Technische Universität Dresden

Resilience of the Critical Communication Networks Against Spreading Failures - Case of the European National Research and Education Networks

Goran Murić

der Fakultät Elektrotechnik und Informationstechnik der Technischen Universität Dresden

zur Erlangung des akademischen Grades

D O K T O R I N G E N I E U R

(Dr.-Ing.)

genehmigte Dissertation

Vorsitzende:	Prof. Dr.-Ing. habil. Christian Georg Mayr
Gutachter:	Prof. Dr.-ing Eduard Jorswieck
	Prof. Dr. Jesus Gomez-Gardeñes
	Prof. Dr. Nataša Gospić

Tag der Einreichung:	29.06.2017
Tag der Verteidigung:	23.08.2017

September 2017

TECHNISCHE UNIVERSITÄT DRESDEN

Abstract

Fakultät Elektrotechnik und Informationstechnik

Doktoringenieur

Resilience of the Critical Communication Networks Against Spreading Failures - Case of the European National Research and Education Networks

by Goran Murić

A backbone network is the central part of the communication network, which provides connectivity within the various systems across large distances. Disruptions in a backbone network would cause severe consequences which could manifest in the service outage on a large scale. Depending on the size and the importance of the network, its failure could leave a substantial impact on the area it is associated with. The failures of the network services could lead to a significant disturbance of human activities. Therefore, making backbone communication networks more resilient directly affects the resilience of the area. Contemporary urban and regional development overwhelmingly converges with the communication infrastructure expansion and their obvious mutual interconnections become more reciprocal.

Spreading failures are of particular interest. They usually originate in a single network segment and then spread to the rest of network often causing a global collapse. Two types of spreading failures are given focus, namely: *epidemics* and *cascading failures*. How to make backbone networks more resilient against spreading failures? How to tune the topology or additionally protect nodes

or links in order to mitigate an effect of the potential failure? Those are the main questions addressed in this thesis.

First, the epidemic phenomena are discussed. The subjects of *epidemic modeling* and *identification of the most influential spreaders* are addressed using a proposed Linear Time-Invariant (LTI) system approach. Throughout the years, LTI system theory has been used mostly to describe electrical circuits and networks. LTI is suitable to characterize the behavior of the system consisting of numerous interconnected components. The results presented in this thesis show that the same mathematical toolbox could be used for the complex network analysis.

Then, cascading failures are discussed. Like any system which can be modeled using an interdependence graph with limited capacity of either nodes or edges, backbone networks are prone to cascades. Numerical simulations are used to model such failures. The resilience of European National Research and Education Networks (NREN) is assessed, weak points and critical areas of the network are identified and the suggestions for its modification are proposed.

Acknowledgements

Completing a PhD is a rewarding process which requires dedication and sacrifise not only by the candidate. I am grateful to my wife Maja who have provided me through moral and emotional support. None of this would be possible without her continuous encouragement throughout my years of study.

I would like to thank my thesis supervisor Prof. Dr.-Ing Eduard Jorswieck of the Faculty of Electrical and Computer Engineering at TU Dresden. He supported me through the whole research and steered me in the right direction whenever he thought I needed it. I would also like to thank my mentor Dr.-Ing. Christian Scheunert for his support especially at the beginning. Special thank goes to my second supervisor Prof. Dr. Nataša Gospić who was there for me since I was a bachelor student and who motivated me to pursue the PhD at the first place.

A very special gratitude goes out to Leibniz Institute of Ecological Urban and Regional Development for helping and providing the funding for the work and particularly to the institute's director Prof. Dr. h.c. Bernhard Müller.

I would also like to acknowledge all the people from the IOER and DLGS who supported me logistically and morally, including Paulina Schiappacasse and Sabine Scharfe and all the colleagues from the office: Sara, Shikha, Nakul, Jiaying, Linh, Neelakshi, Amsalu, Hanna, Martin, Stefan, Nelya, Julia, Benjamin, Patrick...

Finally, I must express my very profound gratitude to my parents and to my sister for providing me with unfailing support and continuous encouragement throughout my life. This accomplishment would not have been possible without them.

Contents

Abstract iii

Acknowledgements v

Contents vi

List of Figures xi

List of Tables xiii

Abbreviations xv

1 Introduction 1

1.1 Introduction to Networks 1
1.1.1 Terms and notations 3
1.1.2 Communication networks 6
1.1.3 Backbone networks 7
1.1.4 Complex and complicated systems 8
1.2 Resilience from Engineering Perspective 9
1.2.1 Four Dimensions of Resilience 14
1.2.2 The resilience as a function of time 16
1.3 Network Resilience 17
1.4 Introduction to Network Failures 20
1.5 Spreading Failures 22
1.5.1 Cascading failures 24
1.5.2 Epidemics in networks 24
1.6 Scientific Contribution of the Thesis 25

2 Theoretical Background **27**
2.1 Modeling Networks . . . 27
2.1.1 Random Graph Models . . . 28
2.1.2 Small-World Property . . . 30
2.1.3 Scale-Free Property . . . 31
2.2 Centrality Measures . . . 35
2.3 Network Resilience . . . 37
2.4 Spreading Failures . . . 41
2.4.1 Epidemics in Networks . . . 42
2.4.2 Cascading Failures . . . 46

3 Problem Statement and Methods **51**
3.1 Problem Statement . . . 51
3.2 Methods . . . 56
3.2.1 The methods for modeling . . . 57
3.2.2 The methods for analysis . . . 57
3.2.3 Metaheuristics . . . 58
3.2.4 Obtaining acyclic graphs . . . 59
3.2.5 The methods for design . . . 61
3.2.6 Analytical Approach in Complex Networks Modeling . . . 61
3.2.7 Numerical Approach . . . 69
3.3 European NRENs . . . 70

4 LTI System Theory and Spreading Phenomena in Networks **75**
4.1 Linear Time-Invariant Representation of Networks . . . 76
4.2 Modeling Epidemics by Virtual Network Expansion . . . 82
4.2.1 Almost Certain Transmission, p=1 . . . 83
4.2.2 Uncertain Transmission and Virtual Network Expansion 88
4.2.3 Numerical Simulations . . . 91
4.2.4 Network Data . . . 95
4.2.5 An Example of Network Analysis . . . 96
4.3 Identifying the Influential Spreaders . . . 99
4.3.1 Calculating the NiR . . . 100
4.3.2 Small Network Example . . . 102
4.3.3 Simulation Results . . . 103
4.3.4 Reasoning Behind . . . 108

4.3.5 Spreading Models . . . 112
4.3.6 Limits of the NiR . . . 113
4.3.7 Network Data . . . 114
4.3.8 Conclusion . . . 117

5 Cascading Failures Analysis Within the European NREN 119
5.1 Motter-Lai Model . . . 123
5.2 The Most Critical Nodes . . . 124
5.2.1 Individual Failure . . . 126
5.2.2 Multiple Simultaneous Failures . . . 130
5.2.3 Geographically Correlated Failures . . . 138
5.2.4 External Risks to European NRENs . . . 143
5.3 Crucitti-Latora-Marchiori Model and Simulation Results . . . 151
5.3.1 Single Node Failure . . . 153
5.3.2 Multiple Nodes Failure . . . 158
5.4 Active (Costless) Protection Strategies . . . 161
5.4.1 Removing Nodes . . . 161
5.4.2 Removing Nodes in the Vicinity of Failure . . . 165
5.5 Passive (Costly) Protection Strategy . . . 169
5.6 Limitations of the Proposed Models . . . 174
5.7 Conclusion . . . 175

Bibliography 177

List of Figures

1.1 System's disruption and bounce-back 12
1.2 Resilience state space 14
1.3 Damage network caused by the various attack strategies 23

2.1 Random rewiring procedure in Watts-Strogatz model 32
2.2 Simulation of epidemics following SI model 44

3.1 Types of edges in a graph 64
3.2 System states of a line graph after a virus infection 65
3.3 Three networks with associated system states 66
3.4 NREN topology in Germany 71
3.5 NREN topology in Germany and Poland 73
3.6 Complete European NREN topology 74

4.1 Small directed network and its corresponding adjacency matrix 79
4.2 Minimum Spanning Tree in a small network 84
4.3 Simple graph and corresponding system response 86
4.4 Virtual network expansion 88
4.5 Probability mass function of the geometric distribution 90
4.6 Spreading dynamics in the original vs. extended network 92
4.7 Spreading simulation vs. system response 94
4.8 Simple network and corresponding LTI step responses 97
4.9 Network protection example 98
4.10 The NiR of all nodes in the network 102
4.11 Nodes in the small network classified by the importance 104
4.12 Correlation of NiR and centrality measures: SI model 106
4.13 Correlation of NiR and centrality measures: SIR model 107
4.14 Corr. between centralities and spreading potential for various p 109
4.15 Expected time of infection and step response: small networks . 110
4.16 The comparison of the step response and the infection time . . 111

5.1 Removing nodes with various betweenness centralities 125
5.2 Cascade anomaly in the European NRENs case 127
5.3 The most critical nodes for various α 129
5.4 Multiple failures - two nodes removed 131
5.5 The sets of n critical nodes 134
5.6 Various sets of five critical nodes 136
5.7 Circular cuts on the map 138
5.8 Critical areas 140
5.9 Location of the most critical areas 141
5.10 Distribution of critical area damage 142
5.11 Map of earthquake magnitudes used in the SHARE model . . . 144
5.12 Combined risk 147
5.13 Averaging the M_W 148
5.14 Two-dimensional analysis - earthquakes and circular zones . . . 150
5.15 Efficiency deterioration in the case of a single node failure . . . 154
5.16 Impact of α on network efficiency after node removal 155
5.17 Effect of the random failure on the network's efficiency 156
5.18 The effect of the removal of three nodes with different loads . . 158
5.19 The most critical nodes within Europen NRENs 159
5.20 Multiple failures - two nodes removed 160
5.21 Effect of the selective removal of least loaded nodes 163
5.22 Nodes candidates for intentional removal 164
5.23 Protective zone 166
5.24 Protective zones identified 168
5.25 Congested nodes in the case of a single failure 171

List of Tables

3.1 Relevant research questions for the spreading failures 53

4.1 Generated and extracted networks 115

5.1 Impact of individual nodes removal 128
5.2 The solution space for n simultaneous failures 132
5.3 Critical groups 137
5.4 The most important protective hubs 172

Abbreviations

ABM	**A**gent **B**ased **M**odel
AC	**A**lgebraic **C**onnectivity
AMF	**A**verage value of **M**aximum **F**low between all pairs of nodes
ARPANET	**A**dvanced **R**esearch **P**rojects **A**gency **Net**work
AS	**A**utonomous **S**ystem
ATTR	**A**verage **T**wo **T**erminal **R**eliability
BA	**B**arabási-**A**lbert scale-free network model
BIBO	**B**ounded-**I**nput **B**ounded-**O**utput
CLM	**C**rucitti-**L**atora-**M**archiori model
DDoS	**D**istributed **D**enial of **S**ervice
ER	**E**rdős-**R**ényi random graph model
GA	**G**enetic **A**lgorithm
IC	**I**ndependent **C**ascade model
ICT	**I**nformation and **C**ommunications **T**echnology
LT	**L**inear **T**hreshold model
LTI	**L**inear **T**ime **I**nvariant
MANET	**M**obile **a**d hoc **N**etwork
ME	**M**aster **E**quation
MIMO	**M**ultiple **I**nput - **M**ultiple **O**ut
MFST	**M**aximum **F**low **B**etween nodes s and t
NC	**N**etwork **C**riticality
NiR	**N**ode **i**mposed **R**esponse
NREN	**N**ational **R**esearch and **E**ducation **N**etwork
NSFNET	**N**ational **S**cience **F**oundation **Net**work
OSI	**O**pen **S**ystems **I**nterconnection model
SHARE	**S**eismic **H**azard **H**armonization in **E**urope
SI	**S**usceptible **I**nfected model
SIR	**S**usceptible **I**nfected **R**emoved model
SIS	**S**usceptible **I**nfected **S**usceptible model
TEC	**T**otal **E**xpected **C**apacity
UPS	**U**ninterruptible **P**ower **S**upply
WSD	**W**eighted **S**pectral **D**istribution

Dedicated to networked humanity...

Chapter 1

Introduction

1.1 Introduction to Networks

The Networks play a significant role in our everyday lives. We are surrounded by various types of networks, and we are the part of many of them as well. Omnipresent communication networks such as World Wide Web or cellular phone networks define the way we live and interact with other people. Furthermore, various real-world entities and processes demonstrate the networked structure, starting from the systems already designed as a network such as electric power grids, highway systems, road networks to the less obvious networks defined in more abstract space such as social networks or collaboration patterns between individuals. Other examples of structures which could be described as networks include organizational networks, business relations between companies, neural networks, protein interaction networks and others.

Infrastructures consist of various networks such as highways, ports, electrical and information and communication networks. Urban, regional and national infrastructures contribute to the improvement in quality of life, supporting the economic and social growth. Furthermore, all those networks are connected between each other, making it an infrastructural network of networks. An example of a critical part of one region's infrastructure is the communication network. It is not an exaggeration to say that ICT has become vital to the functioning of modern-day society. Therefore, the underlying communication

system has the vital role in supporting the variety of improvements which enhance quality of life, including the increasing the efficiency of economic and social systems, raising productivity, and conserving energy. The result of the urbanization, modernization, industrialization, and other forms of progress is an evergrowing demand for more integrated infrastructures. It subsequently requires even better understanding of complex dependencies between all infrastructural elements.

The first systematic analysis of networks in the form of the mathematical graph theory originated in year 1735, when Euler solved famous problem of Seven Bridges of Königsberg[1]. The solution is considered to be a first valid proof in the graph theory and therefore in the network theory as its subgroup. The mathematical foundation of the graph theory paved the way for its application in some more specialized context, as for instance in medicine or social science. One of the early usages of the graph theory in medicine is related to the epidemic spreading problem. In 1927, W. O. Kermack and A. G. McKendrick [2] created a Susceptible Infected Removed (SIR) model for epidemic spreading, which became a base for further research in that area. Social network analysis is one of the areas where network theory is extensively used. Mathematically based quantitative analysis of social interactions dates back to the early 1920s. There has been an intensive cross-interaction between research across various scientific disciplines as sociology, anthropology, psychology and statistics with its own specialized journals as *Social Networks* [3]. Following the rapid evolution of online based social network platforms, the research on the topic gained a new momentum. For many companies, development of new efficient marketing solutions through influential figures in social networks became an important aspect of their business. Therefore, the need for an identification of central individuals within the network revived the research on centrality measures introduced by Freeman in 1979 [4]. The network communication follows a certain process of mass communication. Thus,

[1]The Königsberg bridge problem asks if the seven bridges of the city of Königsberg, over the river Preger can all be traversed in a single trip without doubling back, with the additional requirement that the trip ends in the same place it began. This problem was answered in the negative by Euler [1].

finding the "influential" or "central" person in the network in order to feed them with the information for further transfer to "followers" is of a great practical value [5].

1.1.1 Terms and notations

In order to deal with the topic of network resilience, some of the most commonly used terms and notations from the network science are listed. The following terms are mentioned multiple times throughout the thesis, hence the proper definition for each of them is necessary. Some of the definitions are a part of a general knowledge from graph theory and some of them adapted from the well known works of Newman [6] and Diestel [7]:

Vertex/Node: The fundamental unit of a network which represents a single non-dividable entity.

Edge/Link: The connection of two vertices. It is referred to as a link in computer science. The edge is usually depicted as a line in graphical representation of the network.

Directed/Undirected: An edge is directed if the goods or information runs in only one direction (such as a one-way road), and undirected if it runs in both directions. Directed edges are often represented as a sporting arrows indicating their orientation. An undirected edge might be represented as a directed edge with two directions. A graph is directed if all of its edges are directed.

Component: The component to which a vertex belongs is a set of vertices that can be reached from it by paths running along edges. In a directed graph a vertex has both an in-component and an out-component, which are the sets of vertices from which the vertex can be reached and vertices which can be reached from it.

Degree: The number of edges connected to a vertex. Note that the degree is not necessarily equal to the number of vertices adjacent to a vertex, since there may be more than one edge between any two vertices. Let $G = (V, E)$ be a non-empty graph. The set of neighbors of a vertex v in G is denoted by $N_G(v)$, or briefly by $N(v)$. The *degree* of a vertex v is the number $d = |E(v)|$ of edges at v. This is equal to the number of neighbors of v. In directed networks there is a difference between *indegree* $d_{in}(i)$, showing the number of immediate links directed towards the node i and *outdegree* $d_{out}(i)$ counting the number of links directing away from the node i. For the undirected networks there is one degree measure $d(i) = d_{out}(i) = d_{in}(i)$.

Paths : A path can be considered as any possible way from one vertex to another which might include other edges and vertices but without cycles. A *path* is a non-empty graph $P = (V, E)$ of the form

$$V = \{x_0, x_1, \ldots, x_k\}\ E = \{x_0x_1, x_1x_2 \ldots, x_{k-1}, x_k\},$$

where the x_i are all distinct. The number of edges in a path is its *length*.

Geodesic path: A geodesic path is the shortest path through the network from one vertex to another. Sometimes referred to as the *geodesic distance* it is the number of edges in a shortest path. Note that there may be and often there are more than one geodesic path between two vertices.

Cycles: If there is a path which starts and ends with the same vertex, that path is considered to be a cycle. Cycles are sometimes called *circuits* when closed path is specified in cyclic order but no first vertex is explicitly identified.

Diameter : The diameter of a network is the length, measured in number of edges, of the longest geodesic path between all pairs of vertices.

Centrality : Centrality measures in general try to identify the most important (central) vertices in the graph. Centrality concepts were initially developed in social network analysis. There are many centrality measures in use, and each of those measures has its value in certain applications. It is common

thing that centrality which is optimal for one application is often sub-optimal for another. That is the main reason for an existence of numerous different centralities: *Degree centrality*, *Closeness centrality*, *Betweenness centrality*, *Eigenvector centrality*, *Katz centrality*, *Alpha centrality* and others. For more details about the centrality measures, see Section 2.2.

Connected graph: If there is a path between every pair of vertices the graph is connected. In a connected graph, there are no unreachable vertices.

Edge weight: Each edge of a graph can have an associated numerical value called *weight*. In practice, the weight or cost of an edge is a measure of the length of a route, the capacity of a line or the cost required for data transport, etc. Usually, the edge weights are non-negative values from the set of integers or real numbers between 0 and 1.

Node weight: Similarly to the edge weight, an assorted numerical value to the node (vertex) is called *node weight*. The weight of the node could represent the capacity of the node, routing priority, etc.

Clustering coefficient: A clustering coefficient is a measure of the degree to which nodes in a graph tend to cluster together. In directed networks the clustering coefficient C_n of a node n is calculated as $C_n = l_n/(k_n(k_n - 1))$, where k_n is the number of neighbors of n and l_n is the number of connected pairs between all neighbors. The clustering coefficient represents the ratio between the number of edges between the neighbors of n and the maximum number of edges between the neighbors of n.

Component: The *component* or *connected component* is a subgraph of a greater graph in which all vertices are connected and none of them is connected to any of the remaining vertices from the greater graph.

Giant component: A *giant component* is a connected component of a graph which size is relative and constant compared to the greater graph.

1.1.2 Communication networks

At the beginning of the 20th century, the network theory has been primarily in use in social sciences and medicine. Technological progress introduced the development of interconnected communication devices such as large telephone networks and the computer networks. The expansion of large-scale computer networks and finally the Internet has put a great challenge to the scientific community dealing with the network theory. The requirements for more sophisticated and advanced tools emerged. The research and development in the field of network theory, together with the advancement in electronics have simultaneously led to the highly interconnected world as we know it today.

The pace of the urban and regional development is highly dependent on the telecommunication infrastructure. Setting up a reliable telecommunication networks goes hand in hand with all other infrastructural projects. Nowadays, in the high-tech society, people's lives have become more and more dependent on communication networks, either for business or leisure purposes: all financial transactions are conducted using telecommunication networks, industrial processes are dependent on information exchange, science is unimaginable without the cluster computing over the network. The society and communications technology are tied up like never before. The affiliation is so ingrained that society's daily procedures depend on the reliable communications such as railway signaling systems, traffic control and many other vital services.

Moreover, this dependency is expected to grow considering the new emerging technologies and services such as smart-cities, cloud computing, e-health, the internet of things, and MANETs [8]. The region development is highly associated with its communication infrastructure. For this reason, communication networks are considered one of the critical national infrastructures upon which society depends on [9]. Hence, it is imperative that communication networks should be robust enough to withstand failures and designed to respond adequately to damages and attacks [10].

1.1.3 Backbone networks

One of the most important components of all telecommunication networks is its backbone. A backbone network, or network core, is the central part of a telecommunication network which provides various services to customers who are connected by the access network. Backbone networks are designed to provide reliability and performance of large-scale communications across large distances. There is a huge amount of digital information created each second, and thus the demand for high-capacity data storage and processing locations constantly rises. These networks usually exploit optical technology to carry huge aggregated data as optical networks have demonstrated a capability to satisfy the demands. All higher layer services are highly dependent on the backbone network, as they are unable to operate without the physical links which backbone networks provide.

The backbone network corresponds to the lowest level physical topology which consists mostly of fiber cables[2] and all devices necessary for routing the high loads of traffic such as ADMs[3], layer-2 switches[4] and core routers[5]. The primary focus of previous studies has been on the logical aspects of the internet [11]. Since the backbone networks operate mostly on the lower OSI layers, the topic of the network resilience was often neglected. However, considering the fact that all higher level network services depend on backbone network, thus the study of physical connectivity is an important area of the research, especially modeling challenges dependent on the topology and geographical distribution of the network, such as power failures and severe weather conditions [12].

[2] Copper cables are still in use in certain networks, but mostly due to the traditional legacy. The backbone cabling is mostly fiber already.

[3] An add-drop multiplexer (ADM) is an element of an optical fiber network which combines, or multiplexes, several lower-bandwidth streams of data into a single beam of light.

[4] Layer 2 switches work on the data link layer of the OSI model.

[5] A core router is designed to support multiple telecommunications interfaces of the highest speed and must be able to forward IP packets, which means it can operate on the network layer of the OSI model.

Regarding the amount of complexity and scale of optical backbone networks and the mentioned dependency, in the event of a disaster, communication networks might suffer huge data loss and interruption of high-bandwidth channels, causing disruptions of essential services for weeks and severely complicate recovery operations. These outages may affect many applications/services, irrespective of the importance of the service or sensitivity of the carried data. Therefore, it is crucial to understand the vulnerability of backbone networks to disasters and design appropriate countermeasures and risk mitigation strategies [13].

1.1.4 Complex and complicated systems

In the literature which deals with networks theory, the term *complex networks* often appears. Naturally, the questions emerges: "What is *complex network* exactly?", "Why the networks are referred to as *complex*?".

In order to answer those questions, it is necessary to clarify the main distinction between the *complex* and *complicated* systems. Let us consider a simple system, for example a spring. The behavior of the spring is described by well-known laws (equations) and therefore it is fully predictable. We can easily predict the position or a shape of the spring in any given moment. To spice up things a little bit, we can add one more spring on top of the first one. The task of describing the behavior of such system is not that simple any more. However, it is still possible just with more effort. Adding more springs makes the analysis of such a system increasingly harder up to the point when finding analytical exact solution becomes impractical. Then we can consider the resulting system to be *complex*.

Another example is a system with more elements working together, for example a modern car, with all of its subsystems. It is very hard to assemble all parts and form a controllable system with such a specific function as driving. Modern car could be considered as a *complicated* system. But it is still not a *complex* one. It is not a *complex* system, because the behavior of car is still

highly predictable. If certain part of the engine is broken or the wheels are missing we know it will not be able to move any more.

The system becomes *complex* when we are not able to predict its behavior with given certainty anymore. The *complex* systems usually emerge as the spontaneous outcome of the interactions among many constituent units [14]. This implies that by observing a single constituent element one would not be able to describe the system as a whole, since its self-organizing principles are formed according to the collective and unsupervised dynamics of many elements. It is not easy to come up with a single definition of *complex* systems. Amaral and Ottino suggested a following definition: *A complex system is a system with a large number of elements, building blocks or agents, capable of interacting with each other and with their environment* [15]. The main difference between *complicated* and *complex* system is that former are planned following certain blueprint where each element has its own purpose and latter is made by evolving without the centralized plan.

The majority of real-world networks fall within the scope of the definition of *complex systems*, as they usually grow in time following complicated, not structured, decentralized rules. Measuring even the simple properties of complex systems could be challenging. Therefore, predicting more complicated behavior or measuring the resilience of such systems could be extremely demanding. Systems with a huge number of components interacting trivially are explained by statistical mechanics, and systems with precisely defined and constrained interactions are the concern of fields like chemistry and engineering. In so far as the domain of Complex Systems Science overlaps these fields, it contributes insights when the classical assumptions are violated [16].

1.2 Resilience from Engineering Perspective

The resilience is a term usually used to describe an ability of a certain entity to bounce back or recover from a shock. Therefore, it is used throughout time

across many scientific disciplines to explain the particular process within the area of its expertise. It is often used in psychology to describe the complex processes within the human mind when the individual or a group of people recover from a trauma. It is used in economy as well, as markets which recover fast from a crisis are referred to as a resilient.

According to the [17], the introduction of the resilience to the scientific world came through the seminal work of Holling named "Resilience and Stability of Ecological Systems" in 1973. However, the quick search could show us even older papers dealing with this term. For example, the work of Zakharov[6] published in the U.S.S.R. in 1965 deals with the resilience in polymers. In recently published work, D.E. Alexander [18] was dealing with an etymology of the word *resilience* through the scientific history. Although the term has been existing for many centuries and was used in various but similar contexts in art, literature, law, science and engineering, the work of Holling brought it to prominence in the modern scientific community, especially within the ecological sciences.

Even if the meaning of the term itself is clear for the most of the people, there are various interpretations of the resilience regarding the scientific area. Naturally, the psychologist would have a different approach from the economist, as the people behave in different way from the markets. Anyway, there are some underlying principles behind this concept that should be accepted universally regardless of the specific scientific area. The engineering perspective to the system resilience in its essence is trying to establish this ubiquitous principles, which could be adapted and expanded regarding the needs.

The most general engineering view of the resilience could be regarded as a resilience of a system. By observing a hypothetical system with measurable inputs and outputs throughout the time, we should be able to measure its resilience. Measuring the resilience could be performed by studying the system's

[6] S.K. Zakharov, L.I. Medvedeva, I.A. Arbuzova, Ye.V. Kuvshinskii, *The softening, high-elastic resilience and structure of crosslinked copolymers of methyl methacrylate and styrene with diolefinic monomers*, Polymer Science U.S.S.R., 1965

delivery function over time. But, in order to measure it, we have to define it first. A time dependent quantifable metric for a general system resilience is introduced relatively recently in the work of Henry and Ramirez-Markez in 2012 [17]. Their initial formulation of resilience comply with the basic concept of the word "resilience", which describe the ability of a system to "bounce back", and it is interpreted as a ratio of recovery over loss at time. The basic formula is as following:

$$\Re(t) = Recovery(t)/Loss(t) \tag{1.1}$$

Where $\Re(t)$ is a resilience of a system at time t.

From the resilience perspective the system S has three main states in which it can operate:

- S_0 - Original state (undisrupted)
- S_d - Disrupted state
- S_f - Stable recovered state

Furthermore, there are two main events responsible for triggering the mid-processes that will lead the system from one state to another, and those are: a disruptive event and resilience action.

Measuring the system as a whole actually has no practical meaning. Hence the specific function of a system should be measured, and the resilience of system should be evaluated based on the specific values of a system function over time. The quantifiable "figure of merit" should be selected, so it could represent the main operational level of a system. The *system service function* or *system delivery function* $\varphi(t)$ describes the behavior of the system over time as any state of S should be represented by the value of $\varphi(t)$ at time t. For example, $\varphi(t)$ could describe traffic flow, the number of active nodes or a delay.

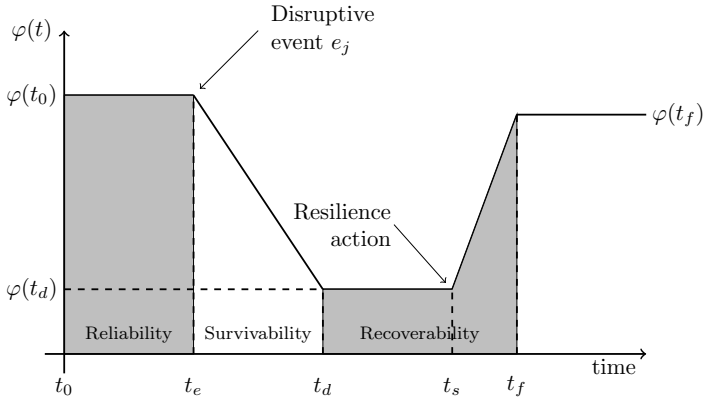


FIGURE 1.1: **System's disruption and bounce-back.** A change of the system's state transition over time in the case of a disruption. The value of a system's service function $\varphi(t)$ varies over time. It deteriorates after a disruptive event e_j and bounce-back after the resilience action. Three main dimensions of the resilience which affect different phases of the process are identified: *reliability*, *survivability* and *recoverability* (adapted from [17] and [19]).

In order to quantify the resilience of a system, the system service function $\varphi(t)$ of a hypothetical system changing over time is depicted in Figure 1.1. It is shown how the *original system state* is getting disrupted by a certain disruptive event e_j in time t_e causing the system function $\varphi(t)$ to degrade during the period of length $t_d - t_e$. The system enters the *disrupted system state*, where the system service function value remains on the level $\varphi(t_d)$. The systems stays in this degraded state until the moment t_s, when resilience action is carried out. The resilience action restores the system to the new equilibrium level of $\varphi(t_f)$. The final value of observed system service function is not necessarily the same as the value on the beginning (before the disruptive event). The value of $\varphi(t_f)$ could be lower or even higher, because the resilience action could bring the system to another stable operational level. According to this description, the basic general formula for a system resilience is derived [17, 19]:

$$\Re_F(t_r|e^j) = \frac{\varphi(t_r|e^j) - \varphi(t_d|e^j)}{\varphi(t_0) - \varphi(t_d|e^j)}, t_r \in (t_s, t_f) \tag{1.2}$$

This approach could be considered as a simple one, as we observe how a single service function of a system changes over time. Ideally, we would like to have one simple measure $\Re$ preferably bounded in the range $0 \leq \Re \leq 1$ which could be able to fully describe the resilience of a system. Unfortunately, the systems are not always simple, and most of the real world systems have two or more service functions. Those functions could be highly dependable, and measuring the single one would not provide an objective assessment. Thus the multicriteria analysis have to be used to estimate the real resilience level of a system with more than one dependable service functions.

Furthermore, the resilience could be modeled as a two-dimensional state space where the vertical axes represent the measure of a system function when the operational state is challenged. In Figure 1.2 it is shown how the system goes from the acceptable function level S_0 to the degraded level S_c. Through the process of *remediation*, service function is improved to the level S_r, and finally recovered back to S_0. A remediation is triggered by an alarm which turns on when the value of the service function becomes too low. For example a remediation action could be a change in the routing policy in the case of the communication network congestion, or modification of the firewall configuration in the cases of DDoS attack. In the case of the latter, the newly configured firewall blocks the traffic considered to cause the denial of service. In this approach, the resilience of a particular system is measured as an area under this trajectory [20].

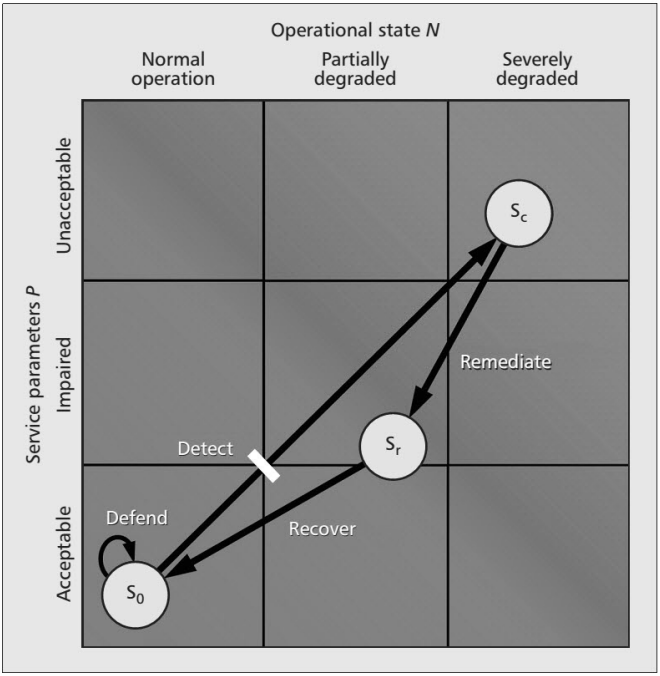


FIGURE 1.2: **Resilience state space.** The resilience is modeled as a two-dimensional state space. The system moves from the initial function level S_0 to the degraded level S_c. Through the process of *remediation*, service function is improved to the level S_r, and finally recovered back to S_0 [20].

1.2.1 Four Dimensions of Resilience: Reliability, Vulnerability, Survivability, and Recoverability

Besides the three dimensions of the resilience depicted in the Figure 1.1: reliability, survivability, and recoverability, a vulnerability is additionally recognized. Reliability is defined as "the ability of a system or component to perform its required functions under stated conditions for a specified period of time" [21]. In the absence of the significant external or internal disruptive event, the system in the period of $t_e - t_0$ is characterized by its *reliability*.

Reliability is highly coupled with the *availability* of a system - as reliability increases, so does availability. Most generally, the availability is described as a ratio of *uptime* of a system to the sum of the *uptime* and *downtime*.

In fact, the *reliability* (and so the *availability*) describes the system while it is in a stable (undisrupted) state. Although the system behavior after the disruption highly depends on its previous state, the measure of the reliability has no significant influence in the assessment of its resilience. The main reason is that reliability is actually the measure used for regular or expected behavior, and the resilience comes after the irregular, unexpected event.

Vulnerability shows in which extent the system is unable to withstand the negative effects of the hostile environment. On the other hand, *survivability* quantifies the ability of a system to "survive", that is to maintain operational level of system function when the system could be recognized as still functional. The survivability is very similar to the notion of *robustness*, which is often used in the literature as a synonym. Although the *vulnerability* and the *survivability* are two distinctive concepts, they are highly related. In general, one can say that vulnerable system has lower chance of survival. In terms of previously described concept of resilience and referring to the Figure 1.1, the vulnerability and survivability levels are relevant for the time period $t_d - t_e$.

Finally, the *recoverability* of a system quantifies the ability of the system to get back to the state prior to disruption or to the next equilibrium level. In the Figure 1.1 the time span $t_f - t_d$ is relevant for the recoverability. In (1.2) the value t_r belongs to the set (t_s, t_f). In the case when $\varphi(t_r|e^j) = \varphi(t_d|e^j)$, the minimum value of $\Re_F(t_r|e^j)$ is reached, and it equals 0. In this case, the system has not recovered and there is no "bounce back" effect. On the other hand, if the $\varphi(t_r|e^j) = \varphi(t_0)$, the resilience value $\Re_F(t_r|e^j)$ equals 1, which means that system has fully recovered.

1.2.2 The resilience as a function of time

In previous chapter, to quantify the *ability* of the system to get back to the previous state, the single measure $\Re$ is proposed. The $\Re$ actually compares the values of system service function in regards to the initial value: prior the resilience action and at the certain point of time. One could intuitively argue if this fairly simple measure could describe the broad concept of resilience *over the time*. In (1.2) the time dimension is not considered, although the *time* in which system reaches the desired level is of a certain importance. More precisely *recoverability* does not quantifies just an ability but the *speed* in which the system will get back to the desired state. The recoverability is directly connected to the measure of the resilience as quickly recoverable systems could be considered as more resilient. Since the recoverability is regarded as a *speed* of returning system back to the previous state, the time span $t_f - t_d$ is important for the assessment of the recoverability. As the difference between t_f and t_d becomes smaller, the time needed for a system to fully recover is shorter.

Furthermore, the(1.2) does not include other time-related aspects of the resilience as it considers the resilience exclusively as a measure of recoverability. The important point it misses to describe is a temporal process succeeding the disruptive event. Right after the disruptive event at the moment t_e, the system function starts deteriorating until it reaches the lowest value $\varphi(t_d|e^j)$ at the time t_d.

Let us consider two systems F_1 and F_2 with observable system functions φ_1 and φ_2. The system functions values of both systems at the time t_e are equal. A disruptive event e^j occurs at the same time, affecting both system equally at the initial moment. Both system functions start to decline until they reach the minimum value of $\varphi(t_{d1})$ and $\varphi(t_{d2})$ at the time t_{d1} and t_{d2} respectively. In the first hypothetical situation $\varphi(t_{d1}) > \varphi(t_{d2})$ and $t_{d1} = t_{d2}$, that is the system F_1 resists the full deterioration of a system function. Therefore it positions itself in a state where further restoration actions are able to restore

the system function value faster. Second situation is when $\varphi(t_{d1}) = \varphi(t_{d2})$ but $t_{d1} > t_{d2}$, that is the system function of F_1 deteriorates slower and gives more time to the possible countermeasures which will stop degradation. In both hypothetical circumstances, we can say that system F_1 is more resilient from a system F_2. Although this cases are not in the scope of (1.2), this basic equation gives us a solid mathematical basis for quantification of certain aspects of resilience.

1.3 Network Resilience

Communication networks have become essential part of the people's daily routines and therefore modern society largely relies on its proper functioning. From the geographical point of view, the communication networks could be designed to provide services on a local as well on the global level. Generally, the larger networks (networks which cover wider geographical area and therefore connecting more users) are considered more important (critical) than smaller ones. The most prominent example of most widely used communication network is the global Internet. The more detailed justification of its importance nowadays become superfluous. It is worth mentioning that Internet is considered as a part of national critical infrastructure and today, when the probability and severity of natural disasters and other threatening events have increased, the protection of communication networks is on the national agenda in many countries [22]. On the other hand, an importance of numerous smaller networks on national or regional level shouldn't be underestimated. There are many private or public owned networks independent from global Internet which provide various services like national telephone networks, military networks or communication networks connecting universities or scientific institutions.

The network could be generally considered as a system with its measurable system functions. The function of a communication network is transfer of

data, hence the most common measures of a system function of a communication network are related to the data quantity transferred through the network. However, the network resilience in general is usually not assessed by a single variable. Multiple control and state parameters of a multi-dimensional complex system make the prediction of the system's resilience difficult. However, there is an analytical framework which allows us to collapse the behaviour of different networks onto a single universal resilience function by systematically separating the roles of the system's dynamics and topology. The formalism proposed in [23] reduces A_{ij} into an 1D system. It is shown that the patterns of the resilience depend only in system's intrinsic dynamic, regardless of the specific topology or weights. All the parameters are condensed in a single β_{eff} and indicate that density, heterogeneity and symmetry are the three key factors to define the system's resilience. This approach provides a tool for an accurate prediction for the system's response to various perturbations. Here, some of the most important measures which could serve as a system functions of a communications networks are discussed.

Throughput. The amount of data which could be successfully delivered over the communication network per certain time slot is called *throughput*. Throughput is commonly measured in amount of data (bits) per second, or in data packets per time slot. The data may be delivered over the physical or logical link and it has to pass through certain number of links and nodes. Therefore, the *throughput* can be measured in the node or on the link. Additionally, one can measure the throughput of the full path from node n to node m or an average throughput of the whole network, etc.

Connectivity. A graph (network) is connected if there exists at least one possible path between any pair of nodes, which means that each node can communicate with any other node in the network. In disconnected graphs, this condition is not met. Disconnected graph is consisted of two or more independent connected graphs. The measure called *connectivity* is defined as a minimum number of elements (nodes or edges) which can be removed to make connected graph disconnected. This measure does not say much about

the throughput or possible congestions in the network. The measure which largely coincides with connectivity was a subject of one of the earliest mathematical proofs in the network theory. In year 1927 Karl Menger showed that the number of node-independent paths between two vertices is always exactly equal to the minimum number of other vertices in the network that must fail in order for those two vertices to become disconnected from each other [24, p. 424]. Let us consider a path $P_1(V_1, E_1)$ from node i to node j in certain network. The path consists of sets of nodes V_1 and links E_1. If there is another path $P_2(V_2, E_2)$ which does not share any common element with path $P_1(V_1, E_1)$, except the first and last node, those paths are described as *node-independent*. The number of node-independent paths in certain network could be used as a indicator of its robustness. This is almost exactly the same as a minimum number of vertices which has to fail in order to make a network disconnected, just applied to the particular pair of nodes. Sometimes, the simple connectivity is not enough to explain robustness or survivability of the real world networks. Therefore, some other measures are introduced which focus particularly on *spatially correlated* or *region based* failures within the network. The *region based connectivity* is introduced as a measure which shows the minimum number of nodes (links) that have to fail within any region of the network before it is disconnected. This measure takes into account not only the topology, but the network's geometry. As an extension of the region based connectivity, there is *region-based component decomposition number (RBCDN)* which measures the number of connected components in which the network decomposes once all the nodes of a region fail [25].

