BUDAPEST UNIVERSITY OF TECHNOLOGY AND ECONOMICS
DEPT. OF TELECOMMUNICATIONS AND MEDIA INFORMATICS

Modeling and Enumerating Geographically Correlated Failure Events in Communication Networks

Ph.D. Dissertation

Balázs Vass

Advisor:

János Tapolcai, D.Sc.
High Speed Networks Laboratory
Dept. of Telecommunications and Media Informatics
Budapest University of Technology and Economics

Budapest, Hungary
2021

Abstract

This dissertation aims to model and list the joint device failures of telecommunication optical backbone networks caused by large-scale regional disasters. The use-cases of these failure lists include helping the operators of modern telecommunication networks to meet the predefined Quality-of-Service (QoS) conditions.

I divided the problem into the following parts: 1) development of failure models that make the best use of the data available, 2) giving fast algorithms for determining the resulting failure lists, 3) providing a theoretical and practical analysis of the complexity of the algorithms and the properties of the failure lists. My thesis relies, in particular, on the foundations of complexity theory, computational geometry, probability theory, and seismology.

The first thesis assumes that while the geographical embedding of the network topology is precisely known, only a maximum size of the disaster-affected area is available as part of the input. Accordingly, I have provided an algorithm for collecting link sets that can be hit by a disk with a radius of r and are maximal among such sets in a *Shared Risk Link Group (SRLG)* list, where an SRLG is defined as a set of links that are considered to have a high probability of failing together. The recommended list is provably short, the algorithm for its enumeration, under certain practical assumptions, has an optimal complexity, that is almost linear in the number of network elements.

According to the assumption of my second thesis, we only know a schematic map of the network (which can be, for example, a rented topology). In this case, I listed the maximum link sets that can be hit by a disk containing a given $k = 0, 1, \ldots$ number of network nodes. This list of SRLGs is provably short and can be determined in low polynomial time of the number of network elements.

The third thesis assumes that, in addition to the geographical embedding of the network, the geographical area affected by the next disaster is known as a random variable. This assumption allows the probabilistic modeling of a set of simultaneously failing network elements. I recommended standard terminology to unify the technical language related to *Probabilistic SRLGs (PSRLGs)*, and built a probabilistic failure model that explicitly takes into account the correlated nature of network element failures. I gave theoretical bounds on the sizes of the recommended PSRLG lists, and tested the model with real seismic input data. According to the test result, the threat to the backbone networks arising from earthquakes can be described by a small, easy-to-query data structure.

For all three of my theses, I created simulation software to answer the arising practical questions.

Kivonat

Disszertációm célja modellezni és felsorolni a távközlési optikai gerinchálózatok nagy kiterjedésű regionális katasztrófák által okozott együttes eszköz-meghibásodásait. Azon linkhalmazokat melyekről azt gondoljuk, hogy érdemi valószínűséggel egyszerre meghibásodhatnak, *közös kockázatú linkcsoportoknak (Shared Risk Link Group, SRLG)* nevezzük. E hibalisták felhasználhatósága szerteágazó, többek közt hozzásegítik a modern távközlő hálózatok operátorait az előre definiált szolgáltatás-minőségi feltételek (QoS - Quality-of-Service) teljesítéséhez.

A problémakört a következő részekre bontottam: 1) az elérhető adatokat legjobban felhasználó hibamodellek kidolgozása, 2) az eredő hibalisták meghatározására szolgáló gyors algoritmusok adása, 3) az algoritmusok bonyolultságának és a hibalisták tulajdonságainak elméleti és gyakorlati elemzése. Tézisem különösen támaszkodik a bonyolultságelmélet, a kombinatorikus geometria, a valószínűségszámítás és a szeizmológia alapjaira.

Az első téziscsoport feltevése, hogy míg a hálózati topológia földrajzi beágyazását pontosan ismerjük, a katasztrófa sújtotta területnek csak egy maximális mérete ismeretes. Ennek megfelelően algoritmust adtam azon linkhalmazok meghatározására, melyek linkenként metszhetők egy r sugarú koronggal, és az ilyen halmazok között maximálisak. Az ajánlott SRLG-lista bizonyíthatóan rövid, a kiszámolására adott algoritmus bizonyos praktikus feltevések mellett optimális, a hálózati elemek számában majdnem lineáris bonyolultságú.

A második téziscsoport feltevése szerint csak egy vázlatos térképét ismerjük a hálózatnak (mely lehet pl. bérelt topológia). Ebben az esetben azon maximális linkhalmazokat listáztam, melyekbe belemetsz egy k darab hálózati csomópontot tartalmazó korong ($k \in \{0, 1, \ldots\}$ adott). Ezen SRLG-k listája bizonyíthatóan rövid, és a hálózati elemek számának alacsony polinomiális idejében meghatározható.

A harmadik téziscsoport felteszi, hogy a hálózat földrajzi beágyazása mellett a következő katasztrófa sújtotta földrajzi terület valószínűségi változóként ismert. E feltevés megengedi hibák probabilisztikus modellezését. Standard terminológiát ajánlottam a *probabilisztikus SRLG-kkel (Probabilistic SRLG, PSRLG)* kapcsolatos szaknyelv egyesítésére. A hálózati elemek korrelált meghibásodását explicit figyelembe vevő hibamodellt ajánlottam, az ajánlott PSRLG-listák méreteire elméleti korlátokat adtam. A modellt valós szeizmikus bemenettel teszteltem, miszerint a gerinchálózatokra a földrengésekből fakadó veszélyt kis méretű, könnyen lekérdezhető adatsruktúrával le lehet írni.

A felmerülő gyakorlati kérdéseket saját szimulációs szoftver segítségével válaszoltam meg.

Acknowledgements

I would like to express my sincere gratitude to my advisor János Tapolcai, for his patient guidance, enthusiastic encouragement, and useful critiques over the past years. His support was indispensable in understanding the research methodologies and becoming a researcher in the field of telecommunications.

I owe special thanks to those who contributed to this thesis work, including Erika Bérczi-Kovács, Lajos Rónyai, David Hay, and Alessandro Valentini. Many thanks to all my co-authors, including, but not limited to, Alija Pašić, Péter Babarczi, Ferenc Mogyorósi, Fernando Kuipers, Jorik Oostenbrink, Teresa Gomes, Stefan Schmid, Gábor Rétvári, Costin Raiciu, József Bíró and Zalán Heszberger. I would also thank Stefan Schmid, David Hay, and Costin Raiciu for hosting me as a guest researcher in 2019.

The work presented here was done at the High Speed Networks Laboratory, Budapest University of Technology and Economics (BME). The instruments of COST CA15127 Action "Resilient Communication Services Protecting End-user Applications from Disaster-based Failures" (RECODIS) and EIT Digital greatly helped my research.

I would like to thank my family, friends, and teachers for guiding me to the Ph.D.: without their continuous support, it would have been hardly possible to write these words of honor. I would like to thank my mother Laura, who taught me to add and multiply numbers until 100 before school, my father Dénes, who explained to me the Earth's movement around the Sun at my age of 4, and my sister Réka, who was my great competitor in chess at home. I would like to thank my (high)school mathematics and physics teachers Géza Dávid, Péterfi Margit, András Ibolya; Zoltán Székely, Károly Cseke-Árkosi and good-old beloved uncle Imre Bardócz (Bokszi) for their continuous support in maths and physics competitions up to national levels both in Romania and Hungary.

Last but not least, I would like to thank my girlfriend Ági for bearing my demand for silence during preparing the first versions of my dissertation while just moving in the same studio flat for the coronavirus quasi-lockdown in the Spring of 2020.

Budapest, Hungary, Balázs Vass
June 18, 2021

Contents

Chapter 1

Introduction

The Internet has become a topmost critical infrastructure. Due to the importance of telecommunication services (as a base for stock market, telesurgery, etc.), improving the preparedness of networks to regional failures is becoming a key issue [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14]. The majority of severe network outages happen because of a disaster (such as an earthquake, hurricane, tsunami, tornado, etc.) taking down a lot of (or all) equipment in a given geographical area. Such failures are called *regional failures*. Many studies have touched the problem of how to prepare networks to survive regional failures, where the first solutions have assumed that fibers in the same duct or within 50 km of every network node fail simultaneously (namely, in a single regional failure) [15, 16]. These solutions were further improved by examining the historical data of the different type of disasters (e.g., seismic hazard maps for earthquakes) and identifying the hotspots of the disasters [2, 5, 6, 8, 9, 11]. The weak point of these approaches is that, during network equipment deployment, many of the risks are considered and compensated (e.g., an earthquake-proof infrastructure in areas with larger seismic intensity), implying that the historical data does not represent the current deployments, and therefore, not the current risks. Thus, it may be more realistic to assume that any physically close-by equipment has a higher chance to fail simultaneously. More recent studies are purely devoted to this particular problem and adapt computational geometric based approaches to capture all of the regional failures and represent them in a compact way [10, 17, 18, 19, C10, C9, C11], where the major challenge is that regional failures can have arbitrary locations, shapes, sizes, effects, etc. Unfortunately, regional failures are not self-discoverable in practice [20]; this, together with the high number of severe network outages witnessed in the last decades [21, 22, 23, 24, 25, 26, 27][1] present clear evidence that selecting the proper list of regional failures is still a challenging problem to solve [2, 3, 5, 6, 7, 8, 9, 10, 11]. To fill this gap in reliable network design, my Ph.D. research is devoted to enhancing the state of the art and suggests unified definitions, notions, and terminology.

The output of the approaches discussed in this Thesis can serve as the input of the network design

[1] A recent example is a few days long telecom outage during Cyclone Amphan in West Bengal in May of 2020 as a result of around 100 fiber cuts due to the falling of trees by the wind speeding up to 190km/h.

and management tools. Currently, network recovery mechanisms are implemented to protect a small set of pre-defined failure scenarios. Each recovery plan corresponds to the failure of some equipment. Informally speaking, when a link (or link set) fails, the network has a ready-to-use plan on how to recover itself. Technically, a set of so-called Shared Risk Link Groups (SRLGs)[2] are defined by the network operators, where each SRLG is a set of links whose joint failure the recovery mechanism should be prepared for. The first part of this Thesis purely focuses on how to define and enumerate SRLGs that cover all types of disasters. In the second part of my Thesis, I address the question of defining a realistic and applicable Probabilistic SRLG (PSRLG) failure model.

It turns out that, surprisingly, in practice, only a small number of SRLGs or PSRLGs are needed to serve as inputs for the higher-layer network management tools. Informally speaking, methods offering lists of SRLGs and PSRLGs translate the composed geometric problem of protecting telecommunication networks against regional failures to small-sized purely combinatorial and probabilistic problems, respectively. These findings open up the possibility of leveraging regional (P)SRLG lists for enhancing network preparedness against disasters.

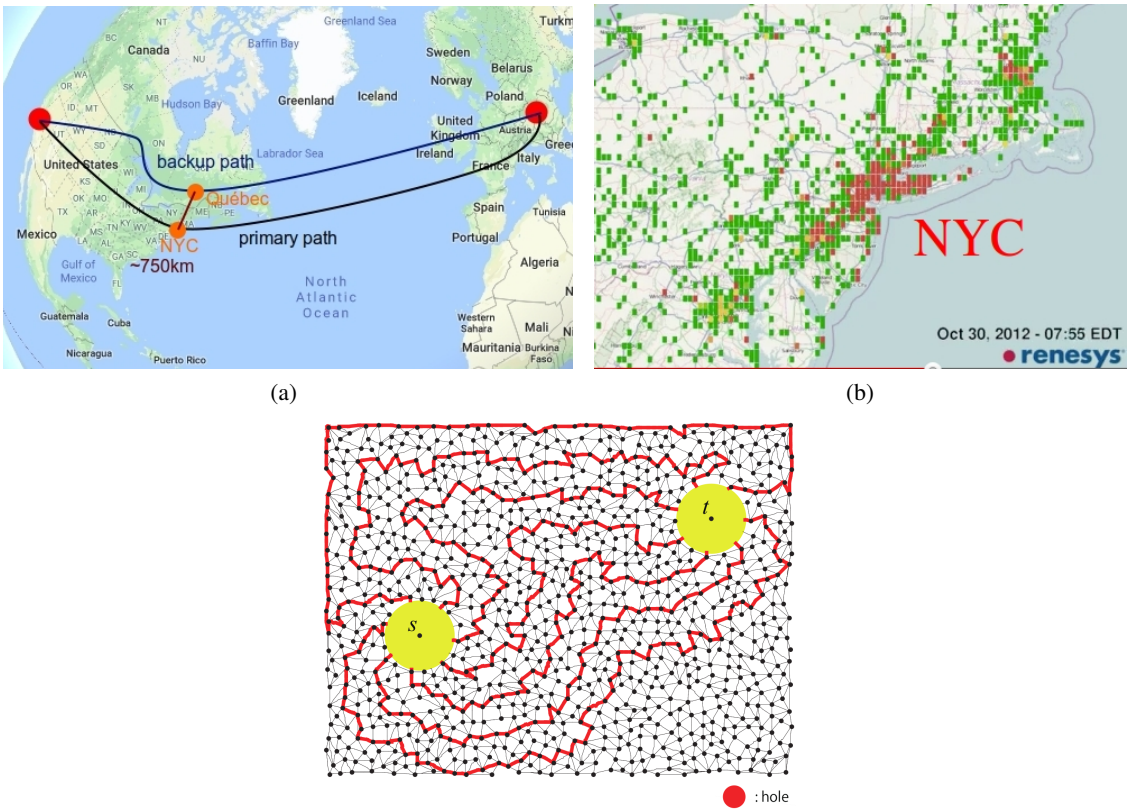
1.1 Example Use-Cases of SRLG and PSRLG Lists

Two basic use-cases of SRLG and PSRLG lists are the resilient routing [29, 30], and determining service availabilities depicted in Fig. 1.1 and 1.2, respectively.

In Fig. 1.1a, we can see a pair of imagined primary and backup paths stretching between Central Europe and California. By demanding a distance of several hundred `km` between the two paths (except their endpoints), we ensure they have a negligible probability of failing together. Fig. 1.1b depicts the state of the routers during Hurricane Sandy that was considered a severe disaster. In Fig. 1.1c, a maximal number (here, 7) of s-t paths are shown, such that there are no two paths that are hit at the same time by any position (outside of the yellow regions) of the red disk depicted in the bottom right corner. Here, in the input, instead of storing the possible disasters and the geometric embedding of the network, one can simply use a list of SRLGs indicating the link sets that can be hit by the same disaster: if path p_1 goes through SRLG S, then path p_2 is forbidden to do so.

The example depicted in Fig. 1.2 underlines difficulty of estimating service availabilities. There, user U reaches her data either in cloud C_1 or in cloud C_2. At the next disaster, the connections to C_1 and C_2 may fail in regions V_1 and V_2, respectively, with an equal chance of $P(V_1) = P(V_2) = 0.001$. If V_1 and V_2 are far from each other (as in Fig. 1.2a), we may suppose the connections fail independently, meaning an unavailability of $P(V_1) \cdot P(V_2) = 0.000001$ of the cloud. If V_1 and V_2 are at the same place (same bridge, valley, etc., Fig. 1.2b), the unavailability of the cloud will be $P(V_1) = P(V_2) = 0.001$. If V_1 and V_2 are 'close' to each other, but not in the same place, the availability of the cloud under the next disaster is difficult to estimate (that could be anything between three- to six-nines). Easing the

[2] An SRLG is a set of links that are considered to have a significant chance of failing together. First introduced in [28].

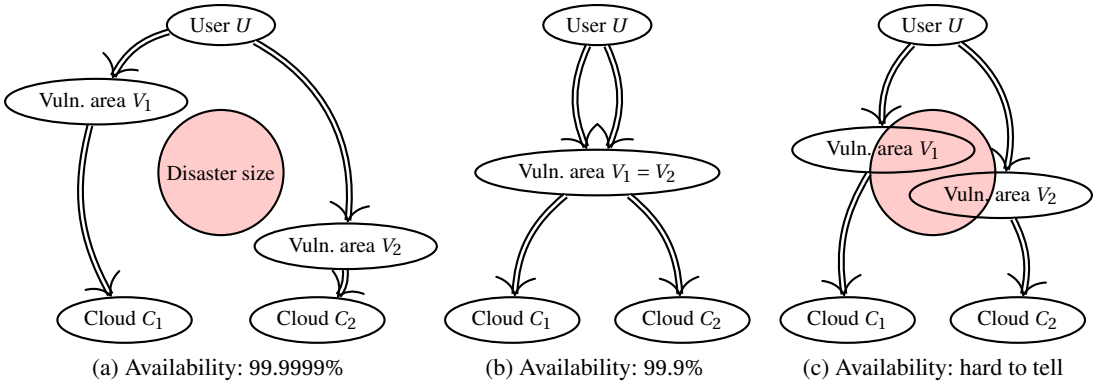


(a) (b)

(c)

Figure 1.1: (a) To avoid most disasters, ensuring several hundred km distance between the primary and the backup paths is enough. (b) The status of the routers during Hurricane Sandy, 2012. Most of the routers in NYC are not functioning, Boston also has problems. (c) For disaster-disjoint routing, storing the disasters and the geometric embedding of the network can be replaced by a short list of SRLGs indicating the link sets that can be hit by the same disaster. Picture credits to [31, 32, 33].

service availability queries demands the investigation of probabilistic extension of the SRLGs, and designing a realistic probabilistic regional failure model.



(a) Availability: 99.9999% (b) Availability: 99.9% (c) Availability: hard to tell

Figure 1.2: User U reaches his data either in cloud C_1 or in cloud C_2. At the next disaster, the connections to C_1 and C_2 may fail in regions V_1 and V_2, respectively, with an equal chance of $P(V_1) = P(V_2) = 0.001$. If V_1 and V_2 may be hit by the same disaster, but are not co-located, the cloud availability under the next disaster is hard to estimate.

1.2 Problem Statement

In this Thesis, I study both the deterministic and probabilistic versions of the problem of representing the effect of regional disasters on telecommunication networks. In the first part, I purely focus on how to define and enumerate the most lifelike (deterministic) SRLGs that cover all types of disasters. Fig. 1.3 depicts the most natural strategies for guaranteeing a level of separation between the primary and the backup path in the absence of the simultaneous presence of both a precise knowledge of the physical positions of the network elements and expertise on possible disasters.

Without any requirements, there might be *no separation* at all between these paths. A common practice is to ensure *link-disjointness* on the paths via enumerating all the single link failures as SRLGs. Compared to this, *node-disjointness* (except for the source and destination nodes s and t) ensures resiliency to any single element failure. An SRLG list providing node-disjointness consists of link sets incident to each network node.

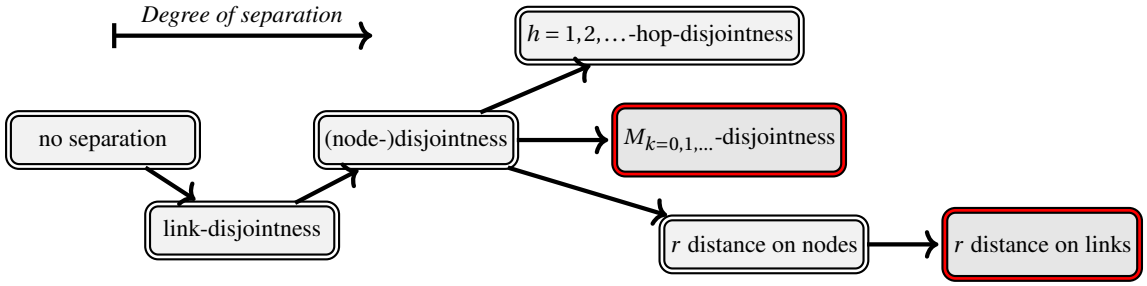


Figure 1.3: Strategies for separating the primary and backup paths in increasing strength (the more right the better the separation is)

To enhance the preparedness granted by node-disjointness, one has to leverage some background information on the geographical embedding of the network. Typically, communication networks have few edge crossings, and links are a few hundred kilometers long. Thus it makes sense to grant a given h hops distance between the primary and backup paths[3]. For this, one may list the links in the vicinity of every network link or node as SRLGs[4]. Unfortunately, a distance of h hops does not necessarily protect the failure of links crossing the same bridge or a bunch of close nodes.

Knowing the exact geographic embedding of the network topology solves this issue: supposing that a disaster may damage the network equipment within a radius r around its epicenter (and the rest of the network is left intact), one only has to list all the maximal link sets which can be hit by a circular disk with radius $r/2$ in a list $M_{r/2}$. Here the challenge is giving fast polynomial algorithms for

[3] Vendor specification of core network equipment suggests to ensure that the primary and backup paths assigned to a connection are edge or node disjoint (e.g., Huawei [34, Section 4.5.4], Alcatel-Lucent [35, pages 46-50], Cisco Systems[36, Chapter 19], Juniper [37, Chapter 3], Infinera[38]). With node-disjointness, operators ensure that the distance between the nodes of the primary and backup paths (except at the terminal nodes) are at least 1-hop-distance from each other.

[4] To ensure an odd number of hops, for every node v, $M_{h=2k-1}$ contains the edges of a tree of shortest paths to v from the nodes not further from v than k hops. Similarly, for every link $e = \{u, v\}$, $M_{h=2k}$ contains the edges of a tree of shortest paths to e from the nodes not further from u or v than k hops.

determining $M_{r/2}$ and showing that $M_{r/2}$ has a manageable size so that we can provide r-disjointness for large network topologies too. Chapter 5 (and Thesis 1) is devoted for this issue.

In many cases, one has only a rough idea of the physical embedding of the network, e.g., when the topology is rented from a Physical Infrastructure Provider [39]. In other words, they have a schematic map of the network, where the scale is not necessarily preserved over the area, and routes of links are only known to be within certain areas. In such circumstances, one can provide a separation which is weaker compared to M_r, but still better than relying only on hop-count: in a list M_k, one can gather the maximal link sets which can be hit by a circular disk hitting k nodes. In Chapter 6 (and Thesis 2), I provide a model to handle this case together with theoretical and experimental upper bounds on the size and construction time of M_k.

Regarding the prior state of the art, there was no PSRLG model, which would take into count that link failures are not independent when a disaster happens. Also, they did not represent the possible disasters as accurately as possible. In the second part of my Thesis (Chapter 7, Thesis 3), I aimed to define a realistic and applicable Probabilistic PSRLG failure model, which takes into count the failures correlation. In the evaluation, we use a seismic hazard representation, which preserves more information on possible future earthquakes than usual hazard maps.

1.3 Overview of this Thesis

At first, in Chapter 2, the (P)SRLG problems studied in this Thesis are formally introduced.

Chapter 3 presents an overview of the state-of-the-art for (P)SRLG modeling and enumeration. In Chapter 4 the necessary algorithmic background for this Thesis is given.

As discussed in the former Subsection, Chapters 5 and 6 present my studies on SRLG modeling an enumeration in case of precise and schematic maps of the network topology given as input, respectively. Chapter 7 presents a stochastic model for PSRLG enumeration. The evaluation of the model is based on real-world seismic data. Finally, Chapter 8 concludes the Thesis.

1.4 Contributions of this Thesis

The contributions of this Thesis are two-fold. Firstly, it offers provably short lists of SRLGs covering all the failures caused by regional disasters. For this, both a model where the exact geographical embedding of the network is known and another model where only a schematic map of the topology is available is given. Fast polynomial algorithms calculating the above lists are offered.

On the other hand, this Thesis provides a model for PSRLG enumeration that produces realistic failure probabilities, the computed data structure can be stored in provably small space in case of circular disasters, and it handles the correlation of link failures better than the prior state-of-art.

Chapter 2

Formal Problem Statement

In the current chapter, I give a general definition of the models and terms covering the remainder of the Thesis. However, since the motivations and models of a different part of this Thesis slightly differ, I will restate or detail some of the notions for the sake of a more fluent first reading of the Thesis. Throughout the dissertation, it will be assumed that basic arithmetic functions $(+, -, \times, /, \sqrt{\ })$ have constant computational complexity.

2.1 Definition of (Probabilistic) Shared Risk Link Groups

When several network elements may fail together as a result of a single event, they are often characterized by *Shared Risk Group*s (SRGs). Each SRG has a corresponding failure event (or events); when such an event occurs, all elements in the SRG fail together. Specifically, the communication network is modeled as a graph $G = (V, E)$, whose vertices are routers, PoPs[1], optical cross-connects (OXC), and users, while the edges are communication links (mostly optical fibers). SRGs are then defined as subgraphs (V', E'), where $V' \subseteq V$ and $E' \subseteq E'$.

In many cases, it is sufficient to consider only *links* in SRGs, and in this case, these groups are called *Shared Risk Link Groups* (SRLGs). For example, an SRLG may contain one edge (to capture a single-link failure) or all edges that touch one vertex (to capture a single-node failure). SRLGs may be more complex and represent simultaneous failures of multiple network elements. In particular, in this chapter, we focus on geographically-correlated failures in which links within a specific region fail together.

A set M of SRLGs can be used as an input to network design and network recovery/protection mechanisms to ensure these mechanisms withstand the failures corresponding to these SRLGs. For example, to ensure connectivity between a specific pair of nodes, protection mechanisms may construct two edge-disjoint paths when $M = \{\{e\} | e \in E\}$, two node-disjoint paths when $M = \{\{(u, v) \in$

[1] A point of presence (PoP) is an artificial demarcation point or interface point between communicating entities.

$E\}|v \in V\}$, or two paths that do not traverse the same geographical region when M corresponds to all sets of links that are physically close-by.

The following definition captures the notion of SRLG introduced by regional failures, such as a natural disaster or an attack. For ease of presentation, we will call these failure events *disasters*, regardless of their cause.

Definition 2.1.1 (SRLG). *A set of links $S \subseteq E$ is an SRLG if we may assume there will be a disaster that can cause all edges in S to fail together. If the disaster can be characterized by a bounded geographical area D, and S is the set of edges that intersect with D, then S is called the* regional SRLG that represents *D, and is denoted by $S = SRLG(D)$. If D is a circular disk, we call $SRLG(D)$ a* circular *SRLG.*

Circular SRLGs, which are the most common in literature, can also be characterized by the failure epicentre $p \in \mathbb{R}^2$ and the failure radius $r \in \mathbb{R}$. In this case $S = \{e \in E|d(e,p) \leq r\}$, where $d(e,p)$ is the Euclidean distance between edge e and point p.

The likelihood of a disaster to occur is not the same at all points of the plane. For example, earthquakes are more likely to occur in rupture zones than in other places, and regions with lower altitudes are more likely to suffer from floods. Thus, the probability of an event occurring is essential. This probability is sometimes given in the form of an *epicenter distribution map*, which gives for each location $p \in \mathbb{R}^2$, the probability that a disaster happened with epicenter p. Moreover, the size (or radius) of the disaster can also be a random variable (e.g., earthquakes with a larger magnitude are less likely to happen than earthquakes with smaller magnitude, even if their epicenters are the same). Thus, it is customary to consider a set $\mathscr{D}$ of disasters D (that can be of infinite size), and attach a probabilistic measure to this set. For simplicity, let us assume that $\mathscr{D}$ is finite, and let $p_D = \Pr[D \in \mathscr{D} \text{ occurs}]$[2]. We note that an SRLG S can represent more than one disaster in $\mathscr{D}$; thus, we denote by the $\text{support}(S) = \{D \in \mathscr{D}|S = SRLG(D)\}$.

Definitions 2.1.2 and 2.1.3 capture the probabilistic nature of disasters and their effect on SRLGs. An FP (Def. 2.1.2) tells the probability that the failed link set will be *exactly* S, while a CFP (Def. 2.1.3) tells the probability that *at least* S will fail:

Definition 2.1.2 (FP). *Given a set $\mathscr{D}$ of disasters D, a probability p_D for each disaster in $\mathscr{D}$, and a link set $S \subseteq E$, the* Link Failure State Probability (FP) *of S is $FP(S) = \sum_{D \in support(S)} p_D$. We note that if a disaster in $support(S)$ actually occurs, then all links in S fail. For a graph G, the collection of all $(S, FP(S))$ pairs is denoted by $FP[G]$.*

Definition 2.1.3 (CFP). *Given a set $\mathscr{D}$ of disasters D, a probability p_D for each disaster in $\mathscr{D}$, and a link set $S \subseteq E$, the* Cumulative Link Failure Probability (CFP) *of S is $CFP(S) = \sum_{T \supseteq S} \sum_{D \in support(T)} p_D$.*

[2]For infinite sets, one can use the corresponding integrals or use discretization and consider only a finite number of sets, albeit with a small error.

We note that if a disaster in $\bigcup_{T \supseteq S}$ support(T) occurs, then all links in S fail. For a graph G, the collection of all (S, CFP(S)) pairs is denoted by CFP[G].

For simplicity, wherever it does not cause any confusion, FP(S) and CFP(S) will also refer to the set-probability couples $(S, \mathrm{FP}(S))$ and $(S, \mathrm{CFP}(S))$, respectively.

Clearly FPs and CFPs are closely interconnected. In a sense, FPs are like probability density functions (PDFs), while CFPs are like their cumulative distribution functions (CDFs).

2.2 Physical Embedding of the Network Topology

As written before, the network is modeled as a geometric graph $G(V, E)$, whose vertices are routers, PoPs, OXCs, and users, while the edges are communication links (mostly optical fibers).

For simplicity, in most cases, G is considered to be embedded in the plane $\mathbb{R}^2$, but in some cases, for accuracy, I use its spherical representation. When not clear out of context, I indicate the type of geometry with parameter $g \in \{p, s\}$, p and s standing for 'planar' and 'spherical', respectively. For simplicity, in the remainder of this Thesis, the terms used in the Euclidean plane might refer to their counterpart in spherical geometry too. That is, 'line segment', 'polygonal chain' (or 'polyline'), and 'containing polygon', will refer to 'geodesic','chain of geodesics', and 'containing closed chain of geodesics' in case of spherical geometry.

Another issue is the geometric embedding of the network links. In this Thesis, I use three different models: 1) links are line segments (sometimes called intervals) between their endpoints, 2) links are polygonal chains between their endpoints, and 3) the exact route of each link $e \in E$ is not known, but contained in a polygonal region p^e (and e is considered to fail if the disaster hits p^e). Mathematically, 3) is more general than 2), which is more general than 1). I use parameter γ to indicate the maximum number of line segments a link or its containing region stands of.

2.3 Disaster Families and Related Induced Failures

Throughout the SRLG enumeration part of my Thesis, I assume that no detailed information is available on the disasters, and I overestimate the disaster areas by circular disks. In both cases of geometries $g \in \{p, s\}$ in which the topology can be embedded, let the set of all circular disks in g be denoted by $\mathscr{C}$.

2.3.1 SRLG Enumeration: Disasters with Limited Size

If the embedding of the network is known precisely (as in Chapter 5), I am interested in the maximal link sets, which can be hit by a $c \in \mathscr{C}$ with a given radius. For this, let $\mathscr{C}_r := \{c \in \mathscr{C} | \text{radius of } c = r\}$, and M_r be the set of SRLGs that can be hit by a $c \in \mathscr{C}_r$. I distinguish lists M_r of the planar and spherical embedding as M_r^p and M_r^s, respectively.

2.3.2 SRLG Enumeration: Disasters for Schematic Embeddings

If only a schematic map of the network is known (as in Chapter 6), I am interested in the maximal link sets, which can be hit by a $c \in \mathscr{C}$ which hits a previously fixed k number of nodes. For this, let $\mathscr{C}_k := \{c \in \mathscr{C} | c \text{ hits } k \text{ nodes }\}$, and M_k be the set of SRLGs that can be hit by a $c \in \mathscr{C}_k$.

2.3.3 PSRLG Failure Modeling

In Chapter 7, I aim to define a generic PSRLG model. Thus, in the model, any disaster shapes are allowed, the disaster does not even need to be a connected subset of the plane/sphere. However, I give theoretical upper bounds on the number of FPs and CFPs supposing that the disasters have a shape of a circular disk (in any given Lp metric). In the simulations, earthquakes also are considered to destroy the network in circular regions.

2.4 General Practices for SRLG Enumeration

As the size of SRLG list $\mathscr{S}$ determines the run-time and complexity of the mechanisms that use it, an important goal is to keep $\mathscr{S}$ as small as possible. For example, when two sets S_1, S_2 are in $\mathscr{S}$ and $S_1 \subseteq S_2$, it is sufficient to include only S_2 in $\mathscr{S}$; omitting S_1 from $\mathscr{S}$ usually does not affect the outcome of the underlying mechanisms[3]. This is due to the monotonicity of network design/recovery mechanisms, where we say, a mechanism is monotone if for any S_1, S_2 such that $S_1 \subseteq S_2$, the actions the mechanism takes in response to S_1 is a subset of the actions it takes in response to S_2.

Moreover, some works use *over-approximation* to reduce the size of $\mathscr{S}$: $\mathscr{S}'$ overapproximates $\mathscr{S}$ is for every $S \in \mathscr{S}$ there exists $S' \in \mathscr{S}'$ such that $S \subseteq S'$. As an over-approximation, instead of including two sets S_1, S_2, one can include a single set $S_1 \cup S_2$ (this is especially appealing if $S_1 \cap S_2$ is of non-negligible size); such over-approximation, however, can degrade the outcome of the underlying mechanisms. For regional SRLGs, over-approximation is achieved by taking a larger failure region. The most common practice is to take a simpler shape that completely contains the original failure region, e.g., *circular disks*, or *fixed shape* bounded by segments and arcs. The presented algorithms are conservative both with the number of listed SRLGs and with the degree of over-approximation in the case of different classes of realistic inputs.

Another widespread practice is to assume that *in the investigated time period, there will be at most one disaster*. If one can enumerate the set $\mathscr{S}$ of SRLGs of single disasters, it is straightforward to compute SRLGs of multiple disaster events. For example, if two disaster can happen simultaneously, one might look at $\mathscr{S}' = \{S_1 \cup S_2 | S_1, S_2 \in \mathscr{S}\}$.

[3]This is true for communication networks, but not for networks in which there is no monotonicity in failures. When attaching probability to the SRLGs, this no longer holds.

2.5 Problem Statement

The somewhat formal statements of the problems I am investigating in this Thesis are the following.

For SRLG enumeration, the problem is defining a model and related algorithms so that, based on the resulting SRLG list, the operators should prepare the network for only a small number of possible regional failure events that cover all the possible disasters. In some cases, the precise embedding of the network is given as part of the input, while in others, only a schematic map is known. With the definitions of this chapter, these problems translate to the following questions:

- Is there a fast polynomial algorithm for enumerating M_r^p? Can tight theoretical upper bounds be given on $|M_r^p|$? Is there an algorithm that, in practice (i.e., for real network topologies), calculates M_r^s efficiently? Is there a significant difference between M_r^p and M_r^s in practice, i.e., in case of real network topologies, and realistic disaster radii ($\leq\sim$ 500km)?
- Can a justifiable model be specified for regional failures if only a schematic map of the network topology is given? Is there a polynomial algorithm for calculating M_k, and related theoretical upper bound on $|M_k|$?

For PSRLG enumeration, the problem is defining a model that yields PSRLG lists that can serve, e.g., service availability queries with great accuracy. This translates to the following questions:

- For PSRLG enumeration, can a failure model be defined such that 1) it produces realistic failure probabilities, 2) correctly captures link failure correlation patterns, 3) the size of the output structures is manageable in practice? Can low theoretical upper bounds be given on the number of the PSRLGs needed to describe the random effect of the next disaster?

2.6 Model Extensions

2.6.1 Segment Link Representation to Polylines, Polylines to Containing Polygons

The network embedding model varies in our papers. In some of them, links are considered as line segments, in others, they are polygonal chains, or even only known to lie in a polygonal region.

Suppose we have an algorithm A_s for the line segment link model. The polygonal chain case can also be handled in polynomial time based on A_s via splitting the polygonal chains up into line segments, running A_s for the resulting problem instance, merging the line segments of each polygonal chain, and finally, filtering out the non-maximal sets.

Now suppose we have an algorithm A_c for the polygonal chain link model. Then, the containing polygon can also be handled in polynomial time using a slight modification of A_c: if in case of a link

e, a disaster d is not hitting the boundary of p^e, one has to decide if d is in the interior of p^e or in its exterior. Fortunately, this can be done in polynomial time (see, e.g., proof of Claim 6.6.1).

2.6.2 Different Link Types

Most of the optical backbone networks consist of multiple types of links, e.g. *aerial*, *buried* and *submarine*. In case of a disaster, these link types have different failure patterns, for example, in case of an earthquake, aerial cables fail in a different region than buried cables, while submarine cables tend to be cut at rupture zones. With this in mind, we can extend our model in the following way. Let L be the set of different link types. For disaster D, let $r(D, l)$ denote the area where links with type l fail in case of D.

Note that I have not paid particular attention to the algorithmic side of more sophisticated failure models in this Thesis. One approach to constructing (P)SRLG lists in the case of these models is taking a sufficiently fine discretization of the original problem [J2, J5].

2.6.3 Mixed Link Types

Different parts of a link e in the input network topology may have different types, e.g., there is a link that is mainly buried, but crosses a river above the water. Such a link can be divided into sections with homogeneous types and fails if one of its section fails. More formally, each link $e \in E$ is partitioned to sections $e^1, \ldots, e^K$ with types $l^1, \ldots, l^K$, respectively. For disaster D, section e^i fails if it has a common point with $r(D, l^i)$, and link e fails if at least one of its sections fails.

2.6.4 Nodes Also Considered Vulnerable

Network nodes have different failure patterns than links, and their probabilistic failures can be represented by (P)SRLGs as follows. For a node $v \in V$ that can fail, the edges incident to v have mixed link types, and in a small vicinity of v are considered to have a type $l_v \in L$ specific to the node, i.e., that tiny parts of the links fail exactly then when the node would have failed. This approach translates to (P)SRLGs as following in the end. We consider that the set of links S incident to v fails because the disaster hits every $l \in S$ or node v.

Chapter 3

Related Work

3.1 Charting the Landscape of (P)SRLG Enumerating Problems

To have a better overview of the problem versions tackled by both other researchers and our group, in the following, a charting of the (PSRLG) enumerating problems is given based on the input data quality/precision.

