UNIVERSIDAD DE LA REPÚBLICA

FACULTAD DE INGENIERÍA
INSTITUTO DE COMPUTACIÓN - INCO
PEDECIBA Informática

P H D   T H E S I S
to obtain the title of

PhD on Informatics

of the Universidad de la República

Defended by

Graciela FERREIRA LEITES MUNDELL

Analysis and Optimization of Highly Reliable Systems

Academic Director: Franco ROBLEDO

Thesis Advisor: Pablo ROMERO

Thesis Advisor: Sergio NESMACHNOW

defended on October 26, 2018

Jury:

Reviewers :	Dr. Bruno TUFFIN	-	INRIA (Francia)
	Dr. Guillermo DURÁN	-	Instituto de Cálculo, Facultad de Ciencias Exactas y Naturales, UBA (Argentina)
President :	Dr. Eduardo FERNÁNDEZ	-	Universidad de la República (PEDECIBA Informática)
Examinators :	Dr. Alvaro PARDO	-	Universidad Católica del Uruguay (Uruguay)
	Dr. Eduardo MORENO	-	Universidad Adolfo Ibáñez (Chile)

Contents

1

Abstract

In the field of network design, the survivability property enables the network to maintain a certain level of network connectivity and quality of service under failure conditions.

In this thesis, survivability aspects of communication systems are studied. Aspects of reliability and vulnerability of network design are also addressed. The contributions are three-fold.

First, a *Hop Constrained node Survivable Network Design Problem* (HCSNDP) with optional (Steiner) nodes is modelled. This kind of problems are $\mathcal{NP}$-Hard. An exact integer linear model is built, focused on networks represented by graphs without rooted demands, considering costs in arcs and in Steiner nodes. In addition to the exact model, the calculation of lower and upper bounds to the optimal solution is included. Models were tested over several graphs and instances, in order to validate it in cases with known solution. An *Approximation Algorithm* is also developed in order to address a particular case of SNDP: the *Two Node Survivable Star Problem* (2NCSP) with optional nodes. This problem belongs to the class of $\mathcal{NP}$-Hard computational problems too.

Second, the research is focused on cascading failures and target/random attacks. *The Graph Fragmentation Problem* (GFP) is the result of a worst case analysis of a random attack. A fixed number of individuals for protection can be chosen, and a non-protected target node immediately destroys all reachable nodes. The goal is to minimize the expected number of destroyed nodes in the network. This problem belongs to the $\mathcal{NP}$-Hard class. A mathematical programming formulation is introduced and exact resolution for small instances as well as lower and upper bounds to the optimal solution. In addition to exact methods, we address the GFP by several approaches: metaheuristics, approximation algorithms, polytime methods for specific instances and exact methods in exponential time.

Finally, the concept of separability in stochastic binary systems is here introduced. *Stochastic Binary Systems* (SBS) represent a mathematical model of a multi-component on-off system subject to independent failures. The reliability evaluation of an SBS belongs to the $\mathcal{NP}$-Hard class. Therefore, we fully characterize separable systems using Han-Banach separation theorem for convex sets. Using this new concept of separable systems and Markov inequality, reliability bounds are provided for arbitrary SBS.

Keywords— Computation Complexity, Survivability, Graph Fragmentation Problem, Stochastic Binary Systems

List of publications issued from this thesis work

This thesis was written following a Swedish PhD style. The chapters of this thesis are based on the following published papers:

1. "A formulation for a Hop Constrained Survivable Network Design Problem ", Graciela Ferreira, Sergio Nesmachnow, and Franco Robledo. Yugoslav Journal of Operations Research, Volume 27, Number 14, Pages 427–438, Year 2017.

2. "An Approximation Algorithm for the Two-Node-Connected Star Problem with Steiner Nodes ", Graciela Ferreira, Franco Robledo, Pablo Romero. To appear in Proceedings of the International Conference 2018 on Applied Combinatorial Optimization, Bologna, Italy (ALIO/EURO 2018). Special Issue of Electronic Notes in Discrete Mathematics, Elsevier.

3. "Graph Fragmentation Problem for Natural Disaster Management ", Natalia Castro, Graciela Ferreira, Franco Robledo, and Pablo Romero. Proceedings of the International Workshop on Machine Learning, Optimization, and Big Data (MOD 2017). In: Nicosia G., Pardalos P., Giuffrida G., Umeton R. (eds) Machine Learning, Optimization, and Big Data. MOD 2017. Lecture Notes in Computer Science, vol 10710. Springer, Cham, pages 496-505.

4. "Graph Fragmentation Problem: Analysis and Synthesis ", Manuel Aprile, Natalia Castro, Graciela Ferreira, Juan Piccini, Franco Robledo and Pablo Romero. To appear. International Transactions in Operational Research.

5. "Building Reliability Bounds in Stochastic Binary Systems ", Héctor Cancela, Graciela Ferreira, Gustavo Guerberoff, Franco Robledo and Pablo Romero. To appear.Proceedings of the International Workshop on Resilient Networks Design and Modeling 2018, August 2018, Longyearbyen - Svalbard (SPITSBERGEN), Norway.

Acknowledgments

Chapter 1

Introduction

This research is framed in the network design area and aspects of survival, reliability and vulnerability are addressed. This thesis follows the Swedish style, and is organized in three parts in correspondence with each of these aspects and its structure is as follows:

- In Part I two papers are included related to subject of Survivability of Network Design Problem (Chapters 2 and 3). Formally, survivability is defined as the capacity of a network to remain operational after disturbances or failures in some components. We work with two different topologies, the first rather general, the second more specific. Both cases addressed are represented exactly by combinatorial optimization problems, using Integer Linear Programming. Being these problems $\mathcal{NP}$-hard, depending on the case, some procedures to calculate lower and upper bounds to the optimal solution, heuristics methods and approximate algorithms are introduced.

- Part II includes two papers related to subject of Graph Fragmentation Problem (Chapters 4 and 5). The GFP is the result of a worst case analysis of a random attack. A fixed number of individuals for protection can be chosen, and a non-protected target node immediately destroys all reachable nodes. The goal is to minimize the expected number of destroyed nodes in the network. GFP also belongs to the class of $\mathcal{NP}$-Hard problems. In this case, a reviewing of main advances in this topic is made, and an exact model, using Integer Quadratic Program, with some bounds are proposed and tested here.

- Part III is dedicated to study of Stochastic Binary Systems (SBS) and its reliability. SBS generalize the static reliability concept to any system composed of a number of components subject to independent failures with known probabilities, and where the operation or failure of the system as a whole is a function of the state of the individual components. In this sense, SBS are a more flexible tool for evaluating and optimizing the reliability of a wider spectrum of real systems, both in the networking area and in other quite different applications areas. Here one paper is presented 6, where the concept of *separable* stochastic binary systems is introduced, reliability bounds for arbitrary SBS are provided inspired by a measure of a distance to a separable system and some experimental results are included.

Main contributions are summarized bellow.

Survivability Network Design:

1. "A formulation for a Hop Constrained Survivable Network Design Problem ". Main contributions are:

   - An integer linear model for the Generalized Steiner Problem with weighted Steiner nodes and diameter constrained is introduced and tested.

   - Lower and upper bounds to the optimal solution are proposed and tested, which allow to approach effectively an optimal solution

2. "An Approximation Algorithm for the Two-Node-Connected Star Problem with Steiner Nodes". Main contributions are:

   - The Two-Node-Connected Star Problem with Steiner Nodes (2NCSP-SN) is here introduced.

   - A generalization of the factor-2 result is offered. Specifically, an approximation algorithm with factor 4α is introduced for the 2NCSP, being $\alpha \geq 1/2$ the relation between the link-costs from the backbone/access network.

   - An exact Integer Linear Programming (ILP) formulation for the 2NCSP-SN is presented.

   - A sensibility analysis is carried out in order to understand the effectiveness of our approximation algorithm.

Graph Fragmentation Problem

3. "Graph Fragmentation Problem for Natural Disaster Management". Main contributions are:

   - A mathematical programming formulation for the GFP is introduced.

   - Lower and upper bounds are obtained.

   - Testing of performance of our exact solution for the GFP under different test cases coming from real-life applications

3. "Graph Fragmentation Problem: Analysis and Synthesis ". Main contributions are:

   - this paper present a review and a comprehensive analysis for the GFP, its relation with Component Order Connectivity problem (COC).

   - a new feasible vulnerability/connectivity metrics is proposed as a synthesis.

   - improvements in IQP modeling.

Reliability in SBS

6. "Building Reliability Bounds in Stochastic Binary Systems". Main contributions are:

   - An efficient representation of separable systems is proposed. It considers $N + 1$ real numbers, being N the size of the system (measured as the number of components subject to failure)

   - The interplay between monotonicity and separability in stochastic binary systems are explored.

6

- A metric to find the closest separable system for any given SBS ϕ is provided. As a consequence, a distant-minimizer separable system ϕ^* was found, as well as upper and lower bounds ($\overline{\phi}$ and $\underline{\phi}$ respectively)

- Corresponding ILP formulations to find ϕ^*, $\overline{\phi}$ and $\underline{\phi}$ are proposed and solved using CPLEX.

- Reliability bounds for arbitrary SBS are found using the previous construction, duality and Chernoff inequality.

- A Proof-of-Concept shows the performance of the reliability bounds and the benefit of our new representation of separable systems.

Part I

Survivability Network Design

Chapter 2

A formulation for a Hop Constrained Survivable Network Design Problem

Survivability is defined as the capacity of a network to remain operational after failures in some components, and it is one of the critical requirements in network planning and design. In this chapter, survivability and quality of service concepts are combined, adding hop-constraints to connectivity aspects. This approach ensures that for every distinct pair of nodes, there exists a predefined number of edge/node disjoint paths, so that each such path does not exceed a given hop limit. The focus is on modelling networks represented by undirected graphs without rooted demands, considering costs in arcs and in optional (Steiner) nodes. An integer linear model is presented in order to solve exactly this problem. The exact model is is complemented with the calculation of lower and upper bounds to the optimal solution and with an heuristic method to effectively solve large instances.

A formulation for a Hop Constrained Survivable Network Design Problem

Graciela Ferreira
Facultad de Ingeniería, Universidad de la República
Montevideo, Uruguay
gferreira@fing.edu.uy

Sergio Nesmachnow
Facultad de Ingeniería, Universidad de la República
Montevideo, Uruguay
sergion@fing.edu.uy

Franco Robledo
Facultad de Ingeniería, Universidad de la República
Montevideo, Uruguay
frobledo@fing.edu.uy

Abstract: This article presents an integer linear model for the hop constrained node survivable network design problem. The formulation is focused on networks represented by undirected graphs with not rooted demands, considering costs in arcs and in optional (Steiner) nodes, too. The proposed model allows setting different values of parameters for constraints between each pair of terminal nodes, including hop length and number of node disjoint paths constraints. This work includes calculating lower and upper bounds to the optimal solution. Since this kind of problems are NP-hard, it is useful to combine the presented formulation with heuristic methods in order to solve effectively large problem instances. The model was tested over the graphs with up to 85 nodes and 148 arcs, in order to validate it in cases with known solution.

1. Introduction

In network design, the survivability property enables the network to maintain a certain level of network connectivity and quality of service under failure conditions. Survivability has been considered as one of the critical requirements in network planning and design [5]. It often involves considering design requirements on the network topology, or in the case of communication networks, constraints could be associated with protocol, bandwidth allocation, etc. For instance, a topology requirement would achieve a design that keeps a minimum two-connected network against any failure of a single link or node. This concept can be applied to multiple types of networks, such as communication, power, transportation network, etc. (see for instance [3] [8]).

Formally, survivability is defined as the capacity of a network to remain operational after disturbances or failures in some components [1]. The survivable network design problem has been extensively studied [5, 3, 8, 10] and it is known to be NP-hard[6]. Survivability properties are usually modeled by requiring a minimal number of node -or edge- disjoint paths between certain pairs of nodes.

In this article, we combine survivability and quality of service concepts for the problem that imposes additionally hop-constraints when designing survivable networks. This approach ensures that for every distinct pair of nodes, there exists a predefined number of edge/node disjoint paths, so that each such path does not exceed a given hop limit.

Recent literature calls this kind of problems as *Hop Constrained Survivable Network Design Problem* (HCSNDP) [5, 2, 7]. The HCSNDP proposes finding the optimal network design with survivability requirements and effectiveness in quality of service (e.g., the maximum length of paths is bounded).

We focus on solving a variant of the HCSNDP that is applied to model networks represented by undirected graphs with not rooted demands, considering costs in arcs and in optional (Steiner) nodes, too. Different values of parameters for constraints between each pair of terminal nodes are allowed in the problem formulation, including hop length and number of node disjoint paths constraints. This is a generalization of HCSNDP that we named as "Generalized Steiner Problem with weighted Steiner nodes and diameter constrained" (GSPWDC).

We introduce an integer linear model for the GSPWDC. The exact model is tested over some graphs in order to perform a validation in cases with known solutions. Since this kind of problems is NP-hard, we propose and test some procedures to calculate lower and upper bounds to the optimal solution, which allow to approach effectively an optimal solution. These bounds are very useful when facing large problem instances.

The paper is organized as follows. Next section introduces the main concepts and definitions related to the problem model and formulations. The proposed problem formulation is presented in Section 3. Lower and upper bounds to the optimal solution are analyzed in Section 4. The experimental evaluation is reported and discussed in Section 5. Finally, Section 6 presents the conclusions and the main lines for future work.

2. Background

A network is represented as a graph, $G = (V, E)$. We consider only two possible types of nodes: *terminal* nodes, for which connectivity requirements are defined (set T), and optional or *Steiner* nodes (set S).

There are two models to specify the survivability conditions [6]. In this work we follow the *Generalizad Steiner Problem* (GSP) approach by Winter [11]: given a network represented by a graph $G = (V, E)$, with costs associated to edges, let $T \subseteq V$ be the set of terminal nodes and let $Q = \{q = (i, j), \forall i, j \in T\}$ be the set of all pairs of nodes in $T \subseteq V$; the problem is to find a subgraph with minimal cost so that $\forall q \in Q$ at least $r_q \in N$ node -or edge- disjoint paths exist. In the case where $T = V$ and $r_q = k, \forall q \in Q$, with cost associated to the edges and node disjoints paths, the problem is also known as $NCON(G, k)$ [9]

In order to model the HCSNDP, we use a variant of the GSP, that we call "Generalized Steiner Problem with weighted Steiner nodes and diameter constrained (GSPWDC)", which is formulated as follows:

Given a undirected simple graph $G = (V, E)$, with: (i) a set of edge cost or weights $C = \{c_{ij} \in R^+\}, \forall (i, j) \in E$; (ii) a set of terminal nodes $T \subseteq V$; (iii) $Q = \{q\}$ the set of pairs of nodes in $T \subseteq V$, (iv) a matrix with node (or arc) connectivity requirements $R = \{R_q\} \forall q \in Q$; (v) a vector with node weights $A = \{a_i \in R^+, i \in S = V \setminus T\}$; and (vi) a matrix of maximum length of paths allowed (*hop requirements*) $L = \{L_q \geq 0, integer, \forall q \in Q\}$; the GSPWDC consist in finding a minimal cost subgraph $H \subseteq G$ such that it covers T and $\forall q \in Q$ exist at least R_q node or arc disjoint paths linking a pair of terminal nodes $q = (i, j)$ in H, so that each one has no more than L_q hops or arcs.

Calling a pair of terminal nodes q as a *demand*, if all q have a common node, then the *demand* is called *rooted*, otherwise it is *unrooted*. According to Mahjoub [7], when $|Q| = 1$, the HCSNDP can be solved in polynomial time for $L \leq 3$, and it is NP-hard for $L \geq 4$. When $|Q|$ is not constrained, the problem is NP-hard, even in simplest case when Q is *rooted*, $R=1$ and $L=2$ $\forall q \in Q$.

The review of related work about models and formulations for the hop-constrained survivable network design problem allows identifying four existing models. The first three models are all variants of the same approach, initially proposed by Gouveia in 1998 [4] and later completed by Botton [1]. All of them were proposed to solve HCSNDP and were also implemented and tested over different networks. The fourth model was presented as an ILP model to solve SNDP in the survey by Kerivin and Mahjoub [6]. This model does not have practical results reported, but it was used to extract some interesting properties of polytopes corresponding to constraints space instead.

All the existing models for hop-constrained survivable networks use the same main idea to represent hop constraints. A set of auxiliary graphs are introduced: an auxiliary graph G^q—also called layered graph due to the method applied for building it— is defined for each $q \in Q$. Each graph G^q contains all existing paths between each pair of terminal nodes $q = (o, d)$ with length not greater than L_q (being L_q the number of hops allowed for $q = (o, d)$). The survivability constraints are formulated over these auxiliary graphs and hop constraints are implicitly considered because in the auxiliary graphs all paths have lengths shorter than L_q.

3. An ILP formulation for the HCSNDP

The proposed formulation is a variant of the existing models known as "*Hop-indexed formulation*", presented by Gouveia et al. [5] and "*Hop multi-commodity flow formulation* (HOP-MCF)" introduced by Botton [1]. The proposed model also incorporates some concepts used in recent works [6, 7, 4]. The model presented here allows considering constraints with different maximal length of paths between each pair of terminal nodes, as well as different numbers of required node-disjoint paths between each pair of terminal nodes (i.e., allowing *heterogeneous* survivability conditions). The model includes the cost of edges and nodes, too. In order to represent hop constraints, just like in the aforementioned previous works, we use extended layered graphs (one for each $q \in Q$) that implicitly guarantees satisfying the maximum path length constraints.

The main idea behind the proposed formulation is to decompose the problem into $|Q|$ subproblems, one for each pair of terminal nodes $q \in Q$. Let $(o(q), d(q))$ be a pair of origin–destination nodes corresponding to q. Fixed q, each subproblem is modeled with a directed graph composed of $L + 1$ layers (according to the transformation proposed by Gouveia [4]): being $G = (V, E)$ the original undirected graph, the alternative representation is $G^q = (V^q, A^q)$ where $V^q = V_1^q \cup \ldots \cup V_{L+1}^q$ / $V_1^q = o(q)$, $V_{L+1}^q = d(q)$ and $V_l^q \subseteq \{V \setminus \{o(q)\}\}$ such that there are a simple path between $o(q)$ and each $v \in V_l^q$ with length at most l, with $l = 2 \ldots L$.

Using the same notation proposed by Botton [3], let v_l^q be the copy of $v \in V$ in the l-th layer of graph G^q, then $A^q = \{(i_l^q, j_{l+1}^q)/(i, j) \in E, i_l^q \in V_l^q, j_{l+1}^q \in V_{l+1}^q, l \in \{1, \ldots, L\}\} \cup \{d(q)^l, d(q)^{l+1}, l \in \{2, \ldots L\}\}$. Details and graphical examples of the extended layered graphs were already provided by Botton [3, 1].

An edge in E with end points i and j is denoted as ij, while the arc between $i_l^q \in V_l^q$ and $j_{l+1}^q \in V_l^q$ in the directed graph is denoted as (i, j, l).

When using the proposed transformation, all paths from $o(q)$ to $d(q)$ in G^q fulfill hop-constraints.

Consider the following set of parameters:

i) a_i denotes the cost associated to each Steiner node i;

ii) c_{ij} is the cost associated to edge $ij, \forall ij \in E$;

iii) R_q is the minimal number of node-disjoint paths required between $o(q)$ and $d(q)$, $\forall q \in Q$;

iv) L_q is the maximum length allowed for paths (hops) between $o(q)$ and $d(q)$, $\forall q \in Q$.

Also, consider the following set of variables:

i) z_{ij} is a binary variable that indicates if edge $ij \in E$ is in the solution;

ii) $x_{ij}^{l,q}$ is the flow through arc (i, j, l), for each q in the layer l of G^q;

iii) N_i is a binary variable that indicates $\forall i \in S$ if the Steiner node i is included or not in the solution. Each N_i is used to allow at most one active outgoing arc from a node i over all layers of G^q, guaranteeing not to repeat nodes in a path between $o(q)$ and $d(q)$.

