



HAL
open science

On the Marshall–Olkin Copula Model for Network Reliability Under Dependent Failures

Omar Matus, Javiera Barrera, Eduardo Moreno, Gerardo Rubino

► **To cite this version:**

Omar Matus, Javiera Barrera, Eduardo Moreno, Gerardo Rubino. On the Marshall–Olkin Copula Model for Network Reliability Under Dependent Failures. *IEEE Transactions on Reliability*, 2019, 68 (2), pp.451 - 461. 10.1109/tr.2018.2865707 . hal-01962918

HAL Id: hal-01962918

<https://inria.hal.science/hal-01962918>

Submitted on 21 Dec 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

On the Marshall-Olkin Copula Model for Network Reliability under Dependent Failures

Omar Matus, Javiera Barrera, Eduardo Moreno, Gerardo Rubino

Abstract

The Marshall-Olkin (MO) copula model has emerged as the standard tool for capturing dependency between components in failure analysis in reliability. In this model shocks arise at exponential random times, that affect one or several components inducing a natural correlation in the failure process. However, because the number of parameter of the model grows exponentially with the number of components, MO suffers of the “curse of dimensionality”. MO models are usually intended to be applied to design a network before its construction, therefore it is natural to assume that only partial information about failure behavior can be gathered, mostly from similar existing networks. To construct such a MO model, we propose an optimization approach to define the shock’s parameters in the MO copula, in order to match marginal failures probabilities and correlations between these failures. To deal with the exponential number of parameters of this problem, we use a column-generation technique. We also discuss additional criteria that can be incorporated to obtain a suitable model. Our computational experiments show that the resulting MO model produces a close estimation of the network reliability, especially when the correlation between component failures is significant.

Index Terms

Reliability, copula theory, optimization methods, failure analysis, network design.

NOMECLATURE

ACRONYMS

MO Marshall-Olkin

NOTATIONS

\mathcal{N}	Set of nodes
\mathcal{C}	Set of components or links
\mathcal{G}	Graph or network, composed by nodes \mathcal{N} and components or links \mathcal{C}
p_i	Marginal failure probability of component i
ρ_{ij}	Correlation between components i and j
n	Number of components
W_V	exponential random time at which a shock affecting components in set V arise
λ_V	Rate at which shocks affecting components in set V arise
T_i	Random lifetime of component i given by the first time a shock affecting it arise
\mathcal{P}	Collection of subsets of \mathcal{C} , where shocks rates are non zero
X_i	State of component i . It could be either $X_i = 1$ (operative) or $X_i = 0$ (inoperative)
$R(t)$	Reliability of the network at time t

O. Matus, J. Barrera and E. Moreno are with the Faculty of Engineering and Sciences, Universidad Adolfo Ibáñez, Santiago, Chile. G. Rubino is with INRIA Rennes – Bretagne Atlantique, Rennes, France.

I. INTRODUCTION

THE topological design and evaluation of telecommunication networks have been addressed in many studies since the 1980s. These systems are designed taking into account performance and dependability goals, and costs. We are concerned here with the dependability side, and we will focus on the central family of metrics collectively known as network reliability. Among the various definitions of network reliability, in this work, we pay attention to the reliability of a network in terms of the probability that a given set of components will remain in the same connected component of the network at a given time, considering that these components may become inoperative over time. This is probably the most used dependability characteristic of communication infrastructures, because loss of connectivity precludes communications, whatever the way information is organized and travels through the network.

To evaluate the reliability of a network, a failure model is required, but even for the simplest of them, in which each component fails independently with some probability, it is difficult to directly evaluate the reliability. More precisely, the exact computation of the different reliability metrics is a #P-complete problem (see [1], [2]). Hence, Monte Carlo simulation techniques have become indispensable for computing such reliabilities (see, for instance, [2], [3], or the on-line proceedings of the Winter Simulation Conference).

One of the main applications of network reliability evaluation tools is network design. The problem consists in building the graph defined by the nodes composing the networks and the links between them, minimizing the total cost of the infrastructure and maximizing its reliability, clearly contradictory goals. The problem is a complex one, since in its most general aspect, the task has a huge number of variants in practice, and the networks are obviously designed taking not only dependability criteria into account but also efficiency, low latencies, etc. In some cases part of the structure is already known, that is, we must add nodes and links to an existing graph of connections, sometimes the designer must start from scratch. If the evaluation of reliability is already a difficult problem, the design problem of maximizing reliability is even more so; thus, to tackle the latter problem, many simplifications are assumed, the most common being the independent failures of the components.

However, empirical analyses of networks have revealed that a significant dependence often exists among the failures of different components, which could considerably affect reliability computations. In [4] authors analyzed data from a Norwegian optical network studying the up/down state of links and routers during a year. They show that for an important number of pairs of components, there is a significant correlation between their states and that this correlation decays when the distance between them increases. Similarly, in [5] authors found that simultaneous failures of two, three or four links represent 40% of the total number of failures. Similar conclusions were obtained in [6] during a study of internet provider networks. These studies show that the dependence among component failures is significant, even for telecommunication networks based on different technologies. Therefore, as mentioned in [7], overlooking the correlation between failure events could lead to the undervaluation of the risk incurred in the network, and thereafter to the inappropriate decision for responses, mitigation and design. This issue is also discussed in [8].