Informally speaking, the most important input information parts are the 1) geometric embedding of the network, and the 2) (probabilistic) disaster effects. Unfortunately, in practice, it is far not obvious that this information is available with high precision. As depicted in Fig. 3.1, we might distinguish three levels of information quality both on the geographic embedding and on the disaster effects, and classify the offered (P)SRLG approaches according to these. I briefly depict and reference all the related problems.

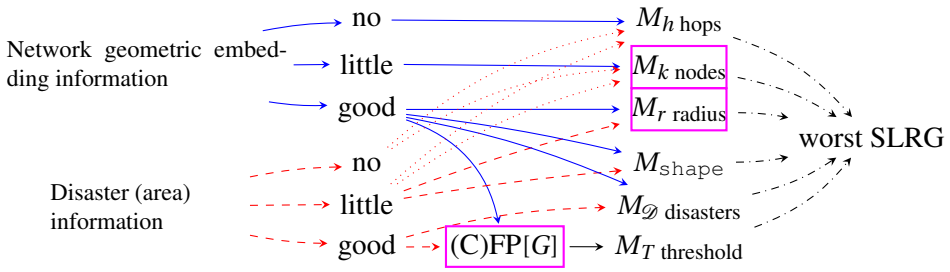


Figure 3.1: A mind map of SRLG and PSRLG problems related to the quality of input data. For a graph G, (C)FP[G] stores PSRLGs, while lists M_* consists of SRLGs. Problems studied in this Thesis ($M_{k \text{ nodes}}, M_{r \text{ radius}}$, (C)FP[$G$]) are highlighted with purple rectangles.

In case of the geographic embedding, we may encounter the following cases:

- *no information on the embedding*: this can happen, e.g., when dealing with a network rented from a Physical Infrastructure Provider. In this case, one may list the links in the

h-neighborhoods of every node or link as an SRLG for a given h. These lists M_h will ensure some hops distance between primary and the backup paths, hopefully translating to a decent physical distance (see [B1, Sec. 1.3.2A]).

- *little information on the embedding*: in many cases, one has a schematic map of the physical topology. Using these, we can compute SRLG list M_k of links sets hit by disks hitting k nodes. Compared to M_h, M_k also protects failures of close nodes and parallel (close) links (see Chapter 6).

- *good information on embedding*: if a precise map of the network topology is part of the input, one can leverage his knowledge on disaster effects too:

  – *no information on disasters:* one can suppose a disaster will do harm only within a disk of radius r around its epicenter, and compute the list M_r of maximal link sets hit by disks with radius r (see Chapter 5). Alternatively, one may assume that the disaster has a fixed shape, e.g., an equilateral triangle of any orientation, and calculate list $M_{\texttt{shape}}$ of maximal link failures caused by this shape [40]. Another possibility is that a set $\mathscr{D}$ of disaster areas is given[1], and computes the list $M_{\mathscr{D}}$ of maximal SRLGs caused by these disasters (see, e.g., a non-probabilistic version of [41]).

  – *good information on disasters:* having detailed probabilistic disaster data allows us to compute PSRLG lists. Here the challenge is to create a model that correctly captures the joint failure probabilities of network elements while producing an output of affordable size (see Chapter 7 or [17]).

Having a list of PSRLGs enables collecting those SRLGs that have a failure probability above a threshold T. From among these, one can collect the maximals in a list M_T [C5].

We note that in the natural condition when the vulnerability metric or a protection mechanism is monotone[2], the worst SRLG fulfilling a given criteria c (e.g., SRLGs that can be hit by circular disks with a range r or hitting k nodes) will be part of the set of exclusion-wise maximal SRLGs fulfilling c. Thus a worst SRLG can be found by simply searching for the worst SRLG in the list of maximal SRLGs fulfilling c (e.g., in M_r, M_k)[B1]. This way, the results presented in this Thesis are firm base ground for solving a whole family of related problems.

Computational complexity and precision are two additional criteria along which the studies of the field can be separated. In case of complexity, clearly, polynomial algorithms outperform their super-polynomial counterparts. Sometimes, to achieve a manageable problem space, studies take

[1] Where in the interior of the disaster area, everything is damaged, while in the exterior, no failure happens

[2] We say vulnerability metric or protection mechanism μ is *monotone*, if, according to μ, for any link set $E_1 \subseteq E_2$, the failure of E_2 is *worse* than the failure of E_1.

discretized input data, e.g., they take a sufficiently fine grid over the topology and assume the disasters can have their epicenter in these grid points. The discretization yields some imprecision, but given our knowledge, e.g., on seismic hazard, this imprecision is often affordable [J5]. In the following, I will refer to algorithms using this kind of discretization as 'approximate' algorithms.

3.2 (P)SRLG Enumeration or Finding Worst (P)SRLG

Tables 3.1 and 3.2 give an overview on studys dedicated to different versions SRLG and PSRLG enumeration, respectively. Note that [B1] gives an overview on these papers.

3.2.1 SRLG Enumeration

Table 3.1 lists SRLG enumerating papers. I marked papers used as a material for this Thesis with ✓in the second column (in rows 2,3,5). Algorithmically, determining M_h is not challenging (row 1). Once we have a disaster set $\mathscr{D}$, enumerating $M_{\mathscr{D}}$ is straightforward too (row 4, see e.g., [B1]).

Table 3.1: Papers enumerating regional SRLGs. While the rest of the papers consider deterministic disaster scenarios, in [C8, C5] SRLGs are obtained from PSRLG lists.

Paper	In this Thesis	Goal	Geometric info.		Assumptions		Algorithms		
			Physical network	Planar/ Spherical	Disaster shape	Single disaster	Precise/ approximate	Poly- nomial	Para- metrized
Vass et al. [B1]/1.3.2A	-	SRLG list	no	plane	-	✓	precise	✓	✗
Vass et al. [C11, J1]	✓	SRLG list	poor	plane	circular	✓	precise	✓	✓
Tapolcai et al. [J4, C10]	✓	SRLG list	good	plane	circular	✓	precise	✓	✓
Vass et al. [B1]/1.3.2E	-	SRLG list	good	any	set of known disasters	✓	precise	✓	✗
Vass et al. [C7, J5]	✓	SRLG list	any	plane+ shpere	bounded by segments+arcs	✓	precise+ approximate	✓	✓
Iqbal et al. [42]	✗	SRLG list	good	plane	-	✓	precise	✗	✗
Trajanovski et al. [19, 40]	✗	most vulne- rable point	good	plane	any	✓	precise	✓	✗
Neumayer et al. [10, 43]	✗	most vulne- rable point	good	plane	circular or line segment	✓	precise	✗	✗
Pašić et al. [C8, C5, J2]	✗	SRLG list	good	plane	any	✓	approximate	✓	✗

(a) There are 3 SRLGs each are a pair of links. (b) There is a single link set with three links

Figure 3.2: Example SRLGs according to (a) Def. 2.1.1, and (b) [42].

A problem closely related to (regional) SRLGs is investigated in [42] (row 6 of Table 3.1). The paper proposed to call a pair of fibers *spatially-close* if their distance is at most r', i.e., they can be covered with a circular disk of radius $\frac{r'}{2}$. They propose to define SRLGs as sets of fibers where any pair of fiber are spatially-close, in other words, any pair of fibers can be covered with a circular disk of radius $\frac{r'}{2}$. Unfortunately, [42] ends up at an NP-complete problem while grouping all fibers that are spatially close to each other, such that the number of distinct link sets is minimized. Furthermore, the resulting link sets may not be even possible to hit by the same disaster (i.e., they are not necessarily SRLGs according to Def. 2.1.1)[3]. See Fig. 3.2 highlighting the difference between this model and the one presented in [42].

In [19, 40], polynomial algorithms are given for finding the most vulnerable point (called the critical region) of the network in case of disasters with a fixed polygonal or elliptic shape and size. Although some arguments of the paper are inaccurate or even conflicting, its reasonings could be healed.

With somewhat different motivations, similar computational geometric ideas were used in papers focusing on the most vulnerable points (worst SRLGs) of physical infrastructure (communication networks or power grids [44]) to regional failures or attacks. In this Thesis, our objective is more general as we want to enumerate all (the maximal) candidate failures instead of searching for the most vulnerable according to some metric. In these works, the network is embedded in the Euclidean plane and the failures are modeled either as a disk around its epicenter (circular) [10, 17], line segments [10], ellipse [19] or polygons (rectangle, square, or equilateral triangle) [19]. Technically these papers also list the candidate failures (SRLGs) and evaluate the vulnerability metric of the residual network in case of each candidate failure.

For (implicit) worst SRLG computation, the following vulnerability metrics were investigated. (1) the point with the maximum number of affected links [10, 17], which is ρ_r. (2) the point with the maximum average two terminal reliability between every node-pair [10, 17, 19]. Here the max-flow algorithm runs $O(nm)$ [45] which we need to run for $O(n)$ in practice. (3) the point with the maximum average all-terminal reliability [8], which allows the identification of network areas that

[3]Note that based on $M_{r=r'/2}$, treating the corner points as degree 2 nodes, the grouping of the spatially-close fiber segments can be directly computed.

can disconnect any component in the network. (4) the point with the maximum average value of the maximum flow between a given pair of nodes [10]. (5) the point with maximal average shortest path length between every pair of nodes [8]. (6) the point with maximal average shortest path length between every pair of nodes [8, 19], (7) survivability as a measure of the weighted spectrum based on the eigenvalues of the normalized Laplacian of a graph [8], (8) network criticality, which is determined from the trace of the inverse of the Laplacian matrix and can be related to the node and link betweenness [8], (9) momentary chance of cable cut caused by a landslide in case of heavy rain [46, 47].

It is a natural idea to list in a container M_T the (maximal) link sets, which have a probability of failing together higher than a given threshold T (like in [C5] or [48]). Obviously, for this, as an intermediate step, one has to generate a set of probabilistic SRLGs. More precisely, CFP is the most useful structure in this context, since, by definition, for a link set S, the Cumulative Failure Probability CFP(S) is the probability that at least the links of S will fail. The advantage of this approach is that SRLG lists can be generated based on sophisticated objectives.

3.2.2 PSRLG Enumeration

Table 3.2 depicts papers dedicated to PSRLG enumeration. [41] assumes there is a set $\mathscr{D}$ of disasters given, each disaster $D \in \mathscr{D}$ is associated with a probability of happening, and based on this, calculates link failure state probabilities FP(S) straightforward, naming them simply as PSRLGs. Although papers [17, 49] aim only to find the most vulnerable point of the network, their tools are suitable for PSRLG enumeration. Our papers [C9, C4, J3] are, up to my knowledge, the first to explicitly take into count the correlated nature of link failures in the presence of a disaster. Also, alongside our book chapter [B1], these works are the first to offer a unified terminology on PSRLGs.

Table 3.2: Papers enumerating regional PSRLGs.

Paper	In this Thesis	Goal	Correlated link failures inside the disaster	Natural disaster / attack
Oostenbrink et al. [41]	✗	FP list	(✓)	-
Tapolcai,Valentini, Vass et al. [C9, C4, J3]	✓	FP list + CFP list	✓	natural disaster (earthquake)
Agarwal et al. [17, 49]	✗	most vulnerable point	✗	attack

The tools presented in papers cited in the previous Subsection for determining a worst SRLG can be used for PSRLG enumeration (like in the case of [49], and [17]). Additional related papers

implicitly listing (P)SRLGs addressed specific sub-problems in network planning, like finding the most vulnerable part(s) of the network [8, 50], studying the impact on the network of a randomly placed disaster [51, 52], designing a network and its services with disaster resiliency in mind [53], and (re)routing of connections to minimize service impact due to a disaster [54]. Some work has considered probabilities, either in the context of a disaster having a certain probability of disconnecting a link, e.g., [17], or in the context of only having partial (probabilistic) information on the geographical layout of a network, e.g., [55].

While the above-mentioned papers considered geographically correlated failures, a common property of the targeted sub-problems is to search for the location(s) where a disaster will cause the maximum *expected* damage to the network. This is a crude averaging process that is unable to exhibit correlations among many important failure events. The problem of precisely and quickly calculating the correlations between link failures to conduct a more thorough network vulnerability assessment had been insufficiently addressed in the past.

3.3 (P)SRLGs as Input

There are several papers refraining from making up their own failure models and, instead, taking lists of (P)SRLGs as input. A good example of this is [56], providing diverse routing algorithms and being the first paper to introduce PSRLGs. In the field of fault-tolerant virtual network mappings, (k)-content connectivity is calculated based on SRGs in [57, 58, 59]. Papers [60, 61, 62] offer SRLG-disjoint routing. Other examples are [2, 5, 11, 63].

There are a number of studies that could have taken (P)SRLGs as input, especially if their failure modeling is less realistic than the state-of-the-art. Some of the problems tackled in these studies are the following. SRLG disjoint routing [64], content connectivity against double-link failures [65], logical survivable topologies against multiple failures [66], flow availability in two-layer networks [67], tunable protection for single and dual link failures, respectively [68, 69], evaluating geographic vulnerabilities of multilayer networks [18, 70], network virtualization design for regional disaster resiliency [6], increasing availability between two nodes after disaster [52], improving connectivity resilience through third-party networks[71], survivable network design [53, 72].

Some surveys, summaries or tutorials tackling (P)SRLGs are [9, 73, 74, 75, B1]. A recent comprehensive guide book for the broader topic of disaster resilience of communication networks is [76] (that also includes our chapter [B1]).

3.4 Computational Geometry and Seismology

The work presented in the Thesis heavily relies on computational geometry. Books [77] and [78] cover most of the used background of this field. In [77], the concept of sweep line algorithms is discussed in detail. [78] is a survey on Voronoi diagrams and Delaunay triangulations, two related classes of geometric proximity graphs that are in great service of regional failure modeling.

Some of our advanced questions in this field are answered in recent papers. These are like the edge count of k-Voronoi diagrams for line segments [79], or construction time and edge number of higher-order Delaunay graphs (for a point set) [80, 81]. Some of the computational geometric tools used in our research were developed in our papers, like the sweep-disk algorithm for Thesis 2.

More related topics are Stereographic projection [82] combined with the the Apollonius problems [83], or the smallest intersecting ball problem [84], [85], which has its origins in the classical 19th-century problem of Sylvester [86] about the smallest enclosing circle for a given set of points in the plane.

The other related field worth mentioning is seismology, however, I only use it in Subsec. 7.6.1, to represent the seismic hazard less blurry than the usual hazard maps. The approach presented in Subsec. 7.6.1 for transforming the raw earthquake catalogs to earthquake activity rate maps is essentially the same as the one used to create the SHARE European Earthquake Catalogue [87]. Additional references will be provided in the mentioned subsection.

Chapter 4

Algorithmic Background

4.1 The Big O Notation

Let $F_{\mathbb{N}} = [\mathbb{N} \to \mathbb{N}]$ denote the space of all functions on natural numbers and let $f : \mathbb{N} \to \mathbb{N} \in F_N$ be a specific function. We employ the Landau symbol O to denote the following class of functions:

$$O(f) = \left\{ g \in F_{\mathbb{N}} \middle| \limsup_{n \to \infty} \frac{g(n)}{f(n)} < \infty \right\}.$$

Writing '$g = O(f)$' instead of '$g \in O(f)$' is also widespread. To bypass this issue, in the followings, mostly we will simply say 'g is $O(f)$'. If g is $O(f)$, we say f is $\Omega(g)$. Additionally, if both f is $O(g)$ and g is $O(f)$, we say f is $\Theta(g)$ and g is $\Theta(f)$ to denote that these are asymptotically equal.

4.2 Time and Space Complexity

The worst-case time complexity (or simply, complexity, or running time) of an algorithm is estimated by counting the maximum number of elementary operations performed by the algorithm, given an input of an arbitrary size s. In this Thesis, I prove theoretical bounds on running time and output size of algorithms that are *polynomial*, that is, there exists a constant c such that they perform at most $O(s^c)$ elementary operations. Sometimes I prove bounds that leverage some parameters depending on the input. E.g., by Cor. 5.1.14, the number $|M_r|$ of maximal link sets that can be hit by a disaster with radius r is $O((|V| + x)\rho_r)$, where x is the number of link crossings, and ρ_r is the maximum number of links such a disaster can hit *in the particular problem input*.

4.3 Computational Geometry: Sweep Line Algorithms

In our studies, we rely on multiple computational geometric tools. In the following, we briefly depict the principle of sweep line algorithms, one of those tools which are used repeatedly in later chapters.

In computational geometry, a sweep line algorithm or plane sweep algorithm is an algorithmic paradigm that uses a conceptual sweep line or sweep surface to solve various problems in Euclidean space. It is one of the key techniques in computational geometry.

In sweep line algorithms, it is imagined that a line is moved across the plane, keeping its orientation and stopping at some event points. Geometric operations are restricted to the immediate vicinity of the sweep line whenever it ends, and the complete solution is available once the line has passed over all objects. This principle is mostly manifested with the help of efficient data structures based on binary search trees or similar.

As an example, given a set E of line segments embedded in the Euclidean plane, the *line segment intersection problem* asks whether there exist two intersecting line segments in E. Clearly, the question can be answered in polynomial time (in the function of $|E|$), the challenge is to answer it 'as fast as possible'. A bound is given as follows:

Proposition 4.3.1 (Theorem 5 of [88]). *All k pairwise intersections among n segments in the plane can be computed in $O(n \log n + k)$ time. The running time is optimal. The storage requirement is $O(n + k)$. If so desired, the algorithm will compute the vertical map of the set of segments within the same time and space bounds.*

We can conclude that the fastest algorithm deciding whether there exist any line segment intersection runs in $\Theta(|E| \log |E|)$. A former and simpler algorithm called Bentley-Ottmann [77] also solves the problem. This algorithm, loosely speaking while swiping a vertical line from left to right maintains an ordered balanced binary search tree storing the line segments $e \in E$ according to their ordinates y_e at the current abscissa x. It has the following complexity:

Proposition 4.3.2 (Theorem 2.4 of [77]). *All intersection points of E, together with the segments giving the intersection, can be reported in $O((m + I) \log m)$ time and $O(m)$ space, where I is the number of intersection points*[1].

We can see that both [88] and the Bentley-Ottmann algorithm solve a more difficult problem: report the line segment intersections. This can be seen as a special case of the problem of reporting the maximal link sets, which can be hit by a circular disk (shaped disaster) of radius r, which we study in Sec. 5.1. In Chapter 6, we also rely on the concept of sweep surface algorithms. We note that although the algorithm of [88] is asymptotically faster than the Bentley-Ottmann in some settings, it looses its competitive edge if k is $\Omega(I \log n)$.

Another field where sweep line algorithms are useful is the construction of Voronoi diagrams [78] and Delaunay triangulations, two closely related classes of geometric proximity graphs that we will leverage throughout this Thesis.

[1] Note that in Prop. 4.3.2, in each intersecting point, an arbitrary number of line segments can intersect each other.

Chapter 5

Maximal SRLGs Induced by Disks with Radius r

5.1 Planar Regional Link Failures Caused by Disasters with Radius r

5.1.1 Problem Definition and Basic Results

The input is a real number $r \geq 0$ and an undirected connected graph $G = (V, E)$ embedded in the $2D$ plane, where V denotes the set of nodes and E the set of edges (which are also called links). Let $n := |V|$ and $m := |E|$. I assume $n \geq 3$. The edges of G are embedded as line segments, which I call *intervals* in the geometric proofs[1]. A *disk* with centre point p *hits* an edge e if its distance to p is at most r.

Definition 5.1.1. *A **regional failure** F is a non-empty subset of E, for which there exists a disk with radius r hitting every edge in F.*[2]

Note that the failure of node v is modeled as the failure of all edges incident to node v. Therefore listing the failed nodes beside listing failed edges would not give us additional information from the viewpoint of connectivity.

Definition 5.1.2. *Let F_r be the **set of regional failures** of a network for a given radius r.*

According to Def. 5.1.1, a subset of a regional failure is also a regional failure. Thus, F_r is a downward closed set minus the empty set.

[1] The case, when edges are considered to be embedded as polygonal chains between their endpoints consisting of at most a constant number of line segments, can also be handled in polynomial time based on the presented results via splitting the polygonal chains up into line segments, running the presented proposed algorithm (sketched in Table 5.2) for the resulting problem instance, merging the line segments of each polygonal chain, and finally, filtering out the non-maximal sets.

[2] Thus, what we call a regional failure is the worst-case outcome of a disaster damaging an area. F can be seen a compact representation of all of its subsets.

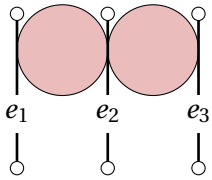


Figure 5.1: In the figure above, the solid circular disks are disasters with radius r, $d(e_1, e_2) = d(e_2, e_3) = 2r$, while $d(e_1, e_3) = 4r$. The set of regional failures is $F_r = \{\{e_1\}, \{e_2\}, \{e_3\}, \{e_1, e_2\}, \{e_2, e_3\}\}$. The set of maximal regional failures is $M_r = \{\{e_1, e_2\}, \{e_2, e_3\}\}$.

Recall the network can recover if an SRLG or a subset of links (and nodes) in the SRLG fail simultaneously. In other words, if a regional failure F is listed as an SRLG, then there is no need to list any subset of the links $F' \subsetneq F$ as a new SRLG. The goal is to define a set of SRLGs which covers every possible regional failure and which is of minimal size.

Definition 5.1.3. *Let $M_r \subseteq 2^E$ denote the set of SRLGs, for which*

$$M_r = \{F \quad \textit{is a regional failure and there is no regional failure } F' \textit{ such that } F' \supsetneq F\} \ . \quad (5.1)$$

In other words, the set of SRLGs M_r is a set of failures caused by disks with radius at most r in which none of the failures is contained in another. Figure 5.1 illustrates Definitions 5.1.1-5.1.3. Note that F_r is the set of regional failures, which is the downward closed extension of M_r minus the empty set. A family of sets from the power set of E in which none of the sets is contained in another is called an *antichain* (in the inclusion lattice over 2^E). This antichain is also sometimes called a Sperner system, independent system or a clutter. Note that, M_r is an antichain. Due to the minimality of SRLGs, the following holds.

Proposition 5.1.1. *For each SRLG $F \in M_r$, $F \subseteq E$, there is a circular disk c of radius r such that F is exactly the set of edges hit by c.*

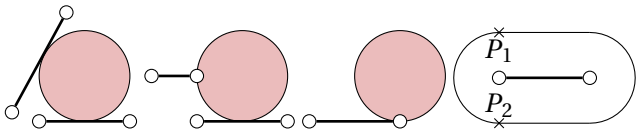
Let r be a tiny positive number. In this case, the list of possible regional failures consists of every *single link or node failure* and link crossings. In other words, this model is a generalization of the 'best practice.' The corresponding antichain can be the set of single node failures, i.e., $|M_r| = n + x$, where x is the number of edge crossings. Informally speaking, protecting node failures is sufficient to protect link failures as well.

In the following, the aim is to determine the set M_r. At first glance, it is not clear that the cardinality of M_r is 'small.' I will prove polynomial upper bounds on $|M_r|$.

To estimate the size of the SRLG list, let ρ_r denote the maximum number of edges a disk with radius r can hit in the plane, i.e., for every failure F caused by a disk with radius r, $|F| \leq \rho_r$. An observation is that if $\rho_r = O(\log n)$ then there is a polynomial blowup when switching from M_r to

Table 5.1: Table of symbols for Sec. 5.1

Notation	Meaning
	General
$G(V,E)$	the network modeled as an undirected connected geometric graph
n, m	number of nodes $\|V\| \geq 3$ and edges $\|E\|$, respectively
r	disaster range ($r \geq 0$)
F	regional failure, i.e.is a non-empty subset of E, for which there exists a disk with radius r hitting every edge in F
F_r	set of regional failures of a network for a given radius r
M_r	F is in M_r if it is is a regional failure and there is no regional failure F' such that $F' \supsetneq F$
c_F	smallest hitting disk of F, where a disk c is smaller than disk c', if c has a smaller radius than c', or if they have equal radius and the center point of c is lexicographically smaller than the center point of c'
X	set of points p which are not in V and there exist at least 2 non-parallel edges crossing each other in p
E_w	$:= \{e \in E \mid d(w,e) \leq 3r\}$; the edges in E_w are in sorted order with respect to the lexicographic ordering of their endpoints
V_e	$:= \{w \in V \cup X \mid d(e,w) \leq 3r\}$
$\mathscr{C}_{r,w}$	The set of the following disks: for $e, f \in E_w$, disks c of radius r (if exist) according to Thm. 5.1.2: either case a) applies if e and f are not parallel, and c intersects them in two different points, or case b) when c intersects e and f in two different points, one being an endpoint of e, or case c) when c touches e at an endpoint; moreover we require that formerly computed disks c have centers not farther than $2r$ from w.
$\mathscr{L}_{r,w}$	list of set of edges hit by an element of disk set $\mathscr{C}_{r,w}$
	Parameter
ρ_r	link density of the network, which is measured as the maximal number of links that could be hit by a circular disk shaped disaster of radius r
x	number of link crossings of the network G
μ	square mean of numbers v_e for all $e \in E$, where v_e is the number of $w \in V \cup X$ such that $d(w,e) \leq 3r$
ϕ_r	maximum number of nodes in the $3r$-neighborhood of a link of the input graph G



(a) (b) (c) (d) $N(e,r)$ of an edge e

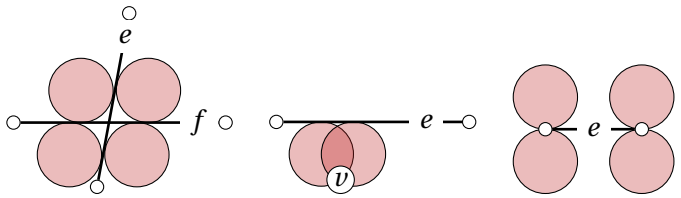
Figure 5.2: Case (a),(b) and (c) of Thm. 5.1.2 and the neighbourhood $N(e,r)$ of an edge e.

F_r, as $|F_r| \leq |M_r| 2^{\rho_r}$. M_r can be treated as a compact representation for F_r. It is also immediate that from F_r one can obtain M_r by $O(|F_r|^2)$ comparisons of subsets of E.

I say a disk c hits a set of edges E_c if it hits all the edges in E_c. Note that several disks can hit the same set of edges.

First, a slight variant of Lemma 9 from [10] is given. This study's assumptions allow somewhat more general topologies with more than 2 collinear points. The segments $e \in H$ are assumed to be nondegenerate.

Theorem 5.1.2. *Let r be a positive real, and H be a nonempty set of intervals (i.e., edges) from $\mathbb{R}^2$ which is hit by a circular disk of radius r. Then there is a disk c of radius r which hits the intervals of H such that at least one of the following holds (see Fig. 5.2 for illustrations).*
(a) There are two non-parallel intervals in H such that c intersects both of them in a single point.



(a) $\forall e \in E$ and $\forall f \in E$ (b) $\forall e \in E$ and $\forall v \in V$ (c) $\forall e \in E$

Figure 5.3: The circular disasters examined in Thm. 5.1.3

These two points are different.
(b) There are two intervals in H such that c intersects both of them in a single point. These two points are different, and one of them is an endpoint of its interval.
(c) Disk c touches the line of an interval $e \in H$ at an endpoint of e.

Proof. For a line segment e on the plane and a nonnegative real number r the *r-neighborhood*[3] $N(e, r)$ of e is defined as the set of all points P on the plane which have distance at most r to (some point of) e. It is immediate that $N(e, r)$ is a closed convex subset (see Fig. 5.2d) of the plane.

Consider the boundary B of the intersection

$$\cap_{e \in H} N(e, r). \tag{5.2}$$

The points of B are obviously in the union of the boundaries of the neighborhoods $N(e, r)$, where $e \in H$. The union is composed of a finite number of line segments and half circles. The circular arcs belong to circles of radius r centered at endpoints of line segments $e \in H$. We distinguish two cases.

(1) B has a point R which is on a halfcircle arc of the boundary on $N(e, r)$ for some $e \in H$. Let c_R be the disk of radius r centered at R. If R is an endpoint (P_1 or P_2 in Fig. 5.2d) of the halfcircle, then (c) is satisfied for c_R. We can thus assume that R is an inner point of the halfcircle connecting P_1 and P_2, and $P_i \notin B$. From the fact that B is closed, we obtain that there exists a point R' on the circular arc RP_2 which is in B, but no point of the open $R'P_2$ arc is in B. Then there must be an $f \in H$ such that $N(f, r)$ passes through R' but does not contain a larger arc $R'R''$ from $R'P_2$. Then R' is on the boundary of $N(f, r)$. We argue that (b) holds for $c_{R'}$ and the intervals e, f. This is immediate if the tangent lines to $N(e, r)$ and $N(f, r)$ at R' are different. If they are the same line ℓ then e and f must be in different halfplanes defined by ℓ, hence $e \cap f = \emptyset$ and hence (b) holds for $c_{R'}$. This reasoning settles the case (1). Note that we can also assume now that $|H| > 1$.

(2) No point of B is on a circular arc form the boundary of $N(e, r)$, with $e \in H$. Then B is a (possibly degenerate) polygon composed of some line segments. Let R be a vertex of polygon B, and $e \in H$ be a segment such that R is an interior point of one of the line segments on the border of $N(e, r)$. Let ℓ

[3] called hippodrome in [17].

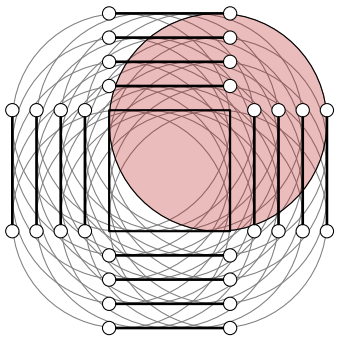


Figure 5.4: An example topology ($k = 4$) where the number of maximal SRLGs hit by circular disk shaped disasters is $\Omega(m^2)$ or $\Omega((n+x)\rho_r)$.

be the line of this latter segment. The fact that R is a vertex of B implies that there must be another segment $f \in H$ such that one of the line segments on the boundary of $N(f, r)$ passes through R and the line ℓ' of this segment is different from ℓ. Indeed, otherwise, for every $g \in H$ there would be an open interval form ℓ containing R in $N(g, r)$, which contradicts the extremality of R. As e is parallel to ℓ and f is parallel to ℓ', we infer that (a) holds for c_R. □

5.1.2 Bounds on the Number of SRLGs

Lemma 5.1.3. *Let H' be a set of intervals from $\mathbb{R}^2$, $|H'| \leq 2$, and r be a positive real number. Then every circular disk described in Thm. 5.1.2 for $H = H'$ can be determined in $O(1)$ time.*

Proof. Easy elementary geometric discussion of cases (a), (b) and (c) of Thm. 5.1.2. See Fig. 5.3 for illustration. Note that there can be at most 4 circles that intersect two line segments, as shown in Fig. 5.3(a), and at most two circles intersecting a line segment and a single point, as shown in Fig. 5.3(b), and four circles can touch a line at endpoints, as shown in Fig. 5.3(c). □

From Thm. 5.1.2 and the argument of Lemma 5.1.3 we obtain the following upper bound on the number of SRLGs.

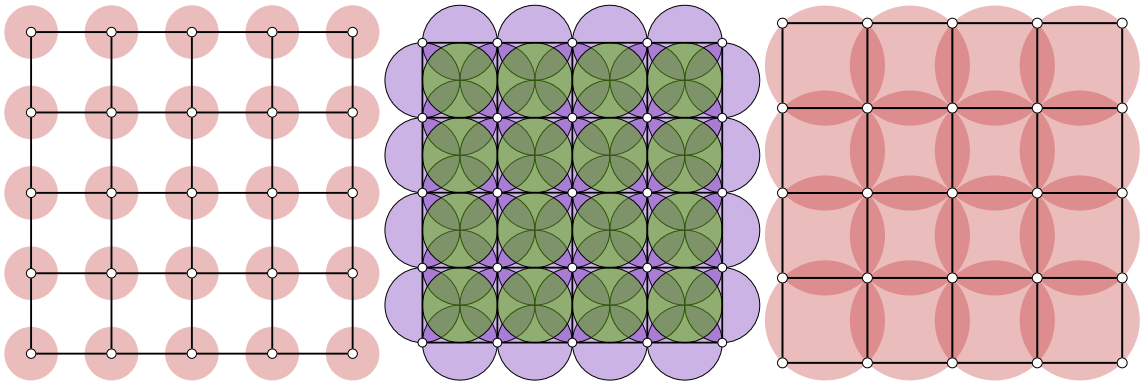
Corollary 5.1.4. $|M_r| \leq 4\binom{m}{2} + 4m + 2mn$.

Note that, the graphs of Claim 5.1.5 demonstrate that the above bound is asymptotically tight.

5.1.2.1 Worst Case Graph

Claim 5.1.5. *The graph sketched in Fig. 5.4 has at least $\frac{n^2}{64}$ maximal regional failures of a radius k.*

Proof. Here we construct a set of n segments whose graph is planar (there are no edge intersections), and for a suitable radius r it has at least $\frac{n^2}{64}$, in particular a quadratic number of, incomparable failure



(a) $r = 0.3$, $|M_r| = 25$ (b) $r = 0.5$, $|M_r| = 56$ (c) $r = \sqrt{2}/2$, $|M_r| = 16$

Figure 5.5: The set of SRLGs of a 5×5 grid network.

events.[4]

Let k be a positive integer. We consider a collection of $4k$ axis parallel line segments in $\mathbb{R}^2$. We start out with the four edges of the square of edge size k whose bottom left corner is at the origin $O = (0,0)$. We consider the bottom edge connecting O to $(k,0)$, and put its copies translated i units downwards, for $i = 1,\ldots,k$ into our set of segments. For example for $i = 2$ we obtain the segment from $(0,-2)$ to $(k,-2)$. This way we obtained k segments. Similarly we translate the upper edge (from $(0,k)$ to (k,k)) of the square by i units upwards for $i = 1,\ldots,k$. These are k additional horizontal segments. We do the same in the vertical direction: we consider k translates to the left of the left edge of our starting square, and k translates to the right of the right edge of the square. We have $4k$ nonintersecting line segments of length k. The configuration for $k = 4$ is shown in Fig. 5.4. Consider now a disk $c = c(i,j)$ of radius k centered at the point (i,j), where i,j are integers, $0 \le i,j \le k$. We readily see that c intersects exactly i of the right vertical segments and $k-i$ of the left vertical segments. Similarly, c intersects exactly j of the upper horizontal edges and $k-j$ of the lower horizontal edges. We infer that no two disks of the form $c(i,j)$ can hit the same set of edges. This implies that there are at least $(k+1)^2$ maximal failure events with radius k. The number of vertices is $n = 8k$. The number of such maximal failures is at least $\frac{n^2}{64}$. □

5.1.2.2 Circular Disk Failures with Radius at Most r

In this subsection, we take a more general model and assume that the radius of the failure is not a network-wide parameter but depends on the area. Our goal is to enumerate every circular disk failure for any radius at most r.

[4] No attempt has been made to optimize the constant. In fact, a more elaborate variant of the preceding construction gives $\frac{n^2}{16}$ maximal failures.

Definition 5.1.4. *Let a disk c be **smaller** than disk c', if c has a smaller radius than c', or if they have equal radius and the centre point of c is lexicographically smaller than the centre point of c'.*

Definition 5.1.5. *Let $F \subseteq E$ be a finite nonempty set of edges (not necessarily a failure). We denote the smallest disk among the disks hitting F by c_F, and we say c_F is the **smallest hitting disk** of F.*

It is not difficult to see that c_F always exists. The key idea of our approach that we can limit our focus only on the smallest hitting disks c_F, for $F \in F_r$, and ignore the rest of the disasters. The consequence of the next theorem is that the number of smallest hitting disks c_F, $F \in F_r$ is not too large.

Theorem 5.1.6. *Let H be a nonempty set of intervals from $\mathbb{R}^2$ with smallest covering disk c_H. Then there exists a subset $H' \subseteq H$ with $|H'| \leq 3$ such that $c_H = c_{H'}$.*

Thm. 5.1.6 would be trivial if the smallest hitting disks were defined on sets of nodes because a triplet of non-collinear nodes defines a circle. In the proof in Subsection 5.1.4.1 we show that this property holds for edges (considered as line segments) too. Compared to the algorithm of Thm. 5.1.2 here we not only shift the disks but also shrink them.

Corollary 5.1.7. $\left| \bigcup_{0<r<\infty} M_r \right| \leq \binom{m}{3} + \binom{m}{2} + m = \frac{m^3}{6} + \frac{5m}{6}$.

Theorem 5.1.8 (Thm. 28 of [T1]). *Let H be a set of intervals from $\mathbb{R}^2$, $|H| \leq 3$. Then c_H can be determined in $O(1)$ time.*

Remark. Thm. 5.1.8 outlines an efficient algorithm for c_H in an exact symbolic computational setting. A good numerical algorithm for approximating the radius r of c_H and the center P of c_H is also possible: for a positive real number r' we can efficiently test if $N(e_1, r') \cap N(e_2, r') \cap N(e_3, r') \neq \emptyset$. Indeed $N(e_i, r')$ is a union of two half disks and a rectangle, and the intersection of such objects is easily computable. Using such tests for emptiness, r can be approximated by binary search as the smallest r' providing nonempty intersection.

Since the smallest hitting disk of a triplet of edges can be calculated in $O(1)$ time, we could solve the problem by processing $O(m^3)$ triplets of edges. However, we will achieve better upper bounds on the running time and of $|M_r|$ with the help of some further observations.

5.1.3 Improved Bounds and Algorithm to Enumerate the Set of SRLGs

Next, we define five practical parameters of the input to better estimate the number of SRLGs and computing time.

ρ_r is the *link density* of the network, which is measured as the maximal number of links that could be hit by a circular disk shaped disaster of radius r.

x is the number of link crossings of the network G.

μ is the square mean of numbers v_e for all $e \in E$, where v_e is the number of $w \in V \cup X$ such that $d(w,e) \leq 3r$.

In backbone networks, x is a small number as typically a network node is also installed on each link crossings [89], while the link density ρ_r practically should not depend on the network size. We also know that ρ_r is at least the maximal nodal degree in the graph. For simplicity, we assume that edges intersect in at most one point.

