}^{\infty} \frac{k p_k}{< k >} x^{k-1} = \frac{G^{'}(x)}{G^{'}(1)}. \tag{3.5}$$

It must be highlighted that all these functions are based on the assumption that the states of links are independent from each other [37].

3.3. SSNs WITH k-REGULAR OUT DEGREE

It is shown in Liu *et al.* [39] that the minimum number of driver nodes can be obtained by using the following set of generating functions

$$G_{out}(x) = \sum_{k_{out}=0}^{\infty} P_{out}(k_{out}) x^{k_{out}}, \tag{3.6}$$

$$G_{in}(x) = \sum_{k_{in}=0}^{\infty} P_{in}(k_{in}) x^{k_{in}}, \tag{3.7}$$

$$H_{out}(x) = \sum_{k_{out}=1}^{\infty} \frac{k_{out} P_{out}(k_{out})}{< k_{out} >} x^{k_{out}-1}, \tag{3.8}$$

$$H_{in}(x) = \sum_{k_{in}=1}^{\infty} \frac{k_{in} P_{in}(k_{in})}{< k_{in} >} x^{k_{in}-1}, \tag{3.9}$$

where $P_{out}()$ and $P_{in}()$ denote the probability distribution function of the out-degree and in-degree, respectively, and $< k_{out} >$ and $< k_{in} >$ denote the average out-degree and in-degree, respectively.

The general expression for the minimum fraction N_D of driver nodes obtained by Liu *et al.* [39] reads

$$n_D = \frac{N_D}{N} = \frac{1}{2}\{G_{in}(w_2) + G_{in}(1-w_1) - 2 + G_{out}(\hat{w}_2) + G_{out}(1-\hat{w}_1) + k(\hat{w}_1(1-w_2) + w_1(1-\hat{w}_2))\}, \tag{3.10}$$

where $w_1, w_2, \hat{w}_1$ and $\hat{w}_2$ satisfy

$$w_1 = H_{out}(\hat{w}_2), \tag{3.11}$$

$$w_2 = 1 - H_{out}(1-\hat{w}_1), \tag{3.12}$$

$$\hat{w}_1 = H_{in}(w_2), \tag{3.13}$$

$$\hat{w}_2 = 1 - H_{in}(1-w_1). \tag{3.14}$$

By construction, the out-degree distribution for the SSN suggested in [68], is a Dirac delta function, i.e.

$$P_{out}(k_{out}) = \delta(k - k_{out}), \tag{3.15}$$

where k is the fixed out-degree for every node. It is also shown in [68] that, for sufficiently large SSN's, the in-degree distribution closely resembles a Poisson distribution, with average k, i.e.

$$P_{in}(k_{in}) = \frac{k^{k_{in}}}{k_{in}!} e^{-k}. \tag{3.16}$$

Using the degree distributions in Eqs.(3.6)-(3.9) it follows

$$G_{out}(x) = x^k, \tag{3.17}$$

$$G_{in}(x) = e^{-k(1-x)}, \tag{3.18}$$

$$H_{out}(x) = x^{k-1}, \tag{3.19}$$

$$H_{in}(x) = e^{-k(1-x)}. \tag{3.20}$$

Therefore, the parameters $w_1, w_2, \hat{w}_1$ and $\hat{w}_2$ satisfy

$$w_1 = \hat{w}_2^{k-1}, \tag{3.21}$$

$$w_2 = 1 - (1 - \hat{w}_1)^{k-1}, \tag{3.22}$$

$$\hat{w}_1 = e^{-k(1-w_2)}, \tag{3.23}$$

$$\hat{w}_2 = 1 - e^{-kw_1}. \tag{3.24}$$

For the trivial case $k = 0$ it is easy to see that the above set of equations leads to $n_D = 1$, i.e. all agents in the swarm need to be controlled, which makes sense because the out-degree of every node is 0 in this case. Also, for the case $k = 1$, Eqs.(3.21)-(3.24) are solved for $w_1 = 1, w_2 = 0, \hat{w}_1 = e^{-1}$ and $\hat{w}_2 = 1 - e^{-1}$. Hence, for $k = 1$, it holds that $n_D = e^{-1}$.

For the case $k > 1$, [68] argues that the smallest solution of the pair of Eqs.(3.21) and (3.24) is given by $w_1 = \hat{w}_2 = 0$, and assuming that w_1 and $\hat{w}_2$ are indeed zero, the following expression for the fraction of driver nodes is derived:

$$n_D = \frac{1}{2}\{(1 - e^{-k(1-w_2)})^k - 1 + e^{-k(1-w_2)} + k(1 - w_2)e^{-k(1-w_2)}\}, \tag{3.25}$$

where w_2 is the solution of the implicit equation

$$1 - w_2 = (1 - e^{-k(1-w_2)})^{k-1}. \tag{3.26}$$

From Eq.(3.25) the asymptotic behaviour of n_D for large k can also be derived:

$$n_D \approx \frac{1}{2}e^{-k}. \tag{3.27}$$

However, upon simulation of SSN's, determining the fraction of driver nodes by applying the maximum matching algorithm, as described in [39], we found a discrepancy between Eq.(3.25) and the simulation results, see Figure 3.1.

We generate 10000 directed networks with $N = 10000$ for each out-degree k whose value ranges from 1 to 8. The fraction n_D of driver nodes is the average fraction of driver nodes over 10000 networks for each out-degree k. As shown in Figure 3.1, the result from Eq.(3.25) fits well with the simulation result at $k = 1$. However, the difference between Eq.(3.25) and simulation results are obvious for other values of k. For example, at all $k > 1$, the results from the simulation are about two times the results given by Eq.(3.25).

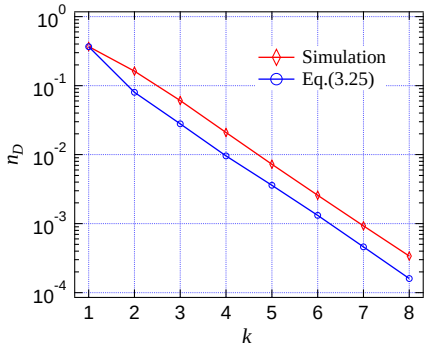


Figure 3.1: Comparing Eq.(3.25) with simulation results.

The discrepancy is due to the assumption that one can choose the solution of Eq.(3.21) and Eq.(3.24) given by $w_1 = \hat{w}_2 = 0$. One can also argue that the pair Eq.(3.21) and Eq.(3.24) is equivalent with the pair Eq.(3.22) and Eq.(3.23). If we assume

$$w_1 = 1 - w_2 \tag{3.28}$$

and

$$\hat{w}_2 = 1 - \hat{w}_1, \tag{3.29}$$

then the pair of equations Eq.(3.22)-Eq.(3.23) follows from the pair of equations Eq.(3.21)-Eq.(3.24). As a result, applying Eq.(3.10) leads to the following expression for the fraction of driver nodes:

$$n_D = ((1 - e^{-k(1-w_2)})^k - 1 + e^{-k(1-w_2)} + k(1 - w_2)e^{-k(1-w_2)}), \tag{3.30}$$

where w_2 is still the solution of Eq.(3.26).

The asymptotic behaviour of n_D for large k becomes:

$$n_D \approx e^{-k}. \tag{3.31}$$

Note that Eq.(3.30) also holds for $k = 1$, another indication for its correctness.

Table 3.1 shows the comparison between the approximations in Eqs.(3.30) and (3.31) and the simulations.

We generate 10000 directed networks with $N = 10000$ for each out-degree k whose value ranges from 1 to 8. The fraction of driver nodes n_D is the average fraction of driver nodes in 10000 networks. Then we calculate the analytical results from Eq.(3.30) and Eq.(3.31) and also the corresponding absolute relative error r. As shown in Table 3.1, the absolute relative errors of our approximation are less than 1% for k from 1 to 6. For the case where $k = 7$ and $k = 8$, the absolute relative errors are still small which are less than 6%. When the values of k are small, the absolute relative errors of Eq.(3.31) are large.

We conclude from Table 3.1 that the simulations are an excellent fit with our approximation Eq.(3.30). Also, the asymptotic approximation Eq.(3.31) is increasingly accurate for increasing k.

Table 3.1: Comparing Eqs.(3.30)-(3.31) with simulation results.

k	Eq.(3.30)		Eq.(3.31)		Simulation
	value	r	value	r	
1	0.367879	0.0079%	0.367879	0.0079%	0.36782
2	0.161903	0.40%	0.135335	16.07%	0.162003
3	0.060759	0.29%	0.049787	17.82%	0.06068
4	0.020916	0.28%	0.018316	12.18%	0.020943
5	0.007262	0.93%	0.006738	6.35%	0.007221
6	0.002578	0.23%	0.002479	4.06%	0.002561
7	0.00093	2.76%	0.000912	0.77%	0.000929
8	0.000339	5.93%	0.000335	4.69%	0.000346

3.4. SSNs WITH k-REGULAR OUT DEGREE AND RANDOM LINK FAILURES

In this section we generalize the results of the previous section by considering again SSNs with k-regular out-degree, but now we assume that a fraction p of the links is removed at random. This assumption is in accordance with some real-life scenarios, such as the communication disconnection between robots in swarm robotic networks because of the limited range of communication.

We will show that the analysis that led to our implicit approximations can also be conducted for this case. A crucial step is to find expressions for the generating functions Eqs.(3.6)-(3.9) for this case.

Instrumental in this is the following Lemma, see [70] which gives an expression for the degree distribution, after removing m links uniformly at random.

Lemma 1. *After removing m links in a uniform and random way from a network $G_0(N,L)$, with degree distribution $Pr[D_{G_0}=j]$, the degree distribution $Pr[D_G=i]$ of the new network G satisfies:*

$$Pr[D_G=i]=(1-p)^i\sum_{j=i}^{N-1}\binom{j}{i}p^{j-i}Pr[D_{G_0}=j], \tag{3.32}$$

where $p=\frac{m}{L}$ denotes the fraction of removed links in the original network G_0

Theorem 2. *Consider a directed network with k-regular out-degree and Poisson in-degree with average k. Then, after removing uniformly at random a fraction p of the links, the generating functions $\bar{G}_{out}(x)$ and $\bar{G}_{in}(x)$ of the out- and in-degree, respectively, satisfy*

$$\bar{G}_{out}(x)=(p+(1-p)x)^k \tag{3.33}$$

$$\bar{G}_{in}(x)=e^{-k(1-p)(1-x)} \tag{3.34}$$

The proof of Theorem 2 is given in Appendix A. Note that for the case without link removals, i.e. $p=0$, Eqs.(3.33)-(3.34) reduce to Eqs.(3.17)-(3.18) Also, we can deduce from Eqs.(3.33)-(3.34) directly that both the average out- and in-degree after link removals, which we will denote by $\bar{k}$, equal

$$\bar{k}=k(1-p). \tag{3.35}$$

Theorem 3. *Consider a directed network with k-regular out-degree and Poisson in-degree with average k. Then, after removing uniformly at random a fraction p of the links, the generating functions $\bar{H}_{out}(x)$ and $\bar{H}_{in}(x)$ of the excess out- and in-degree, respectively, satisfy*

$$\bar{H}_{out}(x) = (p + (1-p)x)^{k-1} \tag{3.36}$$

$$\bar{H}_{in}(x) = e^{-k(1-p)(1-x)} \tag{3.37}$$

The proof of Theorem 3 is given in Appendix A. Note that for the case without link removals, i.e. $p = 0$, Eqs.(3.36)-(3.37) reduce to Eqs.(3.19)-(3.20).

The results in Theorems 2 and 3 can also be directly deduced by using a result from [70]: if the generating function for the degree distribution for a network is given by $G(x)$, then the generating function $\bar{G}(x)$ for the resulting network after a fraction p of links are randomly removed, satsifies:

$$\bar{G}(x) = G(p + (1-p)x). \tag{3.38}$$

In addition, Theorem 3 can also be established directly by applying Eq.(3.5) to Eqs.(3.33)-(3.34).

We are now in a position to state the following result.

Theorem 4. *Consider a directed network with k-regular out-degree and Poisson in-degree with average k. Then, after removing uniformly at random a fraction p of the links, the fraction of minimum number of driver nodes is given by*

$$\begin{aligned} n_D = (p + (1-p)(1 - e^{-k(1-p)(1-w_2)}))^k - 1 + e^{-k(1-p)(1-w_2)} + \\ k(1-p)(1-w_2)e^{-k(1-p)(1-w_2)}, \end{aligned} \tag{3.39}$$

where w_2 satisfies

$$1 - w_2 = (p + (1-p)(1 - e^{-k(1-p)(1-w_2)}))^{k-1}. \tag{3.40}$$

The asymptotic behaviour of n_D for large k is given by

$$n_D \approx e^{-k(1-p)}. \tag{3.41}$$

For the case without link removals, i.e. $p = 0$, Eqs.(3.39)-(3.40)-(3.41) reduce to Eqs.(3.30)-(3.26)-(3.31), respectively. The proof of Theorem 4 is given in Appendix A.

Table 3.2 shows the comparison between the approximations in Eqs.(3.39) and (3.41) and simulations, for the cases $p = 0.2$ and $p = 0.5$.

We generated 1000 directed networks with $N = 10000$ with out-degree k, where$k \in \{1,2,3,4,5,6,7,8\}$. For each network with the same out-degree k, we randomly removed a fraction p of links and get the value of n_D, and then repeat this process for 1000 times. Thus, the fraction of driver nodes n_D for a combination (k, p) is the average fraction of driver nodes in 10^6 realizations.

As shown in Table 3.2, the absolute relative errors r of our approximation Eq.(3.39) are small which are less than 4% for all considered k values when $p = 0.2$ or $p = 0.5$. By

Table 3.2: Comparing Eqs.(3.39)-(3.41) with simulation results.

k	Eq.(3.39)				Eq.(3.41)				Simulation	
	$p=0.2$	r	$p=0.5$	r	$p=0.2$	r	$p=0.5$	r	$p=0.2$	$p=0.5$
1	0.442926	1.41%	0.584101	3.72%	0.449329	0.019%	0.606531	0.021%	0.449321	0.606622
2	0.238827	0.30%	0.410116	0.12%	0.201897	15.21%	0.367879	10.20%	0.238905	0.410229
3	0.116278	0.30%	0.279218	0.24%	0.090718	21.75%	0.22313	19.89%	0.116176	0.279108
4	0.050341	0.38%	0.183439	0.19%	0.040762	18.72%	0.135335	26.08%	0.050167	0.183421
5	0.021143	0.96%	0.112696	0.29%	0.018316	12.53%	0.082085	26.95%	0.021215	0.112680
6	0.009002	0.13%	0.065394	0.55%	0.00823	8.70%	0.049787	23.45%	0.009041	0.065339
7	0.003902	0.20%	0.037384	1.18%	0.003698	5.41%	0.030197	18.92%	0.003915	0.03736
8	0.001714	2.50%	0.021502	0.20%	0.001662	5.5%	0.018316	14.65%	0.001706	0.021533

contrast, the relative errors of the asymptotic approximation Eq.(3.41) are large for most cases.

We conclude from Table 3.2 that the simulations are an excellent fit with our approximation Eq.(3.39). Also, the asymptotic approximation Eq.(3.41) is increasingly accurate for increasing k, as expected .

Finally, Figure 3.2 shows the fraction of driver nodes n_D as function of the out-degree k for several values of p. The value of n_D decreases as the average degree of networks increases for a specific p. For the same k value, a larger value of p leads to a larger value of n_D.

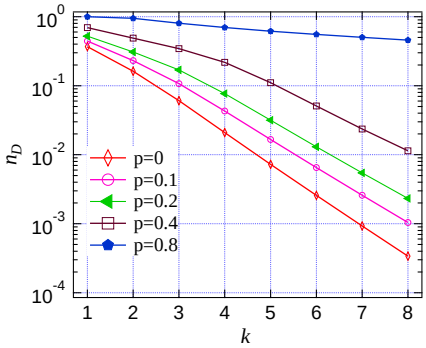


Figure 3.2: Fraction of driver nodes as function of the out-degree k for several values of the fraction of removed links p.

3.5. SSNs WITH BI-MODAL OUT-DEGREE

In this section we generalize the results of one of the previous sections by considering SSNs with bi-modal out-degree, i.e. we assume that for a fraction α of nodes the out-degree is k_1, while for the remaining $1-\alpha$ fraction of nodes, the out-degree equals k_2. We will assume $k_1 \neq k_2$ and both k_1 and k_2 are larger than 0.

Theorem 5. *Consider a directed network with bi-modal out-degree* $\alpha\delta(k_{out}-k_1)+(1-$

$\alpha)\delta(k_{out}-k_2)$, *with average out-degree*

$$k = \alpha k_1 + (1-\alpha)k_2 \tag{3.42}$$

and a Poisson in-degree distribution with average k. *The generating functions* $\hat{G}_{out}(x)$ *and* $\hat{G}_{in}(x)$ *of the out- and in-degree, respectively, satisfy*

$$\hat{G}_{out}(x) = \alpha x^{k_1} + (1-\alpha)x^{k_2}, \tag{3.43}$$

$$\hat{G}_{in}(x) = e^{-k(1-x)}. \tag{3.44}$$

The proof of Theorem 5 is given in Appendix B.

Theorem 6. *Consider a directed network with bi-modal out-degree* $\alpha\delta(k_{out}-k_1)+(1-\alpha)\delta(k_{out}-k_2)$, *with average out-degree*

$$k = \alpha k_1 + (1-\alpha)k_2 \tag{3.45}$$

and a Poisson in-degree distribution with average k. *Then, the generating functions* $\hat{H}_{out}(x)$ *and* $\hat{H}_{in}(x)$ *of the excess out- and in-degree, respectively, satisfy*

$$\hat{H}_{out}(x) = \frac{\alpha k_1 x^{k_1-1} + (1-\alpha)k_2 x^{k_2-1}}{k}, \tag{3.46}$$

$$\hat{H}_{in}(x) = e^{-k(1-x)}. \tag{3.47}$$

The proof of Theorem 6 is given in Appendix B. The proof also can be established by applying Eq.(3.5) directly to Eqs.(3.43)-(3.44). Note for the case $k_1 = k_2 = k$, where the out-degree reduces to a Dirac function, Eqs.(3.43)-(3.47) reduce to Eqs.(3.17)-(3.20).

Theorem 7. *Consider a directed network with bi-modal out-degree* $\alpha\delta(k_{out}-k_1)+(1-\alpha)\delta(k_{out}-k_2)$, *with average out-degree* $k = \alpha k_1 + (1-\alpha)k_2$ *and Poisson in-degree with average* k. *Then, the fraction of minimum number of driver nodes is given by*

$$n_D = \alpha(1-e^{-k(1-w_2)})^{k_1} + (1-\alpha)(1-e^{-k(1-w_2)})^{k_2} - 1 + e^{-k(1-w_2)} + ke^{-k(1-w_2)}(1-w_2), \tag{3.48}$$

where w_2 *satisfies*

$$1 - w_2 = \frac{\alpha k_1 (1-e^{-k(1-w_2)})^{k_1-1} + (1-\alpha)k_2(1-e^{-k(1-w_2)})^{k_2-1}}{k}. \tag{3.49}$$

The asymptotic behaviour of n_D *for large* k *is given by*

$$n_D \approx e^{-k}. \tag{3.50}$$

Note for the case $k_1 = k_2 = k$, where the out-degree reduces to a Dirac function, Eqs. (3.48)-(3.49)-(3.50) reduce to Eqs.(3.30)-(3.26)-(3.31), respectively.

The proof of Theorem 7 is given in Appendix B.

Table 3.3 shows the comparison between the approximations in Eqs.(3.48) and (3.50) and simulations.

We generate 1000 directed networks with $N = 10000$ for each out-degree combination (k_1, k_2, α). For each network with the same out-degree combination (k_1, k_2, α), we randomly remove a fraction p of links and get the value of n_D, and then repeat this process for 1000 times. Thus, the fraction of driver nodes n_D for a combination (k_1, k_2, α) is the average fraction of driver nodes from 10^6 realizations. As shown in Table 3.3, the absolute relative errors r of our approximation Eq.(3.48) are small indicating a good fit with simulations. The absolute relative errors of Eq.(3.50) are larger, especially for small average degree.

Table 3.3: Comparing Eqs.(3.48)-(3.50) with simulation results.

$\mathbf{k_1}$	$\mathbf{k_2}$	**k**	α	**Eq.(3.48)**		**Eq.(3.50)**		**Simulation**
				value	r	**value**	r	
1	3	2.5	0.25	0.107746	0.51%	0.082085	23.43%	0.107795
1	3	2	0.5	0.183062	0.020%	0.135335	26.09%	0.181395
1	3	1.5	0.75	0.273670	0.040%	0.223130	18.44%	0.273455
2	4	3.5	0.25	0.036402	0.56%	0.030197	16.58%	0.036705
2	4	3	0.5	0.063648	0.27%	0.049787	21.57%	0.06352
2	4	2.5	0.75	0.106955	0.25%	0.082085	23.44%	0.106735
2	6	5	0.25	0.007355	1.04%	0.006738	9.30%	0.007315
2	6	4	0.5	0.022172	0.76%	0.018316	16.76%	0.022335
2	6	3	0.75	0.071349	0.19%	0.049787	30.09%	0.071875
2	8	6.5	0.25	0.001555	3.81%	0.001503	0.33%	0.001595
2	8	5	0.5	0.007556	2.20%	0.006738	8.86%	0.007745
2	8	3.5	0.75	0.045382	0.35%	0.030197	33.69%	0.04665
4	6	5.5	0.25	0.004324	0.68%	0.004087	4.84%	0.004362
4	6	5	0.5	0.007293	0.97%	0.006738	6.71%	0.007181
4	6	4.5	0.75	0.012357	1.40%	0.011109	8.34%	0.01228
4	8	7	0.25	0.000931	3.22%	0.000912	5.20%	0.000962
4	8	6	0.5	0.002593	4.18%	0.002479	4.36%	0.002706
4	8	5	0.75	0.007354	1.17%	0.006738	7.56%	0.007269

We conclude from Table 3.3 that the simulations are an excellent fit with our approximation Eq.(3.48). Also, the asymptotic approximation Eq.(3.50) is increasingly accurate for increasing k.

3.6. SSNs WITH BI-MODAL OUT-DEGREE AND RANDOM LINK FAILURES

In this section we generalize the results of the previous section by considering again SSNs with bi-modal out-degree, but now we assume that a fraction p of the links is removed at random. We will show that the analysis that led to our implicit approximations can also be conducted for this case. Similar to the case for regular out-degree, a crucial step is to

find expressions for the generating functions Eqs.(1)-(4) for this case.

Based on Lemma 1, we get:

Theorem 8. *Consider a directed network with bi-modal out-degree* $\alpha\delta(k_{out}-k_1)+(1-\alpha)\delta(k_{out}-k_2)$*, with average out-degree*

$$k = \alpha k_1 + (1-\alpha)k_2 \tag{3.51}$$

and Poisson in-degree with average k*. Then, after removing uniformly at random a fraction* p *of the links, the generating functions* $\tilde{G}_{out}(x)$ *and* $\tilde{G}_{in}(x)$ *of the out- and in-degree, respectively, satisfy*

$$\tilde{G}_{out}(x) = \alpha(p+(1-p)x)^{k_1} + (1-\alpha)(p+(1-p)x)^{k_2} \tag{3.52}$$

$$\tilde{G}_{in}(x) = e^{-k(1-p)(1-x)} \tag{3.53}$$

By applying the generating function $\tilde{G}(x)$ for the resulting network after a fraction p of links are randomly removed [70], the theorem also follows directly from $\tilde{G}_{out}(x) = \hat{G}_{out}(p+(1-p)x)$ and $\tilde{G}_{in}(x) = \hat{G}_{in}(p+(1-p)x)$. Note that for the case without link removals, i.e. $p=0$, Eqs.(3.52)-(3.53) reduce to Eqs.(3.43)-(3.44). Also, we can deduce from Eqs.(3.52)-(3.53) directly that both the average out- and in-degree after link removals, which we will denote by $\tilde{k}$, equal

$$\tilde{k} = k(1-p) \tag{3.54}$$

Theorem 9. *Consider a directed network with bi-modal out-degree* $\alpha\delta(k_{out}-k_1)+(1-\alpha)\delta(k_{out}-k_2)$*, with average out-degree*

$$k = \alpha k_1 + (1-\alpha)k_2 \tag{3.55}$$

and Poisson in-degree with average k*. Then, after removing uniformly at random a fraction* p *of the links, the generating functions* $\tilde{H}_{out}(x)$ *and* $\tilde{H}_{in}(x)$ *of the excess out- and in-degree, respectively, satisfy*

$$\tilde{H}_{out}(x) = \frac{\alpha k_1(p+(1-p)x)^{k_1-1} + (1-\alpha)k_2(p+(1-p)x)^{k_2-1}}{k} \tag{3.56}$$

$$\tilde{H}_{in}(x) = e^{-k(1-p)(1-x)} \tag{3.57}$$

The proof of Theorem 9 can be obtained by combining the proofs of Theorems 3 and 6. By applying the generating function $\tilde{G}(x)$ for the resulting network after a fraction p of links are randomly removed [70], the theorem also follows directly from $\tilde{H}_{out}(x) = \hat{H}_{out}(p+(1-p)x)$ and $\tilde{H}_{in}(x) = \hat{H}_{in}(p+(1-p)x)$. Note that for the case without link removals, i.e. $p=0$, Eqs.(3.56)-(3.57) reduce to Eqs.(3.46)-(3.47).

After obtaining expressions for all required generation functions, we are now in a position to state the following result.

Theorem 10. *Consider a directed network with bi-modal out-degree* $\alpha\delta(k_{out}-k_1)+(1-\alpha)\delta(k_{out}-k_2)$*, with average out-degree* $k=\alpha k_1+(1-\alpha)k_2$ *and Poisson in-degree with average* k*. Then, after removing uniformly at random a fraction* p *of the links, the fraction of minimum number of driver nodes is given by:*

$$\begin{aligned} n_D = \alpha(p+(1-p)(1-e^{-k(1-\omega_2)}))^{k_1} + (1-\alpha)(p+(1-p)(1-e^{-k(1-\omega_2)}))^{k_2} \\ -1+e^{-k(1-p)(1-\omega_2)}+k(1-p)e^{-k(1-\omega_2)}(1-\omega_2) \end{aligned} \tag{3.58}$$

where ω_2 *satisfies*

$$1-\omega_2 = \frac{\alpha k_1(p+(1-p)(1-e^{-k(1-p)(1-\omega_2)}))^{k_1-1}+(1-\alpha)k_2(p+(1-p)(1-e^{-k(1-p)(1-\omega_2)}))^{k_2-1}}{k} \tag{3.59}$$

The asymptotic behaviour of n_D *for large* k *is given by*

$$n_D \approx e^{-k(1-p)} \tag{3.60}$$

For the case without link removals, i.e. $p=0$, Eqs.(3.58)-(3.60) reduce to Eqs.(3.48)-(3.50).

The proof of Theorem 10 is given in Appendix B.

To verify our approximation Eq.(3.58), we generate 1000 directed networks with $N=10000$ for each out-degree combination (k_1,k_2,α). For each network with the same out-degree combination (k_1,k_2,α), we randomly remove a fraction p of links and get the value of n_D, and then repeat this process for 1000 times. Thus, the fraction of driver nodes n_D for a combination (k_1,k_2,α,p) is the average fraction of driver nodes in 10^6 realizations.

Table 3.4 shows the comparison between Eq.(3.58) and simulations. In most cases, the relative errors between Eq.(3.58) and simulations are small. We conclude from Table 3.4 that the simulations are an excellent fit with our approximation Eq.(3.58).

3.7. DISCUSSION

In this chapter, we correct the formula given in [68] for the minimum number of driver nodes for a specific class of swarm signalling networks, which are characterised by a regular out-degree. We then generalize the results by considering SSNs with a regular out degree k where a fraction p of the links is unavailable. For this case we derive an implicit equation, whose solution leads to the minimum number of driver nodes. We find that our approximation fits well with simulation results. Finally, we relax the condition that the out-degree is regular and look into bi-modal out-degree distributions. For this case we also consider scenarios with unavailable links. We derive an implicit equation and verify its accuracy. We find that our approximation for bi-modal out-degree distribution fits well with simulation results.

Table 3.4: Comparing the approximation Eq.(3.58) with simulation results.

$\mathbf{k}_1$	$\mathbf{k}_2$	$\mathbf{k}$	α	**Eq.(3.58)**				**Simulation**	
				$p=0.2$	r	$p=0.5$	r	$p=0.2$	$p=0.5$
1	3	2.5	0.25	0.251484	0.50%	0.541569	0.19%	0.252746	0.540569
1	3	2	0.5	0.340662	0.41%	0.627028	1.29%	0.342065	0.619028
1	3	1.5	0.75	0.431100	0.29%	0.709013	2.21%	0.432370	0.693714
2	4	3.5	0.25	0.122113	0.25%	0.410770	0.29%	0.121813	0.409569
2	4	3	0.5	0.183813	1.32%	0.476848	0.43%	0.186273	0.474822
2	4	2.5	0.75	0.247667	0.80%	0.535501	0.43%	0.245692	0.537824
2	6	5	0.25	0.033257	0.87%	0.299464	0.52%	0.033549	0.297913
2	6	4	0.5	0.094631	1.86%	0.435961	1.48%	0.096426	0.429607
2	6	3	0.75	0.216405	0.93%	0.514376	3.06%	0.218443	0.499125
2	8	6.5	0.25	0.008650	0.06%	0.101497	17.49%	0.008655	0.123010
2	8	5	0.5	0.037450	0.81%	0.406573	0.15%	0.037150	0.405974
2	8	3.5	0.75	0.204397	0.017%	0.505728	1.00%	0.204363	0.510815
4	6	5.5	0.25	0.020441	5.68%	0.163736	0.14%	0.021671	0.163504
4	6	5	0.5	0.032167	4.57%	0.229064	0.44%	0.033706	0.228061
4	6	4.5	0.75	0.050380	1.00%	0.288043	0.80%	0.049880	0.285759
4	8	7	0.25	0.005504	1.47%	0.058532	0.33%	0.005586	0.058338
4	8	6	0.5	0.013368	0.077%	0.135230	0.42%	0.013357	0.134664
4	8	5	0.75	0.033187	0.27%	0.265665	0.93%	0.033275	0.263211

4

USING MACHINE LEARNING TO QUANTIFY THE ROBUSTNESS OF NETWORK CONTROLLABILITY

This chapter presents machine learning based approximations for the minimum number of driver nodes needed for structural controllability of networks under link-based random and targeted attacks. We compare our approximations with existing analytical approximations and show that our machine learning based approximations significantly outperform the existing closed-form analytical approximations in case of both synthetic and real-world networks. Apart from targeted attacks based upon the removal of so-called critical links, we also propose analytical approximations for out-in degree-based attacks.

4.1. INTRODUCTION

In the modern world, we see networks everywhere such as the Internet, transportation networks, and communication networks [72]. It is important that these networks perform their desired functions properly. Naturally, we need to control these networks to ensure their proper functioning and maintenance. Network science offers a way to study and analyze these networks using graph theory. The entities in a network are represented by the nodes and interconnections between the nodes are represented by links. For example, in an air-transportation network, the nodes represent different airports and the links represent the flight paths that connect these airports. Network controllability is the ability to drive a system from an initial state to any other state in a finite time by application of external inputs on certain nodes [26]. For directed networks, Liu *et al.* [39] showed that the minimum number of nodes required to control a network can be identified through the maximum matching of the network. However, Cowan *et al.* [73] pointed out that the results of Liu *et al.* [39] are based on the assumption of no self-links. In

This chapter is based on the published paper [71].

other words, a state of a node can only be changed through interacting with its adjacent nodes. In Chapter 2, we derived closed-form analytical approximations for the minimum number of driver nodes as a function of the fraction of removed links for both random and targeted attacks [36]. However, the approximations sometimes do not fit well with the simulations, especially when the fraction of removed links is not small. Figure 4.1 shows the performance of Sun's approximation as compared to simulation for a Erdős-Rényi network under targeted attack.

The objective of this work is to improve the analytical approximations for both random and targeted attacks using machine learning methods. We will compare our machine learning based approximations with the existing analytical approximations and simulations. Furthermore, we will also derive an analytical approximation for out-in degree-based attacks and evaluate its performance on both synthetic and real-world networks.

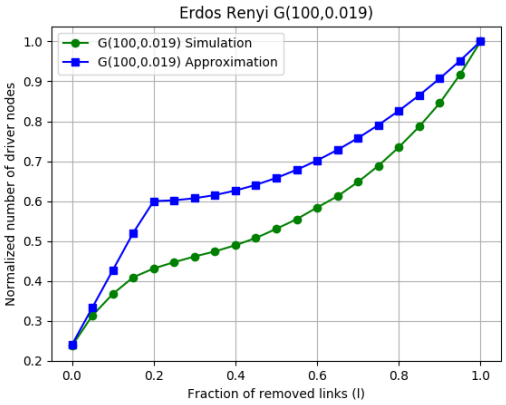


Figure 4.1: Performance comparison of Sun's approximation for the normalized minimum number of driver nodes as a function of the fraction of removed links in a Erdős-Rényi network under targeted attack.

In the remainder of this chapter, in Section 4.2 we describe the concept of network robustness. In Section 4.3, we discuss the closed-form analytical approximations for the minimum number of driver nodes given in [36]. Machine learning methods and information related to training and testing data are discussed in Section 4.4. Machine learning based approximations for both random and targeted attacks are presented in Section 4.5. An analytical approximation for out-in degree-based attacks is also derived in this section. Additionally, we also analyze and compare our machine learning based approximations with the approximations and simulation results obtained in Chapter 2. Finally, in Section 4.6 we conclude this chapter.

4.2. NETWORK ROBUSTNESS

Network robustness is the ability of a network to deal with failures and errors. In real-world networks, we encounter various failures such as power transmission line failures in an electrical network and network disruption due to natural disasters. It is important

to make networks robust to deal with such failures. A generic quantitative definition of network robustness does not exist but there are various metrics to assess network robustness depending on the type of network and its purpose. In this work, we assess network robustness in terms of controllability. Network robustness under perturbations has been studied extensively. Socievole *et al.* [49] studied network robustness in case of epidemic spreads. They investigated Susceptible-Infected-Susceptible (SIS) spreads with N-Intertwined Mean-Field Approximation (NIMFA) epidemic threshold as the robustness metric. Trajanovski *et al.* [52] considered node removals in both random and targeted attacks to study network robustness. They used two metrics to evaluate the network robustness, the size of the giant component and efficiency. Wang *et al.* [20] considered effective graph resistance as the robustness metric to investigate network robustness in case of both synthetic and real-world networks. Koç *et al.* [74] studied the robustness of networks in terms of cascading failures that lead to blackouts in electrical power grids.

Real-world networks are often challenged by perturbations in the form of random and targeted attacks [75]. In this work, we simulate these attacks by removing links. We do not consider node removals. Random attacks are the unintentional failures such as disruption of networks due to natural disasters and failures due to exhausted mechanical parts [15]. Targeted attacks are carried out by people with malicious intent to maximize the damage [47] [76] [22] [77]. In targeted attacks, it is assumed that the attacker has the information related to network topology, functions and vulnerabilities.