Network diameter. Another quantitative measure of topological robustness of the network is *network diameter*. Diameter of a network is defined as a length $d = max_{u,v}d(u, v)$ of a longest shortest path between any two vertices in the network. The *average diameter* is sometimes referred to as a diameter. Average diameter is described as an average length of the shortest paths between any two nodes in a network. If the diameter d is smaller, the nodes are able to communicate between each other more easily. Larger number of nodes does not necessarily mean the network has a large diameter.

Additional measures. Even the throughput and connectivity could be considered as the most important measures of network's operational function, there are other measures which could provide additional information about the network itself. These measures could be particularly important in the times of the disruption and undesired events. For example, the *number of active nodes* after the disruption, the *average congestion* in the network or the *number of infected nodes* in the case of malicious virus attack. Particularly important measure is a size of the *largest connected component* or the *largest connected subgraph*. If the node in the network is damaged (removed), it is usually considered that all associated links are broken. Sufficient number of such damages could make one initially connected graph disconnected, which means that the graph is split into two or more smaller pieces (subgraphs). The simplest measure of such an impact of the network is the relative size of the *largest connected component* remained from the network S_f/S_0 where S_0 is the original size of the network before the disruption. When $S_f \ll S_0$ the network has been broken into many small parts and therefore is not functional any more [14, p. 118].

1.4 Introduction to Network Failures

In the applied network setups typical failure events include cable cuts, hardware malfunctions, software errors, power outages, natural disasters (e.g., flood, fire, and earthquake), accidents, human errors (e.g., incorrect maintenance) and malicious physical/electronic attacks [10]. Since the networks are mathematically modeled as a graphs $G(V, E)$ with V vertices and E edges, all "real world" threats should be modeled in the way that the corresponding graph is damaged appropriately. For example, the single software error would be mapped to the model as a removed node or as a node with changed parameters. Natural disasters affect relatively large areas, and therefore disruptions like that are modeled as a geographically correlated failures. Usually, for such type of failures, a group of geographically close nodes are removed from the

graph. On the other hand, the malicious software attack on a network does not imply the removal of nodes in the mathematical model, but rather changing the states of a node in accordance to the specific spreading rule. The real impact of a certain damage on the network could be assessed by periodical evaluation of network function. There are various types of possible strategies for network damaging[7]. Regarding the type of the failure and the nature of the network it has to be decided which system function from the Section 1.3 should be used.

Single failure event means that only one element (link or node together with its associated links) is removed. Then, the certain network function is measured and the impact of node or link removal is assessed by the ratio of measured function before and after the disruption. Single failure could occur randomly, or could be a result of a targeted attack. The impacts of a random and targeted failure are usually very different. Albert et al. [26] were considering the damage to the network made by removing certain nodes, randomly at first. Then, they started removing nodes in a targeted manner in order to simulate the intentional malicious attack. For the latter, they chose the nodes according to their centrality. As a measure of network function they use the *average diameter* and *size of the largest component* of the network. Two types of the networks were separately assessed: Erdős-Rényi random graph and scale-free networks. The results of the experiment show that scale-free networks are more vulnerable to the deletion of high-degree vertices in comparison to the random graph whose vulnerability to vertices removal is almost independent on the type of nodes chosen to be removed. Therefore, the *targeted attacks* could lead to a rapid collapse of many "real world" networks which usually do not follow the random topology. The reason for that is network design which usually follows the principle of building the hubs, very central nodes with high degrees.

[7]Note that "damaging" network does not necessary mean a harmful activity. If we want to protect the network from possible malicious software attack and we are able to *immunize* certain nodes, the immunization would be considered as a removal of nodes from the network. In this case, the intention is to slow down the unwanted information spreading. On the other hand, the legitimate data should be transferred as quickly as possible.

Usually, single failures occur regularly following certain statistical distribution. The majority of networks are designed in a way that they are resilient to a certain frequency of failures. Despite the random expected failures, the network should remain functional. For example, a network could be designed to withstand just a single failure at the time. If the frequency of failures is $f = 1/t$, the nodes should be designed in a way that each failed node has to recover its functionality in less than t seconds. Otherwise, a second node might fail, which could lead to the unexpected cascading phenomena as covered in Chapter 5. This type of failures falls within the group of *multiple uncorrelated failures*. Those failures are named *uncorrelated* if any node might be affected following the common distribution and if the failure of one node does not affect the failure of another. Therefore, it is obvious that designing the networks which is resistant to the multiple uncorrelated failures requires more resources as there is a demand for more redundant links or additional nodes.

Another type of failures are *multiple correlated failures*. In the case of the uncorrelated failures, the failure of one network element has no impact on failure of another and all failures are not a result of a common event. On the other hand, when multiple network elements go down simultaneously and it is caused by a single common undesired event, those failures are considered to be correlated. The main causes of correlated failures are natural disasters, such as floods, earthquakes or storms. Natural disasters tend to be focused on a certain geographical area and therefore the failures they cause are geographically or spatially correlated.

1.5 Spreading Failures

Single failures within the communication networks are common, and most of those networks are designed to sustain such type of failures. Even the multiple random failures do not make significant damage on modern real world networks. One example is the Internet. A small fraction of network elements within the Internet are always non-functional. Still, it keeps on functioning

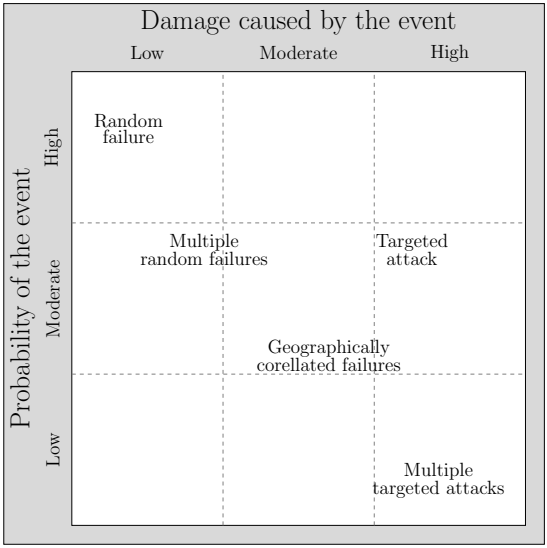


FIGURE 1.3: **The level of damage in the network caused by the various attack strategies.** The majority of man made networks show high level of robustness against the random failures, but high vulnerability to targeted attacks.

despite this level of failure [24]. Targeted attacks, however are the bigger issue, and could cause some significant outages. Beside these types of failures which are caused mostly by some external factor (e.g. undesired event causes one or more nodes to stop working), the failures could be a result of another failure within the network. Only small initial shock, like the breakdown of a single network element could lead to the effect of avalanche and cause the whole communication system to collapse [27]. Those failures which have a tendency to spread through the network are *spreading failures* and they could be both *cascading failures* or *epidemics*. In the Figure 1.3 the comparison of the potential damage to the network made by certain type of failures is depicted.

1.5.1 Cascading failures

The initial failure caused by a single undesirable event or by multiple correlated or uncorrelated events usually causes the chain of failures within the network, which is defined in the literature as a *cascading failure*. The cascading failures could be caused either by the malfunction of node(s) (caused by unintentional human error, random failure, natural disaster...) or intentional malicious electronic attack on network components which would lead to the erroneous element behavior. In communication networks as well in other networks (electrical grids or transport networks) the goods or data is transported through links and nodes which are limited by their own capacity. When failure occurs, the traffic is usually routed through secondary paths. The rerouting subsequently increases the load of associated elements in the network. For too large failure, alternative routes are not able to sustain an additional load, and associated elements become unresponsive. In this case, the failure would spread through the network causing large number of nodes to operate improperly. Furthermore it would consequently lead to the further congestion of the network and finally be manifested in a way that make the network not meeting its service specification or it might result in a complete network outage. More on modelling of cascading failures could be seen in Chapter 5.

1.5.2 Epidemics in networks

An object, commodity, substance or an idea could spread across the network in many ways. For example, the gas is transported through the pipes. In that case, the good (gas) is shipped from one point to another and the same good does not exists any more at its place of origin. For modeling this type of propagation, sometimes referred to as *conservative flow*, we can use diffusion equation or some other equation which describes the similar phenomena. On the other hand, in the case of the communication networks, the "good" which is transported is data. However, digital information is rather copied than

moved. Therefore, the conservative flow models are not suitable for this kind of propagation. Beside the simplest model known as breadth-first search, there are several well known models for representing the epidemics and those are described more thoroughly in Chapter 2.4.1. The data which spread across the network uncontrollably might be unwanted, and that is particularly the case with the malicious software inserted in the network or the virus within the population. In that case, researchers and practitioners are interested in designing the networks in such a way that epidemics do not escalate quickly. Nevertheless, some epidemics are desired, for example a quick spread of news through the large social networks. No matter which model is used, a failure is modeled as a "contagion" spreading over the links of a complex network, altering the "states" of the nodes as it spreads, either recoverable or otherwise.

1.6 Scientific Contribution of the Thesis

The Thesis deals with a phenomena of spreading failures. The main scientific contribution is twofold. First, the spreading failures are discussed from the system theoretic point of view. A Linear Time-Invariant (LTI) system approach is used for modelling *epidemics* and identification of the *influential spreaders*. The topic of LTI system theory and spreading phenomena in complex networks is analyzed in Chapter 4.

Epidemics. The term *epidemic* is often used to describe the spread of a virus within the human population. However, various processes in complex networks show a similar dynamic behavior. Usually the epidemics are analyzed in order to control or limit the infection. Sometimes, however, the goal is just the opposite, e.g. to boost the spread of an information through the communication network. There is a number of models which consider the population to be homogeneous [2, 28], where the connectivity (degree) of all nodes is considered to be equal. However, real networks such as social networks or autonomous system (AS) networks deviate from such homogeneity [29]. The dynamic of the epidemic is defined mostly by the connectivity

pattern, especially in networks with a power law degree distribution [30]. For some networks where the topology is fully disclosed, epidemics can be modeled without assumptions on the connectivity pattern and infections can be simulated on the actual topology. Usually, an agent-based method is used for the simulation of an infection. Here, the alternative approach, which utilizes the LTI systems toolbox is presented. Based on the paper *"On modeling epidemics in networks using linear time-invariant dynamics"* [31] it is shown as effective method for evaluating epidemic dynamics analytically in every time step, omitting agent-based simulation.

Influential spreaders. In order to control or prevent some of the spreading processes, it becomes an imperative to understand the role of certain network elements. Therefore, identifying the most important nodes in regard to the spreading phenomena emerged as an important area of research [32–34]. The solution presented in this Chapter applies the tools from the systems theory to identify the most influential potential spreader in the network. Based on the paper *"Using LTI Dynamics to Identify the Influential Nodes in a Network"* [35] the proposed *Node Imposed Response (NiR)* accurately evaluates node spreading power.

Secondly, the resilience of the European National Research and Education Network (NREN) backbone is assessed against the *cascading failures*. The NREN has a limited capacity of network elements such as nodes and links. Therefore, it is susceptible to cascading failures, which can originate in one point and spread through the network causing numerous breakdowns. In Chapter 5, the behavior of the NREN is examined under various attack strategies: *individual failure*, *multiple simultaneous failures* and *geographically correlated failures*. The external risks to the network are also assessed with the emphasis to the seismic risk. Furthermore, the protection strategies including *active* and *passive* ones which mitigate the cascade are proposed.

Chapter 2

Theoretical Background

The topic of network resilience is broad, interdisciplinary and it combines several approaches. The literature deals with assessment of the resilience[1] of the whole network, assessment of the importance of certain network elements such as nodes or links in terms of network resilience, evaluation of the metrics used in network characterization and finally the methods for addressing the issues detected by the assessments. The network resilience is strongly related to the epidemic spreading modeling and cascading failures.

The literature review presented here represents the most relevant findings regarding the network resilience with emphasis on those dealing with epidemics modeling and cascades. All of these methods were built on the top of the graph theory, so the most notable literature on that topic is listed here as well. Also, the papers that deal with the topic of percolation theory and centrality measures in general are listed and reviewed.

2.1 Modeling Networks

In order to appropriately examine the network behavior both analytically and numerically, the proper mathematical model should be used for network

[1] The resilience is often presented as a measure of vulnerability, reliability, survivability or the robustness. Although all of these terms are not the synonyms for the resilience, they present the subset of broader resilience concept.

formation. Since the real-world networks (communication networks, social networks, transportation networks, neural networks...) are quite diverse by its nature and function, and all of them are designed in a different way following various patterns and needs, the unique mathematical explanation of such complex forms is hardly achievable. Nevertheless, researchers were able to develop a set of models which are successfully used to represent most common real-world networks in the way that they show the similar properties. There are three main families of network models [24] which emphasize various network properties. There are *random graph models* which do not consider the rules of the real network genesis. On the other hand we have models which focus on generating the networks by mimicking *small-world* property or *scale-free* property of the real-world networks.

2.1.1 Random Graph Models

The random graphs models are one of the oldest and best studied models in graph theory. The family of those models originated from a series of papers published by Erdős and Rényi from 1959 onward [36]. There are several versions of the original Erdős-Rényi's (ER) model, although all of them are based on a rather simple principle of generating the edges between vertices in a random manner. The most popular model is one denoted as $G_{n,p}$ where the probability of edge existence between two vertices is p and the probability that there is no edge is $1 - p$. All probabilities are independent, and pairs of vertices are chosen uniformly at random.

The main characteristics of random graphs models is an absolute lack of knowledge regarding the rules of the network genesis. The random graph models are based on a simplest premise that connections between vertices are generated randomly. The main reasons why random graph models are used extensively are their simplicity and the fact that the properties of the ER networks are simple to solve analytically. Since certain networks are often described by its general properties, this characteristic is convenient. For

example, the average degree $\langle k \rangle$ is calculated from the number of edges $\langle E \rangle$ generated in a graph $\langle E \rangle = \frac{1}{2}N(N-1)p$. Since each edge connects two vertices it is a part of a degree calculation for both of them. Therefore, we have

$$\langle k \rangle = \frac{2\langle E \rangle}{N} = (N-1)p \simeq Np. \tag{2.1}$$

In ER model, it possible to derive other network properties such as *clustering coeficient* or *average shortest path length* very easily. By using the $G_{n,p}$ model to generate graph, the resulting graph will have a *binomial degree distribution* and in the case when n is large it becomes the *Poisson distribution*. That is the main disadvantage of this model as in many real-world networks those distributions could not be easily detected. The *degree distribution* is very important property of a network which affects its behavior in many ways [24]. In order to overcome this weakness of ER model, various authors started to generate graphs with different degree distributions. Molloy and Reed [37] proposed an algorithm for generating random graph following preferred degree distribution. They assign the already fixed degree sequence to the graph and generate it in the way such that each vertex has a degree k_i from the set of already formed sequence K and $i = 1, ..., N$. They also showed that it is possible to define the random graph with any degree distribution.

The networks are usually not homogeneous, which means that they consist of various types of nodes. Each node could have its own characteristic, and regarding each node's property, it could be assigned to a different "type". In many networks the nodes of a same type are more likely to form links between each other. For example, the social links between people are usually established based on people's preferences, occupation or social status. The work of Söderberg [38] is based on the idea of a "fitness" value distributed to each node in the network. Based on the assigned "fitness", the edge is formed between each two nodes (i, j) following the certain probability $p_{i,j} = f(x_i, x_j)$ where f is a given function describing the "attraction" between the nodes of

various types. In the case where there is only one type or f is constant, this model produces the ER graph.

Another group of graph models which fit in a random graph family are called *Exponential random graphs*. Holland and Leinhardt [39] introduced an approach at first aimed at the studies of social networks, which was extended by a numerous authors later on, applying the similar solution to other areas. The main idea behind the exponential random graphs model is defining the distribution of probabilities that one graph could be formed in many possible ways. The network of N nodes could be realized in $\binom{N(N-1)/2}{E}$ possible ways. There is a certain distribution of probabilities of these realizations which is developed from the comparison to the real data. The model represented here is de facto random, but with the high probability that any random generated topology resembles the real data.

2.1.2 Small-World Property

In order to avoid the main drawbacks of the random network models, Watts and Strogatz [40] came up with a proposal of the model which will generate the networks which are more similar to the real ones. There are two measures in which real-world and randomly generated networks differ and the Watts-Strogatz (WS) model could be used to solve that. First, it is a *clustering coefficient* (page 5). The values of the clustering coefficients of the real-world networks are generally higher than the randomly generated networks. And secondly, the *average shortest path* (page 4) length of a real network is shorter. That means the small-world networks are highly clustered and have a short path lengths at the same time. The original WS model starts with a ring of N vertices $(i_1, i_2, ..., i_k, ..., i_N)$. Each vertex gets connected with $2m$ nearest neighbors. For example, the vertice i_k will have the links to the 4 vertices: $i_{k-1}, i_{k-2}, i_{k+1}$ and i_{k+2} as shown in Figure 2.1. In the next step, the probability p is introduced. The edge connected to a clockwise vertex of each node is reconnected with the probability p to a random node in the

network, excluding the starting node. This process ensures the probabilistic nature of a resulting network while the number of edges stays constant.

Shortly after the Watts and Strogatz published their model, Barthélemy and Amaral [41] studied the same model more in details. They have focused on a probability value p and finding the critical value of p at which the small-world behavior might be observed. Furthermore they have showed that the actualization of the small-world behavior is not a phase transition but a crossover phenomena. They found out that the smallest value of p needed to transform initial network to the small-world network is related to the size of the network. The relation is presented in a form of a scaling function

$$l(n,p) \sim n^* F\left(\tfrac{n}{n^*}\right)$$

where $F(u \ll 1) \approx u$, $F(u \gg 1) \approx \ln(u)$ is a scaling ansatz[2] and the $l(n,p)$ is the average distance between two vertices in a network with n nodes connected with a probability p. The value n^* is a function of p and it represents the crossover size above which the network behaves as a small world.

The Barrat and Weigt [42] showed the analytical approach for calculating the degree distribution of such networks. Even if a degree distribution remains almost the same as in the random graphs, the clustering coefficient and shortest path lengths have been dramatically changed even for the small values of p.

2.1.3 Scale-Free Property

The *degree distribution* (page: 4) of nodes in networks generated by previously described models sometimes differ from the reality. It is shown that for a number of systems, including the World Wide Web, citation networks and social networks, the degree distribution follows the power law. That is,

[2] Ansatz is an assumed form for a mathematical statement that is not based on any underlying theory or principle.

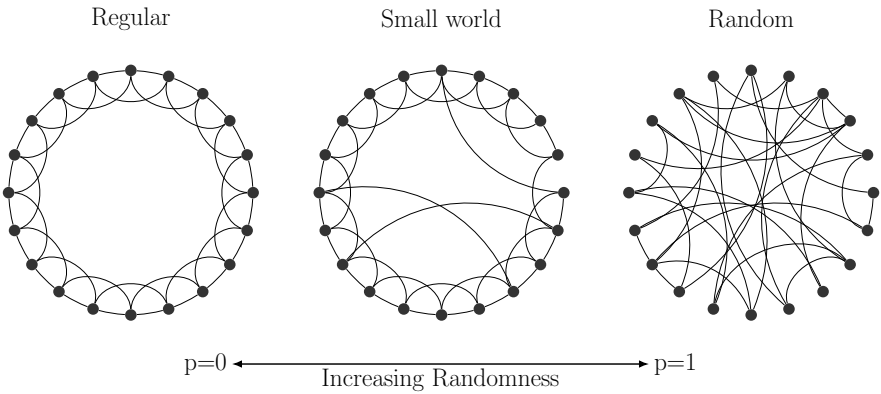


FIGURE 2.1: **Random rewiring procedure in Watts-Strogatz model.** Random rewiring procedure for interpolating between a regular ring lattice and a random network, without altering the number of vertices or edges in the graph. Inspired by [40] and adapted.

the distribution of degrees $P(k)$ of nodes in the network with k connections to other nodes follows the power law function, as defined in (2.2) where γ represents a parameter with a value usually in a range $2 < \gamma < 3$.

$$P(k) \sim k^{\gamma} \tag{2.2}$$

The model and algorithm for generating such type of networks called *scale-free networks* is introduced by Barabási and Albert [43]. In order to recognize the causes for such degree distribution in many real networks, the paper questions the very nature of network genesis. Networks do not emerge instantly, but come out as a result of growth throughout time following certain criteria. How do the majority of networks form? Which criteria people follow individually in their social interactions, which results the social network at the end? Which unwritten rules internet service providers adopt while connecting with others or designing their own network?

The answer on those questions lies in a process known as a *preferential attachment*. The networks are not static, but form dynamically. New vertices

have been attached to the existing ones but with the different probabilities. The insight behind this approach is the fact that newly arrived vertices tend to attach to those vertices in the network which are already highly connected (the nodes with a large degree). This mechanism is rather known as a rich-get-richer phenomenon, the Gibrat principle or cumulative advantage [14]. For example, the person with many social connections is more likely to acquire new friends than a person with a narrow social circle. The same principle apply for other networks as well. Barabási and Albert [43] propose a model which will generate the network with a degree distribution which follows the power law. In the proposed solution, the network starts with a small core of m_0 vertices. The network grows with every time-step by adding one vertex with m edges ($m < m_0$) connecting it to the existing network. The other end of each edge is connected to a node in network chosen randomly with a probability which is proportional to its degree. The probability Π that a new vertex will be connected to already existing vertex i depends on a degree k_i of the vertex i in a way that

$$\Pi(k_i) = \frac{k_i}{\sum_j k_j} \tag{2.3}$$

The authors in [43] have simulated the network formation with and without the preferential attachment and showed that the absence of preferential attachment eliminates the scale-free feature of the network. Therefore, the model they proposed indeed generates the network with scale-free characteristics. At the beginning, the difference in degrees between vertices is small. As the time elapses, the difference becomes larger since the vertices with a higher degree attract more edges from newly arrived nodes, leading eventually to a formation of vertices which are highly connected at the expense of the others. The rate at which each vertex acquires edges is:

$$\frac{\delta k_i}{\delta t} = \frac{k_i}{2t} \tag{2.4}$$

so we can calculate the number of edges k added to one vertex i at any point of time t_i

$$k_i = m\sqrt{\left(\frac{t}{t_i}\right)} \tag{2.5}$$

The probability density $P(k)$ obtained for a long time period leads to the stationary solution

$$P(k) = \frac{2m^2}{k^3} \tag{2.6}$$

This means that the value of an exponent γ equals 3, which is one of the flaws of this particular model, since in the real networks the parameter γ can vary. Barabási and Albert were using just one parameter to define preferences for edge attachment which is node degree. One can assume that there are other measures which have an influence on edge creation or process of rewiring. A degree is considered as one of many *centrality* measures which indicate the importance of the nodes. This led to a number of papers dealing with the preferential attachment and the nature of node's "attractiveness" for a connection. Bianconi and Barabási [44] proposed a *fitness* measurement μ chosen randomly according to some distribution $\rho(\mu)$. The *fitness* might depend on many factors and it represents the more general approach to the previously described model. The probability that newly formed edge will be connected to the node s is

$$\Pi_s = \frac{\mu_s k_s}{\sum_j \mu_j k_j} \tag{2.7}$$

By adjusting the fitness value for the node, one can achieve desired exponent γ, as not all systems follow the "rich-get-richer" principle, but in this case "fittest-get-richer". Therefore, the fitness might be a combination of many properties of a node summed up in a single value.

Klemm and Eguíluz [45] have proposed a model which combines the properties of small world and scale-free networks. They define and analyze a model for network self-formation which creates the network with high clustering

coefficient, small path lengths and scale-free distribution of degrees at the same time.

In order to develop more advanced tools for network optimization and protection, the fundamental step is to understand the underlying principles of their formation. This section presented a small set of solutions. Those are focused on creating the model which generates the networks which accurately represent the networks evolved more-less naturally in the real world. None of these models could generate the real-world networks exactly, but they are able to create graphs with similar properties. Therefore they are able to provide the mathematical basics for analyzing such networks on more applicable level. By properly generating networks we get a pool of experimental "playgrounds" which could help us to understand more complex network properties such as *network resilience*.

2.2 Centrality Measures

The various measures are introduced in early 70's in sociology mostly describing the relations between people. Later, similar measures are used to describe other networks, especially after the development of computer networks. The crucial thing was to identify the most important nodes within the network. The family of those measures are called *centrality measures* and they were developed to determine in which extent a certain node is centrally positioned regarding the topology. Centrality measures are the fundamental tool in assessing all the networked structures.

Degree of a node is the number of edges (links) incident to the node [7]. In directed networks there is a difference between *indegree* $d_{in}(i)$, showing the number of immediate links directed towards the node i and *outdegree* $d_{out}(i)$ counting the number of links directing away from the node i. For the undirected networks there is one degree measure $d(i) = d_{out}(i) = d_{in}(i)$. Degree is the simplest yet most robust measure of the node importance. The

degree can not accurately capture the node influence in networks consisting of large clusters divided by the nodes with low degree, but in most cases degree accurately identifies the most important hubs.

Betweenness represents the number of shortest paths from all nodes to all others that pass through a particular node. The value is usually normalized in the range $[0, 1]$. The betweenness centrality b_k of the node k is defined as follows [46]:

$$b_k = \sum_i \sum_j \frac{g_{ikj}}{g_{ij}} \tag{2.8}$$

where g_{ij} is the number of geodesic paths from node i to node j, and g_{ikj} is the number of geodesic paths from i to j that pass through k. There are some variations in the betweenness centrality based on the various approaches of defining the most desirable path. In some cases the constraints in the network make geodesic path not desirable, since it could be too costly (e.g. congested, expensive etc.). The actual betweenness centrality of the node is then modified taking in account also the weights of the links.

Coreness is the centrality measure derived from the k-core (also called k-shell) decomposition process of the network. The k-core is the largest subgraph comprising nodes of degree at least k [47]. A k-core of the graph can be obtained by recursively removing all the nodes of degree less than k, until all nodes in the remaining graph have at least degree k. The coreness c_i of a node i is k if the node belongs to the k-core but not to the (k+1)-core [48]. By observing the coreness measure, we can identify the best individual spreaders in the network if the spreading process originates in a single node [49].

H-index, or Hirsch index, was originally used to measure the citation impact of the author. The H-index concept was later extended to quantify the importance of the node in the network. The H-index of a node is defined to be the maximum value h such that there exists at least h neighbors of degree no less than h [50]. It is interrelated to *coreness* and the *degree*, and it outperforms both measures in several cases.

Dynamics sensitive centrality integrates the topological features and the dynamical properties at the same time [51]. While the all other centrality measures used for comparison rely solely on the topological features, DS introduces two parameters, β and μ, representing the rate of the infection and the rate of the recovery respectively. The DS centrality is therefore particularly suitable for identifying the most influential spreaders when the SIR epidemic model is concerned. However, to properly assess the node's importance one has to know the spreading dynamics parameters in advance.

Besides the most important centrality measures which are used throughout the thesis for node assessment and comparison, there are others, tailored for various dynamics. For example, Opsahl, Agneessens et al. [52] described the most important centrality measures and proposed a generalized centrality measure which combines the aspects of weights and number of ties. Newman [53] has proposed a new betweenness measure that counts essentially all paths between vertices. The measure is particularly useful for finding vertices of high centrality that do not happen to lie on geodesic paths or on the paths formed by maximum-flow cut-sets. Dolev, Elovici and Puzis [54] defined the Routing Betweenness Centrality (RBC) measure which generalizes previously well known Betweenness measures such as the Shortest Path Betweenness, Flow Betweenness, and Traffic Load Centrality. Other important papers regarding the centrality measures relevant to my research area could be found within the references [46, 55–59].

2.3 Network Resilience

The concept of a network resilience is not clearly defined. The terms and definitions are overlapping and while some authors use different terms to explain the same phenomenon others use the same terms to describe different events. This fuzziness makes categorizing and analyzing of the *resilience literature* from this scientific area demanding. The majority of authors deals with one particular phase in a whole process which characterizes the resilience,

which usually corresponds to the four dimensions of the resilience described in Section 1.2.1. Therefore, there are many research projects not dealing with the resilience per se, but with the phenomena such as *vulnerability*, *reliability*, *survivability* and *error tolerance* or *attack tolerance*. Furthermore, the authors have different opinions when it comes to the measure of network's function, so the system service function $\varphi(t)$ vary from paper to paper. Some authors observe the network's efficiency [60], some network's connectivity [26] and some use the number of centrality measures or the largest connected components [61, 62].

Crucitti, Latora et al. [60] research an impact of errors and attacks on certain networks regarding the network's efficiency. The concept of network efficiency was introduced earlier by the same authors [63]. Instead of the characteristic path length L and the clustering coefficient C the authors proposed a measure called *efficiency* of a network E which indicate how efficiently the information propagate through the network G. The *efficiency* is measured both on the local $E_{loc}(G)$ and global $E_{glob}(G)$ level. The measure of global and local efficiency is based on efficient communication between two nodes, which is inversely proportional to the shortest distance. They have shown that both local and global efficiency of scale-free networks drop rapidly if the nodes with high connectivity are removed, which means that those networks are more sensitive to intentional attack.

In one of the fundamental works in this area, Albert, Jeong and Barabási [26] discussed the network tolerance to the removal of nodes that play a vital role in maintaining network's connectivity. More specifically, they were measuring the diameter of various networks and observing the change in the measured function during the simulated attack. The main conclusion is that deliberately conducted attack on a network which damages the most connected nodes in the scale-free networks causes a big increase in diameter. If the top 5% of nodes with highest connectivity are removed, the diameter of a typical scale-free network is doubled. The removal of most connected nodes drastically alters the network's original topology and therefore prevents the efficient

information exchange.

Latora and Marchiori [64] developed a method to find the critical components of an infrastructure network. They proposed a method to evaluate the importance of an element of the network by considering the drop in the network's performance caused by its deactivation. They were using the network *efficiency* proposed earlier by Crucitti, Latora et al. [60] to evaluate the network performance to evaluate three real networks (two internet backbone networks and one transport network) and show that the damage of the most connected nodes, the hubs, is not always the worst possible scenario, as some other nodes could be more important in terms of the network survivability.

Holme et al. [61] have studied a behavior of a network under the various types of attacks. They examine the real-world[3] and artificially generated[4] networks. In their simulation, the networks had been attacked by different attack strategies, each of which based on the initial degrees (ID), the initial betweenness (IB), the recalculated degree (RD), and the recalculated betweenness (RB). The same procedure was repeated for the edges. The authors made a comparison of the effects of certain attacks on the networks calculating the change of the *normalized average geodesic path* $\widetilde{l^{-1}}$ and the *normalized size of the largest connected component* $\widetilde{S}$.

Mishkovski, Biey and Kocarev [62] used the normalized average edge betweenness to measure *network vulnerability*. They study networks generated artificially by most common models[5] and referring to them as a *synthetic networks*. They have managed to calculate the general *vulnerability index* for certain types of synthetic networks. Furthermore, the authors studied the several real-world networks[6] as well and compared them in terms of vulnerability. Among the synthetic networks, the most robust one (the least

[3] Scientific collaboration network and Computer network from Internet traffic

[4] Erdös-Rényi model of random networks, Watts-Strogatz model of small-world networks, Barabási-Albert model of scale-free networks and Clustered scale-free network model

[5] Erdös-Rényi model for random networks, Geometric random networks, Watts-Strogatz model for small-world networks and Scale free networks

[6] Human brain network, US power grid network, Collaboration network, Urban transport networks and EU power grid network

vulnerable) is the network generated by Watts-Strogatz model. Among the real-world networks, the least vulnerable one is the network of neurons which form the human brain.

Crucitti et al. [65] were dealing with an error and attack tolerance of the complex networks. They study two types of networks (*Erdős-Rényi random graphs* and *Barabási-Albert scale-free network*) and measure the effect in the case of removal of randomly selected nodes. More precisely, the targeted removal is considered as an *attack* and random removal is considered as an *error*. The targeted attacks were focused on nodes with high centrality, and the centrality measures used to identify most important nodes were *degree*, *betweeness* and the *recalculated betweeness*. In order to measure the network behavior during the disruption, they use the *global efficiency* measurement, which represents the average of the efficiency $\varepsilon_{ij} = 1/t_{ij}$ over all couples of nodes:

$$E(G) = \frac{\sum_{i \neq j \in G} \varepsilon_{ij}}{N(N-1)} = \frac{1}{N(N-1)} \sum_{i \neq j \in G} \frac{1}{t_{ij}} \tag{2.9}$$

The authors came up with more-less expected conclusion. The BA networks are very vulnerable to targeted attack, due to the power-law degree distribution, so the removal of very important nodes will cause the collapse of the network in early stage of an attack. On the other hand, the same networks are very resistant when it comes to random errors. In the case of ER networks the differences in network behavior during random or targeted attacks are almost non existing.

Cohen et al. [66] introduced a criterion for the collapse of certain networks under the random attacks. They observe Internet and the conclusion is that Internet topology is highly resilient to the breakdown of nodes. The same conclusion is valid for other networks which demonstrate the similar properties, particularly those networks which connectivity distribution is described by a power-law.

Doyle et al. [67] proposed an optimization-based approach for Internet modeling. They discuss "robust yet fragile" nature of the Internet, which is highly

robust to the perturbations for which it was designed but quite vulnerable to other perturbations. The Internet is generally prone to components failures or removals from the network. However, the protocols and feedback regulations enable some extraordinary robustness, scalability and adaptability even to a drastic network changes. On the other hand, the robustness comes together with the high level of vulnerability to failures caused by malicious actions or hijacking. They also question the scale-free model for Internet analysis and its applicability to highly evolved systems of unstructured, ensemble-based approaches.

Chen et al. [68] proposed a measure of network fragmentation. Authors define the fragmentation F as a relation between the actual number of links in the network compared to a number of all possible links. For systems close to or below criticality, F gives better precision for fragmentation of the whole system compared to P_∞, which is the connectivity of a fragmented network. The P_∞ is a ratio between the largest cluster size N_∞ (called the incipient order parameter) and N (called the infinite cluster), so the $P_\infty \equiv N_\infty/N$.

2.4 Spreading Failures

Various dynamical processes in networks tend to advance from a single origin to the rest of the network. For example, data in communication networks which is intended for broadcast originates from a single transmitter and reaches all other nodes. Often unwanted processes such as the failures are those which also spread. Here, we identify two main types of the spreading failures: *epidemics* and *cascades*. They both exhibit the same basic dynamical property of spreading. However, the causes and the mechanism of the spreading are different.

2.4.1 Epidemics in Networks

Epidemic spreading models originate from the biology, thus many of those models had been described much before the man made communication networks were invented. However, the models could be used to define the behavior of the computer virus in a similar way they define the spreading behavior of virus among humans. The pioneers in mathematical modeling of epidemics are Kermack and KcKendrick, with their work from the year 1927 where they introduced the Susceptible-Infected-Recovered (SIR) model [2]. Later on, many authors from the various scientific disciplines dealt with this issue, combining approaches from various fields such as biology, telecommunications or physics. There are numerous examples of applications where epidemic models may apply. Among others they include computer networks, social networks and power supply networks.

The simplest epidemiological model widely used Susceptible-Infected (SI) model. Each individual in the population can be in one of two possible states: infected or susceptible to infection. The probability that a susceptible node receives an infection from any neighboring infected node in very small time interval dt is βdt where β defines the *spreading rate*. What makes this model simple is that infected individuals remain infected permanently. Therefore, the evolution of epidemic in SI model is fully defined by the number of infected individuals in a point of time $i(t)$. In SI model, all nodes will be infected at the end of the process, but it is interesting to observe the rate in which the epidemics spreads regarding the topology of the network. The growth rate of number of infected individuals is presented in (2.10). It is proportional to the spreading rate λ, the density of susceptible vertices that may become infected, $s(t) = 1 - i(t)$ and the number of infected individuals. The evolution of epidemic following SI model in homogeneous network is presented as follows [69]:

$$\frac{di(t)}{dt} = \lambda \langle k \rangle i(t)[1 - i(t)] \tag{2.10}$$

The simplest situation presented here corresponds to a complete lack of degree information, which means that this equation counts only for the homogeneous networks where all vertices have almost the same number of neighbors, meaning the degrees of all vertices are similar. The solution for (2.10) is:

$$i(t) = \frac{i_0 e^{(t/\tau_H)}}{1 + i_0[e^{(t/\tau_H)} - 1]} \tag{2.11}$$

where the i_0 is the initial density of infected individuals and $\tau_H = (\lambda\langle k\rangle)^{-1}$ is the time-scale of infection growth. However, this calculation is only valid if the degree fluctuations are very small, $k \approx \langle k\rangle$. Many networks encountered in the real-world are far from homogeneous, therefore we face the networks with a heterogeneous connectivity pattern. That means that degree of vertices k is highly fluctuating and difference between degrees in the same network might be substantial. In this case, the SI model could be rewritten as:

$$\frac{di_k(t)}{dt} = \alpha[1 - i_k(t)]k\theta_k(t) \tag{2.12}$$

where the creation term is proportional to the spreading rate α, the degree k, the probability $1 - i_k$ that a vertex with a degree k is not infected, and the density θ_k of infected neighbors of vertices of degree k.

The models analytically describing the epidemic spread through the network could be validated by numerical simulations. Since there is a randomness taking part in the simulation process, the sufficient number of simulation experiments had to be carried out and the results have to be averaged. The (2.11) is checked for validity in [69]. The authors use a network of a size $N = 10^4$ and k ranging from 4 to 20. As an example of homogeneous complex network, the authors choose the ER network constructed from a set of N different vertices with $N(N-1)/2$ possible edges created with a probability p. This results in a random network with average degree $\langle k\rangle = pN$ and a Poisson degree distribution:

$$P(k) = e^{-k}\frac{\langle k\rangle^k}{k!} \tag{2.13}$$

The results of a simulation of SI model (Figure 2.2) show that (2.11) adequately approximates the results acquired during the simulation.

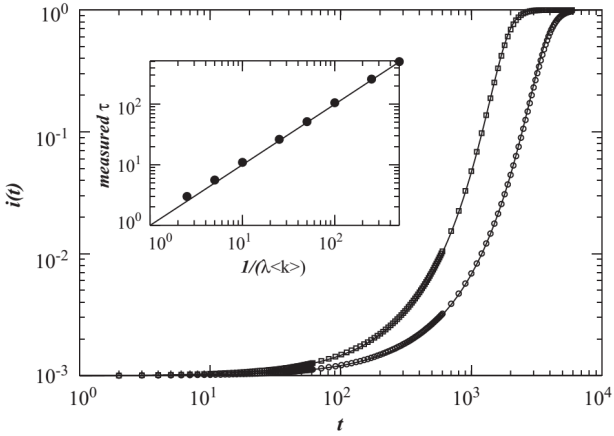


FIGURE 2.2: **Simulation of epidemics following SI model.** Main frame: the symbols correspond to simulations of the SI model with $\lambda = 10^{-4}$ on ER networks with $N = 10^4$, $\langle k \rangle = 20, 40$; the lines are fits of the form of (2.11). Inset: measured time-scale τ; as obtained from fitting, versus the theoretical prediction for different values of $\langle k \rangle$ and λ [69].

In reality, SI model might be used only in some specific cases. A Suspectible-Infected-Recovered (SIR) model, although fairy simple as well, could approximate the real epidemics more accurately. In the SIR model, each individual can be in one of three possible states: *susceptible*, *infected* or *recovered*. For a case of fixed population where $N = S(t) + I(t) + R(t)$ the main equations describing this model are:

$$\frac{dS}{dt} = -\beta \frac{S}{N} I \tag{2.14}$$

$$\frac{dI}{dt} = \beta \frac{S}{N} I - gI \tag{2.15}$$

$$\frac{dR}{dt} = gI \tag{2.16}$$

where S, I and R are number of susceptible, infected and recovered individuals respectively. The parameter β is a contract parameter and the $1/g$ is the mean infectious period. The rate at which the fraction S of individuals in the population which are in the susceptible state decreases is proportional to the number I of infected individuals who are able to transmit the disease $dS/dt = -\beta SI$ [24]. The results of model analysis are presented by Keeling [70] as follows:

1. An epidemic can only occur if $R_0 = \beta/g > 1$;
2. S is monotonically decreasing, R is monotonically increasing and I is unimodal;
3. The epidemic eventually dies out with some proportion of susceptibles, S_∞ remaining: $S_\infty = exp((S_\infty - 1) - R_0)$.

In the case of a malicious virus infection, the telecommunication networks act in the way where the state of one node is directly dependent on the state of its neighbor. Hence, the failure of a single node can cause a failure to spread across the network.

The neighboring node will fail when its load reaches a threshold which is chosen independently using some probability distribution. Since an attack on a small fraction of nodes has the potential to trigger a global failure [71], the fundamental issue is to develop the strategies of defense in order to prevent the failures from propagating through the entire network. Percolation theory is useful in modeling such phenomena.

Piraveenan, Prokopenko et al. [72] give a review of the most important centrality measures and discuss their relevance and importance. Some of the models for cascading failures and spreading of epidemics are shown in their work as well. The new centrality measure called *percolation centrality* was introduced. The other measures quantify the importance of a node in purely topological terms, and the value of the node does not depend on the *state* of

the node in any way. The percolation centrality on the other hand, measures the importance of nodes in terms of aiding the percolation through the network. A percolation state x_i^t of a node i can take any value $0 \leq x_i^t \leq 1$ in time t, where $x_i^t = 0$ indicates a non-percolated state at time t and $x_i^t = 1$ indicates a fully percolated state at time t. The authors show that percolation centrality measure becomes particularly useful in these scenarios when an early intervention is warranted.