Definition 5.1.6. *Let X be the set of points p which are not in V and there exist at least 2 non-parallel edges crossing each other in p. Let $x = |X|$.*

As mentioned before, in backbone network topologies, typically $x \ll n$. This is because a switch is usually installed if two cables are crossing each other[5]. It gives us the intuition that G is "almost" planar, and thus it has few edges.

Claim 5.1.9. *The number of edges in G is $\Omega(n)$ and $O(n+x)$.*

Proof. Since G is connected, $m = \Omega(n)$ is immediate. The upper bound was proved in [T1] as follows. Let $G'(V \cup X, E')$ be the planar graph obtained from dividing the edges of G at the crossings. Since every crossing increases the number of edges by at least two, $|E'| \geq m + 2x$. On the other hand, $|E'| \leq 3(n+x) - 6$ since G' is planar. Thus $m \leq |E'| - 2x \leq 3n + x - 6$. □

Here we add a note on the Crossing Lemma giving a lower bound on x in function of n and m. For a given graph G, let $\mathrm{cr}(G)$ the minimum number of edge crossings over the planar embeddings of G. Thm. 6. of [91] states that $\mathrm{cr}(G) \geq \frac{1}{29}\frac{m^3}{n^2} - \frac{35}{29}n$, and if $m \geq 6.95n$, then $\mathrm{cr}(G) \geq \frac{1}{29}\frac{m^3}{n^2}$.

5.1.3.1 Lower Bound on Computing the Maximal Failures

Now we present a straightforward lower bound on the time needed to determine M_r. As it will turn out (in Cor. 5.1.19), in specific circumstances, this lower bound is asymptotically tight.

Corollary 5.1.10. *The complexity of computing M_r is $\Omega(n \log n)$.*

Proof. By combining Prop. 4.3.1 (Lemma 4 of [88]) and Claim 5.1.9, we get that reporting that there are no intersecting line segments takes $\Omega(n \log n)$. In other words, this means that computing M_r in the special case of $r = 0$ needs $\Omega(n \log n)$ time. □

[5] Recent experimental studies give empirical evidence that real-world road networks typically have $\Theta(\sqrt{n})$ edge crossings [90].

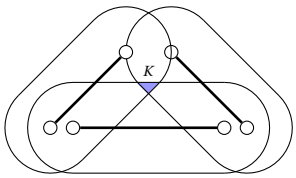


Figure 5.6: Illustration to Thm. 5.1.12

5.1.3.2 Upper Bounds and Algorithm for Computing the Maximal Failures

The set of link intersections X can be computed in near-linear time, for example, with the help of algorithm Bentley-Ottmann [77] briefly explained in Chp. 4.

Claim 5.1.11. *X can be reported in $O((n+x)\log n)$ time and $O(n+x)$ space.*

Proof. To easily distinguish nodes and edge intersections geometrically, edges are shortened in both directions with a tiny fraction of their length. The statement follows by using Proposition 4.3.2 (Theorem 2.4 of [77]) and Claim 5.1.9 by noting also that $O(\log(n+x))$ is $O(\log n)$. □

The next theorem states, it is enough to process the edge triplets in the neighborhood with radius $3r$ of every point in $V \cup X$.

Theorem 5.1.12 (Thm. 32 of [T1]). *For every failure $H \in F_r$ there exists a disk c of radius at most r hitting H with centre point at distance at most $2r$ from $V \cup X$.*

Theorem 5.1.13. *Let r be a positive real number, $F \in M_r$ be a set of line segments which can be hit by a disk of radius r. Then there exists a segment $e \in F$ and a disk c described in Thm. 5.1.2 (disk c has radius r, hits F, intersects e in a single point Q, and (a), or (b), or (c) holds with $H = F$), such that the centre point of c is at distance at most $2r$ from either an endpoint of e or a point of crossing (of e and another segment $f \in F$).*

Proof. We proceed along the lines of the proof of Thm. 5.1.2. If we are in case (1) of the proof of Thm. 5.1.2, then (b) or (c) holds for the statement of the theorem, as Q can be an endpoint of a segment $e \in F$.

We may turn our attention to case (2) from Thm. 5.1.2. Then $K = \cap_{e \in F} N(e, r)$ is a closed bounded convex set on the plane whose boundary is a polygon composed of line segments. If K has no interior points in the plane, then r is an optimal hitting radius for F. Then $c = c_F$ will be a suitable disk. The proof of Thm. 5.1.12 can be extended to show that the requirements of Theorem 5.1.2 will be valid for c_F in the place of c. This follows from a simple but tedious analysis of the Cases 1-4 of Theorem 5.1.8, which we omit here.

We may, therefore, assume that K has an interior point (see also Fig. 5.6). Then K is a proper convex k-gon for some $k \geq 3$, hence there exists a vertex R of K with angle $\alpha \geq \frac{\pi}{3}$. The circle of

radius r centered at R will meet the requirements of the theorem. Indeed, there will be then two segments $e, f \in F$ such that their supporting lines are tangent to c, and c is seen at angle α from their point of intersection. Q will be the point of tangency of e or f with c. See the last case in the proof of Thm. 5.1.12 for further details. □

Next, we will give better upper bounds on the number of SRLGs. As a consequence of Theorem 5.1.13, when considering circular disasters of radius r, then in a sense, we may ignore the points on the edges $e \in E$ which are more than $3r$ away from $V \cup X$. Consider the pairs (e, v) where $e \in E$, $v \in V \cup X$, and $v \in e$. If we have an SRLG of radius r as in Theorem 5.1.13 with edge e such that the distance of c is at most $2r$ from v, then the edges of this SRLG must intersect the disk of radius $3r$ centered at v. This gives at most $15\rho_r$ possibilities for the other edge besides e in Theorem 5.1.13 (a) or (b) (see Fig. 5.7, where 15 circular disks of radius r cover a disk of radius $3r$). The number of pairs (e, v) can be counted by looking at the contribution of node v: it will be $\deg v$, where deg is the degree in the planarized graph. The sum of the degrees is twice the number of the edges of the latter graph, which is $O(n + x)$. Thus we have the following bound:

Corollary 5.1.14. $|M_r| = O((n + x)\rho_r)$.

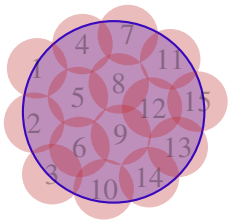


Figure 5.7: A disk with radius $3r$ can be covered with 15 disks with radius r. Generally, covering a disk with a radius ϵ with the fewest possible number of disks with radii 1 is called the disk covering problem [92].

This bound is asymptotically tight[6] on the graphs in Claim 5.1.5 because $\rho_r = \frac{n}{2}$ for $r = k$. Next, we discuss the algorithm to generate the list of SRLGs.

Theorem 5.1.13 together with other formerly presented results inspire an improved algorithm with a running time near linear in n described in Table 5.2. The main idea is to build up local data structures, pre-compute the lists of candidate members of M_r, then merge these lists, all in nearly linear time. With this aim, we make the following definitions.

Definition 5.1.7. *For a given r and $w \in V \cup X$, let $E_w := \{e \in E |\ d(w, e) \leq 3r\}$; and let the edges in E_w be given in sorted order with respect to the lexicographic ordering of their endpoints. For a given $e \in E$, let $V_e := \{w \in V \cup X |\ d(e, w) \leq 3r\}$.*

[6] No attempt have been made to optimize the constant.

Table 5.2: Algorithm for determining M_r and complexity of its tasks

#	Task	Complexity
1	Determine X	$O((n+x)\log n)$
2	For $w \in V \cup X$ determine E_w	$O((n+x)(\log n+\rho_r^2))$
3	For $e \in E$ determine V_e	$O((n+x)(\log n+\rho_r^2))$
4	For $w \in V \cup X$ determine $\mathscr{L}_{r,w}$	$O((n+x)\rho_r^3)$
5	For $e \in E$ for $w_1, w_2 \in V_e$ compare $\mathscr{L}_{w_1}$ with $\mathscr{L}_{w_2}$	$O((n+x)\mu\rho_r^5)$
6	Merge resulting lists in M_r	$O(n+x)$

Theorem 5.1.15. *All the sets E_w for $w \in V \cup X$ can be determined in $O((n+x)(\log n+\rho_r^2))$. Similarly, all the sets V_e for $e \in E$ can be computed in the same time complexity.*

The proof of Thm. 5.1.15 is relegated to Subsection 5.1.4.2.

Lemma 5.1.16. *The set of SRLGs for circular disk shaped disasters of radius r can be computed in $O((n+x)(\log n+\rho_r^3))$.*

Proof. Based on Claim 5.1.11 and Thm. 5.1.15, E_w can be determined in the proposed complexity for all $w \in V \cup X$.

Then for every node w, we compute list $\mathscr{L}_{r,w}$ containing the set of edges hit by an element of disk set $\mathscr{C}_{r,w}$ defined as follows: for $e, f \in E_w$ we compute disks c of radius r (if exist) according to Thm. 5.1.2: either case a) applies if e and f are not parallel, and c intersects them in two different points, or case b) when c intersects e and f in two different points, one being an endpoint of e, or case c) when c touches e at an endpoint; moreover we require that formerly computed disks c have centres not farther than $2r$ from w. These disks are collected in $\mathscr{C}_{r,w}$. This takes $O((n+x)\rho_r^3)$ time, since there are $O(\rho_r^2)$ disks c to determine and store in $\mathscr{C}_{r,w}$, and for each $c \in \mathscr{C}_{r,w}$ the set of edges hit by c can be determined in $O(\rho_r)$ time based on E_w. It follows readily from Thm. 5.1.13 that for every $F \in M_r$ there exists a $w \in V \cup X$ such that F is a subset of an element of list $\mathscr{L}_{r,w}$. □

Please note that lists $\mathscr{L}_{r,w}$ together may contain duplicates and non-maximal sets as well, those will be eliminated later at a subsequent phase.

Finally, based on Cor. 5.1.7 we give an upper bound on the total number of circular disk failures with radius at most r.

Proposition 5.1.17. $\left|\bigcup_{0<r'<r} M_{r'}\right|$ *is* $O((n+x)\rho_r^2)$.

Proof. We can use Theorems 5.1.6 and 5.1.12 and the fact that a disk of radius $3r$ hits $O(\rho_r)$ segments. From Theorem 5.1.6, we see that it suffices to construct disks of the form c_H, for sets of segments H of size at most 3. Then by Theorem 5.1.12 it is enough to calculate for every $v \in V \cup X$

the smallest hitting disk of every set H containing an edge going through v and containing 1 or 2 edges from the $3r$ neighborhood of v. For a fixed v we have $O(\deg v \cdot \rho_r^2)$ SRLGs, and the claim follows. □

As mentioned after Lemma 5.1.16, the final task for determining M_r is to merge lists $\mathscr{L}_{r,w}$ by eliminating duplicates and non-maximal elements. To do this in subquadratic time in n, one must avoid comparing all pairs of lists $\mathscr{L}_{r,w_1}, \mathscr{L}_{r,w_2}$.

Definition 5.1.8. *Let μ be the mean square of numbers $|V_e|$ for all $e \in E$, i.e. $\mu := \frac{\sum_{e\in E}|V_e|^2}{m}$.*

Theorem 5.1.18. *The maximal circular disk failures with radius exactly r can be computed in time $O((n+x)(\log n + \mu\rho_r^5))$ and this is tight in n.*

Proof. According to Lemma 5.1.16, all sets of failures $\mathscr{L}_{r,w}$ can be determined in time $O((n+x)(\log n + \rho_r^3))$.

We observe that it is enough to compare lists $\mathscr{L}_{r,w_1}$ and $\mathscr{L}_{r,w_2}$ for possible containment or duplicates only if $E_{w_1} \cap E_{w_2} \neq \emptyset$, or in other words there exists an $e \in E$ for which $\{w_1, w_2\} \subseteq V_e$. We deduce that it is enough to compare for all $e \in E$ and $w_1, w_2 \in V_e$ list pairs $\mathscr{L}_{r,w_1}, \mathscr{L}_{r,w_2}$. This means comparing at most

$$\sum_{e\in E} \frac{|V_e|(|V_e|-1)}{2} < m\frac{\sum_{e\in E}|V_e|^2}{m} = m\mu \overset{\text{Claim 5.1.9}}{=} O((n+x)\mu)$$

pairs of lists, with each list having $O(\rho_r^2)$ elements. Taking into consideration that a comparison of two elements (SRLG candidates) can be done in $O(\rho_r)$, we obtain a complexity of $O((n+x)\mu\rho_r^5)$, confirming the claim for the total complexity. The lower bound is provided by Corollary 5.1.10. □

Table 5.2 summarizes the steps of our proposed algorithm. Note that parameters ρ_r, x, and μ are theoretically upper bounded by m, $\frac{m(m-1)}{2}$, and $(n+x)^2$, respectively, meaning that our algorithm for determining M_r is clearly polynomial in n or m. Furthermore, based on Thm. 5.1.18 using that x is $O(n)$ in practice, and that ρ_r is more or less proportional to $\frac{2r}{diam}m$ ([C10])) in the interval $(0, diam/2]$, where *diam* is the geometric diameter of the network, we get a complexity bound of $O\left(n(\log n + \mu(\frac{r}{diam})^5)\right)$ for determining M_r. Also, as in practice $x = O(n)$, and for r much smaller than network diameter, $\rho_r = O(1)$, and $\mu = \log(n)$ we can state that:

Corollary 5.1.19. *If $\rho_r = O(1)$, $\mu = O(\log n)$, and x is $O(n)$, M_r can be calculated in $O(n \log n)$ optimal time. These assumptions hold in practice when r is much smaller than the geographical network diameter.*

Proof. Combining Thm. 5.1.18 and Cor. 5.1.10 yields the proof. □

In the phrasing of Thesis 1.1, instead of μ, I use a more intuitive parameter, namely, ϕ_r:

Definition 5.1.9. *Parameter ϕ_r denotes the maximum number of nodes in the $3r$-neighborhood of a link of the input graph G.*

Corollary 5.1.20. *M_r^p can be determined in $O\big((|V|+x)\big(\log|V|+\phi_r^2\rho_r^5\big)\big)$.*

Proof. Combining Thm. 5.1.18 and the definition of parameters μ and ϕ_r completes the proof. □

5.1.4 Auxiliary Proofs of Section 5.1

5.1.4.1 Proof of Thm. 5.1.6

We need the following simple lemma.

Lemma 5.1.21. *Let C_1, C_2, C_3 be convex subsets of the plane $\mathbb{R}^2$ such that $t = C_1 \cap C_2 \cap C_3$ is a line segment with more than 1 point. Then there exist two indices i, j such that $C_i \cap C_j$ is collinear.*

Proof. Let R, S be two different points of t. If the statement is false, then the pairwise intersections $C_i \cap C_j$ all contain points not on the line of t. Without loss of generality we may assume that $C_1 \cap C_2$ and $C_1 \cap C_3$ contain points P_2 and P_3 from the same open halfplane defined by the line of t. If $P_3 = P_2$ then we obtain $P_2 \in t$ which is a contradiction. We infer that P_3, P_2, R, S are four different points. Radon's lemma (Theorem 1.3.1 in [93]) can be applied to them. The Radon point X will be on one hand on the open halfplane defined by the line of t and containing the P_i. On the other hand $X \in C_1 \cap C_2 \cap C_3$. This gives a contradiction. □

Proof of Thm. 5.1.6. Let r be the radius of c_H. We have then

$$\cap_{e\in H} N(e, r) \neq \emptyset, \tag{5.3}$$

but $\cap_{e\in H} N(e, r') = \emptyset$ for any $r' < r$.

The statement of the theorem is immediate if H has at most 2 sets. Suppose now that $|H| \geq 3$. If for every 3-element subset H' of H there exists a radius $r_{H'} < r$ such that $\cap_{e\in H'} N(e, r_{H'}) \neq \emptyset$, then with $r^* = \max r_{H'}$ we have $\cap_{e\in H} N(e, r^*) \neq \emptyset$ by the planar Helly's theorem (Theorem 1.3.2 in [93]) applied to the convex sets $N(e, r^*)$, hence H can be hit by a disk of radius r^*, which is impossible. We obtain that there exists a 3 element subset H' of H such that the radius of $c_{H'}$ is r.

Note also, that the intersection on the left of (5.3) is necessarily a (possibly degenerate) nonempty closed bounded line segment s. This follows from the fact that the intersection is a nonempty closed bounded convex subset *without an interior point*. Indeed an interior point would allow a hitting radius for H, which is less than r. Note also that the lexicographically smallest (end)point P of s is the center of c_H.

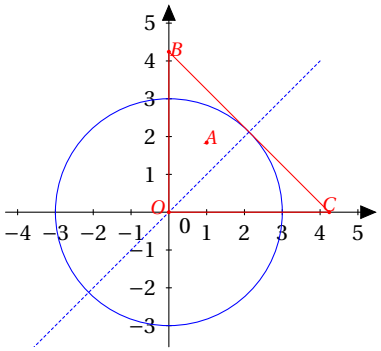


Figure 5.8: Illustration for proof of Lemma 5.1.22

We observe next that for H' above the hitting radius r is also minimal, hence the intersection $\cap_{e\in H'} N(e,r) = s'$ is also a line segment which contains s. If the smallest point of s' is P then we are done, as P will be the center of $c_{H'}$. We may therefore suppose that s' contains a point Q smaller than P.

Suppose that $H' = \{e_1, e_2, e_3\}$. We verify that there exist i, j, $1 \le i < j \le 3$, such that the intersection $N(e_i, r) \cap N(e_j, r)$ is a subset of the line of s'. Indeed, this follows from Lemma 5.1.21 applied to the neighborhoods $N(e_i, r)$ and $t = s'$.

We conclude by noting that there exists an edge $f \in H$ such that $N(f, r)$ does not have a point on the line of s' which is smaller than P (otherwise s itself had such a point). These imply that the lexicographically smallest point of $N(e_i, r) \cap N(e_j, r) \cap N(f, r)$ is P and the proof is complete. □

5.1.4.2 Proof of Thm. 5.1.15

We need the following three simple lemmas.

Lemma 5.1.22. *Let $A = (x, y)$ be a point in the plane of distance at most 3 from the origin. Then, any line going through A intersects either the x- or the y-axis not farther than $3\sqrt{2}$ from the origin.*

Proof. Without loss of generality, we can assume A is in the first quadrant of the plane. Let $B = (0, 3\sqrt{2})$ and $C = (3\sqrt{2}, 0)$, respectively. Then line of BC is tangent to the circle centered at the origin O and having radius 3 (at the point $\left(\frac{3}{\sqrt{2}}, \frac{3}{\sqrt{2}}\right)$, see Fig. 5.8). Now any line ℓ passing through A must intersect a side of the triangle OBC, hence it intersects at least two sides (Pasch's axiom), therefore ℓ intersects either OB or OC. □

Definition 5.1.10. *Let ρ' be the maximum number of edges of E' intersecting a disk with radius $3r$.*

Lemma 5.1.23. *ρ' is $O(\rho_r)$.*

Proof. For any point p, the number of edges of G' hit by the disk with radius r and center point p is less or equal to the number of edges of G hit by the disk with radius $(1 + 3\sqrt{2})r$ and center point p,

which is $O(\rho_r)$ since a disk with radius $(1+3\sqrt{2})r$ clearly can be covered by a constant number of disks with radius r. □

Lemma 5.1.24. *There are $O((n+x)\rho_r^2)$ intersecting link pairs in link set E' resulting from elongating each edge of E by $3\sqrt{2}r$ in both directions.*

Proof. Let $\{e, f\} \in E$ be two links such that only their elongated versions $\{e', f'\} \in E'$ are crossing in a point z. We claim that this z is on the elongated part of at least one of e' or f', i.e., considering the edges as geometric intervals, $z \in e' \setminus e$ or $z \in f' \setminus f$. Also, for each $e' \in E'$, there are $O(\rho_r)$ edges of E' that cross $e' \setminus e$, since, for an edge $f' \in E'$ to cross $e' \setminus e$, $d(e' \setminus e, f)$ has to be $\leq 3\sqrt{2}r$, and the $3\sqrt{2}r$ neighborhood of $e' \setminus e$ (where $e' \setminus e$ stands of two $3\sqrt{2}r$ long intervals) can be covered with a constant number of disks with radius r. Based on these, and using that $|E'|$ is $O(n+x)$ (Claim 5.1.9), we can deduce that there are $O((n+x)\rho_r)$ newly appearing crossing link pairs in E' in addition to those that are crossing in E in a point of $V \cup X$. Regarding to the number of these 'old crossings', in each point of $V \cup X$, there are at most ρ_r links of E crossing (and those links of E' that cross in $V \cup X$, were already counted), meaning $O((n+x)\rho_r^2)$ crossing link pairs. This means a total of $O((n+x)\rho_r^2)$ crossing link pairs in E'. □

Proof of Thm. 5.1.15. First, let us concentrate on determining sets E_w for $w \in V \cup X$. Let $G'(V, E')$ be the graph resulting from elongating the edges of E by $3\sqrt{2}r$ in both directions. For reporting link intersections in some slightly modified versions of G', we shall use the Chazelle algorithm [88] that, out of m links, reports all the k intersecting pairs in $O(m \log m + k)$ (Prop. 4.3.1).

The most important observation is that, based on Lemma 5.1.22, if an edge $e \in E$ is also part of E_w for a $w = (x, y) \in V \cup X$, then the corresponding edge e_{3r} in E' (that was extended in length by $3\sqrt{2}r$ in both directions) intersects either $I_w^| := [(x - 3\sqrt{2}r, y), (x + 3\sqrt{2}r, y)]$ or $I_w^- := [(x, y - 3\sqrt{2}r), (x, y + 3\sqrt{2}r)]$. Here we use also the simple fact that the diameter of a square (the length of the longest segment within the square) of side length $3r$ is $3\sqrt{2}r$.

Let $G'^|$ be the graph resulting by adding intervals $I_w^|$ to G' for every $w \in V \cup X$ as edges of the graph. Let $E_w'^|$ denote the set of edges (of E') intersecting $I_w^|$. $G_w'^-$ and $E_w'^-$ can be defined similarly. It is easy to see that $E_w'^| \cup E_w'^-$ contains all the edges, which in the original graph G are not farther from w than $3r$, however, it may contain some outliers. Thus in order to get E_w, one can check the distance of the original (i.e., not extended) edges from w, which correspond to edges in $E_w'^| \cup E_w'^-$ from w.

It is easy to see that $G'^|$ has still $O(n+x)$ edges. We count the number of pairwise intersections in $G'^|$ as follows. By Lemma 5.1.24, in G', there are $O(n+x)\rho_r^2$ link pairs crossing. In addition to these, each of the $(n+x)$ new edges (intervals) in $G'^|$ intersect $O(\rho_r)$ other edges (since the $3\sqrt{2}r$ neighborhood of each of these $6\sqrt{2}r$ long edges can be covered with a constant number of disks of radius r, and in case of an $e \in E$ elongated as $e' \in E'$, e' crossing a $| \in E'^|$ means $d(e, |) \leq 3\sqrt{2}r$). This

sums up to $O((n+x)\rho^2+(n+x)\rho_r)$, that is $O((n+x)\rho_r^2)$. Thus, by Prop. 4.3.1, the intersections of $G'^{|}$ can be determined in $O((n+x)\log n+(n+x)\rho_r^2)$, that is $O((n+x)(\log n+\rho_r^2))$ time, alongside with the sets $E'^{|}_w$ for $w\in V\cup X$. The same reasoning applies to the sets E'^{-}_w.

For any given $w\in V\cup X$, $E'^{|}_w\cup E'^{-}_w$ contains $\leq 2\rho'$ edges, this way based on Lemma 5.1.23, E_w can be determined in $O(\rho_r\log\rho_r)$ time in such a way that the edges are given in E_w in sorted order with respect to the lexicographic ordering of their endpoints. This means a total complexity of $O((n+x)\rho_r\log\rho_r)$ for this second phase.

The inverse mapping, i.e., sets of nodes V_e for $e\in E$, can be done in the course (or after) the preceding algorithm. Let V_e be initialized as empty set for all edges e, then, when an E_w is confirmed, w is added to sets V_e for all $e\in E_w$. Clearly, this also can be done in the proposed complexity. □

5.2 Spherical Regional Link Failures of Disasters with Radius r

5.2.1 Model and Assumptions

However, in the Section, we are more interested in the spherical representations, throughout the Section, we will consider two types of embeddings of the network: embedding in Euclidean planar and spherical geometry. Thus, the Section will provide, in fact, a heuristic algorithm for determining both the planar and spherical maximal link failures caused by disaster zones having a radius of r.

The network is modeled as an undirected connected geometric graph $G=(V,E)$ with $n=|V|\geq 3$ nodes and $m=|E|$ edges stored in a lexicographically sorted list. The nodes of the graph are embedded as points in the Euclidean plane or sphere, and their precise coordinates are considered to be given in 2D and 3D Cartesian coordinate system in the planar and spherical case, respectively. Note that if coordinates are given in polar system (in the case of spherical geometry), one can easily transform them to Cartesian at the very beginning.

When speaking of planar geometry, for each edge e there is a *polygonal chain* (or simply *polyline*) e^p in the plane in which the edge lies (see Fig 5.9). Parameter γ will be used to indicate the maximum number of line segments a polyline e^p can have. Naturally, in spherical case, the polyline of an edge refers to a series of geodesics. Note that this model covers special cases when edges are considered as line segments (geodesics).

For simplicity, I assume that nodes of V and the corner points of the containing polygons defining the possible route of the edges are all situated in general positions of the plane, i.e., there are no three such points on the same line and no four points on the same circle, and in the spherical case there are no antipodal nodes or breakpoints and no great circles of geodesics of polylines cross the North pole.

In this study, my goal is to generate a set of SRLGs, where each SRLG is a set of edges. Note that from the viewpoint of connectivity, listing failed nodes besides listing failed edges has no additional information.

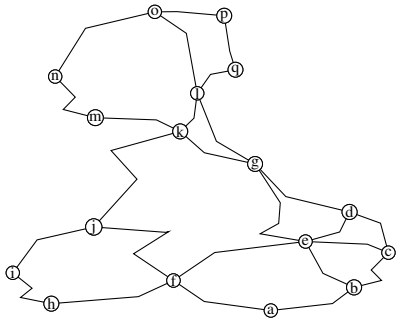


Figure 5.9: Input graph $G(V,E)$ with polylines, $n = 17$, $\gamma = 4$

5.2.1.1 Model for Circular Disk Shaped Disasters

In most of this study, it will be assumed that disasters are either having a shape of a circular disk or they are overestimated by a circular disk.

I will often refer to circular disks simply as disks, and I assume that all network elements that intersect the interior of a circle c are failed, and all other network elements are untouched.

Definition 5.2.1. *A circular disk disaster c hits an edge e if the polyline of the edge e^p intersects disk c. Similarly node v is hit by disk c if it is in the interior of c. Let E_c (and V_c) denote the set of edges (and nodes, resp.) hit by a disk c.*

I emphasize that in this model, when I say e is hit by c, it does not necessarily mean that e is destroyed indeed by c, instead, it means that there is a positive chance for e being in the destroyed area. In other words, this modeling technique does not assume that the failed region has a shape of a disk but overestimates the size of the failed region in order to have a tractable problem space.

Definition 5.2.2. *Let $\mathscr{C}^p$ and $\mathscr{C}^s$ denote the set of all disks in the plane and the set of all disks on the sphere, respectively. For both geometry types $g \in \{p,s\}$, let $\mathscr{C}_r^g$ denote the set of disks part of $\mathscr{C}^g$ having radius at most r.*

Based on the above definition, I define the set of failure states that a network may face after a disaster with a maximal radius.

Definition 5.2.3. *For all geometry types $g \in \{p,s\}$, let set $F(\mathscr{C}_r^g)$ denote the set of edges which can be hit by a disk $c \in \mathscr{C}_r^g$, and let $M_t^g = M(\mathscr{C}_r^g)$ denote the set of maximal edge sets in $F(\mathscr{C}_r^g)$.*

5.2.2 Heuristic Algorithm for Enumerating Maximal Circular Disk Failures

In this Section, we present a heuristic approach suitable for computing both M_r^p and M_r^s.

Definition 5.2.4. *For a point P (in the plane or on the sphere) and node $v \in V$, let the node-distance couple be $[v, d(v,P)]$, where $d(v,P)$ is the distance of v and P. Let $e(P)$ be the list consisting of the*

Table 5.3: Table of symbols for Sec. 5.2

Notation	Meaning
	General
$G(V,E)$	the network modeled as an undirected connected geometric graph, E stored as an ordered list
n, m	number of nodes $\|V\| \geq 3$ and edges $\|E\|$, respectively
e^p	for edge e, in case of planar geometry, there is a *polygonal chain* (or simply *polyline*) e^p in the plane in which the edge lies (see Fig 5.9); on the sphere e^p is a similar region enclosed by a closed sequence of geodesics
E_c and V_c	set of edges and nodes, resp. hit by a disk c, where a circular disk disaster c hits an edge e if the polyline of the edge e^p intersects disk c. Similarly node v is hit by disk c if it is in the interior of c.
$\mathscr{C}^p$ and $\mathscr{C}^s$	set of all disks in the plane and the set of all disks on the sphere, respectively; for both geometry types $g \in \{p,s\}$, let $\mathscr{C}_r^g$ denote the set of disks part of $\mathscr{C}^g$ having radius at most r
$\mathscr{C}_r^g$	for both geometry types, denotes the set of disks part of $\mathscr{C}^g$ having radius at most r
$F(\mathscr{C}_r^g)$	set of edges which can be hit by a disk $c \in \mathscr{C}_r^g$
M_t^g	set of maximal edge sets in $F(\mathscr{C}_r^g)$
$[v, d(v,P)]$	node-distance couple for a point P (in the plane or on the sphere) and node $v \in V$, where $d(v,P)$ is the distance of v and P
$e(P)$	list consisting of the link-distance pairs of all links $e \in E$, sorted according to the lexicographical order of the links
$e(P)_{\text{hit}}$	sorted list of links not further from P than r
$\mathscr{P}$	set of points P for which we want to construct the link-distance lists $e(P)$.
$d_{\mathscr{P}}$	maximal distance of any geometric location from the (closed) convex hull of the geometric embedding of graph G to the closest point of set $\mathscr{P}$
$\sqsupseteq$	taken two set of sets E_1 and E_2, we denote the relationship of the sets with $E_1 \sqsupseteq E_2$ if and only if for all $e_2 \in E_2$ there exists an $e_1 \in E_1$, such that $e_1 \supseteq e_2$
H_r^g	output of Algorithm 1 (a failure list approximating M_r^g)
	Parameter
γ	maximum number of line segments (geodesics) a polyline e^p can have
λ	maximum cardinality of the list of candidate maximal failures detected so far in Alg. 1

link-distance pairs of all links $e \in E$, sorted according to the lexicographical order of the links. Let $e(P)_{hit}$ be the sorted list of links not further from P than r.

Proposition 5.2.1. *For a given point P, both $e(P)$ and $e(P)_{hit}$ can be computed in $O((n+x)\gamma)$.* □

Clearly, both node-distance lists and edge-distance lists can be determined quickly. Informally speaking, the plan is to determine these lists for a point set that is 'dense enough' to be able to determine the maximal SRLG lists based on these node-distance and edge-distance lists.

Definition 5.2.5. *Let $\mathscr{P}$ denote the set of points P for which we want to construct the link-distance lists $e(P)$.*

Let us stick to planar geometry for a moment. Intuitively, we can calculate M_r^p by including the grid points of a sufficiently fine grid (let's say containing 1 km × 1 km squares) in $\mathscr{P}$. On the sphere, we should choose a similar nice covering[7].

[7] In other words, $\mathscr{P}$ is an ϵ-*net* [94] of the network area, for some $\epsilon > 0$, and the range space of the network area paired with the set of closed circular disks. Here, to design an approximation algorithm, we exploit that the network area has a finite Vapnik–Chervonenkis (VC) dimension [95] both in the plane and on the sphere.

Definition 5.2.6. *Let $d_{\mathscr{P}}$ be the maximal distance of any geometric location from the (closed) convex hull of the geometric embedding of graph G to the closest point of set $\mathscr{P}$, i.e. $d_{\mathscr{P}} := \max_{t \in conv(G)} \min_{p \in \mathscr{P}} dist(p,t)$.*

Definition 5.2.7. *Taken two set of sets E_1 and E_2, we denote the relationship of the sets with $E_1 \sqsupseteq E_2$ if and only if for all $e_2 \in E_2$ there exists an $e_1 \in E_1$, such that $e_1 \supseteq e_2$.*

Algorithm 1 is an example heuristic algorithm for determining M_r^g. We will refer to the output of the algorithm as H_r^g. Starting from Thm. 5.2.2, we use an additional parameter λ:

Definition 5.2.8. *Let λ be the maximum cardinality of the list of candidate maximal failures detected so far in Alg. 1.*

The intuition behind defining λ is, that in practice, $|M_r^g|$ is $O(n)$ (as presented in planar case in [T1]), thus in Alg. 2 typically there has to be done only $O(n)$ comparisons.

Algorithm 1: Heuristic algorithm for determining the maximal r-range SRLG lists

```
   Input: G(V,E), r, 𝒫, geometry type g,
     coordinates of nodes and polylines of edges
   Output: H_r^g
   begin
1      for P ∈ 𝒫 do
2          determine e(P)_hit
3          if e(P)_hit ≠ ∅ then
4              refresh M_r^g with e(P)_hit
                   // according to Alg. 2
5      return H_r^g
```

Algorithm 2: Refreshing SRLG list M with failure f

```
   Input: SRLG list M, failure f
   Output: M refreshed with f
   begin
1      maximal:=True
2      for f_M ∈ M do
3          if f ⊆ f_M then
4              maximal:=False
5      if maximal then
6          M := M ∪ {f}
7          for f_M ∈ M do
8              if f ⊃ f_M then
9                  M := M \ {f_M}
10     return M
```

Theorem 5.2.2. *Alg. 1 determines H_r^g in $O(|\mathscr{P}|[(n+x)\gamma + \lambda\rho_r])$. Furthermore, $M_r^g \sqsupseteq H_r^g \sqsupseteq M_{r-d_P}^g$.*

Proof. Regarding to the complexity, for an element P of $\mathscr{P}$ we have to construct $e(P)_{\text{hit}}$, which can be done in $O((n+x)\gamma)$, then refresh the list of suspected maximal failures with $e(P)_{\text{hit}}$ in $O(\lambda\rho_r)$, since the list contains at most λ ordered lists consisting of at most ρ_r edges.

On the other hand, $M_r^g \sqsupseteq H_r^g$ is immediate, since the algorithm investigates only a subset of disks with radius r, while for every point t in the r-neighborhood of conv(G), there exists a $p \in \mathscr{P}$ such that disk $c(t, r - d_{\mathscr{P}}) \subseteq c(p,r)$, yielding $H_r^g \sqsupseteq M_{r-d_p}^g$, from where the proof follows. □

Using the fact that the shape of the disasters is a closed disk we get the following corollary:

Corollary 5.2.3. $\lim_{d_{\mathscr{P}} \to 0} H_r^g = M_r^g$, *for any fixed network.*

Corollary 5.2.4 (of Thm. 5.2.2). *$M_r^g \supseteq H_r^g \supseteq M_{r-d_P}^g$. Furthermore, if both of x and λ is $O(n)$, the resulting list H_r^g of running Alg. 1 is determined by the algorithm in $O(|\mathscr{P}|n(\gamma+\rho_r))$. If in addition, γ is $O(1)$, and ρ_r is $O(r/diam)$, H_r^g is determined by Alg. 1 in $O(|\mathscr{P}|n\frac{r}{diam})$.*

Based on Thm. 5.2.2, if one wants to protect disasters caused by disks with radius r (i.e., over-approximate the failures caused by them), it is only needed to run Alg. 1 initializing the radius as $r + d_{\mathscr{P}}$. This way, adding the fact that $\lambda \leq |\mathscr{P}|$, we have:

Corollary 5.2.5 (of Thm. 5.2.2). *$M_{r+d_P}^g \supseteq M_r^g \supseteq H_r^g$. Trivially, if $\mathscr{P}$ is such that $H_{r+d_P}^g = H_r^g$, then Alg. 1 calculates M_r^g in $O(|\mathscr{P}|[(n+x)\gamma+\lambda\rho_r])$, that is $O(|\mathscr{P}|[(n+x)\gamma+|\mathscr{P}|\rho_r])$.*

This gives us a *heuristic* way to calculate M_r^s: if, for $\mathscr{P}$, $H_{r+d_P}^g = H_r^g$, Alg. 1 calculates M_r^g in polynomial time; else, we provide a denser $\mathscr{P}$, and try again.

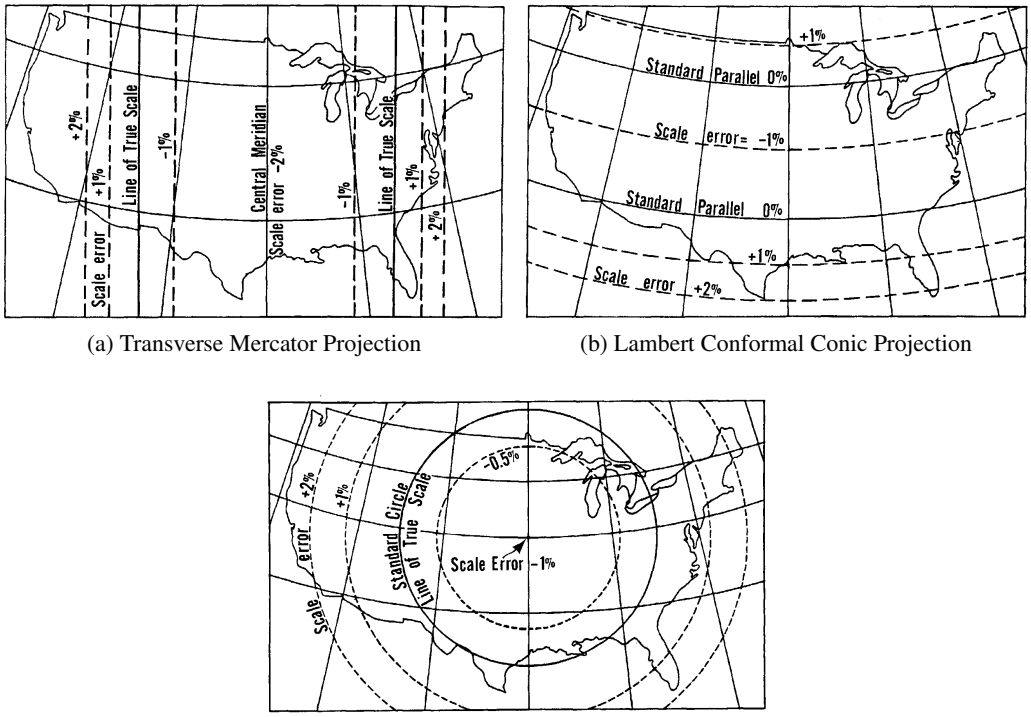
5.3 Are SRLG lists for Spherical and Planar Network Representation the Same?