The ILP-HCSNDP formulation is presented next.

$$(ILP\text{-}HCSNDP) \quad \min \sum_{(i,j) \in E} c_{ij}.z_{ij} + \sum_{i \in V} a_i.N_i \text{ with:} \tag{1}$$

$$\sum_{j:(o(q),j,1) \in A^q} x^{1,q}_{o(q),j} = R_q \tag{2}$$

$$\sum_{j:(j,d(q),L_q) \in A^q} x^{L(q),q}_{j,d(q)} = -R_q \tag{3}$$

$$\sum_{j:(j,i,l-1) \in A^q} x^{l-1,q}_{ij} - \sum_{j:(i,j,l) \in A^q} x^{l,q}_{ji} = 0, \text{ for all } q \in Q, l \in \{2,\dots,L_q\}, i \in V^q_l \tag{4}$$

$$\sum_{l=1,\dots,L_q} x^{l,q}_{ij} + x^{l,q}_{ji} \le z_{ij}, \text{ for all } (i,j) \in E, q \in Q \tag{5}$$

$$z_{ij} \in \{0,1\}, \text{ for all } (i,j) \in E \tag{6}$$

$$x^{l,q}_{ij} \ge 0 \text{ integer for all } (i,j,l) \in A^q, q \in Q \tag{7}$$

$$\sum_{l=1,\dots,L_q} \sum_{j \in \delta(i)} x^{lq}_{ij} \le N_i \text{ for all } q \in Q, (i,j) \in E, \text{ and } i \in S \tag{8}$$

$$N_i \in \{0,1\}, \text{ for all } i \in S \tag{9}$$

In the ILP-HCSNDP formulation, Equation (1) is the objective function: it proposes minimizing the costs associated to arcs and Steiner nodes. Regarding the constraints, the network flow over G^q (Equations (2), (3), and (4)) assure that there are R_q paths from $o(q)$ to $d(q)$. Equation (5) does not enable using multiple times a given edge ij on a path in G^q, thus guaranteeing the edge-disjointness property, while they also link variables z and flow variables of copies of the same arc in different layers, which means that the total unimodularity property of matrix restriction is lost [1]. As a consequence, a set of constraints (defined in Equation (7)) must be explicitly introduced to obtain a feasible solution.

Equations (2)–(7) are present in the model by Botton; our formulation includes new constraints (defined in Equation (8) and Equation (9)) in order to guarantee the existence of node-disjoint paths and to allow the model to represent the costs associated to Steiner nodes. Note that $x^{l,q}_{ij} \le 1$, $i \ne j$, as stated by Equation (5) and Equation (6).

Constraints (stated in Equations (2), (3), and (4)) represent $|Q|$ independent sets of network flow constraints, one set defined for each $q \in Q$. Then, if constraints defined in Equations (5)–(9) are relaxed, there will be $|Q|$ independent network flow problems to solve, and the solution for each variable x^{lq}_{ij} will be integer. So, the model can take advantage of following a constraint decomposition approach, but in this case, the difficulty

is that in the objective function, variables x_{ij}^{lq} have no costs.

A simple idea to easily test the feasibility of constraints related to maximum number of node-disjoint paths allowed, is to consider for each $q \in Q$ a set $CS(q)$ where $CS(q) \subseteq V \setminus \{o(q), d(q)\}$ such that nodes $o(q)$ and $d(q)$ are not connected by a path in subgraph $G' \subseteq G$ induced by $V' = V \setminus CS(q)$. So, CS is a cut set between $o(q)$ and $d(q)$ nodes. Applying Menger's theorem [6], for each q, the minimum size of $CS(q)$ indicates the maximum number of node-disjoint $o(q)$-$d(q)$ paths.

4. Lower and upper bounds

Due to the intrinsic complexity of ILP-HCSNDP, heuristic and approximate approaches have to be used to cope with general and real-world instances of medium and large dimension. This section proposes and describes a set of procedures to calculate bounds to the exact solution of the problem.

We propose computing two lower bounds (called LB1, LB2) and four upper bounds (called UB1, UB2, UB3,UB4), which are defined and explained below. The proposal for computing upper bounds is based on a general idea: when fixing values of all variables z_{ij} and N_i, then only x_{ij} variables remain, and the original problem can be separated in $|Q|$ independent subproblems, which can be solved independently, too.

Lower bounds:

- LB1: it is an optimal solution of the ILP-HCSNDP integer relaxation. Here we work over G^q, but it is a linear program working in real variables.

- LB2: it is an optimal solution of the ILP-HCSNDP relaxing constraints of maximal length of paths. Here we work over original graph G, and integer variables.

Upper bounds:

- UB1: it fixes all boolean variables z_{ij} and N_i to one. For each one of $|Q|$ independent subproblems, we find the optimal solutions. Then, we calculate which arcs we must include in a global feasible solution in order to support all optimal solutions of the subproblems. If some of this subproblems are not feasible then, global problem is not feasible.

- UB2: it fixes all boolean variables z_{ij} and N_i, some to one and other to zero. For each one of $|Q|$ subproblems, we find optimal solutions and then we calculate which arcs must be included in a global solution in order to support all optimal solutions of the subproblems. If some of this subproblems are infeasible, the solution is discarded.

- UB3: it is an improvement of UB1 and UB2 computed using a Dantzig-Wolfe decomposition (DW) applied over ILP-HCSNDP problem. Equations (2) are included in DW subproblem (network flow problem) and remaining constraints are in the main problem. Integer conditions should be included in the main problem, because the subproblem always has integer solutions. In this work we do not include integer conditions in the main problem, so we used results of DW decomposition only when solutions or upper bounds are integer and improve UB1 or UB2.

- UB4: A greedy heuristic algorithm that builds a feasible solution.

In UB1 and UB2 cases, as z_{ij} and N_i variables are fixed, we can work with separate $\mid Q \mid$ independent subproblems, so we could profit this condition working in parallel. In UB3, the subproblem of DW decomposition can be separated, too. For UB1,UB2, and UB3 we work over G^q, then may be hard to calculate a feasible solution due to high dimensions.

Algorithm 1 presents a pseudo-code of the heuristic applied to compute UB4. It is a greedy algorithm that include a diversification phase

The proposed heuristic for computing UB4 starts by calculating a minimum float cost problem over G_d, with capacity arcs constraints and one additional constraint (lines 12-13). In this float problem, the pair of nodes in q are taken as source and destination. All edges have maximal capacity equal to one, except the edge that links sink to source that has minimal capacity R_q. This assures to find R_q node-disjoint paths between nodes in q.

Let $x_{i,j}, \forall (i,j) \in E$ be decision variables. The constraint added in line 13, $\sum_{ij \in E} x_{ij} \le b$, with $b = R_q * L_q$ bounds the total path length. This constraint does not ensure that all paths have length lower than L_q, but it decreases the search domain in order to find a feasible solution. For solving this extended float problem, an integer linear programming can be used. It is even necessary because the last constraint breaks integrality property of the float solution.

After that, the float solution $s(q)$ found, which is a binary vector with $|E|$ elements, is tested in order to determine if hop constraints are satisfied (line 17). If there are some path that do not satisfy hop limits in last finding solution $s(q)$, then b is decreased by one (line 18) and the procedure is repeated. If the procedure does not attain a feasible solution and the extended float problem is infeasible for some value of b, the procedure adds cuts in order to exclude paths that do not satisfy hop limits.

The heuristic procedure finishes when it finds an infeasible problem for some q (line 24), i.e., the global problem does not have a solution, or when it finds a global feasible solution. In effect, as b is an integer and it is decreased by one at each iteration where the feasibility test for L_q fails, the procedure does finish in a finite number of steps. If cuts must be added, they are also performed in a finite number of steps.

Algorithm 1 Proposed heuristic algorithm A:

Require: $V, E, T, S, Q, C = \{c_{ij}, \forall(i,j) \in E\}, A = \{a_k, \forall k \in S\}, p, R, L$

Ensure: gs ▷ gs is a binary vector with dimension $|E|$

1: Build data structures: incidence matrix node-node, incidence matrix node-arcs
2: Convert undirected graph to directed, where each arc is replaced by a pair of edges with opposite senses.
3: Replace each Steiner node S by a pair of nodes (S_1, S_2) linked by an edge going from S_1 to S_2, such that all incoming edges to S are incoming to S_1 and all outgoing edges of S are outgoing edges of S_2. Let G_d be this directed graph.
4: Build incidence matrix node-node and incidence matrix node-arcs NEa for G_d.
5: Order Q by preferences: Sort list Q according to our R and L values in descendent order (first pairs q of terminal nodes with high $R(q)$ and $L(q)$) that is, we process first those pair of terminal nodes that allow the longest paths, which requires more path linkings between them.
6: Build an ordered list of Steiner nodes $ListS$, based on preferences calculated according to cost of nodes divided by our degree. Order it by increasing preferences
7: **for** all $q \in Q$ **do**
8:     **if** $ListS \neq \emptyset$ **then,**   ▷ Diversification step
9:         extract first element r of $ListS$ with probability p
10:         $c_r = -1$
11:     **end if**
                                                    ▷ Search a candidate solution
12:     Build a minimum float cost problem over G_d, with capacity arcs constraints
13:     Add a constraint $\sum_{ij \in E} x_{ij} \leq b$, where $b = R_q * L_q$
14:     **while** $b > 0$ **do**
15:          Solve a linear program built in the previous points. Let $s(q)$ be this solution.
16:         **if** $s(q) \neq \emptyset$ **then**
17:             **if** some path $s(q)$ does not satisfy hop limits R_q **then**
18:                 $b \leftarrow b - 1$
19:             **else**
20:                 $b = 0$
21:             **end if**
22:         **else**
23:             **if** $b = R_q * L_q$ **then**
24:                 Display "infeasible problem"
25:                 Stop
26:             **else**
27:                 Add a cut excluding paths that do not satisfy hop limits in $s(q)$
28:             **end if**
29:         **end if**
30:     **end while**
31:     Set to 0 all costs associated with edges in feasible solution corresponding to $s(q)$
32:     Modify $ListS$, extracting all elements with cost ≤ 0
33:     $gs = gs \lor s(q)$ ▷ In order to build a global feasible solution
34: **end for**

---

5. Computational experiments and results

This section presents the experimental evaluation of the proposed ILP-HCSNDP model. The analysis is oriented to validate in cases with known solution and to research their behavior on large instances.

5.1. Development and execution platform

The model was implemented in CPLEX 12.51 MIP solver, and the executions were performed on a eight-cores Intel i7 processor at 3.07 GHz having 16GB RAM.

5.2. Problem instances

The proposed formulation was tested over eight graphs, using different values for the size of sets Q, R, and L. This methodology also allows having heterogenous values in matrices R and L.

We decided to work with simple undirected graphs. Table 1 summarizes the main characteristics of the graphs used in the experimental analysis. Column D is the graph density, defined for this type of graphs as $2|E|/|V|(|V|-1)$. Note that the maximal value for D is 1 when solving a complete graph with $\frac{1}{2}|V|(|V|-1)$ edges.

Table 1: Graphs and instances used in the experimental evaluation

graph	$|V|$	$|E|$	D	instance
FR1	19	43	0.2515	I1
FR2	11	30	0.5454	I2,I3,I4,I10,I11,I12
EON	19	36	0.2105	I5, I13
NFSNET	14	52	0.5714	I6, I14
TA1	24	55	0.1993	I7
B1	50	63	0.0514	I8
B2	50	63	0.0514	I9
RAU2	85	148	0.0415	I15

Instance I1 is built from FR1, a simple graph for which it is easy to find a solution. This problem instance is used to tune the model. Instance FR2 has been studied as a NCON(G, r) instance, with $r_{ij}= 2$, $for\ all\ (i, j) \in E$,

where the edge costs satisfy the triangle inequality. We use this example to test our model in a case with known optimal solution. EON and NFSNET are graphs used in the article by Gouveia et al. [5]. The remaining graphs are taken from libraries of test sets available at Internet: TA1 is from Survivable fixed telecommunication Network Design library (SNDlib, http://sndlib.zib.de), B1 and B2 are instances of the Steiner Tree Problem from SteinLib (http://steinlib.zib.de)

Finally, RAU2 graph is a real-life scenario based on the current Uruguayan academic network (www.rau.edu.uy).

A given graph can be associated with several instances that differ in their parameter values. The last column in Table 1 indicates the problem instances created from each considered graph.

5.3. Numerical results and discussion

Table 2 reports the parameter values and the results obtained for each instance solved. A given value in columns labeled L, R, c_{ij} or a_i means that the respective parameter is constant for all paths between each pair of nodes in T. Otherwise, the "diff" label is used to report when using different values. Instances marked with * are cases with $R = 1$, where optimal solutions or upper bounds are known. Columns labeled *opt*, *const*, *bin*, *int* and *time* report for each instance the optimal value (when attained), the number of constraints, the number of binary variables, the number of integer variables, and the time (in seconds) to solve each instance, without including the time taken to calculate V^q, the set of nodes in the layered graph G^q. The number of constraints and variables before the CPLEX presolve stage are reported; applying a presolve method could significantly reduce that number.

Table 2: Instances details and experimental results

instance	$\lvert T \rvert$	$\lvert Q \rvert$	L	R	c_{ij}	a_i	opt	const	bin	int	time (s)
I1	2	1	5	4	1	1	32	820	60	431	0.04
I2	11	55	7	2	diff	0	25	4497	30	25396	2.58
I3	11	55	8	2	diff	0	24	4992	30	29191	1.45
I4*	11	55	7	1	diff	0	20	4497	30	25396	137.13
I5*	10	45	3	1	1	1	10	16980	45	9887	0.04
I6	6	15	2	2	diff	1	9	7164	60	3169	0.03
I7*	24	396	4	1	1	0	23	24553	51	117201	2.35
I8*	9	36	50	1	diff	0	82	174268	104	238090	1319.00
I9*	13	78	50	1	diff	0	83	339570	79	515729	63177.00
I10	11	55	diff	2	diff	0	24	4713	30	27133	1.55
I11	11	55	diff	2	diff	0	26	4704	30	26983	3.33
I12	11	55	diff	diff	diff	0	25	4704	30	26983	14.28
I13	10	45	diff	diff	diff	diff	12	16909	45	9374	0.06
I14	10	45	diff	diff	diff	diff	18	25519	67	11696	0.14
I15	20	190	diff	diff	diff	diff	6583	1944361	213	582187	8198.33

The experimental evaluation was performed over graphs with up to 85 nodes and 148 arcs. Most instances are solved in a few seconds, only two cases demanded more than an hour: instance I9 (about 20 hours) and instance I15 (about two hours). Instances I8 and I9 took longer to find the optimal solution; these are cases with R=1, and large sets L and T. According to the cases studied, the parameters that most influence the resolution time are $\lvert L \rvert$ and $\lvert T \rvert$; the first is related to the dimensions of G^q and the second to the number of graphs. In effect, the quantity of x variables, which are the most numerous, depends on $\lvert Q \rvert$, $\lvert L \rvert$ and $2 \times \lvert E \rvert$, since each G^q is a directed graph. The number of remaining variables (all binary) depends on the number of Steiner nodes and the quantity of undirected edges.

For a given specific instance (I15), we explored several values of $\lvert T \rvert$ in the experiments. The largest value of $\lvert T \rvert$ for which we were able to obtain results in reasonable execution times (less than 24 hours) was $\lvert T \rvert$= 20. This is a relevant result for our research community since all variants of I15 problem instance are built over the real infrastructure of our Uruguayan academic network, allowing to explore different configurations for survivability and quality of service for academic and research projects.

Table 3 reports the lower and upper bounds computed for each instance solved. The column labeled *Bound*, indicates the kind of bound reported. Possible values are: LB1, LB2, UB1, UB2, UB3, or UB4, whose meaning has already been described in 4.

Table 3: Lower and upper bounds

instance	optimum	LB1	LB2	UB1	UB2	UB3	UB4
I1	32	32	32				32
I2	25	24.14	24				70
I3	24	24	24				70
I4*	20	12.07	20	84	79	26	42
I5*	10	9	9	21	18	15	25
I6	9	9	6		17	24	19
I7*	23	20.8	22	51	51	43	65
I8*	82	72	82	131	141		131
I9*	83	72.5	83				148
I10	24	24	24				62
I11	26	24.36	24				62
I12	25	22.5	22				60
I13	12	11.5	10				24
I14	18	18	17				20
I15	6583	6278.25	4308				13211

6. Conclusions and future work

In this paper we presented and evaluated a formulation for the Hop Constrained Survivable Network Design Problem.

We focused on the node survivability case for networks represented by simple and undirected graphs, not rooted demands, and considering costs in arcs and Steiner nodes. Based on the related previous works, we have developed a new formulation that accounts for specific quality of service and survivability constraints. The proposed model allows a heterogeneous setting for the network by including different values for the length of paths (related to the quality of service) and the number of paths (related to the connectivity demands), between each pair of terminal nodes.

The proposed formulation was evaluated for medium-size instances with up to 85 nodes and 148 arcs. The evaluation accounted for a significantly large number of decision variables. The CPLEX implementation of the proposed formulation was able to effectively solve all but one instance to optimality. Most of the problems were solved in a few seconds. Instance I15, which is based on a real network, was solved in about two hours. No optimal solution was computed for instance I9. The results for instance I15 are relevant, as this case study models the current Uruguayan academic network.

We used relaxation methods to compute lower bounds for the problem. In addition, decomposition techniques and heuristic methods were proposed to find upper bounds. This approach allowed to find accurate lower bounds that are close to the optimal solution for the set of problem instances considered in this article.

In the case of upper bounds, results allowed to conclude that UB1, UB2, UB3 are hard to calculate for some instances. Nevertheless, UB4 can be computed faster and it allows computing results for all instances in less than one hour. In this last case, we find that the proposed heuristic is sensible to preference order in Q, but rather insensitive to chosen diversification technique.

The main lines for the actual and future work are related to: (i) improving UB4, trying to introduce an effective diversification technique or to develop a local search phase that explores neighborhoods of global feasible solutions, (ii) improving the techniques for constructing and managing the graph G^q and applying decomposition algorithms in order to be able to solve significantly larger instances applying the proposed ILP formulation, (iii) try to measure and to assure the distance to optimum of approach solutions, (iv) introducing new instances of graph with special topologies that put to test the algorithms or other instances, particular cases of HCSNDP problems, with known optimal solutions. Being the HCSNDP a NP-hard problem, this kind of (relaxed) exact algorithms may be useful when combined with heuristic methods in order to effectively solve large instances modeling real-life situations.

References

[1] BOTTON, Q. *Survivable network design with quality of service constraints*. PhD thesis, Université Catholique de Louvain, Louvain School of Management, 2010.

[2] BOTTON, Q., FORTZ, B., AND GOUVEIA, L. On the hop-constrained survivable network design problem with reliable edges. Working Paper 4, Centro de Investigação Operacional, Faculdade de Ciencias da Universidade de Lisboa, Lisboa, Portugal, 2014. Available at http://www.optimization-online.org/DB_FILE/2012/04/3424.pdf. Accessed on March 2017.

[3] BOTTON, Q., FORTZ, B., GOUVEIA, L., AND POSS, M. , "Benders decomposition for the hop-constrained survivable network design problem". *INFORMS Journal on Computing 25*, 1 (2013), 13–26.

[4] GOUVEIA, L. , "Using variable redefinition for computing lower bounds for minimum spanning and Steiner trees with hop constraints". *INFORMS Journal on Computing 10*, 2 (1998), 180–188.

[5] GOUVEIA, L., PATRÍCIO, P., AND DE SOUSA, A. , "Hop-constrained node survivable network design: An application to MPLS over WDM". *Networks and Spatial Economics 8*, 1 (2008), 3–21.

[6] KERIVIN, H., AND MAHJOUB, A. , "Design of survivable networks: A survey". *Networks 46*, 1 (2005), 1–21.

[7] MAHJOUB, R., SIMONETTI, L., AND UCHOA, E. Hop-level flow formulation for the hop constrained survivable network design problem. In *Network Optimization*. Springer, 2011, pp. 176–181.

[8] ORLOWSKI, S., AND WESSÄLY, R. The effect of hop limits on optimal cost in survivable network design. In *Telecommunications Planning: Innovations in Pricing, Network Design and Management*, S. Raghavan and G. Anandalingam, Eds. 2006.