By expanding our understanding of epidemics in the networks, we become closer to the solutions for the network protection. Holme [73] studied the dynamics of the epidemics under the various vaccination strategies. In many models, usually the vaccinated node is considered immune to the infection. In certain cases, the immunized node is considered to be immune to infection but able to transfer virus to other nodes. In this paper author study how a fraction of the population should be vaccinated in the most efficient way in order to stop the epidemic. He focused only on local vaccination strategies considering the argument that the global network information is rarely available.

Bauer and Lizier [34] introduced a new method for an approximation of the number of infections in the network of susceptible individuals if an initial infected node is given. They make use of the counting potential infectious walks, the paths the infection is probably going to take starting from the infected node. The method proposed in the paper can give the better estimation accuracy for the same computational cost compared to other methods.

2.4.2 Cascading Failures

Any system that can be modeled using an interdependence graph with limited capacity of either nodes or edges to carry flow will be prone to cascading failure phenomena [74]. There are number of models for the cascading failures described in the literature and few of them are mentioned here. The first one is the Motter-Lai (ML) model [75] which uses a simple traffic pattern between the nodes and includes the capacity of nodes. In ML model, each

pair of nodes in the network is considered to exchange one unit of the relevant quantity (information, energy ...) per time step. The quantity is transmitted through the shortest path connecting them. Each node is loaded by a number of shortest paths traversing it. Furthermore, each node has its limitations regarding the maximum load it can sustain, and it is called the capacity. It is assumed that the capacity C_j of node j is proportional to its initial load L_j,

$$C_j = (1 + \alpha)L_j, \quad j = 1, 2, \ldots N$$

where the α is *tolerance parameter*, and the N is the initial number of active nodes. The removal of nodes alters the topology of the network, and therefore the distribution of shortest paths. In such case some nodes might reach its capacity limits, and fail. After the initial breakdown, the rewiring occurs, which leads to the congestion of other nodes and results in further failures. This process is an example of *cascading failure*. It can stop after few time steps, but it can also propagate and compromise the whole network. The authors show that the heterogeneous networks[7] are resistant to the random failures. But, on the other hand, the cascade failure might easily occur in the event of a targeted attack. The attack on a single important node (the one with high load) may trigger a cascade of overload failures capable of disabling the network almost entirely.

Later on Crucitti, Latora and Marchiori [27] proposed a different model for cascading failures in complex networks. The model is based on a dynamical redistribution of loads. Initially, the network is in the free flow state where no nodes are overloaded. After the initial failure, the distribution of load changes and some of the nodes become loaded above the capacity. It causes further change of information routing, causing further congestion. Finally, the network reaches the new stable state. In contrast to the previously proposed ML model, in the case of congestion the nodes are not removed from the network, but their efficiency is deteriorated. As a system measure they use

[7] Most of the man-made networks and networks which evolve through time are considered to be heterogeneous, as the degree distribution is not random

the overall network efficiency and they prove that the efficiency of a network decay with the removal of nodes, particularly those with high initial load.

Both of the above-mentioned models are more thoroughly explained and simulated in the Chapter 5. However, there are some other models worth mentioning.

The Watts model [76] considers the nodes within the network to maintain the binary state either 0 or 1. The state of node is directly dependable on the state of its neighbors according to simple threshold rule. An individual node observes the current states (either 0 or 1) of k other nodes, called *neighbors*, and adopts state 1 if at least a threshold fraction ϕ of its k neighbors are in state 1, else it adopts state 0. The model is particularly interested in social networks where the interpersonal influence is important to observe. In that case, if the network is sufficiently sparse, the propagation of cascades is limited by the global connectivity and when it is sufficiently dense, cascade propagation is limited by the stability of the individual nodes.

Tran and Namatame [77] research a cascading failure phenomena on certain networks. This is a particularly interesting paper in terms of the resilience of the network to cascading failures. They were investigating the topological robustness of networks. In the topological approach the network is able to rewire itself to be an adaptive topology against cascade failure. They proposed two rewiring protocols, and applied them to capacity cascade model. The first one is Preserving Rewiring, in which a network is able to rewire itself but not affect its property and the second is Random Rewiring in which network change its topology toward random networks. The results indicate that proposed rewiring protocols can dramatically reduce the average size of large cascading failures.

Another two important diffusion models in computer science are the *independent cascade* and *linear threshold* models [78]. They are convenient to describe the spread of rumors or ideas within the social networks. Independent Cascade (IC) model generalizes the SIR model. Instead of a single probability

infection, each edge could have a different transmission probability. The probability $P_{u,v}$ is the probability of u infecting v. This probability can be assigned based on frequency of interactions, geographic proximity, or historical infection traces. In the Linear Threshold (LT) model, each node has its threshold θ_v in the interval $[0, 1]$. Additionally, each directed edge $(u, v) \in E$ has a non-negative weight $b(u, v)$. In each time step, the inactive node becomes active if the sum of all the incoming weights from the neighboring active nodes reach the threshold.

All the cascade models mentioned here have their applications in various networks. The particular model to be used to describe a diffusion process depends on the type of network and the specific dynamic. Although, all of the models could be used for modeling a process which spreads through the network, some of them are more suitable for a particular type of network. For example LT or IC models are mostly used for the social networks, while ML and CLM are common for modeling failures in technological networks characterized with loads, such as power grids.

Chapter 3

Problem Statement and Methods

3.1 Problem Statement

As the main topic of the research is oriented toward the protection of the critical communication networks, the central focus is on the resilience of the core part of all the communication networks, the backbone. Therefore, the main research question is: **How to make backbone networks more resilient against spreading failures?**

The primary goal of the research is the development of models and tools for making the backbone networks more resilient against spreading failures. Two main dynamics of spreading failures are addressed: *cascades* and *epidemics*. Both of them share two characteristics: they usually originate in a small fraction of nodes; they spread through the network and could cause global outage. However, the mechanism and the consequences of failures differ. The reason for cascade is capacity deficiency and the epidemics are driven by the spreading property of a virus. Furthermore, each of the failures are caused by a different trigger:

Cascade is triggered by the failure of the nodes or links which are caused either by the random failure, geographically correlated failures or an intentional attack. Although this type of failure is related to a part or section

of a graph, it could easily spread across the whole network. The reason for spreading has been known as a cascade effect. Cascading failures are a result of the increased traffic load in links and nodes after an initial failure. The nodes or links could not support the excess of load, and therefore, they deny traffic or even collapse. After an initial failure within the communication networks, the average traffic load increases. The increase of information flow leads to the effect known as *buffer overflow* which causes delays and service denial.

Epidemic is triggered by the malicious virus infection which is caused by the intentional, deterministic attack on well chosen network nodes with malware that spread to physically and logically connected neighbors. This also leads to cascading failures on a much faster time scale then cascades by node or link failures and buffer overflows. Simple failures are not the only hazardous effect in the case of malicious virus attack. The function of a malware can be of another harmful nature such as to compromise the data or to eavesdrop the confidential communication. Therefore, the adequate measures should be conducted to slow down the contagion process and evade the potential adverse effects on the network and the services provided by it.

In addition to Barker's model (Section 1.2), and in accordance with the work of Johansson and later Ouyang [79, 80], certain phases could be perceived in regard to the disruption of the system. In the Fig. 1.1 one can observe three main phases when it comes to the undesired event, and those are: *preparedness*, *response* and *recovery* phase. The choice of a strategy to make the network more resilient depends on the phase when the given strategy will be used. Based on this, some research sub-questions are formulated. The Table 3.1. represents the set of possible research questions and tasks relevant to each phase and the type of the failure.

In this thesis, the focus is on the following research questions from the Table 3.1:

Phase	*Cascades*	*Epidemics*
Preparedness	Which are the nodes or links that should be additionally protected to keep the network more resilient to cascades?	Which are the nodes that should be additionally protected to keep the network more resilient to epidemics?
	How to tune topology and improve the node properties in order to mitigate the effect of a potential failure?	How to tune topology and improve the node properties in order to slow down the spread of a potential virus infection?
Response	Develop strategies for active topology control in order to prevent further buffer overflows and local congestion.	How the topology of the network might be modified after the infection to slow down the contagion spreading?
Recovery	Prioritize nodes or links reparation in order to regain the functionality of the network.	Prioritize nodes or links reparation in order to regain the functionality of the network.

TABLE 3.1: **Relevant research questions for the spreading failures protection.** The small set of possible research questions relevant to each phase and the type of the failure.

1. Which are the nodes or links that should be additionally protected to keep the network more resilient to cascades?

2. How to tune topology and improve the node properties in order to mitigate the effect of a potential failure?

3. Develop strategies for active topology control in order to prevent further buffer overflows and local congestion.

4. Which are the nodes that should be additionally protected to keep the network more resilient to epidemics?

5. How to tune topology and improve the node properties in order to slow down the spread of a potential virus infection?

The focus is mostly to preparedness and response phase. Tackling wide range of all the problems throughout all the phases would be out of the scope of a single PhD thesis.

Addressing the Research Questions

More detailed information on methods that are used to address each of the research questions are listed below:

1. *Which are the nodes or links that should be additionally protected to keep the network more resilient to cascades?*

   A failure of one set of nodes could have a bigger impact on the network than the failure of some other set. Usually, the removal or failure of nodes or edges, either by random breakdown or intentional attack, within a stressed distributed system, will trigger a subsequent redistribution of stress within the system [75]. The goal is to identify the most important nodes regarding the impact of such failures on the network. The measures which are used to assess the network function are the *relative size of the largest connected component* in case of the Motter-Lai (ML) model (Section 5.1) and the *network efficiency* in case of the Crucitti-Latora-Marchiori (CLM) model (Section 5.3). Beside others, the main centrality measure exploited in both of the models is *betweenness centrality* (Section 2.2). The failures on the network are simulated and the impact on the network is measured after and before the failures. The disruption modeling is also modified depending on the cause of the failure, weather it is random failure, geographically correlated failure or intentional attack. In order to validate some of the results for the case of very large solution space, an appropriate genetic algorithm is used (Section 3.2.3). This way, the survivability of the network is assessed, and the most important nodes or group of nodes are identified.

2. *How to tune topology and improve the node properties in order to mitigate the effect of a potential failure?*

   The assessment of the network survivability gives us an insight in the distribution of important network elements. Some alteration in the network topology could be made to reduce the criticality of certain nodes and increase the overall survivability of the network. The link relocation or addition is not considered as a feasible strategy in infrastructural network such as communication backbone. Therefore, the focus is on the improvement of node properties, particularly the *capacity* of node. Numerical simulations are used to evaluate the strategies for optimal capacity increase.

3. *Develop strategies for active topology control in order to prevent further buffer overflows and local congestion.*

   The strategies for active topology control involve the real time modifications in links and nodes arrangement, which are triggered by the initial event. The objective of the alteration process is to mitigate the effect of the failure which already happened. It should be done quickly and efficiently. Creating new links and nodes in an already established infrastructural network is not feasible strategy due to the relatively short time the cascading failure needs to propagate. However, intentional node and link removal is a reasonable strategy to minimize the cascade. Numerical simulations are used to identify the nodes candidates for removal in case of the critical failure.

4. *Which are the nodes that should be additionally protected to keep the network more resilient to epidemics?*

   The nodes which should be additionally protected are the most critical nodes or links in the network responsible for the contagion spread. Most of the approaches used in the literature are based on various *centrality measures* and their variations. Here, an alternative method for evaluation of the node spreading power is proposed. The *Node Imposed Response (NiR)* measure utilizes concepts from the LTI systems theory.

More specifically, the *NiR* measure is based on the value of the system response to the input step function. Extensive numerical simulations are performed in order to validate the accuracy of the *NiR*.

5. *How to tune topology and improve the node properties in order to slow down the spread of a potential virus infection?*

   The main characteristic of the spreading process which defines the rate of an infection is the transmission probability. It is the probability a susceptible node will be infected by a single infected neighbor in one time step. Modifying the transmission probability for the specific pairs of nodes alters the epidemic dynamic. However, the constraints in the form of costs are usually present, and the best solution implies optimal resource allocation, so the epidemic behaves as expected. Here, the LTI system approach is used to identify the critical links. The results are validated by numerical simulations.

3.2 Methods

The research results which correspond to the sub-questions are compiled in a group of findings which fit the main research question. All the sub-questions regarding the cascade failures are answered through the assessment of the European National Research and Educational Network (NREN) in Chapter 5. The NREN topology is used as a test bed for various protection strategies against the cascades. All the network data are publicly available. Hence the results of the research would be easily reproducible and cross-checked by other authors. These data are used in other publications for numerical assessment and it is possible to perform a calibration and comparison of the results. The sub-questions related to the epidemics are addressed in Chapter 4. The focus is on epidemic modeling and node assessment using tools from the LTI systems theory.

Some of the methods used in the research are roughly grouped as follows:

3.2.1 The methods for modeling

Two supplementary methods used simultaneously are the *numerical* and *analytical* methods. They are applied to cross-check and compare the results acquired by either of them and to refine the models according to observations and derivations.

Analytical methods. The analytical approach helps to get a solution to a system state avoiding simulations and without applying a lot of computing power. The methods from the system theory will be used to analyze the networks. The networks are considered as LTI systems and the responses of a corresponding system for certain inputs are evaluated. The acquired results are used to assess the node spreading power and therefore to identify the most critical elements.

Numerical methods. The simulations are extensively used to observe the dynamics within the networks. The network simulation is widely spread method for networks research. Simulations can give an insight into the dynamics of a processes and provide plenty of information that could not be easily anticipated beforehand. All the simulation code was written in MATLAB, a mathematical software with features which cover many aspects of mathematics and could be used for network simulations and computations.

3.2.2 The methods for analysis

The main method used for the analysis is a *correlation* measure. The measure of correlation is used to validate the assumptions in the case of node assessment. For all the analyses we use Kendall's Tau rank correlation coefficient. It is a non-parametric measure of relationship between ranked data and a powerful tool to compare the results obtained by various modeling methods. The correlation coefficient τ takes a maximum value of 1 if the observations have identical rankings and a minimum value of -1 if observations have dissimilar rank. The first observation is ranked by the values of the vector $\overline{x}$

and the second observation is ranked by the values of the vector $\overline{y}$. Then the rankings are compared using Kendall's Tau [81] as:

$$\tau = \frac{2}{n(n-1)} \sum_{i<j} sgn[(x_i - x_j)(y_i - y_j)], \quad sqn(y) = \begin{cases} 1, & y > 0 \\ -1, & y < 0 \\ 0, & y = 0 \end{cases}$$

In case of the spreading influences explained more in details in Chapter 4, a value x_i is calculated for each node i. In the case of SI model, x_i is time needed to infect the 50% of the network. For SIR model, x_i is the outbreak size. The value y_i is calculated independently for each of the centrality measures. As a result we have a single (primary) vector $\overline{x}$ and six vectors for comparisons $\overline{y}_{nir}$, $\overline{y}_{bet}$, $\overline{y}_{cor}$, $\overline{y}_{deg}$, $\overline{y}_{h}$, $\overline{y}_{ds}$ for six centrality measures: NiR, betweenness, coreness, degree, H-index and DS, respectfully. Then, τ is calculated for each of the centrality measures to estimate its accuracy in assessment of the node spreading power.

Another method for analysis uses numerous *centrality* measures (Section 2.2) to identify the most important nodes. Additionally the *comparisons* are used for a large scale numerical analysis of the impact of node removal. The impacts of various attack strategies are compared to one another.

3.2.3 Metaheuristics

Metaheuristics are general algorithmic frameworks which are usually inspired by the processes in nature. They are used to solve complex optimization problems [82]. Many problems in the area of complex networks are becoming increasingly complex and dynamic and therefore often addressed by some heuristic. Besides the heuristic approach implemented in some algorithms for centrality measures, such as betweenness centrality, a *genetic algorithm* is intensively exploited in the research presented here.

A genetic algorithm (GA) is a metaheuristic inspired by the process of natural selection. Genetic algorithms are usually used to solve optimization and search problems by relying on operations inspired by evolution mechanisms such as mutation, crossover and selection. GA is a method for moving from one population of "chromosomes"[1] to a new population by using some principles of "natural selection" together with the genetics-inspired operators. The selection operator chooses the chromosomes in the population that will be allowed to reproduce, and on average the fitter chromosomes produce more offspring than the less fit ones. Crossover exchanges subparts of two chromosomes, roughly mimicking biological recombination between two single-chromosome organisms and mutation randomly changes the gene of some locations in the chromosome [83]. The genetic algorithm can be applied to solve a variety of optimization problems, including problems in which the objective function is discontinuous, nondifferentiable, stochastic, or highly nonlinear. The genetic algorithm is suitable to quickly identify a relatively good solution from a very big solution space. It can also address problems of mixed integer programming, where some components are restricted to be integer-valued. While the classical algorithms usually generate solution as a single point at each iteration, GA generates a population of points at each iteration. The best point in the population approaches an optimal solution.

Some solutions presented in this thesis are verified by the genetic algorithm which is implemented through the Global Optimization Toolbox in MATLAB [84].

3.2.4 Obtaining acyclic graphs

All undirected networks consist of bidirectional links, which produce cycles between every pair of nodes. To obtain an acyclic graph, a topology has to be modified such that all cycles are removed and number of nodes stays

[1] Chromosome in GA is an array which defines a single solution. Each chromosome consists of "genes" (e.g., bits), each gene being an instance of a particular "allele" (e.g., 0 or 1).

unchanged. In the process of topology modification, the number of removed edges should be minimized in order to maintain the topology as similar as possible to the original. There are two principles used in the proposed algorithm: 1) the most probable path of the infection spreading, such as shortest-path tree with the source node as a parent should be preserved; 2) the edges closer to the source node should be given the priority since the importance of the topology decreases quickly with the distance from the source [85–87]. The algorithm for obtaining acyclic graph considering the importance of the node immediate neighborhood is shown in Algorithm 1.

Algorithm 1 Obtaining acyclic graph

Input: $G(V, E)$, p, i ▷ i - source of infection
extract the shortest path tree (SPT) $G(V, E_{SPT})$
extract the set of remaining edges ▷ $E_{rem} = E - E_{SPT}$
create a hierarchical topology of a SPT
direct all the edges away from the source node
sort remaining edges by the distance from the source node ▷ E'_{rem}
while *no remaining edges* **do**
    return excluded edge $E'_{rem}(i)$
    check for cycles
    if number of cycles ≥ 1 **then**
        remove returned edge
end
return *the acyclic graph* $G(V, E_{acyc})$

The algorithm for generating acyclic graphs presented here does not guarantee the minimal number of removed links. Still, it optimizes a characteristic relevant to the spreading phenomena, which is a local topology around the source node. It attempts to preserve as many links as possible in the source's immediate neighborhood. The further away from the source node we go, the higher the chance that a link will be removed. However, the removal of distant edges has a very limited impact to spreading dynamics.

3.2.5 The methods for design

After modeling and analysis, the improvements to network design are proposed. In order to introduce the design solution which has to fulfill several criteria, the multi-criteria analysis is used. The multi-criteria analysis[2] is applied to achieve the optimal solution in the circumstances where many objectives have to be considered (multi-dimensional design space). Those objectives are usually conflicted (service level, resiliency against failures, cost...) and the multi-objective programming is used to achieve the Pareto boundary of the system.

3.2.6 Analytical Approach in Complex Networks Modeling

One of the focuses of this thesis is the development of an appropriate analytical approach for complex networks modeling. Generally, developing the exact analytical solution to a system with a complex dynamics is a difficult task as it is based on solving the resulting system of differential equations which is usually prohibitively large. The ultimate goal for many scientists is to develop the master equation for a system. Master Equation (ME) describes the time-evolution of a system that can be modeled as being in exactly one of the states at any given time, and where switching between states is treated probabilistically. The system equations are usually a set of differential equations describing the system variation over time. For some systems, the ME exists and is simple, while for other systems, the ME could be derived, but is excessively difficult. Generally, the MEs of a complex networks are derived in the following way [14]: if we denote the state variable as σ_i where all possible states are $\sigma_i = 1, 2, \ldots, k$ for each node, we can denote the particular configuration of the network at time t as the set $\sigma_t = (\sigma_1(t), \sigma_2(t), \ldots, \sigma_N(t))$, where $i = 1, 2, \ldots, N$, and N is the number of nodes in the network. The dynamical

[2] Multi-criteria analysis is sometimes referred to as *multi-objective programming*.

evolution of a system is given by the dynamics of the configuration $\sigma(t)$ defined by the all possible configurations σ. The ME approach studies the probability $P(\sigma, t)$ that a system is in the particular state at the time t. Therefore, the evolution equation for $P(\sigma, t)$ in continuous time approximation is:

$$\delta_t P(\sigma, t) = \sum_{\sigma'} [P(\sigma', t) W(\sigma' \to \sigma) - P(\sigma, t) W(\sigma \to \sigma')], \tag{3.1}$$

where the sum runs over all possible configurations σ' and $W(\sigma' \to \sigma)$ represent the transition rates from one configuration to another.

Except for the very small systems, solving the ME could be very difficult task. For the large systems, fully analytical solution becomes almost impossible. Therefore, other approaches have been used to tackle this issue, like different variations of numerical ME solutions [88].

Two different analytical approaches for analysis of spreading phenomena within complex networks are briefly presented here. The first one uses *Markov chain*, already known method for processes in discrete set of times, successfully tackling the random processes that undergo transitions from one state to another on a state space. We can show that the Markov chain approach quickly becomes impractical with network size. Another solution is based on systems theory, using the set of tools from the theory of LTI systems to tackle the spreading phenomena in complex networks. The brief introduction to LTI approach is presented here, and more detailed explanation is given in the Chapter 4.

Modeling Epidemics in an Undirected Network Using Markov Chain

As an example of a complexity of the analytic solutions, the most common approach to probabilistic problems is presented. Here, we show a possible solution for modeling epidemics in an undirected network using *Markov chain*.

Although, this model is relatively simple and straightforward, the complexity of a system makes the solution computationally demanding and for a large systems practically impossible.

The nature of all the virus-like dynamics is that they are usually transmitted in one direction. If there is one node initially infected, the infection will spread from that node to its neighbors but not vice-versa. The backward transmission of the malicious data from one infected node to another previously infected could be neglected in the modeling as it does not make any change is the network behavior. Furthermore, the repeated infection of already infected node anywhere between last infected and a source could be neglected. It implies that all cyclic paths could be consequently removed. Therefore, for the purpose of modeling virus infection, the network should be considered as directed and acyclic.

The directions of edges in the network depend solely on the source of the infection. All directions on the network are formed depending on the node which is considered to be a source. Thus, there is no single solution for altering the network topology as different set of directions on the edges will be applied for each node chosen to be a source and finally there will be as many topologies as there are possible source nodes.

Four types of links are recognized in the graph: *Tree edges*, *Back edges*, *Forward edges* and *Cross edges* as shown in Figure 3.1.

In order to make graph acyclic and directed, we remove all back edges and keep the others. Since the graph is originally undirected, the same could be done by choosing the right direction of edges in the way that back edges are omitted.

Here, two possible solutions for making graph directed and acyclic are proposed:

1. The first solution is based on a Depth-First Search Algorithm explained in details in the book *Algorithms in a Nutshell* [90] which investigates

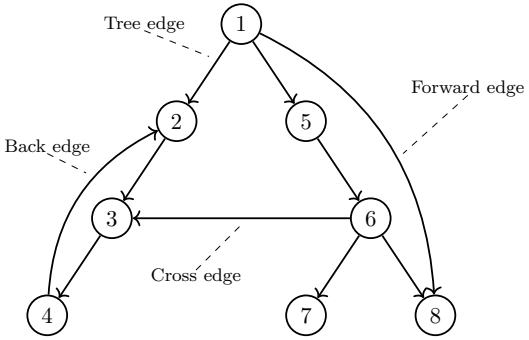


FIGURE 3.1: **Types of edges in a graph.** Here we recognize four types of edges relevant for spreading modeling: *Tree edges*, *Back edges*, *Forward edges* and *Cross edges* [89].

the graph and marks nodes and links. It can be used to identify MST of the graph as well as back links which could cause the loops. Except the back links, all other links are allowed, so the direction of all links which are not a part of a tree should be chosen as follows:

- **The tree edges** - after the minimum spanning tree is identified, the edges should be directed in the way that the chosen source node becomes the root.
- **The back edges** should be removed, so the forward direction stays active.
- **Forward edges** are allowed, but since the minimum spanning tree is formed, all forward edges are going to become a part of the tree.

2. The second solution is based on some of the algorithms for finding minimum spanning tree. For example Kruskal's algorithm is among the most popular ones [91]. After the minimum spanning tree is formed and directions of the links in the tree chosen, comparing original graph with the minimum spanning tree we can identify other links which are allowed to be restored following the Algorithm 1.

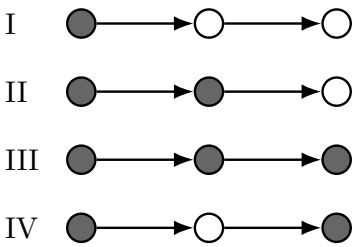


FIGURE 3.2: **System states of a line graph after a virus infection.** The infection originates in the node on the left and it spreads following the SI model. The infected node is marked in gray. All the possible system states are listed from I to III. Note that the state IV is not possible.

After the link removal procedure, we identify all possible system states. In the case of the virus infection, each node in the network could be in one of two possible states: *infected* or *healthy*. In order to implement an analysis using Markov chain, all possible system states are identified and all probabilities of transitions from one state to another are calculated. In case of the SI epidemic model, there are some constraints involved regarding the number of possible system states, which highly reduce the number of states and therefore simplify the analysis process. These constraints are related to the direction of possible infection as depicted in Figure 3.2.

Let us consider a simple network consisting of three nodes in a line topology with directed links as shown in the Figure 3.2. If we assume the infection originates in the node 1, there are three possible system states (state I, II and III). The state IV is not possible as the recovery of nodes is not considered and the infection could not bypass the node in one path. Therefore, we can exploit this property in order to reduce the number of states in the analysis.

The examples of three simple networks with associated system states are presented in Figure 3.3. The probabilities of virus transmission from one to another neighbor node are independent and all equal to is p. The matrix P_x is a transition matrix of system states for each network.

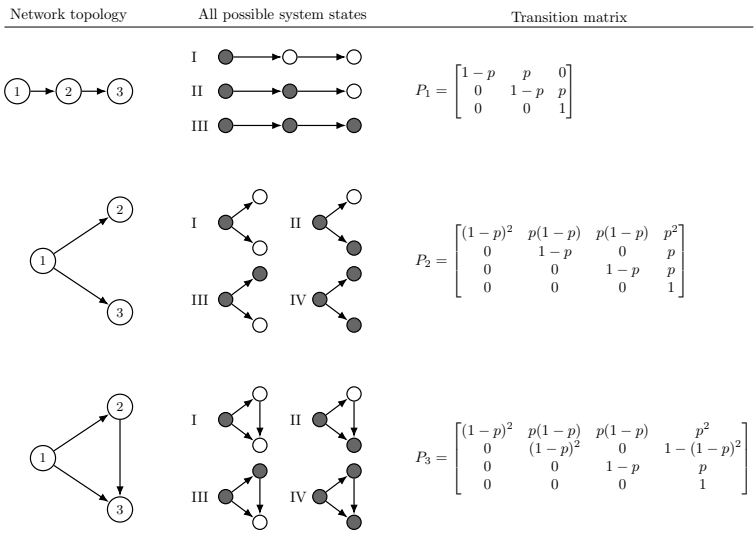


FIGURE 3.3: **Three simple networks with associated system states and transition matrices.** For a given network topology (on the left) and infection origin in node 1, all possible system states are presented (middle). The associated transition matrices for all of the system states are on the right. As the network size increases, the analysis of a stochastic process grows in its complexity.

Solving a problem with small number of nodes is relatively easy, but as the number of nodes in the network becomes larger, the number of possible system states increases. In case of the relatively large networks, identification of all system states becomes difficult. Therefore, we can undertake certain steps to simplify the identification process as much as possible. One possible solution is to split a network into a series of consecutive sections and analyze all system states for each segment independently. Then, all states from each segment can be merged into a single transition matrix. In the model we are considering, the node cannot recover from the infection, and the infection will spread following a certain direction. This property of an epidemic allows splitting network for the analysis. In the case of tree network structure without forward and cross edges, the splitting of a network could be done in many possible

ways. Since the transition matrix of the whole network as well as the parts of a network is upper triangular, a single global transition matrix could be formed by expanding the initial one.

In Figure 3.3 a solution complexity for a very small system is presented. It can serve as an illustration of a complexity of some larger networks. Even with methods for reduction of states, such a solution for the large systems remains practically unattainable.

Modelling Epidemics in Networks Using LTI System Approach

A brief introduction of the system theory based approach for modeling the epidemic in the network is presented here.

The network is decoded in the form of an LTI system. We show that observing the particular system response, we could formulate some conclusions about the potential behavior of the network in the case of virus infection.

As the nodes interact with each other, a network could be regarded as a system composed of connected interdependent elements where the output of one element is at the same time the input of another. Therefore, system theory can be used to describe the behavior of such system. The network structure characterizes internal connection pattern. More precisely, it describes in which way the states of nodes affect each other and at the same time affect the outputs, when excited by the inputs. In the case of the infection, the infected node has an altered state and could influence the states of neighboring nodes. Since the epidemic is based on propagation through the network, the transmission of the node's state to its neighbor is a fundamental building block for the epidemic modeling.

The network is usually represented as a graph $G(V, E)$, with $N = |V|$ vertices or nodes and $M = |E|$ edges or links. Furthermore, the network's topology is usually characterized by the adjacency matrix A_{adj}. For a graph with N nodes, the A_{adj} is the $N \times N$ matrix where $a_{i,j} = 1$ if there is a directed

edge between i^{th} and j^{th} node, and $a_{i,j} = 0$ otherwise. The adapted form of adjacency matrix is used as a system state matrix in further analysis.

In order to obtain simplicity and computability of the analysis, the network is regarded as a discrete LTI MIMO (Linear Time-Invariant, Multiple Input-Multiple Output) system. The state-space formulation used to describe the system is:

$$\underline{x}(n+1) = A\underline{x}(n) + B\underline{u}(n) \tag{3.2}$$

$$\underline{y}(n) = C\underline{x}(n) + D\underline{u}(n), \tag{3.3}$$

where $\underline{x}(n) \in \mathbb{R}^N$ is the state vector at discrete time n, $\underline{u}(n) \in \mathbb{R}^M$ is input or control vector, and $\underline{y}(n) \in \mathbb{R}^M$ is the output.

The matrix $A := (a_{ij})_{N\times N} \in \mathbb{R}^{N\times N}$ is the state transition matrix and the matrix $B \in \mathbb{R}^{N\times M}$ is input matrix. The matrix $C \in \mathbb{R}^{M\times N}$ is the output matrix and $D \in \mathbb{R}^{M\times M}$ is the feedforward matrix. The elements of matrix A are denoted as a_{ij}. The graph is considered to be undirected.

The adjacency matrix A_{adj} represents the topology of bidirectional graph, so the feedback loops are possible, and the stability of the system would be hard to achieve. The purpose of designing a state space model of a network is modeling of epidemic and therefore we should take into count the observed dynamics of epidemic phenomena in real networks. In order to solve the State-Space Network Realization Problem it is necessary to identify the internal network structure of the system and encode it in a set of state-space matrices, (A, B, C, D) which produce the input-output behavior.

Detailed explanation of state-space system modeling of the network dynamics is out of the scope of this short introduction. A thorough description with examples and simulation results is presented in the Chapter 4.

3.2.7 Numerical Approach

The complex behavior observed in the number of physical systems makes it very hard to be described analytically. A complicated mutual dynamics of agents can produce practically unsolvable equations. In this case, the *agent-based models* (ABM) sometimes referred to as *microscopic computer models* are used.

This approach is used when the behavior of the single agent is known, but the outcome of interaction of many agents is unknown. It is one of the ways to emerge from the lower (micro) to a higher (macro) level of systems. Usually, there is a limited number of states for each agent. Each agent is given a set of instructions which determine its behavior in regards to various variables, such as time, interaction rules with other agents, external factors, etc. At each time step, the model-specific update procedure is applied to every agent. Then the state of each agent and the system as a whole is evaluated by the computer. The principle where the simple behavioral rules generate complex behavior is known as K.I.S.S. ("Keep it simple, stupid")[3] [92].

One of the basic examples is the agent-based simulation of the epidemic in the network. Let us consider the network where each node can be in one of two states A (healthy) and B (infected). The reaction process is denoted as $A + B \rightarrow 2B$, which means that in each time step the node in the state A will switch to state B in contact with the neighboring node already in state B. Let us consider that at the beginning all nodes except one are in the state A. The simulation procedure is the same for the each time step and runs as follows: each node in state A checks its neighbors; if any of them is in the state B, it updates its state to B.

This way the complex system behavior is recreated inside the computer, providing solutions for certain problems which could be solved only experimentally. Even though this approach is very powerful for forecasting the complex

[3] K.I.S.S. is not meant to imply stupidity. On the contrary, it is usually associated with intelligent systems that may be misconstructed as stupid because of their simplistic design.

systems dynamics, ABMs are often not transparent enough. They give a very detailed cross section of a system in every point in time, but lack the ability to explain certain behaviors. For a deeper understanding of some basic system properties one usually has to analytically describe the system even in the lower scale, and then to use the ABM to confirm the analytical solution. Numerical simulations are extensively used throughout the thesis to recreate various dynamical processes from epidemic spreading in Section 4 to cascade failures in Section 5.

3.3 European NRENs

The abbreviation NREN stands for the National Research and Education Networks. The purpose of the Research and Education networking is to establish high-performance network infrastructure that connects universities and research institutes independently from the other public or private networks which are used for other purposes.

The National Research and Education Networks (NRENs) are dedicated high speed networks that act at the national level to provide connectivity between universities, research institutes, educational hospitals, schools, further education colleges, libraries and other public institutes. These networks are built separately using dedicated fibre optic connections or using already existing high capacity leased connections provided by telecommunication providers. The NRENs allow researchers, faculty, staff, and students to access a broad range of research tools and information resources. NRENs often establish and coordinate distributed computing resources (grids) and operating experimental test-beds for data-intensive applications. Some NRENs have even broader significance as they act as a service providers for the third parties, and even as a service providers for the general population.

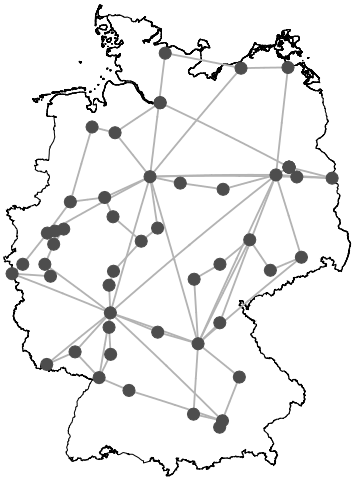


FIGURE 3.4: **NREN topology in Germany.** The illustration of the NREN topology in Germany consists of 51 nodes and 80 links.

Together, these networks connect over 50 million users at 10000 institutions across Europe, delivering a range of networking services for institutions, projects and researchers [93].

The history of the research and education networking starts with the development of the first successful network, known as ARPANET which connected government-sponsored research organization in 1969. In 1984, NSFNET was developed as a general purpose research network which served as the backbone of the Internet. Nowadays, the high speed dedicated research and education networks integrate networking interfaces, switches, and routers and facilitate running computationally intensive R&E applications and services that are often not found on the Internet.

The main drivers for creating NRENs are based on technological, social, and economic factors [94]. The technological factors are there to satisfy high demand of eScience initiatives from multimedia collaboration, distributed high performance computing and requirements of high bandwidth necessary for

the large scientific and experimental facilities like CERN's Large Hadron Collider (LHC). Another motive for the NREN's existence has a social aspect, including virtual organizations, collaborative research or tele-education and facilitating the common culture of R&E community. Finally, NREN helps to develop the capacity for economic prosperity serving as the demand aggregator and consolidating and controling the diverse public expenditures. It serves for the promotion of information society (e-Government, e-Business, e-Health ...) and finally as a stimulation of technological developments and telecom markets.

At the national level it is the network which interconnects the local networks of the research institutes and universities in each country. The National Research and Education Networking organization (the NREN) of the country is responsible of the national network. However, at the international level, connectivity between the European NRENs is provided by the GÉANT network. Some NRENs additionally have their own links to key destinations. Connectivity to the commercial Internet takes place both at the NREN level and, to a limited extent, at the GÉANT level. [95]

The GÉANT network is the pan-European research and education network that interconnects Europe's NRENs. It is co-funded by the European Commission and national bodies responsible for Europe's NRENs. The GÉANT project is a collaboration between 39 partners: 37 European NRENs, the GÉANT Association's Cambridge office (formerly DANTE) and Amsterdam office (formerly TERENA) as well as NORDUnet (representing 5 Nordic countries) [93].

In the analysis of cascading failures later in Chapters 5 and 4, only the backbone of the European NRENs is discussed. The lower level networks connected to the NREN's backbone within the country are therefore excluded[4].

[4]Lower level networks are all MANs and LANs supported by the higher level NREN backbone. NREN backbone usually connects main data centers in the cities, while the MAN and LAN networks are build upon it and cover various research and educational buildings within the city.

FIGURE 3.5: **NREN topology in Germany and Poland.** The illustration of two connected NRENs in Germany and Poland consists of 82 nodes and 116 links in total.

The network of NRENs in Europe consists of 37 national NRENs, with various topologies. They are connected through the GÉANT backbone, making it a single pan-European network. The observed network consists in total of 1157 nodes and 1465 bidirectional links. The topology data is obtained from the network data provided by "The Internet Topology Zoo", an ongoing project to collect data network topologies from around the world supported by the Australian Government through an Australian Postgraduate Award and Australian Research Council Discovery Grants [96].

An example of a single NREN is shown in the Figure 3.4, where the topology of the National Research and Education Network in Germany is depicted. The German NREN backbone consists of 51 nodes and 80 links. The links are presented simply as a straight lines, thus ignoring the actual geometry of the routes, which is in this case not relevant and not a subject of the analysis.

The national RENs are connected between each other. In the Figure 3.5 NRENs from Germany and Poland are shown together with interconnecting

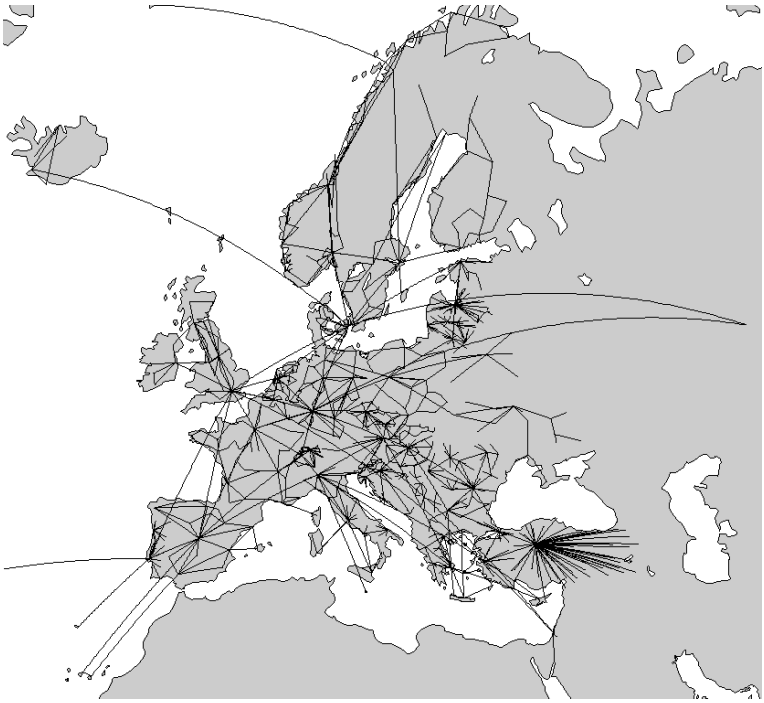


FIGURE 3.6: **Complete European NREN topology.** The overview of the full network of European NRENs with 1157 nodes connected with 1465 links

links. Finally, the complete network overview of European NRENs with 1157 nodes and 1465 links is presented in Figure 3.6.

Chapter 4

LTI System Theory and Spreading Phenomena in Networks

Various natural and artificial systems are characterized by the complex interdependencies of their elements. Those relations are usually modeled as networks. For this reason network analysis became an important tool for studying some of the typical system dynamics such as spreading of information or diseases. Spreading characterizes numerous processes observed in social and communication networks [14], such as the spread of rumors, news and ideas among humans, or data broadcast and cyber attacks on communication networks. Two topics regarding the spreading phenomena are discussed: *epidemic modeling* and *identification of the most influential spreaders*. Both problems are addressed using a proposed Linear Time-Invariant (LTI) system approach.

Throughout the years, LTI system theory has been used mostly to describe the electrical circuits and networks. LTI is suitable to characterize the behavior of the system consisting of numerous interconnected components. In this Chapter we show that the same mathematical toolbox can be used for the complex network analysis.