In many works, regional failures are computed by transforming the geographical coordinates of an existing network into a plane, which introduces distortion. Depending both on the geographical area of the network and on the transforming procedure, this distortion can vary from negligible to significant. For example, the backbone network of a small-to-medium size country is not suffering a significant distortion when compared with the uncertainty of the available geographical data, but when turning to networks covering a large country, a continent, or even multiple continents, there is no projection which can hide the spherical-like geometry of the Earth's surface (see Fig. 5.10 taken from [96]). E.g., while the territory of continental US can be mapped onto a plane with 4% distortion [96], if we want to investigate bigger networks, clearly there is no projection that can hide the spherical-like geometry of the Earth.

There are reasons why one should analyze the global communication network as a whole: Electromagnetic storms induced by the Sun's Coronal Mass Ejections (CMEs) could cause severe simultaneous failures of electric and communication networks all over the Earth.

An important question is that, in practice, under which geographic extension of the network can one say that, in the viewpoint of SRLG enumeration, it is practically indifferent whether we consider a spherical or a planar representation of the network[8]. In other words, focusing now on lists M_r, the question is that under which size of the physical network will M_r^p and M_r^s (maximal link sets which can be hit by a single circular disk with radius r, in the plane and on the sphere, resp.) be precisely the same. The answer depends not only on the physical size but also on the characteristics

[8]Note that in case of disasters hitting a big fraction of the Earth's surface (similarly to the Carrington-event [97]), planar and spherical SRLG lists of fixed disaster shapes differ greatly [J5].



(a) Transverse Mercator Projection

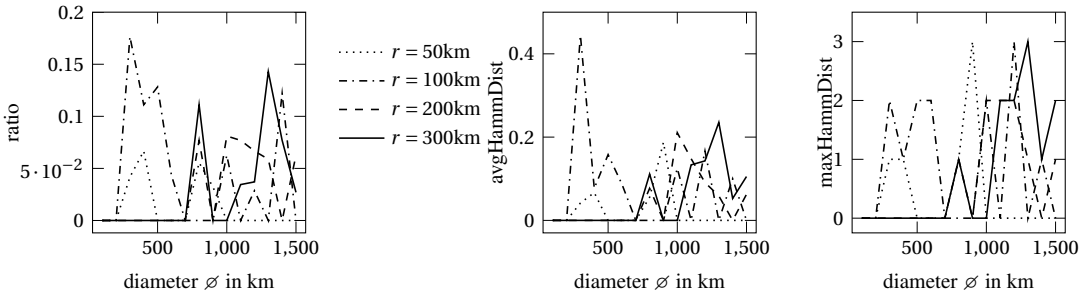
(b) Lambert Conformal Conic Projection

(c) Oblique Stereographic Projection

Figure 5.10: Distortion patterns on common conformal map projections. Projections are shown with a reduction in scale along the central meridian or at the center of projection, respectively. Each of the projections has > 3% scale error over the US. Picture taken from [96].

of the network itself: it can represent a dense metropolitan backbone network with multiple nodes close to each other, but it can also be geographically very sparse. Let τ be the distance of the closest non-adjacent and non-intersecting link of the network, and let D be the diameter of the smallest hitting disk of the network G. We can see that there can be any difference between M_r^s and M_r^p only if $2r \in [\tau, D]$ for either the spherical or the planar representation. Practical radii of circular disasters range from a couple of kilometers to a couple of hundreds of kilometers, which means they might be so small that there cannot be any difference between the SRLG sets (i.e., $2r < \tau$ means $M_r^s = M_r^p$). If τ is smaller than the disaster diameter, then it is easy to find settings, where $M_r^s \neq M_r^p$.

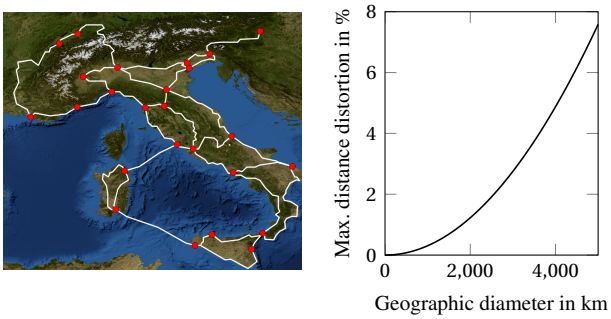
To study the phenomenon more in details, I used two similarity metrics of the SRLG lists: 1) the ratio of SRLGs, which are present in only one of M_r^p and M_r^s, i.e., $\mathscr{M}(r) := |M_r^p \triangle M_r^s| / (|M_r^p| + |M_r^s|) \in [0, 1]$, and 2) the average and maximal Hamming distance of an SRLG from M_r^s to its closest counterpart in M_r^p. I depicted the values of these metrics in Figures 5.11a-5.11c, respectively. As a base of evaluation, I took an Italian topology (Fig. 5.11d, with a diameter $D = 1180$km), and its magnified versions such that the resulting networks have diameters $D = 100, 200, \ldots, 1500$ km on the sphere. It can be seen that, in most cases, all of these metric values are 0 (i.e., $M_r^p = M_r^s$), but one



(a) Ratios of different SRLGs

(b) Average Hamming distance to closest SRLG

(c) Maximum Hamming distance to closest SRLG



(d) The Italian topology (diameter ≃ 1180km)

(e) Scale error of the Orthographic projection

Figure 5.11: Simulation results for comparing SRLG lists resulting from the planar and spherical representation of networks

can witness high spikes of big ratios of different SRLGs (Fig, 5.11a), or spherical SRLGs which have a symmetric difference of 3 links with their closest planar counterpart. This latter phenomenon happens when there are some nodes $u, v \in V$ such that $d(u, v) \leq 2r$ exactly in one of the spherical and planar representations.

The small and inconsistent ratio of different SRLGs in the two studied SRLG lists is due to the fact that though the Earth's surface is curved, this curvature is not practically significant in case of a backbone topology of a small to medium-size country. For example, the maximum distance distortion of the Orthographic projection over Hungary and Italy (having diameters < 530km and < 1250km) is $< 0.1\%$ and $< 0.5\%$, respectively (Fig. 5.11e). Even the contiguous US can be mapped with $< 4\%$ of distance distortion (Fig. 5.10c [96]).

Since the calculation time of M_r^s was approximately twice of the M_r^p in my experience, I concluded as follows. M_r^s and M_r^p can differ, thus it makes sense to compute the SRLG lists with the more precise spherical representation. However, in many of the cases, the distortion yielding from representing the network in the plane causes less inaccuracy than the lack of knowledge on the disaster characteristics, e.g., there can be as much as 10% inaccuracy in determining the disaster radius. In those cases, the planar representation can serve the purpose of SRLG listing well enough.

5.4 Thesis Summary

Thesis 1. *[C7, C10, J4, J5] I proposed polynomial algorithms for enumerating lists M_r^p and M_r^s of maximal link sets (SRLGs) which can be hit by a disaster overestimated by a shape of a circular disk with an arbitrary given radius r, in case of embedding the network in the Euclidean plane and on the sphere, respectively. I gave theoretical upper bounds on the cardinality of both M_r^p and M_r^s. I proved that the proposed algorithm for planar embeddings has a computational complexity which is tight in the number of network nodes. Finally, I compared the similarity of M_r^p and M_r^s in practice.*

Thesis 1.1. *[C10, J4] I proposed an algorithm, which, in case of representing a connected network topology $G(V,E)$ in the Euclidean plane with links considered as line segments, computes the list M_r^p of maximal link sets hit by a circular disk with radius r in $O\left((|V|+x)\left(\log|V|+\phi_r^2\rho_r^5\right)\right)$, where x is the number of link crossings, ρ_r is the maximum number of links which are hit by a circular disk with radius r, ρ_r is the maximum number of links which are hit by a circular disk with radius r, and finally, ϕ_r is the maximum number of nodes in the $3r$-neighborhood of a link. I proved that the complexity of the proposed algorithm is tight in $|V|$. I proved that the cardinality of M_r^p is $O\left((|V|+x)\,\rho_r\right)$, and that this bound is tight. I proved that $\left|\bigcup_{0<r'<r} M_{r'}^p\right|$ is $O((|V|+x)\rho_r^2)$.*

Thesis 1.2. *[C7, J5] I proposed a heuristic algorithm, which, considering a connected network topology $G(V,E)$ on a sphere with links considered as chains of geodesics, and considering a related sufficiently dense set $\mathscr{P}$ of disaster center points, computes list M_r^s of maximal link sets hit by a circular disk with radius r in $O(|\mathscr{P}|[(|V|+x)\gamma+|\mathscr{P}|\rho_r])$, where x is the number of link crossings, γ is the maximal number of geodesics a link stands of, and ρ_r is the maximum number of links which are hit by a circular disk with radius r. Through simulations, I showed that M_r^s and M_r^p can differ in practice, thus it is more precise to compute the SRLG lists with the spherical representation. However, in many of the cases, the distortion yielding from representing the network in the plane causes less inaccuracy than the lack of knowledge on the disaster characteristics. I concluded that, in those cases, the planar representation can serve the purpose of vulnerable region detection well enough.*

Chapter 6

Maximal SRLGs Caused by Circular Disks Hitting k Nodes

6.1 The Limited Geometric Information Failure Model - Informally

In resilient routing, the current best practice is to ensure that the primary and backup paths assigned to a connection are node disjoint. Compared to edge-disjointness, in this way operators ensure that the distance between the nodes of the primary and backup paths (except at the terminal nodes) are at least 1-hop-distance from each other. The intuitive reasoning is that a link in a backbone network is typically a few hundred kilometers long, while natural disasters are never larger than a few hundred kilometers. The root of the outages is usually because:

I) *close nodes* when two nodes are placed close to each other; for example, in highly populated areas. II) *parallel links* when two links are placed close to each other because of some geographic reasons; for example, they traverse the same bridge over a large river or cross a mountain range through the same valley.

We have the following design goals in defining the limited geometric information failure model.

- Do *not underestimate* the set of links involved in a possible regional failure. We believe the operator's damage in case of an unprotected regional failure is much greater than the extra cost of protecting networks against larger SRLGs.
- *Relative link distances* are given, the exact route of the cables are unknown, and the nodes are embedded in a schematic map.
- Provide a *fast and space efficient* way of calculating the set of SRLGs.

According to the first design goal, we deal with circular disk failures and define the size of the regional failure through the number of nodes it covers. Although the regional failures can have any

location, size, and shape, without any background information on the regional failures, it is a common practice to overestimate the size of the regional failure by ignoring its shape and rather focus on its radius only (see [98] and Chapter 5)[1]. According to our second design goal, the scaling of the topology map is not known, thus we cannot define a fixed maximum radius for the regional failures, but instead, we define a limit on the number of nodes interior to the circular disk.

Now we can define the limited geometric information failure model, which is based on the following assumptions:

- The network is a geometric graph $G(V, E)$ embedded in a 2D plane.
- The exact route of the conduits of the network links are not known, but contained by a polygonal region.
- The shape of the disaster is a circular disk with arbitrary radius and center position.
- We focus on *regional link k-node failures*, failures that hit k nodes for $k \in \{0, |V| - 2\}$.

The detailed model description can be found in Sec. 6.2. We argue this failure model can reasonably represent the possible regional failures, without actually requiring to know the scaling of the topology map.

Based on our output, operators can generate SRLG-disjoint primary and backup paths to protect the connection against natural disasters[2]. The distance between the primary and backup paths is a straightforward metric to compare the failure models. Based on the logical topology, the conventional approach to defining the distance is the hop-distance between the nodes traversed by the primary path and the nodes traversed by the backup path, except the terminal nodes. Based on this definition, we can list the failure models in increasing order of their strength (see Subsec. 6.6.3 for the proof):

- Single link failures ($\geq$ 0-hop-distance),
- Single node failures ($\geq$ 1-hop-distance),
- Single regional link 0-node failures,
- Single regional link 1-node failures,
- Single regional link 2-node failures, etc.

Note that, in our experiments with practical network topologies protecting against single regional link 0-node failures resulted in at least 2-hop-distance between the nodes of the primary and backup paths, except the terminal nodes. We believe the proposed approach well captures the possible regional network failures based on the little geographic information available at network devices.

[1] Of course, extreme overestimation of the failed link set should be avoided.

[2] The routing algorithms modify the SRLGs, whose failure isolates the source and destination nodes: those SRLGs are replaced with a smaller non-isolating SRLG according to the failure model.

6.2 Model and Assumptions

The network is modeled as an undirected connected geometric graph $G = (V, E)$ with $n = |V| \geq 3$ nodes and $m = |E|$ edges[3]. The nodes of the graph are embedded as points in the Euclidean plane, and their exact coordinates are considered to be known. In contrast to this, precise positions of edges are not known, instead we assume that for each edge e there is a *containing polygon* (or simply *polygon*) e^p in the plane in which the edge lies (see Fig. 6.2a). Parameter γ will be used to indicate the maximum number of sides a containing polygon e^p can have. Note that this model covers special cases when edges are considered as polygonal chains or line segments (thus at first reading, for ease, the reader may consider the edges as line segments and γ equals two).

For simplicity we assume that nodes of V and the corner points of the containing polygons defining the possible route of the edges are all situated in general positions of the plane, i.e., there are no three such points on the same line and no four on the same circle.[4]

We will often refer to circular disks simply as disks. The disk failure model will be adapted, which overestimates the area of a disaster such that all network elements that intersect the interior of a circle c may fail, and all other network elements are untouched. It is important to note that this modeling technique does not assume that the failed region has a shape of a disk, but overestimates the size of the failed region to have a tractable problem space.

Definition 6.2.1. *A circular disk failure c hits an edge e if the polygon of the edge e^p intersects the interior of disk c. Similarly node v is hit by failure c if it is in the interior of c. Let E_c (and V_c) denote the set of edges (and nodes, resp.) hit by a disk c.*

We emphasize that in this model when we say e is hit by c, it does not necessarily mean that e is destroyed indeed by c, instead, it means that there is a positive chance for e being destroyed.

Definition 6.2.2. *Let $\mathscr{C}$ denote the set of all circular disks in the plane, and let $\mathscr{C}_k \subseteq \mathscr{C}$ denote the set of those hitting exactly k nodes from V.*

Based on the above we can define the set of failure states the network may face after a disk failure hitting exactly k nodes.

Definition 6.2.3. *Let set $F(\mathscr{C}_k)$ denote the set of edge sets which can be hit by a disk $c \in \mathscr{C}_k$, and let $M_k = M(\mathscr{C}_k)$ denote the set of maximal edge sets in $F(\mathscr{C}_k)$.*

Note that for every $l \in \{0, \ldots, k-1\}$ and $f \in M_l$ there is an $f' \in M_{l+1}$ such that $f \subseteq f'$, because any disk hitting $l \leq k$ nodes could be overestimated by a disk hitting k nodes.

[3] Graph $G = (V, E)$ is not necessarily planar.

[4] All of the results of this chapter could be extended to geometric objects in non-general position, however this would complicate our arguments lowering the readability of the Thesis, while by an insignificant perturbation of the data one can make sure that the geometric objects are in general position.

Table 6.1: Table of symbols for Chp. 6

Notation	Meaning
	General
$G(V,E)$	the network modeled as an undirected connected geometric graph
n, m	the number of nodes $\|V\| \geq 3$ and edges $\|E\|$, respectively
e^p	the containing polygon of edge e (see Fig. 6.2a)
V_c, E_c	set of nodes and edges, resp., hit by a disk c
$\mathscr{C}$	the set of all circular disks in the plane
$\mathscr{C}_k$	the set of circular disks in the plane hitting exactly k nodes
$F(\mathscr{C}_k)$	the set of edge sets which can be hit by a disk $c \in \mathscr{C}_k$
M_k	the set of maximal edge sets in $F(\mathscr{C}_k)$
M_k^2	the set of maximal failures which can be hit by a disk from $\mathscr{C}_k$ having 2 nodes on its boundary
M_k^1	the set of maximal failures which can be caused by a half-plane having a node on its boundary hitting exactly k nodes
$\mathscr{C}_k^{u,v}$	the set of disks from $\mathscr{C}_k$ having nodes u, v on their boundary
$\mathscr{C}^{u,v}$	the set of disks from $\mathscr{C}$ having nodes u, v on their boundary
$M_k^{u,v}$	the the set of failures which contain exactly the elements of M_k that can be hit by a disk $c \in \mathscr{C}_k^{u,v}$
	Parameter
k	we are interested in circular disk shaped disasters hitting k nodes
γ	the maximum number of sides a containing polygon e^p can have
ρ_k	the maximum number of edges hit by a disk hitting k nodes.
	Apple
E_k	the set of node-pairs $\{u,v\} \subset V$ for which $\mathscr{C}_k^{u,v} \neq \emptyset$
$D_k(V,E_k)$	the k-Delaunay graph induced by node set V and edge set E_k
$c(x)$, h_+, h_-	For $u, v \in V$, a Cartesian coordinate system is placed in the plane such that line uv be identical to the vertical axis y, u and v have ordinates (y coordinates) 1 and -1, respectively (see Fig. 6.3). In this coordinate system, h_+ and h_+ are the right and left open half plane determined by line uv, respectively, and $c(x) = c(x,u,v)$ denotes the unique disk c in $\mathscr{C}^{u,v}$, which has centre point $(x,0)$, and $c(+\infty) := h_+$ and $c(-\infty) := h_-$, respectively.
$I_k^{u,v}$	the set of those numbers x, for which $c(x) \in \mathscr{C}_k^{u,v}$
x_{max}	the maximum of $I_k^{u,v}$, if exists, else if $I_k^{u,v} \neq \emptyset$, $x_{max} = +\infty$
x_{min}	the minimum of $I_k^{u,v}$, if exists, else if $I_k^{u,v} \neq \emptyset$, $x_{min} = -\infty$
$E_{c(x_{max})}$, $E_{c(x_{min})}$	the edge sets hit by $c(x_{max})$ and $c(x_{min})$, respectively
$\rhd$	for $u, v \in V$, the right side of $c(x_{max})$ cut by the vertical line uv
$\lhd$	for $u, v \in V$, the left side of $c(x_{min})$ cut by the vertical line uv
$x_+(e)$	for edge e, the leftmost disk which hits $e^p \cap \rhd$
$x_-(e)$	for edge e, the rightmost disk which hits $e^p \cap \lhd$
$x_+(v)$, $x_-(v)$	For nodes $w_+ \in h_+$ and $w_- \in h_-$, let $x_+(w_+)$ and $x_-(w_-)$ denote the abcissa of the centre point of circle going through u, v and w_+ or w_-, respectively.
E_+	the list of edges hit by $\rhd$ ordered descending by the x_+ values
E_-	the list of edges hit by $\lhd$ ordered descending by the x_- values
V_+	the list of nodes hit by $\rhd$ ordered descending by the x_+ values
V_-	the list of nodes hit by $\lhd$ ordered descending by the x_- values
V'_+	the list of nodes w in h_+ ordered decreasingly by the abscissa $x_+(z)$ of their leftmost hitting circles (going through u, v, z)
V'_-	the list of nodes z in h_- ordered decreasingly by the abscissa $x_-(z)$ of their rightmost hitting disk (going through u, v, z)
$A_k^{u,v}$	For an edge $\{u,v\} \in E_k$, apple $A_k^{u,v}$ is an ordered system $A_k^{u,v} = (V_+, V_-E_+, E_-)$. For each element of each list its appropriate $x_+()$ or $x_-()$ value is also stored.
A_k	the set of apples $A_k^{u,v}$
	Seesaw *(only in Subsec. 6.5.1)*
For data structure Seesaw for determining M_k^1, please check Subsec. 6.5.1.	

As mentioned before, only the maximal edge sets will be listed as SRLGs. This study aims to offer fast algorithms computing this list for various values of k, more precisely, throughout the chapter we will assume $k \in \{0, n-2\}$ since if a failure hits $n-1$ nodes, there is no node pair to communicate.

6.3 Algorithm for Enumerating Maximal Failures

In this section, a polynomial time algorithm is presented for computing M_k. The basic idea is that determining M_k can be decomposed into several simpler tasks, as illustrated in Fig. 6.1. Informally, after determining the so-called k-Delaunay graph (Def. 6.3.4 in Subsec. 6.3.2), data structures *apple* (Subsec. 6.3.3) and *seesaw* (Subsec. 6.5.1) are computed, and finally M_k is determined by merging lists $M_k^{u,v}$ and M_k^w resulting from querying the apples and seesaws.

6.3.1 Basic Observations

Our first observation is the following.

Claim 6.3.1. *For every $f \in M_k$ ($k \leq n-2$) there exists a disk $c \in \mathscr{C}_k$ such that f is hit by c, and c has at least one node of V on its boundary.*

Proof. Let f be hit by a disk $c_0 \in \mathscr{C}_k$ with centre point p. Since there are nodes of V not inside c_0, c_0 can be magnified from p until its boundary reaches a node u from V. This disk c_1 is also from $\mathscr{C}_k$ and has at least one point on its boundary, still hitting f. □

Disk c_1 described in the proof can be further magnified while keeping its center point on ray $[up$. Here we consider two cases: either there exists a node $v \in V$, which gets on the boundary while magnifying disk $c_1 \in \mathscr{C}_k$, or the open half plane h having p inside, u on the boundary, and having the normal vector $\vec{up}$ hits k nodes.

Definition 6.3.1. *Let M_k^2 be the set of maximal failures which can be hit by a disk from $\mathscr{C}_k$ nodes having 2 nodes on its boundary. Let M_k^1 be the set of maximal failures which can be caused by a half-plane having a node on its boundary hitting exactly k nodes.*

Proposition 6.3.2. *M_k is the set maximal sets in $M_k^1 \cup M_k^2$.* □

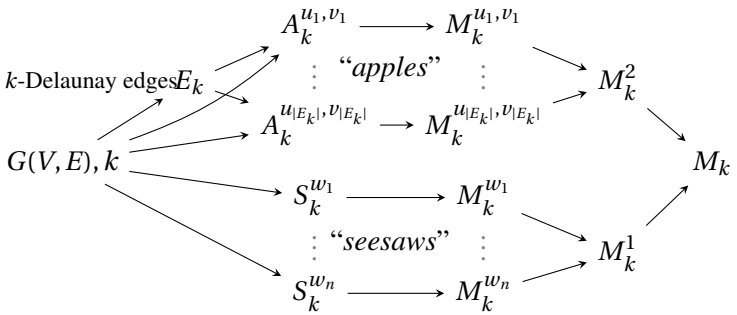


Figure 6.1: Visual sketch of Algorithm 5 for determining M_k

In the followings, we will present a way of computing M_k^2 in details using a data structure called *apple* defined in the studies this Thesis is based on. Determining M_k^1 can be done using similar ideas, thus we present it only briefly in Subsec. 6.5.1-6.5.2, where the same data structure called *seesaw* is defined.

6.3.2 Connection with the k-Delaunay Graphs

Definition 6.3.2. *For a node pair $u, v \in V$ let $\mathscr{C}_k^{u,v}$ denote the set of disks from $\mathscr{C}_k$ having nodes u, v on their boundary. Let $\mathscr{C}^{u,v}$ denote the set of disks from $\mathscr{C}$ having nodes u, v on their boundary. Let $M_k^{u,v} = M\left(\mathscr{C}_k^{u,v}\right)$ be the the set of failures which contain exactly the elements of M_k that can be hit by a disk $c \in \mathscr{C}_k^{u,v}$.*

Discussion after Claim 6.3.1 suggests the following simple method to compute M_k^2. First, for every node pair $\{u, v\} \subset V$, we compute a set of failures $M_k^{u,v}$.

Definition 6.3.3. *Let E_k denote the set of node-pairs $\{u, v\} \subset V$ for which $\mathscr{C}_k^{u,v} \neq \emptyset$.*

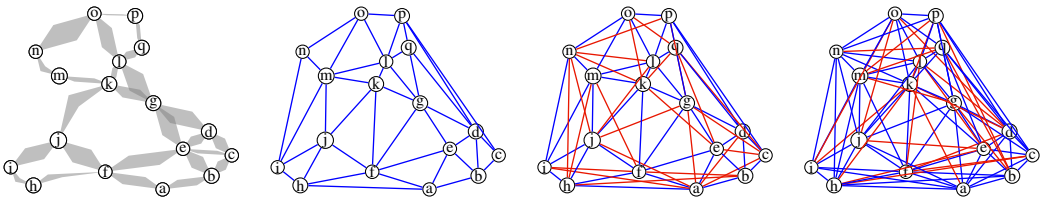
Fig. 6.2 shows an example of the input topology G and the corresponding set of node-pairs E_k for $k = 0, 1, 2$.

We can observe that by definition, M_k^2 can be computed by merging these sets $M_k^{u,v}$, formally M_k^2 is the set of maximal elements from the union of sets $M_k^{u,v}$.

Our second observation is that E_k is the edge set of the so-called k-Delaunay graph [99].

Definition 6.3.4. *Let $D_k = (V, E_k)$ denote the k-Delaunay graph induced by node set V and edge set E_k.*

In other words, a node pair $\{u, v\} \subseteq V$ is a k-Delaunay edge (i.e., $\{u, v\} \in E_k$) if there exists a circle through u and v that has at most k points of the node set V inside. The k-Delaunay graph D_k is a so-called *geometric proximity graph*. There is continuous research on a wide variety of geometric proximity graphs, where two vertices are connected by an edge if and only if the vertices



(a) Input graph $G = (V, E)$ with containing polygons, $n = 17$, $\gamma = 4$.
(b) Graph $D_0(V, E_0)$, $|E_0| = 40$
(c) Graph $D_1(V, E_1)$, $|E_1| = 66$
(d) Graph $D_2(V, E_2)$, $|E_2| = 89$

Figure 6.2: Input topology and k-Delaunay graphs D_k for $k = 0, 1, 2$.

satisfy particular geometric requirements. For example, in [80] the construction time of the k-Gabriel graph is studied, which is known to be a subgraph of D_k. In that paper, the k-Delaunay graph is determined as an intermediate step from the k-Voronoi diagram while determining the k-Gabriel graph in polynomial time. From a theorem in [81] an upper bound on $|E_k|$ can be derived. These statements are the following:

Theorem 6.3.3 (Thm. 2.4 of [80]). *Graph k-Delaunay $D_k = (V, E_k)$ can be constructed in time complexity $O\left(\left(k^2+1\right) n \log n\right)$.*

Theorem 6.3.4 (Thm. 2 of [81]). $|E_k| \leq 3(k+1)n - 3(k+1)(k+2)$.

These theorems give that for small values of k graph D_k is sparse (in other words, $\mathscr{C}_k^{u,v} = \emptyset$ for most node pairs u, v), and it can be computed fast.

6.3.3 Data Structure Apple

Let node pair $\{u, v\} \in E_k$ be given. Let us place a Cartesian coordinate system in the plane such that line uv be identical to the vertical axis y, u and v have ordinates (y coordinates) 1 and -1, respectively (see Fig. 6.3). Obviously, this way the centre point of any disk $c \in \mathscr{C}_k^{u,v}$ has ordinate 0.

Definition 6.3.5. *For a given node pair $\{u, v\}$, the previously described coordinate system and real number x, let $c(x)$ denote the unique disk c in $\mathscr{C}^{u,v}$, which has centre point $(x, 0)$.*

Trivially, $c()$ is a bijective function between $\mathbb{R}$ and $\mathscr{C}^{u,v}$.

Let $I_k^{u,v}$ denote the set of those numbers x, for which $c(x) \in \mathscr{C}_k^{u,v}$. If $\mathscr{C}_k^{u,v}$ is empty, then trivially $I_k^{u,v}$ is empty too. In the case when $\mathscr{C}_k^{u,v}$ is not empty, we can observe that $I_k^{u,v}$ is the union of closed intervals.

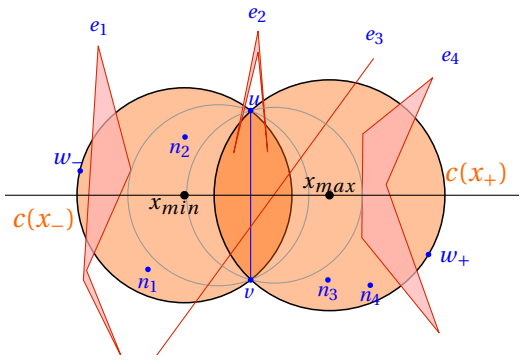


Figure 6.3: Illustration of an apple with $k = 2$. Apple $A_k^{u,v}$ consists of ordered lists of nodes V_+ and V_- and ordered lists of edges E_+ and E_-, where $V_+ = \{n_4, n_3\}$, $V_- = \{n_2, n_1\}$, $E_+ = \{e_4, e_3, e_2\}$ and $E_- = \{e_3, e_2, e_1\}$. Given $G = (V, E)$ and $\{u, v\} \subseteq V$, $A_k^{u,v}$ can be determined in $O\left(n\rho_0\gamma + k \log k + \rho_k \log \rho_k\right)$ (proof of Lemma 6.4.4). By *querying* $A_k^{u,v}$, $M_k^{u,v}$ can be computed in $O\left(\rho_k^3\right)$ (proof of Lemma 6.4.6).

If the number of nodes in both half planes determined by line uv is not equal to k, then there exists a rightmost and a leftmost element of $\mathscr{C}_k^{u,v}$, i.e. there is a maximum x_{max} and minimum x_{min} real number in $I_k^{u,v}$, that is $x_{max} = \max I_k^{u,v}$, and $x_{min} = \min I_k^{u,v}$. Disks $c(x_{max})$ and $c(x_{min})$ have a third node w_+ and w_- on their boundary, respectively. If there are exactly k nodes on the right side of uv, then let h_+ be the right open half-plane determined by line uv, and let x_{max} be sufficiently large to $E_{c(x_{max})}$ to contain all the edges having polygon having a point with positive abscissa. For simplicity, sometimes $x_{max} = +\infty$ and $c(x_{max}) = h_+$ is used. The same applies to the left side of uv.

Let $E_{c(x_{max})}$ and $E_{c(x_{min})}$ denote the edge sets hit by $c(x_{max})$ and $c(x_{min})$, respectively. To compute $M_k^{u,v}$ we use the following observation.

Claim 6.3.5. *For all $f \in F(\mathscr{C}_k^{u,v})$, $f \subseteq E_{c(x_{max})} \cup E_{c(x_{min})}$.*

Proof. It is easy to see that for every disk $c \in \mathscr{C}_k^{u,v}$, $c \subseteq c(x_{max}) \cup c(x_{min})$. □

According to Claim 6.3.5, a first step towards computing $M_k^{u,v}$ is to determine the edge sets hit by $c(x_{max})$ and $c(x_{min})$. Trivially, this can be done in $O(m\gamma)$. The remaining question is how to calculate $M_k^{u,v}$ from $E_{c(x_{max})} \cup E_{c(x_{min})}$. Some additional notations and definitions precede the presentation of the solution.

Let $\rhd$ denote the right side of disk $c(x_{max})$ cut by the vertical line uv, and let $\lhd$ denote the left side of disk $c(x_{min})$ cut by the vertical line uv. For each edge $e \in E_{c(x_{max})} \cup E_{c(x_{min})}$ we will compute two disks: the leftmost disk which hits $e^p \cap \rhd$ and the rightmost disk which hits $e^p \cap \lhd$, which have centre points $x_+(e)$ and $x_-(e)$ respectively.

Let E_+ denote the list of edges hit by $\rhd$, and similarly, let E_- be the list of edges hit by $\lhd$. Thus, we have $E_+ \subseteq E_{c(x_{max})}$ and $E_- \subseteq E_{c(x_{min})}$, and also $E_+ \cup E_- = E_{c(x_{max})} \cup E_{c(x_{min})}$.

Let E_+ and E_- be ordered descending by the x_+ and x_- values of their elements, respectively. Note that according to Claim 6.6.1 from Subsec. 6.6.1, both $x_+(e)$ and $x_-(e)$ can be computed in $O(\gamma)$.

For nodes $w_+ \in h_+$ and $w_- \in h_-$, let $x_+(w_+)$ and $x_-(w_-)$ denote the abscissa of the centre point of circle going through u, v and w_+ or w_-, respectively. We introduce V_+ and V_- similarly to E_+ and E_-, but instead of edges we store nodes hit by $\rhd$ ordered descending by their x_+ values, while in V_- nodes in $\lhd$ are stored ordered also descending, but by their x_- values. Trivially, for a node $v \in V$ both $x_+(v)$ and $x_-(v)$ can be determined in $O(1)$.

Note that while every node $v \in V$ is part of at most one of lists V_+ and V_-, edges can be part of both E_+ and E_-.

Now we can define the data structure *apple* for each edge of the k-Delaunay graph.

Definition 6.3.6. *For an edge $\{u, v\} \in E_k$, apple $A_k^{u,v}$ is an ordered system $A_k^{u,v} = (V_+, V_- E_+, E_-)$, where its composing lists are as described in the subsection before. For each element l of each list we also store its appropriate $x_+(l)$ or $x_-(l)$ value.*

6.3.4 Sweep Disk Algorithms

6.3.4.1 Concept

In this subsection, we introduce the paradigm of sweep disk algorithms, which is similar to the algorithmic paradigm of sweep line (sweep surface) algorithms in computational geometry, briefly presented in Chapter 4.

Our sweep disk algorithms will scan through disk sets $\mathscr{C}^{u,v}$. In this sense, in contrast to the sweep surface paradigm, our disks have different diameters, and instead of keeping orientation, the invariant will be that all disks have u and v on the boundary. Thus our disk to sweep is "elastic," in the sense that it can change its diameter, but not its shape.

6.3.4.2 Example

Our first sweep disk algorithm is used for determining x_{max} and x_{min} for a given $A_k^{u,v}$. The algorithm works as follows. Starting from a disk $c(x) \in \mathscr{C}^{u,v}$ having centre point with abscissa $x = +\infty$ (or sufficiently large), c is swept throughout the elements of $\mathscr{C}^{u,v}$ until $x = -\infty$ (or sufficiently small). Meanwhile the number of nodes hit is followed. Numbers x_{max} and x_{min} can be determined at the first and last state when c hits exactly k nodes, respectively. (Non-existence of such moments would mean that $\{u, v\} \notin E_k$.)

Technically this can be done as follows. Let $V'_+ \subseteq V$ be the list of nodes w from h_+ ordered decreasingly by the abscissa $x_+(z)$ of their leftmost hitting circles (going through u, v, z). Similarly, let V'_- be the list of nodes z in h_- ordered decreasingly by the abscissa $x_-(z)$ of their rightmost hitting disk (going through u, v, z). Applying the fact that a node pair $z_+ \in V'_+$ and $z_- \in V'_-$ can be hit by the same disk $c \in \mathscr{C}^{u,v}$ iff $x_+(z_+) \geq x_-(z_-)$, sweeping can be imitated as in Algorithm 3. Note that for every node z in V'_+ or V'_-, $x_+(z)$ or $x_-(z)$ is stored as part of function x_+ or x_-.

From the following Proposition 6.3.6, one can check that the number of hit nodes can be easily followed with the help of an additional variable.