4.3. ANALYTICAL APPROXIMATIONS

The analytical approximations for random and targeted link removals [36] in Chapter 2 are based on the concept of critical links. If the number of driver nodes required to control a network increases when removing a specific link, then that link is called a critical link. A link that does not belong to any maximum matching is dubbed a redundant link. A link that is neither critical nor redundant is an ordinary link. The initial number of driver nodes N_{DO} i.e. the number of driver nodes before any attack, is calculated using the Hopcroft-Karp algorithm [28]. To find the number of critical links, each link in a network is removed one by one and the Hopcroft-Karp algorithm [28] is applied simultaneously. If the current number of driver nodes N_D exceeds the initial number of driver nodes N_{DO}, then the removed link is a critical link. In a network with N nodes and L links, the Hopcroft-Karp algorithm [28] is applied L times to identify all the critical links.

4.3.1. NUMBER OF DRIVER NODES UNDER RANDOM ATTACKS

As discussed in Chapter 2, for random attacks, the normalized minimum number of driver nodes is expressed as,

$$n_{D,rand} = \begin{cases} \frac{N_{DO}+lL_C}{N}, & l \leq l_C \\ al^2 + bl + c, & l \geq l_C \end{cases} \tag{4.1}$$

where $n_{D,rand}$ represents the normalized value of the minimum number of driver nodes required to fully control a network, L_C represents the number of critical links, l represents the fraction of removed links and $l_C = \frac{L_C}{L}$ represents the fraction of critical links. The values of a, b and c are derived from the boundary conditions described in [36] such that $a = \frac{N-N_{DO}-L_C}{N(l_C-1)^2}$, $b = \frac{L_C}{N} - 2al_C$ and $c = 1 - \frac{L_C}{N} + a(2l_C - 1)$.

4.3.2. NUMBER OF DRIVER NODES UNDER TARGETED ATTACKS

In targeted attacks, first we randomly remove all the critical links and then the remaining links. Sun *et al.* [36] derived the following analytical approximation for targeted attacks.

$$n_{D,crit} = \begin{cases} \frac{N_{DO}+lL}{N}, & l \leq l_C \\ dl^2 + el + f, & l \geq l_C \end{cases} \tag{4.2}$$

where d, e and f are derived from the boundary conditions described in [36] such that $d = \frac{N-N_{DO}-l_C L}{N(l_C-1)^2}$, $e = -2dl_C$ and $f = 1 + d(2l_C - 1)$.

4.4. MACHINE LEARNING

Machine learning is a technique to predict the outcome of a certain event by learning from data. The data could already be available from experiments, data centers or it can be generated through proper simulations. There are numerous applications of machine learning such as predicting customer's buying habits based on historical data in e-Commerce, weather forecasts and Virtual Personal Assistants such as Siri and Alexa. In broader terms, machine learning is classified as supervised learning, unsupervised learning and reinforcement learning. Furthermore, supervised machine learning is divided into classification and regression problems. In this work, we use various supervised learning methods for regression problems to predict the number of driver nodes under various attacks. Specifically, we use Linear Regression, Random Forest and Artificial Neural Networks. Recently, Lou *et al.* [78] also investigated the use of neural networks for network controllability. However, they used another type of neural networks, Convolution Neural Networks.

To develop our machine learning models, various hyper-parameters are used. Table 4.1 and Table 4.2 shows the number of hidden layers and other hyper-parameters that are used to develop our ANN models. For our linear regression model, we use the least-squares to minimize the errors. Additionally, we also use k-fold cross-validation with $k = 10$ to check for over-fitting. In our Random-Forest model, we select the number of trees as 50. Moreover, we also use feature importance scores to determine the features that contribute more to the output. A detailed explanation of the choice of hyper-parameters is presented in the master thesis report [79].

Table 4.1: Selection of ANN size for different networks under targeted, random and out-in degree-based attacks.

Attack	**Number of hidden layers**		
	Real-world	**Erdős-Rényi**	**Barabási-Albert**
Targeted critical link attack	512/512/512	128	512/512/512
Random attack	512/512/512	128/512/512/512	128/512/512/512
Out-in degree based attack	512/512/512	128	512/512/512

4.4.1. DATASET FOR REAL-WORLD NETWORKS

Now we discuss the real-world dataset that we consider to construct our models. For synthetic networks, we generate data through simulations. We use the dataset available at The Internet Topology Zoo [80] for real-world networks. It is a collection of a publicly

Table 4.2: ANN hyper-parameters selection.

Hyper-parameters	Activation Function	Loss Function	Dropout rate	Early Stopping	Patience	Epochs	Batch size
Selection	ReLU	MSE	0.2	Yes	50	300	32

Table 4.3: Properties of 10 real-world networks used for testing our models.

Network	N	L	L_C	N_{DO}
Colt	153	177	38	81
Surfnet	50	68	23	15
EliBackbone	20	30	12	5
Garr200912	54	68	9	30
GtsPoland	33	37	12	14
Ibm	18	24	6	6
Arpanet19706	9	10	6	2
GtsHungary	30	31	8	18
BellCanada	48	64	17	16
Uninet	69	96	19	4

accessible dataset provided by different network operators. As the networks evolve and change, the dataset is updated and in this sense, it is not fixed. Network operators provide maps of their networks and this dataset is interpreted from those maps. However, there are various ambiguities in the dataset as the interpretations are not accurate for some networks. The dataset is available in Graph Markup Language (GML) [81] and GraphML [82] formats. In this work, we consider the dataset that is available in GraphML format as it is easy to parse using python's NetworkX library. We pre-process the data to remove any disconnected networks and multigraphs. After pre-processing of the dataset, we have 232 networks out of which we use 192 networks for training and the remaining 40 networks for testing. The networks in the dataset are not directed, however, we use the information available in two attributes of the GraphML format, edge source and target, to make these networks directed.

The networks in the dataset have small average degrees. The smallest network is the Arpanet196912 network with 4 nodes and 4 links. Cogentco network is the largest network with 197 nodes and 243 links. Additionally, there are some networks that have zero critical links. We conclude that the networks in this dataset vary a lot and machine learning models might have difficulties in learning from such a varying dataset. Table 4.3 lists the properties of some of the real-world networks we use for testing.

4.4.2. DATASETS FOR SYNTHETIC NETWORKS

We generate data for synthetic networks using simulations. We consider two types of synthetic networks, Erdős-Rényi and Barabási-Albert networks. These networks come under the class of random graphs [83]. In Erdős-Rényi (ER) random graphs $G(N, p)$ [84], N denotes the number of nodes and p denotes the probability of an outbound link from a node to another node. For Erdős-Rényi networks, we generate networks with different

values of N and p. For each such network, we generate 100 corresponding networks and determine the average values of network characteristics such as the average degree, the average number of links, the number of critical links and graph metrics such as diameter and clustering coefficient.

In the Barabási-Albert (BA) scale-free model $G(N, M)$ [85] [86], N indicates the number of nodes and M indicates the number of links of a new node that attaches itself to the original network. To generate a BA network, we assume a complete digraph of M_O nodes where M_O equals M. Then we add new nodes one by one with a probability proportional to the number of links of the existing nodes. We generate BA networks with different values of N and M using simulations. For each BA network, we also generate 100 corresponding networks to get the average values of the network characteristics such as the average degree, the average number of links, the average number of critical links and graph metrics such as diameter and clustering coefficient. Moreover, it is to be noted that in a targeted critical link attack, first, the critical links are removed randomly and then the remaining links. For such random removal of links, we use 10000 simulations. Furthermore, in random attacks, all the links are removed uniformly at random and we also use 10000 simulations to get the average values of the minimum number of driver nodes.

4.5. MEASURING THE ROBUSTNESS OF NETWORK CONTROLLABILITY USING MACHINE LEARNING

4.5.1. TARGETED CRITICAL LINK ATTACK

To develop a machine learning based approximation for targeted critical link attack, we predict the difference in the normalized minimum number of driver nodes between the simulation value and the analytical approximation Eq.(4.2) at $l = l_C$. We use various input features such as the number of nodes N, number of links L, number of critical links L_C, clustering coefficient, average degree and diameter. We choose to estimate the difference at l_C as the original approximation fits well with the simulation for $l \ll l_C$ [36], while the difference can be significant at $l = l_C$, see also Figure 4.1 , where $l_c = 0.2$. We subtract this predicted difference to get a new value n_{DX} that is closer to the simulation. We assume a linear relationship similar to the analytical approximation Eq.(4.2) for $l \leq l_C$. The value of the normalized minimum number of driver nodes at $l = 0$ is n_{DO} where, $n_{DO} = \frac{N_{DO}}{N}$ and at $l = l_C$, the value is assumed to be n_{DX}. From these two conditions we get,

$$n_{D,crit,ML} = n_{DO} + \frac{n_{DX} - n_{DO}}{l_C} l, \tag{4.3}$$

where $n_{D,crit,ML}$ gives us the new machine learning based normalized minimum number of driver nodes for $l \leq l_C$. When the fraction of removed links l is greater than or equal to the fraction of critical links l_C i.e. for $l \geq l_C$, we estimate the normalized minimum number of driver nodes using a parabolic approximation of the form,

$$n_{D,crit,ML} = d_{ML} l^2 + e_{ML} l + f_{ML}, \tag{4.4}$$

where d_{ML}, e_{ML} and f_{ML} are derived from the boundary conditions. For the first boundary condition, $n_{D,crit,ML}$ equals n_{DX} at $l = l_C$. When all the links are removed, we need to control all the nodes. Hence, at $l = 1$, $n_{D,crit,ML}$ equals one. Finally, for the third

boundary condition, we assume the derivative of the parabola is zero at $l = l_C$. Using these boundary conditions, we get $d_{ML} = \frac{1-n_{DX}}{l_C^2-2l_C+1}$, $e_{ML} = -2d_{ML}l_C$ and $f_{ML} = 1 + d_{ML}(2l_C - 1)$. Finally, the machine learning based approximation for targeted attacks can be expressed as,

$$n_{D,crit\_ML} = \begin{cases} n_{DO} + \frac{n_{DX}-n_{DO}}{l_C} l, & l \leq l_C \\ d_{ML}l^2 + e_{ML}l + f_{ML}, & l \geq l_C \end{cases} \tag{4.5}$$

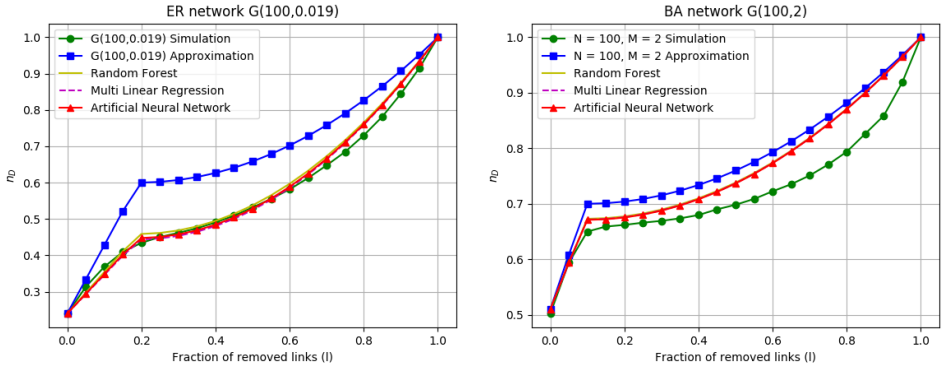


Figure 4.2: Comparison of different methods to get the normalized values of minimum number of driver nodes n_D needed to control the network as a function of the fraction of removed links in synthetic networks under targeted attacks. Simulations are based on 10,000 realizations of attacks.

In Figure 4.2, we compare the performance of linear regression, random forest and artificial neural network models with simulation and analytical approximation Eq.(4.2) for synthetic networks under targeted attacks. We notice that the machine learning based approximation fits better with the simulations than the analytical approximation Eq.(4.2). To further quantify the performance, we use mean absolute errors and mean relative errors to compare the performance of different approximations. Table 4.4 compares the performance of ANN with the analytical approximation Eq.(4.2) for a few synthetic networks. We observe that the mean relative error decreases from 19.07 % to 2.13 % using the ANN-based approximation for ER network with $N = 100$ and $p = 0.019$. For BA network with $N = 100$ and $M = 2$, we see an improvement from 7.04 % to 4.67 %. Furthermore, the mean relative errors are larger for Barabási-Albert networks as compared to Erdős-Rényi networks. This is because, in BA networks, there are a few nodes with high degrees, so even after removal of some links, the minimum number of driver nodes does not change significantly and hence, the curve is less steep in BA networks as compared to ER networks as also evident from Figure 4.2.

Next, we evaluate the performance of machine learning based approximation for real-world networks under targeted attacks. The model is trained on 192 real-world networks and tested on 40 networks. Figure 4.3 shows that machine learning based curves fit better with the simulations than the analytical approximation Eq.(4.2) for Colt and Surfnet network. We also compare the performance of different machine learning models

Table 4.4: Performance indicators for synthetic networks under targeted attacks.

Network	Mean Absolute Error		Mean Relative Error	
	Approximation	ANN	Approximation	ANN
ER(100, 0.019)	0.1000	0.0124	0.1907	0.0213
ER(200, 0.0063)	0.0663	0.0115	0.1008	0.0175
ER(400, 0.0026)	0.0472	0.0046	0.0659	0.0071
BA(50, 2)	0.0590	0.426	0.0821	0.0582
BA(100, 2)	0.051	0.0351	0.0704	0.0467

based on the root mean squared errors (RMSE). The RMSE values are found to be 0.0723, 0.0550 and 0.0430 for linear regression, random forest and artificial neural network model respectively. We observe that the ANN model performs slightly better than the random forest model. The linear regression model performs the least amongst the three machine learning models. This can be explained based on the non-linear relationship between the input features and the difference that we predict.

In Table 4.5, we compare the performance of the ANN-based approximation and the analytical approximation Eq.(4.2) for 10 real-world networks. We notice that machine learning based approximation performs the best in the case of the Colt network with a mean relative error of 1.46 % and the worse in Ibm network with a mean relative error of 8.3 %. Furthermore, we observe that 9 out of 10 networks have mean relative errors of less than 5 %. Among the 40 test networks, the machine learning based approximation performs better than the analytical approximation Eq.(4.2) in 30 networks. For the remaining 10 networks, the analytical approximation performs only slightly better with a difference of less than 2 %.

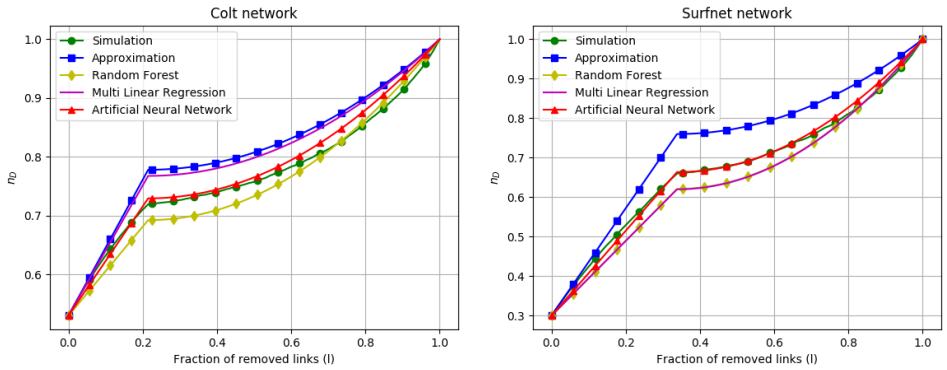


Figure 4.3: Comparison of different methods to get the normalized values of minimum number of driver nodes n_D needed to control the network as a function of the fraction of removed links in real-world under targeted attacks. Simulations are based on 10,000 realizations of attacks.

Table 4.5: Performance indicators for real-world networks under targeted attacks.

Network	Mean Absolute Error		Mean Relative Error	
	Approximation	ANN	Approximation	ANN
Colt	0.0393	0.0116	0.0512	0.0146
Surfnet	0.0597	0.0095	0.0866	0.0151
EliBackbone	0.1468	0.0201	0.2471	0.0376
Garr200912	0.0223	0.0202	0.0277	0.0251
GtsPoland	0.0266	0.0171	0.0335	0.0235
Ibm	0.0595	0.0519	0.0956	0.0832
Arpanet19706	0.0440	0.0255	0.0588	0.0434
GtsHungary	0.0269	0.0321	0.0311	0.0373
BellCanada	0.0502	0.0135	0.0757	0.0230
Uninet	0.1195	0.0309	0.184	0.0485

4.5.2. RANDOM ATTACK

In this section, we develop a machine learning based approximation for the normalized minimum number of driver nodes as a function of the fraction of removed links for random attacks. Furthermore, we compare our approximation with the analytical approximation Eq.(4.1) and simulations. We also evaluate the performance of different machine learning algorithms. For real-world networks, the RMSE comes out to be 0.0165 for the ANN model and 0.0192 for the random forest model. Again, the ANN model performs slightly better in terms of RMSE. In the remainder of this section, we will only consider ANN. For random attacks, we predict the normalized minimum number of driver nodes for different values of the fraction of removed links starting with $l = 0$ to $l = 1$ in steps of 0.05. In other words, for each value of N and p in ER networks, 21 data points are generated for training. The same approach is followed for BA networks for each N and M value. The reason for such an approach is that at l_C, the difference between the approximation value and the simulation value is not significant as the approximation fits well for $l \leq l_C$ [36].

Next, we compare our machine learning based approximation for random attacks with the analytical approximation Eq.(4.1) and simulation. Figure 4.4 shows that the ANN curves fit better with the simulations for both Erdős-Rényi and Barabási-Albert networks. To quantify this improvement, Table 4.6 compares the performance of ANN and analytical approximation Eq.(4.1) based on the mean absolute errors and mean relative errors. We notice a significant improvement in mean relative error from 30.80 % to 6.75 % for ER network with $N = 50$ and $p = 0.082$ using ANN. Similarly, we see an improvement from 13.70 % to 0.44 % in the mean relative error in ER network with $N = 100$ and $p = 0.016$. Furthermore, for BA network with $N = 100$ and $M = 2$, the mean relative error improves from 4.55 % to 0.49 %.

Specifically for ER networks under random attacks, Liu *et al.* [39] also derived an approximation based on generating functions. According to Liu *et al.* [39], the normalized minimum number of driver nodes is given by,

Table 4.6: Performance indicators for synthetic networks under random attacks.

Network	Mean Absolute Error		Mean Relative Error	
	Approximation	ANN	Approximation	ANN
ER(50, 0.082)	0.0712	0.0105	0.3080	0.0675
ER(100, 0.016)	0.0085	0.0024	0.0137	0.0044
BA(50, 2)	0.035	0.0032	0.0517	0.0051
BA(100, 2)	0.032	0.0030	0.0455	0.0049

$$n_D = w_1 - w_2 + k(1-l)w_1(1-w2), \tag{4.6}$$

where k is the average out-degree of an ER network expressed as $k = p(N-1)$. The solution of the implicit equation $w_1 = e^{-k(1-l)e^{-k(1-l)w_1}}$ gives us the value of w_1 and w_2 is given by, $w_2 = 1 - e^{-k(1-l)w_1}$.

Table 4.7: Performance indicators for all three approximations for ER networks under random attacks.

Network	Mean Relative Error		
	Approximation by Sun *et al.* Eq.(4.1)	ANN	Approximation by Liu *et al.* Eq.(4.6)
ER(100, 0.015)	0.0162	0.0084	0.0045
ER(100, 0.017)	0.0156	0.0097	0.0020
ER(200, 0.006)	0.0117	0.0059	0.0018

Now we will compare our ANN-based approximation with Sun's approximation Eq.(4.1), Liu's approximation Eq.(4.6) and simulations. From Table 4.7, it is evident that Liu's approximation Eq.(4.6) outperforms both ANN based approximation and Sun's approximation Eq.(4.1). In $ER(100, 0.015)$ network, the mean relative error using Sun's approximation Eq.(4.1) comes out to be 1.62 %. Our ANN based approximation and Liu's approximation Eq.(4.6) both performs better than Sun's approximation Eq.(4.1) with mean relative errors of 0.84 % and 0.45 % respectively.

We note that Liu's approximation is based upon the use of generating functions for the degree and excess degree distribution, whose expressions are not known for targeted link removals.

For real-world networks under random attacks, we follow a different approach. Here we do not predict the normalized minimum number of driver nodes for the entire range of the fraction of removed links. This is because of the availability of a limited dataset for training and hence, the model always performs worse than the analytical approximation. Moreover, difference estimation at l_C is also not a suitable choice as the original analytical approximation is already good for $l \leq l_C$ [36]. For larger values of the fraction of removed links, the difference in n_D values between the approximation and simulation is significant. So, we choose a point $l = 0.4$ to predict the difference and subtract it from the approximation value to get a new value n_{DX}. Let the value at $l = 0.4$ be l_X. At $l = 0$, the normalized minimum number of driver nodes equals n_{DO} and at $l = 0.4$, n_D equals n_{DX}. From these two points we get,

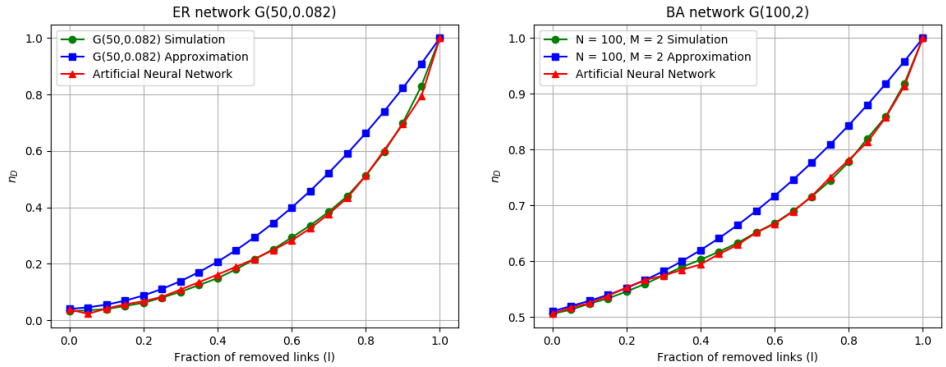


Figure 4.4: Comparison of different methods to get the normalized values of minimum number of driver nodes n_D needed to control the network as a function of the fraction of removed links in synthetic networks under random attacks. Simulations are based on 10,000 realizations of attacks.

$$n_{D,rand,ML} = n_{DO} + \frac{n_{DX} - n_{DO}}{l_X} l, \tag{4.7}$$

where, $n_{D,rand,ML}$ gives the normalized minimum number of driver nodes as a function of the fraction of removed links for $l \leq l_X$. For l values greater than or equal to l_X, we calculate the normalized minimum number of driver nodes using a parabolic approximation,

$$n_{D,rand,ML} = a_{ML} l^2 + b_{ML} l + c_{ML}, \tag{4.8}$$

where we derive the values of a_{ML}, b_{ML} and c_{ML} from the boundary conditions. At $l = l_X$, the value and derivative of Eq.(4.8) equals that of Eq.(4.7). Hence, we get $a_{ML} l_X^2 + b_{ML} l_X + c_{ML} = n_{DX}$ and $2a_{ML} l_X + b_{ML} = \frac{n_{DX} - n_{DO}}{l_X}$. At $l = 1$ i.e. when all the links are removed, we need to control all the nodes. Hence, n_D equals one and we get, $a_{ML} + b_{ML} + c_{ML} = 1$. Using these boundary conditions we get, $a_{ML} = \frac{n_{DO} - 1 + \frac{n_{DX} - n_{DO}}{l_X}}{-l_X^2 + 2l_X - 1}$, $b_{ML} = \frac{n_{DX} - n_{DO}}{l_X} - 2a_{ML} l_X$ and $c_{ML} = 1 + a_{ML}(2l_X - 1) - \frac{n_{DX} - n_{DO}}{l_X}$. Finally, we express machine learning based normalized minimum number of driver nodes for real-world networks under random attacks as,

$$n_{D,rand,ML} = \begin{cases} n_{DO} + \frac{n_{DX} - n_{DO}}{l_X} l, & l \leq l_X \\ a_{ML} l^2 + b_{ML} l + c_{ML}, & l \geq l_X \end{cases} \tag{4.9}$$

Figure 4.5 compares our ANN-based approximation and Sun's approximation Eq.(4.1) with simulations for two real-world networks. We observe that ANN-based approximation fits better with the simulations. To analyze this comparison, Table 4.8 quantifies the performance using mean absolute and mean relative errors for 10 considered real-world networks. It can be noticed that our ANN-based approximation performs the best in the Colt network with a mean relative error of 0.58 % and the least in the Uninet network with

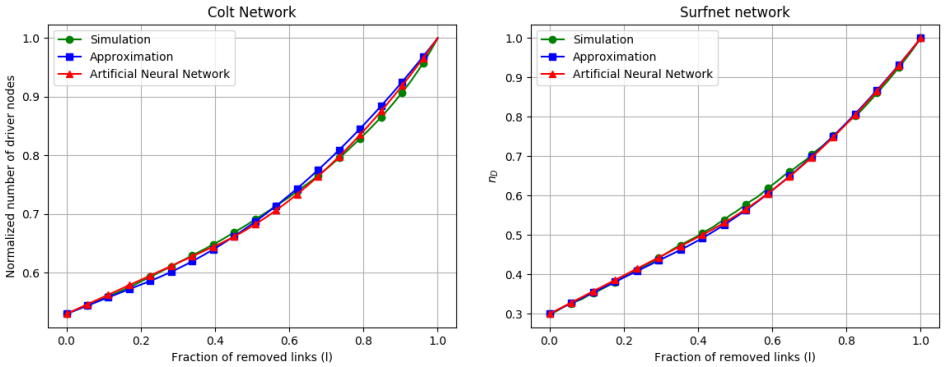


Figure 4.5: Comparison of different methods to get the normalized values of minimum number of driver nodes n_D needed to control the network as a function of the fraction of removed links in real-world networks under random attacks. Simulations are based on 10,000 realizations of attacks.

a mean relative error of 2.75 %. Moreover, the ANN-based model does not always perform better than the analytical approximation. For example, in Ibm and Arpanet19706, the mean relative errors using ANN-based model are larger than the analytical approximation based mean relative errors. This can be explained based on the availability of a limited amount of training dataset for real-world networks. Among the 40 test real-world networks, the machine learning based approximation performs better than the analytical approximation in 28 networks.

Table 4.8: Performance indicators for real-world networks under random attacks.

Network	**Mean Absolute Error**		**Mean Relative Error**	
	Approximation	**ANN**	**Approximation**	**ANN**
Colt	0.0079	0.0043	0.0106	0.0058
Surfnet	0.0072	0.0052	0.0128	0.0090
EliBackbone	0.0256	0.0160	0.0454	0.0274
Garr200912	0.0121	0.0094	0.0156	0.0130
GtsPoland	0.0081	0.0046	0.0127	0.0068
Ibm	0.0072	0.0086	0.012	0.015
Arpanet19706	0.0046	0.0062	0.0073	0.0123
GtsHungary	0.0082	0.0072	0.0098	0.0088
BellCanada	0.0105	0.0071	0.0197	0.0122
Uninet	0.0207	0.0166	0.0338	0.0275

4.5.3. OUT-IN DEGREE-BASED ATTACK

In this section, we will derive an analytical approximation for the normalized minimum number of driver nodes n_D as a function of the fraction of removed links l for out-in degree-based attacks. Out-in degree of a link is defined as the sum of the out-degree

of a source node and the in-degree of a target node. First, we compare different out-in based-attack strategies to select the most efficient one. In the first strategy, we remove links based on the increasing order of out-in degrees, second, if the out-in degrees are the same then links are removed based on the increasing order of out-degrees and finally, in the third strategy, we remove the links based on the decreasing order of out-in degrees. Based on simulations, we found that the first two strategies overlap and are the most efficient ones. So, for the remainder of this section, we will use the first strategy in which we remove links based on the increasing order of out-in degrees. It is to be noted that after removing a link, we re-calculate the out-in degrees in order to determine the next link to be removed.

CASE 1: $l \leq l_C$

Similar to [36], when the fraction of removed links is less than or equal to the fraction of critical links, we assume a linear relationship between the minimum number of driver nodes and the fraction of removed links such that,

$$n_{D,out\_in} = \frac{N_{DO} + lL}{N}. \tag{4.10}$$

CASE 2: $l \geq l_C$

When the fraction of removed links is greater than or equal to the fraction of critical links, we approximate the minimum number of driver nodes using a quadratic equation,

$$f(l) = n_D = gl^2 + hl + i, \tag{4.11}$$

where g, h and i can be derived from the boundary conditions. For the first boundary condition we assume , at $l = l_C$, n_D equals $\frac{N_{DO}+l_C L}{N}$. Second, at $l = 1$, n_D equals one. Third, we assume that the derivative equals zero at $l = 1$. Using these boundary conditions we get, $g = \frac{x-1}{l_C^2 - 2l_C + 1}$, $h = -2g$ and $i = 1 - g - h$ where $x = \frac{N_{DO}+l_C L}{N}$. Finally, for out-in degree-based attacks we can write,

$$n_{D,out\_in} = \begin{cases} \frac{N_{DO}+lL}{N}, & l \leq l_C \\ gl^2 + hl + i, & l \geq l_C \end{cases} \tag{4.12}$$

Figure 4.6 shows the performance of our analytical approximation Eq.(4.12) for Erdős-Rényi and Barabási-Albert networks. We notice that the analytical approximation fits better with the simulations for Barabási-Albert networks. The same is also evident from Table 4.9 in which we show the performance of some synthetic networks. We notice that the mean relative errors are less than 3 % for BA networks and greater than 10 % for ER networks. We also analyze the performance of our approximation in real-world networks. Figure 4.7 shows the performance of our approximation for the Colt and Surfnet networks. It can be observed that the approximation fits fairly well with the simulations. Furthermore, we analyze the performance of 10 considered real-world networks in Table 4.10. We notice that the mean relative errors are less than 10 % in 8 out of 10 real-world networks. Moreover, the approximation performs the best in the GtsHungary network and the least in the Uninet network with mean relative errors of 1.53 % and 13.61 % respectively.

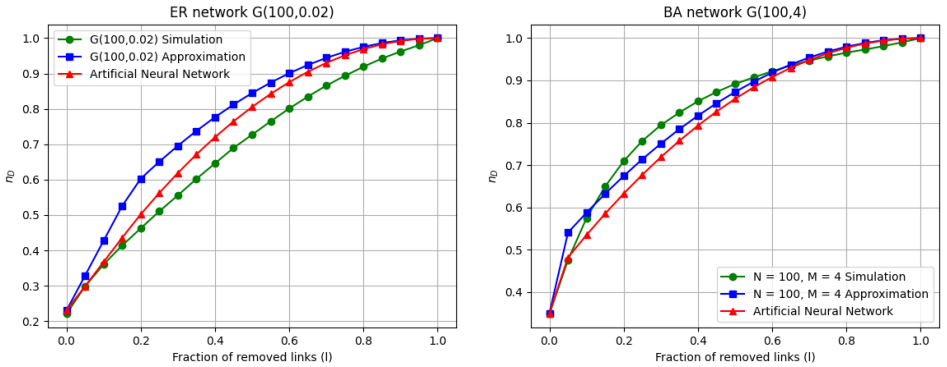


Figure 4.6: Performance comparison of the machine learning based approximation Eq.(4.15) with the analytical approximation Eq.(4.12) to get the normalized values of minimum number of driver nodes n_D needed to control the networks as a function of the fraction of removed links in synthetic networks under out-in degree-based attacks.

Next, we use ANN to further improve the performance of the analytical approximation Eq.(4.12). We will use ANN to predict the difference in the values of the normalized minimum number of driver nodes between the approximation value and the simulation value at l_C. We will then subtract this difference from the approximation value to get a new value n_{DX} that is closer to the simulation. At $l = 0$, the minimum number of driver nodes can be found from Eq.(4.12) and at $l = l_C$, the value is n_{DX}. From these two points, we get,

$$n_{D,out\_in,ML} = n_{DO} + \frac{n_{DX} - n_{DO}}{l_C} l, \tag{4.13}$$

where $n_{D,out\_in,ML}$ gives us the machine learning based normalized minimum number of driver nodes for $l \leq l_C$. For $l \geq l_C$, we assume a quadratic relationship for the normalized minimum number of driver nodes such that,

$$f_{ML}(l) = n_{D,out\_in,ML} = g_{ML} l^2 + h_{ML} l + i_{ML}, \tag{4.14}$$

To get the values of g_{ML}, h_{ML} and i_{ML}, we again use three boundary conditions. n_D equals n_{DX} at $l = l_C$. At $l = 1$, n_D equals one. The derivative $f'_{ML}(1)$ is assumed to be equal to zero at $l = 1$. Using these boundary conditions we get, $g_{ML} = \frac{n_{DX}-1}{l_C^2 - 2l_C + 1}$, $h_{ML} = -2g_{ML}$ and $i_{ML} = 1 - g_{ML} - h_{ML}$. Hence, the machine learning based approximation for the minimum number of driver nodes can be expressed as,

$$n_{D,out\_in,ML} = \begin{cases} n_{DO} + \frac{n_{DX} - n_{DO}}{l_C} l, & l \leq l_C \\ g_{ML} l^2 + h_{ML} l + i_{ML}, & l \geq l_C \end{cases} \tag{4.15}$$

In Figure 4.6, we compare the performance of ANN-based approximation Eq.(4.15) with the analytical approximation Eq.(4.12) and simulations in case of synthetic networks. While we notice that the ANN-based approximation improves the performance in case of

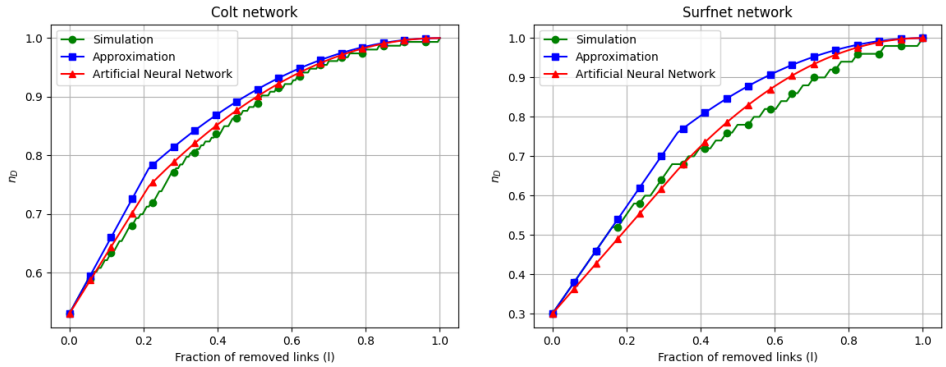


Figure 4.7: Performance comparison of the machine learning based approximation Eq.(4.15) with the analytical approximation Eq.(4.12) to get the normalized values of minimum number of driver nodes n_D needed to control the networks as a function of the fraction of removed links in real-world networks under out-in degree-based attacks.

Erdős-Rényi networks, it does not always improve the performance of Barabási-Albert networks as the original analytical approximation Eq.(4.12) already fits well. In terms of mean absolute errors and mean relative errors, Table 4.9 compares the performance of both approximations. We observe that for $ER(100, 0.02)$ network, the mean relative error decreases from 13.80 % to 6.80 % with ANN-based approximation. We notice similar improvements for other ER networks as shown in Table 4.9. For BA networks, we do not always see an improvement which is also evident in $BA(100, 4)$ network in which the mean relative error increase from 2.76 % to 4.0 % as the original approximation already fits well with the simulations.