4.1 Linear Time-Invariant Representation of Networks

The Linear Time-Invariant systems theory is used to describe a system consisted of many interconnected components which influence each other. One of the most well-known attempts to use LTI to characterize the dynamic in a complex network deals with the specific problem of controllability in dynamic systems [97]. System theory is used to identify the driver nodes which will be able to mainly influence the system dynamic. The approach from [97] addresses the problem of controllability for arbitrary network topologies and sizes also for weighted and directed networks.

The complex networks consist of multiple interconnected components which communicate with each other or influence each other by changing the state of the influenced nodes. Such property has inspired the idea of network conversion to the Multiple-Input and Multiple-Output (MIMO) system. We show that the LTI theoretical approach can be used to capture the most common epidemic dynamics described by *Susceptible-Infected* (SI) and *Susceptible-Infected-Recovered* (SIR) models, more formally explained in Section 2.4.1 and Section 4.3.5.

A network is usually represented as a graph $G(V, E)$, with $N = |V|$ vertices or nodes and $M = |E|$ edges or links. The network topology is usually characterized by the adjacency matrix A_{adj}. For a network with n nodes, A_{adj} is the $n \times n$ matrix where $a_{ij} = 1$ if the i^{th} and j^{th} nodes are connected, and $a_{ij} = 0$ otherwise. This particular graph representation is convenient for a system theoretic approach as it resembles the state matrix A used in the state space representation of a physical system. Alternatively, the topology of the network could be defined by the list of edges, usually represented as the $M \times Z$ matrix where M is the number of edges and Z is the information about the edge. The number of columns is usually $Z = 3$ where the first,

second, and third column are consisted of source node, sink node, and the edge weight respectively.

The state transmission is the main characteristic of epidemic modeling. A certain unwanted information (i.e. virus) enters the network at one or more points and starts spreading. It is subsequently replicated and conveyed from one node to another. The location of the virus, and therefore the state of the network, changes with every transmission in multiple time steps. The topology of the network is considered to be static. These properties allow us to observe the network as a discrete LTI MIMO (Linear Time-Invariant, Multiple Input – Multiple Output) system [31]. The state-space representation describes the system as:

$$\underline{x}(n+1) = A\underline{x}(n) + B\underline{u}(n) \tag{4.1}$$

$$\underline{y}(n) = C\underline{x}(n) + D\underline{u}(n), \tag{4.2}$$

where $\underline{x}(n) \in \mathbb{R}^N$ is the state vector at discrete time n, $\underline{u}(n) \in \mathbb{R}^M$ is input or control vector, and $\underline{y}(n) \in \mathbb{R}^M$ is the output. The matrix $A := (a_{ij})_{N\times N} \in \mathbb{R}^{N\times N}$ is the state transition matrix and the matrix $B \in \mathbb{R}^{N\times M}$ is input matrix. The matrix $C \in \mathbb{R}^{M\times N}$ is the output matrix and $D \in \mathbb{R}^{M\times M}$ is the feedforward matrix.

To represent the dynamics of the system, we use the system matrix A, which can be constructed as a transpose of the adjacency matrix which describes the network topology $A = A_{adj}^T$ [31, 97]. Such a representation implies that a certain signal excites the system by entering one or more input points (nodes). Then, the signal gets conveyed from one node to another. To identify the input nodes, we use the system matrix B. It is the input matrix and it is determined by the system structure. Matrix B is used to identify the input points of the system. There could be one or multiple inputs. Matrix B has dimensions $1 \times N$, thus it is a column vector of N elements. Let us assume the input node is i, then $b_i = 1$, within the matrix $B = (b_i)_{1\times N}$.

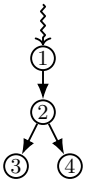
During this process, each node i modulates the signal by amplifying it by a certain parameter $a_{ij} \leq 1$ before transferring it to the adjacent node j. We can choose which nodes to observe, either all or just a fraction of nodes and measure the signal in each of them over time. To identify the observable set of nodes, we use the matrix C. The matrix C is an $M \times N$ matrix of constant coefficients c_{ij} that weight the state variables. Usually C has a size of $M \times 1$. As nodes cannot be partially infected, the state of each node in the network is binary (infected or not infected). Therefore, there is no need for weighting the states variables, i.e. the weights for all variables are the same. Since all weights are the same, in the output matrix $C := (c_i)_M$ all of the elements c_i are equal. We can understand the measurement points as a set of sensors collecting data in every time step. By analyzing the gathered data, we can examine the dynamics in the network and estimate the possible impact of infecting a certain number of nodes. The system matrices are generated the same way, regardless of the network dynamics we want to study. At the end, only the system response to the input signal is used for analysis.

The matrix D, known as a feedthrough (or feedforward) matrix, is an $M \times M$ matrix of constant coefficients d_{ij} that weight the system inputs. In case D is the null matrix, the output equation reduces to a weighted combination of the state variables, i.e. $\underline{y}(n) = C\underline{x}(n)$.

The input vector $\underline{u}(n)$ is an input signal. In case of a discrete system it is referred to as the *input sequence*, as it is a vector of values which are taken consecutively as an input in each time step. Systems are usually excited by *impulse* or *step* function, causing the *impuls* and *step* response respectively.

The *impulse signal*, also known as Dirac delta function is denoted by δ and defined by

$$\delta(n) = \begin{cases} 1, & n = 0 \\ 0, & n \neq 0 \end{cases}$$



$$A_{adj} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

FIGURE 4.1: **An example of a small directed network and its corresponding adjacency matrix.** The zig-zag line shows the signal entering the network. In LTI representation, the node ① is an input point for the signal exciting the system.

Another signal used for analysis is the unit step function, known as the Heaviside step function. An alternative form of the unit step is used as a function of a discrete variable n:

$$\mathbb{1}(n) = \begin{cases} 0, & n < 0 \\ 1, & n \geq 0 \end{cases}$$

After representing a network as a LTI system, we can use various tools developed for a system analysis to gather and analyze the signal response acquired from the network. We can choose to excite the system in one or more points, hence mimic one or multiple initially infected nodes. Also, we can choose to focus the analysis to a designated set of nodes and observe the response only for them.

For example, a small directed network with the corresponding adjacency matrix A_{adj} is shown in Figure 4.1. The state-space representation of the corresponding system with matrices A, B and C is shown in (4.3) and (4.4). In this example the input vector $\underline{u}(n)$ is in the form of the Heaviside (unit) step function. Here we chose to observe the sum of the states of all nodes, and therefore the matrix $C \in \mathbb{R}^{M \times N}$ consists of all ones.

$$\underline{x}(n+1) = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \underline{x}(n) + \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} \times \mathbb{1}(n) \tag{4.3}$$

$$\underline{y}(n) = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \times \underline{x}(n) + \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \times \mathbb{1}(n) \tag{4.4}$$

One can choose any preferred way for solving the system equations [98]. The system could be presented in the form of the transfer function $H(s)$ where

$$H(z) = \frac{Y(z)}{X(z)} = C(zI - A)^{-1}B + D. \tag{4.5}$$

Furthermore, the step response of the system is obtained by

$$y(t) = \mathcal{Z}^{-1}\left(\frac{1}{z}H(z)\right), \tag{4.6}$$

where the $\mathcal{Z}^{-1}\left(F(z)\right)$ is the inverse Z-transform of a function $F(z)$.

However, this method could be challenging for large graphs as finding the inverse of large matrices is computationally expensive. An alternative and more programming-friendly approach would be a recursive solution [31] which avoids the matrix inversion:

State-space equations are a set of linear first order difference equations. This solution is convenient as it could be easily implemented using any mathematical software available. We start from the basic equations (4.1) and (4.2). For an initial point in time n_0 and for every $n > n_0$ we have

$$\begin{aligned} x(n_0+1) &= Ax(n_0) + Bu(n_0) \\ x(n_0+2) &= Ax(n_0+1) + Bu(n_0+1) \\ &= A^2x(n_0) + ABu(n_0) + Bu(n_0+1) \end{aligned}$$

and therefore

$$x(n) = A^{n-n_0}x(n_0) + \sum_{k=n_0}^{n-1} A^{n-1-k}Bu(k). \tag{4.7}$$

The system output is obtained by substituting (4.7) in (4.2):

$$y(n) = CA^{n-n_0}x(n_0) + \sum_{k=n_0}^{n-1} CA^{n-1-k}Bu(k) + Du(n), \tag{4.8}$$

where $y(n)$ is a system response for the input signal in time domain. The response here is denoted as $y(n)$ rather than $y(t)$, as the systems we are observing are discrete, and the response is captured in a set of time intervals, so the $n \in \mathbb{N}$. If $y(n)$ is an impulse response, it is defined as the output signal that results when an impulse $\delta(n)$ is applied to the system input. The same applies for the step response. In that case, the input signal is in form of a step function $\mathbb{1}(n)$. This allows us to predict what the system's output will look like in the time domain. Practically, in the interconnected systems we are dealing with, an input signal enters the network in one or more nodes. The signal then gets conveyed from one node to another. It can be altered by the node it passes through or it can remain unchanged. The output is then red as the sum of the signals in chosen nodes over time. The right-hand side of the (4.8) has three components. The first one identifies the value of the output for the initial system state, $n = 0$. Second part sums the signal in all observable outputs for desired time span n. The third part evaluates the influence of the feedforward matrix D which allows for the system input to affect the system output directly. Systems considered in the analyses in this thesis do not have a feedforward element, and therefore the D matrix is the zero matrix.

4.2 Modeling Epidemics by Virtual Network Expansion

The epidemic process is always characterized by a parameter p which is the probability of transmission. The p represents the probability a virus will be transmitted from a single infected to a single neighboring susceptible node in one time step. Linear Time-Invariant system dynamic describes only deterministic behavior. Therefore, it is necessary to make certain modifications to the network in order to use the LTI approach. One example is to apply *virtual network expansion*, explained here.

A dynamic of the system is determined by the matrix A, constructed in such a way that $a_{ij} = 1$ if there is a link between nodes i and j such that node j affects the node i. The link $(j \to i)$ could be either excitatory or inhibitory [97], i.e. $sgn(a_{ij}) = -1$ or $sgn(a_{ij}) = 1$ respectively. In case of a virus infection in undirected networks, there would be two links between every two connected vertices. The weights of edges in such networks are considered to be equal with unit strength. The links are directed and excitatory, so the matrix A could be obtained as a transpose matrix of the adjacency matrix of the network A_{adj}.

However, If the state transition matrix A would be constructed simply by making the transposed A_{adj} matrix $A = A_{adj}^T$ as stated before, the feedback loops would be imminent, so the BIBO[1] stability of the system would become a severe issue. The idea is to modify the matrix A so the BIBO stability criteria are fulfilled and the dynamic in a system characterize the epidemic phenomenon at the same time.

[1] A system is bounded-input bounded-output (BIBO) stable if its output will stay bounded for any bounded input. In the undirected networks feedbacks surely occur and the outputs of a system are routed back as inputs and form a circuit or loop.

4.2.1 Almost Certain Transmission, p=1

To make a resulting system BIBO stable in the case of the uniform values $a_{ij} = 1, \forall a_{ij} \neq 0$, one approach is to eliminate all possible feedback loops. To keep the epidemic dynamics intact, some assumptions have to be made:

1. *One node will affect another almost surely* - the signal within the LTI system gets transferred from a node to its neighbor without deterioration and with unit strength in each time step. It means the transmission from one node to another would occur almost surely, i.e. probability $p = 1$
2. *The information transmission is not recursive* - a single node can not be infected more than once

Considering the abovementioned assumptions, it is possible to modify the topology in the way that loops get avoided and the epidemic pattern stays unchanged. In the case of the inevitable infection from an infected to neighboring node, meaning the transmission rate is $p = 1$, the infection will spread deterministically. It will follow the shortest path from the source node to all other reachable nodes in the network and the time of the full infection will be minimal, equaling the minimal number of hops from the source to the furthest node. The infection will spread following the shortest-path spanning tree rooted in the initially infected node. Therefore, we construct the minimum spanning tree (MST) [90] subgraph A_{tree} of the undirected graph A_{adj} with $E(A_{tree}) \subseteq E(A_{adj})$ and $V(A_{tree}) = V(A_{adj})$.

Let us take for example a small undirected graph $G(V, E)$ with $M = 4$ nodes and $N = 4$ edges, shown in the Figure 4.2a. The adjacency matrix for that network would be the matrix A_{adj} and one possible MST of A_{adj} for the root node 2 is $A^{(2)}_{tree}$ (Figure 4.2b):

$$A_{adj} = \begin{bmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}, \qquad A^{(2)}_{tree} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

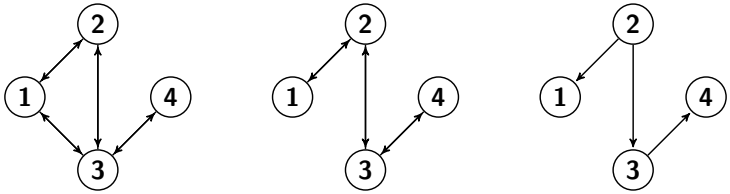


FIGURE 4.2: **Minimum Spanning Tree in a small network.** [31] The example of a network with $M = 4$ and $N = 4$. (a) The original undirected network; (b) Minimum spanning tree with the root at node 2; (c) Directed minimum spanning tree with root at node 2

In this rather trivial example we can easily identify the only MST rooted in node ②. In larger graphs there are usually multiple minimum spanning trees originating from a single node. The same applies also for the weighted graphs as the sum of all weights of tree edges could be the same for multiple trees [99]. However, there is no significance of which MST will be used in the case of almost sure transmission. The path length from the root to all other nodes remains the same, hence the epidemic will spread with the same rate.

The MST of an undirected graph is also undirected. In order to prevent feedback loops in the resulting tree, it has to be transformed to a directed one. In the case of the epidemic modelling, the directions assigned to each edge should reflect the most probable direction of the epidemic spreading. The subgraph $G_T(V, E_T)$ of the graph $G(V, E)$ represents the direction of infection flow through the network. The MST can be aligned in a hierarchical order in the way that the root node is positioned at the top, the first neighbours at the first level, the neighbours of the neighbours at the level below, and so on. Following this convention we can assign the direction to each node so it leads always from higher to lower level. In the working example, the MST $A^{(2)}_{tree}$

with root node 2, becomes $\vec{A}^{(2)}_{tree}$, as shown in Figure 4.2c.

$$\vec{A}^{(2)}_{tree} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \qquad A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

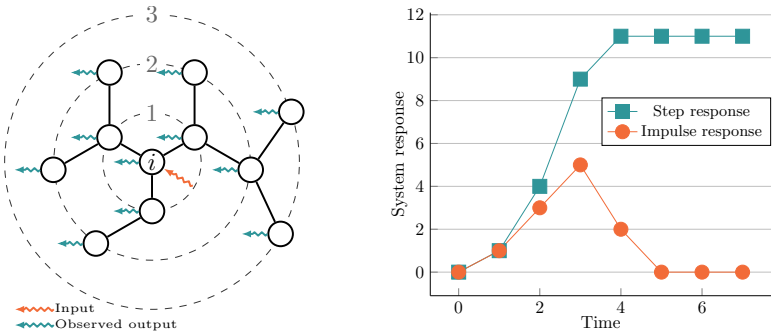
Since the matrix A is a transpose of an adjacency matrix, we have $A = (\vec{A}^{(2)}_{tree})^T$. Also, for the example in Figure 4.2, with input node ② the matrix $B = \begin{bmatrix} 0 & 1 & 0 & 0 \end{bmatrix}'$.

The values of c_i could be determined without any restriction from the set of real numbers, but for the sake of simplicity, here we chose the unit value. If we choose to observe all nodes in the network then $c_i = 1$ for $i = 1, 2, 3 \ldots M$. In the example in Figure 4.2, $C = \begin{bmatrix} 1 & 1 & 1 & 1 \end{bmatrix}$.

Finally the LTI system constructed out of the network from a Figure 4.2 is composed using state-space matrices A, B and C. After the initial system excitement by Dirac-delta $\delta(n)$ or Heavisade-step $\mathbb{1}(n)$ function, the set of values has been assigned to the output vector $\underline{y}(n)$. After applying the recursive solution from (4.8), we get the system output in the form of an impulse and step response for $\delta(n)$ or $\mathbb{1}(n)$ inputs respectively.

- the impulse response of a system is $\underline{y_\delta}(n) = \begin{bmatrix} 1 & 2 & 1 & 0 \end{bmatrix}$
- the step response of a system is $\underline{y_\mathbb{1}}(n) = \begin{bmatrix} 1 & 3 & 4 & 4 \end{bmatrix}$

How those responses correspond with the results of an epidemic simulation? Let us simulate the epidemic spreading in the same network with the same initial node(s) and with the transmission probability $p = 1$. The epidemic model simulated here is a simple SI model. As the output of the simulation we get two vectors: $\underline{v}(n)$ where each value represents the number of infected nodes in the particular time step, and $\underline{v_u}(n)$ where each value represents the total number of infected nodes from the beginning of infection up until the certain time step.



(a) (b)

FIGURE 4.3: **Simple graph and corresponding system response.** [35] (a) An example of a small tree graph where the node i is a source of the infection. The signal in the form of the unit step or impulse function enters the network at node i. Here we observe the state of the system in each time step by measuring the signal strength in all nodes and adding them together. The resulting measurements are step and impulse response respectively. (b) Step and impulse response of a corresponding LTI system with the single input in the position of node i. The response corresponds to the spreading dynamics over time which originates in the source node i. This is the simple case of almost certain infection of the neighboring nodes in each time step over the tree graph. The step response reveals the number of infected nodes over time. Impulse response shows the number of infected nodes in each time step.

- the number of infected nodes at every point in time $\underline{v}(n) = \begin{bmatrix}1 & 2 & 1 & 0\end{bmatrix}$
- the sum of infected nodes at every point in time $\underline{v_u}(n) = \begin{bmatrix}1 & 3 & 4 & 4\end{bmatrix}$

We can see that for this borderline example ($p = 1$), the step response corresponds with the total number of infected nodes so that $\underline{y_{\mathbb{1}}}(n) = \underline{v_u}(n)$ and the impulse response corresponds with the number of infected nodes in each time step so that $\underline{y_\delta}(n) = \underline{v}(n)$.

Another more illustrative example of a small tree graph is presented in the Figure 4.3. To depict the hierarchical structure of a tree, the nodes are positioned on the dotted concentric circles, where the larger circle represents a

single hop distance from the smaller one. For the network from the picture, the corresponding LTI system is created and the step and impulse response are calculated. The A matrix is binary. The output is measured in all nodes simultaneously, which means that we observe all nodes as outputs and calculate the final output as the sum of signal strengths in all nodes over several time steps. The input is a single node i located in the centre, which is a parent node of a tree. The sum output is plotted in Figure 4.3b. Notice the *impulse response* for the unit input. The value of the response over time equals the number of nodes on corresponding circles. For the case of a virus transmission in the example network with the source in the node i, and the almost certain virus transmission from infected to susceptible, the impulse response shows exactly the number of infected nodes over time. Likewise, the *step response* displays the total number of infected nodes.

The example above shows that the approach of using linear-time invariant system analysis can be used to study the epidemics in networks to a certain degree. However, this method of analysis is limited only to tree graph with almost certain transmission from infected to susceptible node. In reality, the contagion spread is characterized with a transmission rate below 100%, and usually $p \ll 1$. For the unlikely case of $p = 1$, the network could be transformed to a shortest path tree with the seed node since the parent as the infection route is known and unnecessary edges could be removed without affecting the infection dynamic. On the other hand, for any $p < 1$, the number of multiple incoming edges and multiple possible paths must not be neglected. Time-invariant system analysis do not consider stochastic dynamics of mutual interaction between elements. Furthermore, the introduction of the probabilities in each iteration of equation 4.7 would produce the time-variant system. A solution proposed next uses a process of *Virtual Network Expansion*, which can transform the original network in the way it can be conveyed into the proper LTI system with respect to the transmission probabilities. It could be used to predict the infection dynamics on an arbitrary topology for any $0 \leq p \leq 1$.

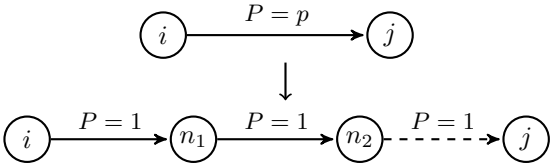


FIGURE 4.4: **Virtual network expansion.** [31] Insertion of additional nodes between a pair of existing ones: an initial edge between nodes i and j has been replaced by a set of consecutive nodes with the transmission probability $P = 1$. The number of additional nodes $n_1, n_2...$ has been derived from the probability density function of geometric distribution using (4.10)

4.2.2 Uncertain Transmission and Virtual Network Expansion

Network epidemics dynamic is almost always characterized by the probability of transmission p. It is a probability of virus transmission from infected to susceptible individual independently in a single time step. However, for the uninfected node i with k_i infected neighbors, the probability of transmission becomes $p_i = 1-(1-p)^{k_i}$. Usually, the p is considered to be uniform regardless of the node, although it is possible to use different p for each pair of nodes, therefore p is a value rather assigned to the *edge*, as it quantifies the relation between two adjacent nodes. It could be alternatively considered as a *weight* of a particular edge.

The LTI approach alone does not give a possibility to introduce probabilities. Therefore, an alternative concept called *Virtual Network Expansion* is introduced to modify the original network before the conversion to LTI suitable form.

Let us consider an undirected network $G(V, E)$ with a probability of virus transmission of $0 < p < 1$. Let us further assume that there is a secondary network $G_E(V_E, E_E)$ with the $p = 1$ which will demonstrate the same properties as the original network $G(V, E)$ in the case of virus infection. In the

proposed solution, a neighboring pair of nodes $i \to j$ with a probability of virus transmission $0 < p < 1$ could be replaced by an appropriate number of nodes $i \to n_1 \to n_2 \to \ldots \to j$ with a probability of virus transmission $p = 1$ between each of them. The process of inserting intermediate nodes is displayed in the Figure 4.4.

Now, let us consider a case where the node i is infected and its neighbor j is susceptible to an infection. In a single time step, the node j will get infected with a probability p. Eventually, the node j will get infected after a certain number of trials. All the trials individually could be represented as an additional intermediate nodes with a transmission probability $p = 1$ between each of them. Therefore it is relevant to find out the number of intermediate nodes $\bar{k}$. The number $\bar{k}$ could be obtained from the discrete geometric probability distribution, as the number of trials before the success.

The geometric distribution, as a special case of the negative binomial distribution is a discrete distribution for $k = 0, 1, 2, \ldots$ with a probability density function

$$P(k) = p(1-p)^k, \tag{4.9}$$

where success probability is denoted by p where $0 \leq p \leq 1$ and the number of trials needed for a single success is denoted by k.

The (4.9), plotted in Figure 4.5 shows the distribution of number of intermediate nodes. For the network expansion procedure, for each edge, the value of $P(k)$ is randomly selected following the uniform distribution from the set of real numbers such that $0 \leq P(k) \leq p$. Then the appropriate number of additional nodes $\bar{k}$ is calculated from

$$k = \frac{\log\left(\dfrac{P(k)}{p}\right)}{\log(1-p)} \tag{4.10}$$

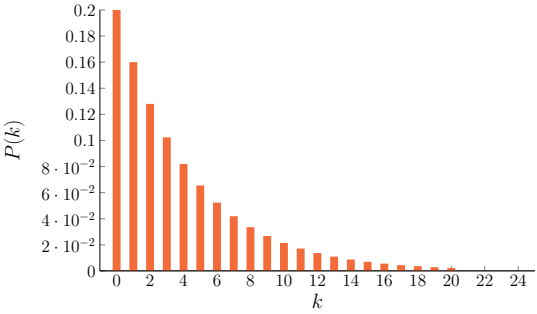


FIGURE 4.5: **Probability mass function of the geometric distribution.** Plot of the probability mass function of the geometric distribution for the probability of success $p = 0.2$. For the network expansion procedure for each edge, the value $P(k)$ is chosen randomly following uniform distribution where $P(k) \in [0, p]$. Then, discrete value of $\bar{k}$ is chosen accordingly.

The value k obtained from (4.10) is a real number $k \in \mathbb{R}$, and the number of trials before success, i.e. number of additional nodes $\bar{k}$ is integer $\bar{k} \in \mathbb{N}$ and obtained as a floor function of k as $\bar{k} = \lfloor k \rfloor$.

By adding an appropriate number of nodes between each pair of nodes in the original network, we can create the extended network $G_E(V_E, E_E)$. The extended network will consist of original and additional nodes $V \in V_E$. The probability of virus transmission in G_E is $p = 1$. We show that the infection dynamic in the original network G corresponds to the step response of the LTI system constructed from the extended network G_E. Note that for the LTI system analysis of the extended network $G_E(V_E, E_E)$ we only observe output at the subset of nodes V from the original network $G(V, E)$, so the system matrix C of the extended network is $C_E = C_E(V)$. The additional nodes are ignored as outputs, and this is the main reason the process is called *virtual network expansion*.

4.2.3 Numerical Simulations

In order to compare the proposed LTI approach against the conventional agent-based dynamics of infection, numerical simulations were performed. As the testbed we use randomly generated networks $G(V, E)$ with $|V| = 100$ nodes and $|E| \approx 200$ edges generated using three different network models. The results of the simulations are averaged over 100 individual runs. Two hypotheses have been tested:

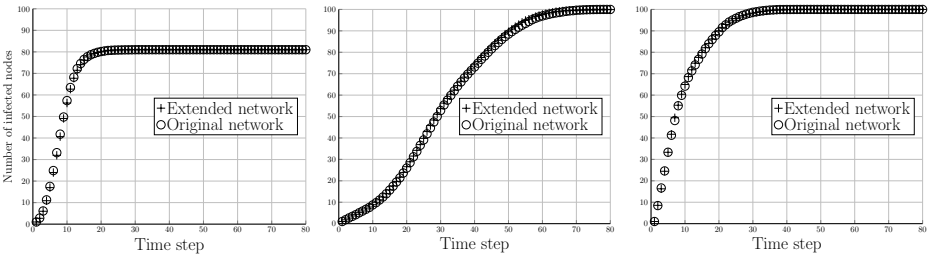
1. Can we transform the original network with $P = p$ to a new network with $P = 1$, which could exhibit the same spreading properties as the original one? A transformation is done by virtual network expansion.
2. Can we model the infection on the original network using LTI dynamics on extended network?

First, we test if the *virtual network expansion* could produce the secondary network with all probabilities of transmission $p = 1$ which performs similarly to the original network in the case of virus infection. For simulating the spreading dynamics in this case we use *susceptible-infected* (SI) epidemic model[2].

We conduct spreading simulations on three random networks generated by the following methods: Erdős-Rényi, Watts-Strogaz and Barabási-Albert graph[3]. The probability of transmission for the original graphs is chosen to be $p = 0.4$. Then, we extend the original networks $G(V, E)$ and conduct the simulations again with probability of infection $p = 1$ on extended networks $G_E(V_E, E_E)$. The virus infection is simulated starting at the same initial node for original and extended networks. We observe the cumulative number of infected nodes over time. In Figure 4.6 we can see almost perfect match of simulation results

[2]For more detailed explanation of all epidemic models used in simulations, see Section 2.4.1 and Section 4.3.5

[3]More information on networks used for the simulation are available in Section 4.2.4



(a) (b) (c)

FIGURE 4.6: **Spreading dynamics in the original vs. extended network.** [31] The simulation of virus spreading on the original $G(V, E)$ and extended network $G_E(V_E, E_E)$. The simulation is conducted on three randomly generated networks (a) Erdős-Rényi network, (b) Watts-Strogaz network and (c) Barabási-Albert network with $N = |V| = 100$ nodes and $M = |E| \approx 200$ edges and probability of infection $p = 0.4$. The plots show the number of infected nodes over time. The extended networks demonstrate very similar properties as the original one.

for the original and extended networks. We show that the method for network extension can be used for epidemic modeling[4].

Secondly, we compare the infection dynamics of the original network $G(V, E)$ and the response of LTI system obtained from the extended network $G_E(V_E, E_E)$. This way we show that it is possible to study spreading processes observing a response of a LTI system. The simulation process in form of a pseudo-code is presented in Algorithm 2.

Similarly to the example shown in the Figure 4.6, first the spreading dynamic is simulated using the SI model. In each time step t, the infected node attempts to infect the susceptible neighbor. The infection gets transmitted with probability $p = 0.4$. Therefore, the susceptible node gets infected with the probability $P = 1 - (1 - p)^k$, where k is the number of infected neighbors.

[4]Notice that the total number of simulated infected nodes in the ER graph is lower than 100. The reason is the method used for generating the network. In the given set of n vertices, each pair is connected with the probability p independently. Therefore, the model allows the creation of multiple connected components. The simulation is conducted on the largest connected component with a size $n \approx 80$

Algorithm 2 Comparison of SIR epidemic and LTI dynamics

procedure SIR SIMULATION
  Input: $G(V, E)$, p, i ▷ i - source of infection
  initialize the infection
  while *all nodes are infected* **do**
    identify all susceptible nodes
    calculate the probability for each susceptible node to get infected
    infect susceptible nodes
    calculate the number of infected nodes $\underline{v_u}(n)$ and $\underline{v}(n)$
  end
  return *the number of infected nodes*

procedure LTI DYNAMICS
  Input: $G(V, E)$, p, i
  for every edge **do**
    extend the edge according to the p
  end
  return *extended network* $G_E(V_E, E_E)$
  create an MST with the root in i
  identify the direction of the edges
  construct an LTI system
  calculate response $\underline{y_H}(n)$ and $\underline{y_\delta}(n)$
  return *step and impulse response*

compare the outputs

Data collected by simulation includes the *cumulative number of infected nodes* $\underline{v_u}(n)$ and the *number of infected nodes in each time step* $\underline{v_u}(n)$ which is the derivative of $\underline{v_u}(n)$.

Then, the LTI system is created from the extended network $G_E(V_E, E_E)$ with the probability of infection $p = 0.4$, following the procedure described in Section 4.2.2. In the Figure 4.7 (first row) the plots show the epidemic dynamic as $\underline{v_u}(n)$ simulated on the original network with the infection probability $p = 0.4$. The values of the $\underline{v_u}(n)$ are compared to the *step* response $\underline{y_H}(n)$ of the LTI system created from the extended original network. The results obtained from

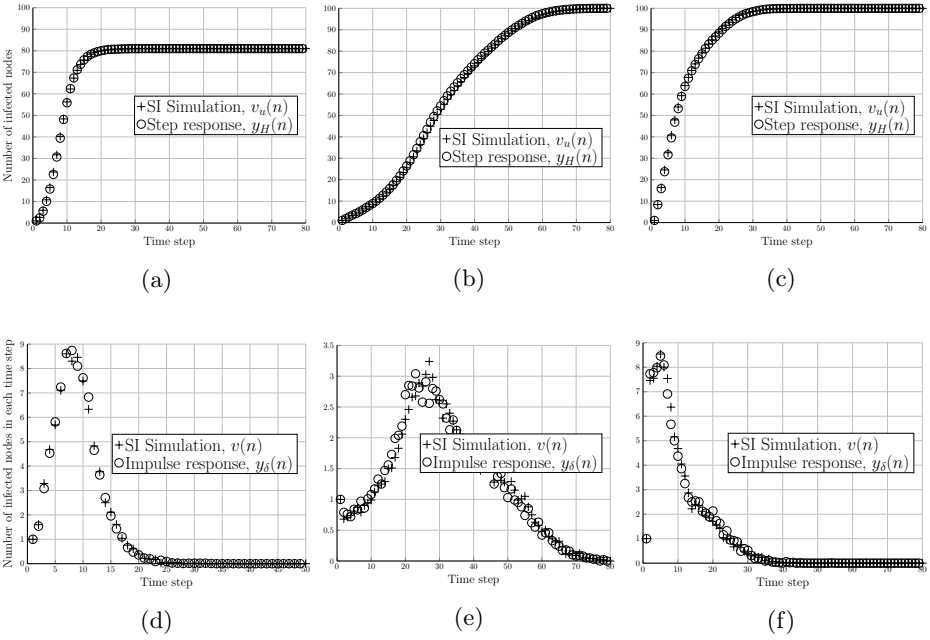


FIGURE 4.7: **Spreading simulation vs. system response.** [31] The simulation of the epidemic on the network $G(V, E)$ with $N = |V| = 100$ nodes and $M = |E| \approx 200$ edges and probability of infection $p = 0.4$. In the first row, the cumulative number of infected nodes is shown against the *step response* of the LTI system created from the extended network $G_E(V_E, E_E)$. In the second row, the number of infected nodes at each time step is compared against the *impulse response* of the LTI system created from the extended network $G_E(V_E, E_E)$. The simulation was conducted on three networks (a)(d) Erdős-Rényi random graph, (b)(e) Watts-Strogaz network and (c)(f) Barabási-Albert network. The results of the simulations are averaged over 100 individual runs. The plots show that proposed LTI model can predict the behavior of the epidemics in the network.

the SI simulation and the system response strongly correlate. The time evolution of epidemic behaves as predicted by proposed LTI model. Similar results are shown in the Figure 4.7 (second row). The results of the SI simulation $\underline{v}(n)$ are now compared against the *impulse* response $\underline{y_\delta}(n)$ of corresponding LTI system. This time, the number of infected nodes at each time step is compared to an output of a system excited by an impulse function. The correlations between the results are significant.

4.2.4 Network Data

Three types of networks used in the simulations are generated randomly following three different network models.

As an example of a simple random graph, we choose the model proposed by Erdős and Rényi [36]. The network is constructed from a set of N nodes and the edges are then generated[5]. The probability of an edge between two nodes is p and the probability that there is no edge is $1 - p$. The average degree $\langle k \rangle$ is calculated from the number of edges $\langle E \rangle$ generated in a graph $\langle E \rangle = \frac{1}{2}N(N-1)p$. Since each edge connects two vertices it is a part of a degree calculation for both of them. Therefore, we have $\langle k \rangle \simeq Np$.

The second group of networks is generated using the Watts-Strogaz model [40] which produces the networks which exhibit the *Small-World*[6] properties. The obtained networks show some properties, namely *clustering coefficient* and *average shortest path*, which are more similar to those manifested by the real-world networks. The small-world networks are highly clustered and have a short path lengths at the same time. The network is generated using random rewiring procedure for interpolating between a regular ring lattice and a random network, without altering the number of vertices or edges in the graph.

[5] For more detailed explanation of Erdős and Rényi model for random graph generation, see Section 2.1.1

[6] For more detailed explanation of small-world networks and Watts-Strogaz model, see Section 2.1.2

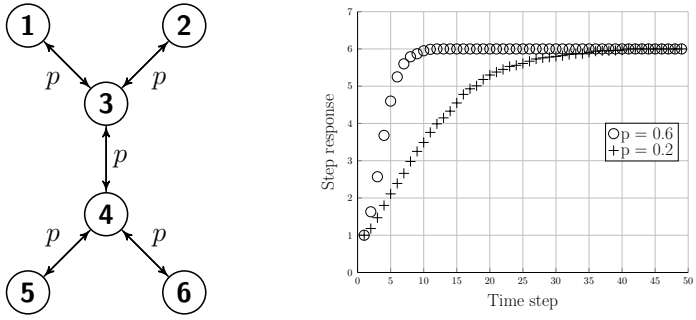
The third model we use, generates the network with *Scale-Free*[7] properties. We use the algorithm introduced by Barabási and Albert [43]. It is shown that for a number of systems, including the World Wide Web, citation networks and social networks, the degree distribution follows the power law. That is, the distribution of degrees $P(k)$ of nodes in the network with k connections to other nodes follows the power law function $P(k) \sim k^{\gamma}$ where γ represents a parameter with a value usually in a range $2 < \gamma < 3$. The network is generated following the principle known as *preferential attachment*, known also as a rich-get-richer phenomenon, the Gibrat principle or cumulative advantage [14].

4.2.5 An Example of Network Analysis

To demonstrate the possible application of the LTI approach in spreading dynamics analysis, here we provide a simple example. We take a small undirected network $G(V, E)$ with $|V| = 6$ nodes and $|E| = 5$ edges (Figure 4.8a) and we use LTI approach in two scenarios.

First, we show how the change of the transmission probability p affects the response of the corresponding LTI system. In the example network from the Figure 4.8a the topology remains unchanged, but the transmission probability changes from $p_1 = 0.6$ to $p_2 = 0.2$. The transmission probability is the same for all pairs of adjacent nodes and the infection is considered to originate from node (1). Based on the topology information $G(V, E)$, we build two LTI systems: the first one with $p_1 = 0.6$ and the second one with $p_2 = 0.2$. Then we calculate the step responses for the acquired systems and plot it in the Figure 4.8b. We observe the difference between the slopes of two obtained curves. The curve with the higher slope represents the step response of a system derived from the network with the *higher* transmission rate. Therefore, analyzing the response of a given system we can estimate the epidemic dynamics in the corresponding network. An inclination of a slope corresponds

[7] For more detailed explanation of Barabási and Albert model and Scale-Free properties, see Section 2.1.3



(a) (b)

FIGURE 4.8: **Simple network and corresponding LTI step responses.** [31] (a) An example of undirected network with $M = 6$ nodes and $N = 5$ links and probability of a transmission p; (b) The step responses of the LTI systems derived from the network using two probabilities of infection $p = 0.6$ and $p = 0.2$

to the rate of epidemic at each point in time. The same value corresponds also to the impulse response.

The response of a corresponding system can be also used for the network optimization in case of the variable transmission rate p. Let us consider the following scenario illustrated in Figure 4.9: a small undirected network $G(V, E)$ with $|V| = 6$ and $|E| = 5$, has an uniform transmission rate $p = 0.6$ between all pairs of nodes. However, we have a possibility to protect the network by changing the transmission rate of only one edge to $p' = 0.1$. Additionally, we are allowed to choose between two edges: $3 \leftrightarrow 4$ and $4 \leftrightarrow 6$. The goal is to optimize the network so the virus percolates slowly, meaning that it needs more time to infect all nodes.

In order to assess the optimal protection strategy, we make three corresponding LTI systems for each of the possibilities: without modification (all transmission probabilities remain unchanged, $p = 0.6$), edge $3 \leftrightarrow 4$ modified and edge $4 \leftrightarrow 6$ modified. Then we calculate the step response of all systems and observe the plots in the Figure 4.9b. The slope of a step response curve

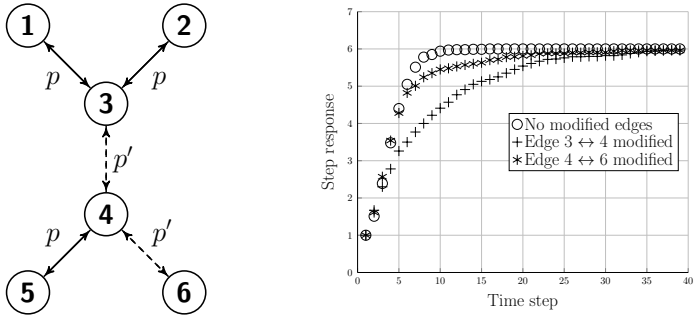


FIGURE 4.9: **Network protection example.** [31] (a) An example of an undirected network with $M = 6$ nodes, $N = 5$ links and two different probabilities of infection $p = 0.6$ and $p_1 = 0.1$. (b) The step response of the LTI systems obtained from a network according to the link chosen to be protected.

suggests the epidemic rate and the moment the curve reaches maximum value suggests the time needed for the full infection. By observing the curves in Figure 4.9b, we can conclude that the best solution is to lower the transmission rate of the edge $3 \leftrightarrow 4$.

The network presented here is rather simple and the optimization decision trivial. Still, this example illustrates how an LTI approach can be used to analyze and optimize the networks against epidemics. This approach has some weaknesses which should be addressed. The main problem is the network modification which has to be performed in order to make this model work. It requires additional topology modification by virtual network extension. However, the results suggest that the state of a network during the epidemic could be calculated for any given time, even without using the agent-based simulation process. Additionally, this approach opens the opportunity to introduce a set of tools from the system theory in epidemic modeling.

4.3 Identifying the Influential Spreaders

There are numerous attempts to precisely evaluate the node importance in the network. Most of the approaches used in the literature are based on various *centrality measures* and their variations. Centrality measures such as *degree*, *betweenness*, *closeness*, *local rank*, *h-index*, *Katz centrality* and *eigenvector centrality*, try to quantify how much the node is centrally positioned within the network regarding the topology [4, 46, 100, 101]. Various centrality measures are designed to quantify the node importance for different spreading processes and not all of them are useful in all types of networks. For example, to identify most influential spreaders in social networks, k-shell is more reliable then degree [102]. Centrality measures easily identify the hubs, but usually fail to capture the spreading potential of the numerous peripheral nodes [103]. Those nodes are in majority and they are most likely to become the source of the infection. To capture the potential influence of non-hubs, other metrics should be used such as *expected force* [85]. There is a growing specific group of measures which all try to explain more specifically the *spreading power* or the *potential influence* of all the nodes in the network. They do not always rely on the path lengths and distances as most of the conventional centralities do. The paramount objective is to determine the important *spreaders* [34, 49], nodes able to spread the infection quickly through the network. Besides the *expected force*, for that purpose we can use: *k-shell* [104], *k-truss* [32], *percolation* [72], *accessibility* [105] or *dynamic influence* [33].