[9] STEIGLITZ, K., WEINER, P., AND KLEITMAN, D. , "The design of minimum-cost survivable networks". *IEEE Transactions on Circuit Theory 16*, 4 (1969), 455–460.

[10] STOER, M. Design of survivable networks, volume 1531 of lecture notes in mathematics, 1992.

[11] WINTER, P. , "Steiner problem in networks: a survey". *Networks 17*, 2 (1987), 129–167.

Chapter 3

An Approximation Algorithm for the Two-Node-Connected Star Problem with Steiner Nodes

The Two-Node Connected Star Problem (2NCSP) is a natural extension of the Ring Star Problem (RSP) where the ring is replaced by an arbitrary two-node connected topology. In this chapter, an extension of the 2NCSP is proposed, allowing the introduction of optional Steiner nodes in the two-connected structure if it were economically convenient. An exact Integer Linear Programming (ILP) formulation for this problem is presented as well as an approximation algorithm.

An Approximation Algorithm for the Two-Node-Connected Star Problem with Steiner Nodes

Graciela Ferreira, Franco Robledo, Pablo Romero [1]

Instituto de Matemática y Estadística
Facultad de Ingeniería. Universidad de la República
Montevideo, Uruguay

Abstract

The goal in topological network design is to build a minimum-cost topology meeting specific real-life constraints. There is a cost-robustness trade-off under single and multiple failures.

Previous works in the field suggest that a backbone composed by a two-node-connected toplogy provides savings with respect to elementary cycles. Consequently, we introduce the Two-Node Connected Star Problem with Steiner Nodes (2NCSP-SN). The goal is to design a minimum-cost topology, where the backbone is two-node connected, the access network is connected in a *star* topology or by direct links to the backbone, and optional nodes (called Steiner nodes) could be included in the solution. The 2NCSP-SN belongs to the class of $\mathcal{NP}$-Hard problems. This promotes the development of heuristics and approximation algorithms.

An approximation algorithm of factor 4α for the 2NCSP-SN is introduced, being $\alpha \geq 1/2$ the cost-ratio between backbone and access links. This is a generalization of the well-known factor 2 for the design of minimum-cost two-connected spanning networks (if we fix $\alpha = 1/2$). Finally, an exact Integer Linear Programming (ILP) formulation is proposed in order to highlight the effectiveness of the approximation algorithm. The results confirm a small gap between the globally optimum solution and the topology offered by our approximation algorithm when the ratio α is close to $1/2$.

Keywords: Network Optimization, Approximation Algorithm, Integer Linear Programming

1 Motivation

The minimum-cost toplogy meeting simple connectivity is the Minimum Spanning Tree, or MST. In this basic setting, Greedy algorithm efficiently finds the MST [3]. However, the design of communication systems is more challenging, since the cost-redundancy trade-off should be tackled. In the physical layer, this means that the network must be robust under single failures of any component, or two-node connected.

A natural way to connect terminal nodes in order to fulfill two-connectivity is to consider an elementary cycle. The design of a minimum-cost elementary cycle is a celebrated problem, known as the Travelling Salesman Problem or TSP. Supported by the concept of reducitibility and Karp list, it is well-known that the TSP belongs to the class of $\mathcal{NP}$-Hard problems [2]. Indeed, a natural reduction to Hamiltonian tour is obtained using 0-1 costs in their links.

A foundational work credited by Clyde Monma et. al. confirms that elementary cycles are sub-optimal for the design of minimum-weight 2-node-connected spanning network problem or MW2CSNP [5]. However, the authors show that cost of the cheapest Hamiltonian tour, $C_{\mathcal{H}}$, always respects the inequality $C_{\mathcal{H}} \leq 4/3OPT$, being OPT the cost of the globally optimum solution. Furthermore, the bound is tight, since they build an asymptotic family of graphs such that the cost-ratio tends to 4/3. The authors find structural properties of the globally optimum solution for the MW2CSNP and, they remark that the optimum 2-edge connected is also 2-node connected. The MW2CSNP is at least as hard as the TSP (using 0-1 costs, OPT is not greater than the number of nodes iff the graph is Hamiltonian). Therefore, several authors address the MW2CSNP using heuristics and approximation algorithms. Christofides provides a 3/2-factor for metric TSP which, together with the 4/3 factor credited by Monma et. al. results in an approximation algorithm with factor 2 for the MW2CSNP [1]. This factor-2 can be extended to several topological network design problems such as Steiner Networks, using either combinatorial analysis or strong duality theorem from linear programming [8,9].

The physical implementation of fiber-optics communication imposed a new challenge, *geographical diversity*. As a consequence, real-life communication networks are hierarchically structured in a backbone and access network, where customers in the last-mile have elementary connectivity requirements [7]. Martine Labbé et. al. introduced a hierarquically organized network, called Ring Star Problem, where the backbone is a ring and the access network is a *star*, this is,

---

[1] Emails:{gferreira,frobledo,promero}@fing.edu.uy

direct links connected to the ring [4]. A natural topological extension called Two-Node Connected Star Problem (2NCSP) has been introduced by Recoba et. al. [6]. The ring is replaced by an arbitrary two-node connected topology, and the authors confirm savings with respect to optimal solutions for the RSP (which are in turn feasible for the 2NCSP). Here, an extension of the 2NCSP is proposed, by the introduction of optional Steiner nodes in feasible solutions.

Complexity: Being RSP and 2NCSP $\mathcal{NP}$-Hard problems, 2NCSP-SN is $\mathcal{NP}$-Hard too.

The contributions of this article are the following:

- The Two-Node-Connected Star Problem with Steiner Nodes (2NCSP-SN) is here introduced.

- A generalization of the factor-2 result is offered. Specifically, an approximation algorithm with factor 4α is introduced for the 2NCSP, being $\alpha \geq 1/2$ the relation between the link-costs from the backbone/access network.

- An exact Integer Linear Programming (ILP) formulation for the 2NCSP-SN is presented.

- A sensibility analysis is carried out in order to understand the effectiveness of our approximation algorithm.

The paper is organized as follows. Section 2 formally presents the 2NCSP-SN using an ILP formulation. An approximation algorithm with factor 4α is introduced in Section 3. A fair comparison between the globally optimum solution and the approximation algorithm is performed in Section 4. Finally, Section 5 presents concluding remarks and trends for future work.

2 Two-Node-Connected Star Problem with Steiner nodes

We are given a simple graph $G = (V, E)$, internal link-costs $IC = \{c_e\}_{e\in E}$, external link-costs $EC = \{d_e\}_{e\in E}$, and $V = S \cup T$, being T the terminal-set and S Steiner nodes with costs $\{a(s)\}_{s\in S}$. The goal in 2NCSP-SN is to build a minimum-cost spanning subgraph $H = (V_H, E_H)$, where $T \subseteq V_H$ and $V_H = I \cup L$, being L the set of leaf-nodes (in the access network), such that $deg_H(v) = 1, \forall v \in L$, and the induced subgraph $H(I)$ is two-node-connected.

Let us develop an ILP for the problem under study. The key idea is to consider connectivity requirement $r_q = 2$ for every pair of terminals q from the backbone, while $r_q = 1$ otherwise. Let $Q = \{q = (i, j), \forall i \neq j, i, j \in T \subseteq V\}$. Consider the following set of binary variables:

- $z_{ij} = 1$ iff $(i, j) \in E$ is in the backbone;

28

- $y_{ij} = 1$ iff $(i, j) \in E$ is in the access network;
- x_{ij}^q is the $i - j$ flow for every pair of terminals q;
- $p_i = 1$ iff the i is included in the access network.

An ILP formulation for the 2NCSP-SN can be expressed as follows:

$$\min_{H \subseteq G} c(H) = \sum_{ij \in E} c_{ij}.z_{ij} + \sum_{s \in S} a_s.p_s + \sum_{ij \in E} d_{ij}.y_{ij} \tag{1}$$

$$\text{s.t.} \sum_{j:(j,i) \in E} x_{ji}^q - \sum_{j:(i,j) \in Ed^q} x_{ij}^q = I(i).r_i \ \forall i \in V, \ \forall q = (q_o, q_d) \in Q \tag{2}$$

$$I(i) = 1 \ \forall i \in V \backslash \{q_o\}, \ I(q_o) = -1 \tag{3}$$

$$r_i = 0 \ \forall i \in V \backslash \{q_o, q_d\} \tag{4}$$

$$\max(1, p_{q_o} + p_{q_d}) \le r_i \le 1 + \min(p_{q_o}, p_{q_d}), \ \forall i \in \{q_o, q_d\}, \ \forall q \in Q \tag{5}$$

$$x_{ij}^q + x_{ji}^q \le z_{ij} + y_{ij}, \ \forall ij \in E, \ \forall q \in Q \tag{6}$$

$$\sum_{j \in \delta(i)} y_{ij} \le 1 + M p_i, \ \forall i \in T \tag{7}$$

$$\sum_{j \in \delta(i)} (z_{ij} + y_{ij}) \le M p_i, \ \forall i \in S \tag{8}$$

$$y_{ij} \le 2 - p_i - p_j, \ \forall ij \in E, i, j \in V \tag{9}$$

$$z_{ij} \le \min(p_i, p_j), \ \forall ij \in E, i, j \in V \tag{10}$$

$$z_{ij} + y_{ij} \le 1 \ \forall ij \in E, i, j \tag{11}$$

$$2p_i \le \sum_{j \in \delta(i)} z_{ij} \le M p_i, \ \forall i \in V \tag{12}$$

Where $\delta(i)$ is neighbor-set for node i, and M is an arbitrarily large integer. The objective function (1) is the contribution of internal/external connections and Steiner nodes. Constraints (2)-(5) ensure connectivity using Kirchhoff equations. Constraints (6) and (7) force one-way flow. By Constraint (8), optional Steiner nodes belong to the backbone, if needed. The definitions of binary variables y_{ij} and z_{ij} are captured by Constraints (9) and (10). Constraints (11) state that either y_{ij} or z_{ij} can be set to 1, but not both. Finally, Constraints (12) state that a terminal node from the backbone could have multiple links, but nodes that belong to the access network must have a single link.

3 Approximation Algorithm

From now on, we assume that the internal/external costs are positive and internal costs satisfy the triangle inequality. Without loss of generality, a complete graph $G = (V, E)$ is considered. Let $\alpha_e = \frac{c_e}{d_e}, \forall e \in E$ be the primary/secondary cost ratio for each arc. In this section we build an approximation algorithm for the the 2NCSP-SN of factor 4α, being α:

$$\alpha = \max_{e \in E} \{\alpha_e\} \tag{13}$$

Recall that Christofides's algorithm is a 3/2-factor for the metric TSP. The key concept of our approximation algorithm is Christofides in order to span the terminal-set with an elementary cycle. Greedy augmentations of the solution including Steiner nodes also takes place, whenever the cost is reduced.

In Line 1, *Christofides* is called in order to build an elementary cycle C that spans the terminal-set T. The corresponding solution is updated in Lines 2-3, where the backbone is C and the access network is empty yet. In the while-loop (Lines 4-11), Steiner nodes are greedily included in the backbone, whenever the cost is reduced (Line 5). If this happens, some terminal node v is included in the access network, and the evidence $s \in S$ is added to the backbone (Lines 6-7). Observe that candidate terminals $t \in T$ are iteratively checked (Line 9), and the condition $|J| \geq 3$ forces to have a cycle in the backbone. The corresponding feasible solution F is finally returned (Line 12).

Require: $G = (T \cup S, E), c(e), d(e) \forall e \in E,$
1: $C \leftarrow Christofides(G, c)$
2: $L \leftarrow \emptyset, I \leftarrow T, E' \leftarrow E(C), J \leftarrow T$
3: $F \leftarrow (L \cup I, E')$
4: **while** $|J| \geq 3$ **do**
5:     **if** there are $s \in S, (t, v), (v, w) \in F$: $c(t, v) + c(v, w) > d(v, s) + c(t, s) + c(s, w)$ **then**
6:         $L \leftarrow L \cup \{v\}, I \leftarrow I \cup \{s\} \setminus \{v\}, J \leftarrow J \setminus \{t, v\}$
7:         $E' \leftarrow E' \cup \{(t, s), (s, v), (s, w)\} \setminus \{(t, v), (v, w)\}$
8:     **else**
9:         $J \leftarrow J \setminus \{t\}$
10:    **end if**
11: **end while**
12: **return** $F = (I \cup L, E')$

Lemma 3.1 *If $L(F) \neq \emptyset$ then $\alpha > 1/2$*

Proof. By the triangle inequality, $c_{(t,v)} \leq c_{(t,s)} + c_{(s,v)}$ and $c_{(v,w)} \leq c_{(v,s)} + c_{(s,w)}$, so $c_{(t,v)} + c_{(v,w)} \leq c_{(t,s)} + 2.c_{(s,v)} + c_{(s,w)}$. If $L \neq \emptyset$, there exists $v \in L$, $s \in S \cap I$, $t, w \in I$ such that $c_{(t,v)} + c_{(v,w)} > d_{(v,s)} + c_{(t,s)} + c_{(s,w)}$,
Therefore $d_{(v,s)} = c_{(v,s)}/\alpha_{v,s} < 2.c_{(s,v)}$, so there exists $e = (v,s)$ such that $\alpha_e > 1/2$, that is $\alpha > 1/2$ with $\alpha = \max_{e \in E}\{\alpha_e\}$.

$\square$

Theorem 3.2 $c(F) \leq \max\{2, 4\alpha\} \times OPT$.

Proof. Let $G^* = (S^* \cup T, E^*)$ be the optimal solution, H the cheapest Hamilton tour spanning T and H^S the cheapest Hamilton tour spanning $T \cup S^*$. Analogously, let us denote TNC (TEC) to the optimal 2-node (resp. 2-edge) connected spanning subgraph for $T \cup S^*$. Recall that F is the output and C the cycle obtained using Christofides algorithm. Combining Monma and Christofides theorems:

$$c(G^*) \leq c(F) \leq c(C) \leq \frac{3}{2}c(H) \leq \left(\frac{3}{2}\right)\left(\frac{4}{3}\right)c(TNC) = 2c(TNC) \qquad (14)$$

Let $G^* = B \cup L$, being B its backbone. Consider an augmentation F' for G^*, doubling edges from L with cost $c_{r,j} = \alpha_{r,j}d_{r,j}$ and adding them. F' is 2-edge connected and

$$c(F') = c(B) + 2\sum_{r \in L} c_{r,j} \leq c(B) + 2\alpha\sum_{r \in L} d_{r,j} = 2\alpha c(G^*) + (1 - 2\alpha)c(B), \qquad (15)$$

If $1 - 2\alpha > 0$, $c(F') \leq 2\alpha c(G^*) + 2(1 - 2\alpha)c(G^*) \leq 2c(G^*)$. In this case a factor 2 is provided, and by Lemma 3.1 F consists of an elementary cycle.

Otherwise, combining (14) and (15) we have that:

$$\begin{aligned} c(G^*) \leq c(F) \leq 2c(TNC) = 2c(TEC) &\leq 2c(F') \\ &\leq 4\alpha c(G^*) + 2(1 - 2\alpha)c(B) \\ &\leq 4\alpha c(G^*). \end{aligned}$$

$\square$

4 Proof-of-Concept

In order to highlight the effectiveness of our approximation algorithm, a sensibility analysis with respect to the ratio α is carried out. We consider a single instance from TSPLIB named berlin52.tsp. This is the case of a real-life network with Euclidean

costs. In order to find the globally optimum solution, an induced subgraph with 22 nodes is considered (with 10 terminal-nodes and 12 Steiner nodes). The ILP has been executed in CPLEX 12.6.3 MIP solver using an Intel i7 processor, 2.30 GHz, 8GB RAM. Table 1 illustrates the performance $c(F)/OPT$ as a function of α. In these examples we take α_e constant: $\alpha_e = \alpha \ \forall e \in E$. The cycle C obtained using Christofides algorithm is $c(C) = 3164.8$, while the cheapest Hamiltonian tour H spanning the terminal-set has a cost $c(H) = 2826.5$. Naturally, since F considers greedy augmentations of C, we get that $c(F) \leq c(C)$ in all cases.

Table 1
Sensibility Analysis as a function of α

α	OPT	$c(F)$	$c(F)/OPT$
10	485.78	3117.3	6.42
4	1110.48	3164.8	2.85
2	2115.02	3164.8	1.50
4/3	2611.45	3164.8	1.21
1	2786.80	3164.8	1.14
4/5	2811.63	3164.8	1.13
2/3	2822.88	3164.8	1.12
4/7	2825.60	3164.8	1.12

Our approximation algorithm outperforms Christofides only when $\alpha = 10$. In this case Steiner nodes are included in the solution. The factor is far away from 4α in all cases. Furthermore, when $\alpha = 4/7$ the ratio 1.12 is the lowest. Curiously enough, the results suggest that the performance is consistently better when the ratio α is decreased.

5 Conclusions and Trends for Future Work

The Two-Node Connected Star Problem with Steiner Nodes (2NCSP-SN) is introduced. Its hardness promotes heuristics and approximation algorithms. Here, we introduce an approximation algorithm of factor 4α for the problem, being $\alpha \geq 1/2$ the minimum ratio among internal/external costs. An exact ILP formulation is also proposed. The celebrated factor 2 from Steiner networks is

retrieved when $\alpha = 1/2$. Furthermore, a proof of concept suggests that the performance of the algorithm is consistently better when α is close to $1/2$.

Currently, we are working to develop a full GRASP methodology enriched with Variable Neighborhood Descent (VND) to find high-competitive solutions for the 2NCSP-SN.

Acknowledgment

References

[1] N. Christofides. Worst-case analysis of a new heuristic for the travelling salesman problem. Technical Report 388, Graduate School of Industrial Administration, Carnegie Mellon University, 1976.

[2] R. Karp. Reducibility among combinatorial problems. In R. Miller and J. Thatcher, editors, *Complexity of Computer Computations*, pages 85–103. Plenum Press, 1972.

[3] Joseph B. Kruskal. On the Shortest Spanning Subtree of a Graph and the Traveling Salesman Problem. *Proceedings of the American Mathematical Society*, 7(1):48–50, February 1956.

[4] Martine Labbé, Gilbert Laporte, Inmaculada Rodríguez Martín, and Juan Jose Salazar Gonzalez. The ring star problem: Polyhedral analysis and exact algorithm. *Networks*, 43(3):177–189, 2004.

[5] Clyde L Monma, Beth Spellman Munson, and William R Pulleyblank. Minimum-weight two-connected spanning networks. *Mathematical Programming*, 46(1-3):153–171, 1990.

[6] Rodrigo Recoba, Franco Robledo, Pablo Romero, and Omar Viera. Two-node-connected star problem. *International Transactions in Operational Research*, 25(2):523–543, 2018.

[7] F. Robledo. *GRASP heuristics for Wide Area Network design*. PhD thesis, IRISA/INRIA Université de Rennes I, 2005.

[8] Vijay V Vazirani. *Approximation algorithms*. Springer Science & Business Media, 2003.

[9] David P. Williamson and David B. Shmoys. *The Design of Approximation Algorithms*. Cambridge University Press, New York, NY, USA, 1st edition, 2011.

Part II

Graph Fragmentation Problem

Chapter 4

Graph Fragmentation Problem for Natural Disaster Management

The Graph Fragmentation Problem (GFP) was defined as a worst case analysis of an abstract epidemic modelling, it can bee seen as a particular case of the Node Immunization Problem (NIP) when the level of virulence is innite. In the GFP a population is represented by a graph, and there is a budget constraint B, that allows to immunize or protect only B nodes or individuals of population. The nature picks a node v uniformly at random from population represented by a graph, if the individual is not protected, the disaster kills all the members of the same connected component as v. The goal is to minimize the expected number of deaths. In this chapter a mathematical programming formulation and exact resolution for small instances are introduced.

Graph Fragmentation Problem for Natural Disaster Management

Natalia Castro, Graciela Ferreira, Franco Robledo, and Pablo Romero

Instituto de Matemática y Estadística, IMERL
Facultad de Ingeniería - Universidad de la República
Montevideo Uruguay
`(ncastro,gferreira,frobledo,promero)@fing.edu.uy,`

Abstract. Natural disasters represent a threaten for the existence of human beings. Given its remarkable importance, operational researchers should contribute to provide rationale decisions.