Table 4.9: Performance indicators for synthetic networks under out-in degree-based attacks.

Network	**Mean Absolute Error**		**Mean Relative Error**	
	Approximation	**ANN**	**Approximation**	**ANN**
ER(50, 0.048)	0.0959	0.0568	0.1786	0.0924
ER(100, 0.02)	0.0828	0.0463	0.1380	0.0680
BA(50, 4)	0.0193	0.0189	0.0278	0.0266
BA(100, 4)	0.0201	0.0308	0.0276	0.0400

Figure 4.7 compares the performance of ANN based approximation Eq.(4.15) with the analytical approximation Eq.(4.12) and simulations for real-world networks. The performance of all the considered 10 real-world networks is shown in Table 4.10. We notice that the ANN-based approximation Eq.(4.15) performs better than the analytical approximation Eq.(4.12) in 7 out of 10 considered real-world networks.

All the simulations are performed on a PC with the following specifications - 8 GB RAM and Intel Core i5 processor with 2 cores. With these specifications, for a dataset consisting of 232 networks, it costs less than 0.6 seconds to train the linear regression and random forest models whereas, it costs approx. 2-3 seconds to train the artificial

Table 4.10: Performance indicators for real-world networks under out-in degree-based attacks.

Network	Mean Absolute Error		Mean Relative Error	
	Approximation	ANN	Approximation	ANN
Colt	0.0210	0.0102	0.0267	0.0129
Surfnet	0.0469	0.0280	0.0609	0.0395
EliBackbone	0.0846	0.0373	0.1188	0.0539
Garr200912	0.0229	0.0213	0.0262	0.0242
GtsPoland	0.0256	0.0357	0.0309	0.0447
Ibm	0.0665	0.0682	0.0922	0.0951
Arpanet19706	0.0416	0.0340	0.0522	0.0519
GtsHungary	0.0140	0.0135	0.0153	0.0148
BellCanada	0.0546	0.0657	0.0742	0.0917
Uninet	0.0956	0.0586	0.1361	0.0829

neural network model. Once the models have been trained, after getting the average values of 10,000 simulations as inputs to the models, it costs less than 0.5 seconds to get the predictions.

4.6. CONCLUSION

In this chapter, we used various machine learning methods to quantify the minimum number of driver nodes N_D as a function of the fraction of removed links l. We studied the robustness of network controllability using machine learning based approximations on both synthetic and real-world networks under random and targeted attacks. We also derived an analytical approximation for out-in degree-based attacks. In case of targeted critical link attack, we first compared the performance of ANN, RF and LR models and conclude that the LR model performs the least due to the nonlinear relationship between the input features and the output difference. ANN model performed slightly better than the RF model. Our machine learning based approximation outperformed the analytical approximation in both synthetic and real-world networks. However, for real-world networks, our approximation performed better than the original analytical approximation in 75 % of the networks. For random attacks our approximation performed better than the analytical approximation in 70 % of the real-world networks. We also compared our machine learning based approximation with Liu's approximation and the approximation provided in Chapter 2 for ER networks under random attacks. Liu's approximation performed better than both machine learning based approximation and the approximation in Chapter 2. We also derived an analytical approximation for out-in degree-based attacks. For synthetic networks, the approximation performed better in case of BA networks than ER networks. Furthermore, in 8 out of 10 considered real-world networks, the mean relative errors are less than 10 %. We further improved our analytical approximation for out-in degree-based attacks using ANN and the mean relative errors reduced to less than 6 % in 7 out of 10 real-world networks.

5

THE REACHABILITY-BASED ROBUSTNESS OF NETWORK CONTROLLABILITY

In this chapter, we propose closed-form analytic approximations for the number of controllable nodes in sparse communication networks, considering link-based random attack, targeted attack, as well as random attack under the protection of critical links. We first compare our approximations with simulation results on communication networks. Results show that our approximations perform well for all three attack strategies as long as the fraction of removed links is small. Only when the fraction of removed links is large, our approximation for targeted attacks does not fit well with simulation results. Finally, we validate our approximations using 200 communication networks and some synthetic networks. Results show that our approximations perform well in most cases.

5.1. INTRODUCTION

In recent years, the analysis of network controllability from a graph theoretic point of view has become an active area of research. Through the control of external inputs [39], a controllable system can be driven from any arbitrary state to any desired state in finite time . For example, a communication network can be controlled externally through input signals such as commands from control units connected to some of the work stations [88].

Most work regarding the robustness of controllability has focused on the number of controls required to maintain network controllability after link or node failures. Lou *et al.* [89] proposed a complex network model called q-snapback network which has the strongest robustness of controllability due to its advantageous inherent structure with many chain and loop motifs, when compared with the multiplex congruence network and the generic scale-free network. Pu *et al.* [54] found that the degree-based node

This chapter is based on the published paper [87].

attack is more efficient than a random failure for degrading the controllability in random and scale-free networks. Nie *et al.* [55] found that the controllability of Erdős-Rényi random graphs with a moderate average degree is not very robust, whereas a scale-free network with moderate power-law exponent shows a stronger ability to maintain its controllability, when these networks are under intentional link attack. Thomas *et al.* [56] identified that the potency of a degree-based attack is directly related to the betweenness centrality of the edges being removed. Chen *et al* [90] evaluated the effect of the number of control inputs on the controllability for random networks and scale-free networks in the process of cascading failure. Lou *et al* [91] proposed a framework of hierarchical attack by means of link- or node-removal attacks and suggest to protect the critical links and nodes to maintain network controllability. Xiao *et al.* [92] proposed a method that modifies any given network with strict structural perturbation to make the network homogenous and effectively enhance its robustness against malicious attacks. Zhang *et al.* [93] optimized the robustness of interdependent network controllability by redundant design including node backup and edge backup. In Chapter 2, we proposed closed-form analytic approximations for the number of controls that are needed to maintain network controllability, where links are removed according to both random and targeted attacks [36].

The above work regarding the robustness of controllability assumes that the network operator has the capability to add additional controls at any location in the network in order to maintain the current network controllable after attacks or failures. In other words, the basic assumption of previous work mentioned above is that network operators have sufficient budget and quantity of resources that can be deployed in response to an attack or failure. However, a more realistic assumption is that network operators have a fixed budget and a limited quantity of resources. Moreover, the increase in additional controls is only a proxy for the most relevant information - how much of the network is still controllable (reachable) after an attack or failure. Parekh *et al.* [94] proposed the number of controllable nodes as a new metric to quantify the robustness of controllability under network perturbations. Thomas *et al.* [56] analyzed the changes in the controllability of synthetic networks from the perspective of reachability and found that scale-free networks evidence higher robustness to random failures than Erdős-Rényi networks. In this chapter, we analyse and measure the robustness of network controllability in terms of reachability. In particular, we determine the maximum number of nodes that are still controllable when the number of driver nodes remains the same during the failure or attack process. Here, the driver nodes are the nodes into which the external control signals are directly injected.

This chapter is organized as follows. In Section 5.2, we introduce some basic concepts and definitions in reachability-based network controllability. In Section 5.3, we analyse the role of critical links in network controllability. In Section 5.4, we compare the robustness of controllability for three cases: random attack, random attack under protection and targeted attack. In Section 5.5, we propose analytic approximations for the number of controllable nodes N_c in these three cases and measure the accuracy of our approximations. Section 5.6 concludes the chapter.

5.2. REACHABILITY-BASED ROBUSTNESS OF CONTROLLABILITY

5.2.1. REACHABILITY-BASED CONTROLLABILITY

So far, most of the existing studies on the robustness of controllability have measured the increase in the minimum number N_d of driver nodes required as a proxy for the reduction in controllability due to a failure. This indirect approach of measuring robustness is referred to as control-based robustness. The robustness of network controllability from the perspective of reachability is also considered by a few authors, see [56] and [94]. Besides, the control-based robustness analysis of network controllability assumes that the network operator has the capability to attach any amount of additional control signals to the nodes in the network. However, network operators normally have limited budget and resources in real life, which constrains the ability to deploy external controls. Based on these considerations, we focus on the reachability-based robustness of controllability, which determines the maximum number N_c of nodes that are still under control when failure or attack occurs, during which the number N_{d0} of driver nodes remains the same [94]. For the reachability-based controllability, there are two cases, namely free control and fixed control [56]. In the free control case, only the number N_{d0} of driver nodes remains the same, but the set of driver nodes can vary. In the fixed control case, both the number and the set of driver nodes are fixed during attacks or failures. In this chapter, we only consider the free control case and delegate the fixed control to future research. For convenience, we use the term reachability to represent reachability-based controllability.

5.2.2. R-VALUE AND CHALLENGES

We inherit the framework and some definitions proposed for network robustness [52,75] to investigate the robustness of reachability. The robustness of a given network determined by a service and an underlying topology is quantified by a robustness value, referred to as the R-value [75]. The R-value is normalized to the interval $[0,1]$. Thus, $R = 1$ reflects complete functionality in an network without failures, and $R = 0$ corresponds to the complete absence of functionality in a severely damaged network. The R-value can be a metric, which is related to network topology and service, such as the size of the giant component [47], the effective graph resistance [20] and network efficiency [95]. In this chapter, we use the normalized maximum number of controllable nodes $n_c = N_c/N$ as the R-value. The number N_c of controllable nodes satisfies $N_{d0} \le N_c \le N$, thus $N_{d0}/N \le n_c \le 1$.

An elementary challenge is an event that changes the network and thus changes the R-value. We assume that a sequence of changes does not coincide in time. In this chapter, we confine an elementary challenge to a link removal in a failure process. A perturbation is a series of m elementary changes, characterized by a sequence of m corresponding R-values $\{R[k]\}_{0<k\le 1}$, where $k = m/L$ is the fraction of removed links, $m \in \{1,\dots,L\}$ is the number of removed links and L is the number of links in the network. In this chapter, we choose the maximum number n_c of controllable nodes as the R-value and observe the impact of link removal on n_c. As shown in Figure 5.1, the maximum number n_c of controllable nodes has a decreasing trend as links are removed one by one.

5.2.3. ROBUSTNESS ENVELOPES

As discussed in the previous part, any realization of failure processes can be expressed as a sequence of R-values denoted $\{R[k]\}_{0<k\leq 1}$ where k is the fraction of removed links and $k \in \{1/L, 2/L, \ldots, 1\}$. Assuming that the nature of the failures is unknown and they occur independently, $R[k]$ is a random variable and can be described by its probability density function (pdf). The pdf of this $R[k]$ is computed using all subsets of $\lfloor kL \rfloor$ links in all possible perturbations. The envelope for a network G is constructed using all $R[k]$ for $k \in \{1/L, 2/L, \ldots, 1\}$, where boundaries are given by the extreme R-values

$$R_{\min}[k] \in \{\min(R[1/L]), \min(R[2/L]) \ldots, \min(R[1])\}, \tag{5.1}$$

$$R_{\max}[k] \in \{\max(R[1/L]), \max(R[2/L]) \ldots, \max(R[1])\}, \tag{5.2}$$

which gives the worst- and best-case of robustness metrics for a network after a given number of challenges [52]. Besides, the expected R-value resulting from $\lfloor kL \rfloor$ perturbations

$$R_{avg}[k] \in \{E(R[1/L]), E(R[2/L]) \ldots, E(R[1])\}. \tag{5.3}$$

Since $R[k]$ defines a probability density function, we are interested in the percentiles of $R[k]$

$$R_{\theta\%}[k] \in \{R_{\theta\%}[1/L], R_{\theta\%}[2/L] \ldots, R_{\theta\%}[1]\} \tag{5.4}$$

where $R_{\theta\%}[k]$ are the points at which the cumulative distribution of $R[k]$ crosses $\frac{\theta}{100}$, namely if $R_{\theta\%}[k] = t$, then $\Pr[R[k] \leq t] = \frac{\theta}{100}$. We refer to $R_{\theta\%}[k]$ as a θ-percentile and define $R_{0\%}[k] = R_{min}[k]$, $R_{100\%}[k] = R_{max}[k]$.

We apply the envelope to present the influence of the failure process on a network [52, 75]. The envelope profiles the pdf of the random variables of the R-value, which is the probability of a random variable to fall within a particular region. The area of the envelope can be regarded as the variation of the robustness impact of a certain series of challenges, which quantifies the uncertainty or the amount of risk due to perturbations. The effectiveness of attack strategies can also be measured by comparing with the worst-, best- and average performance provided by robustness envelopes.

5.3. ANALYSIS OF CRITICAL LINKS

Liu *et al.* [39] proved that the minimum number N_d of driver nodes needed for structural controllability, where the external signals are injected to control the directed network, can be obtained through the "maximum matching" of the network. Define the source node of a directed link as the node from which the link originates and the target node as the node where the link terminates. A maximum matching of a directed network is a maximum set of links that do not share source or target nodes [27], which is illustrated in Figure 1.1(a). Such links are coined "matching links". Target nodes of matching links are matched nodes and the other nodes are unmatched nodes. In order to find the maximum number of matching links, so as to determine the minimum number N_d of driver nodes, a directed network G with N nodes and L links can be converted into a bipartite graph $B_{N,N}$ with $2N$ nodes and L links, as shown in Figure 1.1(b). A maximum matching in a bipartite graph can be obtained efficiently by the Hopcroft-Karp algorithm [28]. The unmatched nodes in a maximum matching constitute a minimum set of driver nodes.

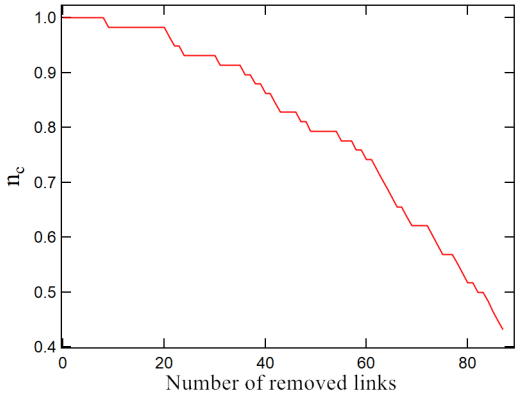


Figure 5.1: The impact of link removal on the normalized maximum number n_c of controllable nodes in a communication network DFN (German optical backbone X-WiN network) with $N = 58$ and $L = 87$.

5.3.1. THE ROLE OF CRITICAL LINKS IN MAXIMUM MATCHING

Links in a network can be classified into three categories: critical, redundant, and ordinary [39]. A link is critical if its removal increases the minimum number of driver nodes N_d by 1 to remain in full control of the system. A link is redundant if it never belongs to a maximum matching. A link is ordinary if it is neither critical nor redundant. In Figure 1.1(a), link a, b, c and d (highlighted in red) are critical links, the removal of any one of them will increase the number of driver nodes by 1, while link e is redundant. The influence of the removal of critical links can be explained by the maximum matching. As shown in Figure 1.1(b), all the critical links a, b, c, d belong to the maximum matching of size 4. If any one of them is removed, there is no alternative link to take its place in the maximum matching. Thus, a new unmatched node will appear and the number of driver nodes will increase by 1. Besides, critical links are conditional and should be updated during attacks. For example, link c is no longer a critical link in the resulting network after link b is removed.

In our previous work [36], we proposed closed-form analytic approximations for the minimum number N_d of driver nodes needed to fully control networks, where links are removed according to both random and targeted attacks.

5.3.2. THE ROLE OF CRITICAL LINKS IN THE STRUCTURE OF CONTROL

In the research concerning reachability-based controllability, Parekh *et al.* [94] found the control structure which consists of a backbone of directed paths, called stems, each driven by an independent control . These paths can then control cycles that are inherently self-regulatory. However, ultimately these stems dictate the need for controls: There must be one control node for each stem in the system in order to guarantee that all nodes in the network are controllable (reachable). In this chapter, we use the algorithm proposed in [56] to find the control structure in the network:

1. Determine the number M of control nodes by the maximum matching introduced in Section III.A.

2. Preprocess the network by adding the fixed number of control nodes and then placing links from each control node to every state node, after which there are N nodes and E links in the network. Then, for all $i, j = 1, ..., N$ and $k = 1, ..., M$:
   (a) Split the nodes into a pair of positive and negative nodes $x_i \Rightarrow x_i^+, x_i^-$, $u_k \Rightarrow u_k^+, u_k^-$.
   (b) Add unit-weight links (x_i^+, x_j^-) and (u_k^+, x_j^-) if the link (x_i, x_j) and (u_k, x_j) exist in the network, respectively.
   (c) Add zero-weight links (x_i^+, x_i^-) and (u_k^+, u_k^-).
   (d) Add zero-weight links (x_i^+, u_k^-).
   (e) Add an extra weight $W \geq E$ to all links.
3. Run the weighted maximum matching algorithm on the bipartite graph generated by step 2 to find the set of matched links. Thus, the total weights of all matched links are maximized. The control structure is then formed by mapping the matched links in the original network. In the implementation, the Fibonacci Heap algorithm [96] is used in the weighted maximum matching, of which the computational complexity of finding the set of matched links is $O(NL + mN^2 + 4n^2 log2N)$.

5

Finally, the number of controllable nodes equals the number of matched nodes in the control structure. Although the concept of critical links was first proposed in control-based controllability analysis which focuses on the number N_d of driver nodes, critical links also play an important role in reachability-based controllability. We found that the removal of a critical link usually decreases the number of controllable nodes by 1 in most cases when the network is sparse. The influence of the removal of critical links can also be explained by the control structure in the network. As shown in Figure 1.1(a), the control structure only consists of one directed path $1 \rightarrow 2 \rightarrow 3 \rightarrow 4 \rightarrow 5$. This path is also formed by four critical links a, b, c, d, which are defined in control-based controllability. Removing each of these critical links will break the path and decrease the number of controllable nodes by 1. However, in some cases, removing a critical link can increase the number of controllable nodes by more than 1. For example, the network constructed by a single path $1 \rightarrow 2 \rightarrow 3 \rightarrow 4 \rightarrow 5 \rightarrow 6$, the number N_c of controllable nodes decreases by 3 after removing the critical link between node 3 and 4. In this chapter, we use the concept of critical links to derive analytical approximations for the decrease in the number N_c of controllable nodes upon link removal.

5.4. NUMBER OF CONTROLLABLE NODES UNDER ATTACKS

In this section, we analyze the normalized number of controllable nodes for three different attack scenarios: (a) random attack, (b) targeted attack and (c) random attack under protection. In a random attack, links are removed from the network uniformly at random. In a targeted attack, we assume that the attacker knows the location of critical links and removes critical links uniformly at random. After all critical links are removed, the attacker randomly removes other links. In a random attack under protection, the network operator takes measures to protect the critical links such that only non-critical links are removed randomly.

We compare the normalized number n_c of controllable nodes for these three attacks in 10 sparse communication networks [80] [97]. Table 5.1 presents the properties of the 10 communication networks: the number N of nodes, the number L of directed links, the initial minimum number N_{d0} of driver nodes and the number L_c of critical links. For a directed network, the average degree $E[D] = 2L/N$, which also equals the sum of the mean out- and in-degree per node. The first 8 networks are small. The other two networks are relatively large, which are an order larger than the average size of the other 8 communication networks. Besides, the last network has a higher average degree, which is more than twice that of the other networks. The number L_c of critical links can be determined by applying the Hopcraft-Karp algorithm L times, by considering all L networks that are obtained by removing exactly one link from the original network. As expected, Figure 5.2 shows that random attack under protection performs the best among the three attack scenarios in maintaining the reachability of the networks. Moreover, we also conclude from Figure 5.2:

1) In the case of random attack under protection, the slope of the decrease in n_c is almost 0 in the beginning for all networks. This emphasizes the importance of protecting critical links.

2) The targeted attack is the most harmful: when the fraction of removed links is smaller than the fraction of critical links, the decrease in n_c is almost linear in the fraction of removed links. When all the critical links are removed, the slope of the decrease in n_c is almost 0 in all 8 networks. Considering that the set of critical links is determined from the initial network, this indicates that the set of critical links of a network does not significantly change during the attack process when the fraction of removed links is small.

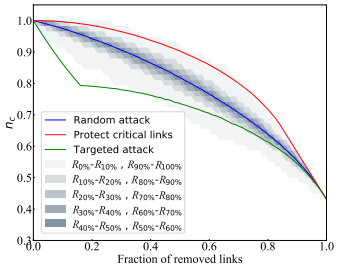
3) The performance of random attack is between targeted attack and random attack under protection. After all links are removed, the normalized number of controllable nodes equals N_{d0}/N.

4) Critical links have a significant impact on the number N_c of controllable nodes upon link removal, which plays a key role to derive analytical approximations for the number N_c of controllable nodes.

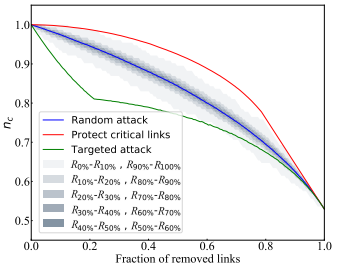
We also use robustness envelopes to evaluate the effectiveness of the three attack strategies. As shown in Figure 5.2, the curves for random attack under protection are quite close to the boundaries represented by the *90*-percentile $R_{90\%}[k]$ among all networks, which means that random attack under protection outperforms 90% realizations of random attack. The curves for targeted attack are much lower than the lower bound of envelopes especially when the fraction of removed links is small in all networks, which again underlines the harm of targeted attack.

5.5. APPROXIMATIONS FOR THE NUMBER OF CONTROLLABLE NODES

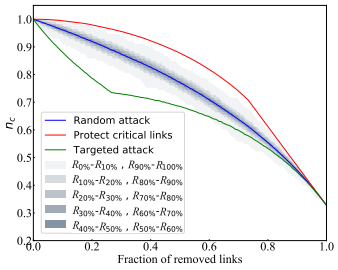
In the previous section, we compared the number of controllable nodes for the three attack scenarios by using a large amount of simulations. In this section, we deduce analytical approximations to quantify the robustness of reachability, expressed in terms of the normalized number n_c of controllable nodes, for the three attack scenarios. Then, we evaluate the accuracy of the analytical approximations in 8 small networks, 2 large networks



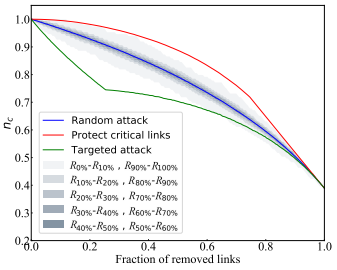
(a) DFN



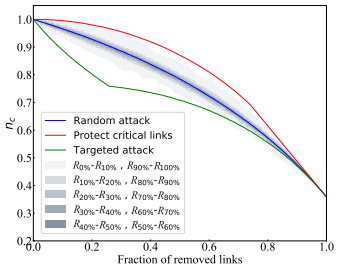
(b) Colt



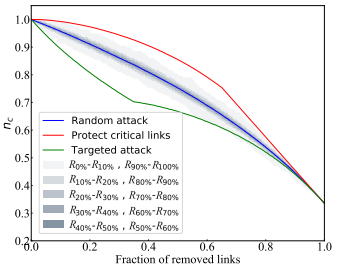
(c) Deltacom



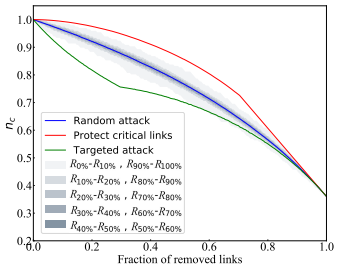
(d) GtsCe



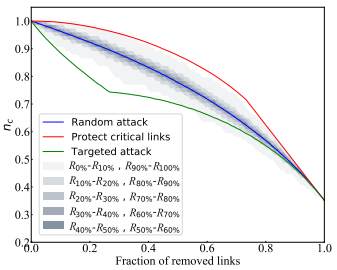
(e) TataNld



(f) UsCarrier



(g) Cogentco



(h) Uninett2010

Figure 5.2: Performance of the normalized number n_c of controllable nodes as a function of the fraction of removed links l for three attack scenarios. The results for each fraction l is based on 1000 simulations. Each envelope of the challenges for the normalized number n_c of controllable nodes is based on 10^4 realizations. In order to compare the scenario for random attack under protection with the other two scenarios in the same sub-figure, we remove critical links uniformly at random after all the other links are removed.

Table 5.1: Properties of the 10 considered communication networks

Networks	N	L	$E[D]$	N_{d0}	L_c
DFN	58	87	3.0	25	14
Colt	153	177	2.3	81	38
Deltacom	113	161	2.8	37	43
GtsCe	149	193	2.6	58	49
TataNld	145	186	2.6	52	48
UsCarrier	158	189	2.4	53	66
Cogentco	197	243	2.5	71	72
Uninett2010	74	101	2.7	26	27
Kdl	754	895	2.4	272	287
Web [97]	643	2280	7.0	324	108

as well as more sparse communication networks. Lastly, we also use synthetic networks to measure the performance of our analytical approximations. Our approximations will be based upon the concept of critical links introduced in [39].

5.5.1. NUMBER OF CONTROLLABLE NODES UNDER RANDOM ATTACKS

1) **The fraction l of removed links is less than the fraction l_c of critical links**
Given a network with N nodes and L links, the initial number N_c of controllable nodes equals N. The number L_c of critical links can be determined by the method we introduced in Section 5.4.

As discussed in Section 5.3.2, the number N_c of controllable nodes decreases by at least one when a critical link is removed. However, we found that the number N_c of controllable nodes only decreases by one for every critical link that is removed in each of the 10 sparse communication networks in Table 5.1. Thus, we heuristically assume that after removing a critical link, the number N_c of controllable nodes decreases by one. If we denote the number of removed links by m, then the fraction of removed links $l = \frac{m}{L}$, while the fraction of critical links l_c satisfies $l_c = \frac{L_c}{L}$. We consider the case $l \leq l_c$, where m links are removed uniformly at random under the condition that the number of removed links obeys $m \leq L_c$. Now assume that of these m links i links are critical ($i \leq m$) and, hence, $m-i$ links are non-critical. We assume that the set of critical links is nearly unchanged when the fraction of removed links is small. Invoking the fact that after removing a critical link, the number N_c of controllable nodes decreases by one, thus, when i critical links are iteratively removed one by one, the number N_c of controllable nodes decreases by one in each iteration. For the $m-i$ removed non-critical links, the number N_c of controllable nodes remains the same based on our assumption that the set of critical links is unchanged when the fraction of removed links is small. Since there are $\binom{L_c}{i}$ possible ways to choose i critical links from L_c critical links and there are $\binom{L-L_c}{m-i}$ possible ways to choose $m-i$ non-critical links from $L-L_c$ non-critical links, the contribution to the decrease in N_c is $i\binom{L_c}{i}\binom{L-L_c}{m-i}$. The average decrease N_c^* of the number N_c of controllable nodes after randomly removing m links, is the sum of this expression for all $i = 1, 2, \ldots, m$

divided by $\binom{L}{M}$.

$$N_c^* = \frac{\sum_{i=1}^{m} i\binom{L_c}{i}\binom{L-L_c}{m-i}}{\binom{L}{m}} \tag{5.5}$$

Using $i\binom{L_c}{i} = L_c\binom{L_c-1}{i-1}$ and Vandermonde's formula $\sum_{j=0}^{k}\binom{a}{j}\binom{b}{k-j} = \binom{a+b}{k}$ for any number a and b, we obtain $L_c\sum_{i=0}^{m-1}\binom{L_c-1}{i}\binom{L-L_c}{m-1-i} = L_c\binom{L-1}{m-1}$. Finally, dividing this expression by $\binom{L}{m}$, leads to the average decrease of controllable nodes

$$N_c^* = lL_c \tag{5.6}$$

When the fraction of removed links is less than, or equal to l_c, we obtain

$$N_c = N - lL_c \tag{5.7}$$

We then normalize the number N_c of controllable nodes to the fraction $\frac{N_c}{N}$ of the minimum number of controllable nodes and denote the obtained approximation as $n_{c,rand}$,

$$n_{c,rand} = \frac{N - lL_c}{N} \tag{5.8}$$

2) **The fraction l of removed links is larger than the fraction l_c of critical links**
Considering that in most cases l_c is quite small, we also estimate the normalized maximum number n_c of controllable nodes when the fraction l of removed links is larger than the fraction l_c of critical links. For $l \geq l_c$, we heuristically propose a simple closed-form approximation for $n_{c,rand}$:

$$n_{c,rand} = al^2 + bl + c \tag{5.9}$$

where the parameters a, b and c will be determined by boundary conditions. For the first two boundary conditions we assume that, for $l = l_c$, Eq.(5.9) has the same value and the same derivative as Eq.(5.8). This leads to the equations $N - l_cL_c = N(al_c^2 + bl_c + c)$ and $-L_c = N(2al_c + b)$, respectively. Finally, if we remove all links, i.e. $l = 1$, only N_{d0} nodes can be controlled. This gives the boundary condition $N_{d0}/N = a + b + c$. Solving for a, b and c and combining with the approximation Eq.(5.8), we obtain the following approximation for $n_{c,rand}$ for all values of l:

$$n_{c,rand} = \begin{cases} \frac{N-lL_c}{N} & l \leq l_c \\ al^2 + bl + c & l \geq l_c \end{cases} \tag{5.10}$$

with, $a = -\frac{N-N_{d0}-L_c}{N(l_c-1)^2}$, $b = -L_c/N - 2al_c$, and $c = (N_{d0} + L_c)/N + a(2l_c - 1)$. Eq.(5.10) respresents a closed-form approximation for n_c, which only depends on N, L, N_{d0} and L_C. The computational complexity of the approximation is $O(\sqrt{N}L^2)$, which is needed for the computation of L_c.

We compare the approximation Eq.(5.10) with simulation results for the 8 relatively small communication networks. Since the simulation settings for large networks are

slightly different, we will evaluate the performance of our approximations for the large networks in Section 5.5.4. Figure 5.3 illustrates that the approximation both under- and overestimates the value of n_c. For moderate values of the fraction of removed links, the approximation exhibits a very good fit for the communication networks. For some networks, such as Deltacom, GtsCe, TataNld and Uninett2010, our approximation Eq.(5.10) fits well with the simulation results regardless of the fraction of removed links.

The performance of our approximations are also measured by three performance indicators:

1) r^* denotes the absolute value of the relative error at $l = 0.2$. We choose the value 0.2 reflecting a relatively large fraction in terms of link-based failures or attacks.

2) l^* represents the smallest value of l, where the relative error between the approximation and the simulated mean exceeds 5%.

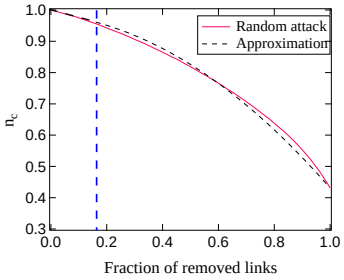
3) γ denotes the fraction of the interval $[0, l_c]$ for which the absolute value of the relative error between the approximation and the mean simulated value does not exceed 5%. The value of γ is computed by K different values of the fraction of removed links, i.e., $v_1, v_2, ..., v_K$, are evenly determined in the interval $[0, l_c]$. Let $n_c^*(v_i)$ and $n_c(v_i)$ denote the mean simulated n_c and the approximation (5.8) at the fraction of removed links $l = v_i$, respectively. Thus, in terms of the indicator function 1_x that equals 1 if the condition x is true, otherwise it is zero,

$$\gamma = \frac{\sum_{i=1}^{K} \mathbf{1}_{\left|\frac{n_c^*(v_i)-n_c(v_i)}{n_c^*(v_i)}\right| \leq 5\%}}{K}.$$

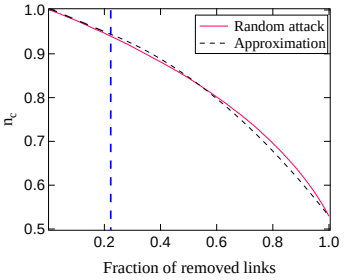
Table 5.2 gives the three performance indicators for Eq.(5.10). As shown in the table, when the fraction of removed links is less than 0.2, the absolute relative error between Eq.(5.10) and the simulated mean is less than 5% for all 8 networks. For most networks, such as Deltacom, GtsCe, TataNld, UsCarrier, Cogentco and Uninett2010, Eq.(5.10) still fits well with simulation results regardless of the fraction of removed links. For the worst performing networks, DFN and Colt, 87% and 81% of the links can be removed before the absolute relative error exceeds 5%, respectively. When the fraction of removed links is less than the fraction l_c of critical links, the absolute value of the relative error between the approximation and the mean simulated value is always less than 5%. Thus, γ equals 100% for all networks.

Table 5.2: Performance indicators for the approximation $n_{c,rand}$ for the 8 communication networks

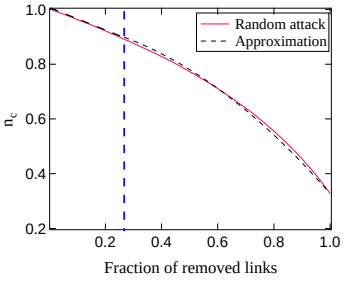
Networks	r^*	l^*	γ
DFN	0.78%	0.87	100%
Colt	1.23%	0.81	100%
Deltacom	1.08%	1.00	100%
GtsCe	1.20%	1.00	100%
TataNld	0.45%	1.00	100%
UsCarrier	1.17%	1.00	100%
Cogentco	0.98%	1.00	100%
Uninett2010	1.24%	1.00	100%



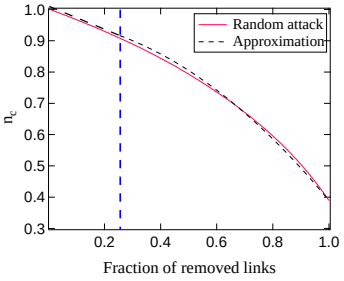
(a) DFN



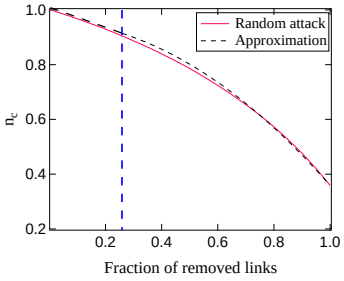
(b) Colt



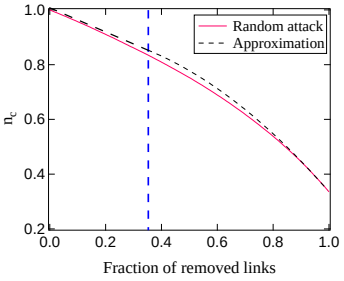
(c) Deltacom



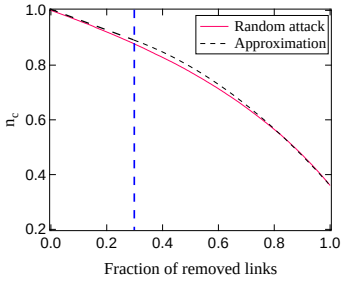
(d) GtsCe



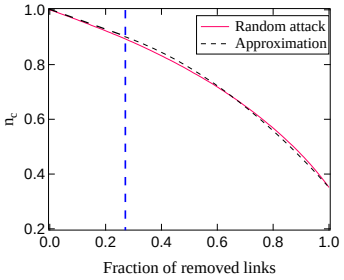
(e) TataNld



(f) UsCarrier



(g) Cogentco



(h) Uninett2010

Figure 5.3: The normalized maximum number of controllable nodes n_c as a function of the fraction of removed links l in communication networks under random attacks. The results for each fraction l is based on 1000 simulations. In each sub-figure, the vertical dashed line marks the position where $l = l_c$.