Another measure named *Node Imposed Response (NiR)* proposed first in [35], which captures the node's spreading potential is explained here. It can accurately classify the most important nodes based on their possible spreading influence. A theoretical background used as an initial rationale behind the proposed approach supports the later acquired simulation results. The measure outperforms *betweenness*, *degree*, *coreness* and *h-index* centrality in identifying the most influential spreaders in case of the SI and SIR spreading processes. Even the *NiR* does not depend on any parameters, its performance

is comparable to the centrality measures such as *dynamic sensitive centrality (DS)* [51] which use parameters for better fitting to the spreading dynamics.

The proposed *NiR* measure utilizes concepts from LTI systems theory. More specifically, the *NiR* measure is based on the value of the system response to the input step function. Here, we show that LTI approach could be used for assessing the node spreading power. Furthermore, it could be fitted to identify various possible influences from a single or multiple sources to a single or multiple end nodes. The modifications to the original measure could be done simply by manipulating the corresponding system matrices.

4.3.1 Calculating the NiR

NiR is the normalized maximum value of the step response S_i for the corresponding LTI system with the node i as the input. Let us define the maximum value of step response for the node i as S_i, then

$$S_i = \max_{1<t<k} y_i(t) \tag{4.11}$$

The function $y_i(t)$, derived from (4.6) is concave and eventually reaches its maximum value for t large enough. Therefore S_i will always exist. Then

$$NiR(i) = \frac{S_i - S_{min}}{S_{max} - S_{min}}, \tag{4.12}$$

where $S_{max} = \max_{j\in\{1,\dots,n\}} S_j$, $S_{min} = \min_{j\in\{1,\dots,n\}} S_j$, and n is the number of nodes in the network.

In order to calculate S_i we have to construct the corresponding LTI system which is defined by system matrices A, B, C and D. All the matrices are created following the procedure explained in the Section 4.1 using the (4.3) and (4.4). Before creating the matrix A, some modifications have to be performed on the original graph. In order to maintain the system's bounded-input, bounded-output (BIBO) stability, the topology should be modified so

the cycles are removed (see Section 3.2 for details). The *NiR* could be calculated only for the acyclic directed graphs. An alternative modification of *NiR* called *reduced NiR*, could be calculated for any graph and it is discussed in Section 4.3.6. For the spreading processes, such as SI and SIR, cycles could be considered as irrelevant as the nodes cannot be infected twice. Also, removing the edges which form the cycles should not significantly influence the spreading dynamic. However, the algorithms for cycle removal do change the topology in the way some paths become excluded, especially for the undirected networks where one has to choose between the edge directions. Therefore, the proper way of removing cycles has to be chosen in order to maintain the most important paths from the source node and to introduce the minimal number of removed edges. An algorithm for making an acyclic graph is described in Section 3.2.

After the manipulation of the original graph we create a system matrix A such that $A = A_{adj}^T$. All non-zero entries are substituted with the value d so $\forall a_{ij} \neq 0 : a_{ij} = d$, and $0 < d < 1$. We can choose any d between 0 and 1. However, choosing $d \ll 1$ is preferred. Supplementary investigation shows the variance between *NiR* values for all the nodes becomes higher for smaller d. In the case of any relatively large network, the variance between node measures become more important for proper node classification, as there is a large fraction of non-hubs, nodes with very similar and small spreading power. In this case, the low variance could lead to false estimation, especially considering the removal of cycles and some of the topology information which gets lost during the process. Therefore, choosing the smaller d is important for proper node differentiation. In our simulations we use $d = 0.1$.

The matrix $B_{[1\times n]}$ is consisted of all zeros except for the input node i which is evaluated, in which case $b_{1,i} = 1$. The matrix $C_{[1\times n]}$ is a vector of all ones, as the output is observed in all nodes and it is not weighted.

The step response of the obtained system will eventually reach the maximum value S_{max}. That particular value calculated for the input node and normalized over all nodes within the range $[0, 1]$ is the *NiR*.

4.3.2 Small Network Example

In Figure 4.10 an example of a small network with $n = 10$ nodes is shown. Each of the nodes have its *NiR* value written above. In order to calculate the *NiR*, the topology has to be modified to make a network acyclic. This procedure depends on the chosen source node (Section 3.2). The modification is performed for every node independently. In this example, we see two versions of the topology with two nodes as the sources: node with ID1 on the left and node with ID10 on the right. The *NiR* value indicates the node's spreading power, which means the node with higher *NiR* will infect the entire or the large fraction of the network faster.

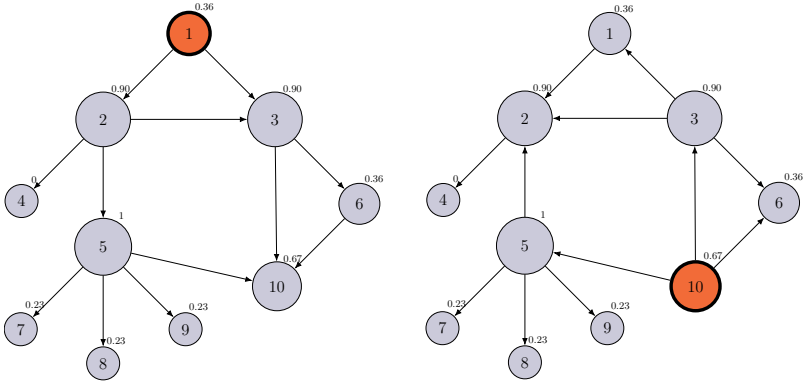


FIGURE 4.10: **The *NiR* of all nodes in the network.** For each node, we can calculate the *NiR*. Initially, the undirected network is made directed and acyclic with the respect to the source node. This is a necessary step in order to maintain the system's stability. The LTI system is then formulated having in mind the new topology and the source. The value of the maximum step response of the corresponding system is the *NiR*. There are two topologies depicted for two observed nodes: node 1 (left) and node 10 (right). The normalized *NiR* values are shown above each node. The radius of the node represents the *NiR* value (the larger the radius, the larger the *NiR*). Here we can identify the node 5 as the one with highest *NiR*.

The claim is supported by simulating the SI spreading dynamics and comparing the results with the obtained NiR values. The simulation was performed as following: the infection originates in a single node; the infection spreads following the transmission rate p, and eventually covers the whole network; the time needed for a full infection is then calculated. If the time of full infection is shorter, the node has a potential to spread the infection faster and is considered more important (i.e. more influential). In order to compare NiR value and simulated spreading potential, we sort nodes both by their NiR and by spreading power obtained by the simulation. This way we identify several distinct groups of nodes with different spreading potentials (Figure 4.11). In case of the small network example, the NiR value accurately captures the spreading potential, as the node grouping is the same as obtained by numerical simulation. It is likely that for large networks where $n \gg 10$, there will be many nodes with very similar NiR values which is in accordance with the innate scale-free principle of many networks, with a large fraction of non-hubs.

4.3.3 Simulation Results

In order to validate the assumptions from the previous section, first we simulate the SI and SIR spreading processes on the set of networks and then compare it to the NiR measure. The results show a significant correlation between the NiR and the simulation results for all the families of networks used for the analysis. The networks used for the simulation are listed in Table 4.1 and more precisely described in Section 4.3.7. The correlation diagrams are shown in Figures 4.12 and 4.13.

The simulations were performed on several networks using SI and SIR models. The benchmark value for the SI model is the time t needed for partial (50% or 70% nodes) infection in the case of a single source node i. For SIR model, the value used for the comparison is an outbreak size (the total number of nodes which got infected) after t time steps. More on spreading models and relevant measures in Section 4.3.5. The results collected by simulations for

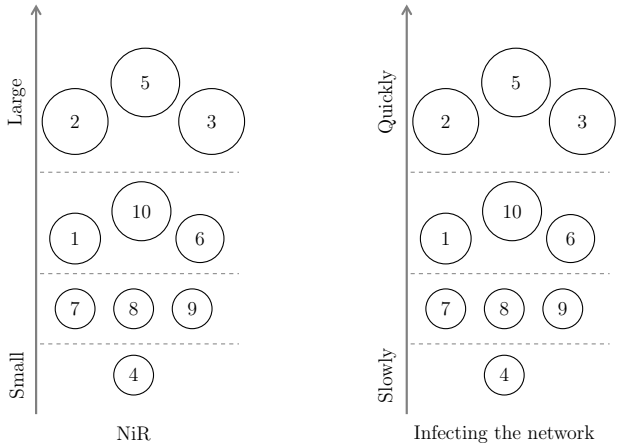


FIGURE 4.11: **Nodes in the small network classified by the importance.** On the left side, the nodes are ranked by their *NiR* value. Then, the SI infection is simulated for each source node and the time measured until the infection reaches all nodes. The right-hand picture shows the node ranking by the time infection needs to spread if the infection starts from that particular node. If the time for full infection is shorter, the node is ranked higher. The *NiR* measure could capture the node's spreading power and put it in the right category. The distinct separation in groups by the spreading potential is made for the sake of the better visual presentation.

each of the nodes are then compared against the *NiR* and five other centrality measures (betweenness, coreness, degree, DS and H-index centrality). The *NiR* measure demonstrates high correlation to simulation results together with the low variance, often outperforming all five measures both in SI and SIR model. The only measure which performs equally is a DS centrality whose parameters depend on the dynamics.

In Figure 4.12 and 4.13 the violin plots represent the correlation distributions. The figures are divided in eight sections for each of the networks used in the analysis. For each centrality measure the violin plot displays the distribution of the correlations acquired during the spreading simulation. The vertical

position of a single violin represents the correlation value, which means the higher the position the stronger the correlation between the particular centrality measure and the simulation results. Likewise, if the plot is positioned low, the correlation is weaker. The width of the violin represents normalized correlation frequency obtained from multiple experiments for the stochastic simulation process. The vertical length of the violin represents the variance which implies the robustness of the measure. If the violin is short, the measure correlates with the spreading dynamics most of the time. Likewise, if the violin is stretched, the variance is high and measure is not always reliable. All the correlations are quantified by Kendall's τ rank correlation coefficient, explained more in details in Section 3.2.

Note that all violin plots are smoothed for the sake of the better presentation. For smoothing, we estimate the probability density function of the observed correlation distribution using normal kernel density with kernel density estimate as $KDE = 0.15$ [106].

Not all the nodes are used for the correlation measurement. In case of large networks, there will be a large fraction of nodes with very similar spreading potential, so the incremental difference in centralities is negligible. Therefore it is justified to take a representative sample of nodes and use them for analysis. Here, we sort nodes based on their NiR value[8]. Then we divide the sorted set into 10 equal pools. From each pool we pick 8 nodes uniformly at random. In total we use 80 nodes for the comparison. This way we choose a representative set of nodes randomly but still avoiding the random generator to choose many similar nodes. We assume that a higher resolution would not add up to the precision while significantly increasing the need for more computing power.

[8] Additional examination shows that choosing another measure for sorting does not make significant difference

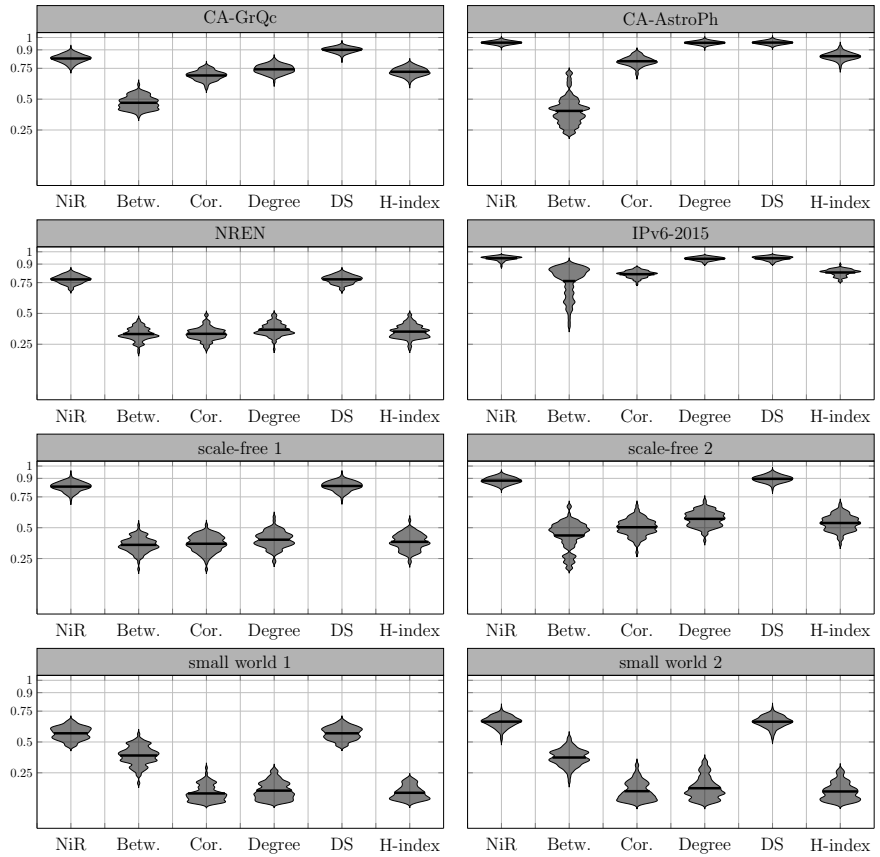


FIGURE 4.12: **Correlation of *NiR* and centrality measures to the spreading outcome on simulated networks - the case of the SI model.** Violin plots show the distribution of correlations between observed spreading dynamic and various centrality measures on 100 generated networks from each network family. Graphs are generated from the sample degree sequence of the real graphs. The correlation with *NiR* is relatively strong and outperforms betweenness, coreness, degree and H-index centrality with higher mean values and low variance. The vertical position of the violin plot demonstrates the correlation coefficient over all observed samples: the higher the position, the stronger the correlation. Additionally, the length of the plot indicates the variance: the more stretched plot, the bigger the variance. Therefore, the preferred plot is narrow and positioned high on the grid. The SI spreading process is used as a reference and it is computed as the time the infection reaches at least 50% of all nodes starting from the source node for the spreading rate of $p = 0.05$. The infection time is calculated as a mean time for 300 simulated processes for each observed node. The horizontal line within the plots shows the arithmetic mean of the correlation coefficients.

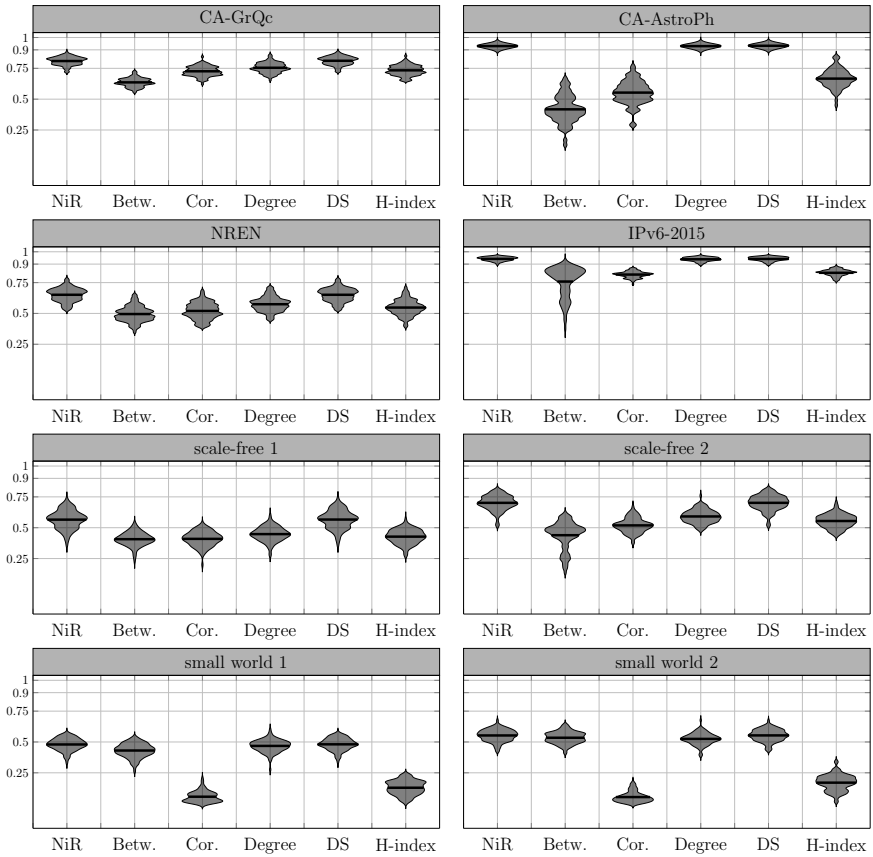


FIGURE 4.13: **Correlation of *NiR* and centrality measures to the spreading outcome on simulated networks - the case of the SIR model.** Violin plots show the distribution of correlations between observed spreading dynamic and various centrality measures on 100 generated networks from each network family. Graphs are generated from the sample degree sequence of the real graphs. The correlation with *NiR* is relatively strong and outperforms betweenness, coreness, degree and H-index centrality with higher mean values and low variance. The vertical position of the violin plot demonstrates the correlation coefficient over all observed samples: the higher the position, the stronger the correlation. Additionally, the length of the plot indicates the variance: the more stretched plot, the bigger the variance. Therefore, the preferred plot is narrow and positioned high on the grid. The SIR spreading process is used as a reference and the benchmark measure is the outbreak size of the infection when the spreading rate is $p = 0.05$ and recovery rate is $\mu = 1$. The outbreak size is calculated as a mean outbreak size for 300 simulated processes for each observed node. The horizontal line within the plots shows the arithmetic mean of the correlation coefficients.

Violin plots represent the correlations for simulation of the epidemic with already fixed parameters. Additionally, we simulate SI and SIR spreading process on two real world networks (*NREN* and *CA-GrQc*) for various values of transmission probability where we choose several incremental values of p ranging between 0.01 and 0.1. As shown in Figure 4.14, *NiR* performs well in both networks for both spreading models, clearly outperforming degree, H-index, coreness, and betweenness centrality. For the SI model, *NiR* performs equally as well as DS centrality, even though it does not use any additional parameters from the spreading model. It is evident that the correlation between the simulated dynamics and centrality measures drops when p increases. This suggests that evaluating the node's spreading potential for large transmission probability becomes more difficult.

4.3.4 Reasoning Behind

The main hypothesis leading to the *NiR* proposal is that the time needed to infect the fraction of the network correlates with the step response of a corresponding LTI system. Here we show the example of three simple graphs given in Figure 4.15 which demonstrate that the value of the step response could be used to predict the simple spreading dynamic. We further argue that the same principle could be used for a directed acyclic graph of an arbitrary size. This is supported by the results of the numerical simulations illustrated in Figure 4.16.

The time delay between the initial infection and the full infection of a network is a function of the probability of virus transmission p between the infected and susceptible node. This relation is reciprocal: the smaller the p, the longer it will take for a full infection. By observing the probabilities of transmission within the graph, we can derive the expected time of all nodes to be infected, $E[X(p)]$. The $E[X(p)]$ is the expected value (mean) of X, where X is a discrete random variable. Expected value is calculated as a weighted average of the possible values that X can take, each value being weighted

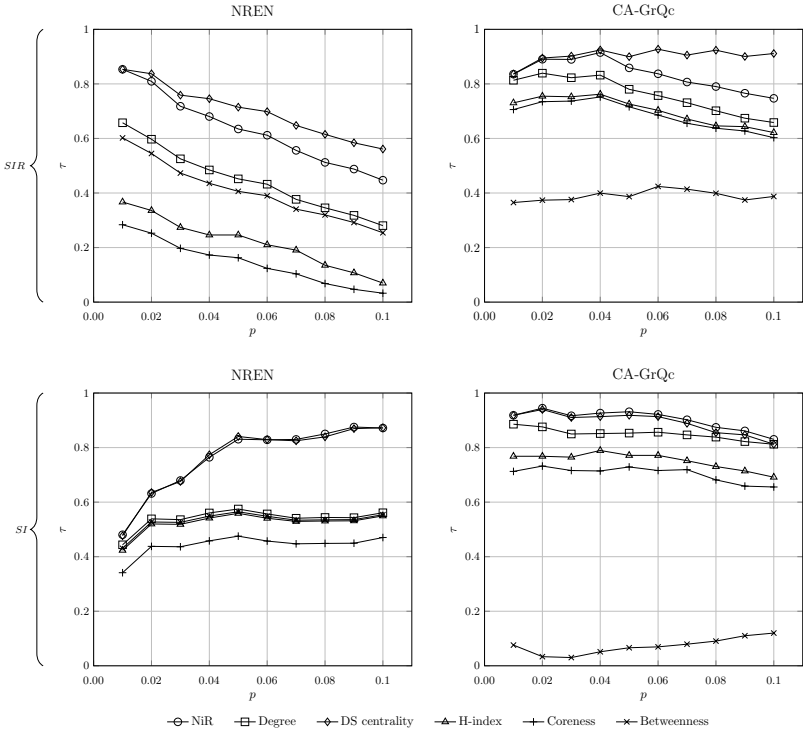
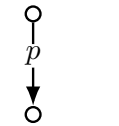


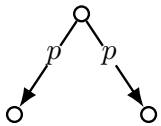
FIGURE 4.14: **Correlation between centrality measures and spreading potential evaluation for various p.** The probability of infection p takes a value from 0.01 to 0.1. Each data point is obtained by averaging over 10^4 individual runs. Plots are generated according to the SIR model (upper half) and SI model (lower half). All correlations are quantified by the Kendall's Tau coefficient τ.

according to the probability of that event occurring. It is a monotonic function for all nodes and therefore it could be used for the node's ordering. On the other hand, the maximum value of the step response of the corresponding LTI system $S_{max}(p)$, presented as a function of p is monotonic as well, although increasing. The corresponding system is derived from the acyclic directed graph with non-negative values of the A matrix elements in the range $0 < a_{ij} < 1$. The example of three small networks in Figure 4.15 and their



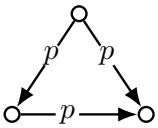
$$E[X_1(p)] = \frac{1}{p}$$

$$S_{max1}(p) = 1 + p$$



$$E[X_2(p)] = \frac{2}{p} - \frac{1}{2p-p^2}$$

$$S_{max2}(p) = 1 + 2p$$



$$E[X_3(p)] = \frac{3}{p} - \frac{1}{2p-p^2} - \frac{1}{3p-3p^2+p^3}$$

$$S_{max3}(p) = 1 + 2p + p^2$$

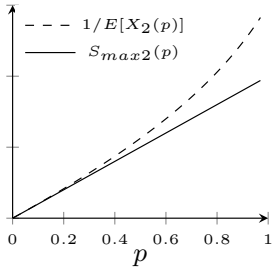


FIGURE 4.15: **Expected time of infection and step response: small networks example.** For three small networks the expected time of full infection, $E[X(p)]$, is calculated. For all networks the source of the infection is the parent node (top). At each time step the parent node tries to infect neighboring susceptible nodes with the probability p. All nodes will be eventually infected and the time of full infection is presented with a certain distribution (i.e. the distribution of the expected number of trials in discrete time for the infection to reach all nodes). The $E[X(p)]$ is the mean of the distribution for each network (the expected number of trials before the success). The $S_{max}(p)$ is the maximum step response value of the corresponding LTI system.

corresponding $E[X(p)]$ and $S_{max}(p)$ demonstrate the monotonic nature of observed functions.

The illustration of the phenomenon becomes clear in this small example. Calculating the $E[X(p)]$ for slightly larger graphs already becomes very complex.

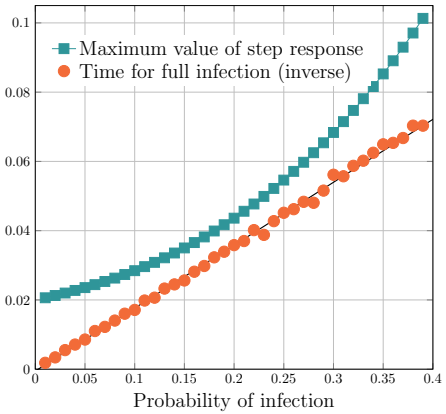


FIGURE 4.16: **The comparison of the step response and the infection time.** Here we compare the time the infection reaches all nodes and the maximum value of step response of the corresponding LTI system for the various probabilities of infection. For the sample random acyclic graph with 100 nodes, the corresponding single input LTI system is constructed. The non-zero values in A matrix range from 0.01 to 0.4, to capture the system behavior with various infection probabilities. The system is excited by step function and the maximum value of the step response is presented with data points as squares. The results indicate the exponential trend. On the other hand, we simulate the SI infection dynamic starting from the same node. The simulation is conducted for 40 different probabilities of infection (0.01-0.4). The time for the infection to reach all nodes is then measured (i.e. the time when the network becomes fully infected). For comparison, the inverse value is plotted with data points as circles and the black line exhibits the linear trend. Both curves are monotonically increasing.

To derive the expected time $E[X(p)]$ of infection for larger non-regular networks rapidly becomes too difficult as the number of nodes increases. However, the simulation results on those networks suggest the same monotonically increasing trend (Figure 4.16). The time the infection will spread rises monotonically with the increased probability of infection p, which is expected. At the same time, the maximum value of the step response follows the similar trend. This leads to the conclusion that those two measures (*expected time*

of the infection and *maximum step response*) could be used interchangeably, except that S_{max} is considerably easier and faster to compute.

4.3.5 Spreading Models

SI (Susceptible-Infected) epidemic model. The SI model differentiate between two possible states of the nodes: *a)* susceptible to infection and *b)* already infected and able to spread the infection. In SI model, the transition from *a)* to *b)* is possible, but not vice-versa. The epidemic process starts with a fraction of initially infected nodes. All the neighboring nodes are considered susceptible. In each time step, the infected node attempts to transmit the infection to all of its susceptible neighbors independently with a transmission rate p. For the healthy (uninfected) node i with k infected neighbors, the probability of infection in each time step is $p_i = 1 - (1 - p)^k$. When the susceptible node gets infected, it remains in that state indefinitely and it is able to spread the infection further. The infection process in the connected graph will eventually affect the whole network until all the nodes switch state from susceptible to infected. We simulate the SI infection which originates from a single node belonging to a set of chosen nodes. Then we measure the time t in a form of a number of time steps needed for infecting the 50% of all the nodes. The spreading process is repeated over 300 times for each of the chosen nodes as source. The average time t is used as a basic benchmark. The NiR values for the same set of chosen nodes is then compared and the correlation is measured. The same process is repeated for 100 different networks derived from the each network family.

SIR (Susceptible-Infected-Recovered) epidemic model. In the SIR model, each node could take one of three states: *a)* susceptible to infection, *b)* already infected and able to spread the infection and *c)* recovered (removed) and not able to infect other nodes. The state transition is unidirectional from *a)* to *b)* to *c)* and not vice-versa. The epidemic process starts with a fraction of initially infected nodes. All the neighboring nodes are considered

susceptible. In each time step, the infected node attempts to transmit the infection to all of its susceptible neighbors independently with a transmission rate p. Simultaneously, all already infected nodes get recovered from infection with the recovery rate μ. For the healthy (uninfected) node i with k infected neighbors, the probability of infection in the next time step is $p_i = 1-(1-p)^k$. At the same time the probability of recovery remains the same regardless of the node's surroundings. We simulate the SIR infection which originates in a single node belonging to a set of chosen nodes. The recovery rate is considered to be $\mu = 1$, which assumes that all of the nodes infected in time step t' will get recovered in $t' + 1$. The results should be very similar for other values of μ [51]. The benchmark value is an outbreak size n_{inf} (the number of infected nodes) after t time steps.

4.3.6 Limits of the NiR

The main drawback of the *NiR* measure is the need to additionally modify the topology in order to obtain the directed acyclic graph. The motive for making the graph acyclic is the BIBO system stability. A system is BIBO stable if there is a bounded output for every bounded input over the time interval $t \in [t_0, \infty)$. Since the *NiR* is defined as a normalized *maximum* value S_{max} of the step response, the system has to have bounded output. The response of the unstable system quickly reaches extremely high absolute value while for the stable system, the response remains within the bounds. However, the systems observed here are not real physical systems, but rather their mathematical model, hence there is no actual danger of producing the unstable electrical circuit. For the sake of the measurement we can allow the system to be unstable and let cycles exist. In that case, we have to read the response quickly after the initial excitement, just after few time steps. That is why this proposed approach is called *reduced NiR*. Reducing the number of steps before measuring the output is justified, and the following analysis supports it.

In order to evaluate the spreading power of the node, the most important area of interest is its surrounding. The importance of topological information decays quickly with the distance from the observed node. Therefore, almost all centrality measures could rely with high confidence on local neighborhood information [85, 107]. The *NiR* shows a similar property. The signal strength decreases with each time step as it is being amplified with the factor of $a_{ij} \ll 1$. Even though multiple incoming edges boost the signal strength by summation, each node decreases the resulting signal with the parameter a_{ij} at the same time. Let us consider the line graph consisting of n nodes connected consecutively with $n-1$ edges. If the corresponding LTI system derived from the network gets excited by an impulse at the first node, and we consider the $a_{ij} = 0.1, \forall a_{ij} \neq 0$, the signal strength after t time steps would be amplified by the order of 0.1^t. The observations made on any nodes positioned even further from the source could be irrelevant for the *NiR* assessment. This property makes it possible to evaluate the node importance with *NiR* using just the knowledge of the local topology.

4.3.7 Network Data

There are two types of networks used in the simulations. The first type are the *scale-free* and *small world* networks constructed randomly using various parameters (Table 4.1). Scale-free networks are constructed based on the Albert-Barabasi model of preferential attachment [43] using the algorithm described by Batagelj [108] and implemented in "A Controllable Test Matrix Toolbox for MATLAB" [109]. The first group of scale-free networks has a minimum node degree of 1, while the second group has a minimum node degree of 2. Those parameters affect the diameter and the density of networks, and therefore the expected dynamic of spreading processes. Similarly, the small-world networks are constructed with the same MATLAB tool using the Watts-Strogatz model [40]. The Watts-Strogatz model is based on two parameters which define the number of nearest neighbors to connect (k) and the probability of adding the shortcut in the given row (p_s). The first group

of generated small world networks has $k = 1$ and $p_s = 0.5$, while the second group has $k = 2$ and $p_s = 0.5$. The variety of initial parameters ensures the generation of networks with different properties such as *diameter*, *density* or *the average degree*. Randomly generated networks are connected, undirected and consisted of 6000 nodes each.

network	*nodes*	*diameter*	*density*	*avg. degree*	*clust. coeff.*	*source*
scale-free 1	6000	14.69 ± 4.68	$7.09e^{-06}$	3.40	0.0944	Generated
scale-free 2	6000	17.78 ± 10.22	$8.75e^{-07}$	4.25	0.1148	Generated
small-world 1	6000	41.46 ± 31.54	$3.95e^{-06}$	3.08	0.1602	Generated
small-world 2	6000	26.77 ± 16.23	$4.99e^{-07}$	5.00	0.2070	Generated
CA-AstroPh	18772	9.54 ± 3.45	$1.10e^{-07}$	21.25	0.2143	SNAP
ca-GrQc	5242	7.93 ± 4.50	$4.17e^{-07}$	6.13	0.5296	SNAP
IPv6-2015	34761	5.19 ± 9.81	$2.37e^{-07}$	10.54	0.0853	UCLA
NREN	1157	20.69 ± 7.31	$2.97e^{-06}$	3.21	0.0994	Topology Zoo

TABLE 4.1: **Generated and extracted networks.** Four networks are generated using Barabási-Albert and Watts-Strogatz models for *scale-free* and *small-world* networks respectively. The rest are the real world networks of various sizes and characteristics taken from: *SNAP - Stanford Large Network Dataset Collection*, *UCLA's Beyond BGP:Internet Topology Project* and *The Internet Topology Zoo*. All data sets are available online. Column *nodes* represents the number of nodes in the original network. Columns *diameter*, *density* and *clust. coeff.* represent the mean values calculated from the set of sampled networks. The *avg. degree* is the same for both the original and sampled networks.

The networks in the second group are derived from the large real-world networks data. These real world networks are taken from various network dataset repositories: *SNAP Datasets: Stanford Large Network Dataset Collection* [110], *UCLA's Beyond BGP:Internet Topology Project* [111], and *The Internet Topology Zoo* [96]. Two networks represent the collaboration pattern between authors of the papers submitted to arXiv; the *ca-GrQc* for the General Relativity and Quantum Cosmology category, and the *ca-AstroPh* for the Astro Physics category. The other two are technological networks which illustrate the topology of networked systems. The Internet AS-level topology network (*IPv6-2015*) is the monthly snapshot of AS-to-AS links as they appeared in the January 2015. The European network of National Research and

Education Networks (*NREN*) is the backbone network managed by GÉANT which connects all European scientific and research institutions.

For all the real world networks except the NREN, the simulations are performed on the network samples obtained from the original network data. The sampled networks are characterized by the degree distribution. For each of the real world networks we generate 100 sampled networks with 1000 nodes each. For each of the random realizations of the topology, we measure the correlation of SI and SIR spreading dynamic and observed measures. Simulated networks are generated by taking 1000 nodes samples from the original network uniformly at random without repetition. The degree sequence is then extracted from the sample. Networks are then constructed from the obtained degree sequence using the Havel-Hakimi algorithm [112]. This algorithm does not guarantee the construction of a connected graph. Therefore, the simulation is performed on the largest connected component on each of the generated networks. Constructing graph from a degree sequence preserves the degree distribution of the original network. Some network characteristics such as communities, get lost during the process. However, the relative size of the generated graphs prevent the replication of community structures anyways [85]. For the type of dynamics simulated here on unweighted undirected networks, some other characteristics such as costs, constraints and direction on edges are irrelevant [113] and therefore ignored.

The node importance is usually characterized by some of the numerous centrality measures. Here we compare our proposed *NiR* measure against the most commonly used centralities such as *degree*, *betweenness*, *coreness*, *h-index* and one parametrized centrality called *dynamic sensitive centrality*. For a detailed explanation on each of the centralities used for the assessment, see Section 2.2. Furthermore, all networks used in simulations are characterized by various global properties. Network attributes such as *diameter*, *density*, *average degree* and *clustering coefficient* are used to recognize the network model and to identify in which extend network topologies contrast to each other.

4.3.8 Conclusion

The proposed NiR metric can successfully capture the possible node importance when it comes to epidemic dynamics for various network models. The results of the numerical simulation show the high correlation with the actual spreading dynamics modelled by SI and SIR processes. The NiR also shows a small variance, which means it is reliable for different topologies. The underlying paradigm of the LTI approach allows the numerous variations of the original metric. For example, the number of input and output points in the system could vary. By choosing the multiple input points, it is possible to estimate the influence of many nodes if they would be excited at the same time. Furthermore, the choice of multiple output points would give an estimate of an exposure of those nodes in case of the spreading process. More exposed nodes are more likely to be reached from the set of chosen input nodes. The analysis is not limited to the unweighted networks. The same approach could be used even for the weighted networks just by including the weights in the system matrix A.

Chapter 5

Cascading Failures Analysis Within the European NREN

The main role of the networks regardless of its nature is to provide a suitable medium to transport a certain type of goods. Whether the goods are vehicles in the case of the transportation networks, data in the case of the communication networks, or even the rumors in the social networks, same fundamental phenomena could be observed. In this context, congestion and breakdowns are in the focus. The analysis shows they can have a major effect to network's efficiency and connectivity.

Designers create the networks having in mind certain requirements, but at the same time they are bounded by various limitations. Capacity requirements for the links in the communication networks have to be met in order for a network to work properly. If the links are designed with insufficient capacity, the network's traffic demands would not be fulfilled. However, the capacity of a link is limited by the cost. Finding the optimal and sustainable solution for a traffic capacity in the communication network is a sophisticated problem. The commonsencial approach would be to tune the links in such a way so that their capacity slightly exceeds the initial traffic load. Although this simple approach would result in the functional network, there are two main reasons why such concept fails in reality. Those are, namely *network growth* and *node failures*. The hypothetical unlimited capacity would solve both problems, but achieving higher capacity comes with a cost and can easily become irrationally

expensive if not planned properly. Furthermore, dense networks are more likely to be robust to cascading failures, but more edges always mean more resources and a higher cost [114].

The natural tendency of all technological networks is to grow or evolve. The most common growth model based on a *preferential attachment phenomena* is explained by the work of Albert and Barabási [43]. Some details on other growth models could be found in the work of Newman [6]. As the growth occurs, with each new node, the traffic demand on already existing nodes increases. Furthermore, the communication networks have to be designed in the way which supports ever rising trends in the service demands even in the cases with constant number of nodes.

Failures are generally inevitable in technological networks. The reasons for failures are various and include random technical failure, geographically correlated failures due to naturally caused disasters or even due to intentionally caused failures as a result of terrorism or diversion. Those failures occur with different frequency and cause various problems. In the functional network, the traffic is distributed following certain routing strategies. Therefore, the loads of the links and nodes are considered balanced and within their limits. In this case, a network is in a stable state. When the breakdown occurs, the topology of a network suddenly changes. The optimal traffic paths through the network subsequently get adjusted. Some nodes possibly get overloaded, which causes congestion at the certain segments of the network. It demands further traffic allocation which could generate additional congestion as a result.

Communication networks are modeled as networks in which the nodes are assumed to be generators, hosts and routers for the information packets. In the majority of the models every node could have any of those three roles, although some models differentiate the nodes by role [115]. The topology could vary, from randomly generated networks to regular lattice topology. If the packets are generated in the network with the rate R, it implies that on average $N \times R$ packets are generated in each time step, where N represents the number of nodes in the network available to generate the packets. Beside

its origin, each packet has its destination node, which also could be decided randomly. After the packets are formed, the routing procedure is initiated and the packets get transported to their destinations following certain routing protocol. For the sake of the modeling simplicity, the most commonly used protocol relies on the *shortest path* between origin and destination node. The moment the packet reaches its destination it gets removed. Routing nodes could also have a buffer, which allows the node to store incoming packets and redistribute them later. Using this relatively simple model, the congestion properties of the network could be derived. Many authors discovered the distinctive phase transition where the network shifts from the free flow to the congested state. It is shown that the phase transition parameter strongly depends on the rate R of generated nodes [116]. Arenas et al. define the global order parameter η:

$$\eta = \lim_{t \to \infty} \frac{n(t + \Delta t) - n(t)}{NR\Delta t} \tag{5.1}$$

If the $\eta \leq 0$, the system is in the free flow phase and is able to route all packets to its destination. Otherwise, if the $\eta > 0$ the system eventually enters the congested phase. The quantity $NR\Delta t$ is the number of packets generated in time Δt and $n(t)$ is the number of packets in the network at time t.

With the increasing value of R, the system has more chance to become congested. Having in mind the typical phase transition behavior from the free flow to congested state, there is a critical value R_c which describes this phase transition. Generally, for all $R < R_c$, $\eta = 0$ the system is in the free flow phase. However, if $R > R_c$, η increases rapidly and the system becomes congested. Therefore, the critical value of R is R_c and it is considered to be an overall capacity of a system [117]. The value R_c could be obtained numerically for any network, but in some special cases it is possible to calculate it.

The congestion phenomena relates to the traffic in the network, and traffic patterns are imposed by routing strategies. For the majority of calculations, the routing policy can be approximated by the *shortest path* between source

and the destination node[1]. In order to analyze the load in the case of the shortest paths strategy, we can make use of the *betweenness centrality* which shows the number of shortest paths through a particular node. The betweenness of the node i is

$$b_i = \sum_{h \neq j \neq i} \frac{\sigma_{hj}(i)}{\sigma_{hj}}, \tag{5.2}$$

where σ_{hj} is the total number of the shortest paths from source h to destination j and $\sigma_{hj}(i)$ is the number of these shortest paths passing through the node i. For the network traffic model presented above with the R generated packets at random, the average number of packets passing through a single node i at each time step is $Rb_i/(N-1)$. If the capacity of the node i is denoted as a c_i, then the condition for not congested system becomes

$$R\frac{b_i}{N-1} \leq c_i \quad \forall i \tag{5.3}$$

In the case of the uniform capacity $c_i = 1$, the congestion threshold becomes [116]

$$R_c = \frac{N-1}{max_i b_i} = \frac{N-1}{b^*}, \tag{5.4}$$

where b^* is the largest value of the betweenness centrality in the network, which represent the node with the largest load. The betweenness centrality is

[1]Such simplified approximation is used mostly for the network analysis. Since the congestion of the nodes depends on topology, in the case of very heterogeneous networks, the load on nodes could diverge with the network size as N^γ with $\gamma > 1$ [118]. There is a variety of different routing strategies which utilize the adaptive (traffic-aware) routing polices, which can increase the overall network capacity. Pure shortest path routing leads to rapid network congestion, while the pure random walk strategies are inefficient. Therefore, the optimal routing strategy is one which follows the shortest path, but at the same time avoids the hubs. In [14, p. 251] some *hub avoidance strategies* are explained.

also used in the following cascading models and numerical simulation on the example of the European NREN network.

In order to understand the cascading robustness of the network, two models are used. The first one is the Motter-Lai model [75] of cascading failures which models the overloaded nodes as non-functional. Motter and Lai in their paper show that the attack on a single important node (one with the high initial load) may trigger a cascading effect which could lead to a collapse of an entire network and subsequently, a severe service outage. The second one is the Crucitti-Latora-Marchiori model [27] based on the dynamical redistribution of the loads of the network. Using this model, the authors were able to show that a breakdown of a single node can cause the collapse of the whole network as a result of the flow redistribution.