In this paper we study a purely combinatorial problem that models management disasters, called Graph Fragmentation Problem, or GFP for short. The problem belongs to the $\mathcal{NP}$-Hard class. As corollary, finding the optimal protection scheme is prohibitive for large populations. First, we review the problem and its properties. Then, we introduce a mathematical programming formulation and exact resolution for small instances. Finally, we discuss feasible model extensions and trends for future work.

1 Motivation

History reveals painful memories full of pandemics, lighting shocks and fires. The Spanish flu from 1918 was deadlier than any war in history, and half the population of the world has been exposed to the virus [14]. An infernal fire in October 1871 ravaged part of Chicago, leaving more than 90.000 homeless and 300 deaths [10].

We encourage operational researchers to be engaged with society, and provide means to cope with natural disasters. In this paper, we follow the research line introduced in [12]. There, a single individual of a population is exposed to a natural disaster, and the disaster is immediately propagated through neighbors. Our task is to determine a sub-population that is protected beforehand, subject to a budget constraint. Clearly, the notion of protection depends on the specific application (location of fire-stations, isolation in electric systems, vaccination against a pandemics).

This paper is organized as follows. Section 2 presents the background of the problem under study, and its origin from epidemic modelling. Section 3 presents a formal definition of the GFP. Theoretical results for the GFP are presented in Section 4, together with the main approaches to address the problem. The main contributions are offered in Section 5, 6 and 7. Specifically, a mathematical programming formulation for the GFP is introduced in Section 5. Lower and upper bounds are obtained in Section 6, inspired by relaxations. An experimental analysis is carried out in Section 7, where we test the performance of our exact solution for the GFP under different test cases coming from real-life applications. It is worth to notice that the literature in the exact analysis of the GFP is scarce, and here we provide the

first steps towards the development of optimal protection schemes under this fundamental model. Section 8 presents feasible model extensions, concluding remarks and trends for future work.

2 Background

A cornerstone in epidemic model is classical SIR (Susceptible -Infected - Removed). In SIR it is assumed a fully-mixed infinite population with random contacts. More realistic models are available from authoritative literature in the field [9, 2]. They consider a graph and epidemic spread governed by probabilistic rules. The authors claim that node-protection (choosing which nodes to remove, so that the epidemic cannot propagate through them) is a presumably hard task, but they do not provide hints nor mathematical proofs.

The Graph Fragmentation Problem, or GFP, represents a worst case analysis of an abstract epidemic modelling. In [12], a realistic SIR-based model is provided, and the Graph Fragmentation Problem (GFP) is introduced as an extremal analysis of highly virulent scenarios. Incidentally, it models other catastrophic events, such as fire-fighting and electric shocks (the formal model is presented in Section 3). There, only Greedy-based heuristics are presented, and there is no complexity analysis. A GRASP heuristic enriched with a path-relinking post-optimization stage is developed in [13].

The first result on computational complexity is offered for the GFP in [11]. The authors prove that the GFP belongs to the class of $\mathcal{NP}$-Hard problems. This theoretical result confirms the intuition from epidemiologists that finding an optimal node-protection mechanism is a hard task.

Curiously enough, in a more recent paper, the optimal protection scheme is found in all acyclic graphs, elementary cycles and some bipartite graphs [1]. In contrast, GFP presents a strong inapproximability result for general graphs. More specifically, there is no approximation algorithm with factor lower than $3/2$, unless $\mathcal{P} = \mathcal{NP}$.

3 Graph Fragmentation Problem

We are given a population represented by a graph $G = (V, E)$, and a budget constraint B, which is a natural number B such that $0 \leq B \leq |V|$. We can choose B nodes and protect them: we delete the nodes from G obtaining a subgraph G', so that the chosen nodes cannot be affected by the disaster. The nature picks a node v uniformly at random from G'. The disaster kills all the members of the same connected component as v.

The goal is to minimize the expected number of deaths. Mathematically, if the subgraph G' has $V' = n$ nodes and k connected components with orders $n_1, \ldots, n_k$, the probability to choose component i is n_i/n. Therefore, the expected number of deaths is $E(G') = \sum_{i=1}^{k} n_i p_i$, with $p_i = n_i/n$. The goal of the Graph Fragmentation Problem

(GFP) is to choose the protected set in order to minimize the expected number of deaths:

$$\min_{U \subseteq V} \sum_{i=1}^{k} \frac{n_i^2}{n}$$
$$s.t. |U| \leq B.$$

Observe that the denominator n is constant for a fixed instance (G, B) in the GFP. Therefore, our problem is to minimize the Euclidean norm of the vector $n = (n_1, \ldots, n_k)$, or Constrained Euclidean Norm Minimization (CENM):

$$\min_{U \subseteq V} \|n_{G-U}\|^2$$
$$s.t. |U| \leq B,$$

where $n_{G-U} = (n_1, \ldots, n_k)$ is the vector with the orders of the connected components from $G' = G - U$. Observe that the objective function $\|n_{G-U}\|^2$ is minimized when the resulting graph $G' = G - U$ has isolated nodes. The reader is invited to consult [1, 6, 4] for a discussion of related vulnerability metrics.

4 Analysis

In this section we highlight the main ideas on the analysis of the GFP for a better understanding of the problem. The following problem will be used to characterize the computational complexity of the GFP.

Definition 1 (Minimum Cardinality Vertex Cover)
Instance: simple graph $G = (V, E)$ and positive integer k.
Does there exist a node-set U such that $|U| \leq k$ and every link is incident to some node from U?

Recall that Minimum Cardinality Vertex Cover belongs to Karp list of 21 $\mathcal{N}\mathcal{P}$-Complete decision problems [8].

Theorem 1. *The GFP belongs to the class of $\mathcal{N}\mathcal{P}$-Hard problems.*

Proof. The graph $G' = G - U$ has isolated nodes if and only if U is a vertex cover, where $|U| \leq B$. Thus, the GFP is at least as hard as Minimum Cardinality Vertex Cover.

The following problem will be considered in order to prove a stronger inapproximability result for the GFP:

Definition 2 (Multiway k-cut)
Instance: simple graph $G = (V, E)$, terminal set $K \subseteq V$ with $|K| = k$, positive integer B.
Does there exist a separator set $U \subseteq V - K$ with $|U| \leq B$ such that each terminal node belongs to different components in $G - U$?

We know that Multiway 2-cut is in $\mathscr{P}$. A polynomial time algorithm is provided by Ford and Fulkerson [5]. However, Multiway k-cut is $\mathscr{NP}$-Complete for every fixed $k \geq 3$ [3].

Theorem 2. *It is $\mathscr{NP}$-hard to approximate GFP within $\frac{5}{3} - \varepsilon$, for any $\varepsilon > 0$.*

Proof. Consider an instance of Multiway 3-cut with ground graph $G = (V, E)$, distinguished nodes $\{v_1, v_2, v_3\}$ and positive integer B. Replace those nodes by large cliques $\{K_N, K_N, K_N\}$, where $N >> |V|$. The order of the new graph G^* is roughly $3N$. If the instance accepts a 3-cut, the cost in the GFP with instance (G^*, B) is roughly N. Otherwise, the expected number of dead nodes is never lower than $\frac{(2N)^2 + N^2}{3N} = \frac{5N}{3}$. Therefore, an approximation algorithm with factor $5/3$ would decide if G with distinguished nodes $\{v_1, v_2, v_3\}$ accepts a 3-cut using B nodes. The existence of such algorithm implies the solution of 3-cut.

Even though the GFP does not accept an optimal solution in polynomial time (unless $\mathscr{P} = \mathscr{NP}$), there exists a dynamic programming-based polynomial time method to find the optimal solution in acyclic graphs:

Theorem 3. *If G is acyclic, there exists a polynomial time algorithm to find the best protection scheme with B nodes.*

Proof. First, consider arbitrary graphs G_1 and G_2 that accept a polynomial time algorithm for any B, then we can solve the problem for $G = G_1 \cup G_2$, using all partitions $B = B_1 + B_2$. This reasoning holds for disjoint branches of a rooted tree (the root is arbitrary in this context). We can consider leaf nodes and their parents, and proceed with disjoint branches as before. The number of stages in a dynamic programming algorithm is not more than the height of the tree (which is not greater than the order of the graph). Finally, the result hold for acyclic graphs. Just connect all the trees by a fixed auxiliary node and consider the previous algorithm for the resulting tree. The reader is invited to consult [1] for technical details.

Theorem 4. *The size of the connected components in $G' = G - U$ must be as even as possible.*

Proof. Let $n_1 \geq n_2 \geq \ldots \geq n_k$ be the orders of the connected components in G'. If $|n_1 - n_k| \geq 2$, a straight calculation shows that $\|(n_1, \ldots, n_k)\|^2 \geq \|(n_1 - 1, \ldots, n_k + 1)\|^2$. This means that the cost in the GFP is reduced whenever the size of the components in G' are as even as possible.

There is no general result for cyclic graphs in general. However, the following result holds for the elementary cycle:

Theorem 5. *The best protection scheme is known for the cycle C_n.*

Proof. Delete an arbitrary node, and obtain an elementary path. Then, protect $B - 1$ nodes in such a way that the resulting sub-paths are as even as possible. By Theorem 4, the resulting graph provides the minimum-cost protection scheme.

41

Let us further analyze the GFP for bipartite graphs. Consider $G = (V_1 \cup V_2, E)$ where $E \subseteq V_1 \times V_2$. Recall that König theorem asserts that the minimum cardinality of a vertex cover in bipartite graphs is precisely the size of the maximum matching L. This number can be found by Ford and Fulkerson algorithm: connect all nodes from V_1 to a source s, all the nodes from V_2 to a sink t, and find the max-flow with unit capacities in the links. If $B \geq L$, all nodes from a vertex cover can be protected, and they can be found in polynomial time. We obtain the following:

Theorem 6. *The optimality for the GFP can be found in polynomial time for all bipartite graphs whenever B is not lower than the maximum matching.*

The computational complexity for the GFP remains open for bipartite graphs in general.

5   Mathematical Programming Formulation

An integer quadratic programming model (IQP) for the GFP is developed. In the model we consider a directed graph $G^d = (V, E')$, where every link from G is replaced by two one-way links. Consider the following model variables:

- n_k: size of connected component k;

- $U_i \in \{0, 1\}$, $i \in V$: node $i \in U$ (or not);

- $x_{ij}^k \in \{0, 1\}$, $(i, j) \in E$: link (i, j) belongs to component k in G;

- $N_i^k \in \{0, 1\}$, $i \in V$: node i belongs to the component k;

- $y_{ij}^{u,v} \in \{0, 1\}$, $(i, j) \in E$, $u, v \in V$: there is some u-v-path that includes (i, j) in the way $i \rightarrow j$.

The mathematical programming model is the following:

$$\min. \sum_{i=1..K} n_i^2/(n-B) \tag{1}$$

$$\text{s.t.} \sum_{j \in V} U_j \leq B, \tag{2}$$

$$\sum_{j \in V} N_j^k = n_k, \forall k = 1 \ldots K \tag{3}$$

$$\sum_{k=1...K} N_j^k = 1 - U_j, \forall j \in V \tag{4}$$

$$N_i^k + N_j^s \leq 1, \forall i,j \in V, i \neq j, (i,j) \in E, \forall k,s \in K, s \neq k, \tag{5}$$

$$\sum_{k=1...K} x_{ij}^k \leq (1 - U_i), \forall (i,j) \in E, i,j \in V \tag{6}$$

$$\sum_{k=1...K} x_{ij}^k \leq (1 - U_j), \forall (i,j) \in E, i,j \in V \tag{7}$$

$$N_i^k + N_j^k \leq 1 + x_{ij}^k, \forall k \in 1 \ldots K, \forall (i,j) \in E, i,j \in V \tag{8}$$

$$y_{i,j}^{u,v} + y_{j,i}^{u,v} \leq \sum_{k=1...K} x_{ij}^k, \forall u,v \in V, \forall (i,j) \in E, i,j \in V \tag{9}$$

$$\sum_{(u,j) \in E'} y_{u,j}^{u,v} \geq N_u^k + N_v^k - 1, \forall u,v \in V, \forall k \in 1 \ldots K \tag{10}$$

$$\sum_{(u,j) \in E'} y_{u,j}^{u,v} = \sum_{(i,v) \in E'} y_{i,v}^{u,v}, \forall u,v \in V \tag{11}$$

$$\sum_{(r,j) \in E'} y_{r,j}^{u,v} = \sum_{(i,r) \in E'} y_{i,r}^{u,v}, \forall r,u,v \in V, r \neq u, r \neq v \tag{12}$$

The objective function captures de cost of the GFP (1). Inequality (2) represents the budget constraint. The size of each connected component is found using Constraint (3). Constraints (4) set $N_j^k = 0$ for every k whenever j is picked for protection. Furthermore, if j is not picked for protection, exactly one member of the variable-set $\{N_j^k\}_{k=1...n}$ must be set to 1. Constraints (5) avoid the existence of a path between different connected components. In Constraints (6)-(7), the variable $x_{i,j}^k$ is set to 0 when at least one of i or j are protected. Constraints (8) respects the definition of the binary variable $x_{i,j}^k$. Constraints (9)-(12) represent Kirchhoff equations, that ensure connectivity in each component. The binary variables $y_{i,j}^{u,v}$ represent the u-v flow that is carried in the link (i,j). Constraint (9) avoids two-way flows. Constraints (10)-(12) model this flow.

This is an IQP formulation or more general, a mixed integer quadratic problem (MIQP). It is well known that it is NP-hard. However, it is important to remark that, differently from MILP or ILP, the source of complexity of IQP is not restricted to the integrality requirement on variables [?].

6   Bounds for the GFP

A lower bound is found by a natural relaxation of the problem, where the variables n_k, x_{ij}^k, N_i^k and y_{ij}^{uv} assume real values. Although this problem is also MIQP, only U_i

variables remain binary. In order to find an upper-bound, the objective function is modified, and as a result we obtain an integer linear program. Observe that all the constraints are linear. Since we preserve all constraints, a feasible solution for the GFP is produced. The new objective function is to minimize the size of the largest component.

The upper bound of GFP is an ILP, also NP-hard. It is modeled as follows:

$$\min Z \text{ with:}$$
$$s.t. \, n_k \leq Z, \, \forall k$$
$$Eq.(2) - (12)$$

7 Proof of Concept

This section presents the exact analysis that is product of our mathematical programming model under selected real-life networks. The model was implemented in CPLEX 12.6.3.0, MIP solver, and the executions were performed on an eight-core Intel i7 processor at 3.07 GHz, 16GB RAM. As a proof-of-concept, four graphs coming from real-life applications were considered:

- The electrical optical network EON considered by Gouveia et. al [7]. See Figure 1(b).
- The National Science Foundation Network form the USA, also considered in the previous study [7]. See Figure 1(c).
- The Uruguayan Academic Network, RAU2, depicted in Figure 1(d).
- ARPANET (Advanced Research Projects Agency Network), depicted in Figure 1(e).

We also considered a toy example in which the analysis is straight (see graph N_1 from Figure 1(a)). Table 1 summarizes the main characteristics of the graphs considered in the experimental analysis. Columns *LB*, *UB* and *Opt* stand for lower-bound, upper-bound and optimal value, respectively. The optimal value was calculated solving the exact model developed in Section 5.

Table 1. Results

graph	$\|V\|$	$\|E\|$	$\|B\|$	$\|V\| - B$	*LB*	*UB*	*Opt*
N1	9	8	1	8	1.03	4.00	3.50
RAU2	10	17	2	8	1.06	2.75	2.75
NFSNET	14	52	5	9	1.13	4.56	4.56
EON	19	36	6	13	1.06	3.46	3.00
ARPANET	20	25	5	15	1.04	2.87	2.60

The gap between the upper bound (UB) and the optimal value in the GFP (opt) is small under all instances. This highlights the fact that the size of the connected components should be as even as possible, in a strict correspondence with Theorem 4.
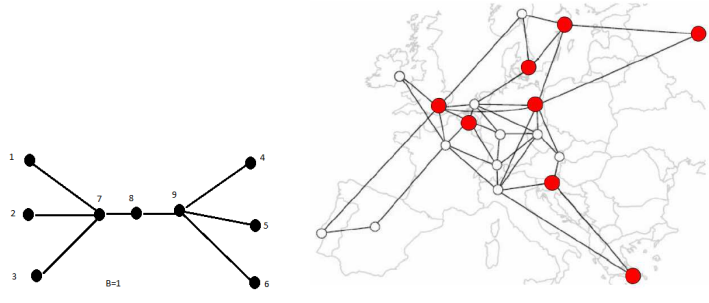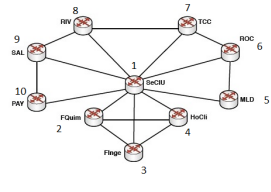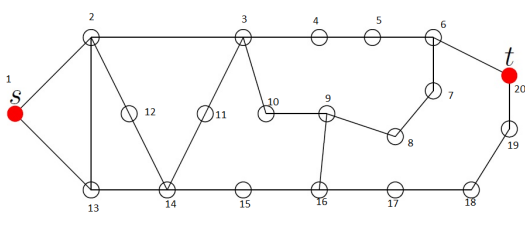
(a) *N1* graph

(b) *EON* graph
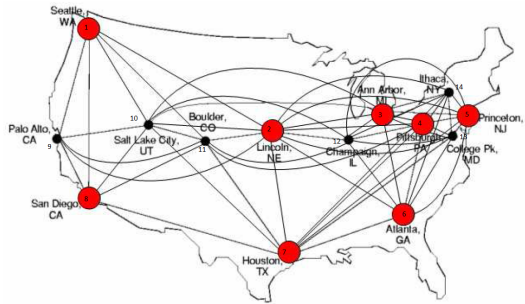
(c) *NFSNET*

(d) *RAU2*

(e) *ARPANET*

Fig. 1. Test graphs

45

Curiously enough, if we consider ARPANET with budget 6 instead, the optimal solution could not be found in a reasonable time (less than 48 hours). However, the bounds are efficiently found in that case, where either some variables assume real values or a the objective is replaced by a linear one.

Note that a trivial lower bound for GFP is 1. In effect, when we protect and remove B nodes in any graph, there are $|V| - B$ remaining nodes and in the best case, these result all disconnected. Then, there would be $|V| - B$ connected components with size 1 and the value of objective function for lower bound is 1. As shown in the Table 1, all values found for proposed lower bound, are very near to the trivial lower bound. It would be desirable to improve these values in future work or to research if this gap has some theoretical basis.

The complexity of the quadratic objective function promotes further research in the analysis of exact and approach algorithms for the GFP.

8   Conclusions and Trends for Future Work

We strongly believe that operational researchers should be engaged with the society in providing means to cope with risk analysis and natural disasters. A purely combinatorial problem is studied in this paper, called Graph Fragmentation Problem or GFP. The GFP belongs to the class of $\mathcal{NP}$-Hard problems, and there is no hope to find efficient algorithms to solve it optimally, unless $\mathcal{P} = \mathcal{NP}$. However, it is possible to solve cases where the population is configured with no cycles or elementary cycles.

A new mathematical programming formulation for the GFP is introduced in this paper, together with bounds. Exact resolutions in CPLEX confirm the fact that optimal solutions can be obtained for the GFP under small populations. Futhermore, there computational efficiency of an integer linear programming relaxation is notorious, and provides feasible solutions with small gaps for the optimal GFP.

Further research includes the development of heuristic methods, extended models, and the interplay between related relaxations. Observe that in the GFP it is assumed that a singleton is picked uniformly at random. This selection law could be modified, and the attacker could select relevant individuals from the system first. As future work, we would like to understand this generalization of the GFP with weighted nodes and adaptive protection schemes. Game theory provides a means to find optimal answers to different attacking systems.

9   Acknowledgements

References

1. Manuel Aprile, Natalia Castro, Franco Robledo, and Pablo Gabriel Romero. Analysis of Node-Resilience strategies under natural disasters. In *International Conference on Design of*

Reliable Communication Networks 2017 (DRCN 2017), pages 93–100, Munich, Germany, March 2017.

2. Frank Ball and David Sirl. Acquaintance vaccination in an epidemic on a random graph with specified degree distribution. *J. Appl. Probab.*, 50(4):1147–1168, 12 2013.