5.5.2. NUMBER OF DRIVER NODES UNDER TARGETED ATTACKS

1) **The fraction l of removed links is smaller than the fraction l_c of critical links**
We assume that, as long as the number of removed links $m \leq L_c$, the removal of each link decreases the number N_c of controllable nodes by one. Consequently, when the number of removed links is smaller than L_c (the fraction l of removed links is smaller than l_c), the approximation for the minimum number N_c of driver nodes decreases linearly with the fraction l of removed links. When the number of removed links equals the number L_c of critical links, the minimum number N_c of driver nodes equals $N - L_c$. Thus, when the fraction l of removed links is no more than the fraction l_c of critical links, we obtain the following approximation for n_c:

$$n_{c,crit} = \frac{N - lL}{N} \tag{5.11}$$

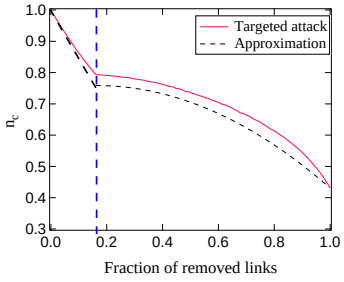
2) **The fraction l of removed links is larger than the fraction l_c of critical links**
We now construct an approximation when the number of removed links is larger than L_c (the fraction l of removed links is larger than l_c), in a similar way as in the previous section. Again assuming that for $l \geq l_c$ it holds that n_c is quadratic in l, we obtain $n_{c,crit} = dl^2 + el + f$. Boundary conditions are now obtained from the assumptions that the parabola passes through $(1, N_{d0}/N)$ and $(l_c, (N - L_c)/N)$ and has a zero derivative at the latter point. This leads to the following approximation for n_c for all values of l:

$$n_{c,crit} = \begin{cases} \frac{N-lL}{N} & l \leq l_c \\ dl^2 + el + f & l \geq l_c \end{cases} \tag{5.12}$$

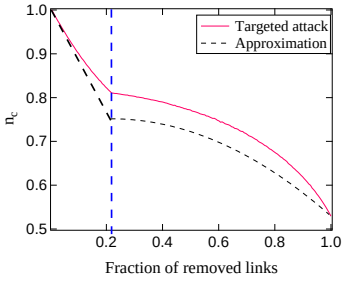
with, $d = -\frac{N - N_{d0} - l_c L}{N(l_c - 1)^2}$, $e = -2dl_c$, and $f = N_{d0}/N + d(2l_c - 1)$.

In Figure 5.4, we compare our approximation Eq.(5.12) with simulation results. Simulation results show that the difference in the curve trend at $l = l_c$, is due to the fact that until $l = l_c$ only critical links are targeted causing a faster descent in the number of controllable nodes. We observe that the approximation Eq.(5.12) fits well with simulation results when the fraction of removed links is sufficiently small in these communication networks. In some networks, such as DFN and UsCarrier, Eq.(5.12) is close to simulation results even when the fraction of removed links is relatively large. When the fraction of removed links is getting larger, the difference between our approximation Eq.(5.12) and simulation results is relatively large. However, approximation Eq.(5.12) always seems to overestimate the impact of targeted attack on the normalized maximum number n_c of controllable nodes, hence, approximation Eq.(5.12) can be considered a worst-case approximation.

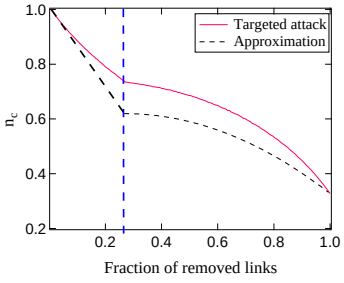
Comparing with the targeted attack, we quantify the performance of the approximation Eq.(5.12) in Table 5.3. For DFN and Colt, Eq.(5.12) is a very good approximation when the fraction of removed links is less than l_c. Eq.(5.12) performs the best for DFN, 23% of the links can be removed before the absolute relative error exceeds 5%. Eq.(5.12) does not perform well for TataNld.



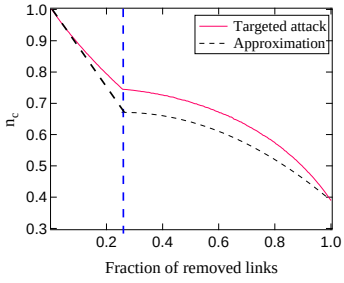
(a) DFN



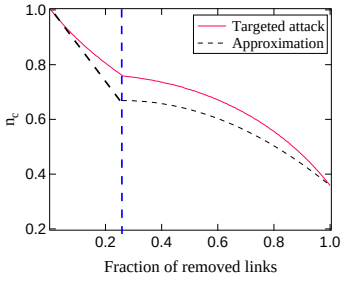
(b) Colt



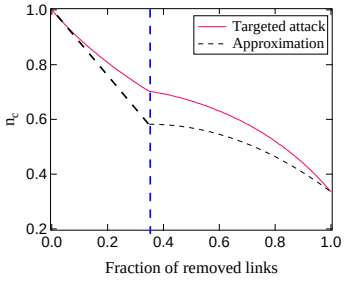
(c) Deltacom



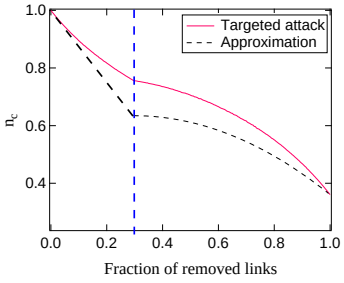
(d) GtsCe



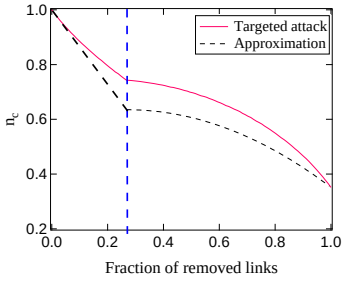
(e) TataNld



(f) UsCarrier



(g) Cogentco



(h) Uninett2010

Figure 5.4: The normalized maximum number of controllable nodes n_c as a function of the fraction of removed links l in communication networks under targeted attacks. The results for each fraction l is based on 1000 simulations. In each sub-figure, the vertical dashed line marks the position where $l = l_c$.

Table 5.3: Performance indicators for the approximation $n_{c,crit}$ for the 8 communication networks

Networks	r^*	l^*	γ
DFN	4.53%	0.23	100%
Colt	4.88%	0.21	97.81%
Deltacom	7.52%	0.18	67.42%
GtsCe	5.06%	0.19	74.80%
TataNld	10.11%	0.12	46.51%
UsCarrier	3.26%	0.22	63.04%
Cogentco	8.75%	0.18	60.81%
Uninett2010	8.86%	0.19	71.08%

5.5.3. NUMBER OF DRIVER NODES UNDER RANDOM ATTACKS WITH PROTECTION

For this scenario, we assume that a fraction of links l_c is protected, then we can only attack a fraction $1-l_c$ of the links. We now construct an approximation for the number $N_{c,prot}$ of controllable nodes when the attack is random under protection. We heuristically assume that the fraction $n_{c,prot}$ of controllable nodes is quadratic in l, we obtain $n_{c,prot} = pl^2 + ql + r$. Boundary conditions are now obtained from the assumptions that the parabola passes through $(1, N_{d0}/N)$ and $(0,1)$ and has a zero derivative at the latter point. This leads to the following approximation for n_c for all values of l:

$$n_{c,prot} = pl^2 + ql + r \tag{5.13}$$

with, $p = N_{d0}/N - 1$, $q = 0$, and $r = 1$.

We compare the approximation Eq.(5.13) with simulation results for the 8 communication networks. The fraction l of removed links in our approximation Eq.(5.13) is from 0 to 1. However, only a fraction $1-l_c$ of links are removed in the simulation for this scenario. Thus, we still remove critical links uniformly at random after all non-critical links are removed, in order to compare the simulation results and our approximation Eq.(5.13) in the same interval $[0,1]$. Figure 5.5 shows that for moderate values of the fraction of removed links, the approximation exhibits an excellent fit for simulation results. For some networks, such as TataNld, UsCarrier and Cogentco, our approximation Eq.(5.13) fits well with the simulation results regardless of the fraction of removed links.

Similarly, the performance of our approximation Eq.(5.13) is measured by three performance indicators. As shown in Table 5.4, when the fraction of removed links is less than 0.2, the absolute relative error between Eq.(5.13) and the simulated mean is less than 5% for all 8 sparse communication networks. For some networks, such as UsCarrier and Cogentco, even when the fraction of removed links is large (0.62 and 0.58, respectively), Eq.(5.13) still fits well with simulation results. Even for the worst performing network, DFN, 28% of the links can be removed before the absolute relative error exceeds 5%.

5.5.4. VERIFICATION BY LARGE NETWORKS

We use the last two large networks, Kdl and Web, from Table 5.1 to further evaluate the accuracy of our approximations. The simulations setting is slightly different from

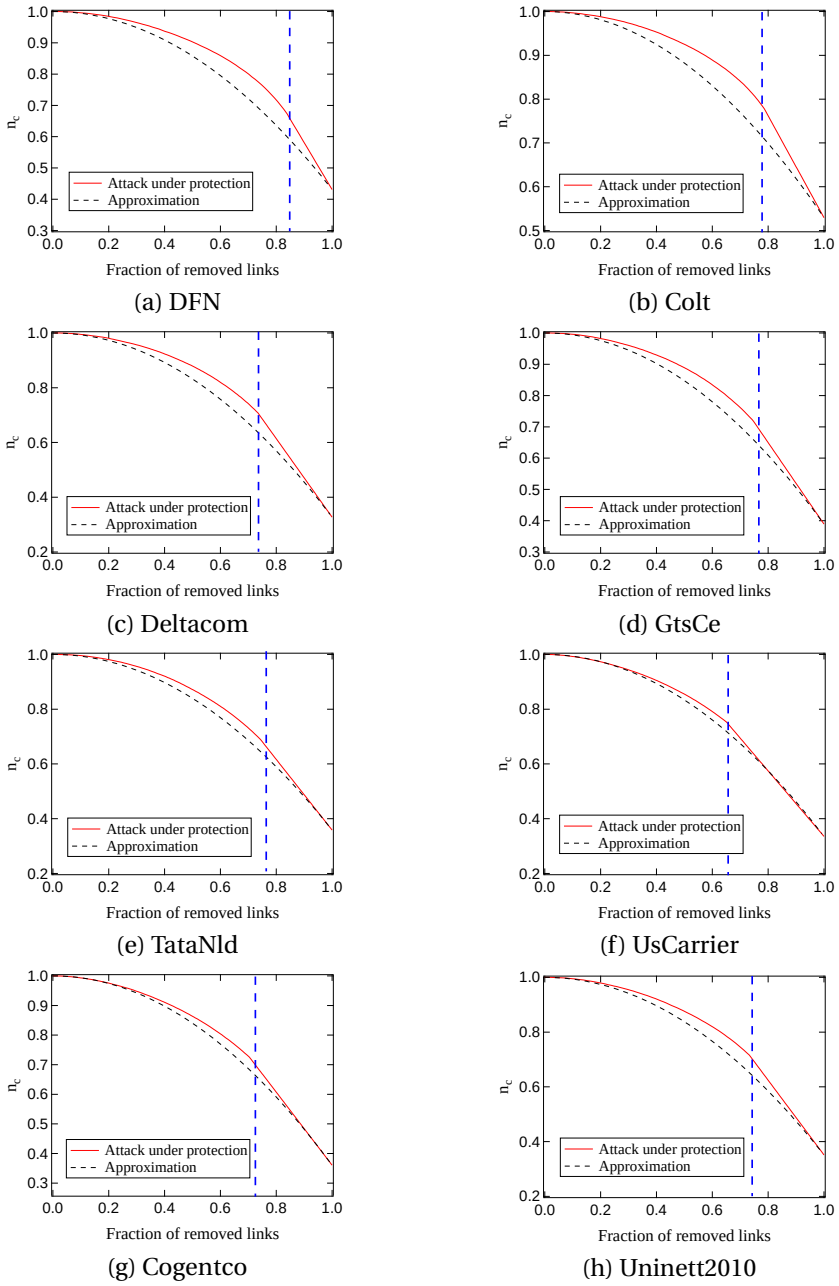


Figure 5.5: The normalized maximum number of controllable nodes n_c as a function of the fraction of removed links l in communication networks under random attacks under protection. The results for each fraction l is based on 1000 simulations. In each sub-figure, the vertical dashed line marks the position where $l = 1 - l_c$. In order to compare the simulation results for random attack under protection with our approximation in the same sub-figure, we remove critical links uniformly at random after all the other links are removed.

Table 5.4: Performance indicators for the approximation $n_{c,prot}$ for the 8 communication networks

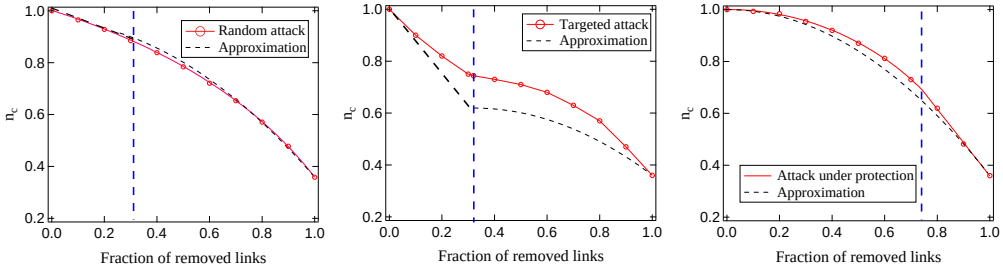
Networks	r^*	l^*	γ
DFN	2.65%	0.28	100%
Colt	1.89%	0.32	100%
Deltacom	1.15%	0.37	100%
GtsCe	1.09%	0.38	100%
TataNld	0.84%	0.42	100%
UsCarrier	0.32%	0.62	100%
Cogentco	0.56%	0.58	100%
Uninett2010	1.22%	0.47	100%

the previous part. The fraction l of removed links is ranging from 0.1 to 1 with a step size 0.1, considering the high computational complexity. As shown in Figure 5.6, the approximation Eq.(5.10) for the random attack and Eq.(5.13) for the random attack with protection perform well in estimating the fraction n_c of driver nodes. We also find that the approximation Eq.(5.12) for targeted attack fits well with simulation results when the fraction of removed links is sufficiently small. Though the approximation Eq.(5.12) does not perform well when the fraction of removed links is large, approximation Eq.(5.12) can be considered a worst-case approximation. Considering the Kdl network and the 8 small networks have similar average degree, the above observation implies that the size of the network does not significantly influence the performance of our approximations. By contrast, Figure 5.7(a) and (c) show that the approximation Eq.(5.10) for the random attack and Eq.(5.13) for the random attack with protection do not perform well for the Web network which has a larger average degree than the above networks. In a network with a higher average degree, there are more alternate matchings which make it more likely for the critical links to change as links are removed. As a result, our approximations do not perform well since our assumption is that the set of critical links is nearly unchanged when the fraction of removed links is small. However, since most communication networks are sparse [98] [99] [100], we can expect that our approximations are applicable for most communication networks.

We then quantify the performance of the approximations for the network Kdl and Web in Table 5.5 and 5.6, respectively. Results show that the network Web has larger r^* values than the network Kdl in all three attacks, which also indicates that our approximation performs better in networks with lower average degree.

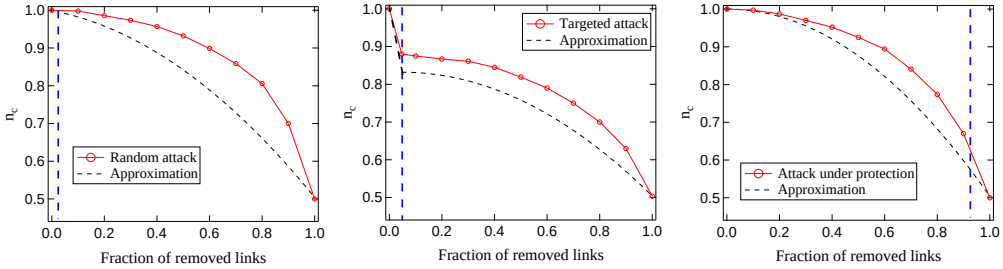
Table 5.5: Performance indicators for the approximation n_c for Kdl

Types of attacks	r^*	l^*	γ
Random attack	2.36%	0.46	100%
Targeted attack	7.54%	0.16	53.65%
Random attack with protection	3.67%	0.38	92.84%



(a) Random attack (b) Targeted attack (c) Random attack with protection

Figure 5.6: Performance of the normalized number n_c of controllable nodes as a function of the fraction of removed links l for three attack scenarios in the Kdl network. The results for each fraction l is based on 1000 simulations. For random attack and targeted attack, the vertical dashed line marks the position where $l = l_c$. For random attack with protection, the vertical dashed line marks the position where $l = 1 - l_c$.



(a) Random attack (b) Targeted attack (c) Random attack with protection

Figure 5.7: Performance of the normalized number n_c of controllable nodes as a function of the fraction of removed links l for three attack scenarios in the Web network. The results for each fraction l is based on 1000 simulations. For random attack and targeted attack, the vertical dashed line marks the position where $l = l_c$. For random attack with protection, the vertical dashed line marks the position where $l = 1 - l_c$.

Table 5.6: Performance indicators for the approximation n_c for Web

Types of attacks	r^*	l^*	γ
Random attack	13.78%	0.11	100%
Targeted attack	9.21%	0.04	54.16%
Random attack with protection	5.25%	0.19	100%

5.5.5. VERIFICATION BY MORE COMMUNICATION NETWORKS

We further use the dataset available at a specialized database - the Internet Topology Zoo [80] to select more communication networks and verify the accuracy of our approximations. The networks in the dataset initially are not directed, however, we use the information available in two attributes, i.e., source node and target node, to make these networks directed. After excluding networks with extremely small size $N < 20$, we have 200 communication networks.

For each attack strategy, we calculate the values of the three performance indicators

for all 200 communication networks and then get the average value for each indicator. As shown in Table 5.7, the approximation Eq.(5.10) for the random attack and Eq.(5.13) for the random attack with protection performs well in estimating the fraction n_c of driver nodes. For the targeted attack, the approximation Eq.(5.12) fits well with simulation results when the fraction of removed links is sufficiently small.

Table 5.7: Performance indicators for the approximation n_c for 200 communication networks

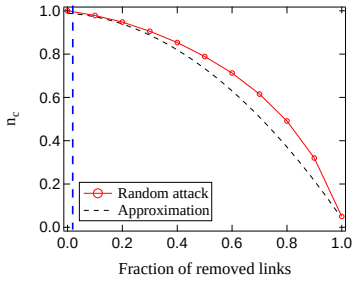
Types of attacks	r^*	l^*	γ
Random attack	4.26%	0.47	98.47%
Targeted attack	10.44%	0.11	47.81%
Random attack with protection	3.52%	0.23	99.36%

5.5.6. VERIFICATION BY SYNTHETIC NETWORKS

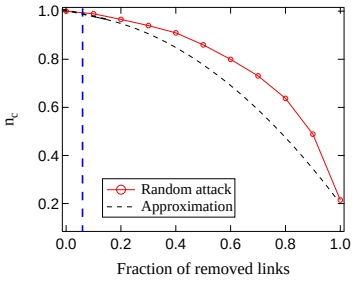
In this section, we test our approximations on two types of synthetic networks, the directed Erdős-Rényi (ER) random network $G_p(N)$ and the Barabási-Albert (BA) scale-free network $BA(N, M_0, M)$. When generating the directed Erdős-Rényi random network $G_p(N)$ with N nodes, the probability that every node has an outbound link to the other nodes is p. We generate the scale-free network $BA(N, M_0, M)$ by using the Barabási-Albert (BA) model, where N is the number of nodes, M is the number of out-going links for each new node added to the current network. We assume that initially the network consists of a complete digraph on M_0 nodes, where M_0 equals M. In the initial complete digraph, every pair of distinct nodes is connected by a pair of unique links (one in each direction). New nodes are added to the network one at a time. Each new node is connected to M existing nodes with a probability that is proportional to the number of links that the existing nodes already have.

In our simulations, we generate Erdős-Rényi (ER) random networks $G_p(N)$ with $N = 100$, $p = 0.05$ and $N = 10000$, $p = 0.0003$, Barabási-Albert (BA) networks with $N = 200$, $M = M_0 = 2$ and $N = 10000$, $M = M_0 = 1$. Figure 5.8 shows that the approximation Eq.(5.10) for the random attack performs well in estimating the fraction n_c of controllable nodes in both types of synthetic networks when the fraction of removed links is small. Figure 5.9 shows that the approximation Eq.(5.12) for the targeted attack performs well as long as the fraction of removed links is sufficiently small. Figure 5.10 shows that Eq.(5.13) for the random attack with protection performs well in both types of synthetic networks when the fraction l of removed links is less than the fraction l_c of critical links. For the large ER and BA networks, Eq.(5.13) fits well with simulation results even when the fraction l of removed links is large. The approximation Eq.(5.12) does not perform well if the fraction l of removed links is large. However, Eq.(5.12) can be considered an approximation for the worst-case scenario.

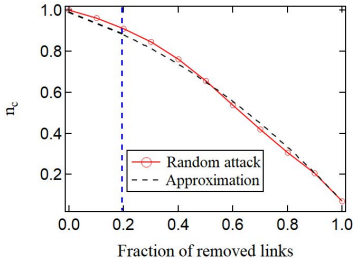
Next we quantify the performance of each approximation for synthetic networks. As shown in Table 5.8, 5.9 and 5.10, the approximation Eq.(5.10) for random attack and Eq.(5.13) for random attack with protection fit well with simulation results even when the fraction l of removed links is relatively large ($l = 0.2$).



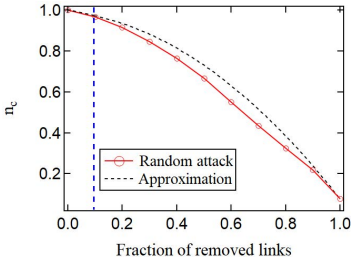
(a) ER: $G_{0.05}(100)$, $E[D]$=5



(b) BA: N=200, $E[D]$=4

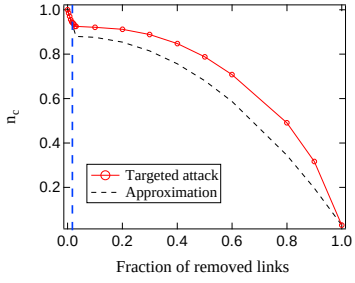


(c) ER: $G_{0.0003}(10000)$, $E[D]$=3

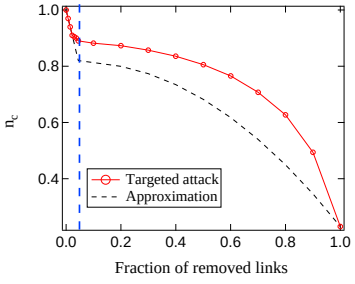


(d) BA: N=10000, $E[D]$=2

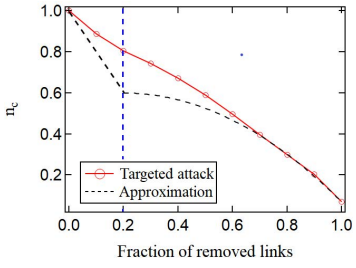
Figure 5.8: The normalized maximum number of controllable nodes n_c as a function of the fraction of removed links l in synthetic networks under random attacks. The results for each fraction l is based on 10000 simulations.



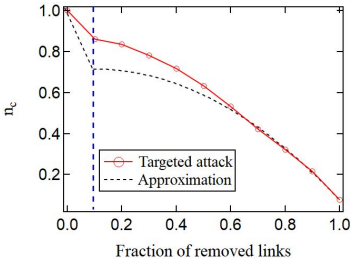
(a) ER: $G_{0.05}(100)$, $E[D]$=5



(b) BA: N=200, $E[D]$=4



(c) ER: $G_{0.0003}(10000)$, $E[D]$=3



(d) BA: N=10000, $E[D]$=2

Figure 5.9: The normalized maximum number of controllable nodes n_c as a function of the fraction of removed links l in synthetic networks under targeted attacks. The results for each fraction l is based on 10000 simulations.

Table 5.8: Performance indicators for the approximation $n_{c,rand}$ for synthetic networks

Types of networks	r^*	l^*	γ
ER: $G_{0.05}(100)$	3.27%	0.34	100%
BA: N=200, $E[D]$=4	6.78%	0.18	100%
ER: $G_{0.0003}(10000)$	8.95%	0.14	85.54%
BA: N=10000, $E[D]$=2	7.63%	0.17	96.23%

Table 5.9: Performance indicators for the approximation $n_{c,crit}$ for synthetic networks

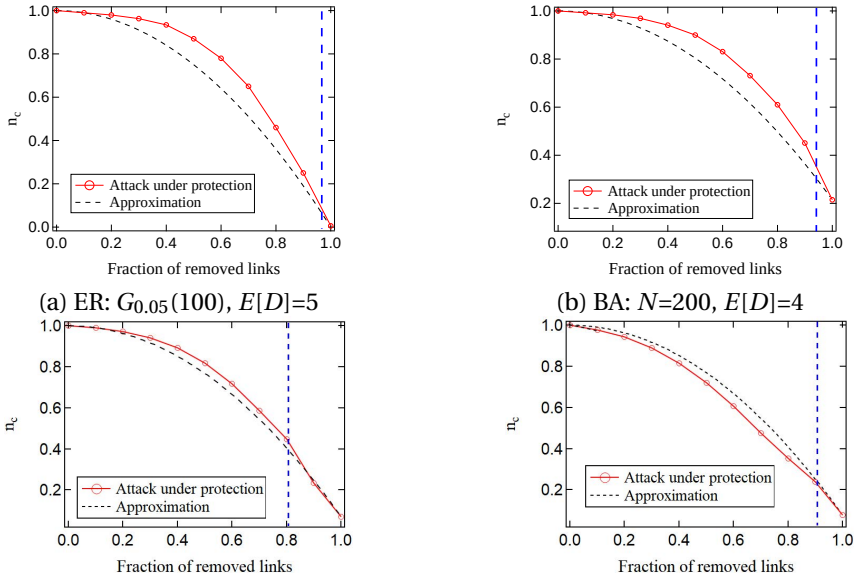
Types of networks	r^*	l^*	γ
ER: $G_{0.05}(100)$	12.64%	0.03	71.97%
BA: N=200, $E[D]$=4	17.36%	0.04	63.91%
ER: $G_{0.0003}(10000)$	23.56%	0.07	34.28%
BA: N=10000, $E[D]$=2	16.28%	0.05	46.76%

Table 5.10: Performance indicators for the approximation $n_{c,prot}$ for synthetic networks

Types of networks	r^*	l^*	γ
ER: $G_{0.05}(100)$	5.21%	0.19	100%
BA: N=200, $E[D]$=4	8.63%	0.17	100%
ER: $G_{0.0003}(10000)$	4.16%	0.23	100%
BA: N=10000, $E[D]$=2	6.94%	0.19	100%

5.6. CONCLUSION

In this chapter, we analyzed the role of critical links in reachability-based network controllability. Simulation results on communication networks have suggested analytical closed-form approximations for the number N_c of controllable nodes. We derived closed-form approximations for the number N_c of controllable nodes as a function of the fraction of removed links, for random attacks, targeted attacks and random attack under protection. Both for random and targeted attacks, our approximation is linear in the fraction l of removed links when this fraction is smaller than the fraction of critical links. When the fraction of removed links is larger than the fraction of critical links, our approximation is quadratic in l. We validated our approximation through simulations on sparse communication networks and synthetic networks. Both for random attacks and random attacks under protection, our approximations for these two cases are always very good, as long as the fraction of removed links is smaller than the fraction of critical links. For some cases, the approximation is still accurate for larger fractions of removed links. For targeted attack, our approximation performs well as long as the fraction of removed links is sufficiently small, whereas our approximation does not perform well when the fraction of removed links is large. However, the approximation for the targeted attack always serves as a worst-case estimate.



(a) ER: $G_{0.05}(100)$, $E[D]$=5

(b) BA: N=200, $E[D]$=4

(c) ER: $G_{0.0003}(10000)$, $E[D]$=3

(d) BA: N=10000, $E[D]$=2

Figure 5.10: The normalized maximum number of controllable nodes n_c as a function of the fraction of removed links l in synthetic networks under random attacks with protection. The results for each fraction l is based on 10000 simulations. In each sub-figure, the vertical dashed line marks the position where $l = 1 - l_c$.

6

THE RECOVERABILITY OF OPTICAL NETWORKS

Optical networks are vulnerable to failures due to targeted attacks or large-scale disasters. The recoverability of optical networks refers to the ability of an optical network to return to a desired performance level after suffering topological perturbations such as link failures. This chapter proposes a general topological approach and recoverability indicators to measure the network recoverability for optical networks for two recovery scenarios: 1) only the links which are damaged in the failure process can be recovered and 2) links can be established between any pair of nodes that have no link between them after the failure process. We use the robustness envelopes of realizations and the histograms of two recoverability indicators to illustrate the impact of the random failure and recovery processes on the network performance. By applying the average two-terminal reliability and the network efficiency as robustness metrics, we employ the proposed approach to assess 20 real-world optical networks. Numerical results validate that the network recoverability is coupled to the network topology, the robustness metric and the recovery strategy. We further show that a greedy recovery strategy could provide a near-optimal recovery performance for the robustness metrics. We investigate the sensitivity of network recoverability and find that the sensitivity of the recoverability indicators varies according to different robustness metrics and scenarios. We also find that assortativity has the strongest correlation with both recoverability indicators.

6.1. INTRODUCTION

High reliability and robustness in optical network backbones play an important role in successfully provisioning high service availability of the Internet and communication systems [102]. In optical networks, disaster-based failures and damages to optical fiber cables can partially overload data delivery, resulting in unavailability of communication services [103]. The causes for such massive failures include: human errors, malicious

This chapter is based on the published paper [101].

attacks, large-scale disasters, and environmental challenges [104]. Calculating the performance of networks under such challenges can provide significant insight into the potential damage they can incur, as well as provide a foundation for creating more robust infrastructure networks.

Network robustness is interpreted as a measure of the response of the network to perturbations, or challenges, imposed on the network [75], which has been studied extensively in recent years. Van Mieghem *et al.* [75] propose a framework for computing topological network robustness by considering both a network topology and a service for which the network is designed. In communication networks, Cholda *et al.* [105] survey various robustness frameworks and present a general framework classification, while Pašić *et al.* [106] present the FRADIR framework that incorporates reliable network design, disaster failure modeling and protection routing. A wide range of metrics based on the underlying topology have been proposed to measure network robustness [107], and further a structural robustness comparison of several telecommunication networks under random nodal removal is presented in [108]. Long *et al.* [109] propose using the maximum variation of the Weighted Spectrum (WS) to measure the survivability of networks to geographic correlated failures. For optical networks applications, Zhu *et al.* [110] investigate the control plane robustness in software-defined optical networks under different link cut attack scenarios and find that control plane enhancements in terms of controller addition do not necessarily yield linear improvements in control plane robustness but require tailored control plane design strategies. Ferdousi *et al.* [111] propose a rapid data-evacuation strategy to move maximum amounts of data from disaster regions using survived resources under strict time constraints for optical cloud networks. Xie *et al.* [112] come up with a robust and time-efficient algorithm to address the emergency backup in inter-datacenter networks with progressive disasters.

The work mentioned above focus on measuring and improving the ability of networks to withstand failures and attacks. However, the recovery process after failures is not considered and the investigation on the ability of a network to recover from failures is lacking. In a broad sense, network robustness is also related to the ability of a network to return to a desired performance level after suffering malicious attacks and random failures [113]. We define such network capability as *network recoverability*[1] in this chapter. As shown in Figure 6.1, recovery measures are taken in order to recover the function or performance of the optical network after the failure process, either by restoring the damaged links or by building new links. The network performance during this period is related to many factors, such as topology, recovery strategy, link adding sequence, etc. Thus, we need an approach to measure the recoverability of optical networks.

Several recovery mechanisms have been investigated under different circumstances [114], particularly in complex networks applications. For example, Majdandzic *et al.* [115] model cascading failures and spontaneous recovery as a stochastic contiguous spreading process and show the occurrence of a phase switching phenomenon. Chaoqi *et al.* [116] construct a dynamic repair model and systematically analyze the energy-transfer relationships between nodes in the repair process of the failure network. Recovery strategies based on centrality metrics of network elements (e.g., nodes or links) are investigated in [113], [117], which show that a centrality metric-based strategy may not

[1] Sometimes also called *network restoration*.

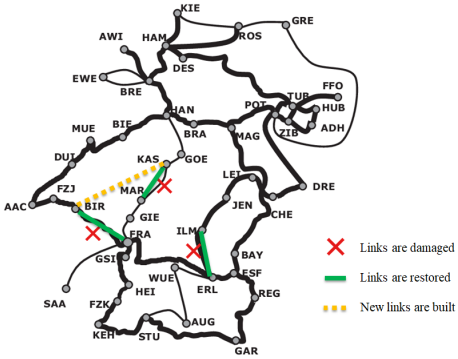


Figure 6.1: Failure and recovery in the DFN (German optical backbone X-WiN network).

exist to improve all the network performance aspects simultaneously.

In optical networks applications, Alenazi *et al.* [118] propose a heuristic algorithm that optimises a network by adding links to achieve a higher network resilience by maximising the algebraic connectivity while decreasing the total cost via selecting cost-efficient links. Natalino *et al.* [119] introduce two heuristics to upgrade Content Delivery Networks (CDNs) and increase content accessibility under targeted link cuts. Hong *et al.* [120] propose a recovery strategy to recover the boundary of the failed nodes in interdependent networks during cascading failures. A progressive recovery approach [121], that consists in choosing the right sequence of links to be restored after a disaster in communication networks, proposes to maximize the weighted sum of the total flow over the entire process of recovery [122], as well as to minimize the total cost of repair under link capacity constraints [123].

Although the above papers [114]– [123] have contributed to a deep understanding of recovery processes in networks, a general framework or methodology for quantifying the recovery capability of a real-world optical network is still lacking. In this chapter, we propose a topological approach and two recoverability indicators to quantify the network recoverability for two different recovery scenarios, which we will denote as Scenario A and Scenario B. For the link-based Scenario A, links can be established between any pair of nodes that have no link between them, after the failure process. The energy-based Scenario B assumes that only the links which are damaged in the failure process can be recovered.

The proposed approach involves three concepts: the network topology, the robustness metric and the recovery strategy. For an optical network G, we apply the average two-terminal reliability $ATTR$ and the network efficiency E_G as the robustness metrics for case studies. The average two-terminal reliability $ATTR$ is defined as the probability that the service between a randomly chosen node pair in the network is available, which also expresses the level of difficulty to disconnect parts of the network. The network efficiency E_G gives an indication of the efficiency of information exchange on networks under shortest path routing [124]. Besides a random recovery strategy and some strategies based on topological properties, we also consider a greedy recovery strategy. In the greedy

strategy, the damaged element (a node or a link) which improves the network performance most has the highest priority to be recovered. Our approach is tested on 20 real-world optical networks, and we verify that the proposed recoverability indicators allow us to compare the performance of different recovery strategies and assess the recoverability of different networks.