Proposition 6.3.6. *Let $c \in \mathscr{C}^{u,v}$. If $V'_+[i-1]$ is not hit by c, then all the preceding elements in V'_+ are not hit by c. If $V'_+[i]$ is hit by c, then all the following elements in V'_+ are hit by c.*

Similarly, if $V'_-[i-1]$ is hit by c, then all the preceding elements are hit by c. If $V'_-[i]$ is not hit by c, then all the following elements are not hit by c. □

Claim 6.3.7. *For a given edge $\{u, v\} \in E_k$, both x_{max} and x_{min} can be determined in $O(n \log n)$.*

Proof. According to those written in this subsection, both V'_+ and V'_- can be determined in $O(n \log n)$ the dominant step being a sorting algorithm. Sweeping can be trivially done in $O(n)$; meanwhile, both x_{max} and x_{min} can be determined. □

Algorithm 3: Determining x_{max} and x_{min} while sweeping through $\mathscr{C}^{u,v}$

Input: V and $u, v \in V$
Output: x_{max} and x_{min}
begin
1 Compute ordered lists V'_+ and V'_-
2 Merge V'_+ and V'_- into descending ordered list $V'_\pm$ using values x_+ for V_+ and x_- for V_-
3 $n_+, n_- \leftarrow 0$
4 **for** $l \in \{1, \ldots, |V'_\pm|\}$ **do**
5 **if** $V'_\pm[l] \in V'_+$ **then** $n_+ \mathrel{+}= 1$
else $n_- \mathrel{+}= 1$
6 $\#_l := |V'_+| - n_+ + n_-$ `// # currently hit nodes`
7 **if** $|V'_+| = k$ **then** $z_+ = v_\emptyset$; $x_{max} = +\infty$
else $w_+ := V'_\pm[\min l : \#_l = k]$; $x_{max} := x_+(w_+)$
8 **if** $|V'_-| = k$ **then** $z_- = v_\emptyset$; $x_{min} = -\infty$
else $w_- := V'_\pm[\max l : \#_l = k]$; $x_{min} := x_-(w_-)$
9 **return** x_{max} and x_{min}

Proposition 6.3.8. *Both V_+ and V_- can be determined in $O(n \log n)$.* □

6.3.5 Determining Apples

Claim 6.3.9. *For a given $\{u, v\} \in E_k$, apple $A_k^{u,v}$ can be determined in $O(m(\log m + \gamma))$.*

Proof. If $\{u, v\} \in E_k$, then x_{max}, x_{min}, V_+ and V_- can be determined in $O(n \log n)$ according to Claim 6.3.7 and Prop. 6.3.8, thus it remains to determine E_+ and E_-. With this aim it is enough to compute the $x_+(e)$ and $x_-(e)$ values for every edge, then collect in E_+ those edges e for which $x_+(e) \leq x_{max}$ and similarly in E_- those edges e for which $x_-(e) \geq x_{min}$. Finally, edges in E_+ and E_- have to be sorted descending according to their x_+ and x_- values, respectively. m polygons of edges (each having at most γ sides) have to be checked and sorted which gives a total complexity of $O(m(\gamma + \log m))$. □

Definition 6.3.7. *Let A_k be the set of apples $A_k^{u,v}$.*

Corollary 6.3.10. *For a given k, knowing E_k, the set of apples A_k can be determined in time complexity $O((k+1)nm(\gamma + \log m))$.*

Proof. Since by Thm. 6.3.4 $|E_k| < 3(k+1)n$, we deduct that $O((k+1)n)$ apples have to be determined. According to Claim 6.3.7, an apple can be constructed in $O(m(\log m + \gamma))$, which completes the proof. □

6.3.6 Computing the Set of SRLGs by Sweeping Through Each Apple

Claim 6.3.11. *Let $e \in E_+$, and $f \in E_-$. They can be hit by the same $c \in \mathscr{C}_k^{u,v}$ if $x_+(e) \leq x_-(f)$ or $x_+(f) \leq x_-(e)$.*

Algorithm 4: Querying an apple

Input: Apple $A_k^{u,v}$ **Output:** Set $M_k^{u,v}$ of locally maximal failures. **begin**

Merge V_+, V_-, E_+ and E_+ into descending ordered list G using values x_+ for V_+ and E_+ and x_- for V_- and E_-;
$n_+, n_-, e_+, e_- \leftarrow 0$;
for $l \in \{1, \ldots, |G|\}$ **do**
  if $l \in V_+$ **then** n_++=1;
  if $l \in V_-$ **then** n_-+=1;
  if $l \in E_+$ **then** e_++=1;
  if $l \in E_-$ **then** e_-+=1;
  $\#_{n,l} := |V_+| - n_+ + n_-$ `// # curr. hit nodes`
  $\#_{e,l} := |E_+| - e_+ + e_-$ `// # curr. hit edges`
  $e_{+,l} := e_+$ `//` $E_+[i]$ `is hit iff` $i \geq e_{+,l}$
  $e_{-,l} := e_-$ `//` $E_-[i]$ `is hit iff` $i \leq e_{-,l}$
Det. L, the set of indexes l, for which $\#_{n,l} = k$;
Det. I_e, the sequence of numbers $\#_{e,l} : l \in L$;
Det. M_e, the set of indexes l of local maximums of I_e;
$(M_k^{u,v})' :=$ hit edge sets in $l \in M_e$ disk positions;
`// Can be det. using` $E_+, E_-, e_{+,l}, e_{-,l}$
$M_k^{u,v} \leftarrow$ maximal elements of $(M_k^{u,v})'$;
return $M_k^{u,v}$

Proof. An edge e can be hit by circle $c(x)$ if $x_+(e) \leq x$ or $x \leq x_-(e)$. □

Determining $M_k^{u,v}$ from apple $A_k^{u,v}$ can be done with the help of a sweep disk algorithm as a subroutine of Algorithm 4 similar to Algorithm 3, the only difference is that here we have to check both the set of currently hit edges and the number of currently hit nodes at the same time.

On the one hand, while sweeping through $\mathscr{C}^{u,v}$ with $c(x)$ (while x decreasing), nodes are also getting hit or not hit by the actual $c(x)$, thus it is not necessarily permanently part of $\mathscr{C}_k^{u,v}$ during the sweep disk algorithm. On the other hand, any edge e having e^p intersecting segment $[u, v]$ or for which $x_+(e) \leq x_-(e)$ should be stored exactly once in any element of $M_k^{u,v}$.

Claim 6.3.12. *Querying $A_k^{u,v}$, Algorithm 4 calculates $M_k^{u,v}$ in $O(m^3)$.*

Proof. The correctness of the algorithm can be easily checked. Since while sweeping an edge can get hit or unhit at most once on one side of line uv, there are at most $O(m)$ failures with locally maximal cardinalities, each of them having $O(m)$ edges, thus $(M_k^{u,v})'$ has $O(m)$ elements of $O(m)$ size. Trivially, the number of currently hitting nodes can be monitored in $O(n)$ total time as in Alg. 3. Since we have an ordering of the edges, every pair of sets from $(M_k^{u,v})'$ can be compared in $O(m)$. This means that from $(M_k^{u,v})'$, $M_k^{u,v}$ can be determined in $O(m^3)$. It can be checked that all the other operations have complexity at most $O(m^3)$. □

Corollary 6.3.13. *Known E_k, lists $M_k^{u,v}$ for all $\{u, v\} \in E_k$, can be determined in $O\big((k+1)\, nm^3\big)$.* □

Algorithm 5: Algorithm for computing M_k with a table on time complexities. (Refer to Table 6.1 for notations.)

Input: $G(V,E)$, k **Output:** M_k **begin**

1 Determining M_k^2;
2 Determine E_k;
3 Determine set A_k of nonempty apples;
4 Query apples from A_k;
5 Merge lists $M_k^{u,v}$ into M_k^2;
6 Determining M_k^1 `// See Sec. 6.5 for details`
7 Determine set S_k of nonempty seesaws;
8 Query seesaws from S_k;
9 Merge lists M_k^w into M_k^1;
10 Merge lists M_k^2 and M_k^1 into M_k;
11 **return** M_k

Complexity	Non-parametrized	Parametrized
STEP 1	$O((k^2+1)n\log n)$	$O((k^2+1)n\log n)$
STEP 2	$O(n(k+1)m(\log m+\gamma))$	$O\big((k+1)\,n\big(n\rho_0\gamma+k\log k+\rho_k\log\rho_k\big)\big)$
STEP 3	$O(n(k+1)m^3)$	$O(n(k+1)\rho_k^3)$
STEP 4	$O(n^2(k^2+1)m^3)$	$O(n^2(k^2+1)\rho_k^3)$
STEP 5	$O(nm\gamma\log m)$	$O(n^2\rho_0\gamma\log(n\rho_0))$
STEP 6	$O(nm^3)$	$O(n\rho_0+\rho_k^3)$
STEP 7	$O(n^2m^3)$	$O(n^2\rho_k^3)$
STEP 8	$O(n^2(k+1)m^3)$	$O(n^2(k+1)\rho_k^3)$
Total for $\gamma=O(1)$	$O(n^2(k^2+1)m^3)$	$O(n^2((k^2+1)\rho_k^3+\rho_0\log(n\rho_0)))$

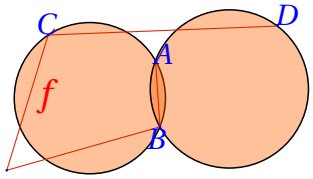


Figure 6.4: In the setting above, circle c_{ABC} hits the whole link set E, while no $c \in \mathscr{C}_0^{B,D}$ hits link f. Thus, however $M_0^{B,D}$ is not empty, it does not contain any elements of M_0.

Here we use the assumption that the corner points of the polygons of the edges are in general position.

6.3.7 Algorithm for Computing Maximal Failures

As presented before M_k^2 can be calculated by determining and querying the apples, and finally merging the obtained lists $M_k^{u,v}$ in M_k^2. Note that there is a valid need of comparing lists of locally maximal failures (see Fig. 6.4). M_k^1 can be computed very similarly to M_k^2 (as shown in Subsec. 6.5.1-6.5.2). Finally, in order to get M_k, M_k^2 and M_k^1 have to be merged. This way the scheme of our algorithm could be written as in Alg. 5.

Complexity bounds on the non-parametrized computing time and length of M_k are summarized as part of the table in Alg. 5. Although in Subsec. 6.6.2 it is shown that there are some artificial networks where these asymptotic bounds are relatively good estimations, we would like to focus on the running time and output size on the real networks, which are nearly planar. Thus, after introducing a new parameter, we present parametrized bounds proven in Subsec. 6.4.1-6.5.2.

Intuitively, a $c \in \mathscr{C}_k$ cannot hit too many edges. Thus we introduce graph density parameter ρ_k, which describes this phenomenon.

Definition 6.3.8. *For all $i \in \{0, n-2\}$, let ρ_i be the maximum number of edges hit by a disk from $\mathscr{C}_i$.*

The parametrized bounds are the following:

Lemma 6.3.14. *M_k^2 can be computed in $O\left(n^2\left(\left(k^2+1\right)\rho_k^3 + (k+1)\left(\rho_0\gamma\right)\right)\right)$. M_k^2 has $O\left(n\left(k+1\right)\rho_k\right)$ elements with at most ρ_k edges.*

Proof of Lemma 6.3.14 can be found in Subsec. 6.4.4.

Besides computing M_k^2, one have to deal with computing M_k^1. When computing M_k^1, the vague idea is to give a geometric algorithm in a way similar to the sweep disk algorithm for querying the apples. Now instead of imaginary sweeping a disk, we rotate a half-plane around every node $v \in V$ until it makes a total turn, and meanwhile, check for hit edge sets with locally maximal cardinalities hit by half-planes hitting exactly k nodes. After this, the maximal elements of the obtained lists are collected in M_k^1. Now follows Lemma 6.3.15 for computing M_k^1. We kindly refer the reader to Subsec. 6.5.1-6.5.2 for its detailed proof.

Lemma 6.3.15. *M_k^1 can be constructed in $O(n^2(\rho_0\gamma\log n\rho_0 + \rho_k^3))$ and has $O(n\rho_k)$ elements, each containing at most ρ_k edges.*

Theorem 6.3.16. *M_k can be computed in $O\left(n^2\left(\left(k^2+1\right)\rho_k^3 + \rho_k\gamma + \left(k+1+\log(n\rho_0)\right)\rho_0\gamma\right)\right)$. M_k has $O\left(n\left(k+1\right)\rho_k\right)$ elements with at most ρ_k edges.*

Proof. Based on Lemmas 6.3.14 and 6.3.15, both M_k^2 and M_k^1 can be computed in the proposed time, have at most the proposed amount of elements containing at most ρ_k edges. The proof will be completed by showing that the merger of M_k^2 and M_k^1 can be done in $O(n^2(k+1)\rho_k^3)$, which is true because of the following. We only have to compare all the pairs $\{p_2, p_1\}$ made up of a $p_2 \in M_k^2$ and $p_1 \in M_k^1$, which means $O((n(k+1)\rho_k)(n\rho_k))$ pairs. Each comparision can be made in $O(\rho_k)$,[5] which gives the proposed complexity. □

Corollary 6.3.17. *If ρ_k is $O(k+1)$, then M_k has $O\left(n\left(k+1\right)\right)$ elements. If in addition γ is upper bounded by a constant, M_k can be computed in $O(n^2(k^5+\log(nk+1)+1))$.*[6]

6.4 Parametrized Complexity Bounds for Enumerating SRLGs of M_k^2

6.4.1 Parametrized Complexity Bounds for Determining Apples

Up to this point, the fact that $G(V,E)$ is, in fact, a graph of a communication network, and thus it is 'almost planar' was not used. Intuitively, an almost planar graph has $O(n)$ edges. Parameters ρ_i

[5] Since link sets are ordered lexicographically.

[6] To be exact, $\gamma = O(\max\{1, \frac{k^5+1}{k+\log n}\})$ still yields the proposed complexity.

($i \in \{0,\dots,n-2\}$) denote the maximum number of edges hit by a disk from $\mathscr{C}_i$ (Def. 6.3.8), thus they help in formalizing this intuition.

Since parameters ρ_i measure local properties of the networks, often it will be assumed that these parameters are not exceeding a constant. For example, ρ_0 is not going to be large since physically close edges are likely to be connected through a node.

Observation 6.4.1. *For any $0 \le i < j \le n-2$, $\rho_i \le \rho_j$.* □

Claim 6.4.2. *In any apple $A_k^{u,v} \in A_k$, $|E_+| \le \rho_k$ and $|E_-| \le \rho_k$.*

Proof. All edges in E_+ are hit by $c(x_{max})$ which by definition hits at most ρ_k edges. Similar for E_-. □

Lemma 6.4.3. *The number of edges is $O(n\rho_0)$, more precisely $m \le (2n-5)\rho_0$.*

Proof. Consider the Delaunay triangulation D_0, which is a planar graph, and thus $|E_0| \le 3n-6$. Since every Delunay triangle has 3 Delaunay edges and a Delaunay edge is the edge of at most 2 Delaunay triangles, and there are at least 3 Delaunay edges on the convex hull of V, the number of Delaunay triangles is at most

$$\frac{2|E_0|-3}{3} \le \frac{2}{3}(3n-6)-1 = 2n-5.$$

Since the polygon of every edge in E intersects at least one triangle, and every triangle can be covered by a disk $c \in \mathscr{C}_0$, which intersects at most ρ_0 polygon of edges of the network, we get that the number m of edges cannot be larger than ρ_0 times the number of the Delaunay triangles. We get $m \le (2n-5)\rho_0$. □

Lemma 6.4.4. *If E_k is given, set A_k of apples can be built in $O\big((k+1)\,n\big(n\rho_0\gamma + k\log k + \rho_k \log \rho_k\big)\big)$.*

Proof. There are $|E_k| \le 3(k+1)n$ apples to determine (Thm. 6.3.4). For each, V_+ and V_- can be determined in $O(n + k\log k)$, then (based on Lemma 6.4.3) $O(n\rho_0)$ edges have to be checked if they are in the apple, each in $O(\gamma)$ time. After this, based on Claim 6.4.2, there are $O(\rho_k)$ edges to order, which gives the proposed complexity. □

Corollary 6.4.5. *If ρ_k and γ is upper bounded by a constant and E_k is given, then A_k can be determined in $O\big(n^2\,(k+\log n)\big)$.* □

6.4.2 Parametrized Bound for Determining $M_k^{u,v}$ for Apple $A_k^{u,v}$

Lemma 6.4.6. *For all the apples $A_k^{u,v}$ in A_k, sets $M_k^{u,v}$ can be determined in $O\big(n\,(k+1)\,\rho_k^3\big)$.*

Proof. Based on Thm. 6.3.4, there are $|E_k| \le 3n(k+1)$ apples to query. We claim that each of them can be queried in $O(\rho_k^3)$. Knowing apple $A_k^{u,v}$, $(M_k^{u,v})'$ can be determined in $O(\rho_k^2)$, following the

steps of Alg. 4. After this, $M_k^{u,v}$ can be determined by comparing each pair of elements of $(M_k^{u,v})'$ and eliminating its non-maximal and redundant members. With this purpose, $O(\rho_k^2)$ comparisons have to be made, each of them has $O(\rho_k)$ complexity. □

6.4.3 Parametrized Complexity Bound on Merging Lists $M_k^{u,v}$

Lemma 6.4.7. *M_k^2 can be computed in $O\left(n^2\left(k^2+1\right)\rho_k^3\right)$ from lists $M_k^{u,v}$.*

Proof. By Thm. 6.3.4, there are $|E_k| \leq 3n(k+1)$ lists containing $O(\rho_k)$ sets containing $O(\rho_k)$ edges. First determine an ordering on the set of edges E, and sort all the candidate sets of edges according to this ordering, each in $O(\rho_k \log \rho_k)$ time. M_k^2 can be computed by comparing all the set pairs (and eliminating the redundant or non-maximal elements), which means $O(n^2(k^2+1)\rho_k^2)$ comparisons. Since (due to the ordering) comparing two sets takes $O(\rho_k)$ time, the total complexity is $O(n^2(k^2+1)\rho_k^3)$. □

6.4.4 Parametrized Complexity for Computing M_k^2

Below is the proof of Lemma 6.3.14 stating that M_k can be computed in $O\left(n^2\left(\left(k^2+1\right)\rho_k^3+(k+1)\,\rho_0\gamma\right)\right)$ and M_k has $O\left(n\,(k+1)\,\rho_k\right)$ elements with at most ρ_k edges.

Proof. As presented previously (Thm. 6.3.3 and Lemmas 6.4.4, 6.4.6 and 6.4.7), each of the corresponding four phases of Alg. 5 can be examined in the proposed complexity. There are $|E_k| \leq 3n(k+1)$ lists $M_k^{u,v}$ to merge, each of them has at most ρ_k edges, completing the proof. □

6.5 Parametrized Algorithm for Enumerating SRLGs in M_k^1

In this section a sketch of an algorithm will be presented for proving Lemma 6.3.15, which states that M_k^1 can be constructed in $O\left(n^2\left(\rho_0\gamma\log\left(n\rho_0\right)+\rho_k^3\right)\right)$ and has $O\left(n\rho_k\right)$ elements, each containing $O\left(\rho_k\right)$ edges.

6.5.1 Data Structure Seesaw

For every node $w \in V$, a data structure is built containing both the direction of every node $z \in V \setminus \{w\}$ related to v and the interval of directions where the polygon of each edge can be seen from w. Nodes and edges also have to be sorted according to this information (similarly to data structure apple). Let us call this data structure *seesaw*, and let S_k^w denote the previously described seesaw. For a given k, let the set of seesaws be denoted by S_k. Let the list of locally maximal failures resulting from querying seesaw S_k^w be denoted by M_k^w.

Claim 6.5.1. *Any seesaw S_k^w can be calculated in $O\big(n\rho_0\gamma\log\big(n\rho_0\big)\big)$ and has a total length of $O\big(n\rho_0\big)$.*

Proof. Trivially, the direction of nodes from $z \in V \setminus \{w\}$ can be determined and sorted in $O(n\log n)$. Also, as by Lemma 6.4.3, the number of edges is $O(n\rho_0)$, intervals of directions corresponding to polygons of edges can be calculated and sorted both by their minimum and maximum values on $O(n\rho_0\gamma\log(n\rho_0))$. The proof follows. □

6.5.2 Querying Seesaws

Claim 6.5.2. *M_k^w can be determined from S_k^w in $O\big(n\rho_0+\rho_k^3\big)$, and has at most $2\rho_k$ elements, each of them containing at most ρ_k edges.*

Proof. Since each set in M_k^w can be hit by a disk in $\mathscr{C}_k$ having w on the boundary, the fact that the elements of M_k^w contain at most ρ_k edges is trivial by definition.

Now we prove $|M_k^w| \le 2\rho_k$. If there is no half-plane having w on the boundary hitting exactly k nodes, then the claim is trivial, otherwise, let h_0^w be such a half-plane. Let h_+^w be the unique half-plane which satisfies the followings: it is the rotation of h_0^w with $\angle \le \pi$, it covers exactly k nodes, and no other half-plane covers exactly k nodes which is the rotation of h_+^w with an angle $\in [\angle, \pi]$. Let h_-^w be similar, but with rotating towards the negative direction. This way, every half-plane going through w covering exactly k nodes is part of $h_-^w \cup h_+^w$, which altogether hit at most $2\rho_k$ edges. Since while turning a half-plane around w, edges are getting hit or not hit one by one, and an edge at most 2 times, there can exist at most $2\rho_k$ hit edge sets with locally maximal cardinalities.

Determining M_k^w can be done by simply turning the half-plane around w. Checking currently hit edge sets, node sets and cardinalities can be done in a total $O(n\rho_0)$ time, list M_k^w can be created in $O(\rho_k^3)$. □

Proof of Lemma 6.3.15: Trivially, every element of M_k^1 contains at most ρ_k elements. Based on Claims 6.5.1 and 6.5.2, lists M_k^w can be determined in $O(n^2\rho_0\gamma\log(n\rho_0)+n\rho_k^3)$ for all $w \in V$. M_k^1 can be calculated by merging the previous lists. $O(n^2\rho_k^2)$ comparisons have to be done, each has a complexity of $O(\rho_k)$. This means a total complexity of $O(n^2(\rho_0\gamma\log(n\rho_0)+\rho_k^3))$, completing the proof. □

6.5.3 Connection Between Seesaws and Apples

It is known that the inverse of the stereographic projection from the North Pole, which maps geometric objects from the plane to the surface of the sphere, has the property that the image of lines and line segments on the plane are great circles and geodesics on the sphere, respectively (corollary of Theorem in [100]). Known this, if we projected our network topology (back) to a spherical surface,

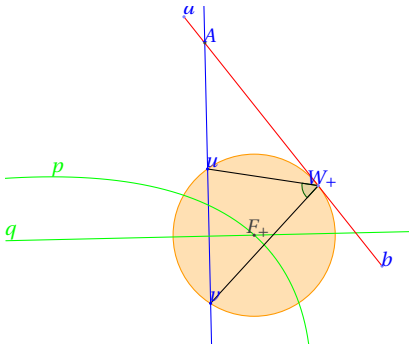


Figure 6.5: Illustration for Subsec. 6.6.1

and if we defined the spherical apple and spherical seesaw data structure, we could see that a spherical seesaw M_k^w is just a spherical apple $M_k^{w,z}$, z being the North Pole, thus sweeping through a spherical apple is the same thing as tilting a spherical seesaw, meaning that on the sphere there would not emerge the need to treat separately these two connected data structures.

6.6 Miscelleneous

6.6.1 Computational Geometry: Determining $x_+(e)$ and $x_-(e)$ in $O(\gamma)$

Claim 6.6.1. *For any edge $e = \{a, b\}$ and node pair $\{u, v\} \subset V$, both $x_+(e)$ and $x_-(e)$ can be calculated in $O(\gamma)$.*

Proof. Let us concentrate on calculation of $x_+(e)$, because $x_-(e)$ can be determined similarly.

Extreme hitting disk of line: First, let us compute the leftmost hitting circle in $\mathscr{C}^{u,v}$ of a point W part of a given line ab. Let $\{A\} = uv \cap ab$ and let $x_+(W)$ be the abscissa of the center point of the leftmost hitting disk of W. Clearly, the function $x_+(W)$ is unimodal in both rays (R_+ and R_-) defined by line ab and point A (i.e., both R_+ and R_- are consisting of an interval where it is strictly monotone increasing and another interval where strictly monotone decreasing).

Let the two points on the line where the local minimum is reached be W_+ on the right side and W_- on the left side of uv. Let the centre point of disks $c(uvW_+)$ and $c(uvW_-)$ be X_+ and X_-, respectively. X_+ is located on the x axis of the coordinate system of apple $A_k^{u,v}$. On the other hand, X_+ is located equidistant from u and line ab, thus it is on parabola p_u defined by point u and line ab. Similarly, let p_v be the parabola defined by v and line ab.

Since p_u and p_v can be characterized with quadratic expressions, which can be solved in $O(1)$, abscissa of X_+ can be found in constant time by determining their common root.

X_- can be determined similarly. Finally, the center point of the desired leftmost hitting disk is the one from X_+ and X_- with smaller abscissa.

Extreme hitting disk of line segment: When restricting the domain of a unimodal function, extremes can appear on the new boundary, thus for segment $[a,b]$, $x_+([a,b]) = \min\big(x(X_+), x(X_-), x(A), x(B)\big)$.

Extreme hitting disk of the polygon of edge: For an arbitrary $e \in E$ we consider two cases regarding the respective position of $[uv]$ and e^p.

In the first case segment $[uv]$ is entirely in the interior of e^p. Trivially, in this case $x_+(e) = -\infty$.

In the second case, segment $[uv]$ is not entirely in the interior of e^p. Since in this model a polygon of edge e^p consists of at most γ line segments, for determining $x_+(e)$ one can determine x_+ for all the line segments that take the minimum value of them in $O(\gamma)$ total time.

Finally, it remains to prove that one can distinguish between the former two cases in $O(\gamma)$. Recall that u and v have ordinates 1 and -1, respectively, both lying on the y axis. Let C be the set of ordinates of the intersection points of axis y and the line segments generating the polygonal chain. If $C \cap [0,1] \neq \emptyset$, than segment $[u,v]$ is neither entirely inside nor entirely outside of the polygon. Otherwise if $|C \cap (1,+\infty)|$ is an even or odd number, than $[uv]$ is outside or inside the polygon, respectively. Note that all the required operations can be done in $O(\gamma)$. □

6.6.2 Extreme cases: Maximum Number of Maximal Failures

In the followings, I restate [J1, Thm 4.] without proof[7] showing that there are some networks for which the running times are asymptotically not significantly better than our non-parametrized bounds. This motivates the introduction of parameter ρ_k, which captures the properties of real-life networks.

Proposition 6.6.2 ([J1], Thm. 4). $\max_{\mathcal{N}} |M_k| = \Theta\left(n^3\right)$ *if* $k = O(1)$*, where* $\mathcal{N}$ *is the set of all networks on* n *points.*

Although $|M_k|$ can be $\Omega(n^3)$, according to this study, in case of many real-life networks it is $O((k+1)n)$.

By combining of Prop. 6.6.2 and Lemma 6.3.14, we have:

Proposition 6.6.3. $|M_k|$ *is* $O(n(k+1)\rho_k)$*, this bound being tight in these parameters for* $k = O(1)$.□

6.6.3 Protecting regional link 0-node failures ensures node disjointness

Claim 6.6.4. *Let* P *and* B *be an SRLG-disjoint primary and backup paths according to regional link 0-node failures. The paths* P *and* B *are node disjoint, apart from the terminal nodes.*

Proof. Assume indirectly that P and B have a common node v in their interior. Let us pick two edges $e_p \in P$ and $e_b \in B$ of G that are adjacent of v. Let p denote the closest corner point of e_p with node v

[7]The reason I omit the proof here is that the proof of [J1, Thm 4.] is only a tiny generalization of the lengthy arguments behind [T1, Claim 5].

if there is any, otherwise its endpoint on P. Similarly we define b for e_b. Note that points v, b and p are not on the same line according to our assumptions in Sec. 6.2; thus, the is a circular disk of any size that covers both e_p and e_b but not node v. We can select a small enough radius for this circular disk, which covers both e_p and e_b and does not have any other nodes interior. The proof follows. □

6.7 Simulation results

In this section, we present numerical results that validate our model and demonstrate the use of the proposed algorithms on some realistic physical networks. The algorithms were implemented in Python version 3.5 using various libraries. No special efforts were made to make the algorithm space or time optimal. The output of the algorithm is a list of SRLGs so that no SRLG contains the other. The network topologies with the obtained list of SRLGs for various k are available online[8].

First, we interpret the input topologies in two ways:

polygon where links are polygonal chains, and

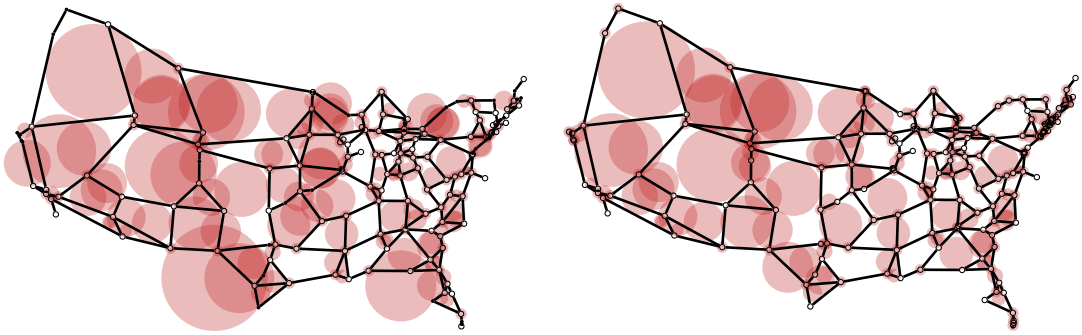
line where the corner points of the polygonal links are substituted with nodes (of degree 2). Here links are line segments.

The second interpretation is artificial, and we mainly use it for verification. Intuitively, the two interpretations result in very different results, as the latter has much more nodes in the network, and thus the regional failures with k nodes interior must be smaller. Fig. 6.6 shows example results for both interpretations of the US ATT-L1 network. The US fiber network has 126 nodes and 208 links as polygonal chains, where the links have 36 corner nodes in total. After transforming it into a network of line segments, we will have a larger network with 162 nodes and 244 links. The transformed network has 30% more nodes; however, the number of SRLGs required for $k = 0$ is just 14% more, which is a sub-linear increase. Surprisingly, after the transformation, the SRLGs became a bit smaller (average number of links $2.98 \rightarrow 2.79$), and the variance in the size of SRLGs is increased from 0.7 to 0.83. It is because, in the transformed, a great number of very small SRLGs appeared, having the two adjacent links for most of the degree 2 nodes.

6.7.1 The list of SRLGs in practice

Table 6.2 shows a comparison for $k = 0$ and $k = 1$ among eleven physical backbone topologies taken from [101]. The columns are: network name, the number of nodes, the number of links, the number of SRLGs, all for both cases where links are polygonal line segments or each corner point of the polygonal links are substituted with a degree 2 node.

[8] `https://github.com/jtapolcai/regional-srlg`



(a) *Polygon*: there are 190 SRLGs with average of 2.98 links and $\rho_0 = 5$.

(b) *Line*: there are 216 SRLGs with average of 2.79 links and $\rho_0 = 5$.

Figure 6.6: The SRLGs of $k = 0$ are visualized for the two cases (a) links are polygonal line segments, and (b) the corner points of the polygonal line segments are treated as degree two nodes and all links are straight lines. In order to have a perspicuous illustration, each SRLG is drawn with the smallest possible circular disk that covers all of its links, even if the disk has nodes interior.

The 126-node US (ATT-L1) network was covered with 190 SRLGs, which is less than listing every single node and link as an SRLG. Fig. 6.6 shows these SRLGs, intuitively each corresponds to a mid-size regional failure. The SRLGs meet our intuition that there are more network nodes in the crowded areas, and thus it generates more SRLGs for them, while in the less crowded areas are covered with SRLGs corresponding to bigger areas. In practice, it is important to have small SRLGs because it strongly influences the performance of the survivable routing algorithms. On Fig. 6.6 the SRLGs are relatively small, each SRLG contains a bit less than 3 links on average. Fig. 6.8c shows how the average size of SRLGs with respect to k for all networks. It has a slightly sub-linear increase with k. Note that the length of the list of SRLGs never exceeded 6000 items in any networks and parameter settings examined.

Next, we have evaluated what would be the radius of the circular disk with $k = 0, 1, 2, 3$ nodes when we know the GPS positions of the nodes (in case of network US ATT-L1). We have performed a Monte Carlo simulation where we pick random locations and compute the maximum radius with $k = 0, 1, 2, 3$ nodes, which is the distance of the closest, the second, third, and fourth closest nodes. Fig. 6.9 shows the cumulative distribution function of the actual radius of the circular disk failures. For example, if we cut the smallest and largest 20% the SRLGs generated for $k = 0$ nodes corresponds to diameter 80-200km.

6.7.2 Tightness of Corollary 6.3.17

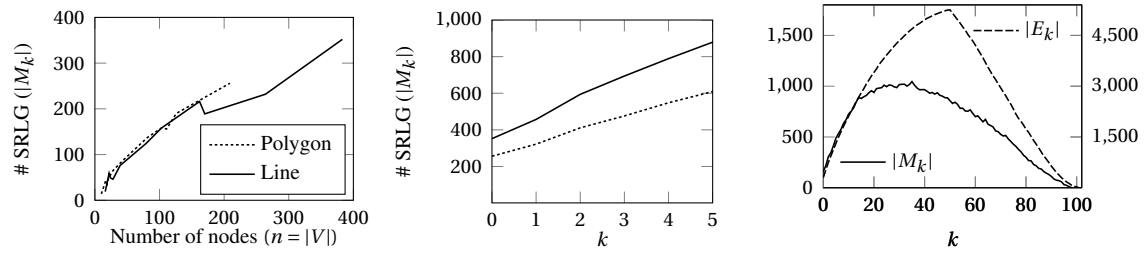
In this subsection we compare the presented parameterized worst case analysis with the obtained simulation results. Corollary 6.3.17 assumes that ρ_k increases linearly with $(k+1)$ for all networks. Fig. 6.8a shows that the edge density increases linearly with $(k+1)$ as we expected. The corollary

Table 6.2: Results of physical backbone topologies of [101].

Name	$\|V\|$		$\|E\|$		# SRLG $k=0$		# SRLG $k=1$	
	Polygon	Line	Polygon	Line	Polygon	Line	Polygon	Line
Pan-EU	10	16	16	22	14	19	27	35
EU (Optic)	17	22	40	45	44	59	57	71
EU (Nobel)	19	28	32	41	36	46	53	81
US [2]	21	24	39	42	48	49	57	64
N.-American	28	39	50	61	65	76	83	97
US (NFSNet)	44	79	73	108	88	124	128	172
US (Fibre)	81	170	141	230	137	189	177	249
US (Deltacom)	103	103	302	302	158	158	218	218
US (Sprint-Phys)	111	264	160	313	156	232	208	307
US (ATT-L1)	126	162	208	244	190	216	255	285
US (Att-Phys)	209	383	314	488	256	352	322	457

provides a linear upper bound on the number of SRLGs with respect to the network size n for fixed k. It is also reflected in Fig. 6.7a which is a graphical illustration for $k = 0$, where we may even have the intuition of a sub-linear growth. Based on Table 6.2 the slope of the curve can be estimated as the number of SRLGs is roughly $\approx 1.2n$ for $k = 0$, and $\approx 2.2n$ for $k = 1$. Corollary 6.3.17 provides a linear upper bound on the number of SRLGs with respect to k if the network (n) is fixed. This is illustrated in Fig. 6.7b where the average number of SRLG is shown for all networks for small k. Here we can experience a slightly sub-linear increase. Fig. 6.8b shows the increase in the number of SRLGs for each network independently for the same range of k. We experienced sub-linearity for all networks, which we further discuss later. Overall, numerical evaluation supports the parameter selection used in the parametrized complexity analysis. We conjecture that Corollary 6.3.17 is close to the experienced performance, and there is little hope for further improving it analytically.

To reveal the reasons why the number of SRLGs increases sub-linearly with respect to k, Fig. 6.7c shows the function graphs of both $|M_k|$ and $\frac{|E_k|}{3}$ for all k values on Network US Deltacom, respectively. As we can see, for $k \leq 15$, $|M_k| \approx \frac{|E_k|}{3}$, while for bigger k values $|M_k| < \frac{|E_k|}{3}$. We can

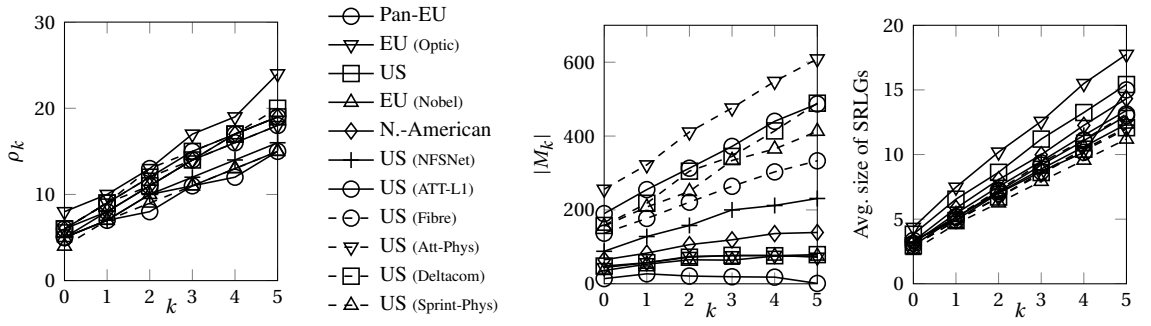


(a) The increase of the number of SRLGs with respect to the number of nodes, for $k = 0$.

(b) The average the number of SRLGs over all networks.

(c) The number of SRLGs and the number of edges in graph k-Delaunay for US Deltacom

Figure 6.7: Comparison of the two interpretation of the input topologies: links are polygonal line segments, or the corner points of the polygonal line segments are treated as degree two nodes and all links are straight lines.



(a) The edge density (ρ_k) (b) The number of SRLGs $|M_k|$. (c) The average size of SRLGs

Figure 6.8: The edge density, number and size of SRLGs for each network and $k = \{0,\ldots,5\}$ in case of polygonal chain links.

deduct that $|M_k| \leq \frac{|E_k|}{3}$ for every $k \in \{0,\ldots,n-2\}$. In other words, the average number of SRLG-s per apple is $\leq \frac{1}{3}$. By Thm. 6.3.4, this also means that $|M_k|$ is $O((k+1)(n-k))$, which induces a sub-linear growth of $|M_k|$ with respect to k.

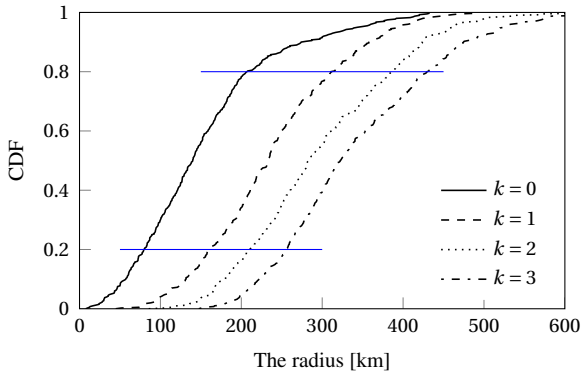


Figure 6.9: The cumulative distribution function of the radius of the disk having a given k number of nodes interior.