3. E. Dahlhaus, D. S. Johnson, C. H. Papadimitriou, P. D. Seymour, and M. Yannakakis. The complexity of multiway cuts (extended abstract). In *Proceedings of the Twenty-fourth Annual ACM Symposium on Theory of Computing*, STOC '92, pages 241–251, New York, NY, USA, 1992. ACM.

4. T. N. Dinh, Y. Xuan, M. T. Thai, P. M. Pardalos, and T. Znati. On new approaches of assessing network vulnerability: Hardness and approximation. *IEEE/ACM Transactions on Networking*, 20(2):609–619, April 2012.

5. L. R. Ford and D. R. Fulkerson. Maximal Flow through a Network. *Canadian Journal of Mathematics*, 8:399–404.

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

7. Luis Gouveia, Luidi Simonetti, and Eduardo Uchoa. Modeling hop-constrained and diameter-constrained minimum spanning tree problems as steiner tree problems over layered graphs. *Mathematical Programming*, 128(1):123–148, 2011.

8. R. M. Karp. Reducibility among combinatorial problems. In R. E. Miller and J. W. Thatcher, editors, *Complexity of Computer Computations*, pages 85–103. Plenum Press, 1972.

9. M. E. J. Newman. The structure and function of complex networks. *SIAM Review*, 45(2):167–256, 2003.

10. John J. Pauly. The great chicago fire as a national event. *American Quarterly*, 36(5):668–683, 1984.

11. J. Piccini, F. Robledo, and P. Romero. Analysis and complexity of pandemics. In *2016 8th International Workshop on Resilient Networks Design and Modeling (RNDM)*, pages 224–230, Sept 2016.

12. Juan Piccini, Franco Robledo, and Pablo Romero. *Node-Immunization Strategies in a Stochastic Epidemic Model*, pages 222–232. Springer International Publishing, Cham, 2015.

13. Juan Piccini, Franco Robledo, and Pablo Romero. Graph fragmentation problem. In *Proceedings of 5th the International Conference on Operations Research and Enterprise Systems*, pages 137–144, 2016.

14. J. K. Taubenberger, A. H. Reid, R. M. Lourens, R. Wang, G. Jin, and T. G. Fanning. Molecular virology: Was the 1918 pandemic caused by a bird flu? Was the 1918 flu avian in origin? (Reply). *Nature*, 440:9–10, April 2006.

Chapter 5

Graph Fragmentation Problem: Analysis and Synthesis

Here the GFP is addressed by several approaches: metaheuristics, approximation algorithms, polytime methods for specic instances, and exact methods for small instances. Also an analysis of GFP computational complexity is included.

Graph Fragmentation Problem: Analysis and Synthesis

Manuel Aprile[a,*], Natalia Castro[b], Graciela Ferreira[b], Juan Piccini[b], Franco Robledo[b] and
Pablo Romero[b]

[a]*DISOPT, École Polytechnique Fédérale de Lausanne. Station 8 CH-1015, Lausanne, Switzerland*
[b]*Laboratorio de Probabilidad y Estadística. Facultad de Ingeniería, Universidad de la República. Julio Herrera y Reissig 565, PC 11300. Montevideo, Uruguay*
E-mail: manuel.aprile@epfl.ch; (ncastro,gferreira,piccini,frobledo,promero)@fing.edu.uy

Abstract

Vulnerability metrics play a key role in the understanding of cascading failures and target/random attacks to a network. The Graph Fragmentation Problem (GFP) is the result of a worst case analysis of a random attack. We can choose a fixed number of individuals for protection, and a non-protected target node immediately destroys all reachable nodes. The goal is to minimize the expected number of destroyed nodes in the network.

In this paper, we address the GFP by several approaches: metaheuristics, approximation algorithms, polytime methods for specific instances, and exact methods for small instances. The computational complexity of the GFP is included in our analysis, where we formally prove that the corresponding decision version of the problem is $\mathcal{NP}$-Complete. Furthermore, a strong inapproximability result holds: there is no polynomial approximation algorithm with factor lower than $5/3$, unless $\mathcal{P} = \mathcal{NP}$. This promotes the study of specific instances of the problem for tractability and/or exact methods in exponential time. As a synthesis, we propose new vulnerability/connectivity metrics and an interplay with Game Theory, using a closely related combinatorial problem called Component Order Connectivity.

Keywords: Vulnerability metrics; Graph Fragmentation Problem; Computational Complexity; Approximation Algorithms; Metaheuristics; Game Theory.

1. Introduction

Modern connectivity theory is largely pressed by social networks analysis, disaster management and business models in telecomunications. In the classical literature, we find minimum-cost k-node-connected spanning topologies as a reference network design model (Monma et al., 1990; Stoer, 1993). This is a notion of survivability under a fixed number of component failures of a system. More

*Pablo Romero (promero@fing.edu.uy).

recently, the focus moved towards disaster management, centrality and vulnerability metrics under random/targeted attacks (Mauthe et al., 2016; Thai and Pardalos, 2011; Gouveia and Leitner, 2017).

Simulation tools were developed in order to capture a large framework of cascading failures in epidemic modelling (Marzo et al., 2017). However, under cascading failures, the system is more robust when the individuals are poorly communicated, in a strong contrast with modern connectivity theory. To the best of our knowledge, there is no simulation tool available for both apparently antipodal scenarios.

The Graph Fragmentation Problem (GFP) is the product of a worst case analysis of a random attack under cascading failures, therefore, it is suitable for pandemic analysis. However, in its min-max version we recover a previous problem called Component Order Connectivity (COC). The corresponding max-min version for COC is a suitable connectivity metric.

The goal of this paper is to present a comprehensive analysis for the GFP, its relation with COC and new feasible vulnerability/connectivity metrics as a synthesis. Both GFP and COC are formally presented in Section 2. Section 3 contains a comprehensive analysis for the GFP. This section covers several approaches for the problem in different subsections, such as Complexity (Subsection 3.1), Approximation Algorithms (3.2), Polytime methods for special graphs (3.3), Exact Analysis (3.4) and Metaheuristics (3.5). Each subsection is enriched with references for further reading. In Section 4 we discuss vulnerability/connectivity metrics suggested by GFP and COC, and a potential interplay with Game Theory. Finally, Section 5 summarizes the conclusions and trends for future work.

2. Graph Fragmentation Problem

This combinatorial optimization problem was introduced by Piccini et al. (2015), originally inspired by epidemic modelling. We are given a population represented by a graph $G = (V, E)$, and a budget constraint B, which is a natural number B such that $0 \leq B \leq |V|$. We can protect (i.e., delete) a node-set $U \subseteq V$ with cardinality $|U| = B$, obtaining a subgraph $G' = G - U$. The nature picks a node v uniformly at random from G'. The disaster kills all the members that belongs to the connected component of v. The goal is to minimize the expected number of deaths.

Mathematically, if the subgraph G' has $|V'| = n$ nodes and k connected components $V_1, \ldots, V_k$ with respective orders $n_1 \geq n_2 \geq \ldots \geq n_k$, the probability to choose component i is n_i/n. Therefore, the expected number of deaths is $E(G') = \sum_{i=1}^{k} n_i p_i$, with $p_i = n_i/n$. The goal of the Graph Fragmentation Problem (GFP) is to protect B nodes in order to minimize the expected number of deaths:

$$\min_{U \subseteq V} E(G - U) = \sum_{i=1}^{k} \frac{n_i^2}{n}$$
$$s.t. |U| = B.$$

Observe that the cost-function in the GFP is precisely $E(G')$, which represents the expected number of deaths. A smart node-protection technique is desirable, since in practice the number of vaccines B (or budget) is much lower than the population $n = |V|$.

The Component Order Connectivity problem (COC) is identical to the GFP, but the goal is to minimize the size of the maximum connected component (the objective function is replaced by $n_1 = |V_1|$).

3. Analysis

In this section we address the GFP using different approaches. The hardness of the corresponding decision version is formally proved in Subsection 3.1. As Corollary, the GFP belongs to the class of $\mathcal{NP}$-Hard $\mathcal{NP}$-Optimization problem. The treatment is kept as simple as possible.

Subsection 3.2 presents a negative result of inapproximability. In fact, there is no feasible approximation algorithm with factor $\alpha < 5/3$, unless $\mathcal{P} = \mathcal{NP}$. This promotes the definition of tractable sub-problems (Subsection 3.3), finding exact solutions for small-sized instances (Subsection 3.4) and the development of metaheuristics (Subsection 3.5).

3.1. Complexity

The hardness of the GFP has theoretical value. Indeed, in Piccini et al. (2016) it is proved that a large set of Node-Immunization Problems are at least as hard as the GFP. Therefore, this proves that immunization is a hard task, and the intuition from epidemiologist is correct. Here, we prove the hardness of the GFP in a more simple way than in Piccini et al. (2016).

The following problem will be used to characterize the computational complexity of the GFP.

Definition 1 (Minimum Cardinality Vertex Cover).
Instance: simple graph $G = (V, E)$ and positive integer k.
Does there exist a node-set U such that $|U| \leq k$ and every link is incident to some node from U?

Recall that Minimum Cardinality Vertex Cover belongs to Karp list of 21 $\mathcal{NP}$-Complete decision problems (Karp, 1972). Observe that if a vertex cover U is found, there is only one dead in $G' = G - U$ for the GFP. This is clearly a globally optimum solution. The optimality for the GFP is strictly related to the determination of a vertex cover.

Theorem 1. *The GFP belongs to the class of $\mathcal{NP}$-Hard problems.*

Proof. The graph $G' = G - U$ has isolated nodes if and only if U is a vertex cover, where $|U| \leq B$. Thus, the GFP is at least as hard as Minimum Cardinality Vertex Cover. $\qquad\square$

3.2. Approximation Algorithm

The following problem will be considered in order to prove a stronger inapproximability result:

Definition 2 (Multiway k-cut).
Instance: simple graph $G = (V, E)$, terminal set $K \subseteq V$ with $|K| = k$, positive integer B.
Does there exist a separator set $U \subseteq V - K$ with $|U| \leq B$ such that each terminal node belongs to different components in $G - U$?

We know that Multiway 2-cut is in $\mathcal{P}$. A polynomial time algorithm is provided by Ford and Fulkerson (1956). However, Multiway k-cut is $\mathcal{NP}$-Complete for every fixed $k \geq 3$ (Dahlhaus et al., 1992). If the answer of Multiway k-cut is affirmative for a given instance (G, K, B), such a separator set $U \subseteq V - K$ with $|U| \leq B$ is called a k-cut. Otherwise, there is no k-cut for instance (G, K, B), and the answer

is negative. In the following, K_n denotes the complete graph or clique with n nodes, where all pair of different nodes are adjacent. Additionally, C_n represents the elementary cycle with n nodes.

Theorem 2. *It is $\mathcal{NP}$-hard to approximate GFP within any factor $\alpha < \frac{5}{3}$.*

Proof. A sketch of the proof is here presented (the interested reader is invited to see the Appendix for further details). Consider an instance of Multiway 3-cut with ground graph $G = (V, E)$, distinguished nodes $K = \{v_1, v_2, v_3\}$ and positive integer B. Replace those nodes by large cliques $\{K_N, K_N, K_N\}$, where $N >> |V|$. The order of the new graph G^* is roughly $3N$. If the instance admits a 3-cut, i.e. a subset whose removal separates v_1, v_2, v_3, then the optimal cost in the GFP with instance (G^*, B) is roughly N. Otherwise, the expected number of dead nodes is never lower than $\frac{(2N)^2 + N^2}{3N} = \frac{5N}{3}$. Therefore, an approximation algorithm with factor strictly less than $5/3$ would determine whether (G, K) admits a 3-cut using B nodes or not, thus solving Multiway 3-cut. $\square$

The interested reader can find the minimum size $N = N(\epsilon, n)$ explicitly in terms of n and ϵ in the Appendix, for any $\epsilon > 0$ and graph-size n.

3.3. Polytime Methods

The main result of the following paragraphs is that the GFP is solvable in polynomial time for all acyclic graphs. The optimal solution for elementary cycles and specific bipartite graphs are provided as well.

First, we informally outline a decomposition result when the input graph has multiple components:

Lemma 1. *If G_1 and G_2 accept polytime algorithms for GFP, also $G = G_1 \cup G_2$ accepts a polytime algorithm.*

Proof. Consider two polytime algorithms $A_1(G_1, B_1)$ and $A_2(G_2, B_2)$, and all natural partitions $B = B_1 + B_2$. Algorithm $A(G, B)$ applies $A_1(G_1, B_1)$ and $A_2(G_2, B - B_1)$ for all feasible numbers $B_1 \leq B$. The minimum objective value is returned. $\square$

Lemma 1 is the key concept to prove the main result of this subsection.

Theorem 3. *The optimal solution for the GFP can be found in polytime for all acyclic graphs.*

Proof. By Lemma 1, it suffices to study a single connected component (i.e., a tree). Choose an arbitrary root-node for the tree. The main idea is to climb from bottom-to-top using a dynamic programming approach. Essentially, we can apply Algorithm A from Lemma 1 to disjoint branches and climb up to the root.

Consider a set of branches $T^1, \ldots, T^m$ with constant size such that they are either disjoint or they only share the root-node. For each T_i, a brute-force algorithm finds the optimal solution in constant time for all possible budget values $B_1, \ldots, B_m$ with $\sum B_i = B$. Iteratively apply Algorithm A from Lemma 1 in order to find the optimum protection scheme in the union $\cup T^i$. The addition of disjoint branches or parent nodes is produced recursively, until the optimum for the whole tree is met. $\square$

The reader is invited to consult Aprile et al. (2017) for technical details of the proof. The result in Theorem 3 can be extended to a larger class of graphs, defined by the notion of treewidth. Given a graph

$G(V, E)$, a tree decomposition of G is a tree whose nodes correspond to subsets of V called bags, and such that the following conditions hold: for each edge uv of G, there is a bag containing both u and v; and for any $v \in V$, the subgraph induced by bags containing v is connected. The treewidth of G is defined as the size of the largest bag in a tree decomposition of G. It is easy to see that trees have treewidth equal to 1 (choose the tree decomposition with a bag for each edge of the tree). Graphs whose treewidth is small, i.e. bounded by a constant, have a "tree-like" structure. For instance, series-parallel graphs and outerplanar graphs have treewidth at most 2. This parameter is widely studied; see Bodlaender and Koster (2007) for a survey. In particular, when a combinatorial optimization problem is polynomially solvable on trees by dynamic programming, this usually extends to all graphs of bounded treewidth. This applies to the GFP as well.

Theorem 4. *The optimal solution for the GFP can be found in polytime for all graphs of bounded treewidth.*

Proof. We only give a sketch of proof. We consider a tree decomposition of the graph G, and we run a dynamic program on the tree such as in Theorem 3 in a bottom-up fashion. An optimal solution for the subgraph of G corresponding to the current branch as in Aprile et al. (2017) for trees. The running time of the algorithm depends polynomially on the size of G and exponentially on the treewidth, hence it is polynomial as long as G has constant treewidth. □

We conclude the section by finding the optimum solution for elementary cycles. First, a basic property of the GFP will be useful. For any set of vertices U, let $n_{G'} = (n_1, \ldots, n_k)$ the vector of the respective orders of the connected components from $G' = G - U$, such that $n_1 \geq n_2 \ldots \geq n_k$. Further, for any real $p \geq 1$, denote $\|n\|_p = (\sum_i |n_i|^p)^{1/p}$ the p-norm. The canonic basis of the vector space $\mathbb{R}^k$ is denoted by $\{e_1, \ldots, e_k\}$. The GFP can be re-written as an euclidean norm minimization:

$$\min_{U \subseteq V} \|n_{G'}\|_2^2$$
$$s.t. \|n_{G'}\|_1 = n$$

Proposition 1. For any graph $G(V, E)$, let k be the number of components of of G' in an optimal solution of GFP. If $U \subset V$ yields the vector $n_{G'} = (n_1, \ldots, n_k)$, then the condition $n_1 - n_k \leq 1$ is a sufficient criterion of optimality for the GFP.

Proof. If $n_1 \geq n_k + 2$, then the vector $u = n - e_1 + e_k$ has lower Euclidean norm. Else, if $n_1 - n_k \leq 1$ there is no feasible improvement, and a global optimum is met. □

Corollary 1. If $G = P_n$, the optimum solution for GFP has r paths P_{d+1} and $B + 1 - r$ paths P_d, where the integers d and r are found using Euclidean division rule: $n - B = d(B + 1) + r$.

Proof. The sufficient criterion from Proposition 1 holds. □

Corollary 2. The optimum for (C_n, B) is the optimum for $(P_{n-1}, B - 1)$.

Proof. Remove an arbitrary node and proceed as in Corollary 1. □

3.4. Exact Analysis

Here we simplify the mathematical programming model for the GFP introduced in Castro et al. (2017). In particular, we reduce the number of binary variables and redundant constraints.
Consider $G' = G - U$, n_k the size of component k and binary variables:

- $U_i \in \{0,1\}$, $i \in V$: node $i \in U$ (or not);
- $x_{ij}^k \in \{0,1\}$, $(i,j) \in E$: link (i,j) belongs to component k in G;
- $N_i^k \in \{0,1\}$, $i \in V$: node i belongs to the component k;

The mathematical programming model for the GFP is the following:

$$\text{min.} \sum_{i=1..K} n_i^2/n \tag{1}$$

$$\text{s.t.} \sum_{j \in V} U_j \leq B, \tag{2}$$

$$\sum_{j \in V} N_j^k = n_k, \forall k = 1 \ldots K \tag{3}$$

$$\sum_{k=1...K} N_j^k = 1 - U_j, \forall j \in V \tag{4}$$

$$N_i^k + N_j^s \leq 1, \forall i,j \in V, i \neq j, (i,j) \in E, \forall k,s \in K, s \neq k, \tag{5}$$

$$\sum_{k=1...K} x_{ij}^k \leq (1 - U_i), \forall (i,j) \in E, i,j \in V \tag{6}$$

$$\sum_{k=1...K} x_{ij}^k \leq (1 - U_j), \forall (i,j) \in E, i,j \in V \tag{7}$$

$$N_i^k + N_j^k \leq 1 + x_{ij}^k, \forall k \in 1 \ldots K, \forall (i,j) \in E, i,j \in V \tag{8}$$

The objective function captures the cost of the GFP (1). Inequality (2) represents the budget constraint. The size of each connected component is found using Constraint (3). Constraints (4) set $N_j^k = 0$ for every k whenever j is picked for protection. Furthermore, if j is not picked for protection, exactly one member of the variable-set $\{N_j^k\}_{k=1...n}$ must be set to 1. Constraints (5) avoid the existence of a path between different connected components. In Constraints (6)-(7), the variable $x_{i,j}^k$ is set to 0 when at least one of i or j are protected. Constraints (8) respect the definition of the binary variable $x_{i,j}^k$.

3.5. Metaheuristics

Naive heuristics for the problem where presented for the first time in Piccini et al. (2015). This is the first evidence that an iterative node-protection picking nodes with maximum degree is sub-optimal. A pure Greedy notion for the problem is presented in Piccini et al. (2016), where the protected nodes are iteratively picked minimizing the objective function as large as possible in a step-by-step fashion.

GRASP is a powerful multi-start metaheuristic, which consists in two phases: a Construction phase and Local Search phase (Resende and Ribeiro, 2016). A full GRASP enriched with a Path-Relinking post-optimization stage has been recently introduced by Piccini et al. (2017). The GRASP iteratively builds the protected-set. In each iteration, it finds the worst/best node for protection, and trades greediness for randomization, choosing a random node from a Restricted Candidate List of the top-ranked nodes (indeed, a fixed percentage $\alpha \in (0, 1)$ of the whole list). After the classical multi-start process of a traditional GRASP implementation, a pool of elite solutions are selected, and a new exploration phase takes place using path-relinking. The performance of the GRASP methodology is studied using optimal solutions of small instances (found by an exhaustive search of the feasible-set). The results confirm that this GRASP heuristic achieves optimal results for small instances. The reader is invited to consult Piccini et al. (2017) for further details.