The rest of this chapter is organized as follows: Section 6.2 introduces the topological approach for measuring the network recoverability for the two considered recovery scenarios. Section 6.3 presents the main concepts in the evaluation of network recoverability. The experimental results are exhibited in Section 6.4. Section 6.5 discusses the sensitivity of the network recoverability on different robustness metric thresholds. Section 6.6 analyzes the correlation of topological metrics with recoverability indicators. Section 6.7 concludes the chapter.

6.2. TOPOLOGICAL APPROACH FOR MEASURING NETWORK RECOVERABILITY

In this section, we introduce an approach for measuring the network recoverability for real-world optical networks for two recovery scenarios.

6.2.1. R-VALUE AND CHALLENGES

We inherit the framework and some definitions proposed for network robustness [52, 75] and extend the methodology for the network recoverability. A given network determined by a service and an underlying topology is translated into a mathematical object, defined as the R-value, on which computations can be performed [75]. The R-value takes the service into account and is normalized to the interval $[0,1]$. Here, $R = 1$ reflects complete functionality in a network without failures, and $R = 0$ corresponds to the complete lack of functionality for a sufficiently degraded network.

An elementary challenge is an event that changes the network and thus possibly changes the R-value. We assume that a elementary changes take place one by one, and thus do not coincide in time. Considering link-based failures and targeted link cuts as common threats to optical infrastructure networks, we confine an elementary challenge to a link removal in a failure process or a link addition in a recovery process. Since every perturbation has an associated R-value, any realization of such a failure process, followed by a recovery process, consists of a number M of elementary challenges and hence can be described by a sequence of R-values denoted $\{R[k]\}_{1\leq k\leq M}$, where k is the sequence number of elementary challenges.

6.2.2. LINK-BASED SCENARIO A: RECOVERY OF ANY ALTERNATIVE LINK

Let $M_{G_0(N,L)}$ denote the robustness metric value of the original network $G_0(N,L)$, with N nodes and L links. Assume that during the failure-recovery process, the resulting graph has L^* links and is denoted by $G(N,L^*)$. We define the R-value R_G as the normalized value of the robustness metric $M_{G(N,L^*)}$, which satisfies

$$R_G = \frac{M_{G(N,L^*)}}{M_{G_0(N,L)}} \tag{6.1}$$

Thus, the R-value R_{G_0} of the original network $G_0(N,L)$ equals 1.

We assume failures in the network only consist of link removals in the network, according to a fixed strategy, such as random failure or targeted link cuts, which usually degrade the robustness of the network. We assume that links are damaged (removed) one by one, until we obtain a graph G_f, whose R-value R_{G_f} first reaches or drops below a constant ρ, where $\rho \in [0,1]$ is a prescribed R-threshold for the robustness metric. Usually this threshold is chosen in such a way that while the R-value is still above it, the service quality remains acceptable [75]. The above process is called the failure process. The number of *failure challenges*, i.e., the number of damaged links in the failure process, is denoted by K_f. For the same network G_0, the smaller the value of K_f, the more effective the failure process is in degrading the R-value [75].

Then we launch the recovery process from the remaining network $G_f(N, L-K_f)$. Scenario A assumes that the recovery links can be established between any two nodes in the complement of the graph after failures. The process of one realization is illustrated in Figure 6.2(a). Specifically, we recover the network by adding links, one by one, to the damaged network G_f by a recovery strategy until the normalized robustness metric R_G first reaches or excesses $R_{G_r} = 1$. The network after the recovery process is denoted by $G_r(N, L-K_f+K_r)$, where K_r is the number of *recovery challenges* (i.e. the number of links that are added during the recovery process). For a given damaged network G_f, the smaller the value of K_r, the more effective the recovery process is. Ideally, the recovery process increases the R-value of the current network exactly to 1. However, the R-value R_{G_r} of the resulting network $G_r(N, L-K_f+K_r)$ is mostly larger than 1, since the robustness metric value of the resulting network $G_r(N, L-K_f+K_r)$ is slightly larger than that of the original network $G_0(N,L)$ in most cases.

We define the *Link Ratio* η_L as the ratio of the number of failure challenges K_f and the recovery challenges K_r, i.e.,

$$\eta_L(G,\rho) = \frac{K_f}{K_r}, \tag{6.2}$$

which indicates the efficiency of the recovery process in one realization. A Link Ratio $\eta_L(G,\rho) > 1$ implies that the network can be recovered by less challenges than the number K_f of failure challenges. Otherwise, the network is more difficult to recover than to destroy.

Scenario A can characterize the recovery process in a connection oriented network with logical connections [125], e.g., a virtual circuit for transporting data or a wireless backhaul network, where the links in a logical network represent the duplex channel between two devices. For example, after channels are interrupted because of signal fading or blocking in a mobile network, one should establish several connections or reconfigure several new channels to maintain the network's overall performance. Besides, Scenario A can also apply to the situation where network operator has the capability to build connections between any node pairs in the network. In this case, the overhead cost of the recovery measures mainly depends on the total number of dispatched connections, which corresponds to the number K_r of recovery challenges in Scenario A.

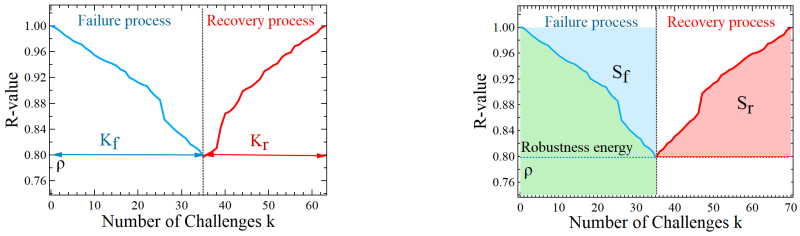


Figure 6.2: Illustration of the failure process and the recovery process in an Erdős-Rényi (ER) random graph $G_{0.1}(100)$ with link density $p = 0.1$ and network size $N = 100$ in one realization. The R-threshold is $\rho = 0.8$.

6.2.3. ENERGY-BASED SCENARIO B: RECOVERY OF FAILED LINKS

The failure process in Scenario B is the same as in Scenario A. In the recovery process in Scenario B, we restore one by one, all the links which were removed during the failure process, until the network is restored to its original topology. Scenario B can be used to describe recovery processes in physical communication networks, e.g., optical backbone networks. In such networks, the recovery measure for each connection, e.g., repairing fiber optic cables, usually requires a relatively long period. During the recovery process, the network still provides services, albeit with a degraded performance. Thus, for this scenario, the network recoverability is related to the network performance (or the robustness metric) throughout the recovery process.

One realization of the failure and recovery process is illustrated in Figure 6.2(b). In Scenario B, the number of failure challenges and the number of recovery challenges are the same, i.e., $K_f = K_r$, and hence, $\eta_L = 1$ in Eq.(6.2). Therefore, we propose another recoverability indicator for Scenario B. The *robustness energy* $S(G,\rho)$ of a network G is the sum of the R-values during the failure process, i.e. $S(G,\rho) = \sum_{k=0}^{K_f} R[k]$, and expresses the robustness performance of the network under successive failures [52]. Thus, the energy of failure challenges is computed by $S_f(G,\rho) = \sum_{k=0}^{K_f}(1-R[k]))$, which indicates the cumulative degradation of the network performance during the failure process. During the recovery process, the energy of the recovery challenges $S_r(G,\rho) = \sum_{k=0}^{K_f}(R[k]-\rho)$ represents the impact of the recovery process on the network performance. For Scenario B we define the *Energy Ratio*, denoted by η_E, as the ratio between the energy of the recovery challenges S_r and the energy of the failure challenges S_f, in each realization for a given R-threshold ρ:

$$\eta_E(G,\rho) = \frac{S_r}{S_f}. \tag{6.3}$$

An Energy Ratio $\eta_E(G,\rho) > 1$ implies the benefit of recovery measures can compensate the loss of network performance by the failures, which indicates a high network recovery capability. Conversely, an Energy Ratio $\eta_E(G,\rho) < 1$ implies a low recoverability.

6.2.4. COMPARISON VIA ENVELOPES AND THE RECOVERABILITY INDICATORS

As we discussed in Section 6.2.1, the impact of any realization of failure and subsequent recovery process on the network's functionality can be expressed as a sequence of R-

values $\{R[k]\}$, where k is the sequence number of elementary challenges. To investigate the recoverability of networks, we need to know the number of challenges needed to make the original R-value (which is normalized to 1) decrease to a predefined R-threshold ρ in the failure process and also the number of challenges needed to increase the R-threshold ρ to the original R-value. This confines us to investigate the number of challenges K as a function of a specific R-value r, i.e., $\{K[r]\}$. Thus, each value in $\{K[r]\}$ is the number of challenges that is needed to change R-value to a specific R-value r for each realization. Considering that it is impossible to list all values of r between the R-threshold ρ and the original R-value, we evenly sampled $H = 1000$ different r values in the interval $[\rho, 1]$. Thus, $r_j = \rho + \frac{(j-1)(1-\rho)}{H-1}$ where j is the jth value of r. The *envelope* is constructed using all sequences $\{K[r]\}$ for $r \in \{r_1, r_2, \ldots, r_H\}$. The boundaries of the envelope are given by the extreme number of challenges K

$$K_{\min}[r] \in \{\min(K[r_1]), \min(K[r_2]) \ldots, \min(K[r_H])\}, \tag{6.4}$$

$$K_{\max}[r] \in \{\max(K[r_1]), \max(K[r_2]) \ldots, \max(K[r_H])\}, \tag{6.5}$$

which gives the best- and worst case values of the robustness metrics for a network after a given number of recovery challenges. The expected number of challenges K leading to the topological approach r_j is

$$K_{avg}[r] \in \{E(K[r_1]), E(K[r_2]) \ldots, E(K[r_H])\}. \tag{6.6}$$

Since $K[r]$ defines a probability density function (pdf), we are interested in the percentiles of $K[r]$

$$K_{m\%}[r] \in \{K_{m\%}[r_1], K_{m\%}[r_2] \ldots, K_{m\%}[r_H]\}, \tag{6.7}$$

where $K_{m\%}[r]$ are the points at which the cumulative distribution of $K[r]$ crosses $\frac{m}{100}$, namely $K_{m\%}[r] = t \Leftrightarrow \Pr[K[r] \leq t] = \frac{m}{100}$.

We apply the envelopes to present the behavior of the failure and recovery processes on a network [52,75]. The envelope profiles the pdf of the random variables of the number of challenges K, which is the probability of a random variable to fall within a particular region. The area of the envelope can be regarded as the variation of the robustness impact of a certain series of challenges, which quantifies the uncertainty or the amount of risk due to perturbations.

We propose two recoverability indicators, the Link Ratio $\eta_L(G, \rho)$ and the Energy Ratio $\eta_E(G, \rho)$, for different scenarios, respectively. Since a failure process and a recovery process could be random under the random strategy, the recoverability indicators are random variables. We compare the recoverability of different networks by the average recoverability indicators for simplicity. For example, the average Link Ratio $E[\eta_L(G_1, \rho)] > E[\eta_L(G_2, \rho)]$ for two different networks G_1 and G_2 implies that the network G_1 usually has a better recoverability than G_2 in Scenario A for a given R-threshold ρ.

Besides the average recoverability indicators, we are also concerned about the variance of the recoverability indicators $Var[\eta(G, \rho)]$. A smaller variance of the recoverability indicators $Var[\eta(G, \rho)]$ implies a narrower uncertainty of the recoverability indicators, thus a better recoverability.

6.3. ROBUSTNESS METRICS AND RECOVERY STRATEGIES

In this section, we introduce the factors which determine specific recovery process, namely robustness metrics, recovery strategies and network topologies.

6.3.1. ROBUSTNESS METRICS

We use two metrics: the average two-terminal reliability $ATTR$ and the network efficiency E_G, as the robustness metrics. These two metrics are closely related to service availability and data delivery on optical networks.

1) **Average two-terminal reliability** $ATTR$**.** In optical networks, the average two-terminal reliability (ATTR) can assess the resilience and vulnerability of a fiber infrastructure [126, 127]. The metric is defined as the fraction of pairs of nodes with a path between them

$$ATTR(G) = \frac{\sum_{i \neq j \in G} 1_{\text{exists a path between}(i,j)}}{\binom{N}{2}}. \tag{6.8}$$

The $ATTR$ measures the reachability fraction of any pair of nodes, but ignores the performance of the information exchange in a network. $ATTR$ equals 1 when the network is fully connected; otherwise $ATTR$ is the sum of the number of node pairs in every connected component, divided by the total number of node pairs in the network. At failure scenarios, the higher the average two-terminal reliability, the higher the robustness [108].

2) **Network efficiency** E_G**.** We assume that the hopcount $h(i, j)$, i.e., the number of links in the shortest path from node i to j, indicates the overhead of data delivery from end to end. Thus, the reciprocal of the hopcount $1/h(i, j)$ implies the amount of packages for one unit overhead, which can be interpreted as the efficiency of data delivery between two nodes in optical networks. If there is no path from i to j, $h(i, j) = \infty$ and $1/h(i, j) = 0$. The efficiency of a given network is defined as the mean of the reciprocals of all the hopcounts $h(i, j)$ in a network, i.e.,

$$E_G = \frac{\sum_{i \neq j \in G} 1/h(i, j)}{\binom{N}{2}}, \tag{6.9}$$

see [124]. Network efficiency E_G quantifies the efficiency of information exchange across the whole network under shortest path routing [128], such as the data transmission between controllers and switches in software-defined optical networks. Network efficiency monotonically decreases with successive link removals.

6.3.2. FAILURE AND RECOVERY STRATEGIES

For simplicity and generality, we consider a *random failure* strategy. The random failure strategy implies that the failures occur independently on links randomly and uniformly, which is consistent with the random failure stage in a product life cycle. The R-value $R[k]$ for a determined number of failure challenges k is a random variable. We consider three different strategies for recovery measures, i.e., random recovery, metric-based recovery and greedy recovery:

1) **Random recovery:** The random recovery strategy refers to the strategy that the links are added randomly and uniformly, one by one, during the recovery process, which can describe a self-repairing process after failures or recovery measures without scheduling.

2) **Metric-based recovery:** The metric-based strategy determines the sequence of adding links by the topological or spectral metrics of links. While there are many relevant metrics, such as closeness and the effective resistance [129] [130], we use three metric-based recovery strategies. The selection criteria of the link between nodes i and j for each strategy are illustrated as follows:

(a) The minimum product of degrees $d_i d_j$. For each challenge in a recovery process, we select and restore the link l_{ij}^* with the minimum $d_i d_j$. If there are multiple node pairs with the same minimum product of degrees, one of these pairs is randomly chosen.

(b) The minimum product $(x_1)_i (x_1)_j$ of the ith and jth components of the eigenvector x_1 belonging to the largest adjacency eigenvalue [131]. For each challenge in a recovery process, we restore the link l_{ij}^* with the minimum $(x_1)_i (x_1)_j$.

(c) The maximum absolute difference $\Delta y = max(|y_i - y_j|)$, where $|y_i - y_j|$ is the absolute difference between the ith and jth components of the Fiedler vector y [20]. For each challenge in a recovery process, we restore the link l_{ij}^* with the maximum Δy.

3) **Greedy recovery:** The greedy recovery strategy involves adding the link l_{max}^* that makes the R-value increase the most in each challenge,

$$l_{max}^* = \arg\max_{l \in G^c} R(G+l) - R(G) \tag{6.10}$$

where G^c is the complement of the current network G. The greedy strategy is a practical and intuitive recovery strategy, where the current optimal link for improving the performance of the network has the priority to be recovered.

4) **Worst case recovery:** The worst case recovery strategy involves adding the link $l^* min$ that makes the R-value increase the least in each challenge,

$$l_{min}^* = \arg\min_{l \in G^c} R(G+l) - R(G) \tag{6.11}$$

where G^c is the complement of the current network G. This strategy is supposed to be an inefficient recovery strategy, where each time the link that contributes the least to the restoration of the network, is recovered.

6.3.3. OPTICAL NETWORKS

As a case study we select 20 real-world optical communication networks. This set of networks was selected from the Internet Topology Zoo [132], covering optical backbone networks located in different regions of the world, see Table 7.1.

The topological properties of the 20 real-world optical networks are described in Table 6.1: the number of nodes N and links L, the average degree $E[D]$, the spectral radius λ_1, the algebraic connectivity μ_{N-1}, the diameter φ and the assortativity ρ_D. As shown in Table 6.1, the average degree $E[D]$ of the 20 optical networks is less than 3. Most of the 20 optical networks have a small value of the algebraic connectivity μ_{N-1}. Besides, 18 out of 20 optical networks have a negative assortativity ρ_D, which signifies a preference of high-degree nodes to connect to other low-degree nodes [133].

6.4. RESULTS AND DISCUSSION

In this section, detailed results and analysis on the real-world optical networks via the proposed approach for assessing network recoverability are presented. For some evaluation items, we only present results for a specific network, i.e., US_signal. We set the R-threshold as $\rho = 0.8$ in the following simulations. The approach translates easily to other networks or other robustness metrics.

Table 6.1: Topological properties of the 20 real-world optical networks.

Networks	Location	N	L	$E[D]$	λ_1	μ_{N-1}	φ	ρ_D
Funet	Finland	26	30	2.31	2.71	0.12	9	-0.31
Intellifiber	US	73	95	2.60	3.55	0.03	15	-0.03
ValleyNet	US	39	51	2.62	3.42	0.03	16	0.10
IowaNet	US, Iowa	33	41	2.48	2.95	0.11	9	-0.32
LambdaNet	Germany	42	46	2.19	2.53	0.04	13	-0.48
Ntelos	US, Virginia	47	58	2.47	3.01	0.04	17	-0.002
PionierL1	Poland	36	41	2.28	2.73	0.08	11	-0.30
RoEduNet	Romania	48	52	2.17	2.95	0.04	13	-0.32
Shentel	US	28	35	2.50	3.14	0.05	13	0.32
US_Signal	US	61	78	2.56	2.89	0.04	14	-0.23
Darkstrand	US	28	31	2.21	2.34	0.07	11	-0.25
Interoute	Europe	110	146	2.67	3.34	0.03	17	-0.20
Missouri	US, Missouri	67	83	2.48	3.09	0.04	14	-0.07
NetworkUSA	US	35	39	2.23	2.63	0.08	10	-0.13
Oteglobe	Europe	83	99	2.39	3.39	0.04	14	-0.22
Palmetto	US, Carolina	45	64	2.84	3.36	0.07	12	-0.15
Sunet	Sweden	26	32	2.46	2.77	0.08	12	-0.42
Switch	Switzerland	74	92	2.49	3.43	0.04	13	-0.37
Syringa	US	74	74	2.00	2.91	0.01	31	-0.35
VtlWavenet	Europe	88	92	2.09	2.32	0.01	31	-0.12

6.4.1. ENVELOPE EXAMPLES AND COMPARISON

Each realization of processes consists of a failure process and a subsequent recovery process. Figure 6.3 exemplifies the envelopes [52] of the challenges in US_signal network for two scenarios and two robustness metrics, $ATTR$ and E_G, respectively, under the random recovery strategy. The envelopes for the failure processes are similar in different scenarios while link-based Scenario A usually needs more challenges to recover the robustness metrics than energy-based Scenario B, if the random recovery strategy is employed. The total number of challenges $K_f + K_r$ could cover a wide range of values since the number of challenges $K_f + K_r$ is influenced by two random processes (i.e., failure and recovery).

Figure 6.3(a) and Figure 6.3(c) also illustrate that the R-value of the average number of challenges $R[K_{avg}]$ for the robustness metric $ATTR$ does not change smoothly with the number of challenges, in both the failure process and the recovery process, because only when a new component appears during the failure process or a component disappears during the recovery process, the $ATTR$ value changes. Furthermore, $R[K_{avg}]$ for $ATTR$ decreases slowly during the initial stage of failure process but increases fast during the initial recovery process. For the robustness metric E_G, the function $R[K_{avg}]$ is slightly concave, illustrated in Figure 6.3(b) and Figure 6.3(d). We will show that the concavity of the function $R[K_{avg}]$ could help to explain the behavior of the recoverability indicators.

6.4.2. COMPARISON OF RECOVERY STRATEGIES

The envelope computation can be applied to compare the performance of different recovery strategies for a specific realization of failures. Figure 6.4 shows different recovery strategies (e.g., random, minimum $d_i d_j$, minimum $(x_1)_i (x_1)_j$, maximum Δy, worst case and greedy) for one realization of failure processes under random failure strategy in the US_signal network. The envelope of recovery processes by random recovery for the average two-terminal reliability $ATTR$ covers a larger surface than that of the network

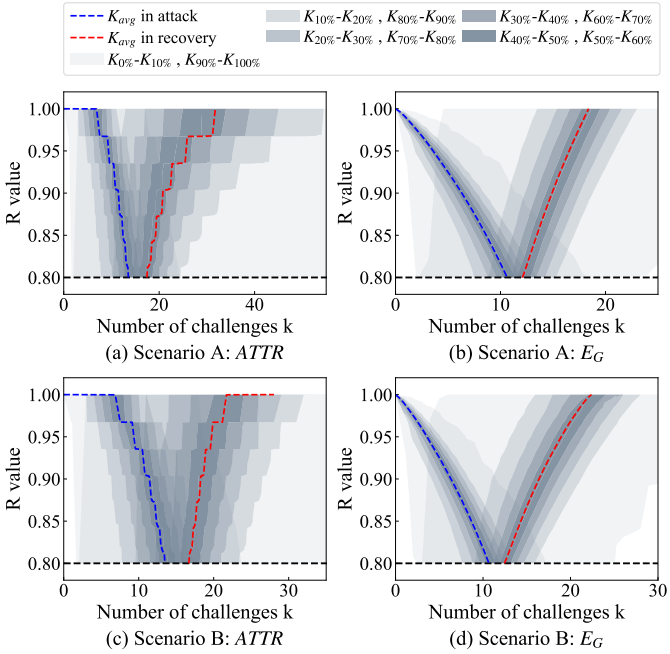


Figure 6.3: Envelopes of the challenges for two scenarios and two robustness metrics (i.e., the average two-terminal reliability $ATTR$ and the network efficiency E_G) in US_signal network, by random recovery strategy. Each envelope is based on 10^4 realizations.

efficiency E_G. This implies that the average two-terminal reliability $ATTR$ in different realizations could deviate more under the random recovery and that the performance of random recovery is more difficult to guarantee. The average challenge sequence $\{K_{avg}\}$ under the random recovery can be a standard to evaluate the performance of other recovery strategies. As shown in Figure 6.4(a) and Figure 6.4(c), the Fiedler vector-based strategy is comparable to the degree-based recovery in Scenario A and the eigenvector-based strategy in Scenario B, which outperforms the average random recovery.

Figure 6.4 also shows that none of the metric-based strategies, with minimum degree product, minimum eigenvector centrality product or maximum absolute difference between Fiedler vector components, can always outperform others for both robustness metrics in both scenarios. Figure 6.4(a) and Figure 6.4(c) exemplify that though the degree-based recovery performs well in link-based Scenario A for $ATTR$, it does not effectively recover the network in energy-based Scenario B. The eigenvector-based strategy outperforms the average behavior of the random strategy in the initial stage of recovery processes but degrades for more recovery challenges in Scenario A. As is shown in Figure 6.4(b) and Figure 6.4(d), these three metric-based recovery strategies are close to and even worse than the average random recovery.

Meanwhile, we notice that the greedy recovery usually upper bounds the random recovery envelopes. The R-value as a function of the number of challenges k under the greedy strategy is concave in the recovery process, which demonstrates the diminish-

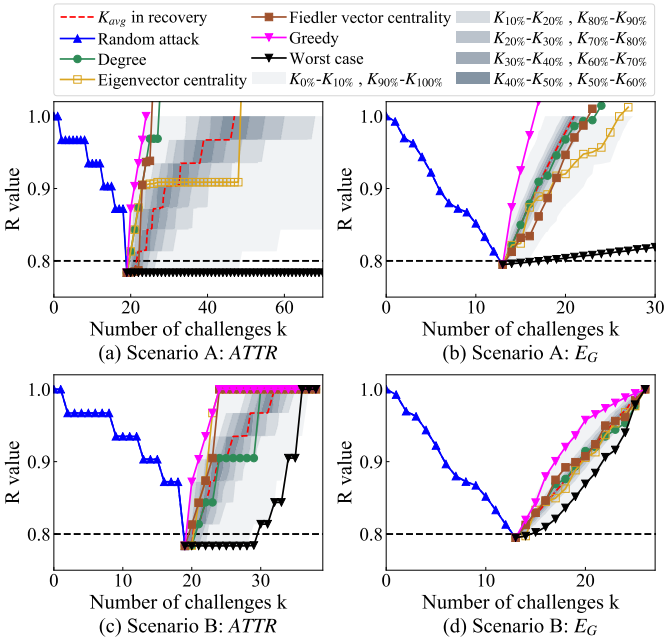
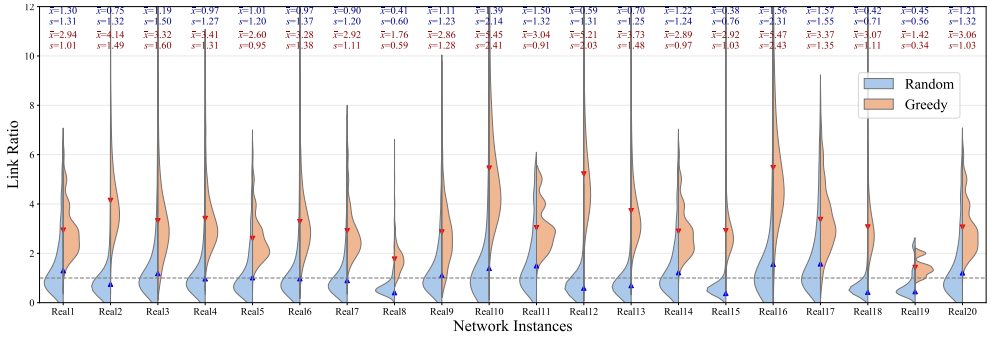


Figure 6.4: Comparisons of different recovery strategies for one realization of failures in US_signal network. Two scenarios and two robustness metrics (i.e., the average two-terminal reliability $ATTR$ and the network efficiency E_G) are applied. Each envelope is based on 10^4 realizations.

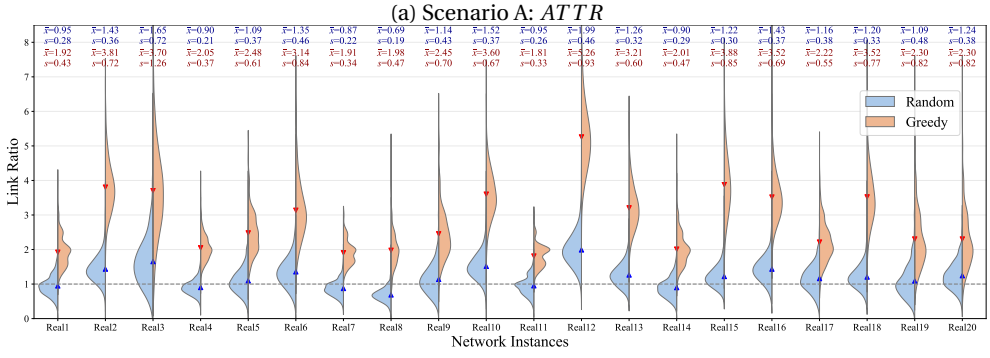
ing returns property of the recovery measures. The greedy recovery provides the most effective way to recover the performance for both robustness metrics, $ATTR$ and E_G, when compared with other listed recovery strategies. The worst case recovery strategy is usually beneath the random recovery envelopes. Among all recovery strategies, the greedy/worst case strategy performs the best/worst. In link-based Scenario A, both for $ATTR$ and E_G, the greedy recovery and the worst case recovery loosely bound the random recovery envelop, because there are much realizations, while envelopes generated by simulation cannot cover all these realizations. The greedy recovery and the worst case recovery tightly bound the random recovery envelop because the number of realizations in energy-based Scenario B is limited.

6.4.3. OVERVIEW OF THE LINK RATIO AND THE ENERGY RATIO

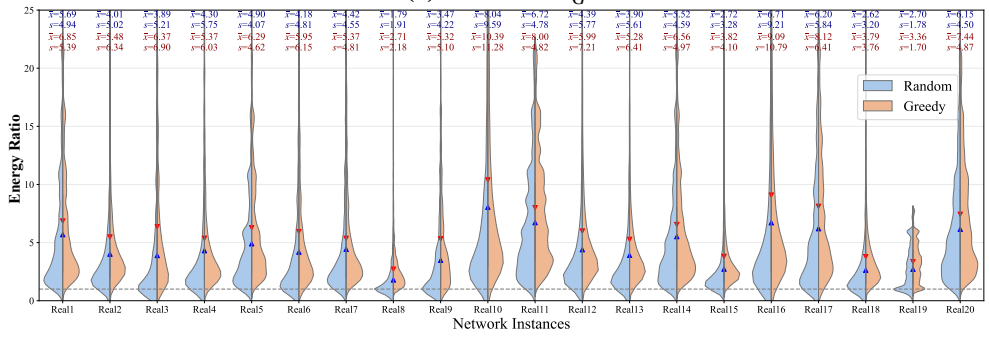
We employ the proposed approach and the recoverability indicators η (including the Link Ratio η_L and the Energy Ratio η_E) to evaluate the 20 real-world optical networks. Figure 6.5 shows the recoverability indicators under two different scenarios, two robustness metrics and two recovery strategies for the 20 considered networks by violin plots. Violin plots are similar to box plots, except that they show the probability density of the ratios η at different values, which presents more insights about the ratios η under random circumstances. Moreover, violin plots can be applied to compare the performance of any two different strategies, in this case the random and the greedy strategy.



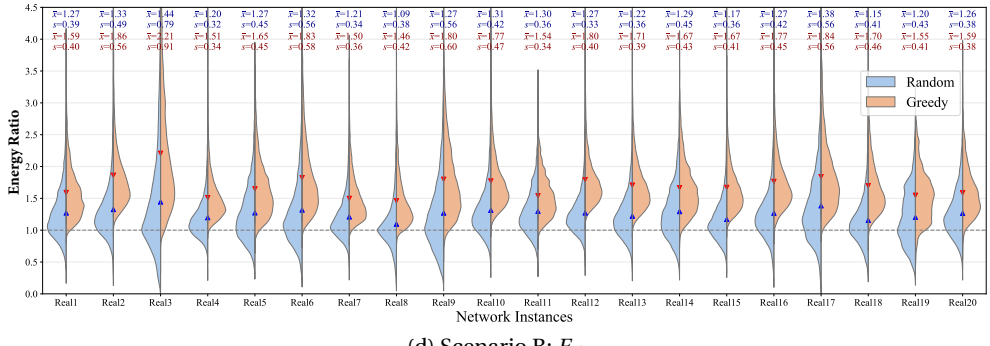
(a) Scenario A: $ATTR$



(b) Scenario A: E_G



(c) Scenario B: $ATTR$



(d) Scenario B: E_G

Figure 6.5: Violin plots of the Link Ratio η_L in Scenario A and the Energy Ratio η_E in Scenario B. The average ratios $\tilde{x} = E[\eta]$ and the standard deviations $s = \sqrt{Var[\eta]}$ are presented on the top of each subplot. The blue surface and values represent the random recovery strategy, and the red surface and values represent the greedy recovery strategy. The average ratios are marked as triangle markers. Each histogram of η is based on 10^4 realizations. For convenience, we use Real1 to Real20 to represent the 20 optical networks in table 6.1.

Figure 6.5 shows that almost all histograms of the ratio η, regardless of the scenarios, the strategies and the metrics, exhibit heavy-tailed distributions, while the greedy strategy presents a heavier tail when compared with random recovery strategy. Also, the ratio η has a wider range of values under the greedy strategy, which implies the greedy strategy has a higher probability to lead to a large ratio η, as well as a better recovery performance.

For both robustness metrics in Scenario A, Real7 (PionierL1) and Real8 (RoEduNet) have an average Link Ratio $E[\eta_L] < 1$ for the random strategy, which implies a relatively low recovery capability. By contrast, Real10 (US_Signal), Real16 (Palmetto) and Real17 (Sunet) have a large average Link Ratio $E[\eta_L] > 1$, which clearly outperform other networks, both for the random strategy and the greedy strategy.

The Energy Ratio η_E exhibits other behaviors than the Link Ratio η_L in Scenario A. The average Energy Ratios $E[\eta_E]$ for the robustness metric $ATTR$ are much larger than 1 under the random strategy, which can be explained by the fact that the function $R[K_{avg}]$ decreases slowly during the initial stage of the failure process but increases fast during the initial recovery process (illustrated in Section 6.4.1). Thus, the energy S_r is much larger than S_f, i.e., the average Energy Ratios $E[\eta_E]$ is much larger than 1 for $ATTR$. Since the function $R[K_{avg}]$ is concave for the robustness metric E_G and thus the energy $S_f < S_r$, the average Energy Ratios $E[\eta_E]$ for different networks are slightly larger than 1. The average Energy Ratio $E[\eta_E]$ in Scenario B under the greedy strategy is usually located in the tail of the distribution of the Link Ratio η_L under the random strategy, which demonstrates that the greedy strategy can increase the recoverability of networks significantly.

6.4.4. RELATION BETWEEN SCENARIO A AND SCENARIO B

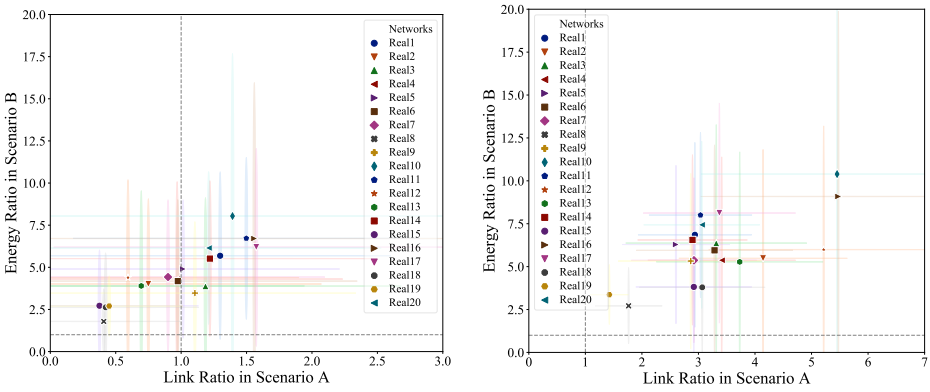
To compare the recoverability between different networks, we employ so-called Scenario A-Scenario B plots, which show the Energy Ratio vs. the Link Ratio, under a given recovery strategy. Scenario A-Scenario B plots are divided into 4 quadrants, by the reference lines $\eta_L = 1$ and $\eta_E = 1$, in order to easily assess the recoverability by the location of the average ratios $E[\eta_L]$ and $E[\eta_E]$. Figure 6.6 shows the average ratios $E[\eta]$ and the standard deviations $\sqrt{Var[\eta]}$ for the real-world networks in Scenario A-Scenario B plots.

Figure 6.6(a) and Figure 6.6(b) show that when the R-value is the average two-terminal reliability $ATTR$, the two recoverability ratios corresponding to two different scenarios have a positive correlation, e.g., a higher Link Ratio η_L in Scenario A typically leads to a higher Energy Ratio η_E in Scenario B, both for random recovery and greedy recovery.