6.8 Thesis Summary

Thesis 2. *[C13, C12, C11, J1] To ensure geographic distance between primary and backup paths when the geographical embedding the network topology is approximate, I proposed a model for enumerating regional SRLGs relying only a schematic map of the network topology. For networks described in this model, I proposed a polynomial algorithm for enumerating list M_k of maximal link sets (SRLGs) which can be hit by a disaster overestimated by a shape of a circular disk hitting an arbitrary number k of nodes. I gave theoretical upper bounds on the cardinality of M_k. Evaluating the model and data structure, I showed that in case of real network topologies as input combined with practical (small) k values, M_k is a reasonably short list of link sets.*

Thesis 2.1 (The Limited Geometric Information Failure Model)**.** *[J1] To ensure geographic distance between primary and backup paths when the geographical embedding the network topology is approximate, I proposed the following model. The (not necessarily planar) network is modelled as an undirected connected geometric graph $G = (V, E)$ with $|V| \geq 3$ nodes. The nodes of the graph are embedded as points in the Euclidean plane, and their exact coordinates are considered to be known. In contrast to this, precise positions of edges are not known, instead, it is assumed that for each edge e there is a* containing polygon *(or simply* polygon*) e^p in the plane in which the edge lies.The disasters are assumed to have a shape of a circular disk with an arbitrary radius and centre position, but hitting at most k nodes for $k \in \{0, |V| - 2\}$. The failures caused by these disasters are called* regional link k-node failures.

Thesis 2.2. *[J1] I proposed an algorithm, which, in case of representing a network topology $G(V, E)$ in the Euclidean plane with each link $e \in E$ being part of a related polygonal region e^p having at most γ sides, computes the list M_k of maximal link sets which can be hit by a circular disk hitting at most k nodes in $O\left(|V|^2 \left(\left(k^2+1\right)\rho_k^3 + \rho_k\gamma + \left(k+1+\log(n\rho_0)\right)\rho_0\gamma\right)\right)$, where ρ_k denotes the maximal number of links hit by a circular disk hitting at most k nodes. I proved that list M_k has $O\left(n\,(k+1)\,\rho_k\right)$ elements, this bound being tight in these parameters for $k = O(1)$.*

Thesis 2.3. *[J1] In case of real network topologies, with their edges considered polyginal chains and line segments between their endpoints, respectively, list M_k of maximal link sets which can be hit by a circular disk hitting at most k nodes has $\approx 1.2 \cdot |V|$ and $\approx 2.2 \cdot |V|$ elements for $k = 0$ and $k = 1$, respectively. Additionally, $|M_k|$ increases sublinearly in function of k. Parameter ρ_k representing the maximal number of hit links by a disaster hitting k nodes was ≤ 10 for all the investigated networks for $k = 0, 1$, and grew to only to < 25 for $k = 5$. I concluded that list M_k has a reasonably small size for practical k values.*

Chapter 7

PSRLGs Modeling Correlated Link Failures Caused by Disasters

7.1 Network Model and Framework to Compute Service Availability

At the very beginning, I would like to include Fig. 7.1 summarising the three layers of the contributions of this Chapter. On the top, there are two data structures analog of the CDF and PDF of the functions, which we believe should be the standard for describing the joint failure probability of network resource sets. The second layer is a stochastic model that explicitly takes into count the correlation between the failures of geographically close network elements in case of disasters. In the third layer, as a concrete example of dealing with a disaster type, we use earthquake catalogs to provide proper input for our model. This way, we describe a method of computing PSRLGs of a network from end to end.

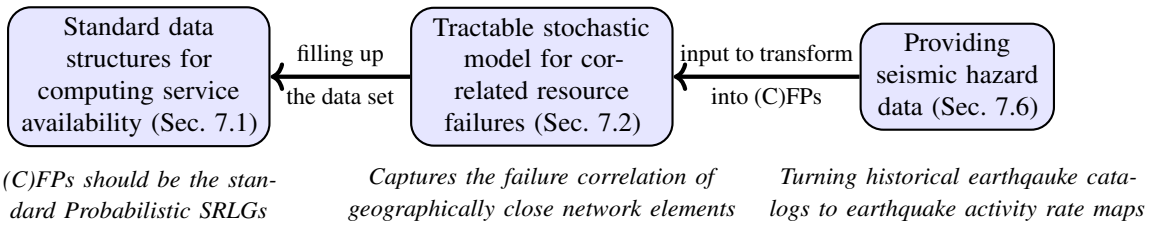
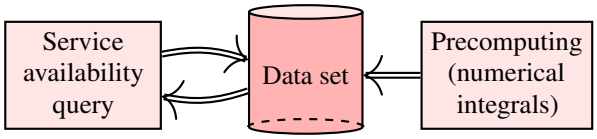


Figure 7.1: Main contributions: We offer 1) standard data structures (for graph G, CFP[G] and FP[G]) for storing joint failure probabilities of link sets, 2) a tractable stochastic model of network element failures caused by disasters, and finally 3) providing the seismic hazard data represented it in a more precise way than the usual hazard maps. Note that our stochastic model can handle the combined inputs of an arbitrary number of disaster families (e.g., tornadoes, earthquakes, tsunamis, etc.). Structures CFP[G] and FP[G] could be established using other models too.



(a) Framework to compute service availability

Data set name	Space complexity	Query time for an arbitrary link set
CFP[G]	$\Omega(2^\rho)$ and $O(2^\rho(n+x)\rho^3\gamma^4)$	hashing: *constant* with high prob. balanced binary tree: $O(\rho\log((n+x)\rho\gamma))$ worst-case
FP[G]	$O((n+x)\rho^3\gamma^4)$	$O((n+x)\rho^3\gamma^4)$

(b) Trade-off between space complexity and query time in case of circular dish shaped disasters

Figure 7.2: Computing service availability via a pre-computed data set: while the disaster hazard can be represented more succinctly using FP[G] for a graph G, with CFP[G] one can achieve lower query times.

7.1.1 Network Model

The network is modeled as an undirected connected geometric graph $G = (V, E)$, with $n = |V|$ nodes and $m = |E|$ links embedded in the Euclidean plane denoted by $\mathbb{R}^2$. The links can be either line segments or polygonal chains (also called 'polylines') built up from at most γ adjacent line segments (where γ is a parameter of our model). The number of link crossings is denoted by x. The geometric *density* of the network topology is the maximum number of links that can be hit by a single disaster and is denoted by ρ. The set of links E is lexicographically sorted, any $S \subseteq E$ is stored as a sorted list.

7.1.2 Framework to Compute Service Availability

We aim to develop a service availability computation engine, where the task is basically to translate the compound problem of simultaneous network failures into a scalar. When setting up an SLA between the user and network provider, the availability of a massive number of network services must be evaluated. Therefore, we need to avoid committing resource-intensive computations at every query. Intuitively, there is much redundancy in these queries. The main idea behind our general framework (depicted in Fig. 7.2a) is to exploit this redundancy by pre-computing some numerical integrals representing failure probabilities of sets of network elements. This, out of the compound geometric and stochastic problem, extracts all the relevant information to a static data set. This data set can address many service availability queries, each of which requiring only lookups and summation.

We propose two standard PSRLG definitions, with different meanings on the probabilities associated with the link sets, to store the failure probabilities of sets of network elements: (1) the Cumulative Failure Probability (CFP), and (2) the Link Failure State Probability (FP). While in this Chapter we focus on failure probabilities of link sets, if necessary, these structures can store failure probabilities of both links *and* node failures (following Sec. 2.6). These data structures were already defined (Def. 2.1.2 and 2.1.3, resp.) in Chapter 2, although the definitions implicitly included the disaster and failure models. In the following, we re-define these models from scratch.

Table 7.1: Table of symbols for Chp. 7

Notation	Meaning
	General
$G(V,E)$	the network modeled as an undirected connected geometric graph, E stored as an ordered list
n, m	number of nodes $\|V\| \geq 3$ and edges $\|E\|$, respectively
$\mathbb{R}^2$	Euclidean plane
CFP(S)	given $S \subseteq E$, the cumulative failure probability (CFP) of S, denoted by CFP(S), is the probability that all links S fail simultaneously (and possibly other links too)
FP(S)	given $S \subseteq E$, the link failure state probability (FP) of S, denoted by FP(S), is the probability that *exactly* the links of S fail simultaneously (and no other links)
FP[G], CFP[G]	the collection of all (S,FP(S)) and (S,CFP(S)) pairs with FP(S) > 0 and CFP(S) > 0, respectively
	Parameter
γ	maximum number of line segments a (polygonal chain shaped) link can have
x	number of link crossings
ρ	maximum number of links that can be hit by a disaster
	Stochastic model
p	epicenter of a disaster, that is a point in the plane $\mathbb{R}^2$
s	shape (and relative size) parameter of a disaster, a real value in $[0,1]$
$h(p)$	hazard of a point $p \in \mathbb{R}^2$, representing the probability that p becomes the epicenter of the next disaster
$r(p,s)$	represents a closed region on the plane (the actual shape of the destroyed area) as a function of epicenter p and the shape/size parameter s; every link having a nonempty intersection with $r(p,s)$ fails
$R(p,s)$	the set of links having a nonempty intersection with $r(p,s)$; set of failed links
$s(p,e)$	smallest size s for which a failure at point p can cover link e
$f(e,p)$	probability that link e fails if a disaster with epicenter p happens
$I_{R(p,s)}(e)$	indicates whether $e \in R(p,s)$
$P(e)$	probability of the failure of link e in the next disaster
$f(S,p)$	probability that a set of links $S \subseteq E$ fail simultaneously, given that the disaster epicenter is $p \in \mathbb{R}^2$
	Precomputation of CPFs and FPs
$\mathscr{P}$	set of points p of the plane such that $dist(p,e) \neq dist(p,e')$ whenever $e \neq e'$ are different segments from E.
$\mathscr{S}(p)$	sequence of link failures for epicenter $p \in \mathscr{P}$, that is an ordered set of links $(e_1, e_2, \ldots, e_l)$, such that $s(p,e_1) \leq s(p,e_2) \leq \cdots \leq s(p,e_l)$, where $l = \|R(p,1)\|$
$\mathscr{S}^j(p)$	first j links of $\mathscr{S}(p)$
$j(S,\mathscr{S}(p))$	ordinal number of set S within $\mathscr{S}(p)$: i, if $S \not\subset \mathscr{S}^{i-1}(p)$ and $S \subseteq \mathscr{S}^i(p)$, 0 otherwise
$f(S,p)$	the conditional probability of a set of links $S \subseteq \mathscr{S}(p)$ failing together, that is $f(e_{j(S,\mathscr{S}(p))}, p)$
$\mathscr{R}_1, \ldots, \mathscr{R}_k$	regions, where each point $p \in \mathscr{R}_i$ has the same sequence $\mathscr{S}_i$ of link failures (see Fig. 7.7)
$\mathscr{S}_i$	for any point $p \in \mathscr{R}_i$, $\mathscr{S}(p) \equiv \mathscr{S}_i$, $i = 1, \ldots, k$.
$e_{i,j}$	j^{th} link of $\mathscr{S}_i$
$P^{i,j}$	$:= \int_{p \in \mathscr{R}_i} h(p) f(e_{i,j}, p) \mathrm{d}p \quad i = 1, \ldots, k,\ j = 1, \ldots, \|\mathscr{S}_i\|$

with these, CFP(S) = $\sum_{i=1}^{k} P^{i,j(S,\mathscr{S}_i)}$, and FP($S$) = $\sum_{i,j}(P^{i,j} - P^{i,j+1})$

	Seismic hazard representation
$c_{i,j}$	epicenter, which represents a latitude-longitude cell on the Earth's surface, taken from a grid of cells over the network area
M_w	moment magnitude $\in \{4.6, 4.7, \ldots, 8.6\} =: \mathscr{M}$
E_{i,j,M_w}	set of earthquakes with centre point in $c_{i,j}$ and magnitude in $(M_w - 0.1, M_w]$
p_{i,j,M_w}	probability that the next earthquake is in E_{i,j,M_w}
r_{i,j,M_w}	activity rates, i.e., arrival rate of earthquakes in E_{i,j,M_w}
t	intensity threshold: in case of a shaking intensity above the threshold, network elements fail; $t \in$ {VI,VII,VIII,IX,X,XI,XII} $:= T$
$disk(c_{i,j}, R(M_w,t))$	the area of destroyed physical infrastructure after each earthquake E_{i,j,M_w}

Definition 7.1.1 (Cumulative Failure Probability (CFP)). *Given a set of links $S \subseteq E$, the cumulative failure probability (CFP) of S, denoted by CFP(S), is the probability that all links S fail simultaneously (and possibly other links too).*

Definition 7.1.2 (Link Failure State Probability (FP)). *Given a set of links $S \subseteq E$, the link failure state probability (FP) of S, denoted by FP(S), is the probability that* exactly *the links of S fail simultaneously (and no other links).*

Sometimes we will refer as 'CFP' to 1) the tuple $(S, \text{CFP}(S))$ for a link set S, or simply, 2) to $\text{CFP}(S)$. For a graph G, we will denote the collection of CFPs with strictly positive probability by $\text{CFP}[G]$. The same applies to the Failure Probabilities ('FP's). We note that the reason behind not referring the tuple of a link set S and $\text{CFP}(S)$ or $\text{FP}(S)$ simply as PSRLGs is that, throughout this Chapter, we need to make a distinction between these two data structures.

Although for some practical tasks, $\text{FP}[G]$ may be a practical input, in the standpoint of availability queries, we mainly look at $\text{FP}[G]$ as a compact representation of structure $\text{CFP}[G]$ (the space complexity of the proposed structures will be investigated in detail in Sec. 7.4).

The *space complexity* of our availability computation engine based on either CFPs or FPs is proportional to the number of link sets S with $\text{CFP}(S) > 0$ (resp., $\text{FP}(S) > 0$). The engine's *time complexity* (namely, its query time) is the time needed to determine the cumulative failure probability of a given link set.

As it turns out, data structures $\text{CFP}[G]$ and $\text{FP}[G]$ present a space-time trade-off: There are more link sets with non-zero CFP than FP, since $\text{FP}(S) > 0$ implies that $\text{CFP}(S') > 0$ for all $2^{|S|} - 1$ nonempty sets such that $S' \subseteq S$. On the other hand, availability queries need to address fewer PSRLGs if they are all expressed as CFPs, and computing these from FPs requires iterating over all FPs in the data set. In Sec. 7.4, we study this trade-off in more detail and give formal bounds on the space complexity and query time for both data structures (see Fig. 7.2b) when applied to our regional failure model.

7.1.3 On Availability Queries when Risk Failures are Correlated

Any availability query can be evaluated by iteratively calling CFP(S) , i.e., the probability of simultaneous failure of all elements in any arbitrary set S. Consider the example network and corresponding CFPs in Fig. 7.3 (non-listed link sets have CFPs of 0). Suppose we need to establish a high-availability connection from the top right node through a working path c and protection path $f - d - e$. The unavailability of the working path is $\text{CFP}(\{c\}) = 0.0113$, and the unavailability of the protection path is $\text{CFP}(\{f\}) + \text{CFP}(\{d\}) + \text{CFP}(\{e\}) - \text{CFP}(\{f,d\}) - \text{CFP}(\{f,e\}) - \text{CFP}(\{d,e\}) + \text{CFP}(\{f,d,e\}) \simeq 0.0275$, by the inclusion-exclusion principle. The total connection availability is $1 - \text{CFP}(\{c,d\}) - \text{CFP}(\{c,f\}) - \text{CFP}(\{c,e\}) + \text{CFP}(\{c,f,d\}) + \text{CFP}(\{c,f,e\}) + \text{CFP}(\{c,d,e\}) - \text{CFP}(\{c,f,d,e\}) \simeq 0.99872$. We can observe that, based on $\text{CFP}[G]$, the connection availability can be computed with the help of CFPs of subsets of $\{c,d,e,f\}$, that is, the union of the links of the working and protection paths.

In contrast, for computing the total connection availability, the $\text{FP}[G]$ data set requires considering a larger number of data set entries. For example, the availability of working path c can be computed as is $1 - \sum_{\{c\} \subseteq S \subseteq \{a,\ldots,e\}} \text{FP}(S)$, i.e., we have to subtract the FP of every link set containing c from 1.

Furthermore, to compute the total availability of the connection, we need to address all nonempty subsets of $\{a, b, c, d, e\}$. The number of links is not part of neither the working nor the protection path; this means up to exponentially more FP[G] queries than CFP[G] queries. Structure FP[G] has an advantage though: it has provably less elements than CFP[G].

By considering joint failure probabilities, we have found that the total connection availability is < 0.9987, i.e., below three nines. For comparison, traditional approaches that assume link failures to be independent, would have estimated the total connection availability to be $1-\text{CFP}(\{c\})\big(\text{CFP}(\{d\})+\text{CFP}(\{e\})+\text{CFP}(\{f\})-\text{CFP}(\{d\})\cdot\text{CFP}(\{e\})-\text{CFP}(\{d\})\cdot\text{CFP}(\{f\})-\text{CFP}(\{e\})\cdot\text{CFP}(\{f\})+\text{CFP}(\{d\})\cdot\text{CFP}(\{e\})\cdot\text{CFP}(\{f\})\big) > 0.99951$, i.e., well above three nines. Even if they correctly compute the availability of each path but assume independent path failures, they estimate the availability by $1 - 0.0113 \cdot 0.0275 > 0.99968$, i.e., even more above three nines. Here, by not considering joint failure probabilities, *the traditional approaches significantly overestimate the total connection availability*, which can lead to more frequent SLA violations and a financial burden on the CSP.

Unfortunately, (correlated) network failures are hard to compute and predict. Nonetheless, to evaluate the expected availability of a service, a network administrator should consider all possible failure scenarios under the specific service availability model stipulated in the corresponding SLA.

7.1.4 Denomination Issues of Probabilistic SRLGs

Probabilistic extensions of SRLGs are called *Probabilistic SRLGs*, PSRLGs. The probabilistic refinement can be defined in multiple ways, thus, in the literature, there are multiple definitions of PSRLGs. E.g., in the first paper considering probabilistic extensions SRLGs (which was [56]), each PSRLG event $r \in R$ occurs with probability π_r, and once a PSRLG event r occurs, link (i, j) will fail independently of the other links with probability $p^r_{i,j} \in [0, 1]$. Thus, we could call the [56]-PSRLGs as 'two-stage PSRLGs'. In contrast with this paper, [56] does not tackle the issue of computing the PSRLGs.

Since both FPs and CFPs are probabilistic extensions of SRLGs, we say that, collectively, these structures are PSRLGs. Moreover, since any version of probabilistic SRLGs can be described with the help of either CFPs or FPs, and due to their natural simplicity, we believe (C)FPs are the right standard way of defining PSRLGs. In the following, we present a model for calculating CFP[G] and FP[G] describing the correlated failure patterns of networks.

7.2 The Regional Failure Model

To compute service availabilities, we need to answer the following question: **what is the probability that a set of links S fails simultaneously?** In other words, we need to find the *cumulative failure probability* of S, i.e., CFP(S), which has a complicated relationship with the correlation structure of

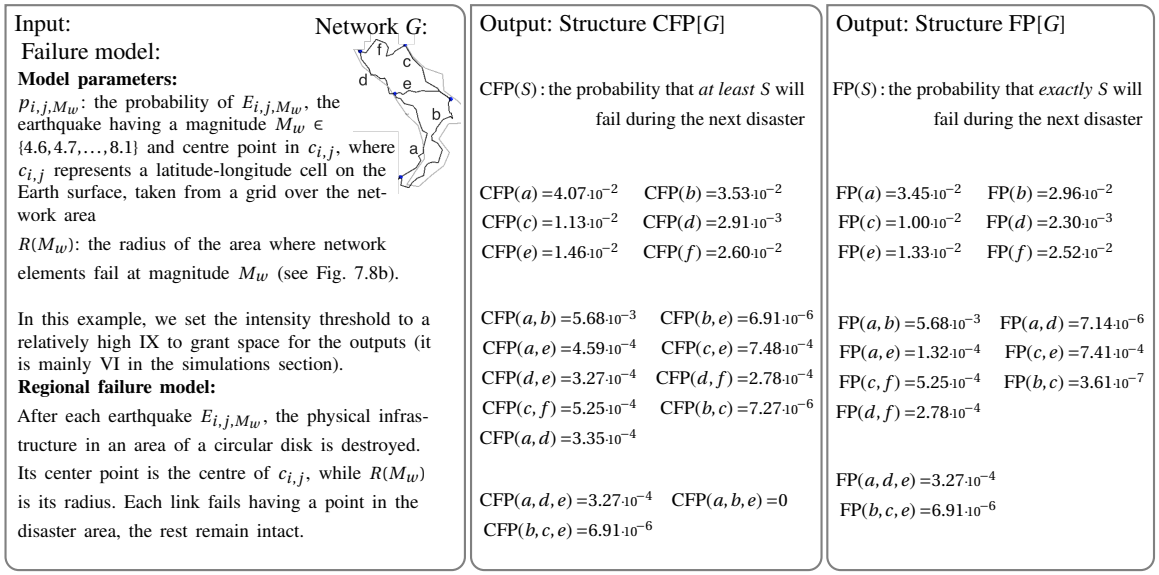


Figure 7.3: An illustration of the problem inputs and outputs. We note that the earthquake failure model depicted here, detailed in Sec. 7.6.1, and used in our simulations, is a special case of our general model presented in Sec. 7.2, that can handle a wide variety of disaster types (including tornadoes, tsunamis, etc.), possibly describing their combined effect.

link failures. Links that lie close together more often fail simultaneously, while further apart links rarely do. To find CFP(S), we first propose a general stochastic model of possible network failure events. After some pre-computation, this will allow us to build a succinct representation of the joint probability distribution of link failures described in the previous section.

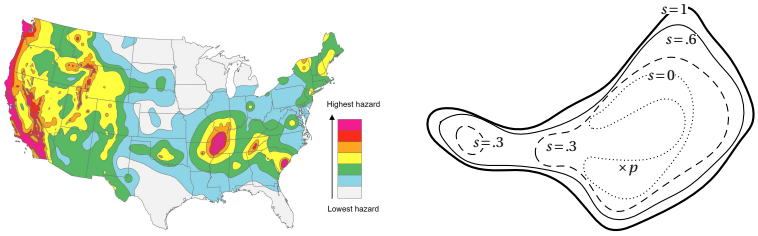
In our model, failures are considered to come solely from disasters affecting a bounded geographical area. This chapter focuses only on link failures (node failures can be translated to the joint failure of the set of all links adjacent to the node). The model could be extended to incorporate node failures as well based on the techniques described in Sec. 2.6.

While traditional approaches focus on single-point failures, which represent hardware/node failures, cable/link cuts, etc., we adopt a model for regional failures and focus on computing the conditional probability $\text{CFP}_d(S)$ that, in a given time period, a set of links S fail together under a disaster of type d (e.g., a tornado, earthquake, Electromagnetic Pulse (EMP), etc.).

Assumption 7.2.1. *We assume that, in the investigated time period, there will be at most one disaster of any type*[1].

In such a case, to obtain the availability values, we need to build a model for each disaster type, and the resulting availability of S can be expressed as $1-\sum_{d\in D} p_d \cdot \text{CFP}_d(S)$, where D denotes the set of modeled failure types and p_d is the probability of disaster d. From now on, for ease of notation,

[1]The case, when more disasters are expected to happen simultaneously, can be handled by defining a new mixed disaster type, see also [102].



(a) Probabilistic hazard map $h(p)$ for earthquakes as function of epicenter p. [103]

(b) Shape of regional disaster $r(p,s)$ for epicenter p and different sizes $s = 0, 0.3, 0.6, 1$.

Figure 7.4: Example of real-world inputs.

we will consider a fixed failure type d, and, therefore, the subscript d is omitted hereafter.

7.2.1 Stochastic Modeling of Regional Failures

In the remainder of the Chapter, we will call events that bring down the network in a geographic area simply as *disasters*, indifferent to their cause. We model regional failures caused by a disaster with the following parameters with randomly chosen values:

epicenter p , which is a point in the plane $\mathbb{R}^2$,

shape (and size) s , which is a real value in $[0,1]$.

Each point $p \in \mathbb{R}^2$ is assigned a **hazard** $h(p)$ representing the probability that p becomes the epicenter of the next disaster (see Fig. 7.4a). Specifically, $h(p)$ is a probability density function on the area $\mathbb{R}^2$, and therefore,

$$\int_{p\in\mathbb{R}^2} h(p)\mathrm{d}p = 1\ . \tag{7.1}$$

After a disaster of the examined type, the physical infrastructure (such as optical fibers, amplifiers, routers, and switches) in some areas is destroyed. The possible shapes for this area are defined by a set $r(p,s)$ that represents a closed region on the plane (the actual shape of the destroyed area) as a function of epicenter p and the shape/size parameter s. This is a general disaster model, where several possible damage areas can be defined by $r(p,s)$.

Definition 7.2.1 (Regional disaster). *We assume a* regional disaster *of epicenter p and shape/size s will result in the failure of exactly those links of network G that have a point in $r(p,s)$.*

Our next assumption is that $r(p,s)$ is monotone increasing in the relative size s, that is, a more severe version of a disaster hits at least the same region of the network, as a weaker disaster (see Fig.

7.4b for an example)[2]. While this assumption holds in general for a variety of disasters, we only use it to achieve 'nicer' equations.

Assumption 7.2.2.

$$r(p,s) \subseteq r(p,s') \text{ if } s < s' \quad \forall p \in \mathbb{R}^2, 0 \le s, s' \le 1 \ . \tag{7.2}$$

For simplicity, we assume $r(p,s)$ for a given p is a result of uniform sampling of damage areas. Namely, for a given p, s has a uniform distribution over interval $[0,1]$, i.e., the probability of the failure to be of size at most s is exactly s. Thus, s is called *relative size* in the remainder of the paper.

Note that, given the disaster epicenter and relative size, the outcome of the attack is deterministic. In other words, any link e within $r(p,s)$ fails with probability 1, if a failure event with parameters p and s occurs. Let us denote the set of failed links by $R(p,s)$. Definition 7.2.1 together with Assumption 7.2.2 imply that, given a point p, $R(p,s) \subseteq R(p,s')$ if $s \le s'$. Let $s(p,e)$ denote the corresponding smallest size s for which a failure at point p can cover link e. Furthermore, we denote by ρ the maximum number of links that can be affected by a single failure (of maximum size $s = 1$):

$$\rho = \max_{p \in \mathbb{R}^2} |R(p,1)| \ . \tag{7.3}$$

7.2.2 The Failure Probability of a Link Set

We first explain how to compute the probability CFP(S) that a set of links $S \subseteq E$ will fail simultaneously in the next disaster.

Let $f(e,p)$ be the **probability** that link e fails if a disaster with epicenter p happens. Note that by Assumption 7.2.2, $f(e,p) > 0$ can occur iff $e \in R(p,1)$. $f(e,p)$ can be computed from $R(p,s)$, where s is in the range $[0,1]$. Hence,

$$f(e,p) = \int_{s=0}^{1} I_{R(p,s)}(e) \mathrm{d}s \ , \tag{7.4}$$

where the indicator function $I_{R(p,s)}(e)$ indicates whether $e \in R(p,s)$. Thus,

$$I_{R(p,s)}(e) = \begin{cases} 1 & \text{if } e \in R(p,s) \ , \\ 0 & \text{otherwise.} \end{cases} \tag{7.5}$$

By Assumption 7.2.2, if $I_{R(p,s)}(e) = 1$, then $I_{R(p,s')}(e) = 1$, for $s \le s'$.

We now extend our notation to capture the probability of the failure of link e in the next disaster:

$$P(e) := \int_{p \in \mathbb{R}^2} h(p) f(e,p) \mathrm{d}p. \tag{7.6}$$

[2]Various failure shapes were studied so far [2, 5, 6, 7, 8, 9, 10, 11, 17, 18, 40, 51, 52, 53, 54, 55], mainly in the form of circular regional disasters or line-segment failures, but in some cases also more general geometric shapes [17, 40]. All of these models meet Assumption 7.2.2.

We denote the probability that a set of links $S \subseteq E$ fail simultaneously, given that the disaster epicenter is $p \in \mathbb{R}^2$:

$$f(S,p) := \int_{s=0}^{1} \prod_{e \in S} I_{R(p,s)}(e) \mathrm{d}s \ . \tag{7.7}$$

In other words, if the sequence of links is $S = (e_1, e_2, \ldots, e_{|S|}) \subseteq R(p,1)$ and $s(p,e_1) \leq s(p,e_2) \leq \cdots \leq s(p,e_{|S|})$, then $\prod_{e \in S} I_{R(p,s)}(e) = 1$ iff $s \geq s(p,e_{|S|})$, otherwise the product is 0. This implies that

$$f(S,p) = f(e_{|S|},p) = \min_{e \in S} f(e,p) \ . \tag{7.8}$$

Finally, using the above results[3]:

$$\mathrm{CFP}(S) = \int_{p \in \mathbb{R}^2} h(p) f(S,p) \mathrm{d}p = \int_{p \in \mathbb{R}^2} h(p) \min_{e \in S} f(e,p) \mathrm{d}p \ . \tag{7.9}$$

For example, on the right of Fig. 7.3, the results of applying the formula to the 5-node network are shown for all the non-zero joint link failure probabilities. In this example, $r(p,s)$ is always a circular disk with a radius computed according to the historical seismic information. Potentially there are exponentially many joint failure events in terms of the network size; however, links far from each other have zero probability of failing jointly because of a single disaster. For example, this holds for links f and b, whose smallest distance is more than the radius of the largest destroyed area.

Former works (e.g., [17, in the proof of Lemma 8]) expressed the joint failure probability of a set S by multiplying the failure probabilities of the links in S, thus implicitly assuming these failures are independent. Unlike [17], our model assumes a deterministic failure outcome (once its epicenter and shape are set). This implies that, in our model, failures are dependent. For example, two lines in the same location (e.g., within the same conduit) always fail together (e.g., when the conduit is cut).

7.2.3 Example of the Geographical Correlation of Failures

In this Subsection, we first consider a simple linear and discrete model for some of the ideas presented so far. We assume that the ground set of our simplified world is the set of 1000 integer points of a line with coordinates between $z_{min} = -499$, $z_{max} = 500$ and we have two links e_0 and e_z, which themselves are integer points from the interval $[-499, 500]$, e_0 is at position 0, and e_z is at position z. Let the probability that i is the location of a disaster be $h_i = 10^{-3}$ for $i = -499, \ldots 500$ so that $\sum_{i=-499}^{500} h_i = 1$. According to Eq. (7.9), the probability of the failure of link e_0 is

$$P(e_0) := \sum_{i=-499}^{500} h_i f(e_0, i) \ , \tag{7.10}$$

[3] Without Assumption 7.2.2, we would have $\mathrm{CFP}(S) = \int_{p \in \mathbb{R}^2} h(p) \int_{s=0}^{1} \prod_{e \in S} I_{R(p,s)}(e) \mathrm{d}s \mathrm{d}p$.

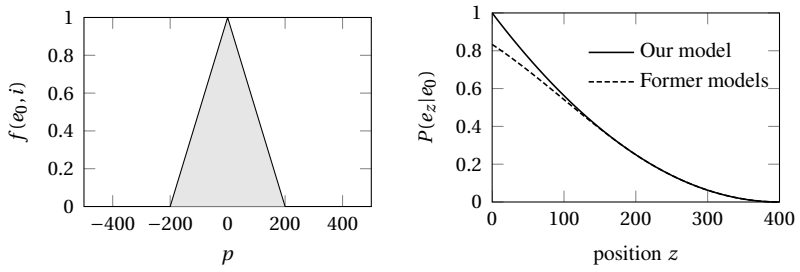


Figure 7.5: An example of $f_i(0)$ at different i positions and the corresponding $P(e_z|e_0)$ depending on z. Former models assumed the link failures are independent given an epicenter of the disaster.

where $f(e_0, i)$ is the conditional probability that link e_0 fails if the failure is at position i. According to Eq. (7.9), the joint probability of the failure of both links e_0 and e_z is

$$P(\{e_0, e_z\}) := \sum_{i=-499}^{500} h_i \min(f(e_0, i), f(e_z, i)) \quad . \tag{7.11}$$

Let $P(e_z|e_0)$ denote the conditional probability that e_z fails, on the condition that e_0 fails. By definition we have

$$P(e_z|e_0) := \frac{P(\{e_0, e_z\})}{P(e_0)} \quad . \tag{7.12}$$

This is a function of z in our setting. Intuitively, $P(e_z|e_0)$ is close to 1 if the two links are exactly in the same location (i.e. $z = 0$).

Additionally, $P(e_z|e_0)$ should be a decreasing function of z in the range of $[0, 500]$. See Fig. 7.5 for an example of $f(e_0, i)$ values and the corresponding $P(e_z|e_0)$.

7.3 Pre-Computation to Speed up Queries

In the previous section, we have described a model that generates a regional disaster according to a hazard density $h(p)$ and a failure shape function $r(p, s)$. Recall that our task is to return CFP(S) for a set of links $S \subseteq E$, which is the probability that links S fail together in case of disaster d.

Unfortunately, the calculation of integrals (7.9) can be a computationally-intensive process. One solution is to calculate some FPs in advance so that when a query comes on the CFP of an arbitrary set of links S, then the task would be summing up some of the pre-computed FP values.

As Lemma 7.4.1 will show, a full list of FPs with non-zero probabilities has $O((n+x)\rho^2\gamma^4)$ items. Every CFP can be derived by summing up

$$\text{CFP}(S) = \sum_{T \supseteq S} \text{FP}(T), \quad \forall S \subseteq E. \tag{7.13}$$

7.3.1 Precomputation of CFPs and FPs

In this subsection, we still rely on Assumption 7.2.2 and make the following additional assumptions to apply some computational geometry results. We emphasize that the second and third additional specifications are technical assumptions to avoid lengthy discussions (see Sec. 7.7).

- Shapes $r(p,s)$ are limited to circular disks centered at p. This corresponds to the case where the failure of a link e depends on the Euclidean distance $dist(p,e)$ of e to the epicenter of the disaster p. In this case, instead of $r(p,s)$, the input is given by radius d as a function of s.
- In our geometric reasoning, we will transform the links of the graph into line segments by slightly shortening them to ensure that no two segments share a common endpoint (see the details of the transformation in Sec. 7.7). We also assume that no more than two links intersect in the same point, and no more than two endpoints lie on the same line.
- The relative size s is a *uniformly Lipschitz continuous function* of radius d. That is, there exists a positive number K such that for every point p in the plane, if we have neighborhoods $r(p,s')$ and $r(p,s)$ with respective radii d' and d, then $|s'-s| \le K|d'-d|$ holds.

For ease of presentation, we slightly reduce the domain we are integrating over. Let $\mathscr{P}$ denote the set of points p of the plane such that $dist(p,e) \neq dist(p,e')$ whenever e and e' are different segments from E. We have that $\mathbb{R}^2 \setminus \mathscr{P}$ is of measure zero, hence in our considerations, integrating over the plane $\mathbb{R}^2$ can be replaced by integrating over $\mathscr{P}$.

Inspired by (7.8), we can now define the sequence of possible link failures (see Fig. 7.6), when the epicenter of the attack is at p:

Definition 7.3.1. *The* sequence of link failures *for epicenter $p \in \mathscr{P}$ is an ordered set of links $\mathscr{S}(p) = (e_1, e_2, \ldots, e_l)$, such that $s(p,e_1) \le s(p,e_2) \le \cdots \le s(p,e_l)$, where $l = |R(p,1)|$. Let $\mathscr{S}^j(p)$ denote the first j links of $\mathscr{S}(p)$, i.e. $\mathscr{S}^j(p) = (e_1, e_2, \ldots, e_j)$.*

Furthermore, the ordinal number of a set S within $\mathscr{S}(p)$ is defined as follows:

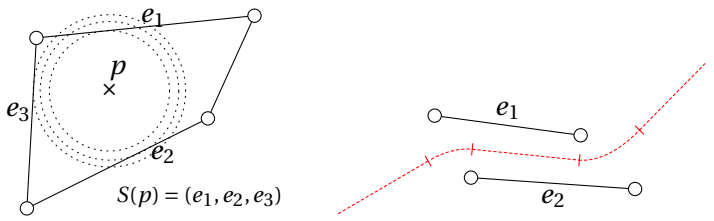
Definition 7.3.2.

$$j(S,\mathscr{S}(p)) = \begin{cases} i, & \text{if } S \not\subset \mathscr{S}^{i-1}(p) \text{ and } S \subseteq \mathscr{S}^i(p) \\ 0, & \text{otherwise.} \end{cases}$$

Due to Assumption 7.2.2 and using also (7.9), if there is a disaster at point p, the conditional probability of a set of links $S \subseteq \mathscr{S}(p)$ failing together is

$$f(S,p) = f(\mathscr{S}^{j(S,\mathscr{S}(p))}(p), p) = f(e_{j(S,\mathscr{S}(p))}, p) \quad . \tag{7.14}$$

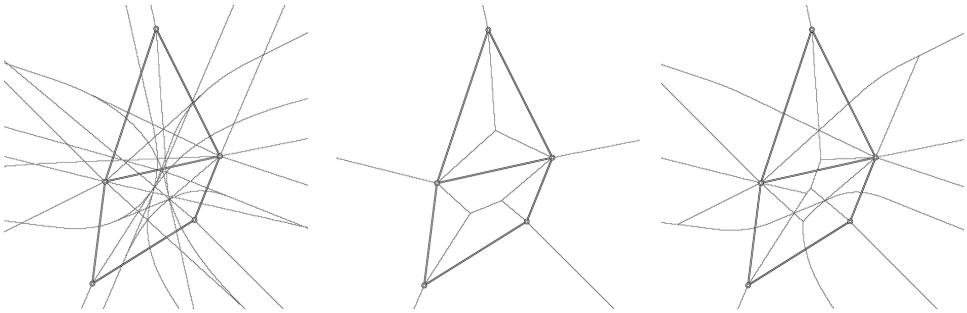
Finally, we use two practical input parameters, x, and ρ, in estimating the space complexity of our approaches. Let x be the number of link crossings in the network G. For backbone networks,



(a) The sequence of link failures for epicenter p.

(b) The bisector curve of two line segments e_1 and e_2.

Figure 7.6: Illustration of link failure sequences



(a) Regions with same sequence of link failures.

(b) Nearest Neighbor Voronoi Diagram

(c) 2-Voronoi diagram

Figure 7.7: An example of different partitions of the plane into regions used in Lemma 7.4.1.

x is a small number, as typically, a switch is also installed on each link crossing [89]. The second parameter is ρ, the *link density* of the network, which is defined as the maximal number of links that could fail together (i.e., could be covered by a circle of radius r). The link density ρ, practically, does not depend on the network size. Moreover, ρ is at least the maximal nodal degree in the graph.