5.1 Motter-Lai Model

The Motter-Lai model [75] presumes that for a given network at each time step one unit of the relevant quantity (e.g. packet) is exchanged between every pair of nodes. The routes of the packets are chosen following the *shortest path* principle. The initial load of a single node in a network is therefore calculated as the number of shortest paths passing through it, that is the betweenness centrality of the node (5.2). The capacity of the node is assumed to be limited and proportional to the initial load $L_i \geq 0$, thus the capacity C_i of the node i is: $C_i = \alpha L_i, \quad i = 1, 2, \ldots, N$. The parameter $\alpha \geq 1$ is again the *tolerance* parameter and the N is the initial number of nodes.

At the time $t = 0$ all nodes are functional and are within the load limit. The removal of a node causes a change in the topology and the distribution of the shortest paths. It eventually causes the load in certain nodes to become higher. The moment the load of a node exceeds its capacity, it is considered as non-functional. Consequently, the loads get distributed again and other nodes get overloaded and therefore removed from the network.

The damage caused by the cascade is measured in terms of the relative size G of the largest connected component of the network, sometimes referred to as a *giant component*:

$$G = N'/N, \tag{5.5}$$

where N' and N are the number of the nodes in the giant component after and before the breakdown respectively.

As expected, the strongest impact on the network is caused with the removal of the node with the highest betweenness centrality. Here, we compare the relative size of the giant component in the network after the breakdown of three nodes with different betweenness centralities (Figure 5.1). After the removal of the most central node, the size of the remaining connected component is only around 50% the size of the original network. We can also see that the removal of the node from the edge of the network doesn't cause significant breakdown.

In light of such general results, one should not jump to conclusion that the most important node in the network is inevitably the one with the largest centrality value. The unpredictable complex dynamics in the networks prevents us to make general assumptions for every network. Although there is a strong correlation between node's centrality and the effect of its failure, it is not necessary that the critical node is one with the largest initial load. The very example of this anomaly is the European NRENs network, illustrated in the Figure 5.2.

5.2 The Most Critical Nodes

Here, the most critical node is the one whose removal would cause the biggest damage to the network, according to the Motter-Lai model of cascading failures. The damage is quantified as the reciprocal value of the largest connected

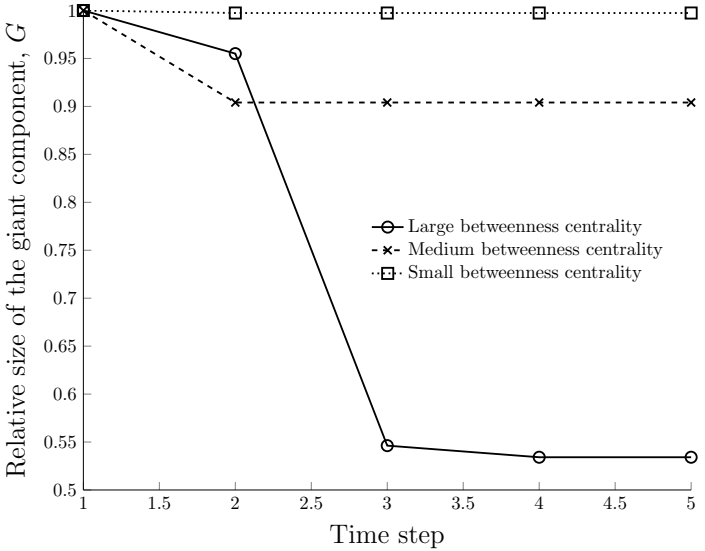


FIGURE 5.1: **Removing nodes with various betweenness centralities.** The removal of nodes with various betweenness centrality causes a different effect on the network. The considerable cascade breakdown is caused by the removal of the node with the highest betweenness (highest initial load). The tolerance parameter α is set to 1.5

component remaining after the simulation of the cascade. After the removal of the node i, the relative size of the remaining largest connected component is G_i and likewise after the removal of j, the relative size of the largest connected component is G_j. If $G_i < G_j$ we conclude that node i is more critical. In the case of the European NREN, the assessment is made based on the numerical simulation of individual and multiple failures. The simulation process is presented in the form of the pseudo code in Algorithm 3.

The $Q(V, E)$ is a graph consisting of set of vertices V and set of edges E. The load and capacity of a node are denoted as L_i and C_i respectively. The remaining size of the giant component after the cascade for the set of removed nodes I is denoted as G_I. The tolerance parameter is α.

Algorithm 3 The Motter-Lai model simulation

Input: $Q(V, E)$, α, I ▷ I - the list of removed nodes
calculate the size of the largest connected component N
calculate the load of each node L_i
calculate the capacity of each node C_i
remove the node(s) ▷ initialize the cascade
while $G_I(t+1) = G_I(t)$ **do** ▷ the new stable efficiency reached
    calculate loads L_i
    if $L_i > C_i$ **then** ▷ the load exceeds the capacity
        $a_{i,[1\ldots N]} = 0$ and $a_{[1\ldots N],i} = 0$ ▷ the node fails, breaking all associated edges
    calculate the size of the giant connected component G_I
return *the relative size of the giant connected component*

5.2.1 Individual Failure

The results of the node failure simulation on European NRENs using the Motter-Lai model show the most critical node is one with the second largest initial load followed by the one with the seventh largest initial load. Supposedly the most critical node (based on its centrality measure) is on the third place of the most important nodes. This "anomaly" is shown in the Figure 5.2, where the node with the smaller initial load causes the heavier fragmentation when removed.

One of the reasons for such an unexpected result is, as mentioned before, the very nature of complex systems where the initial perturbation could cause non-obvious consequences. Another reason is the actual measure used to assess the impact of the node removal. The measure used is the relative size of the giant connected component $G = N'/N$. Although the relative size of the giant component portrays the overall condition of the network, it can be sometimes misleading as it focuses solely on the biggest component, ignoring the rest of the network. It could happen that a network gets fragmented into one big component surrounded by many small chunks. On the other hand, the network could be fragmented in the way such that all small chunks are

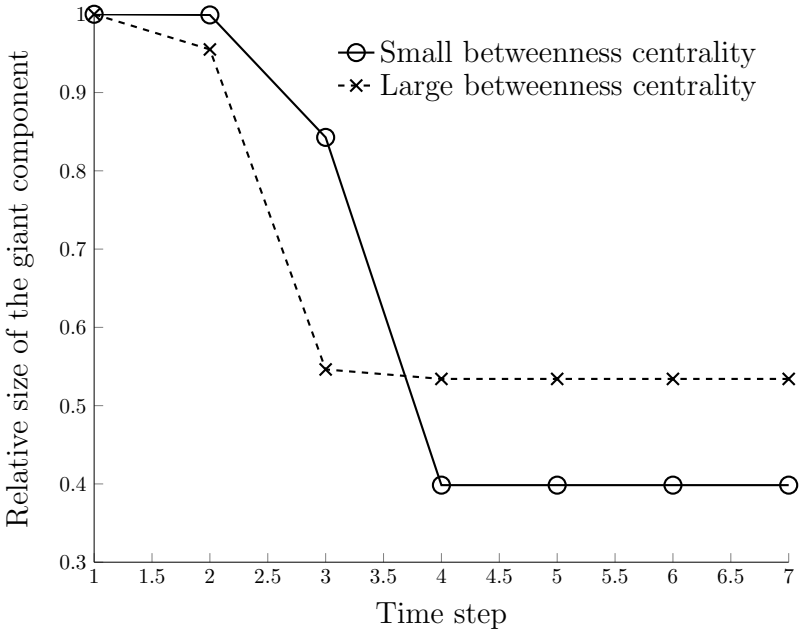


FIGURE 5.2: **Cascade anomaly in the European NRENs case.** The failure of a node with smaller initial load (smaller betweenness centrality) can cause a bigger damage (heavier fragmentation) of the network. The simulation is conducted for $\alpha = 1.01$

connected between each other. The properties of those two resulting networks could be different even if the sizes of the largest components would remain the same.

node ID	$G_{\alpha=1.01}$	$G_{\alpha=1.10}$	$G_{\alpha=1.30}$	$G_{\alpha=1.50}$	*lat.*	*lon.*	*location*
409	0.307	<u>0.242</u>	<u>0.300</u>	0.431	50.08	14.42	Prague, *CZ*
408	0.296	0.351	0.511	0.534	50.11	8.68	Frankfurt, *DE*
414	0.396	0.440	0.410	<u>0.398</u>	48.20	16.37	Vienna, *AT*
181	0.471	0.482	0.903	0.912	50.08	14.42	Prague, *CZ*
413	0.522	0.490	0.697	0.697	47.49	19.03	Budapest, *HU*
15	0.592	0.602	0.909	0.980	48.20	16.37	Vienna, *AU*
16	0.592	0.602	0.909	0.980	48.20	16.37	Vienna, *AU*
404	0.519	0.641	0.676	0.695	55.67	12.56	Copenhagen, *DK*
415	0.848	0.848	0.848	0.848	42.69	23.32	Sofia, *BG*
405	<u>0.267</u>	0.652	0.696	0.700	52.41	16.96	Poznan, *PL*

TABLE 5.1: **Impact of individual nodes removal.** The impact of the removal of individual nodes quantified by size of the largest connected component G after the simulation of the cascade failure using the Motter-Lai model. The simulation is conducted for various values of α. The underlined value is the lowest value of G in the column indicates the biggest damage to the network.

Furthermore, the measure of the largest connected component does not necessarily corresponds to the number of the overloaded (disabled) nodes during the cascade. Sometimes, the small number of failed nodes can cause the drastic fragmentation of the network. That is the case with the European NRENs model. The nodes with higher betweenness centrality cause the failure of other nodes which severely fragment the network.

The Motter-Lai model of cascading failures uses the tolerance parameter $\alpha \geq 1$, which quantifies the extra capacity of the nodes relative to the initial load. The cascading dynamics depends largely on the chosen α value. Theoretical assumption is that the node's capacity is relative to its initial load, although in the case of the real world networks those values could vary significantly.

In the case of the European NRENs, the results of the cascade simulation using the Motter-Lai model show that the most critical node depends on the chosen α. For different α, different nodes emerge as the most critical. The Table 5.1 shows the selected nodes and the respective impact on the network after the node removal. The measure of the damage is the relative size of the

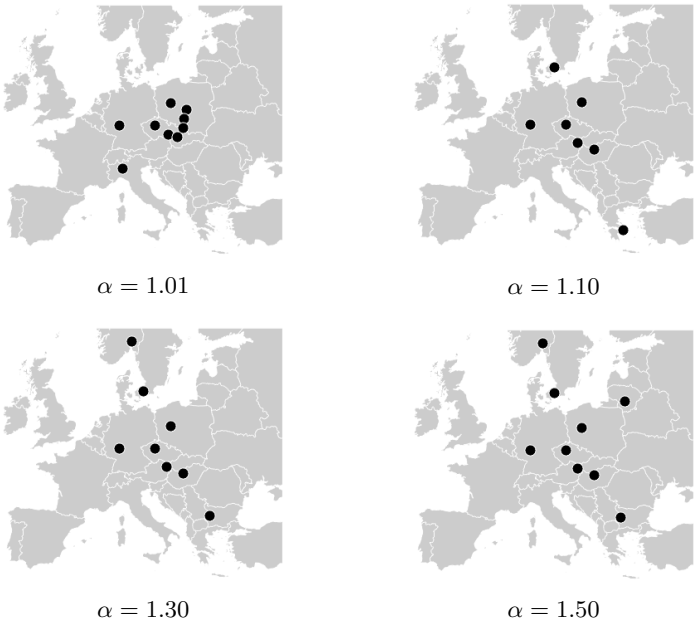


FIGURE 5.3: **The most critical nodes for various α.** The location of the most critical nodes whose *individual removal* would cause the biggest damage, based on the cascading simulation using the Motter-Lai model for various values of tolerance parameter α. The sets of critical nodes are different for various α.

largest connected component G. The smaller the G the bigger the damage caused by the node removal.

The analysis is not straightforward as the importance of the nodes varies with the change of α. The explanation for such a phenomenon lies in the distribution of excess load, which is discrete. In the case of a failure of node i, a node j is endangered as the load could become higher than its capacity. Let us define an excess load on node j as L'_j and the capacity as C_j. The node j remains functional until $L'_j > C_j$. All the values of L'_j which are lower or equal to the capacity are irrelevant, as the node can sustain it. For different tolerance parameter α the capacities of all the nodes are changed. The

excess load, however, remains the same. The change in tolerance parameter affects the capacity distribution. Then, an excess load has to reach a different threshold, so the node fails. This leads to sometimes unexpected dynamics for a slight change in tolerance parameter α. However, we can identify some of the most important nodes which might cause the biggest impact, like the node 408 in Frankfurt (Germany), followed by the node 409 in Prague (Czech Republic) and 414 in Vienna (Austria).

In the case of the European NRENs, the results of the cascade simulation using the Motter-Lai model show that the most critical node varies based on the tolerance parameter. The geographical location of the most critical nodes is shown in Figure 5.3.

In previous example, the α is uniformly distributed. Additional capacity depends on α and the initial load, thus, more central nodes will be allocated more excess capacity. However, such a way of capacity distribution may not be optimal. Wang at al. propose a solution to efficiently distribute the excess capacity among the nodes in a way that the network damage is minimized against any attack strategy. They show that there is an optimal distribution of the defense resource so the network is best protected from cost-based attacks. Cost-based attack assumes that the attacker also has cost, by which he optimizes his attack. If the defense resources of a network are not properly distributed, the attacker could benefit from choosing between the attack strategies [119]. For a particular network, the distribution of the excess load could differ, and the calculation should be performed for each network separately. Implementing the additional protection is usually a cost-benefit multivariate problem. A possible protective strategy for the Europen NREN is discussed in Section 5.5.

5.2.2 Multiple Simultaneous Failures

Random multiple failures are less common than a single failure. However, the multiple simultaneous failures could be also caused by a malicious attack.

The potential damage generated by the removal of two or more nodes could be devastating. The attacker could have a sufficient knowledge on network topology and then perform an optimal attack. If the attacker can chose n nodes to remove, the optimal attack is the one which would cause the biggest damage.

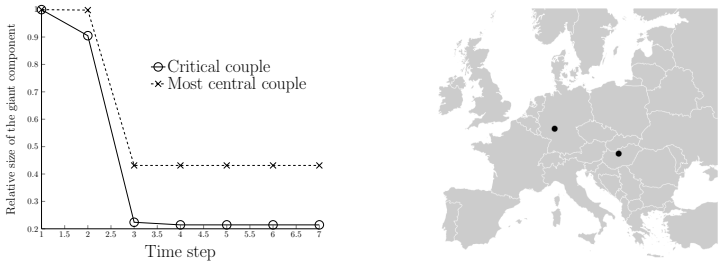


FIGURE 5.4: **Multiple failures - two nodes removed.** (a) Two different pairs of nodes were removed from the network and cascade is simulated. The plot shows the size of the giant component over time. Removing two nodes with the biggest centrality does not affect the network's efficiency as much as the removal of another two critical nodes. If two nodes whose independent removal cause biggest impact (*the most central couple*) are removed simultaneously, such action still does not break the network as a removal of two other nodes (*critical couple*). (b) Geographical location of two critical nodes whose simultaneous removal would significantly damage the network

In addition to single failure analysis, multiple failures are simulated. The sum of damages caused by the removal of two individual nodes at a time is not the same as the damage caused by the simultaneous nodes removal. Therefore, the conclusions on the most critical n nodes can not be drawn solely on the analysis of the n individual nodes failures. First, the simulation of a double node failure is conducted. The results show that the biggest damage is caused by the removal of the node with highest betweenness centrality together with the node with the third largest betweenness. The nodes are located in Frankfurt (Germany) and Budapest (Hungary)(Figure 5.4b). Simultaneous removal of those nodes cause significant fragmentation of the network (Figure 5.4a).

5.2.2.1 Computational Complexity and Solution Space

The Motter-Lai model simulation is computationally inexpensive. The most computationally demanding step within the ML algorithm (Algorithm 3) is to determine the load L of the node in each time step. It relies on the computation of the betweenness centrality, whose runtime is $O(VE)$ and $O(VE + V(V + E)\log(V))$ for unweighted and weighted graphs respectively, where V is the number of vertices and E is the number of edges [120].

However, in order to evaluate the impact of the removal of n nodes, the solution space becomes extremely large even for the small values of n. Even though cascading simulation would take a relatively short time, multiplying it with a number of elements in the feasible set would produce a practically unsolvable problem. The size of the feasible set could be calculated as a *combination* with n distinct elements in the set of size N:

$$F = \frac{N!}{n!(N-n)!}, \quad n \leq N \tag{5.6}$$

where F is a size of the feasible set, n is the number of nodes to chose from the larger set $N = 1157$. For example, solving the problem of finding the most critical couple of nodes, $n = 2$, means examining the set of $F = 668746$ possible solutions. It requires repeating the independent cascading simulations for exactly F times, and then comparing the results.

n	2	3	4	5	6	7	8	9	10
F	668746	2.57×10^8	7.42×10^{10}	1.71×10^{13}	3.28×10^{15}	5.4×10^{17}	7.77×10^{19}	9.92×10^{21}	1.13×10^{24}

TABLE 5.2: The solution space for examining n simultaneous node failures in a network consisting of $N = 1157$ nodes

To estimate the most critical set of 10 nodes whose failure would cause the biggest damage, one has to search through the solution space as large as

$F = 1.13 \times 10^{24}$, which is an incredibly large number[2]. The sizes of the feasible sets for various n are shown in Table 5.2.

5.2.2.2 Finding the Set of n Most Critical Nodes

Tackling the computational problems of this size requires another approach besides the brute force check. First, we should try to reduce the solution space. In the case of fixed n and fixed constraints, it could be done by decreasing N. By measuring the impact of individual node removal, we can intuitively exclude a certain subset of nodes whose impact is negligible. The values distribution of nodes importance is long tailed and follows the power law phenomena observed in many networks [44]. Based on the power law principle, we can exclude a large portion of observed nodes, because their individual removal has almost no impact. However, there is a set of nodes whose removal has big or moderate impact, and all of them should be included in the analysis. Thus, for the initial brute force search we chose only 25 most important nodes. In this case, for $n < 10$, it is feasible to check all possible solutions. For $n \geq 10$ and accordingly large N, brute force approach becomes again computationally expensive.

The following analysis shows that there is usually not one but many sets of n nodes whose simultaneous removal would cause the *same* or very *similar* damage. It means there are multiple solutions for a single value of n. However, the full insight of all possible solutions could be obtained only by using the brute force approach. In Figure 5.5 we can see the sets of removed nodes and the impact their removal makes. The damage is depicted in the form of a bar which represents the largest remaining connected component after the cascade. Notice that the removal of six and more nodes simultaneously does not make a big difference. The critical group may optimally have six nodes. The attacker could focus all of his resources to this very small set of nodes and

[2]It is a 113 followed by 22 zeros. The proper name for this number would be: one septillion one hundred thirty sextillion. It is comparable to the mass of the planet Earth in kilograms (5.98×10^{24}).

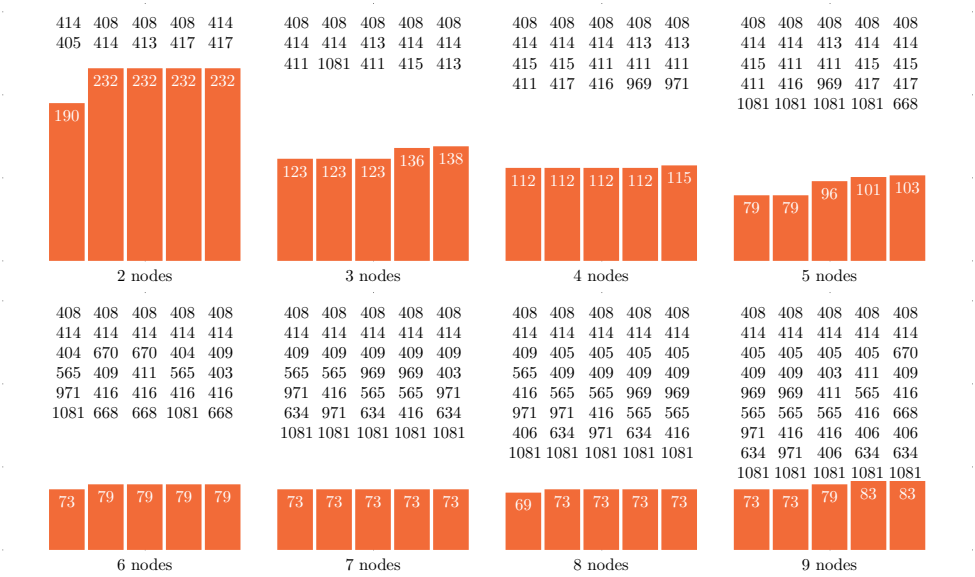


FIGURE 5.5: **The sets of n critical nodes.** Each bar represents the largest remaining connected component after the cascade in the case of group node removal. The higher the bar, the smaller the impact. Above the bars, there are IDs of removed nodes. Notice that the removal of six and more nodes makes no significant impact. However, removing some more nodes could even reduce the cascade, and that phenomenon is explained in the Section 5.4.

still cause a substantial damage. Notice that for the removal of a certain set of 9 nodes the damage is lower than with 8 nodes. It shows that removing some nodes could also decrease the damage. The reason is the actual decrease of traffic coupled with the removal of highly congested edges. This phenomenon when the cascade is reduced by the intentional node removal is explained more in details in Section 5.4. The locations of various sets of $n = 5$ most critical nodes are presented on the map in the Figure 5.6.

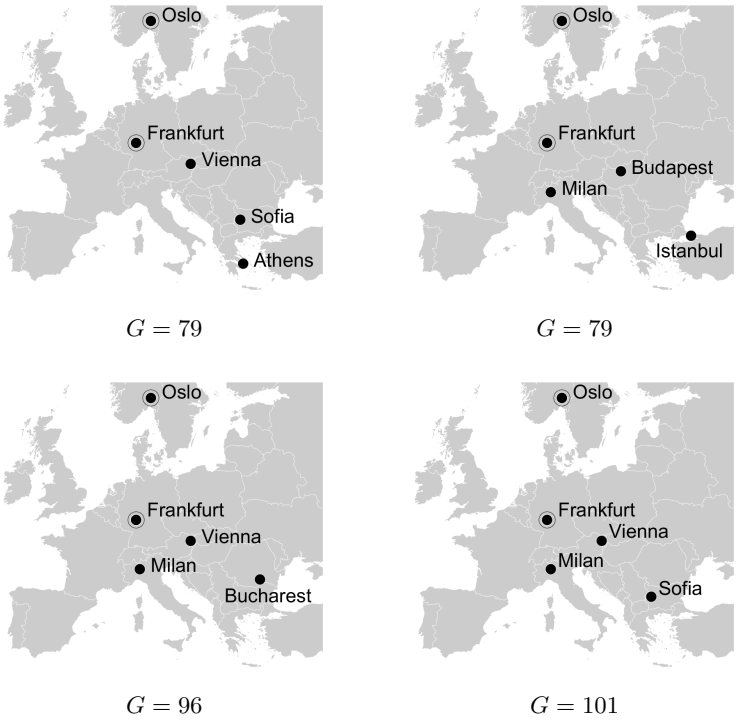
The brute force analysis is performed on a limited set of 25 most important nodes. To support the results, we can extend the solution space and perform

an additional analysis. However, with larger set of possible solutions, the analysis quickly becomes too complex. Hence, some of the heuristic methods has to be used. The method of choice is the Genetic Algorithm described in 3.2.3. The set of 100 most critical nodes is included in the additional evaluation. The original solution space is drastically reduced, but still of a substantial size ($F_{n=10} = 1.73 \times 10^{13}$).

The genetic algorithm tends to find a relatively good single solution through graduate improvement of fitness of the whole generation. Even it does not always get stuck in the local minimum, it pursues the best solution, often omitting other minima. The additional analysis with genetic algorithm is therefore used to confirm only a single critical group. The optimization performed was an integer problem, where a solution is an array of n integers ranging from 1 to 100, and each value is mapped to the appropriate node ID. The maximum of 100 nodes could be combined in groups consisting of n elements each. The set of 100 nodes has been identified by the previous analysis of the individual node impact. The Genetic Algorithm approach for finding the critical group is presented in form of a pseudo code in Algorithm 4.

Algorithm 4 Finding the most critical group using GA

1: **Input:** $G(V, E)$, α
2: Initialize parameters: *The population size is set to pop* $= 200$ *with a limit of maximum* $N_{gen} = 1200$ *generations*
3: Generate the initial population: *The initial population is created randomly with uniform distribution*
4: **while** The number of generations reaches maximum **do** ▷ N_{gen}
5: perform crossover
6: **while** for all solutions in population **do**
7: remove the nodes ▷ initialize the cascade
8: perform evaluation ▷ The evaluation is performed by the value of fitness function which is the size of the remaining largest connected component.
9: Sort the solutions from the final generation
10: **return** *the critical group of nodes*



$G = 79$ $G = 79$

$G = 96$ $G = 101$

FIGURE 5.6: **Various sets of five critical nodes.** Various sets of five critical nodes whose removal would cause the similar damage. Bellow every picture is size of the largest remaining connected component G after the cascade simulation. Smaller G implies bigger damage. The nodes marked with the additional circle are the intersection of all the sets and therefore are the most important nodes in the group of any five.

The GA approach produces the same or worse results for all the critical groups. The additional check with Genetic Algorithm supports a decision to focus the search on relatively small number of critical nodes. More detailed information on critical groups could be seen in Table 5.3.

n	no.	Node Location	G
1	1	Prague	280
1	2	Frankfurt	407
1	3	Vienna	510
1	4	Budapest	568
1	5	Copenhagen	742
2	1	Vienna, Poznan	190
2	2	Frankfurt, Vienna	232
2	3	Frankfurt, Budapest	232
2	4	Frankfurt, Athens	232
2	5	Vienna, Athens	232
3	1	Frankfurt, Vienna, Milan	123
3	2	Frankfurt, Vienna, Oslo	123
3	3	Frankfurt, Budapest, Milan	123
3	4	Frankfurt, Vienna, Sofia	136
3	5	Frankfurt, Vienna, Budapest	138
4	1	Frankfurt, Vienna, Sofia, Milan	112
4	2	Frankfurt, Vienna, Sofia, Athens	112
4	3	Frankfurt, Vienna, Milan, Bucharest	112
4	4	Frankfurt, Budapest, Milan, Istanbul	112
4	5	Frankfurt, Budapest, Milan, Ankara	115
5	1	Frankfurt, Vienna, Sofia, Milan, Oslo	79
5	2	Frankfurt, Vienna, Milan, Bucharest, Oslo	79
5	3	Frankfurt, Budapest, Milan, Istanbul, Oslo	96
5	4	Frankfurt, Vienna, Sofia, Athens, Oslo	101
5	5	Frankfurt, Vienna, Sofia, Athens, Stockholm,	103
6	1	Frankfurt, Vienna, Copenhagen, Riga, Ankara, Oslo	73
6	2	Frankfurt, Vienna, Copenhagen, Prague, Bucharest, Stockholm	79
6	3	Frankfurt, Vienna, Copenhagen, Milan, Bucharest, Stockholm	79
6	4	Frankfurt, Vienna, Copenhagen, Riga, Bucharest, Oslo	79
6	5	Frankfurt, Vienna, Prague, Kaunas, Bucharest, Stockholm	79
7	1	Frankfurt, Vienna, Prague, Riga, Ankara, Budapest, Oslo	73
7	2	Frankfurt, Vienna, Prague, Riga, Bucharest, Ankara, Oslo	73
7	3	Frankfurt, Vienna, Prague, Istanbul, Riga, Budapest, Oslo	73
7	4	Frankfurt, Vienna, Prague, Istanbul, Riga, Bucharest, Oslo	73
7	5	Frankfurt, Vienna, Prague, Kaunas, Ankara, Budapest, Oslo	73
8	1	Frankfurt, Vienna, Prague, Riga, Bucharest, Ankara, Amsterdam, Oslo	69
8	2	Frankfurt, Vienna, Poznan, Prague, Riga, Ankara, Budapest, Oslo	73
8	3	Frankfurt, Vienna, Poznan, Prague, Riga, Bucharest, Ankara, Oslo	73
8	4	Frankfurt, Vienna, Poznan, Prague, Istanbul, Riga, Budapest, Oslo	73
8	5	Frankfurt, Vienna, Poznan, Prague, Istanbul, Riga, Bucharest, Oslo	73
9	1	Frankfurt, Vienna, Poznan, Prague, Istanbul, Riga, Ankara, Budapest, Oslo	73
9	2	Frankfurt, Vienna, Poznan, Prague, Istanbul, Riga, Bucharest, Ankara, Oslo	73
9	3	Frankfurt, Vienna, Poznan, Kaunas, Milan, Riga, Bucharest, Amsterdam, Oslo	79
9	4	Frankfurt, Vienna, Poznan, Milan, Riga, Bucharest, Amsterdam, Budapest, Oslo	83
9	5	Frankfurt, Vienna, Copenhagen, Prague, Bucharest, Stockholm, Amsterdam, Budapest, Oslo	85
10	1	Frankfurt, Vienna, Copenhagen, Poznan, Riga, Bucharest, Stockholm, Ankara, Budapest, Oslo	85
10	2	Frankfurt, Vienna, Poznan, Prague, Riga, Bucharest, Stockholm, Amsterdam, Budapest, Oslo	85
10	3	Frankfurt, Vienna, Poznan, Prague, Istanbul, Riga, Bucharest, Ankara, Budapest, Oslo	85
10	4	Frankfurt, Vienna, Poznan, Prague, Istanbul, Riga, Bucharest, Ankara, Amsterdam, Oslo	85
10	5	Frankfurt, Vienna, Poznan, Prague, Kaunas, Riga, Bucharest, Amsterdam, Budapest, Oslo	85

TABLE 5.3: **Critical groups.** A detailed information on critical groups of various sizes ($1 \leq n \leq 10$). For each size, five critical groups are shown. The rightmost column represents the impact of removal, measured in the largest remaining connected component after the cascade. Notice that the removal of 6 and more nodes does not cause significantly bigger damage.

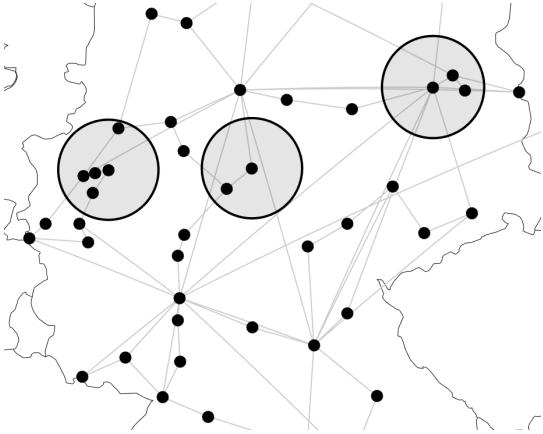


FIGURE 5.7: **Circular cuts on the map.** An example of circular areas around three nodes in Germany displayed on the map. All nodes and adjacent links which fall within the area are considered as failed. For the analysis, the simulation of a cascading process is performed independently for each cut. This example illustrates the circular cuts which cover: Berlin area, Ruhr metropolitan area and Lower Saxony.

5.2.3 Geographically Correlated Failures

There are numerous reasons that network failures could occur. It is a consequence of certain events such as cable cuts, hardware malfunctions, software errors, power outages, natural disasters (e.g., flood, fire, and earthquake), accidents, human errors (e.g., incorrect maintenance) and malicious physical/electronic attacks [10]. One approach for network survivability analysis is to focus on *isolated* independent failures, which could be single or multiple, but not correlated. Independent, uncorellated failures are modelled usually by a random process. However, communication networks are generally very robust against the single random failures, and multiple simultaneous failures are less likely to happen. Another approach is to model a targeted attack, which is directed towards single or multiple important nodes. As shown in Figure 5.6, the critical group of nodes could consist of geographically dispersed points, hence the physical reasons such as natural disaster, accident or power

outage could not cause their simultaneous failure. However, many physical risks could affect relatively large areas and could interrupt the operation of numerous nodes in the vicinity.

Geographically correlated failures analysis relies on the position of nodes and links on a geographical plane. The geographical location of the network elements is usually mapped to the cartesian coordinate system [10]. Then, various strategies could be used to damage the network and the disturbance could be assessed by many means. Most commonly researchers use cuts of various sizes and shapes to remove the links [121], including circular and other two-dimensional figures, such as ellipses and various polygons. Although circles could fairly approximate the affected geographical area, some failures demand other shapes in order to identify a *critical region* [122]. The resulting cut damages the network by altering its topology. The damage assessment is usually performed using several measures including [10]: weighted spectral distribution (WSD), algebraic connectivity (AC) and network criticality (NC). Neumayer et al. use additional measures to study the impact of geographically correlated cuts such as: total expected capacity of the intersected links (TEC), average two terminal reliability of the network (ATTR), maximum flow (MFST) and the average value of maximum flow between all pairs of nodes (AMF) [121].

Here, the shape of a spatial disturbance is modeled as a circle. For a node i, the circular area with radius r is marked. Then, all nodes within the radius are considered as failed. The Figure 5.7 illustrates an example of chosen areas around three nodes as centers The cascading failure is simulated for all nodes and circular areas around them for various r, where $r = [5, 10, 20 \dots 100]$ measured in kilometers Note that links which seemingly fall within the radius remain intact in this simulation and only nodes are considered as failed. The reason for that is the unknown actual geographical path of links. In all figures, links are represented as straight lines, but their real path could take any shape, usually following major infrastructures such as roads or power lines. The data

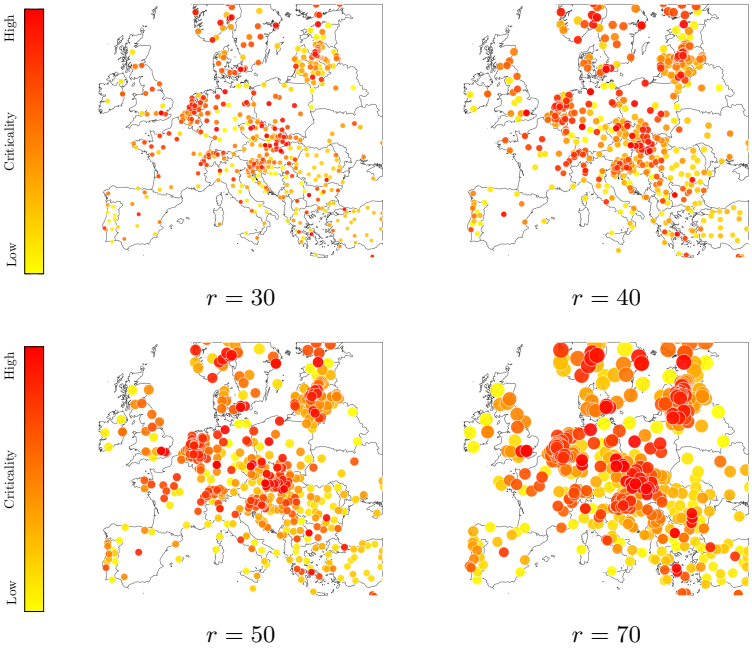


FIGURE 5.8: **Critical areas on the map.** The areas are defined as circular cuts. The network damage is simulated in a way that all nodes within the area are removed together with associated links. The *criticality* of the area is then measured as the size of largest connected component after the cascade. The most critical areas are around the most critical nodes with some exceptions.

about the actual paths is unknown and for that reason, only nodes within the radius and their adjacent links are removed.

This way we can identify the *critical areas* of various sizes. A measure of the damage is the size G of the largest connected component remaining after the cascade. This simulation is a variation of multiple simultaneous node failures, where the nodes are chosen by its geographical location. In Figure 5.8 the critical areas are plotted on the map. The most critical areas are concentrated around critical nodes. That is aligned with the innate characteristic of the

communication network as an example of a scale-free network. The power law is everpresent in almost all analyses regarding the node importance. The European NREN network is not an exception. There is a relatively small number of very important nodes and many nodes whose removal would cause negligible damage.

$r = 50$ $r = 70$

FIGURE 5.9: **Location of the most critical areas.** The most critical areas are located around the most critical nodes. Depending on the r of the circular cut, many areas could include the same important node, which can cause the group of geographically close areas to be identified as critical. The smaller the r the geographical distribution of critical areas becomes more dispersed.

The damage distribution of critical areas and critical nodes are similar but with slightly larger exponent value (Figure 5.10). It means that there are more important areas than important nodes. However, the relative number of important areas still remains small even for the large r. Depending on the radius of the circular cut, many areas could include a very important nodes, whose removal would cause a substantial damage. In Figure 5.9, the concentration of critical areas around the most important nodes are evident. With small r, critical areas tend to become more geographically dispersed and with even smaller r, the geographical distribution of critical areas becomes very similar to the distribution of individual nodes. The reason for such geographical distribution is the fact that for relatively large r, the individual critical nodes become covered with circular cuts which are centered in many

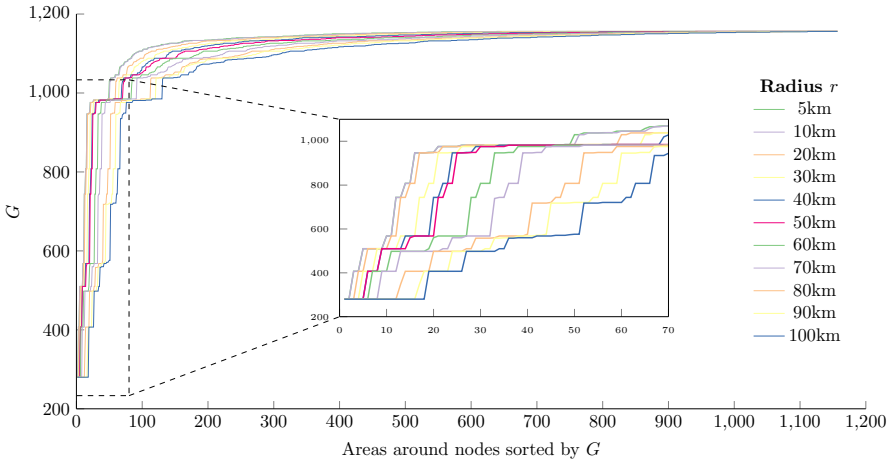


FIGURE 5.10: **Main frame:** The distribution of the extent of damages caused by the circular cuts of various radii. The damage is quantified as the size of the largest remaining connected component G after the cascade. The majority of cuts do not cause greater damage, but the small number of critically positioned cuts could cause devastating impact. **Inset:** An excerpt from the greater graph, showing the damage of the most critical areas. Even the areas with small radius could potentially cause big damage. Note that lower value of G represents bigger damage.

neighboring nodes. The node which is in the "epicenter" of the cut, might not be important but the cut affects a central node within the radius. The most critical node in the area adds the most to network damage. Therefore, to avoid a substantial damage in case of a geographically correlated failures, the most important nodes should not be concentrated in the narrow geographical area.

The damage caused by the circular cuts around nodes with various radii is plotted in Figure 5.10. For each radius, the areas are sorted by the damage the cut makes. The damage is quantified as the largest remaining connected component G after the cascade. This figure illustrates that there is a relatively small number of critical areas. The vast majority of areas do not cause the extensive damage, even for the large r.

An interesting phenomenon has been noticed. Sometimes, damaging the wider area around a single node could cause the smaller cascading effect than damaging narrower area. This phenomenon which appears to be some sort of a paradox could be used as a mean for active network protection against cascades. After the initial failure of the node or group of nodes within the geographical area, deliberately shutting down the nodes in their vicinity can stop the cascade or decrease the overall damage. More details about that and other means of active protection is provided in Section 5.4.

5.2.4 External Risks to European NRENs

The communication networks, as a part of critical infrastructure are exposed to numerous external physical factors that could interfere with their regular operation. For example, a disturbance in power network could also cause the failure in the interconnected communication network. However, the impact of small scale power outages could be mitigated by uninterruptible power supply (UPS) or small independent generators. Generally the frequency of external impact occurrence in a certain area is unaffected by the network design. After all, it is of a great importance to recognize and understand the possible external risks and quantify them. The proper awareness of the external factors and their impact could help network designers to reinforce the topology or to implement measures to mitigate the possible negative impact.