Compared with Figure 6.6(a) and Figure 6.6(b), Figure 6.6(c) and Figure 6.6(d) show that when adopting the network efficiency as the R-value, the two recoverability ratios have a weak correlation, e.g., a higher Link Ratio η_L in Scenario A typically does not lead to a higher Energy Ratio η_E in Scenario B both for random recovery and greedy recovery. This implies that the R-value influences the correlation between Scenario A and Scenario B.

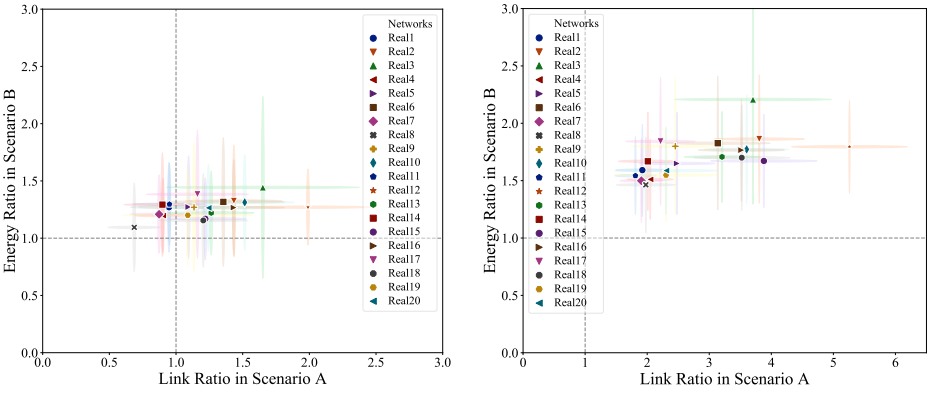
Figure 6.6 shows that all the average Energy Ratios $E[\eta_E]$ are located in the first and the second quadrant, which demonstrates a good recoverability of tested networks in Scenario B. However, for the random recovery, the average Link Ratios $E[\eta_L]$ of some networks are in the second quadrant, which suggests these networks have low recoverability in Scenario A.

Both the average Link Ratio $E[\eta_L]$ and the Energy Ratio $E[\eta_E]$ can be increased by



(a) Random recovery: $ATTR$

(b) Greedy recovery: $ATTR$



(c) Random recovery: E_G

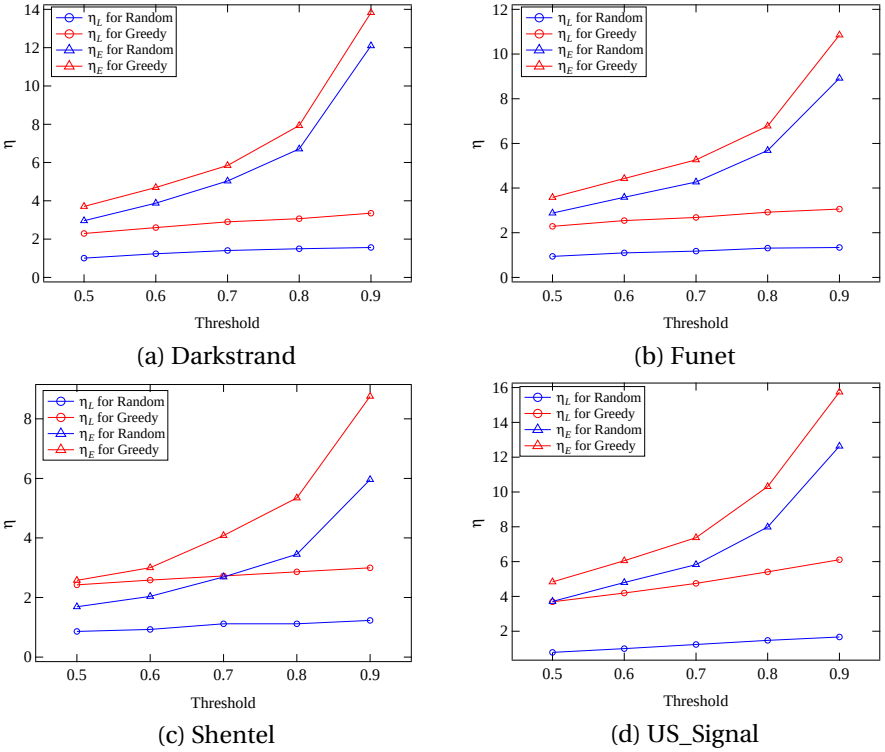
(d) Greedy recovery: E_G

Figure 6.6: Scenario A-Scenario B plots of the Link Ratio η_L and the Energy Ratio η_E for two robustness metrics (i.e., the average two-terminal reliability $ATTR$ and the network efficiency E_G) based on 20 optical networks. The solid markers represent the average ratios $E[\eta]$, and the crosses indicate the value ranges $[E[\eta] - \sqrt{Var[\eta]}, E[\eta] + \sqrt{Var[\eta]}]$.

applying the greedy strategy, but the performance can be different. For example, the average Link Ratio $E[\eta_L]$ of network Real14 (NetworkUSA) is smaller than that of network Real11(Darkstrand) under the random strategy but larger than that of network Real11 under the greedy strategy, which implies that the performance of a recovery strategy strongly depends on the network topology.

6.5. SENSITIVITY ANALYSIS OF NETWORK RECOVERABILITY

In previous sections, the R-threshold was fixed at the vaulue $\rho = 0.8$. In this section we investigate the influence of different R-thresholds on the Link Ratio η_L and the Energy Ratio η_E. Figure 6.7 and Figure 6.8 show the impact of different R-thresholds on recoverability indicators η for 4 optical networks, for the average two-terminal reliability $ATTR$ and the network efficiency E_G, respectively.



(a) Darkstrand (b) Funet

(c) Shentel (d) US_Signal

Figure 6.7: The impact of thresholds on recoverability indicators for the average two-terminal reliability $ATTR$ in 4 optical networks.

We conclude from Figure 6.7 that when the R-value is $ATTR$ the following: 1) a larger R-threshold dramatically increases the average Energy Ratio $E[\eta_E]$ in Scenario B, emphasizing the importance of diagnosing and recovering the network in the early period. 2) The average Link Ratio $E[\eta_L]$ in Scenario A increases slightly with a larger R-threshold (i.e., a lower damage level). Thus, the average Energy Ratio $E[\eta_E]$ in Scenario B is more sensitive than the average Link Ratio $E[\eta_L]$ in Scenario A. 3) The increase of the two recoverability ratios, especially for the average Energy Ratio $E[\eta_E]$ in Scenario B, can be explained by the curvature of the function $R[K_{avg}]$ in the random failure process. As illustrated in Figure 6.3(a) and Figure 6.3(c), the function $R[K_{avg}]$ is approximatively concave when the average number of challenges is small (corresponding to a high R-threshold). As the number of challenges increases in order to degrade the R-value to a lower R-threshold, the function $R[K_{avg}]$ gradually becomes more convex, which is in line with the results obtained in [108]. Thus, the Energy Ratio η_E, which equals the energy of recovery challenges S_r divided by the energy of failure challenges S_f, tends to become larger as the R-threshold increases.

Figure 6.8 shows that when the R-value is the network efficiency E_G we can conclude the following: 1) the average Energy Ratio $E[\eta_E]$ and the average Link Ratio $E[\eta_L]$ are not

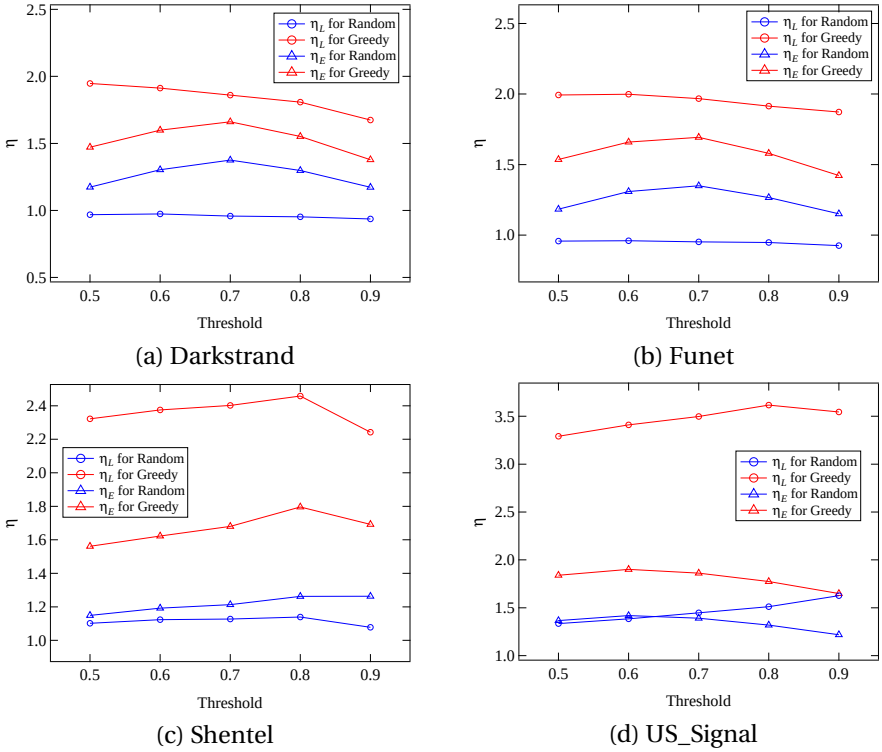


Figure 6.8: The impact of thresholds on recoverability indicators for network efficiency E_G in 4 optical networks.

always monotonically changing as the R-threshold increases. Specifically, for networks Darkstrand and Funet, the average Link Ratio $E[\eta_L]$ for the greedy recovery is slightly decreasing with a higher R-threshold, while for networks Shentel and US_Signal, the average Link Ratio $E[\eta_L]$ is increasing when the R-threshold increases from 0.5 to 0.8. Nevertheless, the average Energy Ratio $E[\eta_E]$ first increases and then decreases with the increment of the R-threshold, both for random recovery and greedy recovery, which may imply an optimal R-threshold for Scenario B exists. 2) Compared with Figure 6.7, the average Energy Ratio $E[\eta_E]$ in Scenario B for network efficiency is less sensitive than that for $ATTR$. This reveals that the sensitivity of recoverability indicators largely depends on the choice of the R-value. 3) For both average two-terminal reliability $ATTR$ and network efficiency E_G, the greedy recovery exhibits a better performance than random recovery, for different R-thresholds. Thus, we propose to use the greedy recovery strategy.

6.6. CORRELATION OF METRICS WITH RECOVERABILITY INDICATORS

In this section, we explore the correlation between recoverability indicators in the random recovery scenario and 10 widely studied network metrics: the average degree $E[D]$, the

spectral radius λ_1, the diameter φ, the algebraic connectivity μ_{N-1}, the assortativity ρ_D, the average hopcount $E[H]$, the clustering coefficient c_G, the ratio μ_1/μ_{N-1}, the effective graph resistance r_G and the global efficiency $E[1/H]$. Results are shown in Table 6.2 and Table 6.3, which are based on 200 optical backbone communication networks in the specialized database [132].

We use the Spearman's rank correlation coefficient ρ_s [133] to evaluate the correlation between the recoverability indicators and the 10 network metrics. The Spearman's rank correlation coefficient ρ_s is less restrictive than the Pearson's correlation coefficient ρ_p since the latter only estimates the linear correlation between two variables. The Spearman's rank correlation coefficient ρ_s measures the strength and direction of monotonic association between two variables X and Y, i.e.,

$$\begin{aligned}\rho_s(X,Y) &= \rho_p(F_X(X), F_Y(Y)) \\ &= \frac{\mathrm{Cov}[F_X(X), F_Y(Y)]}{\sigma_{F_X(X)}\sigma_{F_Y(Y)}},\end{aligned} \tag{6.12}$$

where $F_X(X)$ and $F_Y(Y)$ are the probability distribution of the variable X and Y, respectively. $\rho_p(F_X(X), F_Y(Y))$ is the Pearson's correlation coefficient between $F_X(X)$ and $F_Y(Y)$.

Table 6.2 illustrates the Spearman's rank correlation coefficient ρ_s between the 10 network metrics and the two recoverability indicators, when the R-value is the average two-terminal reliability $ATTR$. As shown in Table 6.2, assortativity ρ_D has the strongest positive correlation with both the average Link Ratio $E[\eta_L]$ and the average Energy Ratio $E[\eta_E]$. Out of the 200 optical networks, 175 networks have negative assortativity, which suggests that a negative assortativity value close to 0 corresponds to a large average Link Ratio $E[\eta_L]$ and Energy Ratio $E[\eta_E]$. The average hopcount $E[H]$ has the weakest correlation with the average Link Ratio $E[\eta_L]$, while the algebraic connectivity μ_{N-1} has the weakest correlation with the average Energy Ratio $E[\eta_E]$. In addition, the effective graph resistance r_G has a relatively strong negative correlation for the average Link Ratio $E[\eta_L]$.

Table 6.3 illustrates the Spearman's rank correlation coefficient ρ_s between the 10 network metrics and the two recoverability indicators, where the R-value is the network efficiency E_G. Assortativity ρ_D still has the strongest positive correlation with both the average Link Ratio $E[\eta_L]$ and the average Energy Ratio $E[\eta_E]$. Since the assortativity ρ_D of most backbone networks (175 out of 200) is negative, this finding suggests that optical networks with an assortativity value closer to 0 has a higher recoverability for random recovery. The average degree $E[D]$ also has a relatively strong correlation with the average Energy Ratio $E[\eta_E]$, suggesting denser network may have a better recoverability. Furthermore, the effective graph resistance r_G has the weakest correlation with the average Link Ratio $E[\eta_L]$, while the algebraic connectivity μ_{N-1} still has the weakest correlation with the average Energy Ratio $E[\eta_E]$.

6.7. CONCLUSION

This chapter proposes a topological approach for evaluating the network recoverability in two scenarios, the link-based Scenario A and the energy-based Scenario B. We found that

Table 6.2: The Spearman's rank correlation coefficient ρ_s between 10 network metrics and the two recoverability indicators. The R-value considered here is the average two-terminal reliability $ATTR$. Results are based on 200 real-world optical networks.

Metrics	ρ_s for $E[\eta_L]$	ρ_s for $E[\eta_E]$
Average degree $E[D]$	0.5119	0.4784
Spectral radius λ_1	-0.4045	-0.4223
Diameter φ	0.1534	0.3239
Algebraic connectivity μ_{N-1}	0.1580	-0.0168
Assortativity ρ_D	**0.5460**	**0.5912**
Average hopcount $E[H]$	0.0353	0.2326
Clustering coefficient c_G	0.3534	0.2616
Ratio μ_1/μ_{N-1}	-0.4783	-0.2831
Effective graph resistance r_G	-0.5246	-0.2766
Global efficiency $E[1/H]$	0.1552	-0.0764

Table 6.3: The Spearman's rank correlation coefficient ρ_s between 10 metrics and two recoverability indicators. The R-value here is network efficiency E_G. Results are based on 200 real-world optical networks.

Metrics	ρ_s for $E[\eta_L]$	ρ_s for $E[\eta_E]$
Average degree $E[D]$	0.4833	0.5380
Spectral radius λ_1	-0.3787	-0.3185
Diameter φ	0.6297	0.2869
Algebraic connectivity μ_{N-1}	-0.3773	-0.0264
Assortativity ρ_D	**0.6677**	**0.5708**
Average hopcount $E[H]$	0.5555	0.1930
Clustering coefficient c_G	0.2665	0.3593
Ratio μ_1/μ_{N-1}	0.1181	-0.2190
Effective graph resistance r_G	0.1006	-0.2518
Global efficiency $E[1/H]$	-0.4178	-0.0433

all the optical networks have a healthy recovery capability in Scenario B under the random recovery strategy, i.e., the average Energy Ratio $E[\eta_E] > 1$, while two of the networks (PionierL1 and RoEduNet) suggest topological improvements for the recoverability in Scenario A, i.e., the average Link Ratio $E[\eta_L] < 1$. The performance of the recoverability in Scenario B can be explained by the concavity of the R-value as a function of the number of challenges. There is also a strong correlation between the network recoverability and the recovery strategy. The greedy recovery strategy exhibits a good performance for the investigated robustness metrics and thus improves the network recoverability. The network efficiency is less sensitive to different R-value thresholds while the Energy Ratio $E[\eta_E]$ for the average two-terminal reliability increases significantly with increasing thresholds in Scenario B. The assortativity has the strongest correlation with the average Link Ratio and the average Energy Ratio, when the robustness metric is either the average two-terminal reliability or the network efficiency.

7

THE RECOVERABILITY OF NETWORK CONTROLLABILITY

In this chapter, we adopted the framework of network recoverability and investigate the recoverability of network controllability for two recovery scenarios: 1) only the links which are damaged in the failure process can be recovered and 2) links can be established between any pair of nodes that have no link between them after the failure process. By applying the normalized value of network controllability as the robustness metric, we employ the proposed approach to assess swarm signalling networks with regular out-degree, and networks with bi-modal out-degree distributions. Furthermore, we also deduced the analytical results of the recoverability indicators by generating functions, which are close to the results based on simulations.

7.1. INTRODUCTION

The secure, reliable and effective operation of critical infrastructures such as power grids, telecommunications and the Internet relies on the ability to control the state of a given system or network. Network controllability offers a graph theoretical interpretation for control systems as first described by Kalman, which is particularly suitable for studying sets of nodes offering the ability to control an entire network. Network controllability has been a hot research topic in recent years [39] [40] [41] [42]. A system is considered controllable if it can be driven from any initial state to any desired final state by external inputs in finite time [24]. Let the $N \times N$ matrix A represent the wiring diagram of a network with N nodes, while the connection of M input signals to the network is described by the $N \times M$ input matrix B, where $M \leq N$. The nodes injected by the input signals are called driver nodes, which steer the state of the network. Then, the system characterized by (A, B) is structurally controllable, if it is possible to find the non-zero parameters in A and B in such a way that the obtained system (A, B) is controllable in the classical sense of satisfying Kalman's rank condition.

The robustness of the network controllability can be measured by quantifying the

increase in the minimum number of driver nodes N_D, under perturbation of the network topology. The impact of topological perturbations on the controllability of networks has been investigated extensively in recent years. Pu *et al.* [54] found that the degree-based node attack is more efficient than a random attack for degrading the controllability in directed random and scale-free networks. Nie *et al.* [55] found that the controllability of Erdős-Rényi random graphs with a moderate average degree is less robust, whereas a scale-free network with moderate power-law exponent shows a stronger ability to maintain its controllability, when these networks are under intentional link attack. Thomas *et al.* [56] identified that the potency of a degree-based attack is directly related (on average) to the betweenness centrality of the edges being removed. Lu *et al.* [57] discovered that a betweenness-based strategy is quite efficient to harm the controllability of real-world networks. Mengiste *et al.* [58] introduced a new graph descriptor, 'the cardinality curve', to quantify the robustness of the control structure of a network to progressive link pruning. In Chapter 2, we proposed closed-form analytic approximations for the number of driver nodes that are needed to maintain network controllability, where links are removed according to both random and targeted attacks [36].

There is also some research concerning the recovery of controllability in networks. Alcaraz *et al.* [134] investigated algorithms for the efficient restoration of controllability following attacks and attacker-defender interactions in power-law networks. Results highlighted that the use of a network diameter can be a suitable option to establish control with low computational and storage costs. In [135], four reachability-based restoration strategies were presented to find optimal solutions that guarantee control at all times and without damaging the structural controllability properties. Zhang *et al.* [136] proposed a maximum matching-based method to recover the controllability of random digraphs in linear time.

The work mentioned above either focus on the robustness of the network controllability under perturbations or specific methods to restore network controllability. However, the recovery process after failures is not considered and the investigation on the capability of a network to recover its controllability from failures is still lacking. We define such network capability as *recoverability of network controllability* in this chapter. Recovery measures are taken in order to recover the controllability of a network after the failure process, either by restoring the damaged links or by building new links. Considering the network performance is related to many factors, such as topology, recovery strategy, link adding sequence [101], we need an approach to measure the recoverability of network controllability.

Based on our previous work [95], we propose a topological approach and define two recoverability indicators to quantify the recoverability of controllability for two different recovery scenarios denoted as Scenario A and Scenario B. For the link-based Scenario A, links can be established between any pair of nodes that have no link between them, after the failure process. The energy-based Scenario B assumes that only the links which are damaged in the failure process can be recovered. Our approach is tested on swarm signalling networks with regular out-degree, networks with bi-modal out-degree distributions as well as some real-world networks. We verify that the proposed recoverability indicators allow us to assess the recoverability of controllability in different networks. For some networks with specific degree distribution, such as regular out-degree and bi-modal

out-degree, we manage to deduce the analytical approximations of the recoverability indicators and measure their accuracy using the simulation results as the benchmark.

The rest of this chapter is organized as follows: Section 7.2 provides the estimations for the recovery of network controllability in Scenario A and Scenario B. Section 7.3 concludes this chapter.

7.2. RECOVERABILITY OF NETWORK CONTROLLABILITY

In Chapter 3, we proposed analytical expressions to determine the minimum fraction n_D of driver nodes during the random failure process. However, it is natural and common in real life to consider recovering a network after failures occur in the network. In this section, we adopt the framework of investigating the recoverability of networks introduced in Chapter 6 and investigate the recoverability of network controllability.

7.2.1. R-VALUE

As discussed in Chapter 6, the robustness of a network can be expressed in a mathematical way, through the so-called R-value, which quantifies the robustness of a network [75]. To normalize the value of n_D as the R-value whose value is between 0 and 1, we define the R-value as:

$$R = \frac{1 - n_D}{1 - n_{D_0}} \tag{7.1}$$

where n_{D_0} is the fraction of driver nodes in the original network, n_D is the fraction of driver nodes during the attack phase and recovery phase. When n_D is equal to n_{D_0}, R equals 1, which reflects the network's controllability does not change. When the R-value equals 0, it means that the network controllability is completely destroyed, and all nodes need to be controlled ($n_D = 1$) to control the whole network.

7.2.2. RECOVERY IN SCENARIO A

As discussed above, the R-value is the controllability metric of a network $G(N, L)$. Attacking this network would make its minimum fraction n_D of driver nodes increase, which is shown in Chapter 2. Thus, the R-value decreases, which denotes the degradation of network controllability. The links are removed one by one until the R-value reaches a predefined threshold $R_{threshold}$. The number of removed links that makes the R-value reach the predefined threshold is denoted as K_f. Then the recovery process starts from the remaining network $G_{attacked}(N, L - K_f)$. Scenario A assumes that the recovered links can be added between any two nodes in the complement of the graph after attacks if the elementary challenges are link-based removals and additions.

According to [133], we deduce the degree distribution for randomly removing a fraction p of links in the attack phase, where $p = i/L$, i is the number of removed links and L is the initial number of links in the network. By adopting the deduced degree distribution, we construct the generating function for the resulting network. Given the generating function $G(x)$ for the initial network, the generating function $\bar{G}(x)$ for the resulting network after removing a fraction of p links satisfies:

$$\bar{G}(x) = G(p + (1 - p)x). \tag{7.2}$$

In the resulting network after removing a fraction p of links, a fraction f of links are randomly recovered, where $f = \frac{K}{N(N-1)-(1-p)L}$, K is the number of recovered links, $N(N-1)-(1-p)L$ is the number of all possible links to recover. Then we can deduce that the generating function for the final network follows:

$$\begin{aligned}\hat{G}(x) &= (1-f(1-x))^{N-1} * \bar{G}(\frac{x}{1-f(1-x)}) \\ &= (1-f(1-x))^{N-1} * G(p+(1-p)\frac{x}{1-f(1-x)}).\end{aligned} \tag{7.3}$$

The proof of Eq.(7.3) is given in Appendix C. Thus, we obtain the generating functions for the random attack process and the random recovery process in Scenario A:

$$\begin{cases} \text{Attack process: } \bar{G}(x) = G(p+(1-p)x), \\ \\ \text{Recovery process:} \\ \hat{G}(x) = (1-f(1-x))^{N-1} \cdot \bar{G}(\frac{x}{1-f(1-x)}), \end{cases} \tag{7.4}$$

By applying Eq.(7.4) to the general formula Eq.(3.10), we can approximate the fraction n_D of driver nodes and the corresponding R-values in the random attack and recovery process.

For swarm signalling networks (SSNs), when a fraction p of links is randomly attacked, the approximation for the fraction n_D of driver nodes follows Eq.(3.39). When a fraction f of links is randomly recovered in the network after attack, the approximation for the fraction n_D of driver nodes in SSNs with regular out-degree k follows:

$$n_D = \omega_1(1-\hat{\omega}_2)\cdot(k(1-p^*)+f(N-1-k(1-p^*)))+G_{in}(1-\omega_1)-1+G_{out}(\hat{\omega}_2), \tag{7.5}$$

where $p^* = K_f/L$,

$$G_{out}(x) = (1-f(1-x))^{N-1}\cdot(p^*+(1-p^*)\frac{x}{1-f(1-x)})^k, \tag{7.6}$$

$$G_{in}(x) = (1-f(1-x))^{N-1}\cdot e^{-k(1-p^*)(1-\frac{x}{1-f(1-x)})}. \tag{7.7}$$

ω_1 and $\hat{\omega}_2$ can be calculated by solving Eq.(3.11) and Eq.(3.14).

For real-world networks, the original generating function of the degree distribution satisfies:

$$G(x) = \frac{\mathcal{N}(0)+\mathcal{N}(1)\cdot x+\cdots+\mathcal{N}(n-1)\cdot x^{n-1}}{N} \tag{7.8}$$

where N is the total number of nodes in the network, $\mathcal{N}(m)$ is the number of nodes whose degree equals m.

In our simulation, the R-threshold is set to 0.9. For each $SSN(N,k)$ with regular out-degree distribution or $SSN(N,k_1,k_2,\alpha)$ with bi-modal out-degree distribution, we generate 100 corresponding swarm signalling networks. For each network, we repeat the random attack and random recovery process for 500 times. Thus, the curve for the random attack and the random recovery in Scenario A is based on the average value of 50000 realizations.

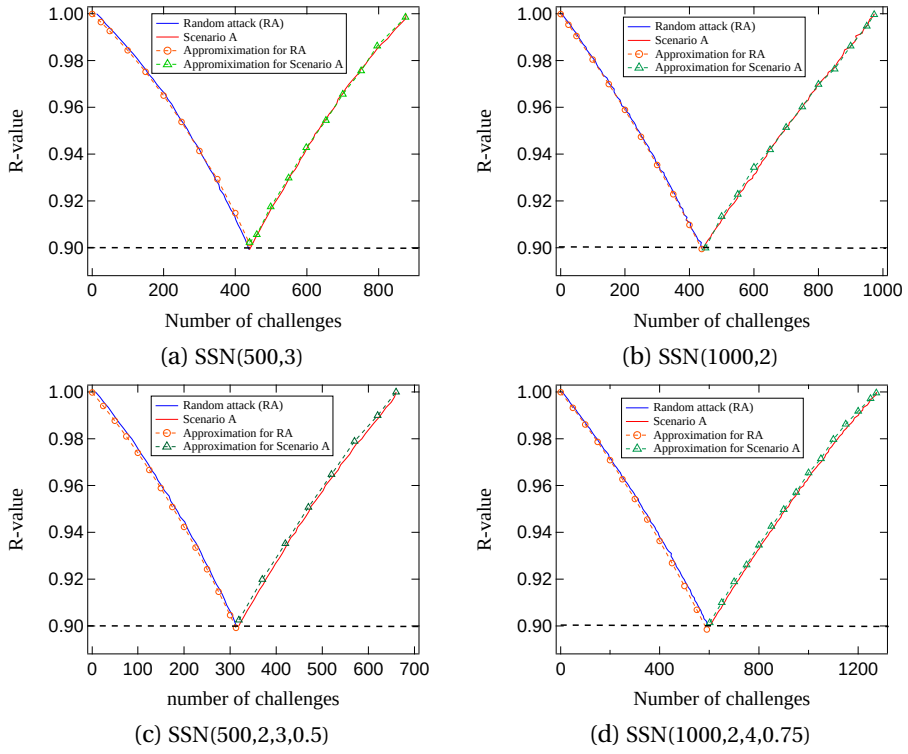


Figure 7.1: The impact of random attack and random recovery on R-values in Scenario A. In each sub-figure, we generate 100 corresponding swarm signalling networks. For each network, we repeat the random attack and random recovery process for 500 times. Thus, the curve for the random attack and Scenario A is based on the average value of 50000 realizations.

Figure 7.1 shows the impact of random attack and random recovery on R-values in Scenario A. In each sub-figure, the solid lines are the average simulation results which consist of R-values of the expected number of challenges $R[K_{avg}]$. Specifically the solid blue line denotes the decrease of the R-value in the random attack process while the solid red line shows the increase of the R-value in the random recovery process in Scenario A. As shown in each sub-figure, the R-value decreases slowly during the initial stage of the random attack process but increases fast during the initial recovery process.

We also calculate the R-values analytically in the attack and recovery process and compare with simulation results in Figure 7.1. To get the analytical R-values for a chosen number of challenges k, we first get the value $p = k/L$ if k belongs to the random attack process. Then, we obtain the generating function provided by Eq.(7.2). By applying Eq.(7.2) into Eq.(3.10), we get the fraction n_D of driver nodes. Finally, the R-value equals $\frac{1-n_D}{1-n_{D0}}$. When k belongs to the random recovery process in Scenario A, we adopt the generating functions Eq.(7.3) and then follow the same methodology to get the R-value. As shown in Figure 7.1, the analytical approximations for the R-values fit well with the simulation results both for the random attack and the random recovery in Scenario A.

The results show that our analytical method has a high accuracy for calculating network controllability in random attack and random recovery process for swarm signalling networks.

The top two sub-figures in Figure 7.2 exemplify the envelopes of the challenges in SSN for the controllability metric R-value in Scenario A, under the random attack and recovery strategy. The approximation fits very well with the simulation, which indicates again that the general formula Eq.(7.4) works well. As shown in the bottom two sub-figures of Figure 7.2, our approximation also fits well with the simulation results in real-world networks. We notice that our analytical approximations for network controllability perform better for *kdl* than *Cogentco*, as the method is based on statistical physics and performs better for large networks.

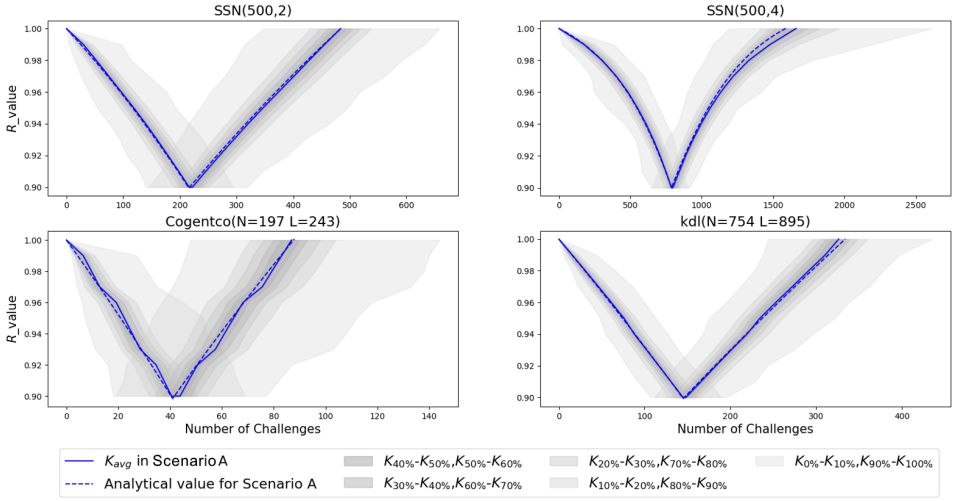


Figure 7.2: Envelopes of the challenges for SSNs with 500 nodes and different average out-degree ($k_{out} = 2$ and $k_{out} = 4$) and two real-world networks (*Cogentco* and *kdl*) in Scenario A, by random attack and random recovery strategy. The threshold of the R-value is 0.9. Each envelope is based on 10^4 realizations.

7.2.3. RECOVERY IN SCENARIO B

The attack process in Scenario B is the same as in Scenario A. In the recovery process in Scenario B, all the links that are removed in the attack process are randomly added until the network returns to the original state under the link-based recovery. A symmetric method is used in Scenario B to express the generating function in the recovery process. By using the same notation as before, $\bar{G}(x)$ and $\hat{G}(x)$ refer to the generating functions in the attack process and the subsequent recovery process, respectively.

$$\begin{cases} \text{Attack process: } \bar{G}(x) = G(p + (1-p)x), \\ \\ \text{Recovery process: } \bar{\hat{G}}(x) = G(p^* + (1-p^*)x) \end{cases} \tag{7.9}$$

In the link-based recovery process, $p^* = \frac{2K_f - i}{L}$, where K_f is the number of removed

links that makes the R-value reach the R-threshold in the attack process, i is the number of challenges which is between K_f and $2K_f$. After applying Eq.(7.9) to Eq.(3.10), we can approximate the fraction n_D of driver nodes and the corresponding R-values for Scenario B.

When a fraction p of links is randomly attacked, the approximation for the fraction n_D of driver nodes in SSNs still follows Eq.(3.39). When the attacked links are randomly recovered, the approximation for the fraction n_D of driver nodes in SSNs with regular out-degree k follows:

$$n_D = G_{in}(1-\omega_1) - 1 + G_{out}(\hat{\omega}_2) + k(1-p^*)\cdot\omega_1(1-\hat{\omega}_2), \tag{7.10}$$

where

$$G_{out}(x) = (p^* + (1-p^*)x)^k, \tag{7.11}$$

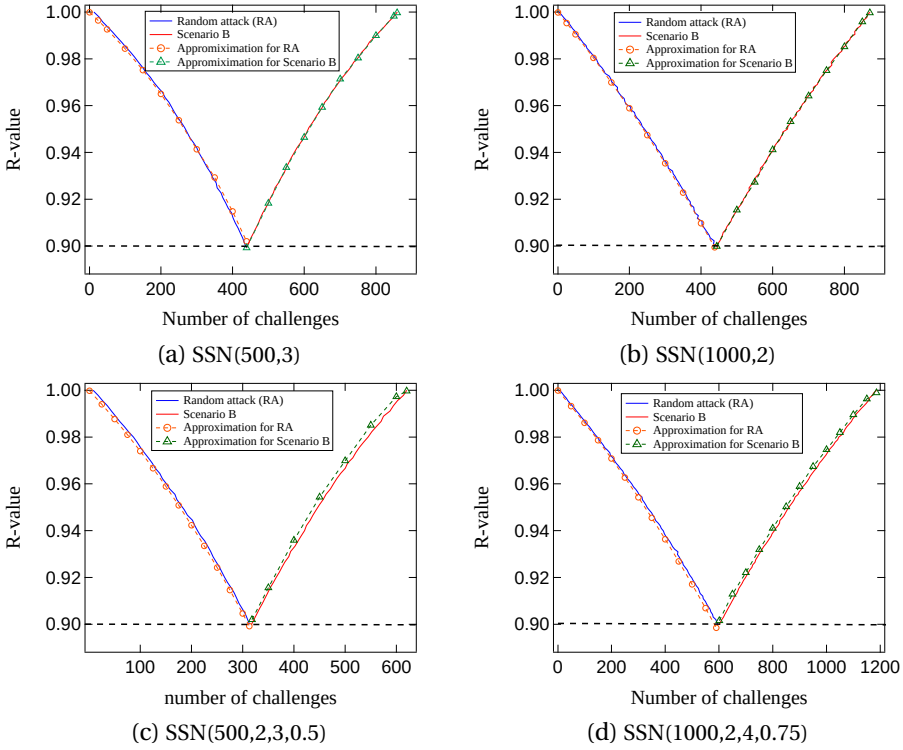
$$G_{in}(x) = e^{-k(1-p^*)(1-x)}. \tag{7.12}$$

Figure 7.3 shows the impact of random attack and random recovery on R-values in two types of swarm signalling networks. The failure process is the same as in Scenario A. However, we found that the average number of challenges needed to restore the R-value is less than that in Scenario A, which means that the efficiency of recovery in Scenario B is higher than in Scenario A. Similarly, we also calculate the R-values analytically in the attack and recovery process in Scenario B and compare with simulation results. As shown in Figure 7.3, the analytical approximations for the R-values fit well with the simulation results for the random recovery in Scenario B. The results indicate that our analytical method has a high accuracy for calculating network controllability in random attack and random recovery process for swarm signalling networks. Figure 7.4 illustrates that our method predicts the change of R-value well during the whole process, not only for SSNs, but also for real-world networks.