3.6. Proof-of-Concept

We perform a fair comparison between the optimal solution provided by the IQP model and our GRASP methodology. As a proof-of-concept, we considered 9 graphs. These graphs include real-life networks such as the US National Science Foundation Network, the electrical optical network considered by Gouveia et al. (2011), the Uruguayan Academic Network depicted in Figure 1(a) and ARPANET. The IQP formulation model was implemented in CPLEX 12.6.3.0, MIP solver, and the executions were performed on an Intel(R) Core(TM) i7 CPU 3.33GHz, 16GB RAM. Graphs are sketched in Figure 1. Table 1 summarizes the main characteristics of the graphs considered in the experimental analysis.

Table 1: Results

Case	Graph	$|V|$	$|E|$	B	$|V| - B$	$\sharp bin$	$\sharp int$	$\sharp const$	COC	$\sharp f.soln$	UB	$Opt\,GFP$	$opt\,known$	$GRASP$	UB (sec)	GFP (sec)	$GRASP$ (sec)
1	Rau2	10	17	2	8	290	11	435	3	23	2.750	2.750	2.750	2.750	2.03	5	0.08
2	NSfNET	14	52	5	9	952	15	1645	5	25	4.556	4.556	4.556	9.000	12.93	1341	0.16
3	Eon	19	36	6	13	1083	20	1555	4	20	3.154	3.000	3.000	8.077	8.28	22689	0.22
4	Arpanet	20	25	6	14	940	21	1171	3	25	2.000	2.000*	2.000	2.857	14.02	7200*	0.24
5	Dodecaedro	20	30	6	14	1040	21	1381	7	26	6.429	6.429	6.429	12.143	18.94	30816	0.24
6	LATA5S	40	72	6	34	4560	41	6145	12	32	9.706	9.706*	9.353	24.882	554.14	18000*	0.41
7	es10fst01	18	20	5	13	720	19	869	3	21	2.077	2.077	2.077	2.385	27.14	630298	0.19
8	es10fst09	21	29	6	15	1092	22	1403	4	23	2.867	2.867*	2.867	3.000	20.93	7200*	0.23
9	se03	13	21	4	9	468	14	667	5	26	4.556	4.556	4.556	9.000	9.61	295	0.13

The respective columns $|V|$, $|E|$ and B from Table 1 summarize the order, size and budget of the input; $|V| - B$ the order of the remaining graph G', $\sharp bin$ the number of binary variables, $\sharp int$ the number of integer variables, and $\sharp const$ the number of constraints. The column COC is the optimal value for COC problem and $\sharp f.soln$ is the number of simultaneously feasible solutions for both COC and GFP. Column UB is the optimal value for COC and $Opt\,GFP$ the optimal value for GFP by means of the IQP formulation. Column $opt\,known$ is the known optimal value by an enumeration of all feasible solutions, and $GRASP$ the output of our GRASP methodology. The last three columns show the CPU time for UB, GFP and $GRASP$, measured in seconds.

Here GRASP is executed without the path-relinking post-optimization stage. It is worth to remark that

(a) $RAU2$ (b) $NFSNET$ (c) EON graph   (d) $ARPANET$

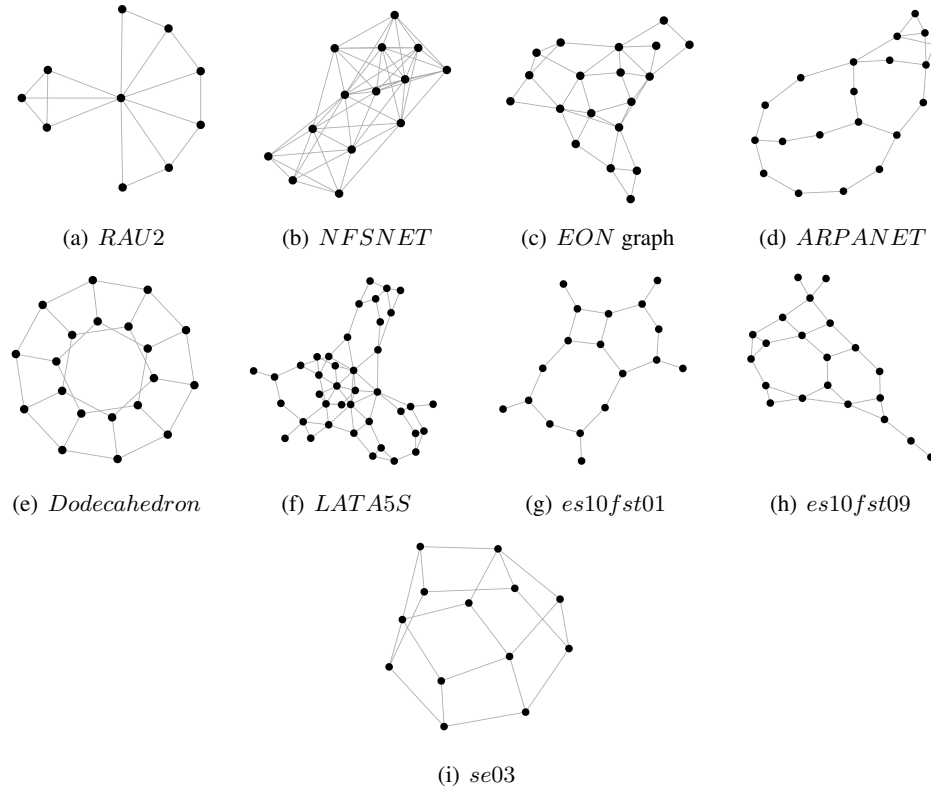(e) $Dodecahedron$ (f) $LATA5S$ (g) $es10fst01$ (h) $es10fst09$

(i) $se03$

Fig. 1: Test graphs

the full GRASP presented in Piccini et al. (2017) has reached optimality in practically all instances.

The gap between COC (UB) and the optimal value for the GFP ($Opt\ GFP$) is small under all instances. This reinforces the fact that optimal solutions for the GFP must be balanced, according to Proposition 1. The timeout has been elapsed for $Arpanet$, $LATA5S$ and $es10fst09$. They are marked with an "*". The upper bound differs from the optimal value only under $LATA5S$.

The complexity of the quadratic objective function promotes further research in the analysis of algorithms for the GFP. Close upper-bounds where found using linear programming (i.e., COC problem), in few seconds.

4. Synthesis

4.1. Vulnerability and Connectivity Metrics

The COC provides a metric of survivability under intentionally malicious attacks, since the attacker picks a node from the largest connected component. On the other hand, the GFP is a metric of survivability under random attacks. Clearly, the effect of intentionally malicious attacks is more adverse, and this explains the fact that the optimal value for COC is an upper-bound for the GFP.

We can find an alternative reading of both problems in terms of connectivity. Let us consider a simple graph G with node-connectivity $\kappa(G) > B$. In this case, the network is *robust under B node failures*. Therefore, strong connectivity implies survivability. If $\kappa(G) < B$, there are many ways to disconnect the graph, and both problems slightly diverge (observe that a balance in the size of the components is desirable for COC as well). The case $\kappa(G) = B$ is critical, and both problems look for node-separator sets with cardinality B.

4.2. Discussion: GFP and Game Theory

Since COC is a min-max problem, we recall the celebrated Minimax Theorem on zero-sum games Nash (1951). This theorem is a consequence of linear programming theory as well Dantzig (1963).

We would like to find an interplay between attacker/defender roles and zero-sum games, in order to find a Nash equilibrium. If COC is a zero-sum game, the optimal value for the corresponding max-min version must be identical to the optimal value of the COC. However, this is not the case for COC.

For example, if we are given an elementary cycle with 6 nodes $G = C_6$ and budget $B = 2$, in COC we protect two nodes that achieve the diameter and the optimum value is 2. In the corresponding max-min problem we should protect two adjacent nodes, and the score is precisely, 4.

In fact, if we stick to the terminology with component sizes $n_1 \geq n_2 \ldots \geq n_k$ of the resulting graph $G' = G - U$, the following inequality holds in general:

Proposition 2. $\min_{|U|=B} n_1 \leq \max_{|U|=B} n_k$.

Indeed, the max-min version is trivial for connected graphs:

Proposition 3. If G is connected, the score for max-min is $n = |G| - B$.

Proof. If G is connected, consider a spanning tree $T \subseteq G$. Since T is a tree, it has some leaf-node, $v_1 \in T$. The resulting subgraph $T - v_1$ is another tree. If we iteratively remove B leaf-nodes, we reach a node-set U such that $T - U$ is connected. Therefore, $G' = G - U$ is also connected (since it has possibly more links), and the optimal value for the max-min problem is precisely $n = |G - U| = |G| - B$. $\square$

Proposition 4.2 discards a potential interplay between COC and zero-sum games. Currently, we consider an extension for the GFP with roles of atacker and defender, called Stochastic Weighted GFP (SWGFP), where:

- And attacker picks the singleton $v \in V$ following some probability law $p(v)$ in the nodes, and
- Non-negative real weights $w(v)$ are assigned to the nodes of the system.

Nodes are assigned weights related with their importance in the network, and the attacker follows some specific probability law, which is not necessarily uniform. The goal is to choose some defense strategy (or node-protection scheme) $U \subseteq V$ with $|U| = B$ such that the expected loss $E(G - U) = \sum_{i=1}^{k} W_k p_k$ is minimum, being W_i the weight of component V_i and $p_i = P(v \in V_i)$. Greedy notions for random and targeted attacks were also found. Nevertheless, a bridge between Game Theory and the GFP is still an open problem.

5. Conclusions and Future Work

The Graph Fragmentation Problem has been analyzed using several approaches. It is not only hard in a strong sense, but also there is no approximation algorithm with factor $\alpha < 5/3$, unless $\mathcal{P} = \mathcal{NP}$. However, the optimality in graphs with bounded treewidth is guaranteed.

Powerful heuristics for the problem are already available. In fact, a GRASP methodology enriched with a path-relinking post-optimization stage shows close-to-optimality results under instances with hundreds of nodes. This is confirmed finding the optimal values with a brute-force approach, and an IQP formulation. As a synthesis, we show the interplay between the GFP and vulnerability/connectivity metrics. Additionally, potential extensions for the GFP are discussed, including the Stochastic Weighted GFP (SWGFP).

As future work, we would like to develop competitive heuristics for the SWGFP, and have a comprehensive understanding of node-protection schemes on the lights of Game Theory.

6. Appendix

Theorem 5. *It is $\mathcal{NP}$-hard to approximate GFP within $\frac{5}{3} - \epsilon$, for any $\epsilon > 0$.*

Proof. A reduction from Multiway 3-cut is provided (Aprile et al., 2017).

Consider an instance G, $K = \{t_1, t_2, t_3\}$ and B for Multiway 3-cut, and assume there is an approximation algorithm $\mathcal{A}$ with factor $5/3 - \epsilon$ for GFP, for some $\epsilon > 0$. Let $n = |V(G)|$. We produce a corresponding instance for the GFP by means of the polynomial reduction $(G, K, B) \to (G^{(K)}, B)$, where the new graph $G^{(K)}$ is identical to G, but the terminal nodes t_1, t_2, t_3 are replaced by cliques K_1, K_2, K_3 with $M = cn$ nodes each (for a constant c to be specified later) and each node in a clique is connected in the same way as the corresponding terminal in G. The intuition behind the reduction is that the large cliques corresponding to the terminals are the "heaviest" parts of the graph and to minimize the objective function of GFP one must ensure that the cliques end up in different components. We now run the algorithm $\mathcal{A}$ on input $(G^{(K)}, B)$. Let S be S the output of the algorithm, $\mathcal{A}_{G^{(K)}}$ its objective value and $\tilde{n} = n + 3M - 3 - B$ the number of nodes in $G^{(K)} - S$. There are two cases:

1. If $\mathcal{A}_{G^{(K)}} < \frac{5M^2}{\tilde{n}}$ the we claim that the answer to Multiway 3-cut is YES and the certificate can be obtained in polynomial time from S.
2. Otherwise, we claim the answer to Multiway 3-cut is NO.

This implies that we can solve Multiway 3-cut through the algorithm $\mathcal{A}$, hence we will be done once we prove the two claims.

1. First, suppose that

$$S \subset V(G^{(K)}) \setminus \{K_1, K_2, K_3\} = V(G) \setminus \{t_1, t_2, t_3\}.$$

Assume by contradiction that removing S from G does not disconnect all the terminals, i.e. there is a component of $G \setminus S$ that contains at least two of them. If the component contains all three cliques, the objective value $\mathcal{A}_{G^{(K)}}$ is at least $9M^2/\tilde{n}$, a contradiction. Otherwise, the component in $G^{(K)} \setminus S$ has size at least $2M$, and there is another component containing the third clique of size at least M, hence

$$\mathcal{A}_{G^{(K)}} \geq \frac{(2M)^2 + M^2}{\tilde{n}} = \frac{5M^2}{\tilde{n}},$$

again a contradiction. Now, consider the case in which some of the vertices of S belong to some clique K_i, we will show that we can obtain a YES-certificate from $S \cap V(G)$. Assume that removing S from $G^{(K)}$ leaves a component C such that $C \cap K_i \neq \emptyset$ for $i = 1, 2, 3$, but then $|C| \geq 3M - B \geq 3M - n$ and $\mathcal{A}_{G^{(K)}} \geq 5M^2/\tilde{n}$ if we choose for instance $c \geq 2$, hence we get a contradiction. So the only possibility, without loss of generality, is that removing S from $G^{(K)}$ leaves a component C such that $C \cap K_i \neq \emptyset$ for $i = 1, 2$, and $C \cap K_3 = \emptyset$. Then, if $|S \cap K_i| = B_i$ for $i = 1, 2, 3$ and $t = |C \cap V(G)|$, we have $|C| = 2M - B_1 - B_2 + t$, and there is another component containing vertices of K_3 of size at least $M - B_3$. Now, we must have $t \leq B_1 + B_2 + B_3$, otherwise it is immediate to verify that

$$\mathcal{A}_{G^{(K)}} \geq \frac{(2M - B_1 - B_2 + t)^2 + (M - B_3)^2}{\tilde{n}} \geq \frac{5M^2}{\tilde{n}},$$

a contradiction. Hence we have that $S' = S \cup (C \cap V(G)) \setminus \cup_i K_i$ is a YES-certificate for G.

2. We have

$$GFP(G^{(K)}, B) \geq \frac{1}{\frac{5}{3} - \epsilon} \mathcal{A}_{G^{(K)}} \geq \frac{1}{\frac{5}{3} - \epsilon} \frac{5M^2}{\tilde{n}}.$$

We will now choose c so that

$$\frac{1}{\frac{5}{3} - \epsilon} \frac{5M^2}{\tilde{n}} > \frac{2M^2 + (M + n - B)^2}{\tilde{n}}. \tag{9}$$

Assuming the latter is true, we have that the three terminals cannot be disconnected, hence the answer to Multiway 3-cut is NO. Indeed, the value $\frac{2M^2 + (M+n-B)^2}{\tilde{n}}$ is the maximum value that is achievable having the three cliques in different components (the value is obtained if all other nodes belongs to the same clique, and is smaller otherwise), and the optimal value is larger than that. To conclude the proof, we need to find c such that Inequality (9) holds. This is an elementary calculation that we omit. It is easy to verify that, for instance, $c = 55/\epsilon$ satisfies Inequality (9).

$$\square$$

References

Aprile, M., Castro, N., Robledo, F., Romero, P.G., 2017. Analysis of node-resilience strategies under natural disasters. In *International Conference on Design of Reliable Communication Networks 2017 (DRCN 2017)*, Munich, Germany, pp. 93–100.

Bodlaender, H.L., Koster, A.M., 2007. Combinatorial optimization on graphs of bounded treewidth. *The Computer Journal* 51, 3, 255–269.

Castro, N., Ferreira, G., Robledo, F., Romero, P., 2017. Graph fragmentation problem for natural disaster management. In *Proceedings of the Third International Conference on Machine Learning, Optimization and Big Data. Springer LNCS*, Springer International Publishing, Cham.

Dahlhaus, E., Johnson, D.S., Papadimitriou, C.H., Seymour, P.D., Yannakakis, M., 1992. The complexity of multiway cuts (extended abstract). In *Proceedings of the Twenty-fourth Annual ACM Symposium on Theory of Computing*, ACM, New York, NY, USA, pp. 241–251.

Dantzig, G., 1963. *Linear Programming and Extensions*. Rand Corporation research study. Princeton University Press.

Ford, L.R., Fulkerson, D.R., 1956. Maximal Flow through a Network. *Canadian Journal of Mathematics* 8, 399–404.

Gouveia, L., Leitner, M., 2017. Design of survivable networks with vulnerability constraints. *European Journal of Operational Research* 258, 1, 89 – 103.

Gouveia, L., Simonetti, L., Uchoa, E., 2011. Modeling hop-constrained and diameter-constrained minimum spanning tree problems as steiner tree problems over layered graphs. *Mathematical Programming* 128, 1, 123–148.

Karp, R.M., 1972. Reducibility among combinatorial problems. In Miller, R.E. and Thatcher, J.W. (eds), *Complexity of Computer Computations*. Plenum Press, pp. 85–103.

Marzo, J.L., Cosgaya, S.G., Scoglio, C.M., 2017. Network robustness simulator: a case study on epidemic models. In *9th International Workshop on Resilient Networks Design and Modeling (RNDM)*, pp. 1–6.

Mauthe, A., Hutchison, D., etinkaya, E.K., Ganchev, I., Rak, J., Sterbenz, J.P.G., Gunkelk, M., Smith, P., Gomes, T., 2016. Disaster-resilient communication networks: Principles and best practices. In *2016 8th International Workshop on Resilient Networks Design and Modeling (RNDM)*, pp. 1–10.

Monma, C., Munson, B., Pulleyblank, W., 1990. Minimum-weight two-connected spanning networks. *Mathematical Programming* 46, 1-3, 153–171.

Nash, J., 1951. Non-cooperative games. *Annals of Mathematics* 54, 2, 286–295.

Piccini, J., Robledo, F., Romero, P., 2015. Node-Immunization Strategies in a Stochastic Epidemic Model, Springer International Publishing, Cham. pp. 222–232.

Piccini, J., Robledo, F., Romero, P., 2016. Graph fragmentation problem. In *Proceedings of 5th the International Conference on Operations Research and Enterprise Systems*, pp. 137–144.

Piccini, J., Robledo, F., Romero, P., 2017. Complexity among combinatorial problems in epidemics. *International Transactions in Operational Research* 25, 1, 295–318.

Resende, M., Ribeiro, C., 2016. *Optimization by GRASP - Greedy Randomized Adaptive Search Procedures*. Springer.

Stoer, M., 1993. *Design of Survivable Networks*. Lecture Notes in Mathematics. Springer.

Thai, M.T., Pardalos, P.M., 2011. *Handbook of Optimization in Complex Networks: Communication and Social Networks*. Springer Publishing Company, Incorporated.

Part III

Reliability in Stochastic Binary Systems

Chapter 6

Building Reliability Bounds in Stochastic Binary Systems

In system reliability analysis, the goal is to nd the probability of correct operation of a system subject to component failures. Stochastic Binary System (SBS) is a mathematical model of multi-component on-off systems subject to random failures. SBS models extend classical network reliability models (where the components subject to failure are nodes or links of a graph) and are able to represent more complex interactions between the states of the individual components and the operation of the system under study. In this chapter, the concept of separable stochastic binary systems are introduced, together with their efcient representation, and reliability bounds for arbitrary SBS, exploiting duality and Chernoff inequality, are calculated.

Building Reliability Bounds
in Stochastic Binary Systems

Héctor Cancela, Graciela Ferreira, Gustavo Guerberoff, Franco Robledo and Pablo Romero

Facultad de Ingeniería

Universidad de la República

Julio Herrera y Reissig 565, PC 11300

Montevideo, Uruguay

Emails: (cancela,gferreira,gguerber,frobledo,promero)@fing.edu.uy

Abstract—**A Stochastic Binary System (SBS) is a mathematical model of multi-component on-off systems subject to random failures. SBS models extend classical network reliability models (where the components subject to failure are nodes or links of a graph) and are able to represent more complex interactions between the states of the individual components and the operation of the system under study.**

The reliability evaluation of stochastic binary systems belongs to the class of $\mathcal{NP}$-Hard computational problems. Furthermore, the number of states is exponential with respect to the size of the system (measured in the number of components). As a consequence, the representation of an SBS becomes a key element in order to develop exact and/or approximation methods for reliability evaluation.