One of the most devastating events which could cause the disruption of various infrastructures and therefore in critical communication infrastructure are the seismic hazards. All around the world many small earthquakes happen frequently. Still the communication infrastructure usually remains functional and without major failures. By observing the data from the past earthquakes, communication network components perform relatively well under seismic conditions. Nonetheless, failures of some components are still found following earthquake events. Common failures found in telecommunication network components are failures of electronic equipment, such as computers,

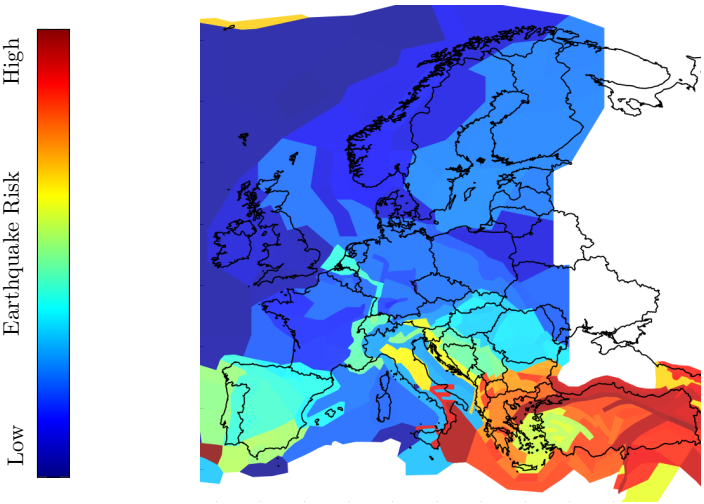


FIGURE 5.11: **Map of highest maximum moment magnitude (M_{max}) used in the SHARE model.** Here, 423 zones with estimated yearly number of seismic events with magnitudes $M_W \geq 4.5$ are plotted. Each area is a polygon with slightly different colour representing magnitudes in the range $6.0 \geq M_{max} \leq 8.8$. The value M_{max} is chosen as the maximum value of four maximum magnitude values estimated early.

server cabinets, switch boards, circuit boards, and battery racks [123]. Besides the direct physical damage to network components, the network failure could be caused by the lack of power supply, since the power lines could be cut. On top of that, the congestion usually occurs as the communication demand during the natural disasters reaches its peak. Congestion caused by the surge in call attempts to verify people's safety or to let people know of one's own safety, could result in usage restrictions being applied from 80% to 90% of telephone calls. No communication system is designed to serve such a large traffic demand.

Larger disasters do not occur very often, but the impact on the communication infrastructure could be devastating. For example the earthquake in Japan in

2011 severely impacted the communications infrastructure. The mobile and landlines were almost completely out of order. Even submarine cable landing stations were damaged and fibre optic cables were cut, severely affecting international services [124].

The European continent is less prone to devastating power of earthquakes. However, certain areas are of a high risk and some of the devastating seismic hazards happen in the last two decades. The most prominent are the events in Turkey[3] in 1999 when two major earthquakes severely damaged the existing communication network. Both fixed lines and GSM communications were not operational. One of the major fibre-optic cables had multiple ruptures and one of the major switches was destroyed completely [124].

Here, a possible approach for seismic risk analysis of the European NREN is presented. First, the risk of each node is assessed based on its location considering the seismic risk of a particular area. Then, the resulting measure is compared to the possible damage caused by the node removal.

The measures of seismic risk are taken from the publicly available database developed within the Seismic Hazard Harmonization in Europe (SHARE) project [125]. For SHARE, an Area Source Model has been constructed, stretching from the Mid-Atlantic Ridge and Iceland in the west, to Romania and Turkey in the East, from Norway in the North to the southernmost Islands of Italy and Greece. The whole SHARE area is covered with 423 zones. The maximum moment magnitude value M_{max} for $M_W \geq 4.5$, where the M_W is the measure of the earthquake intensity on the moment magnitude scale.[4] All areal zones are displayed in the Figure 5.11. For each area, the various parameters are calculated by processing the subset of events (from regional,

[3]The Kocaeli-Golcuk Earthquake of August 17, 1999 and The Duzce-Kaynasli Earthquake of November 12, 1999

[4]The moment magnitude (M_W) scale is based on the concept of seismic moment. It is uniformly applicable to all sizes of earthquakes. Seismic moment is calculated from the amplitude spectra of seismic waves which is a curve showing amplitude and phase as a function of frequency [126]. It is a logarithmic scale and the increase of one step corresponds to $10^{1.5}$ (about 32) times increase in the amount of energy released. Thus, an earthquake of M_W of 7.0 releases about 32 times as much energy as one of 6.0.

national or international catalogues) occurring within the polygon. Then, the certain statistical analysis is used to guarantee homogeneity within the areas and to increase the accuracy of the forecast. More details on datasets and the methods for their acquiring can be found on the project's website (http://www.share-eu.org).

The multidimensional risk analysis is performed on the European NREN. Multi dimensional analysis examines multiple properties of a system. The properties are presented in a form of n variables of various nature and sizes. They could be either weighted or non-weighted or normalized or non-normalized. Usually, the cumulative assessment is calculated as an Euclidean distance from the origin in n-dimensional space.

Here, only two dimensions are examined. The first variable is the node importance D_i measured as the potential damage in the case of the node removal. For each node, the cascading failure is simulated using the Motter-Lai model (Section 5.1) with tolerance parameter $\alpha = 1.1$. The resulting damage is quantified as $D_i = \frac{1}{G_i}$ which is the inverse size of the largest connected component remaining after the cascade. The variable D_i is normalized and its value ranges from 0 to 1, with the majority of nodes taking value between 0 and 0.2. This correlates with the results from the Section 5.2.1: only small number of failed nodes could cause the substantial damage. The second variable M_{Wi} is the earthquake risk quantified for each node i. There are 423 geographical polygons in total with different observed maximum moment magnitude values M_{W_j}. The value M_{W_j} is assigned to all the nodes falling within the geographical polygon j, such that if node i is within the boundaries of polygon j, then $M_{Wi} = M_{W_j}$. Only the most devastating earthquakes with $M_W \geq 4.5$ are observed. The majority of nodes fall within the areas with $7 \leq M_W \leq 8$.

The resulting positions of all nodes in two observed dimensions are plotted in Figure 5.12. The cumulative risk of a single node could be measured as an Euclidean distance from the origin. However, some thresholds should be introduced in order to filter out the nodes which are away from the origin but at the same time have very low value of one of the variables. The most critical

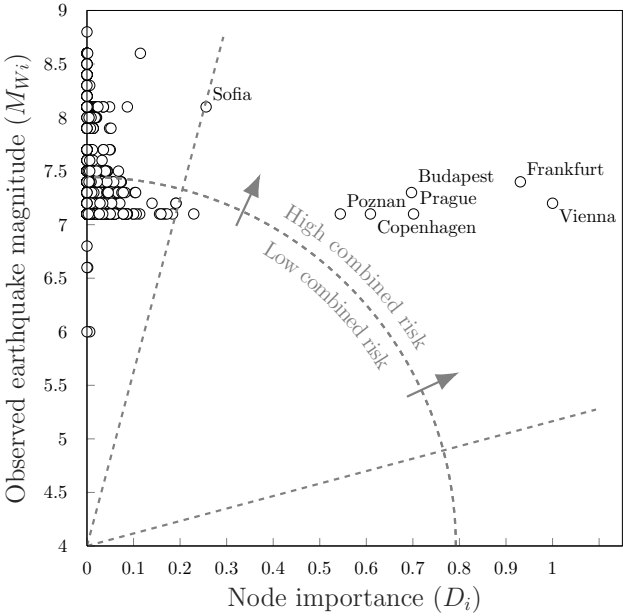


FIGURE 5.12: **Combined risk.** The nodes are plotted with respect to the combined risk in two dimensions. The risk of the seismic activity around the node's location is on the y axis. The x axis represents the potential damage to the network in the case of the node failure. The value on the x axis is the inverse of the size of the largest remaining connected component after the cascade caused by the node removal. The further the position of a node from the origin, the more critical it is. The arbitrary thresholds are marked as dotted lines. The area of high combined risk is in the upper right region.

nodes based on two dimensional risk analysis are positioned in the upper right region in the cartezian coordinate system.

Devastating natural disasters are usually not confined within the small limited area but rather spread across the wide territories. Therefore, the impact analysis should not be always restricted to a single point in space. In the case of severe natural disaster, not all nodes are necessarily directly physically affected. However, the supporting infrastructure such as power distribution

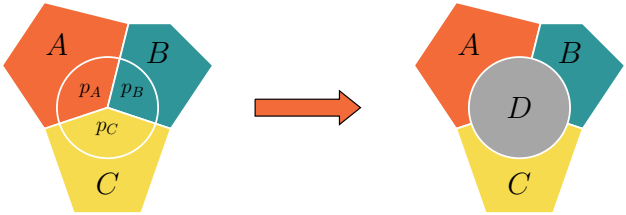


FIGURE 5.13: **Averaging the M_W.** The circular area often includes more than one polygons with different seismic characteristics such as M_W. The resulting $\overline{M_{W_D}}$ of the area D (on the right) is obtained as the weighted average of all the enclosed polygons M_{W_A}, M_{W_B} and M_{W_C}, considering the areas p_A, p_B and p_C respectfully. The weighted average is calculated using (5.7).

system will most likely be severely damaged and the service degradation would consequently cause the failure within the communication infrastructure. The following analysis takes into account the wider geographical area which can be affected by the earthquake. The area is again defined as a circular cut with the center in a single point and radius r.

Regarding the observed seismic activity of the circular zones, an additional adaptations are made. Depending on the size, each zone can cover multiple seismic polygons defined in the SEIFA model. Analysis considers the failure within the zone and for each zone individually. Therefore, the following averaging method was used to obtain the uniform M_W value for each zone:

$$\overline{M_{W_i}} = \frac{\sum_j M_{W_j} \times p_j}{\sum_j p_j}, \tag{5.7}$$

where M_{W_j} is the observed maximum moment magnitude for the polygon j, which is partially or fully included in the circular zone j. The p_j is the area of the intersection between the polygon and the circular area. The sum of areas of all intersections is equal to the area of the circular zone:

$$\sum_j p_j = r^2\pi \tag{5.8}$$

The process of averaging is displayed in the Figure 5.13. Left part of the figure shows the circular zone partially covering polygons A, B and C claiming the areas p_A, p_B and p_C respectively. On the right side, the circular zone D has an uniform averaged M_W, calculated using (5.7). In fact, the resulting M_W is a weighted average of all the M_{W_j} enclosed within the circular cut.

At the beginning of the simulation, all nodes within the circular zone are considered inactive. Depending on the position and the radius of the zone, cascades of various severity occur. The zone importance D_{Z_i} is quantified as the inverse of the size of the largest connected component remaining after the cascade. The variable D_{Z_i} is normalized and could take any value from 0 to 1.

A multi dimensional analysis is performed, similarly to the previous one with a single node failure. Here, the first variable is the damage of the possible failure of all nodes within the circular cut, denoted as D_{Z_i}. Its value is plotted on the x axis on the plots shown in the Figure 5.14. The analysis is performed for various radii. The second variable is the maximum moment magnitude for a circular zone i, denoted as $\overline{M_{W_i}}$ and obtained from (5.7). Here, it is called an *earthquake risk* and it is plotted on the y axis. Each point in the graph represents a single circular area defined by the two above mentioned variables.

The most critical areas are positioned in the seismically inactive regions of the central Europe with small frequency of destructive earthquakes. The number of circular zones are positioned high regarding the earthquake risk, but all of them have no important place in the European NREN topology. Those areas are mostly in Turkey, southern Italy and some Portuguese territories in the Atlantic ocean. None of them have crucial role in overall functioning of the European NREN. There is a lack of points in the graphs positioned in the right corner. Therefore, we can conclude that in general the European

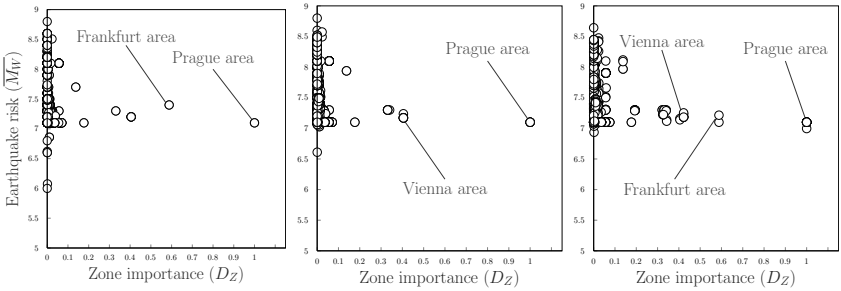


FIGURE 5.14: **Two-dimensional analysis of earthquake risk and damage within the circular zones.** The simulation of a cascade is performed for a number of circular zones of radius r, and the damage D_Z is plotted on the x axes. The *earthquake risk* quantified as the maximum moment magnitude and averaged for each zone $\overline{M_W}$ is plotted on the y axes. Each point represents a single circular area with radius r. The areas closer to the upper right corner of the plot are considered to be more critical. Those plots suggest that European NREN should not be severely affected by the earthquakes even for the large observed areas.

NREN is not highly affected by the earthquakes. Some seismic activities could separate large components, but that should not cause the network collapse.

A more detailed analysis of seismic hazard on interconnected critical infrastructures including power lines and communication network is out of the scope of this chapter. I will also not go into details of seismic risk mitigation methods and the efficient recovery strategies as well as details on requirements for public telecommunication networks in disaster relief.

5.3 Crucitti-Latora-Marchiori Model and Simulation Results

Crucitti, Latora and Marchiori proposed an alternative model for cascading failures [27]. A short analysis of the possible Europen NREN cascade failures is performed following the proposed model.

In the CLM model, each node is characterized by a given capacity for handling the traffic. Initially, the network is in the stationary state and the initial node capacity is larger than the traffic passing through it. After the node breakdown, the distribution of the loads commence and load gets distributed among other nodes in the network. If the traffic at the node after the distribution becomes higher than the capacity, the node is considered to be congested. The efficiency of the node is changed and subsequently the new distribution of loads in the network occurs. The main differences to Motter-Lai model are as follows:

- The congested nodes are not removed from the network. The efficiencies e_{ij} of adjacent links are changed, instead. Congested nodes are considered as less efficient in transferring data to other nodes.
- The damage in the network cause a change in the network's average *efficiency*. The average network efficiency $E(\boldsymbol{G})$ is the system service function, a measure used to evaluate the damage during the cascade.

The communication network is represented as a weighted undirected graph $\boldsymbol{G}$ with N nodes and K arcs (links). $\boldsymbol{G}$ is described by $N \times N$ adjacency matrix e_{ij}. The e_{ij} has the value in the range (0,1] if there is a link between i and j. Otherwise, the $e_{ij} = 0$. The value of e_{ij} is the measure of link's efficiency. The smaller the link efficiency is, the longer it takes to exchange an information unit between i and j. Initially, at the time $t = 0$, $e_{i,j} = 1, \quad \forall i, j$, meaning all the links have the same efficiency 1. The load redistribution

Algorithm 5 The Crucitti-Latora-Marchiori model simulation

```
1: Input: G(V, E), α, I                          ▷ I - the list of removed nodes
2: calculate efficiency E(G)
3: calculate the capacity of nodes C_i
4: remove the node(s)                            ▷ initialize the cascade
5: while E(G)(t + 1) = E(G)(t) do                ▷ the new stable efficiency reached
6:     calculate loads L_i
7:     if L_i > C_i then                         ▷ the load exceeds the capacity
8:         e_ij(t + 1) = e_ij(0)C_i/L_i(t)       ▷ efficiency of the links changes
9:     else
10:        e_ij(t + 1) = e_ij(0)
11:    calculate the new efficiency of the network
12: calculate the ratio between new and original efficiency
13: return the ratio between new and original efficiency
```

will cause the efficiency of the links to change. Since it is assumed that the communication between links takes the most efficient path, altering the links efficiency will usually cause further change of the most efficient path as well. The efficiency of the path is calculated as a harmonic mean of the efficiencies of the component links[5]. The efficiency of the most efficient path between i and j is denoted as ϵ_{ij}. Then, the average efficiency of the network is

$$E(\boldsymbol{G}) = \frac{1}{N(N-1)} \sum_{i \neq 1 \in \boldsymbol{G}} \epsilon_{ij}. \tag{5.9}$$

The average efficiency $E(\boldsymbol{G})$ is used as a measure of the network performance of a $\boldsymbol{G}$.

The load $L_i(t)$ of the node i at the time t equals the total number of shortest (most efficient) paths passing through i. Each node is characterized by the *capacity* C_i which is proportional to the initial load of the node:

[5]The harmonic mean of N numbers $x_1, x_2, \ldots, x_N$ is defined as $H = n \left(\sum_{i=1}^{N} (1/x_i) \right)^{-1}$

$$C_i = \alpha L_i(0), \quad i = 1, 2, \ldots, N, \tag{5.10}$$

where $\alpha \geq 1$ is the tolerance parameter. The initial network operates in the stable stationary state with an efficiency $E(0)$. After the breakdown of a fraction of nodes, the traffic gets rerouted through the network, altering the loads of the remaining nodes. After an initial rerouting, certain nodes get congested and the links leading outwards of the congested nodes become less efficient following the iterative rule:

$$e_{ij}(t+1) = \begin{cases} e_{ij}(0)\dfrac{C_i}{L_i(t)} & \text{if } L_i(t) > C_i \\ e_{ij}(0) & \text{if } L_i(t) \leq C_i, \end{cases} \tag{5.11}$$

where j extends to all first neighbors of i. This way, the nodes don't have to be removed from the network. The efficiency of the corresponding links deteriorate, and the node becomes less preferable choice for the most efficient route in the next iteration.

5.3.1 Single Node Failure

First, the simulation of the efficiency deterioration of the European NRENs is conducted after a single node failure. The simulation setup follows the Crucitti-Latora-Marchiori model. The pseudo code for the simulation process is shown in the Algorithm 5. For a single node failure, two strategies are chosen. The first one is *random failure strategy*, and the second one is the strategy of a *targeted attack*.

For the random failure strategy, the probability of the failure of the each node in the network is equal. Therefore, a single node is randomly chosen and removed from the network and the behavior of the network is observed over time. When the network reaches the next stable state (e.g. the efficiency

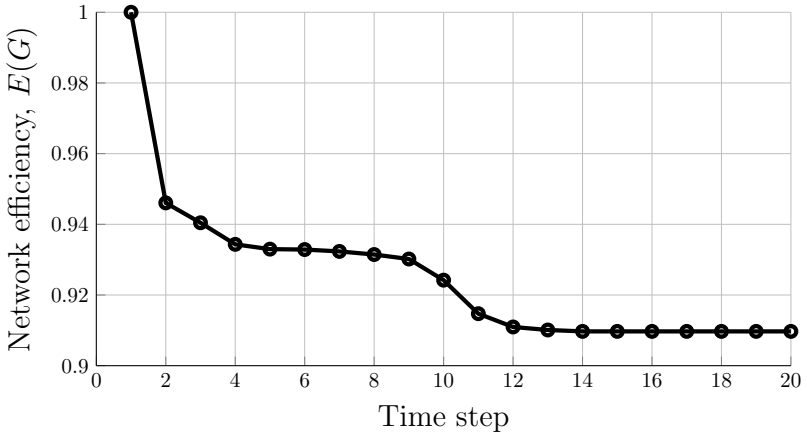


FIGURE 5.15: **Efficiency deterioration in the case of a single node failure.** The particular node located in London is removed. The tolerance parameter is chosen as $\alpha = 1.02$. This is a plot of the overall efficiency of the network over time. It exhibits a substantial deterioration. The NREN network is vulnerable to targeted attacks.

reaches the lowest stable value), the ratio of start and end efficiency is calculated. That ratio represents the impact of the removal of the certain node on network efficiency. The tolerance parameter $\alpha \geq 1$ represents the redundant capacity of all the nodes in the network. As described in (5.10), the additional capacity depends on the initial load $L_i(0)$ and tolerance parameter α. Therefore, the more the node is initially loaded, the more additional capacity it will get.

In Figure 5.15 the particular node located in London is removed. The tolerance parameter is set to $\alpha = 1.02$. The initial load of the node (betweenness centrality) was 86230 (0.129 normalized). After 20 time steps, the overall network's efficiency drops by 10%. Note that there could be more nodes within the same city and also at the same location. In this example, only one particular node in London is removed while other nodes in the same geographical area stay intact.

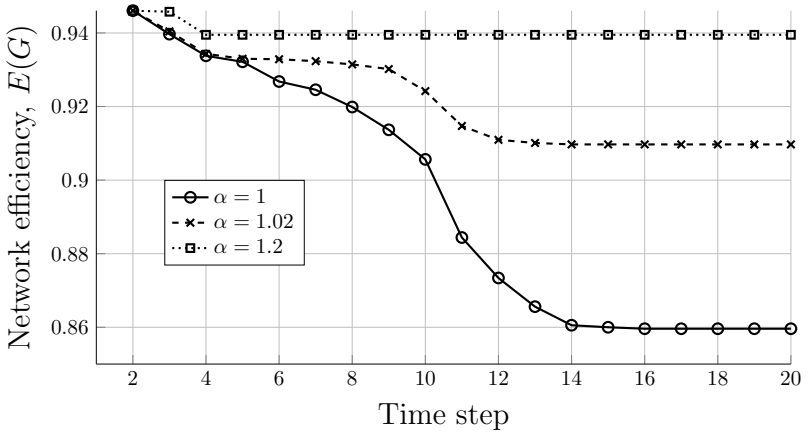


FIGURE 5.16: **Impact of the parameter α on network efficiency after a breakdown.** After the node removal, the overall network efficiency drops. Here, the impact of the parameter α on the network behavior after the breakdown is shown. If α is bigger and the redundant capacity of nodes higher, the impact of a single node failure is less significant. However, increased α implies higher costs.

For smaller α the network's efficiency gets even more affected. Let us consider the case when the overall system operates at its limits where no extra capacity is allowed ($\alpha = 1$). For the same example and same circumstances, the efficiency drops by almost 20%. The dependence of the network efficiency after the node removal in regard to the value of the parameter α is shown in Figure 5.16.

To measure the impact of the random node removal in this case we assess the average impact of removing each node independently. For the each node removal, the impact is measured as the efficiency deterioration over time. Then, all measures in all time steps are averaged over the total number of nodes. The resulting value is the network's *robustness* to the random node removal (Figure 5.17). The tolerance parameter α is set to 1.02. As expected, the random node removal in case of the networks with scale-free properties

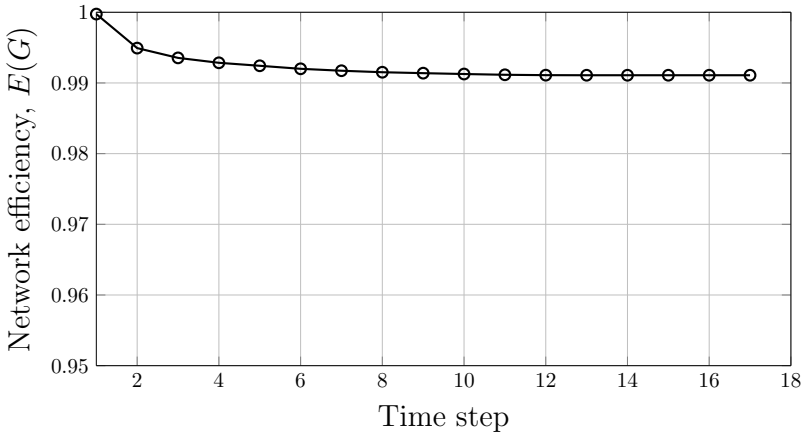


FIGURE 5.17: **Effect of the random failure on the network's efficiency.** The resulting efficiency is an average efficiency of a network after the simulation of failures of all nodes individually. This plot represents the most likely outcome after a random breakdown. Notice that the average efficiency after ten time steps reaches its minimum, which does not go below 0.99. The European NREN is highly resilient against random failures.

such as European NRENs network, does not cause a significant damage. This property is the main reason for the robustness of the Internet as well.

What is the significance of removing single random node out of the network? It is already shown in [27] and confirmed here with simulation that for the scale-free networks, random removal usually does not have a major impact. A degree distribution in scale-free networks follows a power law form $P(k) = k^{\gamma}$, which means that in such networks there is a small number of nodes with a very high degree and at the same time very large number of nodes with relatively small degree. The same applies for the betweenness centrality distribution. Only around 5% nodes have a normalized betweenness value between 0.1 and 1. The rest 95% are in the range between 0 and 0.1.

While homogeneous networks suffer approximately the same damage regardless of the attack strategy, heterogeneous networks appear resistant to random failures, but in contrary show extremely low resistance to directed attack on the highly connected nodes.

The network of European NRENs we observe has the most important properties of the scale-free network, and the most important is the degree distribution which clearly follows the scale-free pattern. This is by all means an expected property as the majority of computer networks follow the power law degree distribution which is caused by the *preferential attachment* growing principle [43]. The network of European NRENs is not an exception. It is relatively immune to random failures, but prone to the significant service outages in the case of the targeted attack. After the simulation of the targeted attack on three nodes with large, medium and small betweenness centrality, the undoubtedly highest impact has the removal of the most central node, as shown in Figure 5.18. At the same time, that is the node with the highest initial load.

The most intuitive topological measure for node importance is node degree, but focusing solely on a degree, we overlook potentially very important nodes who act as a links between large connected segments of a network. Those nodes usually have small degree, but are still very important. A detailed study on various attack strategies which deal with this problem is published by Holme et al. [61]. In many network models, like in Barabási Albert model, the betweenness centrality and degree are highly correlated. Although, in many real networks the situation is more complex as large fluctuations are observed and some nodes might have large betweenness centrality but small degree. The betweenness centrality actually covers the importance of both nodes with high degree and critical nodes with the small degree.

Hence, the most effective strategy for a single node attack would be targeting the nodes with the highest initial load (betweenness centrality). In the case of the network of the European NRENs, five most critical nodes are located in Frankfurt (Germany), Vienna (Austria), Budapest (Hungary), Copenhagen

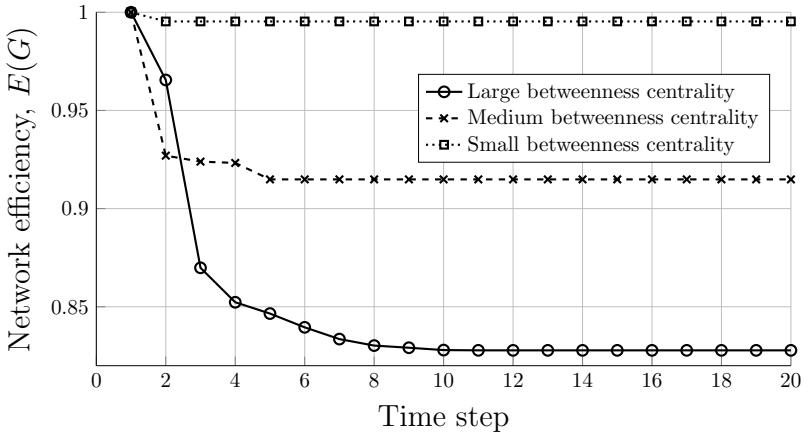


FIGURE 5.18: **The effect of the removal of three nodes with different loads.** The simulation results after removal of three nodes with high, medium and low initial load respectively with $\alpha = 1.02$. The impact on the overall network efficiency is the strongest after removal of node with the highest initial load.

(Denmark) and Poznan (Poland). Note that this analysis is conducted entirely on the information on the network's topology. Therefore, it neglects all details related to the technical aspects of the system elements. Real loads could vary, as they can be adjusted on different levels of network control. Furthermore, the excess capacity is usually not simply governed by a single parameter, but rather by the operational decisions of network designers. However, such approach can give an overview of the potential damage caused by the sudden change of topology.

5.3.2 Multiple Nodes Failure

The potential damage of a multiple failure within the complex networks is not easy to estimate in advance. A damage caused by the failure of multiple elements is not the same as the sum of damages caused by a single failure

(a) (b)

FIGURE 5.19: **The most critical nodes within the European NRENs.** (a) The individual failure of each of the five nodes on the map would significantly affect the network's efficiency. (b) Geographical location of two critical nodes whose simultaneous removal would significantly damage the network.

of each element. Another difficulty to predict a damage is the case of non-simultaneous failures. After the failure of the first node, the topology of the network changes and the failure of the following node could cause an unforeseen outcome.

In the case of the European NRENs, the two-node failure is simulated. The failures are considered to be simultaneous, and tolerance parameter is set to $\alpha = 1.02$. For the sake of the shortest computational time, only 30 nodes with highest betweenness centrality are assessed. In the simulation, all nodes are coupled with each other. The couples of nodes are removed from the network and the drop of the efficiency is observed over time. An expected result would be that the biggest damage is caused by removal of two nodes with highest betweenness centrality. However, the unpredictable property of the complex networks appears again. The nodes which would cause the biggest deterioration of network's efficiency when removed together are nodes with fourth and eighth biggest betweenness centrality. They constitute a critical pair. The removal of a critical pair cause the drop of the efficiency

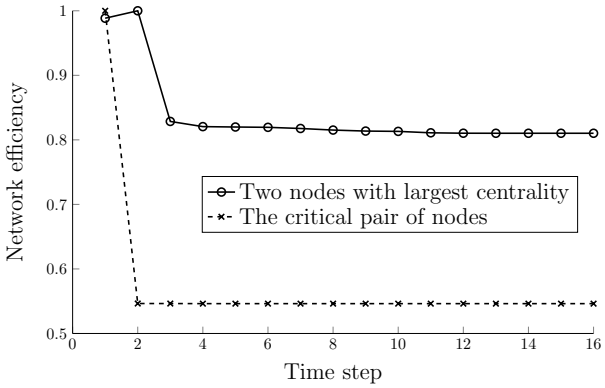


FIGURE 5.20: **Multiple failures - two nodes removed.** The comparison of the impact of removing of two pairs of nodes. The removal of two nodes with the biggest centrality doesn't affect the network's efficiency as the removal of another two critical nodes.

for 45%, while the removal of two nodes with highest betweenness makes the network 20% less efficient (Figure 5.20). The reason for such an unexpected result is the fraction of congested nodes after the failure. Following an initial failure, some nodes get congested, the efficiency of the adjacent links get worse and they stop being part of desired shortest paths. During the cascade, certain paths get restored and belonging nodes recover. After some time, the network enters again the stable state without further changes in efficiency. The resulting efficiency is, however lower than the initial one. The failure of a critical pair causes such an initial congestions which consequently affects the desired shortest paths in a way that many of them do not recover. The excess load gets distributed to many remaining nodes which get congested simultaneously, not allowing the efficient distribution of loads to occur. The geographical location of the critical couple of nodes is shown in Figure 5.19b.

5.4 Active (Costless) Protection Strategies

A costless protection strategy involves a set of actions on a network topology in order to mitigate the impact of a failure of the fraction of nodes. Those strategies are costless because they do not require any substantial investment in the network infrastructure such as additional links or increased capacity. However, they require implementation of some measures which could enable the network to protect itself using active means. An active change of topology in this case considers a deliberate removal of certain nodes in order to stop the cascade or to minimize the impact of the failure. Here, we investigate two strategies. The first one is based on an intentional removal of a fraction of nodes after the initial failure regardless of the actual position of the node. The second one considers the geographical location of failed and removed nodes and gives a possible solution for more localized active protection strategies.

5.4.1 Removing Nodes

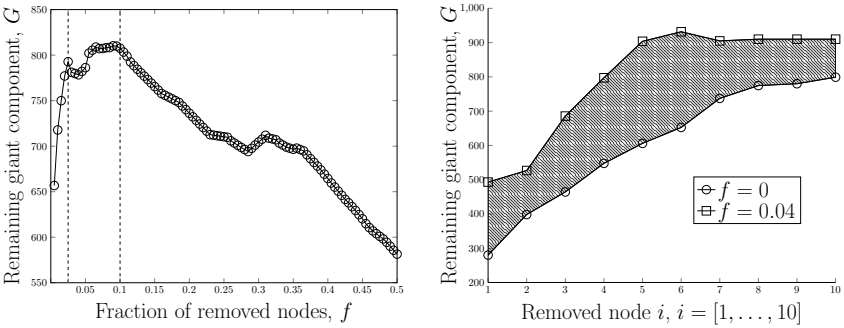
The Europen NREN is a complex network with heterogeneous distribution of loads and therefore it is susceptible to cascading failures. Motter [71] proposed a defense strategy based on a selective further removal of nodes and edges, right after the initial attack or failure. This strategy, with slight modification, is shown to be effective for protection of the particular European NREN network.

The cascade in the network is divided in two major stages: (1) the initial attack, where the fraction of nodes fails, and (2) propagation phase, where further failures happen due to the congestion. A propagation phase occurs in numerous time steps until all the loads of the nodes in the remaining network become smaller than their respective capacities. The size of the cascade is measured as the ratio $G = N'/N$, where N and N' are the sizes of the largest connected component before and after the cascade respectively. The defense mechanism takes place after the phase (1) but before the phase (2). It is

assumed that the only operation allowed after the attack and before the larger cascade is an *intentional removal* of nodes or edges. An intentional removal (IR) of carefully chosen nodes could reduce the cascade. The nodes chosen for the removal should have relatively *small load*. The rationale behind this assumption is that the nodes in the network equally contribute to the traffic load, but are not equally congested[6]. Therefore, a removal of the nodes with small load would decrease the overall traffic without the need for further load distribution.

Intentional removal of nodes with small load is not a straightforward process, as the removal itself contributes negatively to the size of the giant component. The IR should be carefully limited to the certain number of nodes so that cascade is suppressed and the remaining giant component stays relatively large. For the random scale-free network with a random attack on 0.1% of nodes, the optimal fraction of IR nodes is $f \approx 0.4$ [71]. However, for targeted attack on European NREN, simulation shows that the fraction of the removed nodes could be much smaller, $0.02 \leq f \leq 0.1$. In a Figure 5.21a size of the remaining giant component G is plotted as a function of the fraction f of removed nodes with small load. The nodes have been removed after the initial attack in the way that those with smaller load are removed first. The simulation is conducted for the set of ten most critical nodes in the NREN network, and the average G is plotted. The removal of only 2% of the least loaded nodes, is able to reduce the cascade drastically. From 2% to 10% the size of the giant component does not change considerably. However, after the 10% threshold, G starts to drop linearly. It means that further node removal does not add to the network protection but influences negatively to the size of the giant component. The analysis is not performed for the random failure nor averaged over all nodes in the network. The focus is on the set of critical nodes, those whose removal would cause the biggest damage. For the

[6]In ML and CLM model a traffic in the network is modelled in a way that all nodes communicate with every other node by exchanging a bit of information in a single time step. All the nodes supply the network with the same amount of data. On the other hand, a distribution of loads follows the long-tail distribution, where small fraction of nodes are highly congested and the vast majority is not.



(a) (b)

FIGURE 5.21: **Effect of the selective removal of the fraction of the least loaded nodes.** After the initial attack on a critical node, the protective measure is carried out. The fraction f of the least loaded nodes is removed in order to mitigate the cascade failure. The Figure (a) illustrates the change of the size of the largest remaining connected component G after the cascade as a function of f. The maximum G is reached for the $0.02 \leq f \leq 0.1$. For values of f greater of 0.1, the impact of the intentional removal begins to affect negatively to the resulting G. The resulting plot is averaged for ten most critical nodes within the European NREN. The second picture (b) shows the effect of the protective measures for the $f = 0.04$. For each removed critical node i, the protection mechanism keeps the remaining G higher. A difference between two corresponding points on two plots is a measure of the protection effectiveness. All the expected values of G for $0 < f < 0.04$ are the most likely to be found within the shaded area.

European NREN, in the case of a critical node failure, any fraction between 0.02 and 0.1 of the least loaded nodes could be removed in order to prevent further cascade.

The Figure 5.21b displays the comparison of the size of the giant component after the cascade with and without the protection measures for ten most critical nodes. The fraction of removed nodes after the initial attack is chosen to be $f = 0.04$. For each critical node i the cascade is mitigated so that resulting G is always greater if the protective measure is properly implemented.

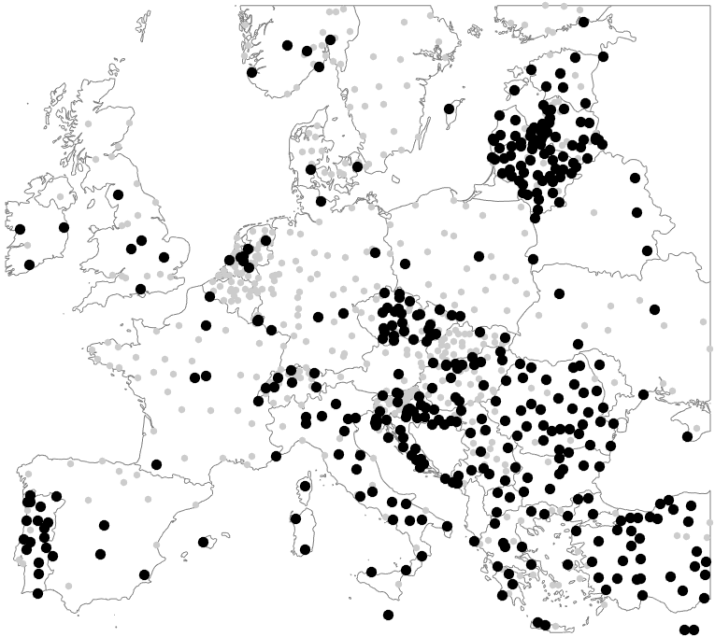


FIGURE 5.22: **Nodes candidates for intentional removal.** Those are the nodes which appear frequently in the lists of the least loaded nodes in the network. To mitigate the cascade, the fraction of removed nodes after the initial attack should be $0.02 \leq f \leq 0.1$. In absolute numbers, the number of intentionally removed nodes is $23 \leq n_{ir} \leq 115$. Any chosen n_{ir} from the set of candidate nodes displayed on the map, should be sufficient to mitigate the cascade in the case of the critical node failure.

The idea is to identify the set of nodes which should be prepared to be removed in the case of the most dangerous failures. The failure of one of the ten most critical nodes from the Table 5.1 would cause the biggest damage. Therefore, the following analysis is performed: For each of the most critical nodes, the failure is simulated and the least loaded nodes are chosen. Those are the nodes *candidates* for intentional removal after the initial attack. A certain number of nodes appears often in the list of candidates for various failed i. Those are

the nodes which are most likely going to have small load in the case of the intentional attack. The most common *candidate* nodes are displayed on the map in the Figure 5.22. The protective mechanism should remove a fraction f of all the nodes but from the subset of *candidate* nodes. In absolute numbers, the number of nodes chosen to be intentionally removed is $23 \leq n_{ir} \leq 115$. A decision maker is free to chose which nodes he will remove from the set of candidate nodes. It does not matter which particular node is removed until the number n_{ir} is within the limits.

5.4.2 Removing Nodes in the Vicinity of Failure: Protective zone

Geographically correlated failures within the European NRENs are analyzed in the Section 5.2.3. The analysis is performed by simulating the failures of the nodes within the geographical area. The areas are chosen to be of a circular shape with various radii r, measured in kilometers where $r \in [5, 10, 20, \ldots 100]$. That way, the most critical areas around nodes have been identified. The damage is quantified as the size of the largest connected component G after the cascade simulation. It is preferable that G remains large, therefore we also introduce the additional damage measure $D = \frac{1}{G}$, thus the larger the D the larger the damage.

Usually, the damage to the network correlates with the radius of the affected area. The larger the area around a node, the bigger damage is caused by the node removal. However, there are examples which do not follow this rule. For a certain critical area of a radius r it is possible to identify the wider critical area with radius $R > r$ whose failure would cause smaller damage. The band between the areas of radii r and R we call a *protective zone*. Deliberately disabling the nodes inside the protective zone could serve as a defensive mechanism against the cascade failure. This should not be associated with the immunization strategy where the nodes around the source of the infection are protected.

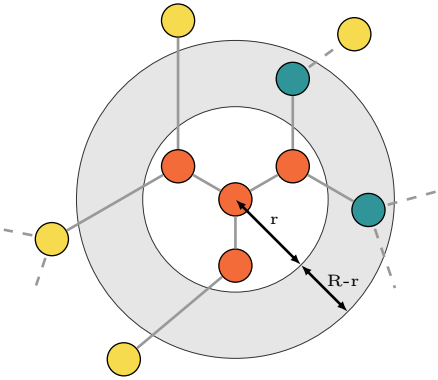


FIGURE 5.23: **Protective zone.** In the case of a node failure within the area of radius r, there is an additional number of nodes in their vicinity, whose simultaneous failure would reduce the overall damage to the network. Those additional nodes fall within the protective zone of radius R where $r < R$. Red nodes are initially failed and blue nodes fall in the protective zone. However, yellow nodes should stay intact, as disabling them could cause larger damage. The protective zone is grey. The damage is quantified by the size of the largest remaining connected component after the cascade.

Let us identify the first circular area A_1 with the radius r and center in x, y and let us quantify the damage the removal of all nodes within the area could cause as D_1. Likewise, let's identify the second area A_2 around the same centre with the radius R and the damage D_2. We say that A_1 and A_2 form a *protective zone* Z if $D_2 < D_1$. Protective zone is fully defined with center, smaller radius and larger radius as $Z(x, y, r, R)$. In the Figure 5.23 the protective zone around the initial critical area is colored gray.

The protective zone Z depends on the geographical location of nodes and the radius of the initial failure. Therefore, the existence and the size of Z have to be measured using numerical simulations for various radii. The resolution of the results depends on the discrete space between radii chosen for the simulation.

The simulation is performed as follows: for chosen r, the areas of radius r around all nodes are identified. There are N circular areas for each r. All the nodes which fall within the circular area A_i are removed from the network. The cascade is then simulated for each area A_i, $\quad i \in [1, \ldots, N]$, and the damage is calculated. Every cascade is simulated using the Motter-Lai model of cascading failures explained in Algorithm 3. The simulation is then repeated for all discrete values of r_j, $\quad j \in [10, 20, \ldots 100]$.

Protective zones around nodes are then identified by substituting the acquired values of $G_{i,j}$ for all consecutive r_j. If $G_{i,j} - G_{i,j-1} > 0$, there is a protective zone around node i defined as $Z(i, r_{j-1}, r_j)$. More informally, the damage caused by removing the nodes from a certain area is larger than removing nodes from the 10km wider area. The Figure 5.24 identifies some protective zones around nodes in European NREN.