7.2.4. ESTIMATIONS FOR RECOVERABILITY INDICATORS

The method using the generating functions to calculate the fraction n_D of driver nodes can also estimate the recoverability indicators in Scenario A and Scenario B. To obtain the *Link Ratio* η_L which is the ratio of the number of failure challenges K_f and the recovery challenges K_r, we need to calculate the value of K_f and K_r, respectively. Since the R-threshold ρ is given, the fraction p of removed links is the numerical solution of Eq.(3.10) after applying Eq.(7.2) into Eq.(3.10). Then we get $K_f = p * L$. Similarly, we can also get the fraction f of added links in the recovery process by solving Eq.(3.10) and calculate the value $K_r = f * (N(N-1) - (1-p)L)$ for Scenario A. For Scenario B where the *Energy Ratio* η_E is the ratio between the energy of the recovery challenges S_r and the energy of the failure challenges S_f, we need to calculate the value of S_f and S_r, respectively. S_r equals the integral of the R-value in the interval $[K_f, 2K_f]$ minus $\rho * K_f$ while S_f equals K_f minus the integral of the R-value in the interval $[0, K_f]$.

Table 7.1 illustrates the estimations for recoverability indicators and the absolute relative errors between the estimations and simulation results. As shown in Table 7.1, all estimations for the average link ratio $E[\eta_L]$ fit well with simulation results with small



(a) SSN(500,3) (b) SSN(1000,2)

(c) SSN(500,2,3,0.5) (d) SSN(1000,2,4,0.75)

Figure 7.3: The impact of random attack and random recovery on R-values in Scenario B. In each sub-figure, we generate 100 corresponding swarm signalling networks. For each network, we repeat the random attack and random recovery process for 500 times. Thus, the curve for the random attack and Scenario A is based on the average value of 50000 realizations.

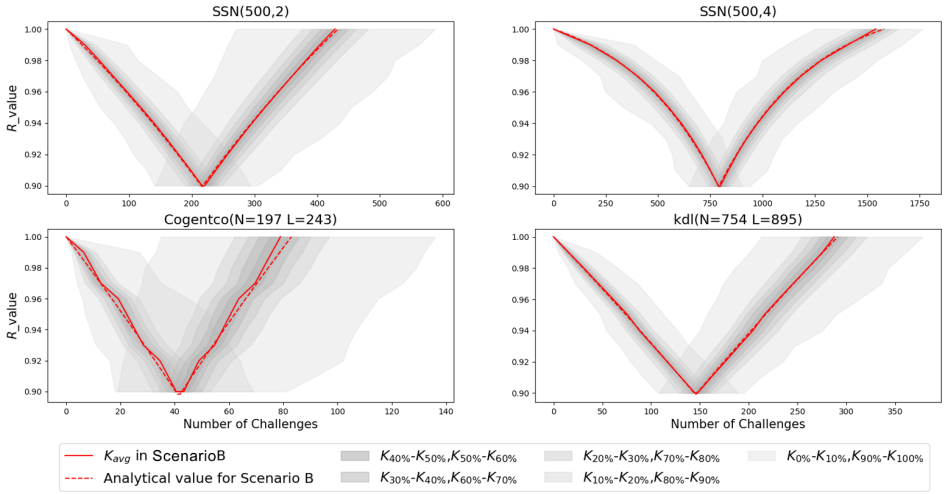


Figure 7.4: Envelopes of the challenges for SSNs with 500 nodes and different average out-degree ($k_{out} = 2$ and $k_{out} = 4$) and two real networks (*Cogentco* and *kdl*) in Scenario B, by random attack and random recovery strategy. The threshold of the R-value is 0.9. Each envelope is based on 10^4 realizations.

absolute relative errors, which further indicates that our analytical method has a high accuracy for calculating network controllability for scenario A. Furthermore, we notice that for the networks with the same size, a higher average out-degree indicates a larger link ratio η_L. However, the average energy ratio $E[\eta_E]$ does not have this trait. Besides, when the network size is large, our estimation for the average energy ratio $E[\eta_E]$ has relatively large relative error compared with simulation results.

Table 7.1: Estimations for the recoverability indicators

Networks	$E[\eta_L]$	Estimation	$\|RE\|$	$E[\eta_E]$	Estimation	$\|RE\|$
SSN(500,2)	0.8526	0.8247	3.27%	1.2423	1.3678	9.68%
SSN(500,3)	0.9436	0.9382	0.57%	1.5802	1.6932	6.96%
SSN(500,4)	0.9736	0.9682	0.55%	1.6988	1.8714	10.16%
SSN(1000,2)	0.7615	0.7865	3.28%	1.4107	1.3379	5.16%
SSN(1000,3)	0.8352	0.8679	3.92%	1.6174	1.7998	11.28%
SSN(1000,4)	0.9678	0.9831	1.58%	1.8277	1.9654	7.53%
SSN(500,2,3,0.5)	0.8734	0.8807	0.83%	1.4920	1.2516	16.11%
SSN(500,2,4,0.25)	0.9613	0.9842	2.38%	1.8763	1.6482	12.16%
SSN(500,3,4,0.75)	0.9567	0.9809	2.53%	1.7529	1.8374	4.82%
SSN(1000,2,4,0.25)	0.8812	0.9206	4.47%	1.6806	1.9413	15.51%
SSN(1000,2,4,0.5)	0.8256	0.8514	3.13%	1.4673	1.7540	19.53%
SSN(1000,2,4,0.75)	0.8073	0.8398	4.03%	1.6278	1.8334	12.63%

7.3. CONCLUSION

This chapter applies the framework of network recoverability to evaluate the recoverability of network controllability in two scenarios, the link-based Scenario A and the energy-based Scenario B. We assess the recoverability of two types of swarm signalling networks and real-world networks. Results show that swarm signalling networks have low recoverability in Scenario A but have high recoverability in Scenario B. Moreover, we propose an analytical method to estimate the fraction of driver nodes in face of random attack and random recovery, which fits well with simulation results. Furthermore, it is convenient to estimate the values of recoverability indicators with high accuracy, by using the analytical method without setting up simulations.

8

CONCLUSION

You only live once, but if you do it right, once is enough.

Mae West

8.1. MAIN CONTRIBUTIONS

This thesis provides original methods and new insights into the investigation on network resilience, encompassing the robustness of controllability and the recoverability of real-world networks. We are devoted to find analytical approximations to efficiently estimate the impact of topological perturbations on the performance of the network. Network topology, structural properties of the network, types of attack and recovery strategy, all need to be taken into account to investigate the network resilience better. The main contributions of each chapter are as follows:

In Chapter 2, we derived analytical closed-form approximations for the minimum number of driver nodes N_D needed to control networks, as a function of the fraction of removed links, both for random and targeted attacks. Both for random and targeted attacks, our approximation is linear in the fraction of removed links l, as long as this fraction is smaller than the fraction of critical links. For fractions of removed links larger than the fraction of critical links, our approximation is quadratic in l. We validated our approximation through simulations on real-world and synthetic networks. For random attacks, the approximation is always very good, as long as the fraction of removed links is smaller than the fraction of critical links. For some cases, the approximation is still accurate for larger fractions of removed links. The approximation for attacks targeting the critical links is also accurate, as long as the fraction of removed links is sufficiently small. The approximation for the targeted attack always serves as a worst-case estimate. Finally, we found that the critical link attack is the most effective among 4 considered attacks, as long as the fraction of removed links is smaller than the fraction of critical links.

In Chapter 3, we correct the formula given in [68] for the minimum number of driver nodes for a specific class of swarm signalling networks, which are characterised by a regular out-degree. We then generalize the results by considering SSNs with a regular out degree k where a fraction p of the links is unavailable. For this case we derive an implicit equation, whose solution leads to the minimum number of driver nodes. We find that our approximation fits well with simulation results. Finally, we relax the condition that the out-degree is regular and look into bi-modal out-degree distributions. For this case we also consider scenarios with unavailable links. We derive an implicit equation and verify its accuracy. We find that our approximation for bi-modal out-degree distribution fits well with simulation results.

In Chapter 4, we show that machine learning is applicable to improve the performance of the analytical method in measuring the robustness of network controllability. By using machine learning, we are able to further improve the accuracy of our approximations for the number of driver nodes. We also derive an analytical approximation for out-in degree-based attacks. Our machine learning based approximations outperform the analytical approximations in both synthetic and real-world networks.

In Chapter 5, we analyze the role of critical links in network controllablity. Simulation results on communication networks have suggested analytical closed-form approximations for the number N_c of controllable nodes. We derive closed-form approximations for the number N_c of controllable nodes as a function of the fraction of removed links, for random attacks, targeted attacks and random attack under protection. We validate our approximation through simulations on sparse communication networks and synthetic networks. Both for random attacks and random attacks under protection, our approximations for these two cases are always very good, as long as the fraction of removed links is smaller than the fraction of critical links. For some cases, the approximation is still accurate for larger fractions of removed links. For targeted attack, our approximation performs well as long as the fraction of removed links is sufficiently small, whereas our approximation does not perform well when the fraction of removed links is large. However, the approximation for the targeted attack always serves as a worst-case estimate.

In Chapter 6, we propose a topological approach for evaluating the network recoverability in two scenarios, the link-based Scenario A and the energy-based Scenario B. We found that all the optical networks have a healthy recovery capability in Scenario B under the random recovery strategy, i.e., the average Energy Ratio $E[\eta_E] > 1$. The performance of the recoverability in Scenario B can be explained by the concavity of the R-value as a function of the number of challenges. There is also a strong correlation between the network recoverability and the recovery strategy. The greedy recovery strategy exhibits a good performance for the investigated robustness metrics and thus improves the network recoverability. The assortativity has the strongest correlation with the average Link Ratio and the average Energy Ratio, when the robustness metric is either the average two-terminal reliability or the network efficiency.

Chapter 7 applies the framework of network recoverability to evaluate the recoverability of network controllability in two scenarios, the link-based Scenario A and the energy-based Scenario B. We assess the recoverability of two types of swarm signalling networks and real-world networks. Results show that swarm signalling networks have low recoverability in Scenario A but have high recoverability in Scenario B. Moreover, we

propose an analytical method to estimate the fraction of driver nodes in face of random attack and random recovery, which fits well with simulation results. Furthermore, it is convenient to estimate the values of recoverability indicators with high accuracy by using the analytical method.

8.2. DIRECTIONS FOR FUTURE WORK

The research questions of this thesis and above insights obtained from the results lead to several future research directions.

In Chapter 2, we use critical links-based approximations to estimate the number of driver nodes in face of random attacks and targeted attacks. By analyzing the results, we found that for dense networks, there are barely critical links in these networks. In this case, our approximation does not perform well. It is meaningful to propose another method which is applicable for dense networks to better capture the change of the number of driver nodes in failures.

In our research on network controllability, we proposed well-performed approximations in Chapter 2, 3 and 4. However, the exact application of driver nodes in real-world systems is still not clear, let alone analyzing the dynamic process under the control of driver nodes. It is promising if we can find real systems or set up test beds to make the research on network controllability more practical.

In Chapter 4, we use several machine learning models, such as Linear Regression, Random Forest and Artificial Neural Networks, to learn the gap between our approximations with simulation results. However, we still need to select different properties of the network manually which is inefficient. Besides, the properties we choose cannot fully describes the network. Thus, a lot of information is missing which degrades the final results obtained. In this case, graph neural network is a promising tool since it is graph-oriented and sensitive to network structure.

In Chapter 6 and 7, we propose the topological approach and apply it to quantify the recoverability of networks under perturbations by using recoverability indicators. As the next step, it is essential to find effective methods to improve the recoverability of networks and validate the performance in real-world networks.

ACKNOWLEDGEMENTS

First of all, I owe my deepest gratitude to my two promotors, Prof. Piet Van Mieghem and Prof. Robert Kooij. It has been my honour and luck to have been working with you both. Piet has always impressed me with his passion and his rigorous attitude towards scientific research. During the early stage of my research, he always reminded me to improve my English and emphasized the importance of precise academic writing. I will bear it in mind and regard it as a dogma for my future career. My sincere gratitude goes to Rob as he has been daily supervising me from the beginning. He is always helpful and supportive in my research. He gave me the freedom and encouragement to pursue my ideas and his kind advises helped me to formulate interesting questions in my research. I am grateful for his patience and tolerance when I am not working efficiently. He always has innovative sometimes crazy ideas which inspired me a lot and promoted my research. He is not only a promotor to me, but also a friend, I really enjoyed listening to his adventure in China and talking with him in Mandarin. His humours always makes me feel relaxed and shows me that a professor can be an attractive and interesting person. His passion in life will always inspire me.

During my Ph.D. studies, I have been fortunate enough to work together and collaborate with some distinguished researchers and master students, Dr. Zhidong He, Ashish Dhiman, Anqi Chen, Hanshu Yu, and Prof. Roland Bouffanais. I am very grateful for the experience discussing research issues with them. Dr. Xiangrong Wang helped me a lot when I was stuck in my research. I feel grateful for Prof. Josep Marzo from University of Girona, I really enjoyed working together with his group in a short-term project and I had a great time in Girona. I would also like to express my gratitude to the committee members of my thesis defence for their time and effort spent on my thesis.

Next, I would like to thank my office mates. Dr. Zhidong He is the office mate with me for the longest time. He is such a nice person and he is always supportive. In my mind, he is also the third promotor to me. He taught me a lot and gave me a lot of guidance based on his research experience, which made my research much easier. His unique attitude towards life and marriage never failed to make my office life enjoyable. I hope he will find the one and get married soon. Besides, I also enjoyed sharing my Hi-Fi experience with him since both of us are Hi-Fi enthusiasts. As a senior colleague, Dr. Hale Çetinay helped me a lot and gave me many useful ideas at the beginning of my research. I also enjoyed talking with Misa Taguchi about Japanese culture. Ivan Jokić likes to share his new findings with me. His serious working attitude and humbleness impressed me. Gabriel Budel can speak Mandarin well and he is always ready to help me and others. He is quite patient to answer any questions. Fenghua Wang always has questions to ask. Her funny comments make the working atmosphere less stressing.

I am also gratitude to other supportive colleagues in NAS group: Dr. Edgar van Boven, Dr. Remco Litjens, Dr. Eric Smeitink, Dr. Maksim Kitsak, Dr. Mattia Sensi, Rogier Noldus, Dr. Jaron Sanders, Karel Devriendt, Dr. Bastian Prasse, Maria Raftopoulou, Qingfeng Tong,

Massimo Achterberg and Albert Senen-Cerda. I owe special thanks to Dr. Qiang Liu and Long Ma. I have the most memorable experience with them in both academic study and spare-time. I am also grateful for the former secretary Joyce van Velzen, Laura de Groot and the current secretary Trisha de Jonge, their hard work guaranteed the operation of NAS group.

My sincere gratitude also goes to the friends I made after I came to the Netherlands. Sheila Sang, my girlfriend, is an angel who brought me a lot of joy and made my life more colorful. She is always supportive and she has the magic to make me happy. From Etienne Borgart and Mery Sang, I learned more about the Dutch culture. They make me feel at home and I am sincerely grateful for their hospitality. Their cat Meili is also adorable. Next up, Lixia Liang, is always young and full of energy. She is good at cooking and always makes big meals for me. I also learned a lot of medical tips from her. I am grateful for all that she has done for me. I would also like to thank Yisu Cheng. She is considerate and often gave me free cakes which contributed to the growth of my belly. Lastly, Leon Tsang, my fellow gamer, I had a lot of fun talking with him about video games. They all made my time here more enjoyable. I will treasure all the unforgettable memories with them.

Finally, I would like to thank my parents. I am grateful to my family for their unconditional love, care and emotional support.

Peng Sun
Delft, January 2022

APPENDIX

A. APPENDIX FOR CHAPTER 2

Since there are $\binom{L_c}{i}$ possible ways to choose i critical links from L_c critical links and there are $\binom{L-L_c}{m-i}$ possible ways to choose $m-i$ non-critical links from $L-L_c$ non-critical links, the contribution to the increase in N_D for each i is $i\binom{L_c}{i}\binom{L-L_c}{m-i}$. The expectation of the increase N_D^* of the minimum number of driver node N_D after randomly removing m links, is the sum of this expression for all $i = 1, 2, \ldots, m$ and divide it by $\binom{L}{M}$.

$$N_D^* = \frac{\sum_{i=1}^{m} i\binom{L_c}{i}\binom{L-L_c}{m-i}}{\binom{L}{m}} \tag{A1}$$

We rewrite the numerator of the right hand site of Eq. (A1):

$$\begin{aligned}\sum_{i=1}^{m} i\binom{L_c}{i}\binom{L-L_c}{m-i} &= \sum_{i=1}^{m} \frac{L_c!}{(i-1)!(L_c-i)!}\binom{L-L_c}{m-i} \\ &= L_c \sum_{i=1}^{m} \binom{L_c-1}{i-1}\binom{L-L_c}{m-i} \\ &= L_c \sum_{i=0}^{m-1} \binom{L_c-1}{i}\binom{L-L_c}{m-i-1}\end{aligned}$$

By using Vandermonde's formula: $\sum_{j=0}^{k}\binom{a}{j}\binom{b}{k-j} = \binom{a+b}{k}$, we obtain $L_c\sum_{i=0}^{m-1}\binom{L_c-1}{i}\binom{L-L_c}{m-1-i} = L_c\binom{L-1}{m-1}$. Finally, dividing this expression by $\binom{L}{m}$, we obtain

$$N_D^* = lL_c \tag{A2}$$

When the fraction of removed links is less than, or equal to l_c, we obtain

$$N_D = N_{D0} + lL_c \tag{A3}$$

Normalizing Eq.(A3) we obtain $n_{D,rand}$ in Eq.(2.1).

B. APPENDIX FOR CHAPTER 3

PROOF THEOREMS 2-4

We first give the proof of Theorem 2. The out-degree distribution $P_{out}(\cdot)$ for the unperturbed network is given in Eq.(3.15). Let us denote the out-degree distribution for the perturbed network by $\bar{P}_{out}(\cdot)$. Then it follows from Lemma 1 and Eq.(3.15) that

$$\bar{P}_{out}(k_{out}) = (1-p)^{k_{out}} \sum_{j=k_{out}}^{N-1} \binom{j}{k_{out}} p^{j-k_{out}} \delta(k-j). \tag{B1}$$

Therefore we obtain

$$\bar{P}_{out}(k_{out}) = 0, \tag{B2}$$

if $k_{out} > k$ and

$$\bar{P}_{out}(k_{out}) = (1-p)^{k_{out}} \binom{k}{k_{out}} p^{k-k_{out}} \tag{B3}$$

if $k_{out} \leq k$. From this we get

$$\begin{aligned} \bar{G}_{out}(x) = \sum_{k_{out}=0}^{\infty} \bar{P}_{out}(k_{out}) x^{k_{out}} = \sum_{k_{out}=0}^{k} (1-p)^{k_{out}} \binom{k}{k_{out}} p^{k-k_{out}} x^{k_{out}} = \\ \sum_{k_{out}=0}^{k} \binom{k}{k_{out}} ((1-p)x)^{k_{out}} p^{k-k_{out}} = (p+(1-p)x)^k. \end{aligned} \tag{B4}$$

This proves that Eq.(3.33) holds.

We assumed that the in-degree distribution of the original graph follows a Poisson distribution, see (3.16) but for finite N the actual distribution is binomial. However, for $N \longrightarrow \infty$ the limiting distribution is indeed Poissonian. Therefore, for proving that Eq.(3.34) holds, we will use Lemma 1 with $N = \infty$. The in-degree distribution $P_{in}(\cdot)$ for the unperturbed network is given in Eq. (3.16). Let us denote the in-degree distribution for the perturbed network by $\bar{P}_{in}(\cdot)$. Then it follows from Lemma 1 and Eq.(3.16) that

$$\bar{P}_{in}(k_{in}) = (1-p)^{k_{in}} \sum_{j=k_{in}}^{\infty} \binom{j}{k_{in}} p^{j-k_{in}} \frac{k^j}{j!} e^{-k}. \tag{B5}$$

From this we get

$$\begin{aligned} \bar{G}_{in}(x) = \sum_{k_{in}=0}^{\infty} \bar{P}_{in}(k_{in}) x^{k_{in}} = \sum_{k_{in}=0}^{\infty} (1-p)^{k_{in}} \sum_{j=k_{in}}^{\infty} \binom{j}{k_{in}} p^{j-k_{in}} \frac{k^j}{j!} e^{-k} x^{k_{in}} = \\ e^{-k} \sum_{k_{in}=0}^{\infty} \left(\frac{(1-p)x}{p}\right)^{k_{in}} \sum_{j=k_{in}}^{\infty} \binom{j}{k_{in}} \frac{(pk)^j}{j!} = \\ e^{-k} \sum_{k_{in}=0}^{\infty} \left(\frac{(1-p)x}{p}\right)^{k_{in}} \sum_{j=k_{in}}^{\infty} \frac{1}{k_{in}!} \frac{(pk)^j}{(j-k_{in})!} = \\ e^{-k} \sum_{k_{in}=0}^{\infty} \left(\frac{(1-p)x}{p}\right)^{k_{in}} \frac{1}{k_{in}!} \sum_{j=k_{in}}^{\infty} \frac{(pk)^{j-k_{in}} (pk)^{k_{in}}}{(j-k_{in})!} = \\ e^{-k} \sum_{k_{in}=0}^{\infty} \left(\frac{(1-p)x}{p}\right)^{k_{in}} \frac{(pk)^{k_{in}}}{k_{in}!} \sum_{\hat{j}=0}^{\infty} \frac{(pk)^{\hat{j}}}{\hat{j}!} = \\ e^{-k} \sum_{k_{in}=0}^{\infty} \frac{(k(1-p)x)^{k_{in}}}{k_{in}!} e^{pk} = \\ e^{-k} e^{k(1-p)x} e^{pk} = e^{-k(1-p)(1-x)}. \end{aligned} \tag{B6}$$

This proves that Eq.(3.34) holds.

Next we will prove Theorem 3. Using the same notation as before, it follows from Eq.(3.8) that for the perturbed system the generating function $\bar{H}_{out}(x)$ is given by

$$\bar{H}_{out}(x) = \sum_{k_{out}=1}^{\infty} \frac{k_{out}\bar{P}_{out}(k_{out})}{< k_{out} >} x^{k_{out}-1} \tag{B7}$$

Then, using Eqs.(B2)-(B3) we obtain

$$\begin{aligned} \bar{H}_{out}(x) = \sum_{k_{out}=1}^{k} \frac{k_{out}(1-p)^{k_{out}}\binom{k}{k_{out}}p^{k-k_{out}}}{k(1-p)} x^{k_{out}-1} = \\ \sum_{k_{out}=1}^{k} \binom{k-1}{k_{out}-1} p^{k-k_{out}}((1-p)x)^{k_{out}-1} = \\ \sum_{m=0}^{k-1} \binom{k-1}{m} p^{k-1-m}((1-p)x)^{m} = (p+(1-p)x)^{k-1}. \end{aligned} \tag{B8}$$

Finally, we prove Eq.(3.37).
Using the same notation as before, it follows from Eq.(3.9) that for the perturbed system the generating function $\bar{H}_{in}(x)$ is given by

$$\bar{H}_{in}(x) = \sum_{k_{in}=1}^{\infty} \frac{k_{in}\bar{P}_{in}(k_{in})}{< k_{in} >} x^{k_{in}-1}. \tag{B9}$$

Then, using Eq.(B5) we obtain

$$\begin{aligned} \bar{H}_{in}(x) = \sum_{k_{in}=1}^{\infty} \frac{k_{in}(1-p)^{k_{in}}}{k(1-p)} \sum_{j=k_{in}}^{\infty} \binom{j}{k_{in}} p^{j-k_{in}} \frac{k^j}{j!} e^{-k} x^{k_{in}-1} = \\ e^{-k} \sum_{k_{in}=1}^{\infty} \frac{k_{in}(k(1-p)x)^{k_{in}}}{xk(1-p)k_{in}!} e^{pk} = e^{-k+pk} \sum_{k_{in}=1}^{\infty} \frac{(k(1-p)x)^{k_{in}-1}}{(k_{in}-1)!} = \\ e^{-k+pk} \sum_{m=0}^{\infty} \frac{(k(1-p)x)^{m}}{m!} = e^{-k+pk+k(1-p)x} = e^{-k(1-p)(1-x)}. \end{aligned} \tag{B10}$$

This finishes the proof of Theorem 3.

Proof of Theorem 4.
Using Theorem 2 and 3, the set of equations (3.11)-(3.14) becomes

$$w_1 = (p+(1-p)\hat{w}_2)^{k-1} \tag{B11}$$

$$\hat{w}_2 = 1 - e^{-k(1-p)w_1} \tag{B12}$$

$$w_2 = 1 - (p+(1-p)(1-\hat{w}_1))^{k-1} \tag{B13}$$

$$\hat{w}_1 = e^{-k(1-p)(1-w_2)} \tag{B14}$$

By setting $\hat{w}_2 = 1 - \hat{w}_1$ and $w_1 = 1 - w_2$, it follows that the pair of Eqs.(B11)-(B12) is equivalent to the pair of Eqs.(B13)-(B14) .
From this it follows that n_D in Eq.(3.10) becomes

$$n_D = \bar{G}_{out}(1 - \hat{w}_1) + \bar{G}_{in}(w_2) - 1 + k(1-p)\,\hat{w}_1(1 - w_2) \tag{B15}$$

Using Eqs.(3.33), (3.34) and (B14), this leads to Eq.(3.39). Furthermore, Eq.(3.40) follows from the substitution of $\hat{w}_1$ given in Eq.(B14) into Eq.(B13).

Finally, we prove that Eq.(3.41) holds. First, we rewrite Eq.(3.39) as

$$n_D = (p + (1-p)(1 - \hat{w}_1))^k - 1 + \hat{w}_1 + k(1-p)(1 - w_2)\,\hat{w}_1, \tag{B16}$$

where $\hat{w}_1$ satisfies

$$\hat{w}_1 = e^{-k(1-p)(p+(1-p)(1-\hat{w}_1))^{k-1}}. \tag{B17}$$

Therefore, for large k we obtain

$$\hat{w}_1 \approx e^{-k(1-p)}, \tag{B18}$$

while from Eq.(B13) we get

$$1 - w_2 = (p + (1-p)(1 - \hat{w}_1))^{k-1} \approx 1 - (1-p)(k-1)\,\hat{w}_1. \tag{B19}$$

Then plugging Eqs,(B18) and (B19) into Eq.(B16) yields

$$\begin{aligned} n_D \approx 1 - (1-p)k\,\hat{w}_1 - 1 + \hat{w}_1 + k(1-p)(1 - (1-p)(k-1)\,\hat{w}_1)\,\hat{w}_1 = \\ 1 - (1-p)k\,\hat{w}_1 - 1 + \hat{w}_1 + k(1-p)\,\hat{w}_1 - (1-p)^2 k(k-1)\,\hat{w}_1^2 \approx \hat{w}_1 \approx e^{-k(1-p)}. \end{aligned} \tag{B20}$$

This completes the proof of Theorem 4.

PROOF FOR THEOREM 5 AND 7

Proof of Theorem 5.
Let us denote the out-degree distribution for the considered network by $\hat{P}_{out}(\cdot)$. Then it holds that

$$\hat{P}_{out}(k_{out}) = \alpha\delta(k_{out} - k_1) + (1-\alpha)\delta(k_{out} - k_2). \tag{B21}$$

Then, denoting the generating function for the out-degree distribution by $\hat{G}_{out}$, we get

$$\begin{aligned} \hat{G}_{out}(x) = \sum_{k_{out}=0}^{\infty} \hat{P}_{out}(k_{out})\,x^{k_{out}} = \\ \sum_{k_{out}=0}^{\infty} (\alpha\delta(k_{out} - k_1) + (1-\alpha)\delta(k_{out} - k_2))\,x^{k_{out}} = \alpha x^{k_1} + (1-\alpha)x^{k_2}. \end{aligned} \tag{B22}$$

Let us denote the in-degree distribution for the considered network by $\hat{P}_{in}(\cdot)$, which for large N will approach a Poisson distribution with average $k = \alpha k_1 + (1-\alpha)k_2$. Then it holds that

$$\hat{P}_{in}(k_{in}) = \frac{k^{k_{in}}}{k_{in}!}e^{-k}. \tag{B23}$$

Then, denoting the generating function for the in-degree distribution by $\hat{G}_{in}$, we get

$$\hat{G}_{in}(x) = \sum_{k_{in}=0}^{\infty} \hat{P}_{in}(k_{in}) x^{k_{in}} = \sum_{k_{in}=0}^{\infty} \frac{k^{k_{in}}}{k_{in}!} e^{-k} x^{k_{in}} = e^{-k} \sum_{k_{in}=0}^{\infty} \frac{(kx)^{k_{in}}}{k_{in}!} = e^{-k} e^{kx} = e^{-k(1-x)}. \tag{B24}$$

This finishes the proof of Theorem 5.

Proof of Theorem 6.
Using the same notation as before, it follows from Eq.(3.8) that the generating function $\hat{H}_{out}(x)$ is given by

$$\hat{H}_{out}(x) = \sum_{k_{out}=1}^{\infty} \frac{k_{out} \hat{P}_{out}(k_{out})}{< k_{out} >} x^{k_{out}-1} \tag{B25}$$

Then, using Eqs.(B21) we obtain

$$\hat{H}_{out}(x) = \sum_{k_{out}=1}^{\infty} \frac{k_{out}(\alpha \delta(k_{out} - k_1) + (1-\alpha)\delta(k_{out} - k_2))}{k} x^{k_{out}-1} = \frac{\alpha k_1 x^{k_1 - 1} + (1-\alpha) k_2 x^{k_2 - 1}}{k}. \tag{B26}$$

Finally, we prove Eq.(3.47).
Using the same notation as before, it follows from Eq.(3.9) that for the perturbed system the generating function $\hat{H}_{in}(x)$ is given by

$$\hat{H}_{in}(x) = \sum_{k_{in}=1}^{\infty} \frac{k_{in} \hat{P}_{in}(k_{in})}{< k_{in} >} x^{k_{in}-1} \tag{B27}$$

Then, using Eq.(B23) we obtain

$$\bar{H}_{in}(x) = \sum_{k_{in}=1}^{\infty} \frac{k_{in} k^{k_{in}} e^{-k} x^{k_{in}-1}}{k k_{in}!} = e^{-k} \sum_{k_{in}=1}^{\infty} \frac{k^{k_{in}-1} x^{k_{in}-1}}{(k_{in}-1)!} - e^{-k} \sum_{i=0}^{\infty} \frac{(kx)^i}{i!} = e^{-k} e^{kx} = e^{-k(1-x)}. \tag{B28}$$

This finishes the proof of Theorem 6

Proof of Theorem 7.
Using Theorems 5 and 6, the set of Eqs.(3.11)-(3.14) becomes

$$w_1 = \frac{\alpha k_1 \hat{w}_2^{k_1 - 1} + (1-\alpha) k_2 \hat{w}_2{}^{k_2 - 1}}{k} \tag{B29}$$

$$\hat{w}_2 = 1 - e^{-k w_1} \tag{B30}$$

$$w_2 = 1 - \frac{\alpha k_1 (1-\hat{w}_1)^{k_1-1} + (1-\alpha) k_2 (1-\hat{w}_1)^{k_2-1}}{k} \tag{B31}$$

$$\hat{w}_1 = e^{-k(1-w_2)} \tag{B32}$$

By setting $\hat{w}_2 = 1 - \hat{w}_1$ and $w_1 = 1 - w_2$, it follows that the pair of Eqs.(B29)-(B30) is equivalent to the pair of equations Eqs.(B31)-(B32).
From this it follows that n_D in Eq.(3.10) becomes

$$n_D = \hat{G}_{out}(1-\hat{w}_1) + \hat{G}_{in}(w_2) - 1 + k\hat{w}_1(1-w_2) \tag{B33}$$

Using Eqs.(3.43), (3.44) and (B32), this leads to Eq.(3.48). Furthermore, Eq.(3.49) follows from the substitution of $\hat{w}_1$ given in Eq.(B32) into Eq. (B31). Finally, we prove that Eq.(3.50) holds. First, we rewrite Eq.(3.48) as

$$n_D = \alpha(1-\hat{w}_1)^{k_1} + (1-\alpha)(1-\hat{w}_1)^{k_2} - 1 + \hat{w}_1 + k(1-w_2)\hat{w}_1, \tag{B34}$$

where $\hat{w}_1$ satisfies

$$\begin{aligned} \hat{w}_1 = e^{-(\alpha k_1 (1-\hat{w}_1)^{k_1-1} + (1-\alpha) k_2 (1-\hat{w}_1)^{k_2-1})} \approx \\ e^{-(\alpha k_1 + (1-\alpha) k_2) + (\alpha k_1 (k_1-1) + (1-\alpha) k_2 (k_2-1))\hat{w}_1} = \\ e^{-k} e^{(\alpha k_1 (k_1-1) + (1-\alpha) k_2 (k_2-1))\hat{w}_1} \end{aligned} \tag{B35}$$

Therefore, for large k we obtain

$$\hat{w}_1 \approx e^{-k}, \tag{B36}$$

while from Eq.(B31) we get

$$\begin{aligned} w_2 \approx 1 - \frac{\alpha k_1 (1-(k_1-1)\hat{w}_1 + (1-\alpha) k_2 (1-(k_2-1)\hat{w}_1}{k} = \\ 1 - \frac{k - (\alpha k_1 (k_1-1) + (1-\alpha) k_2 (k_2-1))\hat{w}_1}{k} = \\ \frac{\alpha k_1 (k_1-1) + (1-\alpha) k_2 (k_2-1)}{k} \hat{w}_1 \equiv \sigma \hat{w}_1. \end{aligned} \tag{B37}$$

Then plugging Eqs.(B36) and (B37) into Eq.(B34) yields

$$\begin{aligned} n_D \approx \alpha(1-k_1\hat{w}_1) + (1-\alpha)(1-k_2)\hat{w}_1 - 1 + \hat{w}_1 + k(1-\sigma\hat{w}_1)\hat{w}_1 = \\ \alpha - \alpha k_1 \hat{w}_1 + 1 - \alpha - k_2(1-\alpha)\hat{w}_1 - 1 + \hat{w}_1 + k\hat{w}_1 - k\sigma\hat{w}_1^2 \approx \hat{w}_1 = e^{-k}. \end{aligned} \tag{B38}$$

This completes the proof of Theorem 7.