We introduce the concept of *separable* stochastic binary systems, whose structure can be efficiently represented. Reliability bounds for arbitrary SBS are provided inspired by a measure of a distance to a separable system, duality and Chernoff inequality. Opportunities for future work arising from this representation are also discussed.

Index Terms—**Stochastic Binary System, Network Reliability, Computational Complexity, Chernoff Inequality.**

I. Introduction

In system reliability analysis, the goal is to find the probability of correct operation of a system subject to component failures. A common practical problem is to design a system with maximum reliability meeting budget constraints [1], [2], [3], [4].

Classical network reliability analysis shaped the body of this field. In this basic setting, we are given a connected graph G with perfect nodes, and the links work independently with identical probability r. The *all-terminal reliability*, $R_G(r)$, is the probability that the resulting subgraph remains connected. This model and some variants (such as perfect links and nodes subject to failure) has been employed to model reliability of classical communications networks, where the emphasis was on a fixed infrastructure of sites holding communication equipment and of fixed links connecting them. Nevertheless, these models have limitations to represent the more diverse landscape of communication networks infrastructure, relying on different equipments, paradigms, and particularly in the case of wireless networks, where usually there does not exist a fixed, predetermined topology. Stochastic binary systems (SBS) generalize the static reliability concept to any system composed of a number of components subject to independent failures with known probabilities, and where the operation or failure of the system as a whole is a function of the state of the individual components. In this sense, SBS are a more flexible tool for evaluating and optimizing the reliability of a wider spectrum of real systems, both in the networking area and in other quite different applications area [5], [6], [7], [8]. At the same time, SBS present their own challenges in terms of computational analysis, as the evaluation of the reliability a general stochastic binary system belongs to the class of $\mathcal{NP}$-Hard problems. This has motivated different research efforts, tackling efficient exact methods for some subclasses of SBS, as well as approximations for the general case [9], [10], [11], [12].

In this paper, we propose a novel representation of a special subset of stochastic binary systems, called *separable systems*. This representation is then exploited in order to find reliability bounds for arbitrary stochastic binary systems. The contributions can be summarized by the following items:

- An efficient representation of separable systems is proposed. It considers $N + 1$ real numbers, being N the size of the system (measured as the number of components subject to failure).
- We explore the interplay between monotonicity and separability in stochastic binary systems.
- A metric to find the closest separable system for any given SBS ϕ is provided. As a consequence, we find a distance-minimizer separable system ϕ^*, as well as upper and lower bounds ($\overline{\phi}$ and $\underline{\phi}$ respectively).
- Corresponding ILP formulations to find ϕ^*, $\overline{\phi}$ and $\underline{\phi}$ are proposed and solved using CPLEX.
- Reliability bounds for arbitrary SBS are found using the previous construction, duality and Chernoff inequality.
- A Proof-of-Concept shows the performance of the reliability bounds and the benefit of our new representation of separable systems.

This paper is organized as follows. Section II presents fundamental concepts of stochastic binary systems. Separable systems are introduced in Section III. They can be represented more efficiently using $N + 1$ real numbers

instead of 2^N numbers that are used to represent arbitrary SBS. Reliability bounds for arbitrary SBS are found in Section IV using separable systems. A proof-of-concept is presented in Section V. Finally, Section VI has concluding remarks and trends for future work.

II. Stochastic Binary Systems

The following terminology is adapted from [13].

Definition 1 (Stochastic Binary System). A *stochastic binary system* is a triad (S, r, ϕ):

- $S = \{1, \ldots, N\}$ is a ground set of *components*,
- $r = (r_1, \ldots, r_N)$ are their *elementary reliabilities*, and
- $\phi : \{0,1\}^N \to \{0,1\}$ is the *structure*.

The concept of reliability is generalized to arbitrary stochastic binary systems.

Definition 2 (Reliability/Unreliability). Let $\mathcal{S} = (S, p, \phi)$ be a stochastic binary system, and consider a random vector $X = (X_1, \ldots, X_N)$ with independent coordinates governed by Bernoulli random variables such that $P(X_i = 1) = r_i$. The *reliability* of $\mathcal{S}$ is the probability of correct operation of the system:

$$R_{\mathcal{S}} = P(\phi(X) = 1) = E(\phi(X)) = \sum_{x : \phi(x) = 1} P(X = x). \quad (1)$$

The *unreliability* of $\mathcal{S}$ is $U_{\mathcal{S}} = 1 - R_{\mathcal{S}}$.

A stochastic binary system is *homogeneous* if the elementary reliabilities are identical (i.e., $r_i = r$ for all i). In this paper we deal with homogeneous SBS.

Definition 3 (Pathsets/Cutsets). Let $\mathcal{S} = (S, r, \phi)$ be a stochastic binary system. A possible state or configuration $x \in \{0,1\}^N$ is a *pathset* (resp. *cutset*) if $\phi(x) = 1$ (resp., if $\phi(x) = 0$).

The binary set $\{0,1\}$ is equipped with the partial order, defined by $0 \leq 0$, $0 \leq 1$ and $1 \leq 1$. The set $\{0,1\}^N$ inherits a natural order in the Cartesian product. Given two partially ordered sets A and B, a function $f : A \to B$ is monotonically increasing if $f(a_1) \leq f(a_2)$ whenever $a_1 \leq a_2$. As usual, we denote $y < x$ if $y \leq x$ and $y \neq x$. Let us denote by $\overline{0}_N$ (resp. $\overline{1}_N$) the binary word with all bits set to 0 (resp. to 1), and by δ_i the binary word with all bits in 0 except the bit in position i which is set to 1.

Definition 4 (Stochastic Monotone Binary System (SMBS)). The triad $\mathcal{S} = (S, r, \phi)$ is a *stochastic monotone binary system* if the structure function $\phi : \{0,1\}^N \to \{0,1\}$ is monotonically increasing, $\phi(\overline{0}_N) = 0$ and $\phi(\overline{1}_N) = 1$.

Observe that SMBS represent *well-behaved* SBS, in the sense that, given a working configuration, the system can fail after the removal of some components, but can not fail if some failed components start to work. Additionally, the system does not work if it has no operational components, and the full-system works.

Definition 5 (Minpaths/Mincuts/Rays). Let $\mathcal{S} = (S, r, \phi)$ be an SMBS:

- A pathset x is a *minpath* if $\phi(y) = 0$ for all $y < x$.
- A cutset y is a *mincut* if $\phi(x) = 1$ for all $x > y$.
- The x-ray is the set $S_x = \{y \in \{0,1\}^N : y \geq x\}$.

It is worth to remark that an SMBS is fully characterized by its mincuts (or its minpaths). In fact, if we are given the complete list of minpaths, then the complete list of pathsets is precisely the union of the x-rays for some minpath x.

We will denote by $\overline{x}$ the state complementary to x in bits (i.e., 0 in x are set to 1 in $\overline{x}$, and vice-versa). In particular, $\overline{\phi(x)} = 1 - \phi(x)$. The following definition of duality will be useful for our later analysis of monotonicity and bounds [14]:

Definition 6 (Duality). The dual of a stochastic binary system $\mathcal{S} = (S, r, \phi)$ has identical ground set S, elementary reliabilities $r_i^d = 1 - r_i$, and structure $\phi^d(x) = 1 - \phi(\overline{x})$, for all possible states $x \in \{0,1\}^N$. The dual is denoted by $\mathcal{S}^d = (S, 1 - r, \phi^d)$.

The following examples provide an insight of the different applications of stochastic binary systems. Classical examples include a reference in the field for the interested reader.

1) All-Terminal Reliability: the ground set is precisely the links of a simple graph. The system is up if the resulting random graph is connected.
2) K-Terminal Reliability: in the same random graph, the system is up if some distinguished node-set K, called terminals, belong to the same connected component [15].
3) Diameter Constrained Reliability: a diameter constraint d is added to the K-Terminal Reliability. The system is up if every pair of terminals are connected by paths whose length is not greater than the diameter [16], [17].
4) Node-Reliability: the ground set is the set of the nodes of a simple graph. The system is up if the resulting random graph is connected.

There exists an interplay between SBS and propositional logic. Recall that a theorem-proving procedure is the first $\mathcal{NP}$-Complete decision problem established by Stephen Cook [18]. In other words, the recognition of a tautology is a hard decision problem from propositional logic.

Theorem 1. *The reliability evaluation of an arbitrary SMBS belongs to the class of $\mathcal{NP}$-Hard problems.*

Proof. Arnie Rosenthal formally proved that the reliability evaluation for the K-terminal reliability model belongs to the class of $\mathcal{NP}$-Hard computational problems [19]. Since K-Terminal is a particular SMBS, the result follows by inclusion. □

Corollary. *The reliability evaluation of an arbitrary SBS belongs to the class of $\mathcal{NP}$-Hard problems.*

Les us close this section with three elementary properties of the dual system that will be useful in our analysis.

Lemma 1. *The dual of the dual is the original system.*

Proof. $\phi^{dd}(x) = 1 - \phi^d(\overline{x}) = 1 - (1 - \phi(x)) = \phi(x)$. $\square$

Lemma 2. *The dual of an SMBS is another SMBS.*

Proof. Consider arbitrary states $x \leq y$ and a monotone structure ϕ. Since $\overline{x} \geq \overline{y}$, we get that $\phi(\overline{y}) \leq \phi(\overline{x})$. Therefore: $\phi^d(x) = 1 - \phi(\overline{x}) \leq 1 - \phi(\overline{y}) = \phi^d(y)$. $\square$

Tue dual system has complementary reliability with respect to the original one:

Lemma 3. *If $\mathcal{S} = (S, \phi, p)$ is an SBS, then $R_{\mathcal{S}^d} = 1 - R_{\mathcal{S}}$.*

Proof. Recall that the dual system has complementary probabilities in every component. Therefore: $P^d(X = x) = \prod_{i: x_i = 1}(1 - r_i) \prod_{i: x_i = 1}(r_i) = P(X = \overline{x})$. Let $\mathcal{P}$ denote the path-sets of the original SBS. Then:

$$
\begin{aligned}
R_{\mathcal{S}^d} &= \sum_{x: \phi^d(x) = 1} P^d(X = x) = \sum_{x: \phi(\overline{x}) = 0} P^d(X = x) \\
&= 1 - \sum_{x: \phi(\overline{x}) = 1} P^d(X = x) \\
&= 1 - \sum_{x: \phi(\overline{x}) = 1} P(X = \overline{x}) \\
&= 1 - P(x \in \mathcal{P}) = 1 - R_{\mathcal{S}}.
\end{aligned}
$$

$\square$

III. SEPARABLE SYSTEMS

Observe that $\{0, 1\}^N$ is the set of the extremal points of the unit hypercube $Q_N \subseteq \mathbb{R}^N$. Let us assign labels to the extremal points of Q_N according to a given structure ϕ. Every hyperplane defines a partition of $\mathbb{R}^N$ into two subsets. Consider the family of hyperplanes $\mathcal{H}$ such that $\overline{0}_N$ and $\overline{1}_N$ lie on different sides. For any member H of $\mathcal{H}$, denote by $Q_0 \subseteq Q_N$ the extremal points of the hypercube that belong to the side of $\overline{0}_N$; and $Q_1 = Q_N - Q_0$. Define a structure function ϕ_H such that its cutsets are precisely Q_0, and its pathsets are Q_1. Consider an equivalence relation $(\mathcal{H}, \sim)$ such that $H_1 \sim H_2$ if and only if $\phi_{H_1} = \phi_{H_2}$.

Recall that in the Euclidean space $\mathbb{R}^N$, a hyperplane is fully characterized by a normal vector $\vec{n}$ and a point P that belongs to the hyperplane: $\langle \vec{n}, X - P \rangle = 0$, where $\langle x, y \rangle = \sum_{i=1}^{N} x_i y_i$ is the inner product. If we denote $\vec{n} = (n_1, \ldots, n_N)$ and $\langle \vec{n}, P \rangle = \alpha_0$, the hyperplane can be written as $\sum_{i=1}^{N} n_i x_i = \alpha_0$.

Lemma 4. *If $\phi = \phi_H$ for some hyperplane H, then there exists $H_2 \sim H_1$ with non-negative normal vector such that $\|\vec{n}\|_1 = \sum_{i=1}^{N} n_i = 1$.*

Proof. Let $\phi = \phi_H$ for the hyperplane $H) \sum_{i=1}^{N} n_i x_i = \alpha_0$, and suppose that there exists some index j such that $n_j < 0$. There are two exhaustive and mutually disjoint cases:

i There exists some mincut $x = (x_1, \ldots, x_N)$ such that $x_j = 0$: in this case, we know that $x + \delta_j$ is a minpath, so, $\phi(x + \delta_j) = 1$. By the definition of the hyperplane, we

get that $\sum_{i=1}^{N} n_i x_i \leq \alpha_0$ but $\sum_{i=1}^{N} n_i x_i + n_j > \alpha_0$. The only possibility is that $n_j > 0$. But we assumed $n_j < 0$; this is a contradiction.

ii All mincuts verify $x_j = 1$: Consider an alternative hyperplane $H_2) \sum_{i \neq j}^{N} n_i x_i = \alpha_0 - n_j$. We will prove that $H_2 \sim H$. If x is a mincut, then $\sum_{i=1}^{N} n_i x_i \leq \alpha_0$, and therefore $\sum_{i \neq j}^{N} n_i x_i \leq \alpha_0 - n_j$. If x is a minpath, it must have $x_j = 1$. Since $\sum_{i=1}^{N} n_i x_i > \alpha_0$ we get that $\sum_{i \neq j}^{N} n_i x_i > \alpha_0 - n_j$. Observe that $n_j = 0$ in the new hyperplane H_2, and $H_2 \sim H$ as desired.

By an iterative replacement of all the negative coordinates we obtain an equivalent hyperplane $H_2 \sim H$ with non-negative vector $\vec{n'}$, expressed by $H_2) \sum_{i=1}^{N} n'_i x_i = \alpha'$ for some real number α'. Finally, observe that $\overline{0}_N$ is always a cutset, so $0 \leq \alpha'$. Analogously, $\overline{1}_N$ is always a pathset, so $\sum_{i=1}^{m} n'_i > \alpha' \geq 0$. The result is obtained by a normalization of the normal vector $\vec{n}_2$, which is possible since $\sum_{i=1}^{N} n'_i > 0$. $\square$

Even though there exist infinite equivalent hyperplanes, using Support Vector Machine (SVM) it is possible to find a single hyperplane with the largest gap (this is, with the largest distance to any of the vertices in the hypercube). Using Lemma 4, we can replace it by an equivalent hyperplane with non-negative versor. Without loss of generality, we will assume a non-negative normal vector with unit 1-norm.

Proposition 1. *The structures ϕ_H are monotone.*

Proof. By Lemma 4, in particular we can choose $n_i \geq 0$ in the hyperplane $H) \sum_{i=1}^{N} n_i x_i = a_0$. Let us denote $f(x) = \sum_{i=1}^{N} n_i x_i$. If $x_1 \leq x_2$, then $f(x_1) \leq f(x_2)$, and therefore $\phi_H(x_1) \leq \phi_H(x_2)$. $\square$

A subtlety is that the *mincuts* from Lemma 4 are indeed the points $Q_0 \subset Q_N$ that are closer to the original hyperplane. A natural question is to determine if all SMBS can be represented by a hyperplane. The answer is negative:

Proposition 2. *There exist SMBS that cannot be represented by a hyperplane.*

Proof. Consider the SMBS defined by the mincuts $M = \{(1, 1, 0, 0), (0, 0, 1, 1)\}$. Observe that the set of states $P = \{(0, 1, 0, 1), (1, 0, 1, 0)\}$ is a subset of minpaths. Suppose for a moment that there exists some separator $H) \sum_{i=1}^{4} n_i x_i = \alpha$ for some real numbers $\alpha, n_1, \ldots, n_4$. Since $(1, 1, 0, 0)$ and $(0, 0, 1, 1)$ are mincuts, we get that $\sum_{i=1}^{4} n_i \leq 2\alpha$. However, $(1, 1, 0, 0)$ and $(0, 0, 1, 1)$ are minpaths, so $\sum_{i=1}^{4} n_i > 2\alpha$; a contradiction. $\square$

Definition 7 (Separable System). An SBS is *separable* if the cutsets/pathsets can be separated by some hyperplane.

An interpretation of separable systems recalls Riesz representation theorem for Hilbert spaces [20]. Indeed, the structure of a separable system can be written as an indicator

that an inner-product exceeds some threshold in a Hilbert space:

$$\phi(x) = 1_{\langle x, \vec{n} \rangle \geq \alpha_0}. \tag{2}$$

A natural question is to characterize separable systems in terms of pathsets and cutsets. Let us denote $CH(\mathcal{P})$ and $CH(\mathcal{C})$ the convex hull of the pathsets and cutsets respectively.

Proposition 3. *An SBS is separable iff* $CH(\mathcal{P}) \cap CH(\mathcal{C}) = \emptyset$.

Proof. If the intersection is empty, Hahn-Banach separation theorem for convex sets asserts that there exists a hyperplane H that separates those convex sets [20]. As a consequence, $\phi = \phi_H$ for some hyperplane H.

For the converse, we know that the SBS is separable. Therefore, there exists some hyperplane H) $\sum_{i=1}^{N} n_i x_i = \alpha_0$ such that $\sum_{i=1}^{N} n_i x_i \leq \alpha_0$ for cutsets, and $\sum_{i=1}^{N} n_i x_i > \alpha_0$ for pathsets. Suppose for a moment that $CH(\mathcal{P}) \cap CH(\mathcal{C}) \neq \emptyset$. There exists some element $z \in \mathbb{R}^N$ such that:

$$z = \sum_{j=1}^{r} \alpha_j x_j = \sum_{k=1}^{s} \beta_k y_k, \tag{3}$$

for some states $x_1, \ldots, x_r \in P$, $y_1, \ldots, y_s \in C$, and non-negative numbers such that $\sum_{j=1}^{r} \alpha_j = \sum_{k=1}^{s} \beta_k = 1$. If we denote $x_j = (x_{j1}, \ldots, x_{jN})$ we know that $\sum_{i=1}^{N} n_i x_{ji} > \alpha_0$. Therefore, for $z = (z_1, \ldots, z_N)$ we get that:

$$\begin{aligned}
\sum_{i=1}^{N} n_i z_i &= \sum_{i=1}^{N} n_i (\sum_{j=1}^{r} \alpha_j x_{ji}) \\
&= \sum_{j=1}^{r} \alpha_j [\sum_{i=1}^{N} n_i x_{ji}] \\
&> (\sum_{j=1}^{r} \alpha_j) \alpha_0 = \alpha_0.
\end{aligned}$$

Analogously, using the fact that $z = \sum_{k=1}^{s} \beta_k y_k$ we get that $\sum_{i=1}^{N} n_i z_i \leq \alpha_0$, which is a contradiction. Therefore we must have $CH(\mathcal{P}) \cap CH(\mathcal{C}) = \emptyset$, and the result holds. $\quad\square$

IV. RELIABILITY BOUNDS

In this section we exploit the properties shared by separable systems in order to find reliability bounds for arbitrary SBS. The strategy is the following:

- First, we find an upper bound for the reliability of separable systems using Chernoff inequality [21].
- For any given structure ϕ, we find the closest separable systems $\overline{\phi}$ and $\underline{\phi}$ such that $\underline{\phi} \leq \phi \leq \overline{\phi}$.
- By means of Chernoff-upper bound for separable systems and $\overline{\phi}$, we produce an upper bound for the reliability of the original SBS.
- By means of Chernoff-upper bound for separable systems, $\underline{\phi}$ and duality, we produce an lower bound for the reliability of the original SBS.

We describe each step in the following subsections.

A. Chernoff Bound

Lemma 5. *For all separable systems* $\mathcal{S}$ *we have:*

$$R_{\mathcal{S}} \leq e^{-\sup_{t>0} \{t\alpha_0 - \sum_{i=1}^{m} c(i,t)\}}, \tag{4}$$

being $c(i,t) = \log \mathbb{E}\left(e^{t x_i n_i}\right) = \log(p.e^{t n_i} + 1 - p)$.