Let us divide the plane into disjoint regions $\mathscr{R}_1, \ldots, \mathscr{R}_k$, where each point $p \in \mathscr{R}_i$ has the same sequence $\mathscr{S}_i$ of link failures (see Fig. 7.7, and [104] for efficient algorithms calculating these regions). Here, k is the number of possible failure sequences. For any point $p \in \mathscr{R}_i$, we introduce notation $\mathscr{S}(p) \equiv \mathscr{S}_i$, $i = 1, \ldots, k$.

Based on Equation (7.14), it is sufficient to pre-compute and store the following integrals:

$$P^{i,j} = \int_{p \in \mathscr{R}_i} h(p) f(e_{i,j}, p) \mathrm{d}p \quad i = 1, \ldots, k, \; j = 1, \ldots, |\mathscr{S}_i|, \tag{7.15}$$

where $e_{i,j}$ denotes the j-th link in $\mathscr{S}_i$.

Finally, since the regions are mutually disjoint as subsets of $\mathscr{P}$ and cover it entirely, equation (7.9) can be written as a sum and, according to (7.14), the failure probability of any link set $S \subseteq E$ can be evaluated as

$$\mathrm{CFP}(S) = \sum_{i=1}^{k} \int_{p \in \mathscr{R}_i} h(p) f(S, p) \mathrm{d}p = \sum_{i=1}^{k} P^{i, j(S, \mathscr{S}_i)} \;, \tag{7.16}$$

where we define $P^{i,0} := 0$ for every $i = 1, \ldots, k$. Based on Eq. (7.13) and (7.16), one can derive that:

$$\mathrm{FP}(S) = \sum_{i,j} \left(P^{i,j} - P^{i,j+1} \right), \tag{7.17}$$

where the summation is for those pairs (i, j) for which $1 \le i \le k$ and $j(S, \mathscr{S}_i) = |S| > 0$. As a default, we set $P^{i,|\mathscr{S}_i|+1} = 0$.

7.4 Space and Time complexity of Structures CFP$[G]$ and FP$[G]$

7.4.1 Cardinality of Structures FP[G] and CFP[G]

In our basic model, considering the case of the disaster shapes being circular disks in a given L_p metric, (where, for $p = 2$, we get back the usual Euclidean circles, for $p = 1$ or $p = \infty$, we have a family of parallel-sided squares, and, for $p = 2/3$, astroids, that are specific 4-cornered stars), the number of FPs can be upper bounded as follows[4].

Lemma 7.4.1. *In case of a set of circular disk shaped disasters (i.e., $r(p, s)$ is circular) in a given L_p metric, and the edges of the network being in general position, there are $O((n + x)\rho^2\gamma^4)$ FPs with non-zero probability.*

Proof. Let us concentrate on line segment links for a moment. According to Claim 5.1.9, the number of links, m, is $O(n + x)$ for line segment links. We know from [79, Thm. 6] that the number of k-Voronoi cells in L_p norm for line segments is $O(k(m - k) + x)$, or alternatively, $O(k(n + x - k) + x)$ thus disasters hitting k links can hit at most this many link sets. Since a circular disk can hit at most ρ links, this sums up to $O(\rho^2(n + x + x)$, which is $O(\rho^2(n + x))$.

If links can be polygonal chains consisting of at most γ line segments, there are $O(\gamma(n + x))$ segments with $O(\gamma^2 x)$ crossings, meaning $O(k\gamma^2(n + x))$ k-Voronoi regions. By counting the k-Voronoi regions for $k \in \{1, \ldots, \gamma\rho\}$, this yields an upper bound of $O((n + x)\rho^2\gamma^4)$ for the number of FPs. □

In the same setting, the number of CFPs can be very large:

Lemma 7.4.2. *The number of CFPs with non-zero probabilities is lower-bounded by $\Omega(2^\rho)$. In case of a set of circular disk shaped disasters in a given L_p metric, and the edges of the network being in general position, the number of CFPs with non-zero probabilities is upper-bounded by $O(2^\rho(n + x)\rho^2\gamma^4)$.*

[4] We note that a similar upper bound can be shown for a more general family of disaster sets for the joint failure probability of the nodes. Specifically, taking a set V of points of the plane (i.e., the nodes instead of the links), it is known that, for any convex compact S subset of the plane, the number of subsets V' of V with cardinality k fulfilling $V' = V \cap S'$ for some homothetic copy of S (i.e., S' is a copy of S shrinked/enlarged and shifted) is less than $(2k - 1)|V| - k^2$ [105].

Proof. By the definition of ρ, there is a link set S with CFP$(S) > 0$ and $|S| = \rho$. As, for any $S' \subseteq S$, CFP$(S) > 0$ implies CFP$(S') > 0$, implying the lower bound. By Lemma 7.4.1, there are at most $O((n+x)\rho^2\gamma^4)$ non-zero FPs, each having at most 2^ρ subsets, yielding the upper bound. □

Every FP and CFP can be stored in $O(\rho)$ space, since it contains a link set of at most ρ links, alongside with a related probability. This way, the space requirement of FP[G] and CFP[G] is upper bounded by $O((n+x)\rho^3\gamma^4)$ and $O(2^\rho(n+x)\rho^3\gamma^4)$, respectively.

7.4.2 Query Time of Structures FP[G] and CFP[G]

When storing the non-zero FPs in a list, by Eq. (7.13), querying the FP[G] structure for CFP(S) requires iterating over all non-zero FPs and summing up all FP(T) such that $T \supseteq S$. Thus, S has to be compared with $O((n+x)\rho^2\gamma^4)$ (Lemma 7.4.1) other sets, and each comparison can be made in $O(\rho)$. The number of possible additions is also $O((n+x)\rho^2\gamma^4)$, thus the query time of the FP[G] structure is upper-bounded by $O((n+x)\rho^3\gamma^4)$. Alternatively, if we stored the FPs in an ordered balanced binary tree, we would need to lookup all the exponential number of $T \supseteq S$.

The query time of CFP$[G]$ also depends on the data structure used for storing CFPs. For example, if we store all non-zero CFPs in a list, the query time would be $\Omega(2^\rho)$ (Lemma 7.4.2). In contrast, by hashing all CFP(S) on S, we reduce the query time a constant with very high probability. Last, when storing all non-zero CFPs in a self-balancing binary tree, the worst-case query time would be $O(\rho + \log((n+x)\rho\gamma))$ (Lemma 7.4.2). Although the CFP structure can achieve impressive query times, this comes at the cost of its space complexity ($\Omega(2^\rho)$), which makes it inefficient for larger network densities.

7.5 Implementation Issues

The approaches and performance guarantees we gave in Sections 7.3 and 7.4 are valid under the assumption that the shape of a regional failure is always a circular disk. In this section, we propose a heuristic that (1) can accommodate any disaster shape; (2) does not require advanced geometric algorithms; and (3) is more suitable for digital inputs, as it uses discrete functions instead of continuous ones.

We discretize the problem by defining a sufficiently fine grid over the plane such that for each grid cell c, the disaster regions $r(p,s)$ and hit link sets $R(p,s)$ are "almost identical"[5] for all $p \in c$. This reduces the integration problem from Sec. 7.2 to a summation[6].

[5] In particular, we may assume that $f(e,p)$ is independent of p as long as it is in c and denote this common value by $f(e,c)$.

[6] [51] uses a similar grid approach.

We consider $\mathbb{R}^2$ as a Cartesian coordinate system. Let r denote the absolute maximum range of a disaster in km. Let (x_{min}, y_{min}) be the bottom left corner and (x_{max}, y_{max}) the top right corner of a rectangular area in which the network lies. It is sufficient to process each c in the rectangle of bottom left corner $(x_{min} - r, y_{min} - r)$ and top right corner $(x_{max} + r, y_{max} + r)$, and we denote by $c_{i,j}$ the grid cell in the i-th column and j-th row of this rectangle. We assume we are given the probability $h_{i,j}$ of the next disaster epicenter p lying in cell c: $h_{i,j} = \int_{p \in c_{i,j}} h(p) \mathrm{d}p$.

Now, for each c, we can compute the sequence of link failures and store the link sets as follows.

7.5.1 Structure CFP[G]

For our CFP[G] structure, we use an associative array CFP[G], which can be addressed by a set of links $S = \{\ell_1, \ell_2, \ldots, \ell_k\}$ and returns its cumulative failure probability. In the pre-computation process, we have to extract the contribution of $c_{i,j}$ to the failure probability of every subset S of links. To do so, we process the sequence of link failures $\mathscr{S}_{i,j} = (e_1, e_2, \ldots, e_l)$ attached to disaster epicenters which are in $c_{i,j}$[7], and increment the CFP[G] values accordingly: CFP($\{e_1\}$)+ = $h_{i,j} \cdot f(e_1, c_{i,j})$, CFP($\{e_2\}$)+ = $h_{i,j} \cdot f(e_2, c_{i,j})$, CFP($\{e_1, e_2\}$)+ = $h_{i,j} \cdot f(e_2, c_{i,j})$, etc. By default, for every link set S, we set initially CFP(S) = 0.

To obtain CFP(S), we look it up in the associative array. If S is not found, then CFP(S) = 0.

7.5.2 Structure FP[G]

For our FP[G] structure, we take a similar approach as for the CFP[G] structure and use a list of 'S, FP(S)' set-failure probability pairs.

In the pre-computation process, we have to extract the contribution of $c_{i,j}$ to the link failure state probability of every subset S of links. As in the case of the CFPs, we do so by iterating over the sequence of link failures $\mathscr{S}_{i,j} = (e_1, e_2, \ldots, e_l)$ and incrementing the FP values accordingly: FP($\{e_1\}$)+ = $h_{i,j} \cdot \big(f(e_1, c_{i,j}) - f(e_2, c_{i,j})\big)$, FP($\{e_1, e_2\}$)+ = $h_{i,j} \cdot \big(f(e_2, c_{i,j} - f(e_3, c_{i,j})\big)$, FP($\{e_1, e_2, e_3\}$)+ = $h_{i,j} \cdot \big(f(e_3, c_{i,j} - f(e_4, c_{i,j})\big)$, etc.

To obtain CFP(S), we sum up $\sum_{T \supseteq S} \mathrm{FP}(T)$.

7.6 Model Evaluation Based on Seismic Hazard Data

In this section, we present numerical results that validate our model and demonstrate the use of the proposed algorithms on real backbone networks (taken from [106]) accompanied with real seismic hazard inputs. The algorithms were implemented in Python 3.6., using its various libraries[8],

[7] Here, we represent $c_{i,j}$ by its center p. According to Def. 7.3.1, for $i < j$, link e_i is closer to p than e_j, i.e., $s(p, e_i) < s(p, e_i)$.

[8] The simulation data can be downloaded from [106].

respecting the regional failure model presented in Section 7.2, and following the implementation principles of Section 7.5. Run-times were measured on a commodity laptop with a Core i5 CPU at 2.3 GHz with 8 GiB of RAM.

As a practical scenario, the simulations presented in this paper focus on transforming the seismic hazard on network topologies to PSRLGs. As a first step, we need to convert the historical seismic hazard data into a regional failure model for our framework. Subsec. 7.6.1 discusses our earthquake representation based on epicenter and moment magnitude. In a nutshell, the model translates the seismic hazard data to a set of circular disk shaped disaster areas with radii depending on the actual moment magnitude (Fig. 7.8). Note that the epicenter distribution is non-uniform here.

We are taking this probabilistic earthquake set as input, Subsec. 7.6.2 presents our simulation results validating our PSRLG model.

7.6.1 Seismic Hazard Representation

We are investigating the failures caused by the next earthquake within a given geographic area; thus, we assume there is exactly one earthquake in the investigated period. Each earthquake is uniquely identified by its epicenter and moment magnitude [107]:

epicenter $c_{i,j}$, which represents a latitude-longitude cell on the Earth's surface, taken from a grid of cells over the network area.

moment magnitude $M_w \in \{4.6, 4.7, \ldots, 8.6\} =: \mathscr{M}$.

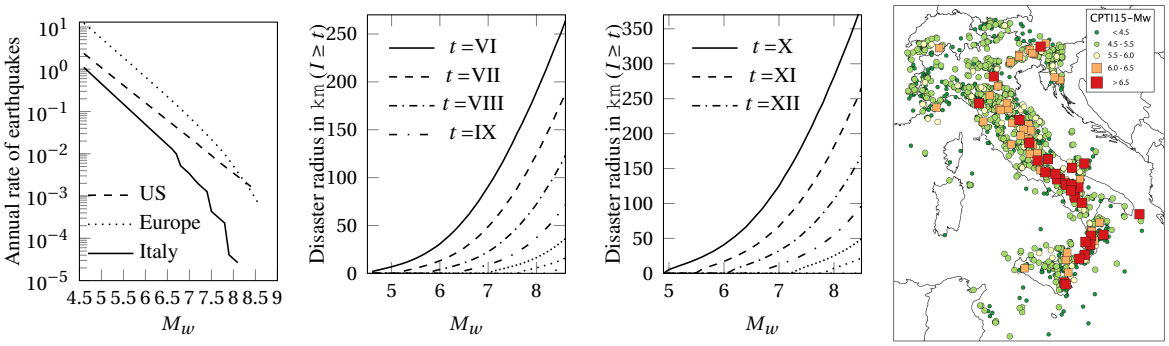
We index the grid cells such that $i \in \{1, \ldots, i_{max}\} =: \mathscr{I}_i$, $j \in \{1, \ldots, j_{max}\} =: \mathscr{I}_j$.

Let E_{i,j,M_w} denote the set of earthquakes with centre point in $c_{i,j}$ and magnitude in $(M_w - 0.1, M_w]$. As cells and magnitude intervals are small enough that the failures caused by each earthquake in E_{i,j,M_w} will often be identical[9], we will represent all E_{i,j,M_w} with a single earthquake having a center point in the center of $c_{i,j}$ and a magnitude of M_w. Let the probability that the next earthquake is in E_{i,j,M_w} be p_{i,j,M_w}. Note that these probabilities add up to 1, i.e. $\sum_{i,j \in \mathscr{I}_i \times \mathscr{I}_j} \sum_{M_w \in \mathscr{M}} p_{i,j,M_w} = 1$.

Our initial input are the activity rates r_{i,j,M_w} of earthquake types (see Fig. 7.8a) instead of the p_{i,j,M_w} values, so we first have to translate these rates to probabilities. We claim that under the assumption that each kind of earthquake E_{i,j,M_w} arrives according to independent Poisson arrival processes with parameters r_{i,j,M_w}, the rates of earthquakes E_{i,j,M_w} can be transformed to probabilities p_{i,j,M_w} as follows:

$$p_{i,j,M_w} = r_{i,j,M_w} \Big/ \sum_{i,j \in \mathscr{I}_i \times \mathscr{I}_j} \sum_{M_w \in \mathscr{M}} r_{i,j,M_w}. \tag{7.18}$$

[9]The sides of grid cells used in our simulations were 0.05° long in the Italian rate map, and 0.1° in case of the EU and the USA, meaning 4km to 10km of cell side length.



(a) Rate vs. magnitude

(b) Disaster ranges $R(M_w, t)$ for Italy/Europe

(c) Disaster ranges $R(M_w, t)$ for the USA

(d) Historical earthquakes from the Italian catalogue [108].

Figure 7.8: Seismic input data

We assign each network element e an **intensity threshold** $t(e)$. If the intensity I of the ground shaking reaches this threshold ($I \geq t(e)$) at any point of the physical embedding of e, the element fails. In our simulation, every network element has the same threshold $t(e) := t$, where $t \in \{\text{VI,VII,VIII,IX,X,XI,XII}\} := T$ according to the Mercalli-Cancani-Sieberg (MCS) scale [109][10].

After each earthquake, E_{i,j,M_w}, the physical infrastructure (such as optical fibers, amplifiers, routers, and switches) in an area $disk(c_{i,j}, R(M_w, t))$ of a circular disk is destroyed. The center point of $disk(c_{i,j}, R(M_w, t))$ is the center of $c_{i,j}$, while its radius $R(M_w, t)$ is monotone increasing in the magnitude M_w, and decreasing in the intensity threshold t (see Fig. 7.8b and 7.8c). As earthquakes can occur anywhere in the cell, we increase the radius by the distance between the center of the cell and its outer corners. This way, the disk is always an overestimate of an earthquake's damaged area in cell $c_{i,j}$ with magnitude M_w.

7.6.1.1 Earthquake Activity Rates

These are the occurrence rates of earthquake events as a function of space, time, and magnitude. To obtain them, we need to define an earthquake source model, defined as an area or an active fault that could host earthquakes as testified by instrumental seismic activity, historical seismicity, geomorphological evidence, and regional tectonics. The choice of the earthquake source model is strongly driven by the available knowledge of the area and by the scale of the problem. It may range from well-defined active faults, especially when working at a local scale, to less understood and wider scale seismotectonic provinces. When the catalog of earthquakes covers a long period, it can be used to compute earthquake activity rates without any information of seismotectonic provinces and/or active faults, via, for example, a smoothed seismicity approach. In this work, we evaluated the earthquake source model for Italy and the USA from the most recent published earthquakes

[10] Intensity $I \leq$V does not cause structural damage, while $I =$XII means total damage.

catalogs ([108], and [110], for Italy and the USA, respectively) that cover a long period and can be used to obtain earthquake source model without other information. Although earthquakes can be clustered in time and space with their distribution that is over-dispersed if compared to the Poisson law[111], a common way to treat this problem (i.e., cluster in time and space) is to de-cluster the earthquake catalog, i.e., removing all events not considered mainshocks, via a declustering filter [112]. Here, both catalogs are considered de-clustered. The standard methodology to estimate the density of seismicity in a grid, and used in this work, is the one developed by [113]. The smoothed rate of events in each cell is determined as follows:

$$Sr_i = \frac{\sum_j r_j \exp\left(\frac{-d^2(c_i,c_j)}{d_c^2}\right)}{\sum_j \exp\left(\frac{-d^2(c_i,c_j)}{d_c^2}\right)}, \tag{7.19}$$

where r_j is the cumulative rate of events with magnitudes greater than the completeness magnitude M_c in each cell c_i of the grid and computed from the historical catalogue of earthquakes, $d(c_i, c_j)$ is the distance between the centers of grid cells c_i and c_j. The parameter d_c is the correlation distance (for Italy, 30km [114] and for the USA, 75km [115]). Then, the earthquake activity rates for each node of the grid are computed following the Truncated Gutenberg-Richter model [116]:

$$\lambda(M) = \lambda_0 \frac{\exp(-\beta M) - \exp(-\beta M_u)}{\exp(-\beta M_0) - \exp(-\beta M_u)} \tag{7.20}$$

for all magnitudes M between M_0 (lower or minimum magnitude) and M_u (upper or maximum magnitude); otherwise $\lambda(M)$ is 0. The upper and lower magnitude bounds represent, respectively, the maximum magnitude, or the largest earthquake considered for a particular source model, which depends on the regional tectonic context (in our case, M_w is at most 8.1, 8.6 and 8.3 for Italy, Europe, and the US, respectively), and the minimum magnitude, or threshold value, below which there is no engineering interest or lack of data (in this study, $M_w > 4.5$)[11]. Additionally, λ_0 is the smoothed rate Sr_i of earthquakes at $M_w = 4.5$ and $\beta = b\ln(10)$, where b is the b-value of the magnitude-frequency distribution. For Italy, we calculated the b-value of the distribution on a regional basis using the maximum-likelihood method from [117], while for the USA, it comes from [110]. While for Italy and the USA, we computed the earthquake rates (Fig. 7.8a) following this approach and with the referenced data, for Europe, we used the already published SEIFA model ([118], and [87]), a kernel-smoothed, zonation-free stochastic earthquake rate model that considers seismicity and accumulated fault moment. In this model, activity rates are based on the SHARE European Earthquake Catalogue frequency-magnitude distribution model. The spatial distribution of model rates depends on the density distributions of earthquakes and fault slip rates. A magnitude-frequency distribution indicates

[11] Fig. 7.8a shows that, in the investigated range of magnitudes, the global rate of earthquakes dips exponentially in the function of the magnitude.

the probability that an earthquake of a size within the upper and lower bound of the distribution may occur anywhere inside the source during a specified period.

While this does give us the rates for all combinations of epicenters and magnitudes for Italy, the USA, and Europe (Fig. 7.8a), we still need the relation between magnitude and disaster area to be able to apply these rates to the network resiliency models.

7.6.1.2 The Radius of the Damaged Zone

The only earthquake effect that can be quantified at the scale of the whole country is ground shaking because quantifying any other earthquake effects requires a site investigation. Shaking intensity is localized and is generally diminishing with distance from the earthquake's epicenter. At the scale of a whole country, we can assume that soil and topographic conditions are relatively homogeneous. The seismic intensity only depends on the distance from the earthquake epicenter.

Here, we assume all links (and nodes) inside the area with a given MCS intensity $I \geq t$ (where $t \geq$VI) are damaged, while all components outside of this area remain functioning. Thus, to obtain all disaster areas, we now only need the disaster area radius for each magnitude $M_w \in \{4.6, 4.7, \ldots, 8.6\}$. For this purpose, we used the intensity prediction equation of [119] and [120], for Italy/Europe, and the USA, respectively, where the expected intensity I at a site located at epicentral distance R is:

$$I_{\text{It,EU}} = 1.621 \cdot M_w - 1.343 - 0.0086(D - h) - 1.037(\ln D - \ln h), \tag{7.21}$$

$$I_{\text{US}} = 0.44 + 1.70 \cdot M_w - 0.0048 \cdot D - 2.73 \cdot \log_{10} D, \tag{7.22}$$

where $D = \sqrt{R^2 + h^2}$ is a sort of hypocentral distance, and h represents the hypocentral depth, which may be viewed as the average depth of the apparent radiating source [119], h equaling 3.91km and 10km for Italy/Europe and the USA, respectively. In this way, it is possible to compute for each M_w and intensity threshold t the site-distance $R(M_w, t)$ from the epicenter of the desired intensity threshold level. It is worth noting that Eq. (7.21) has been obtained using only the Italian earthquake historical catalog, and so it is not entirely correct to use it for the entirety of Europe. However, the Italian catalog is one of the more complete catalogs in Europe. There is no similar equation in the literature for the entire continent (to the best of our knowledge), and its development is beyond the paper's scope. We assume that the application of Eq. (7.21), as a first approximation, can be considered correct for entire Europe.

7.6.2 Simulation results

We consider seven topologies: one Italian topology, three other European topologies, and another three US topologies. Unless otherwise stated, we set the intensity tolerance threshold, t, to VI according to the MCS scale. The node and link counts, as well as the number of CFPs and FPs with

non-zero probability, of all topologies are given in Table 7.2 both for t = VI and t = VII.

Interestingly, although the US network has slightly more nodes and links than the Italian network, it has much less CFPs (946 compared to 12106). This difference is easily explainable when we consider our theoretical results from Sec. 7.4: the number of non-zero CFPs is lower-bounded by $\Omega(2^{\rho})$ (Lemma 7.4.2), which means an exponential growth with the maximal number of hit links, ρ. Since the Italian network has much shorter links than the American network, its hit link sets tend to be larger. We can observe this same exponential increase with the maximal number of hit links when we decrease the threshold from t = VII to t = VI. For example, the number of CFPs of NFSNET is 1762 at t = VII, but explodes to 14199 if we decrease this threshold to t = VI. In contrast, the number of FPs makes a much smaller jump, from 523 to 969.

By only storing the x largest CFPs, we can trade in some precision in exchange for a significant reduction in memory usage. Fig. 7.9a shows the precision of this approach versus x. For the Italian topology, the highest probability among the omitted edge sets is 5.4×10^{-4} or 1.7×10^{-5} if we store only the top 100 or 1000 CFPs respectively. Furthermore, increasing the precision by order of magnitude requires only a bit more than an order of magnitude more CFPs. Similarly, in the case of the other networks, storing the first 100 or 1000 CFPs means that the highest probability among the omitted edge sets is below 5×10^{-4} or 1×10^{-5}, respectively; and increasing the number of CFPs by order of magnitude is more than enough for increasing the precision by a factor of 10.

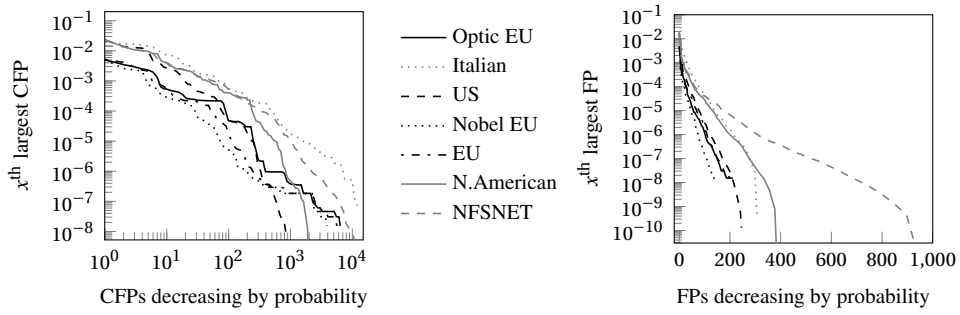
Speaking of the precision-memory trade, omitting some of the FPs is also possible. In this case, the imprecision in the value of CFP(S) for some S can be upper bounded by the sum of probabilities stored in the omitted FPs. On Fig. 7.9b, we can see the probability assigned to the x^{th} most probable FP. Fortunately, the highest number of non-zero FPs was low, 969 in our experience, meaning that, most probably, no omission is needed.

As mentioned before, the difference in the number of non-zero CFPs can partly be explained by a difference in hit link set sizes. Fig. 7.9c shows the maximal number of hit links, ρ, versus the intensity threshold, t. We can confirm that, at t =VI, the Italian network has a much higher density than the US network (13 compared to 7).

We have also investigated the average CFP of a set of links with given cardinality. Fig. 7.9d shows the average failure probability concerning the number of links failing together. Single links have

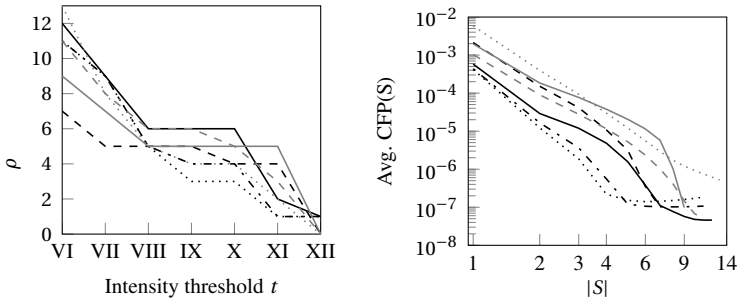
Table 7.2: The investigated network topologies

Network name	n	m	# CFPs at t =VI	# FPs at t =VI	# CFPs at t =VII	# FPs at t =VII
Optic EU	22	45	6377	202	1369	135
Italian	25	34	12106	308	676	200
US	26	43	946	246	260	164
Nobel EU	28	41	3867	149	680	94
EU	37	57	5634	212	745	133
N.-American	39	61	2024	394	556	257
NFSNET	79	108	14199	969	1762	523



(a) Structure CFP[G], t =VI

(b) Structure FP[G], t =VI



(c) Link density ρ

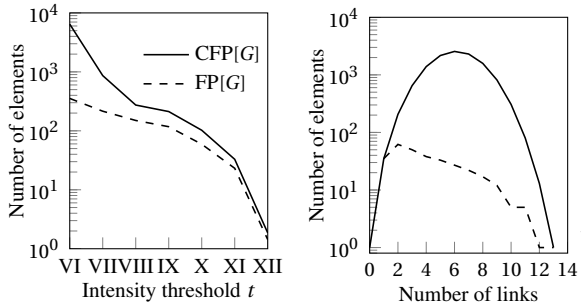
(d) The avg. CFP of listed edge sets with given cardinality, t =VI

Figure 7.9: The space and complexity of the data structures for the examined network topologies.

an average failure probability between $[4.2 \times 10^{-4}, 2.1 \times 10^{-3}]$, depending on the network topology. The average failure probability for double and triple link failures lies in $[1.2 \times 10^{-5}, 3.9 \times 10^{-4}]$ and $[1.9 \times 10^{-6}, 9.3 \times 10^{-5}]$, respectively. These averages meet our expectations that the correlation between link failures is significant. By our observations, the combination of link failures with the highest CFPs is predominantly the combined failure of links incident to a single node.

Fig. 7.10 further investigates the relationship between the space requirements of CFP[G] and FP[G]. In Fig. 7.10a, we show the space requirement of structures CFP[G] and FP[G] as a function of the intensity threshold t. As expected, the number of CFPs drops quickly with the intensity threshold. Our results show that, especially at lower thresholds, choosing the FP structure can significantly reduce space requirements. This phenomenon is even stronger in case of Italy_995, a network with 32 nodes and 70 links over Italy, that we decided to exclude from most of the simulation presentations. The reason for this is its unusually high density: at intensity tolerances of t = VI and ρ = VII, it has densities $\rho_{\text{VI}} = 31$ and $\rho_{\text{VII}} = 19$, yielding $> 10^9$ and 1153294 CFPs, while the number of its FPs is only 2011 and 1090, respectively.

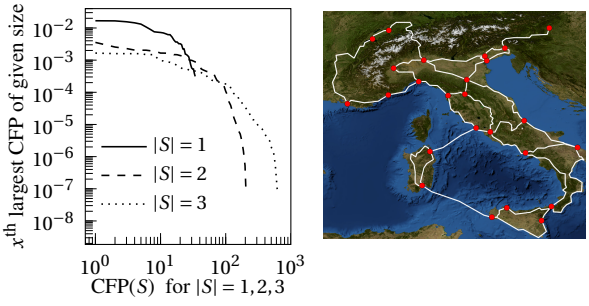
Fig. 7.10b depicts the number of CFPs and FPs with given cardinality for the Italian. Since there is a link set of cardinality 13 with positive FP, there must be over 1700 subsets of cardinality 6 with non-zero CFP. In comparison, the number of FPs peaks at 71 for cardinality 4.



(a) Average CFP/FP number vs. intensity threshold

(b) CFP/FP number vs. # links contained for Italian, t =VI

Figure 7.10: Comparison of space efficiency of structures CFP[G] and FP[G]



(a) CFPs of Italian for t =VI

(b) Italian topology

Figure 7.11: CFP comparison of single, double and triple link failures for Italian

Continuing our study of the cardinality of failed link sets, Fig. 7.11a investigates the dependency between CFP(S) and $|S|$ in detail, for $|S| = 1$, 2 and 3. There are 34 single link failures in the Italian network whose CFPs range between $[0.0003, 0.019]$; it has 205 dual link failures with non-zero probabilities between $[7 \times 10^{-8}, 0.0037]$, and there is a number of 648 triple link failures with strictly positive probabilities, ranging between $[7 \times 10^{-8}, 0.0019]$. Here we can see that some CFPs with size l are less probable than some other CFPs containing $l+1$ links. Thus, only storing CFPs with at most l links rarely yields the same result as only storing the most probable CFPs. Also, we can observe that the CFPs of the most probable triple link sets are not much smaller than the CFPs of the most probable link pairs. This is another sign that the most probable double and triple link failures are failures of the links incident to the same network node.

7.7 An End Note on the Geometric Transformation of the Network

In the geometric reasonings of the current Chapter, we transform the links of the graph into line segments. We also need to ensure that no two segments share a common endpoint. In the network, the adjacent links terminate in a single node; thus, we need to perform a minor transformation as follows.

Let $S \subseteq E$ be a set of segments and $\epsilon > 0$ a small number. Suppose that we shorten some segments e of S, in a way that we delete ϵ long subsegment from both ends, in such a way that the deleted intervals do not overlap. Let S' denote the set of segments S after shortening.

Lemma 7.7.1. *We have* $f(S,p) \geq f(S',p)$ *and* $f(S,p) - f(S',p) \leq \epsilon K$ *hold for every point* p.

Proof. For the first inequality note that

$$f(S,p) = \int_{s=0}^{1} \prod_{e\in S} I_{R(p,s)}(e)\mathrm{d}s \geq \int_{s=0}^{1} \prod_{e'\in S'} I_{R(p,s)}(e')\mathrm{d}s = f(S',p) \tag{7.23}$$

because $I_{R(p,s)}(e) \geq I_{R(p,s)}(e')$ holds for every s, whenever $e \in S$.

We turn now to the second inequality. Let s be the smallest value such that $\prod_{e\in S} I_{R(p,s)}(e) = 1$ (if there is any), and set $s' = s + \epsilon K$. Let d and d' be the radii of $r(p,s)$ and $r(p,s')$, resp. By the Lipschitz property we have $\epsilon K = s' - s \leq K(d' - d)$ giving that $d' > d + \epsilon$. We know by the definition of s that $r(p,s)$ intersects every segment $e \in S$ in some point Q_e. But then $r(p,s')$ intersects e'. This holds, because the larger disk $r(p,s')$ clearly contains the disk of radius ϵ centered at Q_e, and the latter disk must intersect e' because we deleted disjoint subintervals of length at most ϵ from e to obtain e'. We have therefore $\prod_{e'\in S'} I_{R(p,s')}(e') = 1$, hence

$$f(p,S) - f(p,S') = \int_{y=0}^{1} \left(\prod_{e\in S} I_{R(p,y)}(e) - \prod_{e'\in S'} I_{R(p,y)}(e') \right) \mathrm{d}y \leq \int_{y=s}^{s'} 1\mathrm{d}y = \epsilon K. \tag{7.24}$$

□

We transform our set of segments into one, where no segment e has an endpoint A on any other segment. If we have such a segment, then we carry out the transformation by deleting an ϵ long subsegment of e starting at A. Lemma 7.7.1 gives that if we set ϵ sufficiently small, then all the values $f(p,S)$ and $f(p,S')$ will be very close to each other, hence CFP(S) and CFP(S') will be very close to each other. Moreover, for any two segments $e_1, e_2 \in E$, we have that either $e_1 \cap e_2 = \emptyset$, or $e_1 \cap e_2$ is an interior point of both segments.

As a simple example illustrating the Lipschitz condition 2) from 7.3.1, suppose that $r(p,s)$ is a disk centered at p having radius sR_p, where R_p is the radius of $r(p,1)$. Then for radii $d = sR_p$ and $d' = s'R_p$ we have $|s' - s| = \frac{1}{R_p}|d' - d|$. The Lipschitz condition then holds if there exists a $k > 0$ such that $R_p \geq k$ for every p.

7.8 Thesis Summary

Thesis 3. *[C4, C9, J3] I defined a stochastic model of link failures caused by disasters, which considers the correlation between failures of links which are geographically close to each other. To unify the notions and terminology on Probabilistic SRLGs, I proposed standard data structures for containing the disaster probabilities. In case of circular disk shaped disasters, for the size and query time of these data structures, I proposed theoretical upper bounds. Evaluating the model and data structures, I showed that in case of taking real seismic data as input, these data structures have a manageable size.*

Thesis 3.1. *[C4, C9, J3] Inspired by earthquake behaviours, I defined a stochastic model of link failures caused by disasters. This model is the first to explicitly consider the correlation between failures of links which can be subject to the same disaster.*

To unify the notions and terminology linked to probabilistic extensions of Shared Risk Link Groups, I proposed two standard data structures for describing the disaster probabilities. Namely, for a graph G, these structures are called FP[*G*] *and CFP*[*G*]*, respectively. In FP*[*G*]*, for each link set S, the probability that* exactly *S will fail is stored as FP(S) , while in CFP*[*G*]*, the probability that* at least *S will fail is stored as CFP(S) .*

Thesis 3.2. *[C4, C9, J3] In case of disasters having shapes of circular disks in a given L_p metric, representing the network topology $G(V,E)$ in the Euclidean plane with links considered as polygonal chains consisting of at most γ line segments, denoting the number of link crossings by x, and the maximum number of links which are hit by one of the disasters by ρ, I proved the followings. There are $O((|V|+x)\rho^2\gamma^4)$ FPs with nonzero probability. The number of CFPs with positive probability is lower bounded by $\Omega(2^\rho)$ and upper bounded by $O(2^\rho(|V|+x)\rho^2\gamma^4)$. Storing all the positive CFPs in a balanced binary tree, the worst-case query time of the CFP of a given link set is $O(\rho\log((|V|+x)\rho\gamma))$. Storing all the positive FPs in a list, the query time of the CFP of a given link set is $O((|V|+x)\rho^2\gamma^4)$.*

Thesis 3.3. *[C4, C9, J3] Using real-world seismic hazard data combined with Italian, European and contiguous US network topologies, I found the followings. Assuming network equipment fails only at a shaking of intensity VIII of the MCS scale, there is no significant difference in the cardinality of CFPs and FPs with positive probability. The number of CFPs becomes unacceptably large and slow to compute only at the combined presence of strong earthquakes (with $M_w \geq 8$), short network links ($\leq\sim 50$ km), and network resources poorly resistant to ground shaking (failing at intensity VI). Structure FP has a low cardinality and can be computed in some minutes in these circumstances too, even on a commodity laptop. Finally, listing CFPs with at most l links rarely yields a list equivalent to keeping some of the most probable CFPs.*

Chapter 8

Conclusion

8.1 Summary

This Thesis is dedicated to prove that the effect of regional disasters (natural on man-made) can be modeled with a low number of SRLGs or PSRLGs. These carefully constructed lists of (P)SRLGs can be used as input for e.g., network recovery/planning mechanisms.

In Chapter 5 (Thesis 1), I showed that, known the geometric embedding of the network topology, and overestimating the disaster area by a circular disk with radius r, the list M_r^p of worst-case SRLGs can be calculated in low-polynomial time, and has $O(|V|\rho_r)$ elements. In case of spherical embedding I also proved low-polynomial bounds on the size and computing time of M_r. I showed that in case of real-world network topologies and disasters with a radius $\leq\sim$ 500km, the difference between lists M_r^s and M_r^p is often less than the difference due to the disaster size estimation.

Chapter 6 (Thesis 2) offered a regional failure model for the case when only a schematic map of the network is given as input. The resulting list M_k of maximal failures caused by circular disks hitting at most k nodes has at most $O(|V|k\rho_k)$ elements, and can be calculated in low-polynomial time.