PROOF FOR THEOREM 8

Using Theorems 8 and 9, the set of Eqs.(3.11)-(3.14) becomes

$$\omega_1 = \frac{\alpha k_1 (p + (1-p)\hat{\omega}_2)^{k_1-1} + (1-\alpha) k_2 (p + (1-p)\hat{\omega}_2)^{k_2-1}}{k} \tag{B39}$$

$$1-\omega_2 = \frac{\alpha k_1 (p+(1-p)(1-\hat{\omega}_1))^{k_1-1} + (1-\alpha) k_2 (p+(1-p)(1-\hat{\omega}_1))^{k_2-1}}{k} \tag{B40}$$

$$\hat{\omega}_1 = e^{-k(1-p)(1-\omega_2)} \tag{B41}$$

$$1-\hat{\omega}_2 = e^{-k(1-p)\omega_1} \tag{B42}$$

By setting $\hat{\omega}_2 = 1-\hat{\omega}_1$ and $\omega_2 = 1-\omega_1$, it follows that the pair of Eqs.(B40)-(B41) is equivalent to the pair of Eqs.(B39)-(B42). Then by using Eq.(3.10), we get

$$\begin{aligned} n_D = \alpha (p+(1-p)(1-e^{-k(1-\omega_2)}))^{k_1} + (1-\alpha)(p+(1-p)(1-e^{-k(1-\omega_2)}))^{k_2} \\ -1+e^{-k(1-p)(1-\omega_2)} + k(1-p)e^{-k(1-\omega_2)}(1-\omega_2) \end{aligned} \tag{B43}$$

where w_2 is the solution of Eqs.(B40)-(B41). This proves that Eq.(3.58) holds.

Finally, we prove that Eq.(3.60) holds. From Eqs.(B40)-(B41) it follows that

$$\begin{aligned} \hat{\omega}_1 = e^{-(1-p)(\alpha k_1 (p+(1-p)(1-\hat{\omega}_1))^{k_1-1} + (1-\alpha) k_2 (p+(1-p)(1-\hat{\omega}_1))^{k_2-1})} \approx \\ e^{-k(1-p)} e^{(1-p)^2(\alpha k_1(k_1-1)+(1-\alpha)k_2(k_2-1))\hat{w}_1} \end{aligned} \tag{B44}$$

Therefore, for large k we obtain

$$\hat{w}_1 \approx e^{-k(1-p)}, \tag{B45}$$

Similarly, from Eq.(B40) we can deduce

$$w2 \approx \frac{(1-p)(\alpha k_1(k_1-1)+(1-\alpha)k_2(k_2-1))}{k} \hat{w}_1 \equiv \sigma \hat{w}_1 \tag{B46}$$

Substitution of Eq.(B45) and Eq.(B46) into Eq.(B43), we obtain

$$n_D \approx e^{-\bar{k}(1-p)} \tag{B47}$$

This completes the proof of Theorem 10.

C. APPENDIX FOR CHAPTER 7

As deduced in [70],

Lemma 11. *after adding a fraction f of links in a uniform and random way to a network $G_0(N,L)$, with degree distribution $Pr[D_{G_0}=j]$, the degree distribution $Pr[D_G=k]$ of the new network G satisfies:*

$$Pr[D_G=k] = (1-f)^{N-1-k} \sum_{j=0}^{N-1} \binom{N-1-j}{k-j} f^{k-j} Pr[D_{G_0}=j], \tag{C1}$$

where $f = \frac{m}{N(N-1)-L}$ denotes the fraction of added links in the original network G_0. Then, the corresponding generating function $\bar{G}(x)$ is

$$\begin{aligned}\bar{G}(x) = \sum_{k=0}^{N-1}\sum_{j=0}^{k}\binom{N-1-j}{k-j}(1-f)^{N-1-k}f^{k-j}Pr[D_{G_0}=j]x^k = \\ \sum_{j=0}^{N-1}\sum_{k=j}^{N-1}\binom{N-1-j}{k-j}(1-f)^{N-1-k}f^{k-j}Pr[D_{G_0}=j]x^k\end{aligned} \tag{C2}$$

Let $\alpha = k - j$, we get

$$\begin{aligned}\bar{G}(x) = \sum_{j=0}^{N-1}\sum_{\alpha=0}^{N-1-j}\binom{N-1-j}{\alpha}(1-f)^{N-1-j-\alpha}f^{\alpha}Pr[D_{G_0}=j]x^{\alpha+j} = \\ \sum_{j=0}^{N-1}\sum_{\alpha=0}^{N-1-j}\binom{N-1-j}{\alpha}(1-f)^{N-1-j-\alpha}(fx)^{\alpha}Pr[D_{G_0}=j]x^{j} = \\ \sum_{j=0}^{N-1}(1-f(1-x))^{N-1-j}Pr[D_{G_0}=j]x^{j} = \\ (1-f(1-x))^{N-1}G(\frac{x}{1-f(1-x)})\end{aligned} \tag{C3}$$

BIBLIOGRAPHY

[1] A. Mislove, M. Marcon, K. P. Gummadi, P. Druschel, and B. Bhattacharjee, "Measurement and analysis of online social networks," in *Proceedings of the 7th ACM SIGCOMM conference on Internet measurement*, 2007, pp. 29–42.

[2] E. Rodríguez-Núñez and J. C. García-Palomares, "Measuring the vulnerability of public transport networks," *Journal of Transport Geography*, vol. 35, pp. 50–63, 2014.

[3] S. Giovinazzi, A. Austin, R. Ruiter, C. Foster, M. Nayyerloo, N.-K. Nair, and L. Wotherspoon, "Resilience and fragility of the telecommunication network to seismic events," *Bulletin of the New Zealand Society for Earthquake Engineering*, vol. 50, no. 2, pp. 318–328, 2017.

[4] W. T. Miller, P. J. Werbos, and R. S. Sutton, *Neural networks for control*. MIT Press, 1995.

[5] S. Sherwin, V. Franke, J. Peiró, and K. Parker, "One-dimensional modelling of a vascular network in space-time variables," *Journal of Engineering Mathematics*, vol. 47, no. 3, pp. 217–250, 2003.

[6] H. J. Herrmann, C. M. Schneider, A. A. Moreira, J. S. Andrade Jr, and S. Havlin, "Onion-like network topology enhances robustness against malicious attacks," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2011, no. 01, p. P01027, 2011.

[7] K. Park and H. Lee, "On the effectiveness of route-based packet filtering for distributed dos attack prevention in power-law internets," *ACM SIGCOMM Computer Communication Review*, vol. 31, no. 4, pp. 15–26, 2001.

[8] Y. Koç, M. Warnier, P. Van Mieghem, R. E. Kooij, and F. M. Brazier, "A topological investigation of phase transitions of cascading failures in power grids," *Physica A: Statistical Mechanics and its Applications*, vol. 415, pp. 273–284, 2014.

[9] M. N. Kabir, M. A. Rahman, S. Azad, M. M. A. Azim, and M. Z. A. Bhuiyan, "A connection probability model for communications networks under regional failures," *International Journal of Critical Infrastructure Protection*, vol. 20, pp. 16–25, 2018.

[10] P. Erdös and A. Rényi, "On random graphs I," *Publicationes Mathematicae Debrecen*, vol. 6, pp. 290–297, 1959.

[11] D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small-world' networks," *Nature*, vol. 393, no. 6684, p. 440, 1998.

[12] A.-L. Barabási and E. Bonabeau, "Scale-free networks," *Scientific American*, vol. 288, no. 5, pp. 60–69, 2003.

[13] R. Albert, H. Jeong, and A.-L. Barabási, "Error and attack tolerance of complex networks," *Nature*, vol. 406, no. 6794, pp. 378–382, 2000.

[14] D. S. Callaway, M. E. Newman, S. H. Strogatz, and D. J. Watts, "Network robustness and fragility: Percolation on random graphs," *Physical Review Letters*, vol. 85, no. 25, p. 5468, 2000.

[15] R. Cohen, K. Erez, S. Havlinl, M. Newman, A.-L. Barabási, D. J. Watts *et al.*, "Resilience of the internet to random breakdowns," in *The Structure and Dynamics of Networks*. Princeton University Press, 2011, pp. 507–509.

[16] R. Cohen, D. Ben-Avraham, and S. Havlin, "Percolation critical exponents in scale-free networks," *Physical Review E*, vol. 66, no. 3, p. 036113, 2002.

[17] A. Vázquez and Y. Moreno, "Resilience to damage of graphs with degree correlations," *Physical Review E*, vol. 67, no. 1, p. 015101, 2003.

[18] H. S. Wilf, *Generating Functionology*. CRC Press, 2005.

[19] M. E. Newman, S. H. Strogatz, and D. J. Watts, "Random graphs with arbitrary degree distributions and their applications," *Physical Review E*, vol. 64, no. 2, p. 026118, 2001.

[20] X. Wang, E. Pournaras, R. E. Kooij, and P. Van Mieghem, "Improving robustness of complex networks via the effective graph resistance," *The European Physical Journal B*, vol. 87, no. 9, p. 221, 2014.

[21] Z. He, K. Navneet, W. van Dam, and P. Van Mieghem, "Robustness assessment of multimodal freight transport networks," *Reliability Engineering & System Safety*, vol. 207, p. 107315, 2021.

[22] X. Huang, J. Gao, S. V. Buldyrev, S. Havlin, and H. E. Stanley, "Robustness of interdependent networks under targeted attack," *Physical Review E*, vol. 83, no. 6, p. 065101, 2011.

[23] M. Parandehgheibi and E. Modiano, "Robustness of interdependent networks: The case of communication networks and the power grid," in *2013 IEEE Global Communications Conference (GLOBECOM)*. IEEE, 2013, pp. 2164–2169.

[24] A. Lombardi and M. Hörnquist, "Controllability analysis of networks," *Physical Review E*, vol. 75, no. 5, p. 056110, 2007.

[25] I. D. Couzin, J. Krause, N. R. Franks, and S. A. Levin, "Effective leadership and decision-making in animal groups on the move," *Nature*, vol. 433, no. 7025, pp. 513–516, 2005.

[26] R. E. Kalman, "Mathematical description of linear dynamical systems," *Journal of the Society for Industrial and Applied Mathematics, Series A: Control*, vol. 1, no. 2, pp. 152–192, 1963.

[27] Y. Yang and G. Xie, "Mining maximum matchings of controllability of directed networks based on in-degree priority," in *2016 35th Chinese Control Conference (CCC)*. IEEE, 2016, pp. 1263–1267.

[28] J. E. Hopcroft and R. M. Karp, "An nˆ5/2 algorithm for maximum matchings in bipartite graphs," *SIAM Journal on Computing*, vol. 2, no. 4, pp. 225–231, 1973.

[29] W. Zhang, N. Wang, and C. Nicholson, "Resilience-based post-disaster recovery strategies for road-bridge networks," *Structure and Infrastructure Engineering*, vol. 13, no. 11, pp. 1404–1413, 2017.

[30] L. Sun, J. An, Y. Yang, and M. Zeng, "Recovery strategies for service composition in dynamic network," in *2011 International Conference on Cloud and Service Computing*. IEEE, 2011, pp. 60–64.

[31] Y. Almoghathawi, A. D. González, and K. Barker, "Exploring recovery strategies for optimal interdependent infrastructure network resilience," *Networks and Spatial Economics*, vol. 21, no. 1, pp. 229–260, 2021.

[32] J. P. Sterbenz, E. K. Çetinkaya, M. A. Hameed, A. Jabbar, S. Qian, and J. P. Rohrer, "Evaluation of network resilience, survivability, and disruption tolerance: analysis, topology generation, simulation, and experimentation," *Telecommunication Systems*, vol. 52, no. 2, pp. 705–736, 2013.

[33] J. E. Ramirez-Marquez, C. M. Rocco, K. Barker, and J. Moronta, "Quantifying the resilience of community structures in networks," *Reliability Engineering & System Safety*, vol. 169, pp. 466–474, 2018.

[34] J. Gao, B. Barzel, and A.-L. Barabási, "Universal resilience patterns in complex networks," *Nature*, vol. 530, no. 7590, pp. 307–312, 2016.

[35] M. A. Di Muro, C. E. La Rocca, H. E. Stanley, S. Havlin, and L. A. Braunstein, "Recovery of interdependent networks," *Scientific Reports*, vol. 6, no. 1, pp. 1–11, 2016.

[36] P. Sun, R. E. Kooij, Z. He, and P. Van Mieghem, "Quantifying the robustness of network controllability," in *2019 4th International Conference on System Reliability and Safety (ICSRS)*. IEEE, 2019, pp. 66–76.

[37] M. E. J. Newman, "The structure and function of complex networks," *SIAM Review*, vol. 45, pp. 167–256, 2003.

[38] R. Cohen, K. Erez, D. ben Avraham, and S. Havlin, "Resilience of the internet to random breakdowns," *Physical Review Letters*, vol. 85, pp. 4626–4628, Nov 2000.

[39] Y.-Y. Liu, J.-J. Slotine, and A.-L. Barabási, "Controllability of complex networks," *Nature*, vol. 473, no. 7346, p. 167, 2011.

[40] Z. Yuan, C. Zhao, Z. Di, W.-X. Wang, and Y.-C. Lai, "Exact controllability of complex networks," *Nature Communications*, vol. 4, p. 2447, 2013.

[41] T. Jia, Y.-Y. Liu, E. Csóka, M. Pósfai, J.-J. Slotine, and A.-L. Barabási, "Emergence of bimodality in controlling complex networks," *Nature Communications*, vol. 4, p. 2002, 2013.

[42] T. Nepusz and T. Vicsek, "Controlling edge dynamics in complex networks," *Nature Physics*, vol. 8, no. 7, p. 568, 2012.

[43] Ching-Tai Lin, "Structural controllability," *IEEE Transactions on Automatic Control*, vol. 19, no. 3, pp. 201–208, June 1974.

[44] N. J. Cowan, E. J. Chastain, D. A. Vilhena, J. S. Freudenberg, and C. T. Bergstrom, "Nodal dynamics, not degree distributions, determine the structural controllability of complex networks," *PLoS One*, vol. 7, no. 6, pp. 1–5, 2012.

[45] J. Ruths and D. Ruths, "Control profiles of complex networks," *Science*, vol. 343, no. 6177, pp. 1373–1376, 2014.

[46] G. Yan, G. Tsekenis, B. Barzel, J.-J. Slotine, Y.-Y. Liu, and A.-L. Barabási, "Spectrum of controlling and observing complex networks," *Nature Physics*, vol. 11, no. 9, p. 779, 2015.

[47] H. Cetinay, K. Devriendt, and P. Van Mieghem, "Nodal vulnerability to targeted attacks in power grids," *Applied Network Science*, vol. 3, no. 1, p. 34, 2018.

[48] B. Berche, C. von Ferber, T. Holovatch, and Y. Holovatch, "Resilience of public transport networks against attacks," *The European Physical Journal B*, vol. 71, no. 1, pp. 125–137, 2009.

[49] A. Socievole, F. De Rango, C. Scoglio, and P. Van Mieghem, "Assessing network robustness under SIS epidemics: The relationship between epidemic threshold and viral conductance," *Computer Networks*, vol. 103, pp. 196–206, 2016.

[50] X. Wang, Y. Koç, S. Derrible, S. N. Ahmad, W. J. Pino, and R. E. Kooij, "Multi-criteria robustness analysis of metro networks," *Physica A: Statistical Mechanics and its Applications*, vol. 474, pp. 19–31, 2017.

[51] X. Wang, R. E. Kooij, and P. Van Mieghem, "Modeling region-based interconnection for interdependent networks," *Physical Review E*, vol. 94, no. 4, p. 042315, 2016.

[52] S. Trajanovski, J. Martín-Hernández, W. Winterbach, and P. Van Mieghem, "Robustness envelopes of networks," *Journal of Complex Networks*, vol. 1, no. 1, pp. 44–62, 2013.

[53] Y. Koç, M. Warnier, P. Van Mieghem, R. E. Kooij, and F. M. Brazier, "The impact of the topology on cascading failures in a power grid model," *Physica A: Statistical Mechanics and its Applications*, vol. 402, pp. 169–179, 2014.

[54] C.-L. Pu, W.-J. Pei, and A. Michaelson, "Robustness analysis of network controllability," *Physica A: Statistical Mechanics and its Applications*, vol. 391, no. 18, pp. 4420–4425, 2012.

[55] S. Nie, X. Wang, H. Zhang, Q. Li, and B. Wang, "Robustness of controllability for networks based on edge-attack," *PLoS One*, vol. 9, no. 2, p. e89066, 2014.

[56] J. Thomas, S. Ghosh, D. Parek, D. Ruths, and J. Ruths, "Robustness of network controllability to degree-based edge attacks," in *International Workshop on Complex Networks and their Applications*. Springer, 2016, pp. 525–537.

[57] Z.-M. Lu and X.-F. Li, "Attack vulnerability of network controllability," *PLoS One*, vol. 11, no. 9, p. e0162289, 2016.

[58] S. Abebe Mengiste, A. Aertsen, and A. Kumar, "Effect of edge pruning on structural controllability and observability of complex networks," *Scientific Reports*, vol. 5, p. 18145, 12 2015.

[59] R. Rossi and N. Ahmed, "The network data repository with interactive graph analytics and visualization." in *Proc. of the 29th AAAI Conference on Artificial Intelligence*, vol. 15, 2015, pp. 4292–4293.

[60] O. Jahn, R. H. Möhring, A. S. Schulz, and N. E. Stier-Moses, "System-optimal routing of traffic flows with user constraints in networks with congestion," *Operations Research*, vol. 53, no. 4, pp. 600–616, 2005.

[61] P. Crucitti, V. Latora, and M. Marchiori, "A topological analysis of the italian electric power grid," *Physica A*, vol. 338, no. 1-2, pp. 92–97, 2004.

[62] J. S. Coleman, *Introduction to Mathematical Sociology*. Collier-Macmillan, 1964.

[63] M. A. Evans and D. Scavia, "Forecasting hypoxia in the Chesapeake Bay and Gulf of Mexico: Model accuracy, precision, and sensitivity to ecosystem change," *Environmental Research Letters*, vol. 6, no. 1, p. 015001, 2010.

[64] D. R. White, "Rethinking the role concept," *Research Methods in Social Network Analysis*, p. 429, 2017.

[65] R. Zhang, X. Wang, M. Cheng, and T. Jia, "The evolution of network controllability in growing networks," *Physica A: Statistical Mechanics and its Applications*, vol. 520, pp. 257–266, 2019.

[66] Y. Lou, L. Wang, K.-F. Tsang, and G. Chen, "Towards optimal robustness of network controllability: An empirical necessary condition," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 67, no. 9, pp. 3163–3174, 2020.

[67] S. Wang, Y. Yang, L. Sun, X. Li, Y. Li, and K. Guo, "Controllability robustness against cascading failure for complex logistic network based on dynamic cascading failure model," *IEEE Access*, vol. 8, pp. 127 450–127 461, 2020.

[68] M. Komareji and R. Bouffanais, "Resilience and controllability of dynamic collective behaviors," *PLoS One*, vol. 8, no. 12, pp. 1–15, 12 2013.

[69] M. Newman, *Networks: An introduction*. Oxford: Oxford University Press, 2010.

[70] P. Van Mieghem, *Performance Analysis of Complex Networks and Systems*. Cambridge University Press, 2014.

[71] A. Dhiman, P. Sun, and R. Kooij, "Using machine learning to quantify the robustness of network controllability," in *3rd International Conference on Machine Learning for Networking, MLN 2020*. Springer, 2021, pp. 19–39.

[72] T. M. Tirpak, "Telecommunication network resource management based on social network characteristics," Nov. 11 2010, US Patent App. 12/463,445.

[73] N. J. Cowan, E. J. Chastain, D. A. Vilhena, J. S. Freudenberg, and C. T. Bergstrom, "Nodal dynamics, not degree distributions, determine the structural controllability of complex networks," *PLoS One*, vol. 7, no. 6, p. e38398, 2012.

[74] Y. Koç, M. Warnier, P. Van Mieghem, R. E. Kooij, and F. M. Brazier, "The impact of the topology on cascading failures in a power grid model," *Physica A: Statistical Mechanics and its Applications*, vol. 402, pp. 169–179, 2014.

[75] P. Van Mieghem, C. Doerr, H. Wang, J. M. Hernandez, D. Hutchison, M. Karaliopoulos, and R. Kooij, "A framework for computing topological network robustness," *Delft University of Technology, Report20101218*, 2010.

[76] P. Holme, B. J. Kim, C. N. Yoon, and S. K. Han, "Attack vulnerability of complex networks," *Physical Review E*, vol. 65, no. 5, p. 056109, 2002.

[77] S. A. Mengiste, A. Aertsen, and A. Kumar, "Effect of edge pruning on structural controllability and observability of complex networks," *Scientific Reports*, vol. 5, no. 1, pp. 1–14, 2015.

[78] Y. Lou, Y. He, L. Wang, and G. Chen, "Predicting network controllability robustness: A convolutional neural network approach," *IEEE Transactions on Cybernetics*, 2020.

[79] A. K. Dhiman, *Measuring the Robustness of Network Controllability*. MSc thesis, Delft University of Technology, 2020.

[80] S. Knight, H. X. Nguyen, N. Falkner, R. Bowden, and M. Roughan, "The internet topology zoo," *IEEE Journal on Selected Areas in Communications*, vol. 29, no. 9, pp. 1765–1775, 2011.

[81] M. Himsolt, "Gml: A portable graph file format," Technical Report, Universitat Passau, Tech. Rep., 1997.

[82] U. Brandes, M. Eiglsperger, I. Herman, M. Himsolt, and M. S. Marshall, "Graphml progress report structural layer proposal," in *International Symposium on Graph Drawing*. Springer, 2001, pp. 501–512.

[83] R. Van Der Hofstad, *Random graphs and complex networks*. Cambridge University Press, 2016, vol. 1.

[84] P. Erdös and A. Rényi, "On the evolution of random graphs," in *the Structure and Dynamics of Networks*. Princeton University Press, 2011, pp. 38–82.

[85] R. Albert and A.-L. Barabási, "Statistical mechanics of complex networks," *Reviews of Modern Physics*, vol. 74, no. 1, p. 47, 2002.

[86] A.-L. Barabási, E. Ravasz, and T. Vicsek, "Deterministic scale-free networks," *Physica A: Statistical Mechanics and its Applications*, vol. 299, no. 3-4, pp. 559–564, 2001.

[87] P. Sun, R. E. Kooij, and P. Van Mieghem, "Reachability-based robustness of controllability in sparse communication networks," *IEEE Transactions on Network and Service Management*, vol. 18, no. 3, pp. 2764–2775, 2021.

[88] Y. Nakayama, K. Mori, K. Takaragi, and S. Domen, "System and method for performing interlocution at a plurality of terminals connected to communication network," Jan. 18 1994, US Patent 5,280,583.

[89] Y. Lou, L. Wang, and G. Chen, "Toward stronger robustness of network controllability: a snapback network model," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 65, no. 9, pp. 2983–2991, 2018.

[90] S.-M. Chen, Y.-F. Xu, and S. Nie, "Robustness of network controllability in cascading failure," *Physica A: Statistical Mechanics and its Applications*, vol. 471, pp. 536–539, 2017.

[91] Y. Lou, L. Wang, and G. Chen, "A framework of hierarchical attacks to network controllability," *Communications in Nonlinear Science and Numerical Simulation*, vol. 98, p. 105780, 2021.

[92] X. Yan-Dong, L. Song-Yang, H. Lv-Lin, and B. Liang, "Optimization of robustness of network controllability against malicious attacks," *Chinese Physics B*, vol. 23, no. 11, p. 118902, 2014.

[93] Z. Zhang, Y. Yin, X. Zhang, and L. Liu, "Optimization of robustness of interdependent network controllability by redundant design," *PLoS One*, vol. 13, no. 2, p. e0192874, 2018.

[94] D. Parekh, D. Ruths, and J. Ruths, "Reachability-based robustness of network controllability under node and edge attacks," in *Signal-Image Technology and Internet-Based Systems (SITIS), 2014 Tenth International Conference on*. IEEE, 2014, pp. 424–431.

[95] Z. He, P. Sun, and P. Van Mieghem, "Topological approach to measure network recoverability," in *2019 11th International Workshop on Resilient Networks Design and Modeling (RNDM)*. IEEE, 2019, pp. 1–7.

[96] M. L. Fredman and R. E. Tarjan, "Fibonacci heaps and their uses in improved network optimization algorithms," *Journal of the ACM (JACM)*, vol. 34, no. 3, pp. 596–615, 1987.

[97] R. Rossi and N. Ahmed, "Network repository," 2013. [Online]. Available: http://networkrepository.com

[98] C. I. Del Genio, T. Gross, and K. E. Bassler, "All scale-free networks are sparse," *Physical Review Letters*, vol. 107, no. 17, p. 178701, 2011.

[99] M. S. Uddin, S. T. H. Murshed, and L. Hossain, "Towards a scale free network approach to study organizational communication network." in *PACIS*, 2010, p. 196.

[100] L. Li, D. Alderson, J. C. Doyle, and W. Willinger, "Towards a theory of scale-free graphs: Definition, properties, and implications," *Internet Mathematics*, vol. 2, no. 4, pp. 431–523, 2005.

[101] P. Sun, Z. He, R. E. Kooij, and P. Van Mieghem, "Topological approach to measure the recoverability of optical networks," *Optical Switching and Networking*, vol. 41, p. 100617, 2021.

[102] J. Tapolcai, "Shared risk link group failure restoration with in-band approximate failure localization," *Optical Switching and Networking*, vol. 10, no. 2, pp. 163–172, 2013.

[103] J. L. Marzo, S. G. Cosgaya, N. Skorin-Kapov, C. Scoglio, and H. Shakeri, "A study of the robustness of optical networks under massive failures," *Optical Switching and Networking*, vol. 31, pp. 1–7, 2019.

[104] E. K. Cetinkaya and J. P. Sterbenz, "A taxonomy of network challenges," in *2013 9th International Conference on the Design of Reliable Communication Networks (DRCN)*. IEEE, 2013, pp. 322–330.

[105] P. Cholda, A. Mykkeltveit, B. E. Helvik, O. J. Wittner, and A. Jajszczyk, "A survey of resilience differentiation frameworks in communication networks," *IEEE Communications Surveys & Tutorials*, vol. 9, no. 4, pp. 32–55, 2007.

[106] A. Pašić, R. Girão-Silva, B. Vass, T. Gomes, and P. Babarczi, "FRADIR: A novel framework for disaster resilience," in *2018 10th International Workshop on Resilient Networks Design and Modeling (RNDM)*. IEEE, 2018, pp. 1–7.

[107] J. L. Marzo, E. Calle, S. G. Cosgaya, D. Rueda, and A. Mañosa, "On selecting the relevant metrics of network robustness," in *2018 10th International Workshop on Resilient Networks Design and Modeling (RNDM)*. IEEE, 2018, pp. 1–7.

[108] D. F. Rueda, E. Calle, and J. L. Marzo, "Robustness comparison of 15 real telecommunication networks: Structural and centrality measurements," *Journal of Network and Systems Management*, vol. 25, no. 2, pp. 269–289, 2017.

[109] X. Long, D. Tipper, and T. Gomes, "Measuring the survivability of networks to geographic correlated failures," *Optical Switching and Networking*, vol. 14, pp. 117–133, 2014.

[110] J. Zhu, C. Natalino, L. Wosinska, M. Furdek, and Z. Zhu, "Control plane robustness in software-defined optical networks under targeted fiber cuts," in *2018 International Conference on Optical Network Design and Modeling (ONDM)*. IEEE, 2018, pp. 118–123.

[111] S. Ferdousi, M. Tornatore, M. F. Habib, and B. Mukherjee, "Rapid data evacuation for large-scale disasters in optical cloud networks," *Journal of Optical Communications and Networking*, vol. 7, no. 12, pp. B163–B172, 2015.

[112] X. Xie, Q. Ling, P. Lu, W. Xu, and Z. Zhu, "Evacuate before too late: distributed backup in inter-dc networks with progressive disasters," *IEEE Transactions on Parallel and Distributed Systems*, vol. 29, no. 5, pp. 1058–1074, 2017.

[113] X. Pan and H. Wang, "Resilience of and recovery strategies for weighted networks," *PLoS One*, vol. 13, no. 9, p. e0203894, 2018.

[114] T. Afrin and N. Yodo, "A concise survey of advancements in recovery strategies for resilient complex networks," *Journal of Complex Networks*, vol. 7, no. 3, pp. 393–420, 2019.

[115] A. Majdandzic, B. Podobnik, S. V. Buldyrev, D. Y. Kenett, S. Havlin, and H. E. Stanley, "Spontaneous recovery in dynamical networks," *Nature Physics*, vol. 10, no. 1, p. 34, 2014.

[116] F. Chaoqi, W. Ying, Z. Kun, and G. Yangjun, "Complex networks under dynamic repair model," *Physica A: Statistical Mechanics and its Applications*, vol. 490, pp. 323–330, 2018.

[117] W. Sun and A. Zeng, "Target recovery in complex networks," *The European Physical Journal B*, vol. 90, no. 1, p. 10, 2017.

[118] M. J. Alenazi, E. K. Cetinkaya, and J. P. Sterbenz, "Cost-efficient algebraic connectivity optimisation of backbone networks," *Optical Switching and Networking*, vol. 14, pp. 107–116, 2014.

[119] C. Natalino, A. Yayimli, L. Wosinska, and M. Furdek, "Infrastructure upgrade framework for content delivery networks robust to targeted attacks," *Optical Switching and Networking*, vol. 31, pp. 202–210, 2019.

[120] S. Hong, J. Zhu, L. A. Braunstein, T. Zhao, and Q. You, "Cascading failure and recovery of spatially interdependent networks," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2017, no. 10, p. 103208, 2017.

[121] J. Wang, C. Qiao, and H. Yu, "On progressive network recovery after a major disruption," in *2011 IEEE INFOCOM Proceedings*. IEEE, 2011, pp. 1925–1933.

[122] K. Al Sabeh, M. Tornatore, and F. Dikbiyik, "Progressive network recovery in optical core networks," in *2015 7th International Workshop on Reliable Networks Design and Modeling (RNDM)*. IEEE, 2015, pp. 106–111.

[123] D. Z. Tootaghaj, H. Khamfroush, N. Bartolini, S. Ciavarella, S. Hayes, and T. La Porta, "Network recovery from massive failures under uncertain knowledge of damages," in *2017 IFIP Networking Conference (IFIP Networking) and Workshops*. IEEE, 2017, pp. 1–9.

[124] V. Latora and M. Marchiori, "Efficient behavior of small-world networks," *Physical Review Letters*, vol. 87, no. 19, p. 198701, 2001.

[125] P. Van Mieghem, *Data Communications Networking*, Delft, 2011, ISBN: 978-94-91075-01-8.

[126] S. Neumayer, G. Zussman, R. Cohen, and E. Modiano, "Assessing the vulnerability of the fiber infrastructure to disasters," *IEEE/ACM Transactions on Networking*, vol. 19, no. 6, pp. 1610–1623, 2011.

[127] S. Rai and D. P. Agrawal, *Distributed computing network reliability*. Los Alamitos, CA (USA); IEEE Computer Society Press, 1990.

[128] Y. T. Woldeyohannes and Y. Jiang, "Measures for network structural dependency analysis," *IEEE Communications Letters*, vol. 22, no. 10, pp. 2052–2055, 2018.

[129] P. Van Mieghem, K. Devriendt, and H. Cetinay, "Pseudoinverse of the Laplacian and best spreader node in a network," *Physical Review E*, vol. 96, no. 3, p. 032311, 2017.

[130] W. Ellens, F. Spieksma, P. Van Mieghem, A. Jamakovic, and R. Kooij, "Effective graph resistance," *Linear Algebra and its Applications*, vol. 435, no. 10, pp. 2491–2506, 2011.

[131] P. Van Mieghem, D. Stevanović, F. Kuipers, C. Li, R. Van De Bovenkamp, D. Liu, and H. Wang, "Decreasing the spectral radius of a graph by link removals," *Physical Review E*, vol. 84, no. 1, p. 016101, 2011.

[132] S. Knight, H. Nguyen, N. Falkner, R. Bowden, and M. Roughan, "The internet topology zoo," *Selected Areas in Communications, IEEE Journal on*, vol. 29, no. 9, pp. 1765 –1775, october 2011.

[133] P. Van Mieghem, *Performance analysis of complex networks and systems*. Cambridge University Press, 2014.

[134] C. Alcaraz and S. Wolthusen, "Recovery of structural controllability for control systems," in *International Conference on Critical Infrastructure Protection*. Springer, 2014, pp. 47–63.

[135] C. Alcaraz and J. Lopez, "Safeguarding structural controllability in cyber-physical control systems," in *European Symposium on Research in Computer Security*. Springer, 2016, pp. 471–489.

[136] S. Zhang and S. D. Wolthusen, "Iterative recovery of controllability via maximum matching," in *2017 13th IEEE Conference on Automation Science and Engineering (CASE)*. IEEE, 2017, pp. 328–333.

CURRICULUM VITÆ

Peng Sun received the bachelor's degree in communication engineering and the master's degree in communication and information system from Shandong University, in 2012 and 2017. Since October 2017, he has been a Ph.D researcher in the Network Architectures and Services (NAS) group, faculty of Electrical Engineering, Mathematics and Computer Science, Delft University of Technology, under the supervision of Prof. Piet Van Mieghem and Prof. Robert Kooij.

His research focuses are robustness and reliability of networked systems, recoverabiity of complex networks.

LIST OF PUBLICATIONS

7. A. Chen, **P. Sun** and R. E. Kooij, *The recoverability of network controllability*, 5th International Conference on System Reliability and Safety (ICSRS 2021), 24-26 November, Palermo, Italy.

6. **P. Sun**, R. E. Kooij and R. Bouffanais, *Controllability of a class of Swarm Signalling Networks*, in preparation.

5. **P. Sun**, R. E. Kooij and P. Van Mieghem, *Reachability-based Robustness of Controllability in Sparse Communication Networks*, IEEE Transactions on Network and Service Management (2021).

4. A. Dhiman, **P. Sun**, R. E. Kooij, *Using Machine Learning to Quantify the Robustness of Network Controllability*, Machine Learning for Networking - Third International Conference, MLN 2020, Springer, p. 19-39.

3. **P. Sun**, Z. He, R. E. Kooij and P. Van Mieghem, *Topological Approach to Measure the Recoverability of Optical Networks*, Optical Switching and Networking, 100617.

2. Z. He, **P. Sun** and P. Van Mieghem, *Topological approach to measure network recoverability*, The 11th International Workshop on Resilient Networks Design and Modeling (RNDM 2019), 14-16 October, Nicosia, Cyprus. (*received the 2019 James P. G. Sterbenz best paper award.*)

1. **P. Sun**, R. E. Kooij, Z. He and P. Van Mieghem, *Quantifying the Robustness of Network Controllability*, 4th International Conference on System Reliability and Safety (ICSRS 2019), 20-22 November, Rome, Italy. (*received the best presentation award*)