Proof. We apply a well-known result introduced by Chernoff and used in the proof of Cramér Theorem for large deviations bounds [21]. For each $t > 0$ we have:

$$\begin{aligned}
R_{\mathcal{S}} &= \mathbb{P}(\phi(x) = 1) = \mathbb{P}(\sum_{i=1}^{m} x_i n_i \geq \alpha_o) \\
&= \mathbb{P}(t \sum_{i=1}^{m} x_i n_i \geq t\alpha_o) = \mathbb{P}(e^{t \sum_{i=1}^{m} x_i n_i} \geq e^{t\alpha_o}) \\
&\leq e^{-t\alpha_o} \mathbb{E}\left(e^{t \sum_{i=1}^{m} x_i n_i}\right) = e^{-t\alpha_o} \prod_{i=1}^{m} \mathbb{E}\left(e^{t x_i n_i}\right) \\
&= e^{-t\alpha_o} \prod_{i=1}^{m} \left(e^{c(i,t)}\right) = e^{-\{t\alpha_o - \sum_{i=1}^{m} c(i,t)\}}, \tag{5}
\end{aligned}$$

where Markov's inequality for positive random variables has been used. The result holds taking the infimum with respect to $t > 0$ on both sides. $\quad\square$

B. Closest Separable Systems

For any given structure ϕ, we build the *closest* separable structures $\overline{\phi}$ and $\underline{\phi}$ such that $\underline{\phi} \leq \phi \leq \overline{\phi}$ in terms of misclassification error:

Definition 8 (Misclassification Error)**.** If we are given two structures ϕ_1 and ϕ_2, the misclassification error is:

$$d(\phi_1, \phi_2) = \sum_{x \in \{0,1\}^m} |\phi_1(x) - \phi_2(x)|. \tag{6}$$

Clearly, d is a metric in the space of all structures. For a rough approximation of ϕ into a separable systems we also consider the closest separable ϕ^* (without bounds):

Definition 9 (Minimum Least Square)**.** Given an arbitrary structure ϕ, the separable structure ϕ^* minimizes the misclassification error.

Proposition 4. *The lower bound for* ϕ *is* $\underline{\phi} = 1 - \phi'$, *being* ϕ' *the upper bound of* $1 - \phi$.

Proof. By duality, $1 - \underline{\phi} \geq 1 - \phi \geq 1 - \overline{\phi}$. The closest upper-bound for $1 - \phi$ is precisely $1 - \underline{\phi} = \phi'$, and the result holds. $\quad\square$

Now, we fully characterize ϕ^* and $\overline{\phi}$ using Integer Linear Programming (ILP) formulations. The lower-bound $\underline{\phi}$ can be

obtained using Proposition 4. The following ILP describes $\underline{\phi}$:

$$\text{min.} \sum_{x \in \{0,1\}^N} d(x) \tag{7}$$

s.t.

$$\underline{\phi}(\overline{0}) = 0, \tag{8}$$

$$\underline{\phi}(\overline{1}) = 1, \tag{9}$$

$$d(x) = \phi(x) - \underline{\phi}(x), \forall x \in \{0,1\}^N \tag{10}$$

$$\sum_{j=1..N} x_{ij}\alpha_j > \underline{\phi}(x_i), \forall i = 1..2^N \tag{11}$$

$$\sum_{j=1..N} x_{ij}\alpha_j \leq 1 + M\underline{\phi}(x_i), \forall i = 1..2^N \tag{12}$$

Where:

- Constraints (8) and (9) establish monotonicity.
- Constraint (10) defines the misclassification error $d(x)$.
- Constraints (11) and (12) state that ϕ^* is separable.

Observe that Constraints (11) are active when $x_i = (x_{i1}, \ldots, x_{iN})$ is a pathset. Note that in this case (12) is superfluous (choosing the constant M large enough). Constraints (12) define a similar condition for cutsets.

The following ILP model provides ϕ^*:

$$\text{min.} \sum_{x \in \{0,1\}^N} d(x) \tag{13}$$

s.t.

$$\phi^*(\overline{0}) = 0, \tag{14}$$

$$\phi^*(\overline{1}) = 1, \tag{15}$$

$$d(x) = \delta^+(x) + \delta^-(x), \forall x \in \{0,1\}^N \tag{16}$$

$$\phi^*(x) = (1 - \phi(x))d(x) + \phi(x)(1 - d(x)), \forall x \in \{0,1\}^N \tag{17}$$

$$\delta^+(y) + \delta^-(x) \geq \phi(x) - \phi(y), \forall (x,y) \in \{0,1\}^N, x \leq y \tag{18}$$

$$\sum_{j=1..N} x_{ij}\alpha_j > \phi^*(x_i), \forall i = 1..2^N \tag{19}$$

$$\sum_{j=1..N} x_{ij}\alpha_j \leq 1 + M\phi^*(x_i), \forall i = 1..2^N \tag{20}$$

Where:

- $\delta^+(x)$ is set to 1 if x is modified from a cutset into a pathset.
- $\delta^-(x)$ is set to 1 if x is modified from a pathset into a cutset.
- $d(x)$ is set to 1 if $\phi(x) \neq \phi^*(x)$.

C. Reliability Bounds

Combining the ILP formulation for $\underline{\phi}$ and Chernoff-bound (Lemma 5), a lower-bound for an arbitrary SBS ϕ is produced. Finally, combining the ILP formulation for $\underline{\phi}$, Chernoff-bound and Theorem 2, an upper-bound for an arbitrary SBS ϕ is produced. First, a technical lemmas:

Lemma 6. *The dual of a separable system is also separable.*

Proof. If we are given a separable system with hyperplane $H) \sum_{i=1}^N n_i x_i = \alpha_0$, being $\vec{n}$ non-negative, then $H^d) \sum_{i=1}^N n_i x_i = 1 - \alpha_0$. In fact, if we are given a pathset from the dual $x \in \mathcal{P}^d$, we know that $\overline{x} = \overline{1}_N - x$ is a cutset in the original system, and $\sum_{i=1}^N n_i x_i = \sum_{i=1}^N n_i(1 - \overline{x_i}) = 1 - \sum_{i=1}^N n_i \overline{x_i} > 1 - \alpha_0$. A similar calculation holds for pathsets. $\square$

Theorem 2. $R_\mathcal{S} \geq 1 - R'$, being R' Chernoff-bound for $(\underline{\phi})^d$.

Proof. By duality we know that $\phi^d \leq (\underline{\phi})^d$. Since the dual of a separable system is also separable, we can apply Lemma 5 in order to find an upper bound R' for the reliability of the system $(\underline{\phi})^d$. We get that $R_{\mathcal{S}^d} \leq R'$. Finally, recall that the reliability of a dual is complementary to the reliability of the original system (Lemma 3). Therefore:

$$R_\mathcal{S} = 1 - R_{\mathcal{S}^d} \geq 1 - R'. \tag{21}$$

$\square$

V. PROOF-OF-CONCEPT

Our goal is to understand the performance of our bounding method for some sample situations. We consider a wireless system subject to node failures. For this reason we consider the all-terminal Node-Reliability model. Recall that Node-Reliability is not an SMBS in general.

We considered the graphs sketched in Figures 1-4. For each graph we consider the SBS given by Node Reliability with structure ϕ. Then, we find the closest SMBS ϕ_m and ϕ^u, the closest separator systems $\underline{\phi}$, $\overline{\phi}$ and ϕ^*, solving the respective ILP formulations and CPLEX optimization engine. Tables I and II report the misclassification errors and reliability bounds respectively, using $r_i = r = 1/2$ for the elementary reliabilities. The asterisk $*$ means that the optimization reached the limit of three hours, and this value is sub-optimal; this is the case of the Icosahedron graph (I).
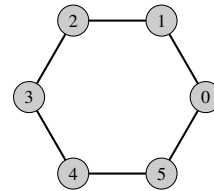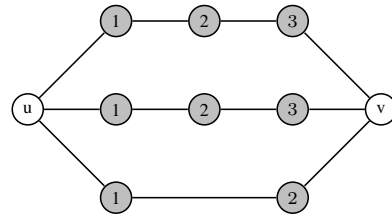

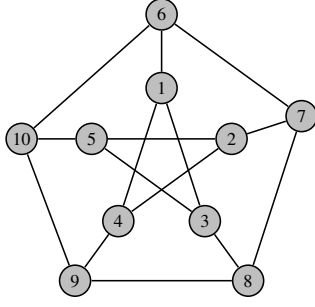Fig. 1. Elementary cycle C_6

Fig. 2. Monma graph $M_{(3,3,2)}$.
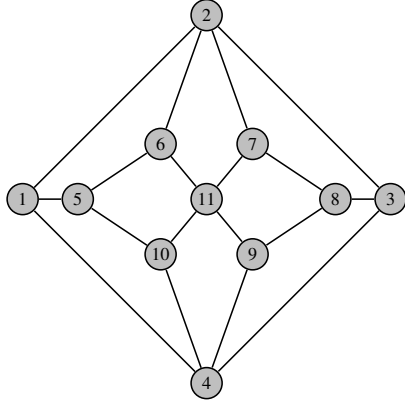
Fig. 3. Petersen graph (P)

Fig. 4. Icosahedron graph (I)

TABLE I
MISCLASSIFICATION ERROR

Case	$d(\phi, \phi_m)$	$d(\phi, \phi^u)$	$d(\phi, \underline{\phi})$	$d(\phi, \overline{\phi})$	$d(\phi, \phi^*)$
C_6	12	21	12	23	12
M_{10}	106	720	113	759	113
P	175	390	314	425	190
I	302	958	537	958	379*

TABLE II
PERFORMANCE OF RELIABILITY AND BOUNDS

Case	$R_\mathcal{S}(\phi)$	$R_\mathcal{S}(\phi^*)$	$R_\mathcal{S}(\underline{\phi})$	$R_\mathcal{S}(\overline{\phi})$	LB	UB
C_6	0.3906	0.2031	0.2031	0.7500	0	1
M_{10}	0.1846	0.0801	0.0742	0.9258	0	1
P	0.5449	0.5059	0.2383	0.9600	0	1
I	0.5317	0.4932	0.2695	0.9995	0	1

Observe that ϕ^* achieves the minimum distance, $d(\phi, \phi_m) \leq d(\phi, \underline{\phi})$, and $d(\phi, \phi^u) \leq d(\phi, \overline{\phi})$, as expected by definition. However, the gaps are small. This suggests that finding bounds for an arbitrary SBS by separable systems (which are advantageous due to their small space requirement for representation) may not entail a large loss of precision when compared by bounds obtained using SMBS approximations (which potentially need exponential space for representation).

From Table II we can check that $R_\mathcal{S}(\underline{\phi}) \leq R_\mathcal{S}(\phi) \leq R_\mathcal{S}(\overline{\phi})$, while $R_\mathcal{S}(\phi^*)$ is closer to $R_\mathcal{S}(\phi)$. The last two columns UB is the upper bound found applying Lemma 5 directly and using the separator hyperplane for $\overline{\phi}$. Column LB is calculated applying Theorem 2 and using the separator hyperplane for $\underline{\phi}^d$. For the four cases under study, these formulations result in trivial bounds.

Figures 5-8 display the exact reliabilities $R_\mathcal{S}(\underline{\phi}) \leq R_\mathcal{S}(\phi) \leq R_\mathcal{S}(\overline{\phi})$, together with Chernoff bounds, for each network. The non-monotonicity of $R_\mathcal{S}(\phi)$ is appreciated Icosahedron network topology. The bounds provided by the separable systems are much tighter than the Chernoff bounds. At the same time, there is a clear gap between lower and upper bounds, specially for medium-range values of p. When p approaches 1, the quality of the bounds improve; and particularly the lower bound is closer to the exact value. This is of interest, as usually when designing or evaluating a system, the goal is to guarantee a certain level of reliability (thus, making the lower bound a relevant approximation).
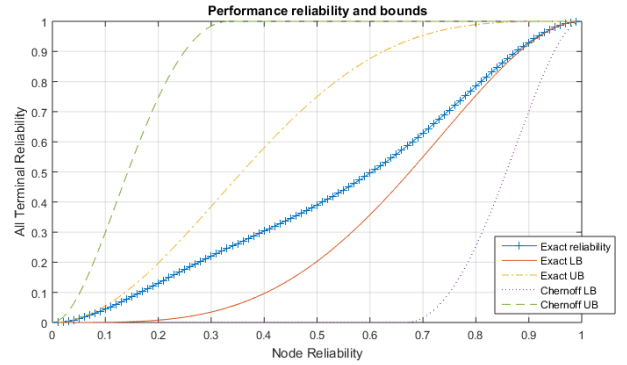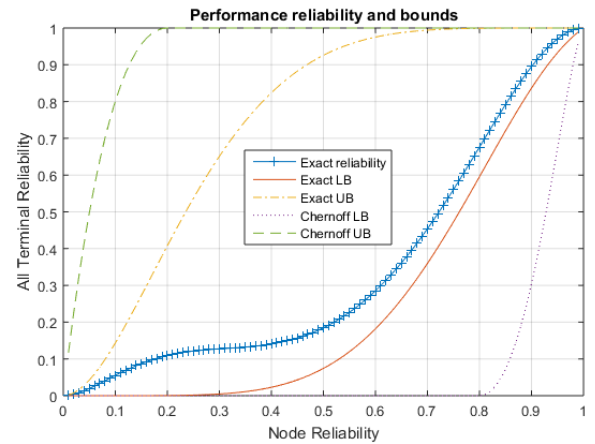

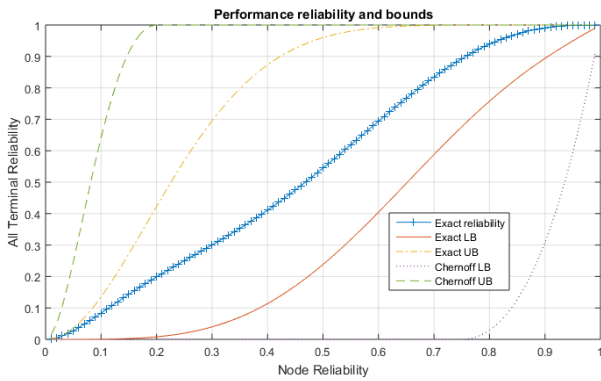Fig. 5. C6 graph

Fig. 6. M10 graph
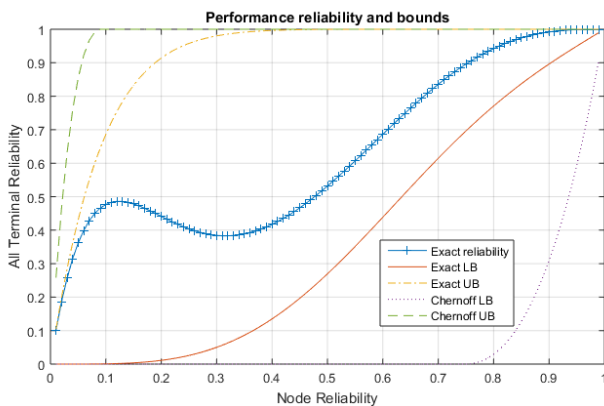
Fig. 7. Petersen graph

Fig. 8. Icosahedron graph

VI. Concluding Remarks

An efficient representation of separable systems is here introduced. This representation is analogous to Riesz Representation Theorem for Hilbert spaces, but for particular SBS, using a simple inner product. Supported by this natural representation, we produce reliability bounds for arbitrary SBS, exploiting duality and Chernoff inequality. The results are highlighted in systems under the node-reliability model.

This interesting interplay between Stochastic Binary Systems and Functional Analysis should be further studied. As a future work, we would like to develop new reliability bounds using the theory of Functional Analysis, and apply these results to potential applications in real-life systems. Other lines of research include taking into account dependencies between the components' states, and studying how SBS structure can be exploited in a dynamic context (i.e, when the time dimension is taken into account so that the components' states are evolving, i.e, failing and being repaired, at different moments of the system evolution).

References

[1] Y. Chopra, B. Sohi, R. Tiwari, and K. Aggarwal, "Nework topology for maximizing the terminal reliability in a computer communication network," *Microelectronics Reliability*, vol. 24, no. 5, pp. 911 – 913, 1984.

[2] H. W. Lee, K. Lee, and E. Modiano, "Maximizing reliability in wdm networks through lightpath routing," *IEEE/ACM Transactions on Networking*, vol. 22, no. 4, pp. 1052–1066, Aug 2014.

[3] J. Barrera, H. Cancela, and E. Moreno, "Topological optimization of reliable networks under dependent failures," *Operations Research Letters*, vol. 43, no. 2, pp. 132 – 136, 2015.

[4] D. Coudert, J. Luedtke, E. Moreno, and K. Priftis, "Computing and maximizing the exact reliability of wireless backhaul networks," in *International Network Optimization Conference (INOC)*, Lisbon, Portugal, 2017.

[5] J. Barta and Montemanni, "Wireless multicasting under probabilistic node failures: a heuristic approach," *Optimization and Engineering*, vol. 13, no. 4, pp. 705–726, 2012.

[6] F. Markowetz and M. Boutros, *Linking Phenotypes and Genotypes*, ser. Cambridge Series in Systems Genetics. Cambridge University Press, 2015.

[7] H. Cancela, G. Guerberoff, F. Robledo, and P. G. Romero, "Reliability maximization in stochastic binary systems," in *14th International Conference on the Design of Reliable Communication Networks (DRCN 2018)*, Paris, France, February 2018.

[8] X. Song, Z. Zhai, Y. Liu, and J. Han, "A stochastic approach for the reliability evaluation of multi-state systems with dependent components," *Reliability Engineering & System Safety*, vol. 170, no. Supplement C, pp. 257 – 266, 2018.

[9] M. Lin, "An efficient algorithm for computing the reliability of stochastic binary systems," *IEICE Transactions*, vol. 87-D, no. 3, pp. 745–750, 2004.

[10] E. Canale, H. Cancela, J. Piccini, F. Robledo, P. Romero, G. Rubino, and P. Sartor, "Recursive variance reduction method in stochastic monotone binary systems," in *2015 7th International Workshop on Reliable Networks Design and Modeling (RNDM)*, Oct 2015, pp. 135–141.

[11] E. Canale, H. Cancela, F. Robledo, P. Romero, and P. Sartor, "Diameter constrained reliability: Complexity, distinguished topologies and asymptotic behavior," *Networks*, vol. 66, no. 4, pp. 296–305, 2015.

[12] G. Levitin, *Computational Intelligence in Reliability Engineering: Evolutionary Techniques in Reliability Analysis and Optimization*, ser. Studies in Computational Intelligence. Springer Berlin Heidelberg, 2006.

[13] M. O. Ball, "Computational complexity of network reliability analysis: An overview," *IEEE Transactions on Reliability*, vol. 35, no. 3, pp. 230 –239, aug. 1986.

[14] P. Romero, "Duality in stochastic binary systems," in *2016 8th International Workshop on Resilient Networks Design and Modeling (RNDM)*, Sept 2016, pp. 85–91.

[15] M. O. Ball, "Complexity of network reliability computations," *Networks*, vol. 10, no. 2, pp. 153–165, 1980.

[16] L. Petingi and J. Rodriguez, "Reliability of networks with delay constraints," in *Congressus Numerantium*, vol. 152, 2001, pp. 117–123.

[17] E. Canale, H. Cancela, F. Robledo, P. Romero, and P. Sartor, "Full complexity analysis of the diameter-constrained reliability," *International Transactions in Operational Research*, vol. 22, no. 5, pp. 811–821, 2015.

[18] S. A. Cook, "The complexity of theorem-proving procedures," in *Proceedings of the third annual ACM symposium on Theory of computing*, ser. STOC '71. New York, NY, USA: ACM, 1971, pp. 151–158.

[19] A. Rosenthal, "Computing the reliability of complex networks," *SIAM Journal on Applied Mathematics*, vol. 32, no. 2, pp. 384–393, 1977.

[20] J. Conway, *A Course in Functional Analysis*, ser. Graduate Texts in Mathematics. Springer New York, 1994.

[21] H. Cramér, "Half a century with probability theory: some personal recollections," *Annals of Probability*, vol. 4, pp. 509–546, 1976.