Lastly, Chapter 7 (Thesis 3) presented 1) a unified terminology on PSRLGs, 2) a tractable stochastic model of disaster failures explicitly taking in count the correlation on the link failures, and 3) an evaluation of the model based on seismic data. I concluded that the risk induced by the seismicity on backbone topologies can be encoded in and quickly looked up from a small-sized CFP or FP list.

These lists of vulnerable regions can be used as input of various problems arising in the field of network resiliency. Some of these problems are resilient (geodiverse) routing, (k-)content connectivity, network failure detection, service availability queries, resilient backbone network planning, disaster avoidance control, resilient SDN, etc.

8.2 Open Problems

I believe for M_r^p, there is little room for making the created algorithms faster for its enumeration. Also, theoretical upper bounds on $|M_r^p|$ are tight. Contrary to this, inspired by the lessons learned on M_r^p, there are some low-hanging fruits for a faster algorithm enumerating M_r^s, and better upper bounds on $|M_r^s|$.

In case of M_k, the presented algorithm computes it in $O\left(n^2\left(k^5 + k\log(nk+1)+1\right)\right)$ (if γ is constant and ρ_k is $O(k+1)$). A natural question is whether it is possible to compute M_k quicker. For $k = 0$ the answer is yes in case of usual networks: as presented in [C13] and [T1], M_0 can be determined in $O(n(\log n + \rho_0^3\tau_0))$, where τ_0 is an additional parameter depending on local properties of the embedding of the network in the plane (and edges are considered as line segments). In other words, there exists an algorithm to compute M_0 in $f(\rho_0, \tau_0)O(n\log n)$. I believe that there exist similar algorithms for determining M_k in $f\left(k, \rho_k, \gamma, \ldots\right) O\left(n\log n\right)$, where f depends only on k and on 'local' properties of the embedding. This near-linear complexity in n would allow M_k to be computed even quicker in case of huge networks too. Due to the limits of my research, this remained an open problem.

It is also an open question if one can enhance the preciseness of our probabilistic disaster failure model presented in Chapter 7 while maintaining its relative simplicity.

8.3 Possible Future Work

Possible future directions of this research include but not restrict to:

- better integration of failure modeling into disaster resilience approaches (FRADIR[C8, C5, J2]-like studys),
- proving our conjecture that the regional SRLG-disjoint routing problem is in P,
- evaluating our probabilistic failure model with more complex real-world inputs,
- as a side-track of a future SRLG list comparing study, creating the 'SRLG-Zoo', a webpage similar to Topologyzoo [121], from where one could download network topologies and related (P)SRLG lists.

The reason I did not mention designing more elaborated SRLG models is twofold. In many cases, the data on the network and its environment is very limited, thus, in my opinion, designing more complex SRLG failure models than the state-of-the-art would be an overkill (e.g., embedding the network to a hyperbolic space). On the other hand, if we do have abundant data, then we should aim for increasingly accurate probabilistic models.

Afterword

PTTR - Post, Telephone, Telegraph, Radio: this good old Romanian telecom brand is still written in front of a post office in Seklerland in 2021 when half of the western civilization is in home office or learns online and 5G networks are being installed worldwide. The speed of technological transformation is never anticipated, indeed. My parents are telling mind-boggling stories of the telecommunication possibilities of the communist Romania of the '80s. They had to wait for hours in a crowded room to get access to a telephone line to home, and when they finally got it, they had to listen to and answer the frequent question of the operator: 'Still speaking?'. The way of sending birthday wishes would also be peculiar today: the post transferred dispatches charged per characters counted by hand.

After the revolution of '89, my father applied for an international telephone line for his freshly founded company. Surprisingly at that time, he got it, and a number of people could manage their foreign affairs at him – with a little bit of luck since successful dials were rare. In the '90s, Romania made a huge leap forward in the technology applied and closed up to the state of the art. The Internet arrived in cities, towns, and lucky villages of the country, and in the absence of regulations enforced, in the 2000s, the country was told to host a major pool of hackers.

By now, the Internet has become a topmost critical infrastructure all over the world that allows not only companies to rely on it and send money instantaneously but also scientists to collaborate in a way unseen before. On 31st July 2019, for example, we were finalizing an INFOCOM submission, working at the same time on the same file from Madrid, Zurich, Vienna, Berlin, and Székelyudvarhely, respectively. I have to admit that despite our work was smooth and effective, I had to use my mobile net because the regular (wired) Internet connection was down in the town. They rumor that the power supply of a nearby router is getting wet at local storms (if that is a common feature of our local infrastructure, then even a mild storm can cause here a *regional failure*). Maybe it was just a simple cable cut. That is not unanticipated in local level either: when they built the sewer pipe to the local spa in 2018, according to the rule, they planned its route to avoid all kind of infrastructure, apparently but for the optical cable of the Internet: they cut it in the first hour of digging.

Conform to the arising requirements, it is likely that the telecommunication networks will continue to evolve to such an extent that in some decades, our global status in 2021 would seem as obsolete the telecom possibilities of the '80s Romania. As discussed in this Dissertation, my findings are part of the work for this vision.

List of Figures

Publications

Book chapters

[B1] B. Vass, J. Tapolcai, D. Hay, J. Oostenbrink, F. A. Kuipers, "How to Model and Enumerate Geographically Correlated Failure Events in Communication Networks", in Guide to Disaster-Resilient Communication Networks, Springer, 2020

[B2] T. Gomes, L. Martins, R. Girao-Silva, D. Tipper, A. Pašić, B. Vass, L. Garrote, U. Nunes, M. Zachariasen, J. Rak, "Enhancing Availability for Critical Services", in Guide to Disaster-Resilient Communication Networks, Springer, 2020

[B3] T. Gomes, D. Santos, R. Girão-Silva, L. Martins, B. Nedic, M. Gunkel, F. Dikbiyik, B. Vass, J. Tapolcai, J. Rak, "Disaster-Resilient Routing Schemes for Regional Failures", in Guide to Disaster-Resilient Communication Networks, Springer, 2020

Journal papers

[J1] B. Vass, E. Bérczi-Kovács, and J. Tapolcai, "Enumerating Maximal Shared Risk Link Groups of Circular Disk Failures Hitting k Nodes", IEEE-ACM Transactions on Networking, 2021

[J2] A. Pašić, R. Girão-Silva, F. Mogyorósi, B. Vass, T. Gomes, P. Babarczi, P. Revisnyei, J. Tapolcai, J. Rak "eFRADIR: An Enhanced FRAmework for DIsaster Resilience", IEEE Access, 2021

[J3] B. Vass, J. Tapolcai, Z. Heszberger, J. Bíró, D. Hay, F. A. Kuipers, J. Oostenbrink, A. Valentini, L. Rónyai, "Probabilistic Shared Risk Link Groups Modelling Correlated Resource Failures Caused by Disasters", IEEE Journal on Selected Areas in Communications, 2021

[J4] J. Tapolcai, L. Rónyai, B. Vass, and L. Gyimóthi, "Fast Enumeration of Regional Link Failures Caused by Disasters with Limited Size", IEEE-ACM Transactions on Networking, 2020

[J5] B. Vass, L. Németh, J. Tapolcai, "The Earth is Nearly Flat: Precise and Approximate Algorithms for Detecting Vulnerable Regions of Networks in Plane and on Sphere", Networks, Wiley, 2020

Conference and workshop papers

[C1] B. Vass, E. Bérczi-Kovács, C. Raiciu, G. Rétvári, "Compiling Packet Programs to Reconfigurable Switches: Theory and Algorithms", P4 Workshop in Europe (EuroP4 '20), Barcelona, Spain, 2020

[C2] B. Németh, Y.-A. Pignolet, M. Rost, S. Schmid, B. Vass, "Cost-Efficient Embedding of Virtual Networks With and Without Routing Flexibility", IEEE IFIP Networking, Paris, France, 2020

[C3] D. Haja, B. Vass, L. Toka, "Towards making big data applications network-aware in edge-cloud systems", IEEE 8th International Conference on Cloud Networking (CloudNet), Coimbra, Portugal, 2019

[C4] A. Valentini, B. Vass, J. Oostenbrink, L. Csák, F. A. Kuipers, B. Pace, D. Hay and J. Tapolcai, "Network Resiliency Against Earthquakes", IEEE Int. Workshop on Resilient Networks Design and Modeling (RNDM), Nicosia, Cyprus, 2019

[C5] A. Pašić, R. Girao-Silva, B. Vass, T. Gomes, F. Mogyorósi, P. Babarczi, J. Tapolcai, "FRADIR-II: An Improved Framework for Disaster Resilience", IEEE Int. Workshop on Resilient Networks Design and Modeling (RNDM), Nicosia, Cyprus, 2019

[C6] D. Haja, B. Vass, L. Toka, "Improving Big Data Application Performance in Edge-Cloud Systems", IEEE 12th International Conference on Cloud Computing (CLOUD), Milan, Italy, 2019

[C7] B. Vass, L. Németh, A. de Sousa, M. Zachariasen and J. Tapolcai, "Vulnerable Regions of Networks on Sphere", IEEE Int. Workshop on Resilient Networks Design and Modeling (RNDM), Longyearbyen (Svalbard), Norway, 2018.

[C8] A. Pašić, R. Girão-Silva, B. Vass, T. Gomes, and P. Babarczi, "FRADIR: A Novel Framework for Disaster Resilience", IEEE Int. Workshop on Resilient Networks Design and Modeling (RNDM), Longyearbyen (Svalbard), Norway, 2018.

[C9] J. Tapolcai, B. Vass, Z. Heszberger, J. Biró, D. Hay, F. A. Kuipers, and L. Rónyai, "A Tractable Stochastic Model of Correlated Link Failures Caused by Disasters", in Proc. IEEE INFOCOM, Honolulu, HI, USA, 2018.

[C10] J. Tapolcai, L. Rónyai, B. Vass, and L. Gyimóthi, "List of Shared Risk Link Groups Representing Regional Failures with Limited Size", in Proc. IEEE INFOCOM, Atlanta, GA, USA, 2017

[C11] B. Vass, E. Bérczi-Kovács, and J. Tapolcai, "Enumerating Shared Risk Link Groups of Circular Disk Failures Hitting k nodes", in Proc. International Workshop on Design Of Reliable Communication Networks (DRCN), Munich, Germany, 2017.

[C12] B. Vass, E. Bérczi-Kovács, and J. Tapolcai, "Enumerating Circular Disk Failures Covering a Single Node", in Int. Workshop on Resilient Networks Design and Modeling (RNDM), Halmstad, Sweden, 2016.

[C13] B. Vass, E. Bérczi-Kovács, and J. Tapolcai, "Shared Risk Link Group Enumeration of Node Excluding Disaster Failures", in Int. Conference on Networking and Network Applications (NaNA), Hakodate (Hokkaido), Japan, 2016. Winner of Best Paper Award.

Posters

[P1] B. Vass, "Shared Risk Link Groups of Disaster Failures", in IEEE Conference on Computer Communications Student Poster (INFOCOM Poster), 2016.

Theses

[T1] B. Vass, "A Combinatorial Geometric Approach for Network Attack Analysis", M.Sc. Thesis (Advisors: E. Bérczi-Kovács and A. Kőrösi), ELTE, Budapest, Hungary

Other publications

[O1] B. Vass, "Hálózatok legfennebb r sugárban ható hibáinak felsorolása" 1st prize at the Scientific Students' Associations conference 2015 organized by the Mathematical Institute of the Eötvös Loránd University (in Hungarian)

Bibliography

[1] A. de Sousa and D. Santos, "The Minimum Cost D-Geodiverse Anycast Routing with Optimal Selection of Anycast Nodes," in *2019 15th International Conference on the Design of Reliable Communication Networks (DRCN)*, March 2019, pp. 21–28.

[2] F. Dikbiyik, M. Tornatore, and B. Mukherjee, "Minimizing the risk from disaster failures in optical backbone networks," *J. Lightw. Technol.*, vol. 32, no. 18, pp. 3175–3183, 2014.

[3] O. Gerstel, M. Jinno, A. Lord, and S. B. Yoo, "Elastic optical networking: A new dawn for the optical layer?" *Communications Magazine, IEEE*, vol. 50, no. 2, pp. s12–s20, 2012.

[4] T. Gomes, J. Tapolcai, C. Esposito, D. Hutchison, F. Kuipers, J. Rak, A. de Sousa, A. Iossifides, R. Travanca, J. André, L. Jorge, L. Martins, P. O. Ugalde, A. Pašić, D. Pezaros, S. Jouet, S. Secci, and M. Tornatore, "A survey of strategies for communication networks to protect against large-scale natural disasters," in *8th International Workshop on Resilient Networks Design and Modeling (RNDM)*, Sept 2016, pp. 11–22.

[5] M. F. Habib, M. Tornatore, M. De Leenheer, F. Dikbiyik, and B. Mukherjee, "Design of disaster-resilient optical datacenter networks," *J. Lightw. Technol.*, vol. 30, no. 16, pp. 2563–2573, 2012.

[6] I. B. B. Harter, D. Schupke, M. Hoffmann, G. Carle *et al.*, "Network virtualization for disaster resilience of cloud services," *IEEE Commun. Mag.*, vol. 52, no. 12, pp. 88–95, 2014.

[7] J. Heidemann, L. Quan, and Y. Pradkin, *A preliminary analysis of network outages during hurricane Sandy*. University of Southern California, Information Sciences Institute, 2012.

[8] X. Long, D. Tipper, and T. Gomes, "Measuring the survivability of networks to geographic correlated failures," *Optical Switching and Networking*, vol. 14, pp. 117–133, 2014.

[9] B. Mukherjee, M. Habib, and F. Dikbiyik, "Network adaptability from disaster disruptions and cascading failures," *IEEE Commun. Mag.*, vol. 52, no. 5, pp. 230–238, 2014.

[10] S. Neumayer, G. Zussman, R. Cohen, and E. Modiano, "Assessing the vulnerability of the

fiber infrastructure to disasters," *IEEE/ACM Trans. Netw.*, vol. 19, no. 6, pp. 1610–1623, 2011.

[11] R. Souza Couto, S. Secci, M. Mitre Campista, K. Costa, and L. Maciel, "Network design requirements for disaster resilience in IaaS clouds," *IEEE Commun. Mag.*, vol. 52, no. 10, pp. 52–58, 2014.

[12] A. Xie, X. Wang, and S. Lu, "Risk Minimization Routing Against Geographically Correlated Failures," *IEEE Access*, vol. 7, pp. 62 920–62 929, 2019.

[13] J. Rak, D. Hutchison, E. Calle, T. Gomes, M. Gunkel, P. Smith, J. Tapolcai, S. Verbrugge, and L. Wosinska, "RECODIS: Resilient communication services protecting end-user applications from disaster-based failures," in *18th International Conference on Transparent Optical Networks (ICTON)*, July 10-14 2016, invited paper. [Online]. Available: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7550596&isnumber=7550246

[14] Jacek Rak, et al., "COST Action Resilient communication services protecting end-user applications from disaster-based failures (RECODIS)," http://www.cost.eu/COST_Actions/ca/CA15127.

[15] A. Kushwaha, D. Kapadia, A. Gumaste, and A. Somani, "Designing Multi-Layer Provider Networks for Circular Disc Failures," in *International Conference on Optical Network Design and Modeling (ONDM)*, Dublin, Ireland, May 2018.

[16] H. Zang, C. Ou, and B. Mukherjee, "Path-protection routing and wavelength assignment (RWA) in WDM mesh networks under duct-layer constraints," *IEEE/ACM Transactions on Networking (TON)*, vol. 11, no. 2, pp. 248–258, 2003.

[17] P. K. Agarwal, A. Efrat, S. K. Ganjugunte, D. Hay, S. Sankararaman, and G. Zussman, "The resilience of WDM networks to probabilistic geographical failures," *IEEE/ACM Trans. Netw.*, vol. 21, no. 5, pp. 1525–1538, 2013.

[18] M. T. Gardner and C. Beard, "Evaluating geographic vulnerabilities in networks," in *IEEE Int. Communications Quality and Reliability Workshop (CQR)*, 2011, pp. 1–6.

[19] S. Trajanovski, F. Kuipers, P. Van Mieghem *et al.*, "Finding critical regions in a network," in *IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*. IEEE, 2013, pp. 223–228.

[20] J. Strand, A. L. Chiu, and R. Tkach, "Issues for routing in the optical layer," *IEEE Commun. Mag.*, vol. 39, no. 2, pp. 81–87, 2001.

[21] D. M. Masi, E. E. Smith, and M. J. Fischer, "Understanding and mitigating catastrophic disruption and attack," *Sigma Journal*, pp. 16–22, 2010.

[22] Y. Nemoto and K. Hamaguchi, "Resilient ICT research based on lessons learned from the Great East Japan Earthquake," *IEEE Commun. Mag.*, vol. 52, no. 3, pp. 38–43, 2014.

[23] A. Kwasinski, W. W. Weaver, P. L. Chapman, and P. T. Krein, "Telecommunications power plant damage assessment for Hurricane Katrina – site survey and follow-up results," *IEEE Systems Journal*, vol. 3, no. 3, pp. 277–287, 2009.

[24] Y. Ran, "Considerations and suggestions on improvement of communication network disaster countermeasures after the Wenchuan earthquake," *IEEE Commun. Mag.*, vol. 49, no. 1, pp. 44–47, 2011.

[25] G. O'Reilly, A. Jrad, R. Nagarajan, T. Brown, and S. Conrad, "Critical infrastructure analysis of telecom for natural disasters," in *Int. Telecommunications Network Strategy and Planning Symposium (Networks)*. IEEE, 2006, pp. 1–6.

[26] J. S. Foster Jr, E. Gjelde, W. R. Graham, R. J. Hermann, H. M. Kluepfel, R. L. Lawson, G. K. Soper, L. L. Wood, and J. B. Woodard, "Report of the commission to assess the threat to the united states from electromagnetic pulse (EMP) attack: Critical national infrastructures," DTIC Document, Tech. Rep., 2008.

[27] R. Wilhelm and C. Buckridge, "Mediterranean fibre cable cut – a RIPE NCC analysis," 2008. [Online]. Available: https://www.ripe.net/data-tools/projects/archive/mediterranean-fibre-cable-cut

[28] L. Mollenauer, J. Gordon, P. Mamyshev, I. Kaminow, and T. Koch, "Optical Fiber Telecommunications IIIA," 1997.

[29] J. P. Sterbenz, D. Hutchison, E. K. Çetinkaya, A. Jabbar, J. P. Rohrer, M. Schöller, and P. Smith, "Resilience and survivability in communication networks: Strategies, principles, and survey of disciplines," *Computer Networks*, vol. 54, no. 8, pp. 1245–1265, 2010.

[30] J. Rak, *Resilient routing in communication networks*. Springer, 2015.

[31] Google maps. Accessed: 2021. [Online]. Available: https://www.google.com/maps/

[32] Renesys. Accessed: 2021. [Online]. Available: http://www.renesysgroup.com/

[33] Y. Kobayashi and K. Otsuki, "Max-flow min-cut theorem and faster algorithms in a circular disk failure model," in *IEEE INFOCOM*, 2014, pp. 1635–1643.

[34] "HUAWEI NE20E-S Universal Service Router, Feature Description - MPLS," 2018, v800R010C10SPC500.

[35] "Alcatel-Lucent 7705 SERVICE AGGREGATION ROUTER OS, MPLS GUIDE," 2013,

release 6.0.R4.

[36] "Cisco WAE Design 7.1.2 User Guide, Chapter 19 : LSP Disjoint Path Optimization," 2018.

[37] "Juniper Networks Junos OS MPLS Applications User Guide, Chapter 3 MPLS Traffic, Shared Risk Link Groups for MPLS," 2020.

[38] K. Pithewan, M. Yaaseen, R. P. Ramamoorthy, and M. Misra, "Disjoint path computation for arbitrary directed graph," 2016, US Patent 9,253,032.

[39] "Shaping Europe's digital future, Actors in the broadband value chain," 2019, European Commission, https://ec.europa.eu/digital-single-market/en/actors-broadband-value-chain.

[40] S. Trajanovski, F. A. Kuipers, A. Ilić, J. Crowcroft, and P. Van Mieghem, "Finding critical regions and region-disjoint paths in a network," *IEEE/ACM Trans. Netw.*, vol. 23, no. 3, pp. 908–921, 2015.

[41] J. Oostenbrink and F. Kuipers, "Computing the impact of disasters on networks," *ACM SIGMETRICS Performance Evaluation Review*, vol. 45, no. 2, pp. 107–110, 2017.

[42] F. Iqbal, S. Trajanovski, and F. Kuipers, "Detection of spatially-close fiber segments in optical networks," in *Design of Reliable Communication Networks (DRCN)*, 2016, pp. 95–102.

[43] S. Neumayer, G. Zussman, R. Cohen, and E. Modiano, "Assessing the impact of geographically correlated network failures," in *IEEE Military Communications Conference (MILCOM)*, 2008, pp. 1–6.

[44] A. Bernstein, D. Bienstock, D. Hay, M. Uzunoglu, and G. Zussman, "Power grid vulnerability to geographically correlated failures analysis and control implications," in *IEEE INFOCOM*, 2014, pp. 2634–2642.

[45] J. B. Orlin, "Max flows in O (nm) time, or better," in *Proceedings of the forty-fifth annual ACM symposium on Theory of computing*. ACM, 2013, pp. 765–774.

[46] H. Saito, H. Honda, and R. Kawahara, "Disaster avoidance control against heavy rainfall," in *IEEE INFOCOM 2017-IEEE Conference on Computer Communications*. IEEE, 2017, pp. 1–9.

[47] H. Honda and H. Saito, "Nation-Wide Disaster Avoidance Control Against Heavy Rain," *IEEE/ACM Transactions on Networking*, vol. 27, no. 3, pp. 1084–1097, 2019.

[48] L. Pašić, A. Pašić, F. Mogyorósi, and A. Pašić, "FRADIR Meets Availability," in *2020 16th International Conference on the Design of Reliable Communication Networks DRCN 2020*. IEEE, 2020, pp. 1–6.

[49] P. K. Agarwal, A. Efrat, S. K. Ganjugunte, D. Hay, S. Sankararaman, and G. Zussman, "Network vulnerability to single, multiple, and probabilistic physical attacks," in *Military Communications Conference (MILCOM)*. IEEE, 2010, pp. 1824–1829.

[50] X. Wang, X. Jiang, and A. Pattavina, "Assessing network vulnerability under probabilistic region failure model," in *IEEE High Performance Switching and Routing (HPSR)*, 2011, pp. 164–170.

[51] X. Wang, X. Jiang, A. Pattavina, and S. Lu, "Assessing physical network vulnerability under random line-segment failure model," in *IEEE High Performance Switching and Routing (HPSR)*, 2012, pp. 121–126.

[52] H. Saito, "Analysis of geometric disaster evaluation model for physical networks," *IEEE/ACM Trans. Netw.*, vol. 23, no. 6, pp. 1777–1789, 2015.

[53] ——, "Spatial design of physical network robust against earthquakes," *J. Lightw. Technol.*, vol. 33, no. 2, pp. 443–458, 2015.

[54] F. Iqbal and F. Kuipers, "Spatiotemporal Risk-Averse Routing," in *IEEE INFOCOM Workshop on Cross-Layer Cyber Physical Systems Security (CPSS)*, 2016.

[55] O. Gold and R. Cohen, "Coping with physical attacks on random network structures," in *IEEE ICC*, 2014, pp. 1166–1172.

[56] H.-W. Lee, E. Modiano, and K. Lee, "Diverse routing in networks with probabilistic failures," *IEEE/ACM Trans. Netw.*, vol. 18, no. 6, pp. 1895–1907, 2010.

[57] M. F. Habib, M. Tornatore, and B. Mukherjee, "Fault-tolerant virtual network mapping to provide content connectivity in optical networks," in *Optical Fiber Communication Conference*. OSA, 2013, pp. OTh3E–4.

[58] X. Li, Y. Wang, G. Zhang, X. Gao, Y. Zhao, and J. Zhang, "K-content connectivity in data center networks," in *Opto-Electronics and Communications Conference (OECC)*. IEEE, 2015, pp. 1–3.

[59] X. Li, S. Huang, S. Yin, B. Guo, Y. Zhao, J. Zhang, M. Zhang, and W. Gu, "Shared end-to-content backup path protection in k-node (edge) content connected elastic optical datacenter networks," *Optics express*, vol. 24, no. 9, pp. 9446–9464, 2016.

[60] B. Nedic, M. Gunkel, T. Gomes, and R. Girão-Silva, "SRLG-disjointness and geodiverse routing–a practical network study and operational conclusions," in *2018 10th International Workshop on Resilient Networks Design and Modeling (RNDM)*. IEEE, 2018, pp. 1–8.

[61] R. Girão-Silva, B. Nedic, M. Gunkel, and T. Gomes, "Shared Risk Link Group disjointness and geodiverse routing: A trade-off between benefit and practical effort," *Networks*, vol. 75, no. 4, pp. 374–391, 2020.

[62] K. Xie, H. Tao, X. Wang, G. Xie, J. Wen, J. Cao, and Z. Qin, "Divide And Conquer For Fast SRLG Disjoint Routing," in *48th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN)*. IEEE, 2018, pp. 622–633.

[63] C. Colman-Meixner, F. Dikbiyik, M. F. Habib, M. Tornatore, C.-N. Chuah, and B. Mukherjee, "Disaster-survivable cloud-network mapping," *Photonic Network Communications*, vol. 27, no. 3, pp. 141–153, 2014.

[64] J.-C. Bermond, D. Coudert, G. D'Angelo, and F. Z. Moataz, "SRLG-diverse routing with the star property," in *Design of Reliable Communication Networks (DRCN)*. IEEE, 2013, pp. 163–170.

[65] A. Hmaity, F. Musumeci, and M. Tornatore, "Survivable virtual network mapping to provide content connectivity against double-link failures," in *Design of Reliable Communication Networks (DRCN)*. IEEE, 2016, pp. 160–166.

[66] B. Jaumard and H. A. Hoang, "Design and dimensioning of logical survivable topologies against multiple failures," *Journal of Optical Communications and Networking*, vol. 5, no. 1, pp. 23–36, 2013.

[67] W. Ni, J. Wu, C. Huang, and M. Savoie, "Analytical models of flow availability in two-layer networks with dedicated path protection," *Optical Switching and Networking*, vol. 10, no. 1, pp. 62–76, 2013.

[68] J. Yallouz and A. Orda, "Tunable QoS-aware network survivability," in *INFOCOM*. IEEE, 2013, pp. 944–952.

[69] J. Yallouz, O. Rottenstreich, and A. Orda, "Tunable survivable spanning trees," *ACM SIGMETRICS Performance Evaluation Review*, vol. 42, no. 1, pp. 315–327, 2014.

[70] M. T. Gardner, R. May, C. Beard, and D. Medhi, "Finding geographic vulnerabilities in multilayer networks using reduced network state enumeration," in *Design of Reliable Communication Networks (DRCN), 2015 11th International Conference on the*, March 2015, pp. 49–56.

[71] A. de Sousa, "Improving the Connectivity Resilience of a Telecommunications Network to Multiple Link Failures Through a Third-Party Network," in *2020 16th International Conference on the Design of Reliable Communication Networks DRCN 2020*. IEEE, 2020, pp.

1–6.

[72] P. N. Tran and H. Saito, "Geographical route design of physical networks using earthquake risk information," *IEEE Communications Magazine*, vol. 54, no. 7, pp. 131–137, 2016.

[73] F. A. Kuipers, "An overview of algorithms for network survivability," *ISRN Communications and Networking*, vol. 2012, 2012.

[74] C. Doerr and F. A. Kuipers, "All quiet on the internet front?" *IEEE Commun. Mag.*, vol. 52, no. 10, pp. 46–51, 2014.

[75] M. F. Habib, M. Tornatore, F. Dikbiyik, and B. Mukherjee, "Disaster survivability in optical communication networks," *Computer Communications*, vol. 36, no. 6, pp. 630–644, 2013.

[76] J. Rak and D. Hutchison, Eds., *Guide to Disaster-Resilient Communication Networks*. Springer International Publishing, 2020.

[77] M. de Berg, O. Cheong, M. van Kreveld, and M. Overmars, *Computational Geometry: Algorithms and Applications*. Springer Berlin Heidelberg, 2008.

[78] F. Aurenhammer, "Voronoi diagrams: a survey of a fundamental geometric data structure," *ACM Computing Surveys (CSUR)*, vol. 23, no. 3, pp. 345–405, 1991.

[79] E. Papadopoulou and M. Zavershynskyi, "The higher-order Voronoi diagram of line segments," *Algorithmica*, vol. 74, no. 1, pp. 415–439, 2016.

[80] T.-H. Su and R.-C. Chang, "The k-Gabriel graphs and their applications," in *Algorithms. Lecture Notes in Computer Science, vol. 450.*, Berlin, Heidelberg, 1990.

[81] M. Abellanas, P. Bose, F. H. J. García, C. M. Nicolás, and P. A. Ramos, "On Structural and Graph Theoretic Properties of Higher Order Delaunay Graphs," in *International Journal of Computational Geometry & Applications, Volume 19, Issue 06*, 2009.

[82] B. Rozenfeld and N. D. Sergeeva, *Stereographic projection*. Moscow, USSR: Mir Publisher, 1977.

[83] H. S. M. Coxeter, "The Problem of Apollonius," in *The American Mathematical Monthly*, vol. 75, no. 1. Taylor & Francis, 1968, pp. 5–15.

[84] N. M. Nam, T. A. Nguyen, and J. Salinas, "Applications of Convex Analysis to the Smallest Intersecting Ball Problem," *Journal of Convex Analysis*, vol. 19, no. 2, pp. 497–518, 2012.

[85] S. Jadhav, A. Mukhopadhyay, and B. Bhattacharya, "An optimal algorithm for the intersection radius of a set of convex polygons," *Journal of Algorithms*, vol. 20, no. 2, pp. 244–267, 1996.

[86] J. J. Sylvester, “A question in the geometry of situation,” *Quarterly Journal of Pure and Applied Mathematics*, vol. 1, 1857.

[87] J. Giardini, D. and. Woessner, L. Danciu, H. Crowley, F. Cotton, G. Grunthal, R. Pinho, G. Valensise, S. Akkar, R. Arvidsson, R. Basili, T. Cameelbeeck, A. Campos-Costa, J. Douglas, M. B. Demircioglu, M. Erdik, J. Fonseca, B. Glavatovic, C. Lindholm, K. Makropoulos, C. Meletti, R. Musson, K. Pitilakis, K. Sesetyan, D. Stromeyer, M. Stucchi, and A. Rovida. (2013) Seismic Hazard Harmonization in Europe (SHARE): Online Data. [Online]. Available: http://www.share-eu.org/

[88] B. Chazelle and H. Edelsbrunner, “An optimal algorithm for intersecting line segments in the plane,” *Journal of the ACM (JACM)*, vol. 39, no. 1, pp. 1–54, 1992.

[89] D. Eppstein, M. T. Goodrich, and D. Strash, “Linear-Time Algorithms for Geometric Graphs with Sublinearly Many Edge Crossings,” *SIAM Journal on Computing*, vol. 39, no. 8, pp. 3814–3829, 2010.

[90] D. Eppstein and M. T. Goodrich, “Studying (non-planar) road networks through an algorithmic lens,” in *Proceedings of the 16th ACM SIGSPATIAL international conference on advances in geographic information systems*, 2008, pp. 1–10.

[91] E. Ackerman, “On topological graphs with at most four crossings per edge,” *Computational Geometry*, vol. 85, p. 101574, 2019. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0925772119301154

[92] R. Kershner, “The number of circles covering a set,” *American Journal of Mathematics*, vol. 61, no. 3, pp. 665–671, 1939. [Online]. Available: http://www.jstor.org/stable/2371320

[93] J. Matoušek, *Lectures on discrete geometry*. Graduate Texts in Mathematics 212, Springer-Verlag, New York, NY, 2002.

[94] D. Haussler and E. Welzl, “ϵ-nets and simplex range queries,” *Discrete & Computational Geometry*, vol. 2, no. 2, pp. 127–151, 1987.

[95] V. AYC, “Vapnik," on the uniform convergence of relative frequencies of events to their probabilities,",” *Theory of Probability and its Applications*, vol. 16, 1971.

[96] J. P. Snyder, *Map projections: A working manual*. U.S. Government Printing Office, 1987, vol. 1395.

[97] G. S. Lakhina and B. T. Tsurutani, “Geomagnetic storms: historical perspective to modern view,” *Geoscience Letters*, vol. 3, no. 1, p. 5, 2016.

[98] S. Neumayer, A. Efrat, and E. Modiano, "Geographic max-flow and min-cut under a circular disk failure model," *Computer Networks*, vol. 77, pp. 117–127, 2015.

[99] F. Aurenhammer, R. Klein, and D. Lee, *Voronoi diagrams and Delaunay triangulations*. World Scientific Publishing Co Inc, 2013.

[100] Lecture on stereograhpic circles. Accessed: 2018. [Online]. Available: http://www.cs.bsu.edu/homepages/fischer/math345/stereo.pdf

[101] S. Orlowski, R. Wessäly, M. Pióro, and A. Tomaszewski, "SNDlib 1.0: survivable network design library," *Networks*, vol. 55, no. 3, pp. 276–286, 2010.

[102] M. Rahnamay-Naeini, J. E. Pezoa, G. Azar, N. Ghani, and M. M. Hayat, "Modeling stochastic correlated failures and their effects on network reliability," in *IEEE Int. Conf. on Comp. Comm. and Networks (ICCCN)*, 2011, pp. 1–6.

[103] US National Seismic Hazard Maps. Accessed: 2014. [Online]. Available: https://earthquake.usgs.gov/hazards/hazmaps/conterminous/

[104] C. Bohler, C.-H. Liu, E. Papadopoulou, and M. Zavershynskyi, "A randomized divide and conquer algorithm for higher-order abstract voronoi diagrams," in *Algorithms and Computation*, H.-K. Ahn and C.-S. Shin, Eds. Cham: Springer International Publishing, 2014, pp. 27–37.

[105] M. Axenovich and T. Ueckerdt, "Density of range capturing hypergraphs," *JoCG*, 2016, 21 pages.

[106] Network library. [Online]. Available: https://github.com/jtapolcai/regional-srlg/tree/master/psrlg

[107] H. Kanamori, "The energy release in great earthquakes," *Journal of Geophysical Research (1896-1977)*, vol. 82, no. 20, pp. 2981–2987, 1977. [Online]. Available: https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/JB082i020p02981

[108] A. Rovida, M. Locati, R. Camassi, B. Lolli, and P. Gasperini, "CPTI15, the 2015 version of the Parametric Catalogue of Italian Earthquakes," *Istituto Nazionale di Geofisica e Vulcanologia*, 2016. [Online]. Available: https://doi.org/10.6092/INGV.IT-CPTI15

[109] A. Sieberg, "Erdebeben," *Handbuch der Geophysic*, vol. 4, pp. 552–554, 1931.

[110] C. Mueller, "Earthquake Catalogs for the USGS National Seismic Hazard Maps," *Seismological Research Letters*, vol. 90, 10 2018.

[111] Y. Y. Kagan, "Statistical distributions of earthquake numbers: consequence of branching process," *Geophysical Journal International*, vol. 180, no. 3, pp. 1313–1328, 2010.

[112] J. K. Gardner and L. Knopoff, "Is the sequence of earthquakes in Southern California, with aftershocks removed, Poissonian?" *Bulletin of the Seismological Society of America*, vol. 64, pp. 1363–1367, 1974.

[113] A. Frankel, "Simulating strong motions of large earthquakes using recordings of small earthquakes: the Loma Prieta mainshock as a test case," *Bulletin of the Seismological Society of America*, vol. 85, no. 4, pp. 1144–1160, 1995. [Online]. Available: http://www.bssaonline.org/content/85/4/1144.short

[114] A. Valentini, F. Visini, and B. Pace, "Integrating faults and past earthquakes into a probabilistic seismic hazard model for peninsular italy," *Natural Hazards and Earth System Sciences*, vol. 17, pp. 2017–2039, 2017. [Online]. Available: https://doi.org/10.5194/nhess-17-2017-2017

[115] M. D. Petersen, A. M. Shumway, P. M. Powers, C. S. Mueller, M. P. Moschetti, A. D. Frankel, S. Rezaeian, D. E. McNamara, N. Luco, O. S. Boyd, K. S. Rukstales, K. S. Jaiswal, E. M. Thompson, S. M. Hoover, B. S. Clayton, E. H. Field, and Y. Zeng, "The 2018 update of the us national seismic hazard model: Overview of model and implications," *Earthquake Spectra*, vol. 36, no. 1, pp. 5–41, 2020. [Online]. Available: https://doi.org/10.1177/8755293019878199

[116] Y. Y. Kagan, "Seismic moment distribution revisited: I. Statistical results," *Geophysical Journal International*, vol. 148, no. 3, pp. 520–541, 2002. [Online]. Available: https://doi.org/10.1046/j.1365-246x.2002.01594.x

[117] D. H. Weichert, "Estimation of the earthquake recurrence parameters for unequal observation periods for different magnitudes," *Bulletin of the Seismological Society of America*, vol. 70, no. 4, pp. 1337–1346, 1980.

[118] S. Hiemer, J. Woessner, R. Basili, L. Danciu, D. Giardini, and S. Wiemer, "A smoothed stochastic earthquake rate model considering seismicity and fault moment release for Europe," *Geophysical Journal International*, vol. 198, pp. 1159–1172, 06 2014.

[119] P. C., D. Albarello, P. Gasperini, V. D'Amico, and B. Lolli, "The attenuation of seismic intensity in Italy, Part II: modeling and validation," *Bulletin of the Seismological Society of America*, vol. 98, no. 2, pp. 692–708, 2008.

[120] W. Bakun, "MMI attenuation and historical earthquakes in the basin and range province of western North America," *Bulletin of the Seismological Society of America*, vol. 96, no. 6, pp. 2206–2220, 2006. [Online]. Available: http://pubs.er.usgs.gov/publication/70028691

[121] "The internet topology zoo," accessed: 2021. [Online]. Available: http://www.topology-zoo.org/dataset.html