The explanation of such cascade dynamic could be related to the findings of Motter in [71]. For a limited number of nodes the protective zone around them could be identified. The cascade reduction will commence if the nodes within the protective zone and adjacent edges meet conditions from [71]. The removed nodes should have relatively *small load*, and the belonging edges should have a *large excess of load*. If the nodes and edges within the protective zone meet those conditions in some extent, there is a chance that their removal would reduce the cascade.

In this case, the protection strategy is constrained by the geographical location of nodes. However, using the method proposed here, we can develop the localized protection strategy against cascading failure. The action protocol for identified sub networks could be implemented. In the case of failure, certain local sub networks should be disconnected from the rest of the network. Such an action would damage the big network, but it would also save it from a larger failure. Restoring the links should take place after the network reaches new stable state.

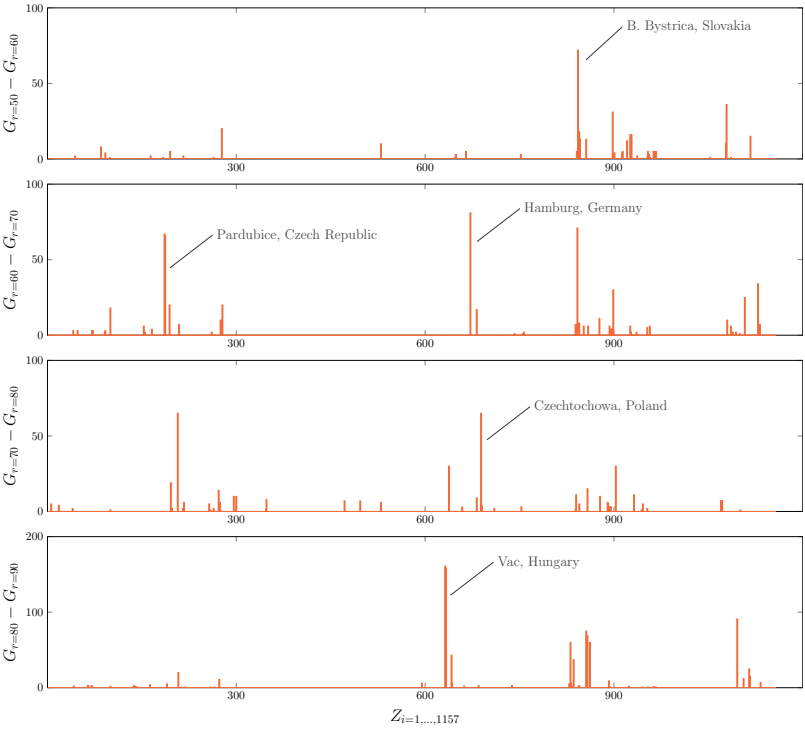


FIGURE 5.24: **Protective zones identified.** By simulating the cascade failure for circular areas of various radii r around nodes, the damages could be quantified as the size of the remaining largest component G. For larger r, the G is smaller, meaning the bigger damage. However, for some consecutive values of r, the damage becomes lower. Here, the values of two consecutive Gs are substituted, and all positive values plotted. If $G_{j+1} - G_j > 0$ for node i, there is a *protective zone* around the node i identified by the radii used to quantify G. The height of the bar represents the difference in potential cascade failure if the protective zone is "activated".

Each point on the x axis represents a single area around the node i.

5.5 Passive (Costly) Protection Strategy

Passive protection strategies refer to the set of preventive measures taken to ensure higher robustness of a network against certain type of failures. In case of cascading failures, the load distribution in the network plays the most important role. Equally distributed load makes a network more robust against cascades. For example, randomly generated ER network is equally robust against random failures and targeted attacks. A uniform link distribution does not produce hubs and every failure gets easily mitigated by other nodes. However, European NREN is not random and similarly to other man made technological networks, it has a long-tail degree distribution. Distributions of other centrality measures such as closeness and betweenness follow the same pattern. Following a theoretical assumption made by Motter and Lai [75] that the load of each node is proportional to its betweenness score, we can conclude that load distribution in European NREN also has a long-tail shape. The same model presumes that each node has an excess capacity proportional to its load.

Having this in mind, there are two possible approaches to make a network more robust against cascades. The first one is to "flatten" the load distribution. If the routing strategy is considered unchanged, it could be done by adding more links and therefore changing the topology of the network. That way the loads would be more equally distributed and the number and importance of hubs would be decreased. The second strategy does not change the topology but adjusts the excess capacity distribution. In the case of a critical node failure, other nodes would be able to withstand the load increase and prevent the congestion. Both solutions are costly and imply a multivariate optimization approach. In this thesis, the focus is on capacity increase, as the effective topology change is unrealistic for an infrastructural network of a size of European NREN.

First, let us consider a case where the goal is to prevent the cascade to happen at all. After the initial failure of a node i, the load L_i carried by the node

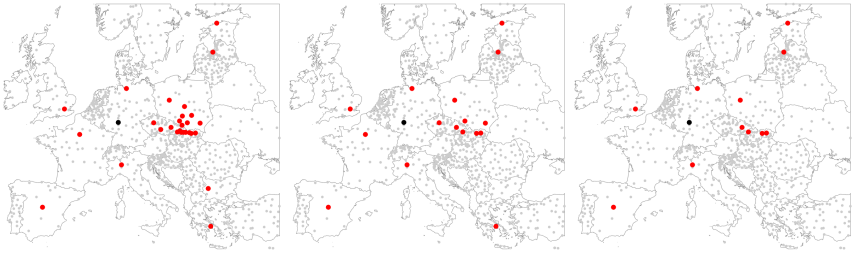
i is going to be distributed among other nodes in the network. A load will most likely not be distributed equally. Since the topology suddenly changes, different nodes take a role of a hub and become responsible for taking over the majority of the load. Those secondary hubs are the most critical nodes in this phase of a cascade. If their capacity is large enough, the cascade would not occur. Therefore, we refer to them as a *protective hubs*, nodes able to withstand an increased amount of traffic if the node i fails. Each node able to cause a cascade has a set of protective hubs. A set of the protective hubs $H_{i,\alpha}$ depends on the node i and tolerance parameter α. By simulating the failures of a single node and identifying the protective sets for various α, we find out that usually $H_{i,\alpha_1} \subseteq H_{i,\alpha_2}$ if $\alpha_1 \geq \alpha_2$.

Based on the chosen tolerance parameter α, a decision maker can chose which nodes to protect having in mind the available budget. In this case, a protection refers to the increased capacity of a designated nodes. The set with the highest priority should be the set $H_{i,\alpha}$ where the parameter α is large (≈ 1.5). Then, a decision maker can chose to protect larger set of nodes which could be identified by lowering the tolerance parameter. A various sets of protective nodes for a single node failure are shown in Figure 5.25. For a relatively small α (≈ 1.02) a large set of protective nodes is identified. If we consider larger α, the set of protective nodes decreases in size.

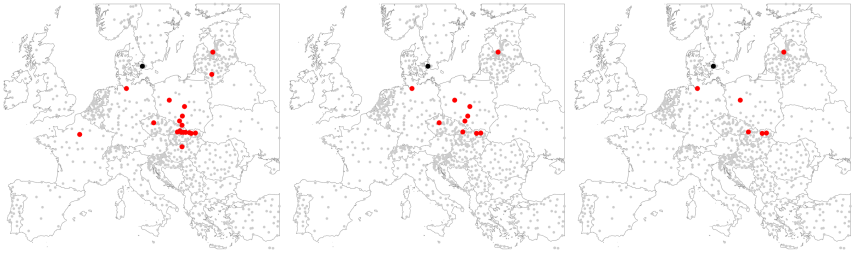
How much the capacity of each protective hub should be increased to prevent a cascade? After the initial failure, all protective hubs become overloaded. In the Motter-Lai model of cascading failures used here for simulation, even a smallest overload would cause a node to fail. Therefore, a capacity of each protective hub should be increased for at least an amount of expected excess of load:

$$C_{prot} = C_{init} + L_{exc}$$

The equation is valid for each protective hub i. The value C_{prot} is a resulting capacity of a protective hub able to withstand the increased load L_{exc} over the capacity C_{init} allocated initially for the node i.



$\alpha = 1.02$ $\alpha = 1.12$ $\alpha = 1.50$



$\alpha = 1.02$ $\alpha = 1.12$ $\alpha = 1.50$

FIGURE 5.25: **Initially congested nodes in the case of a single failure - protective hubs.** After the initial failure of a single node, a traffic load carried by the broken node gets allocated among the rest of the nodes within the network. It often causes a congestion which could cause further failures. The set of initially congested nodes is the most responsible for mitigating the impact and sustain a proper network operation. Increasing the capacity of those nodes could stop a cascading failure. The first row shows congested nodes (red) after a failure of a single node in Frankfurt (black). Three pictures show different sets of congested nodes depending on a chosen tolerance parameter α. The larger the tolerance parameter, the smaller number of congested nodes. The second row shows the same, but for the case of a node failure in Copenhagen.

$\alpha = 1.02$					$\alpha = 1.06$					$\alpha = 1.10$				
lat.	*lon.*	*location*	*budg*	*freq.*	*lat.*	*lon.*	*location*	*budg*	*freq.*	*lat.*	*lon.*	*location*	*budg*	*freq.*
50.8	20.6	Kiellze, PL	1.3	7	49.0	21.2	Prešov, SL	11.1	6	48.7	19.1	B. Bys., SL	1.2	5
49.0	21.2	Prešov, SL	11.5	6	48.7	19.1	B. Bys., SL	1.2	5	48.8	18.0	Trencin, SL	1.2	5
48.7	19.1	B. Bys., SL	1.3	5	48.8	18.0	Trencin, SL	1.2	5	50.2	18.6	Gliwice, PL	1.2	5
48.8	18.0	Trencin, SL	1.3	5	50.2	18.6	Gliwice, PL	1.2	5	50.8	19.1	Czecht., PL	1.3	4
49.8	19.0	B.Biala, PL	1.4	5	50.8	19.1	Czecht., PL	1.4	5	50.8	20.6	Kiellze, PL	1.2	4
50.2	18.6	Gliwice, PL	1.3	5	51.7	19.4	Lodz, PL	1.3	5	53.5	10.0	Hamburg, DE	10.3	4
50.8	19.1	Czecht. PL	1.5	5	49.6	17.2	Olomouc, CZ	35.9	5	49.2	16.6	Brno, CZ	2.2	4
50.0	19.9	Krakow, PL	1.9	5	49.1	18.3	Puchov, SL	7.3	4	49.6	17.2	Olomouc, CZ	34.6	4
51.7	19.4	Lodz, PL	1.4	5	50.8	20.6	Kiellze, PL	1.3	4	48.2	16.3	Vienna, AT	1.8	4
40.4	-3.7	Madrid, ES	3.2	5	53.7	20.4	Olsztyn, PL	1.4	4	48.2	16.3	Vienna, AT	1.8	4

$\alpha = 1.20$					$\alpha = 1.30$					$\alpha = 1.50$				
lat.	*lon.*	*location*	*budg*	*freq.*	*lat.*	*lon.*	*location*	*budg*	*freq.*	*lat.*	*lon.*	*location*	*budg*	*freq.*
49.6	17.2	Olomouc, CZ	31.7	4	49.6	17.2	Olomouc, CZ	29.3	4	49.6	17.2	Olomouc, CZ	25.4	4
48.9	20.5	S.N.Ves, SL	6.7	3	48.9	20.5	S.N.Ves, SL	6.2	3	48.9	20.5	S.N.Ves, SL	5.4	3
49.0	21.2	Prešov, SL	9.8	3	49.0	21.2	Prešov, SL	9.0	3	49.0	21.2	Prešov, SL	7.8	3
49.1	18.3	Puchov, SL	6.4	3	49.1	18.3	Puchov, SL	5.9	3	49.1	18.3	Puchov, SL	5.1	3
53.5	10.0	Hamburg, DE	9.4	3	53.5	10.0	Hamburg, DE	8.7	3	53.5	10.0	Hamburg, DE	7.5	3
37.9	23.7	Athens, GR	1.9	3	56.9	24.1	Riga, LV	11.5	3	56.9	24.1	Riga, LV	9.9	3
56.9	24.1	Riga, LV	12.4	3	49.2	16.6	Brno, CZ	1.9	3	49.2	16.6	Brno, CZ	1.6	3
49.2	16.6	Brno, CZ	2.0	3	48.2	16.3	Vienna, AT	1.5	3	48.2	16.3	Vienna, AT	1.3	3
48.2	16.3	Vienna, AT	1.7	3	48.2	16.3	Vienna, AT	1.5	3	48.2	16.3	Vienna, AT	1.3	3
48.2	16.3	Vienna, AT	1.7	3	48.7	19.1	B. Bys., SL	1.0	2	52.4	16.9	Poznan, PL	3.4	2

TABLE 5.4: **The most important protective hubs.** The list of ten most important protective hubs is shown for various values of α. The hubs are identified by simulating a failure of ten most critical nodes individually. The nodes which appear in the set of protective hubs are then chosen according to the frequency of their appearance. The column *budg* represents the ratio between the original C_{init} and required C_{prot} capacity for avoiding the failure due to congestion. If a single node has a multiple C_{prot}, the one with the maximal value is chosen.

To identify the most important protective hubs, the following numerical analysis is conducted for the European NREN case: Ten most critical nodes (from Table 5.1) are analyzed For each node i, a set of protective hubs $H_{i,\alpha}$ is identified. Then the union of $H_{i,\alpha}$ for all i is produced taking in count also the multiplicity of the elements. This way, we have an occurrence frequency of all the nodes within the various protective sets. If a certain node appears to be in many protective sets, such a node is a candidate for a protection. The list of nodes who appear frequently in protective sets of top ten most critical nodes are shown in Table 5.4. As well as the protective sets for each node change with the tolerance parameter α, the nodes candidates for a protection shown in the table are different for various α.

As mentioned before, the protection of chosen nodes involves a certain capacity increase. In the Table 5.4, the column *budget* shows the ratio between the capacity required for preventing the congestion C_{prot} and the original capacity C_{init}, so $budg = C_{prot}/C_{init}$. For example, to protect a secondary hub in Madrid for $\alpha = 1.02$, the capacity of the node should be increased 3.2 times. A single protective hub has different values of C_{prot} depending on the chosen failed node i. In order to maintain the highest level of connectivity and to avoid any failure due to congestion the budget is chosen so that $C_{prot} = \max(C_{prot}(i))$.

The additional analysis shows that any capacity increase smaller than C_{prot} does not prevent the cascade. Furthermore, increasing the capacity above the C_{prot} does not add to the protection. Decision maker is therefore not in a position to choose how much he should increase the capacity of a certain node, but only to choose which nodes to protect and to pay the appropriate price. This leads to the multivariate cost benefit analysis, which is out of the scope of this chapter. However, the additional analysis shows that protective hubs with larger excess of load and higher frequency should be protected first. The nodes with the high frequency of occurrence within the sets of protective hubs $H_{i,\alpha}$ are more likely to be congested in the case of the important node failure. Furthermore, in the case of congestion of hubs with relatively large C_{prot}, a large amount of traffic needs to be allocated to other nodes. Therefore, the nodes with high C_{prot} should be priority. It is important to consider in the analysis that *budget* value in the Table 5.4 is a relative to the original capacity of the node C_{init}. However, the absolute values of excess of load L_{exc} and required capacity C_{prot} should be considered when deciding on protection strategy. The nodes with high initial capacity could have a small ratio between the initial and required capacity, but the potential failure could produce a considerable excess of load which can cause further congestions.

5.6 Limitations of the Proposed Models

It is important to recognize certain limitation introduced by the models used in this Chapter. The most prominent is the **level of abstraction** which should be considered in assessing real world networks. In all of the used models, the network is considered to consists of certain number of simple elements such as nodes and edges which behave simply and expectedly. The loads and capacities are assumed based on the network topology. In reality those quantities could vary significantly. Furthermore, the mechanism of node failure described in the Motter-Lai model does not correspond fully to what happens when a congestion occurs in reality. Usually, the nodes do not actually fail, but discard some information or keep it in the internal buffer. However, the amount of information needed for fully realistic assessment is very difficult to acquire, and theoretical assumptions should be sufficient for overall network analysis and some of the general conclusions. Another limitation worth addressing is the **shortest path problem**. In all of the models it is assumed that information seeks the shortest route from origin to destination. This behavior quickly produces the highly loaded nodes which behave like hubs. In reality, the routing policies could vary. The information packet could be routed so it avoids congestion. Modern policies include multipath routing where redundancy increases diversity and robustness of the data transfer. Many providers use leased lines whose cost could influence the routing policies, so the information often does not follow the optimal but the least expensive path. On top of that, there is prioritized data which require special treatment. All of this could be more complicated with the dynamical routing, where the policies change periodically or after an event. Still, for the sake of simplicity, but not diverging much from the real network dynamics, researchers legitimately assume the shortest path routes.

5.7 Conclusion

In this Chapter, the particular European NREN network is assessed against cascading failures. Two models for cascades are used, namely Mottel-Lai [75] and Crucitti-Latora-Marchiori [27] model. The most critical nodes, those whose failure would cause the largest cascade, are identified both for the individual and multiple simultaneous failures. Geographically correlated failures are also assessed, having in mind that external factors such as earthquakes, floods or power outages usually affect a group of nodes within the area. Additionally, an example of the external risk analysis is shown. The risk of earthquakes is evaluated using two-dimensional analysis which includes both the node importance and the potential damage caused by the earthquake.

Besides the node analysis, some protection strategies are proposed. Costless strategies imply an active measures after the cascade and they are based on the intentional node removal after the cascade. Additional nodes are removed in order to reduce the overall traffic in the network and limit the cascade which is caused by the excess of loads. Nodes could be removed locally in the vicinity of failed node or globally, which requires full control over all the nodes in network. In addition to costless, there are also costly protection strategies. They imply some sort of investment in the increased capacity which take place before the cascade. Therefore, they are also referred to as passive protection strategies.

The analyses and proposed protection methods are useful to give an insight in the complexity of cascading failures phenomena. There is rarely the universal solution for protection against cascades and the methods usually vary from network to network. Some instances from Section 5.2 could be used as a basis for future research in identifying the most critical network elements. Additionally, the concept of costly protection strategies presented in Section 5.5 open the possibility for implementation of various multivariate optimization methods. Problems discussed here might be considered as a mere scratching

the surface of wide set of possible challenges. However, addressing all the issues more broadly would be out of the scope of this Thesis.

Bibliography

[1] E. W. Weisstein, "Königsberg Bridge Problem."

[2] W. O. Kermack and A. G. McKendrick, "A Contribution to the Mathematical Theory of Epidemics," *The Royal Society A*, vol. 115, 1927.

[3] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D. Hwang, "Complex networks: Structure and dynamics," *Physics Reports*, vol. 424, pp. 175–308, Feb. 2006.

[4] L. C. Freeman, "Centrality in Social Networks Conceptual Clarification," *Social Networks*, vol. 1, pp. 215–239, 1978.

[5] A. Klein, H. Ahlf, and V. Sharma, "Social activity and structural centrality in online social networks," *Telematics and Informatics*, vol. 32, pp. 321–332, May 2015.

[6] Mark Newman, "The structure and function of complex networks," *SIAM Review*, vol. 45, p. 58, 2003.

[7] R. Diestel, *Graph Theory*. Springer-Verlag, Heidelberg, 2006.

[8] J. Ripoll, M. Manzanob, and E. Calle, "Spread of epidemic-like failures in telecommunication networks," *Physica A*, vol. 410, pp. 457–569, 2014.

[9] T. G. Lewis, *Critical Infrastructure Protection in Homeland Security: Defending a Networked Nation*. Hoboken, New Jersey: John Wiley & Sons, Inc., 2006.

[10] X. Long, D. Tipper, and T. Gomes, "Measuring the survivability of networks to geographic correlated failures," *Optical Switching and Networking*, no. 0, p. 17, 2014.

[11] E. K. Çetinkaya, M. J. Alenazi, Y. Cheng, A. M. Peck, and J. P. Sterbenz, "A comparative analysis of geometric graph models for modelling backbone networks," *Optical Switching and Networking*, vol. 14, pp. 95–106, Aug. 2014.

[12] E. K. Çetinkaya, D. Broyles, A. Dandekar, S. Srinivasan, and J. P. G. Sterbenz, "Modelling communication network challenges for Future Internet resilience, survivability, and disruption tolerance: a simulation-based approach," *Telecommunication Systems*, Sept. 2011.

[13] M. F. Habib, M. Tornatore, F. Dikbiyik, and B. Mukherjee, "Disaster survivability in optical communication networks," *Computer Communications*, vol. 36, pp. 630–644, Mar. 2013.

[14] A. Barrat, M. Barthelemy, and A. Vespignani, *Dynamical processes on complex networks*. Cambridge University Press, 2008.

[15] L. A. N. Amaral and J. M. Ottino, "Complex networks," *The European Physical Journal B - Condensed Matter*, vol. 38, pp. 147–162, Mar. 2004.

[16] M. Prokopenko, F. Boschetti, and A. J. Ryan, "An information-theoretic primer on complexity, self-organization, and emergence," *Complexity*, vol. 15, pp. 11–28, sep 2009.

[17] D. Henry and J. Emmanuel Ramirez-Marquez, "Generic metrics and quantitative approaches for system resilience as a function of time," *Reliability Engineering & System Safety*, vol. 99, pp. 114–122, Mar. 2012.

[18] D. E. Alexander, "Resilience and disaster risk reduction: an etymological journey," *Natural Hazards and Earth System Science*, vol. 13, pp. 2707–2716, Nov. 2013.

[19] K. Barker, J. E. Ramirez-Marquez, and C. M. Rocco, "Resilience-based network component importance measures," *Reliability Engineering and System Safety*, vol. 117, pp. 89–98, 2013.

[20] P. Smith, D. Hutchison, J. P. G. Sterbenz, M. Schöller, A. Fessi, M. Karaliopoulos, C. Lac, and B. Plattner, "Network Resilience: A Systematic Approach," *Communication Magazine*, pp. 88–97, 2011.

[21] IEEE, "IEEE Standard Glossary of Software Engineering Terminology," Dec. 1990.

[22] Council of Europe, "Council Directive 2008/114/EC on the identification and designation of European critical infrastructures and the assessment of the need to improve their protection," *Official Journal of the European Union L*, vol. 345, 2008.

[23] J. Gao, B. Barzel, and A.-L. Barabási, "Universal resilience patterns in complex networks," *Nature*, vol. 530, pp. 307–312, feb 2016.

[24] M. Newman, A.-L. Barabási, and D. J. Watts, *The Structure and Dynamics of Networks*. Princeton: Princeton University Press, 2006.

[25] S. Banerjee, S. Shirazipourazad, P. Ghosh, and A. Sen, "Beyond connectivity - new metrics to evaluate robustness of networks," in *2011 IEEE 12th International Conference on High Performance Switching and Routing*, pp. 171–177, IEEE, July 2011.

[26] R. Albert, H. Jeong, and A.-L. Barabasi, "Error and Attack Tolerance of Complex Networks," *Nature*, vol. 406, pp. 378–382, 2000.

[27] P. Crucitti, V. Latora, and M. Marchiori, "Model for cascading failures in complex networks," *Rapid Communications*, 2004.

[28] J. Kephart and S. White, "Directed-graph epidemiological models of computer viruses," in *Proceedings. 1991 IEEE Computer Society Symposium on Research in Security and Privacy*, pp. 343–359, IEEE Comput. Soc. Press, 1991.

[29] D. Chakrabarti, Y. Wang, C. Wang, J. Leskovec, and C. Faloutsos, "Epidemic thresholds in real networks," *ACM Transactions on Information and System Security*, vol. 10, no. 4, pp. 1–26, 2008.

[30] M. Barthélemy, A. Barrat, R. Pastor-Satorras, and A. Vespignani, "Velocity and Hierarchical Spread of Epidemic Outbreaks in Scale-Free Networks," *Physical Review Letters*, vol. 92, p. 178701, apr 2004.

[31] G. Murić, C. Scheunert, and E. A. Jorswieck, "On modeling epidemics in networks using linear time-invariant dynamics," in *The IEEE WiMob 2015 Workshop on Emergency Networks for Public Protection and Disaster Relief*, (Abu Dhabi), pp. 138–146, 2015.

[32] F. D. Malliaros, M.-E. G. Rossi, and M. Vazirgiannis, "Locating influential nodes in complex networks.," *Scientific reports*, vol. 6, p. 19307, jan 2016.

[33] K. Klemm, M. Á. Serrano, V. M. Eguíluz, and M. S. Miguel, "A measure of individual role in collective dynamics," *Scientific Reports*, vol. 2, no. 292, 2012.

[34] F. Bauer and J. T. Lizier, "Identifying influential spreaders and efficiently estimating infection numbers in epidemic models: a walk counting approach," *Europhysics Letters*, vol. 99, no. 68007, 2012.

[35] G. Murić, E. Jorswieck, and C. Scheunert, "Using LTI Dynamics to Identify the Influential Nodes in a Network," *PLOS ONE*, vol. 11, dec 2016.

[36] P. Erdös and a. Rényi, "On random graphs," *Publicationes Mathematicae*, vol. 6, pp. 290–297, 1959.

[37] M. Molloy and B. Reed, "A critical point for random graphs with a given degree sequence," *Random Structures & Algorithms*, vol. 6, pp. 161–180, Mar. 1995.

[38] B. Söderberg, "General formalism for inhomogeneous random graphs," *Physical Review E*, vol. 66, p. 066121, Dec. 2002.

[39] P. W. Holland and S. Leinhardt, "An exponential family of probability distributions for directed graphs," *Journal of the American Statistical Association*, vol. 76, no. 373, pp. 33–50, 1981.

[40] D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small-world' networks.," *Nature*, vol. 393, pp. 440–2, June 1998.

[41] M. Barthélémy and L. A. N. Amaral, "Small-World Networks: Evidence for a Crossover Picture," *Physical Review Letters*, vol. 82, pp. 3180–3183, Apr. 1999.

[42] A. Barrat and M. Weigt, "On the properties of small-world network models," *Europ. Phys. J. B 13, 547*, vol. 13, pp. 547–560, Mar. 2000.

[43] A. Barabási, "Emergence of Scaling in Random Networks," *Science*, vol. 286, pp. 509–512, Oct. 1999.

[44] G. Bianconi and A.-L. Barabási, "Competition and multiscaling in evolving networks," *Europhysics Letters*, vol. 54, p. 13, nov 2000.

[45] K. Klemm and V. M. Eguíluz, "Growing scale-free networks with small-world behavior," *Physical Review E*, vol. 65, p. 057102, May 2002.

[46] S. P. Borgatti and M. G. Everett, "A Graph-theoretic perspective on centrality," *Social Networks*, vol. 28, pp. 466–484, Oct. 2006.

[47] P. Hagmann, L. Cammoun, X. Gigandet, R. Meuli, C. J. Honey, V. J. Wedeen, and O. Sporns, "Mapping the Structural Core of Human Cerebral Cortex," *PLoS Biology*, vol. 6, p. e159, jul 2008.

[48] J. I. Alvarez-Hamelin, L. Dall'Asta, A. Barrat, and A. Vespignani, "k-core decomposition: a tool for the visualization of large scale networks," *Advances in Neural Information Processing Systems*, apr 2005.

[49] M. Kitsak, L. K. Gallos, S. Havlin, F. Liljeros, L. Muchnik, H. E. Stanley, and H. A. Makse, "Identification of influential spreaders in complex networks," *Nature Physics*, vol. 6, pp. 888–893, aug 2010.

[50] L. Lü, T. Zhou, Q.-M. Zhang, and H. E. Stanley, "The H-index of a network node and its relation to degree and coreness," *Nature Communications*, vol. 7, p. 10168, jan 2016.

[51] J.-G. Liu, J.-H. Lin, Q. Guo, and T. Zhou, "Locating influential nodes via dynamics-sensitive centrality," *Scientific Reports*, vol. 6, p. 21380, feb 2016.

[52] T. Opsahl, F. Agneessens, and J. Skvoretz, "Node centrality in weighted networks: Generalizing degree and shortest paths," *Social Networks*, vol. 32, pp. 245–251, July 2010.

[53] M. E. J. Newman, "A measure of betweenness centrality based on random walks," *Social Networks*, vol. 27, pp. 39–54, 2005.

[54] S. Dolev, Y. Elovici, and R. Puzis, "Routing Betweenness Centrality," *Journal of the ACM*, vol. 57, pp. 1–27, Apr. 2009.

[55] H. Wang, J. M. Hernandez, and P. Van Mieghem, "Betweenness centrality in a weighted network," *Physical Review E*, vol. 77, p. 046105, Apr. 2008.

[56] J. D. Noh and H. Rieger, "Random walks on complex networks," *Phys. Rev. Lett.*, vol. 92, no. 118701, p. 5, 2004.

[57] L. C. Freeman, S. P. Borgatti, and D. R. White, "Centrality in valued graphs: A measure of betweenness based on network flow," *Social Networks*, vol. 13, pp. 141–154, June 1991.

[58] S. P. Borgatti, "Centrality and network flow," *Social Networks*, vol. 27, pp. 55–71, Jan. 2005.

[59] C. Dangalchev, "Residual closeness in networks," *Physica A: Statistical Mechanics and its Applications*, vol. 365, pp. 556–564, June 2006.

[60] P. Crucitti, V. Latora, M. Marchiori, and A. Rapisarda, "Efficiency of scale-free networks: error and attack tolerance," *Physica A*, vol. 320, pp. 622–642, 2003.

[61] P. Holme, B. J. Kim, C. N. Yoon, and S. K. Han, "Attack vulnerability of complex networks," *Phys. Rev. E*, vol. 65, 2002.

[62] I. Mishkovski, M. Biey, and L. Kocarev, "Vulnerability of complex networks," *Communications in Nonlinear Science and Numerical Simulation*, vol. 16, pp. 341–349, 2010.

[63] V. Latora and M. Marchiori, "Efficient Behavior of Small-World Networks," *Physical Review Letters*, vol. 87, no. 19, 2001.

[64] V. Latora and M. Marchiori, "Vulnerability and protection of infrastructure networks," *Physical Review E*, vol. 71, 2005.

[65] P. Crucitti, V. Latora, M. Marchiori, and A. Rapisarda, "Error and attack tolerance of complex networks," *Physica A*, vol. 340, pp. 388–394, 2004.

[66] 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.

[67] J. C. Doyle, D. L. Alderson, L. Li, S. Low, M. Roughan, S. Shalunov, R. Tanaka, and W. Willinger, "The "robust yet fragile" nature of the Internet.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, pp. 14497–502, oct 2005.

[68] Y. Chen, G. Paul, R. Cohen, S. Havlin, S. P. Borgatti, F. Liljeros, and H. E. Stanley, "Percolation theory and fragmentation measures in social networks," *Physica A*, vol. 378, pp. 11–19, 2007.

[69] M. Barthelemy, A. Barrat, R. Pastor-Satorras, and A. Vespignani, "Dynamical patterns of epidemic outbreaks in complex heterogeneous networks," *Journal of Theoretical Biology 235*, vol. 235, p. 13, oct 2005.

[70] M. J. Keeling, "The effects of local spatial structure on epidemiological invasions," *Proceedings of the Royal Society B: Biological Sciences*, vol. 266, pp. 859–867, Apr. 1999.

[71] A. E. Motter, "Cascade Control and Defense in Complex Networks," *Physical Review Letters*, vol. 93, p. 098701, Aug. 2004.

[72] M. Piraveenan, M. Prokopenko, and L. Hossain, "Percolation Centrality: Quantifying Graph-Theoretic Impact of Nodes during Percolation in Networks," *PLOS one*, vol. 8, p. e53095, Jan. 2013.

[73] P. Holme, "Efficient local strategies for vaccination and network attack," *Europhysics Letters (EPL)*, vol. 68, pp. 908–914, dec 2004.

[74] E. Hennigan and D. Bren, "Cascade Failure in Distributed Networks," 2009.

[75] A. E. Motter and Y.-C. Lai, "Cascade-based attacks on complex networks," *Phys. Rev. E*, vol. 66, no. 065102, 2002.

[76] D. J. Watts, "A simple model of global cascades on random networks," in *Proceedings of the National Academy of Sciences*, vol. 99, pp. 5766–5771, 2002.

[77] H. A. Q. Tran and A. Namatame, "Improve Network's Robustness against Cascade with Rewiring," *Procedia Computer Science*, vol. 24, pp. 239–248, 2013.

[78] P. Shakarian, A. Bhatnagar, A. Aleali, E. Shaabani, and R. Guo, *Diffusion in Social Networks*. SpringerBriefs in Computer Science, Cham: Springer International Publishing, 2015.

[79] J. Johansson, *Risk and Vulnerability Analysis of Interdependent Technical Infrastructures: Addressing Socio-Technical Systems*. PhD thesis, Lund, Sweden, 2010.

[80] M. Ouyang, L. Dueñas Osorio, and X. Min, "A three-stage resilience analysis framework for urban infrastructure systems," *Structural Safety*, vol. 36-37, pp. 23–31, 2012.

[81] M. G. Kendall, "A New Measure of Rank Correlation," *Biometrika*, vol. 30, p. 81, jun 1938.

[82] L. Bianchi, M. Dorigo, L. M. Gambardella, and W. J. Gutjahr, "A survey on metaheuristics for stochastic combinatorial optimization," *Natural Computing*, vol. 8, pp. 239–287, jun 2009.

[83] M. C. scientist) Mitchell, *An introduction to genetic algorithms*. MIT Press, 1998.

[84] I. The MathWorks, "MATLAB and Global Optimization Toolbox Release 2015a."

[85] G. Lawyer, "Understanding the influence of all nodes in a network.," *Scientific reports*, vol. 5, p. 8665, jan 2015.

[86] M. Benzi and C. Klymko, "On the limiting behavior of parameter-dependent network centrality measures," *SIAM Journal on Matrix Analysis and Applications*, vol. 36, pp. 686–706, dec 2013.

[87] P. Bonacich, "Power and Centrality: A Family of Measures," *American Journal of Sociology*, vol. 95, no. 2, pp. 1170–1182, 1987.

[88] J. Goutsias and G. Jenkinson, "Markovian dynamics on complex reaction networks," *Physics Reports*, vol. 529, pp. 199–264, aug 2013.

[89] E. Gansner, E. Koutsofios, S. North, and K.-P. Vo, "A technique for drawing directed graphs," *IEEE Transactions on Software Engineering*, vol. 19, pp. 214–230, mar 1993.

[90] S. S. George T. Heineman, Gary Pollice, *Algorithms in a Nutshell*. O'Reilly Media, 2008.

[91] J. B. Kruskal, "On the Shortest Spanning Subtree of a Graph and the Traveling Salesman Problem," *Proceedings of the American Mathematical Society*, vol. 7, no. 1, pp. 48–50, 1956.

[92] T. Dalzell and E. Partridge, *The Routledge dictionary of modern American slang and unconventional English*. Routledge, 2009.

[93] G. Association, "GÉANT Association Compendium of National Research and Education Networks in Europe," tech. rep., 2014.

[94] S. Al-Agtash, "Developing a Lebanese National Research and Education Network," tech. rep., 2011.

[95] J. Dyer, "The case for National Research and Education Networks (NRENs)," 2009.

[96] S. Knight, H. Nguyen, N. Falkner, R. Bowden, and M. Roughan, "The internet topology zoo," *Selected Areas in Communications, IEEE Journal on*, vol. 29, pp. 1765 –1775, october 2011.

[97] Y.-Y. Liu, J.-J. Slotine, and A.-L. Barabási, "Controllability of complex networks.," *Nature*, vol. 473, pp. 167–73, may 2011.

[98] J. G. Proakis and D. G. Manolakis, *Digital Signal Processing*. Pearson Education Limited, 2007.

[99] H. Tanner, "On the controllability of nearest neighbor interconnections," in *2004 43rd IEEE Conference on Decision and Control (CDC) (IEEE Cat. No.04CH37601)*, vol. 3, pp. 2467–2472 Vol.3, IEEE, 2004.

[100] E. Estrada and J. A. Rodríguez-Velázquez, "Subgraph centrality in complex networks," *Physical Review E*, vol. 71, p. 056103, may 2005.

[101] L. Lü, D. Chen, X.-L. Ren, Q.-M. Zhang, Y.-C. Zhang, and T. Zhou, "Vital nodes identification in complex networks," *Physics Reports*, vol. 650, pp. 1–63, 2016.

[102] S. Pei and H. A. Makse, "Spreading dynamics in complex networks," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2013, 2013.

[103] M. Šikić, A. Lančić, N. Antulov-Fantulin, and H. Štefančić, "Epidemic centrality — is there an underestimated epidemic impact of network peripheral nodes?," *The European Physical Journal B*, vol. 86, p. 440, oct 2013.

[104] S. Carmi, S. Havlin, S. Kirkpatrick, Y. Shavitt, and E. Shir, "A model of Internet topology using k-shell decomposition," *Proceedings of the National Academy of Sciences*, vol. 104, pp. 11150–11154, jul 2007.

[105] M. P. Viana, J. L. B. Batista, and L. d. F. Costa, "Effective number of accessed nodes in complex networks.," *Physical review. E, Statistical, nonlinear, and soft matter physics*, vol. 85, p. 036105, mar 2012.

[106] H. Hoffmann, "Simple violin plot using matlab default kernel density estimation.," 2015.

[107] E. Estrada, "Generalized walks-based centrality measures for complex biological networks," *Journal of Theoretical Biology*, vol. 263, no. 4, pp. 556–565, 2010.

[108] V. Batagelj and U. Brandes, "Efficient generation of large random networks," *Physical Review E*, vol. 71, p. 036113, mar 2005.

[109] A. Taylor and D. J. Higham, "CONTEST," *ACM Transactions on Mathematical Software*, vol. 35, pp. 1–17, feb 2009.

[110] J. Leskovec and A. Kravl, "SNAP Datasets: Stanford Large Network Dataset Collection," 2014.

[111] Z. L. Zhang B, Liu R, Massey D, "Internet Topology Project."

[112] S. L. Hakimi, "On Realizability of a Set of Integers as Degrees of the Vertices of a Linear Graph. I," *Journal of the Society for Industrial and Applied Mathematics*, vol. 10, no. 3, pp. 496–506, 1962.

[113] M. Milena and V. Nisheeth, "On Generating Graphs with Prescribed Vertex Degrees for Complex Network Modeling," in *ARACNE 2002: 3rd Workshop on Approximation and Randomization Algorithms in Communication NEtworks*, College of Computing Georgia Institute of Technology, 2002.

[114] Z. Chen, J. Zhang, W.-B. Du, O. Lordan, and J. Tang, "Optimal Allocation of Node Capacity in Cascade-Robustness Networks," *PLOS ONE*, vol. 10, p. e0141360, oct 2015.

[115] T. Ohira and R. Sawatari, "Phase transition in a computer network traffic model," *Physical Review E*, vol. 58, pp. 193–195, jul 1998.

[116] A. Arenas, A. Díaz-Guilera, and R. Guimerà, "Communication in networks with hierarchical branching.," *Physical review letters*, vol. 86, pp. 3196–9, apr 2001.

[117] Z. Liu, M.-B. Hu, R. Jiang, W.-X. Wang, and Q.-S. Wu, "Method to enhance traffic capacity for scale-free networks.," *Physical review. E, Statistical, nonlinear, and soft matter physics*, vol. 76, p. 037101, sep 2007.

[118] B. Danila, Y. Yu, S. Earl, J. A. Marsh, Z. Toroczkai, and K. E. Bassler, "Congestion-gradient driven transport on complex networks.," *Physical review. E, Statistical, nonlinear, and soft matter physics*, vol. 74, p. 046114, oct 2006.

[119] X. Wang, S. Guan, and C. Heng Lai, "Protecting infrastructure networks from cost-based attacks," *New Journal of Physics*, vol. 11, p. 033006, mar 2009.

[120] G. David, "MatlabBGL," 2006.

[121] S. Neumayer, G. Zussman, R. Cohen, and E. Modiano, "Assessing the Vulnerability of the Fiber Infrastructure to Disasters," in *IEEE INFOCOM 2009 - The 28th Conference on Computer Communications*, pp. 1566–1574, IEEE, apr 2009.

[122] S. Trajanovski, F. A. Kuipers, A. Ilic, J. Crowcroft, and P. Van Mieghem, "Finding Critical Regions and Region-Disjoint Paths in a Network," *IEEE/ACM Transactions on Networking*, vol. 23, pp. 908–921, jun 2015.

[123] K. Leelardcharoen, *Interdependent Response of Telecommunication and Electric Power Systems to Seizmic Hazard*. PhD thesis, Georgia Institute of Technology, 2011.

[124] ITU-T Focus Group on Disaster Relief Systems Network Resilience and Recovery, "Technical report on Telecommunications and Disaster Mitigation," tech. rep., International Telecommunication Union, 2013.

[125] ETH Zurich on behalf of the EU-FP7 Consortium of SHARE, "Seismic Hazard Harmonization in Europe," 2013.

[126] Earthquake Hazards Program, "Earthquake Hazards Program." Available at `https://earthquake.usgs.gov`.