In order to deal with this underestimated aspect of the global problem, the Marshall-Olkin (MO) copula model has become the standard dependency tool for failure analysis in reliability. The model considers a system with n components that are operative at an initial time. Then, shocks arise at exponentially distributed independent times, each of them simultaneously affecting subsets of components. Hence, a component will remain operative until the first shock affecting it occurs. This model was proposed in [9] for a system with two components. Then, the MO copula model was extended for a system with several components and two types of shocks, those that affect each component independently and one shock that affects all components. This model has been also known in the literature as the β -factor model. Later,

the MO copula model was extended considering shocks that can affect any subset of components (see [10, Ch. 6]).

The problem of reliability evaluation in a system failing according to a MO model has been addressed by different techniques. Analytical expressions can be obtained for specific topologies, for example, for the k -out-of- n system (see [11]–[13]), but for general networks, even in the exchangeable case, we only rely on simulation techniques [14]. Up to our knowledge, the only work dealing with the non-exchangeable case is [15], that proposed different rare-event simulation techniques to compute the static network reliability. For the design problem, in [16] the authors proposed an optimization formulation to design networks under the MO model and causal failure models (the latter being a form of failure propagation in the system).

The tractability of the MO model has inspired several extensions leading to a large body of articles, which has been carefully discussed in [17]. It is known that the MO model suffers from the so-called “curse of dimensionality”, because the number of parameters of the model grows exponentially with the number of components [18]. Sometimes, real-world situations naturally suggest that only a small subset of shocks are possible, reducing the number of parameters in the MO model, but there is not always enough evidence to restrict the possible shocks. Another assumption widely utilized to reduce this number of parameters is to suppose that components are exchangeable; this essentially means that shocks affecting the same number of components have the same rate. However, there are many natural situations when this assumption is not realistic, because real-world applications need non-exchangeable models (see [19]). So, the study of MO models and their generalization to the non-exchangeable case has been an emerging research topic in recent years. In [20], the authors propose a methodology to estimate the parameters of a non-exchangeable MO model, generalizing the exchangeable Lévy-Frailty representation of the MO copula model. This estimation procedure requires a large number of i.i.d. samples of the components failure times.

Reliability evaluation and network design techniques discussed above are usually intended to be applied to design a network before its construction. Therefore, it is natural to assume that only partial information can be gathered, mostly from similar existing networks. As noted by [21] and [7], even in existing telecommunication networks, information of the full failure joint distribution is scarce; hence, suitable models should consider this lack of data.

To underline the importance of taking into account dependencies in dependability analysis, for instance in evaluating network reliability metrics as in this paper, let us briefly look at extreme cases. Assume we have a system with n identical components in series, all with common individual reliability r . If we assume independence between components, that is, between their states (operating or failed), the system’s reliability is r^n . In the extreme case where the components’ states are all the same (imagine a system where the only possible failure event puts all components simultaneously down), the system’s reliability is simply r , and $r > r^n$. Also observe that r is usually close to 1 and that $r^n \rightarrow 0$ as $n \rightarrow \infty$. If instead of being a fragile series architecture, we have a highly redundant one, with all the components working in parallel, again, assuming independence the system’s reliability is $1 - (1 - r)^n$. Now, $r < 1 - (1 - r)^n$ and $1 - (1 - r)^n \rightarrow 1$ as $n \rightarrow \infty$. So, using the classical independence assumption, we can overestimate or underestimate the reliability of the system when there are dependencies, and the induced relative error can be important and even huge if the structure is a series (or something close to, as it is typically the case), as observed in [22].

In summary, we are interested in a MO model that will be used to evaluate the reliability of a network, without assuming exchangeability or any other simplification on the possible shocks affecting the network. Moreover, considering the lack of data available, the parameters of this model should be obtained using only partial information of the network failures. In particular, we assume that for the random vector

representing the state (up/down) of the network at a fixed time, only their marginal probabilities and correlations are available.

In this paper, we propose a methodology for constructing such MO model. To obtain the required marginal failure probabilities and correlations, this methodology propose an optimization model to find the parameters of the shocks in the MO copula. To deal with the exponential number of parameters of this problem, we use a column-generation technique to reduce the complexity. We also discuss additional criteria that can be incorporated to this optimization problem. Our computational experiments show that the resulting MO model produces a close estimation of the network reliability, especially when the correlation between component failures is important.

We focus our work in the context of telecommunication networks. However, we think that it is also of interest for other types of networks, such that electrical networks [23] or flows networks [24], for other dependability metrics, such as the Mean Time To Failure or the highly combinatorial Diameter Constrained Reliability (see [25]), and for other research areas where MO copula models are utilized, like in Finance (see [26] and [17]).

The article is organized as follows. In Section II, we introduce the notation and formally introduce the MO model. In Section III, we present an optimization problem for obtaining a MO copula model using only the marginal probabilities and correlations of failures. In Section IV we use a reference failure model to compare the resulting reliabilities obtained using the proposed optimization model. Finally, in Section V, we present the conclusions of our study. A preliminary conference paper of this work was presented in the Winter Simulation Conference 2016 [22].

II. MARSHALL-OLKIN (MO) COPULA MODEL FOR DEPENDENT FAILURES

We will use the standard model of telecommunication networks in the area, a graph where the elements subject to failures are the links. We denote the graph as $\mathcal{G} = (\mathcal{N}, \mathcal{C})$ where \mathcal{N} is the set of nodes and \mathcal{C} the set of links. We denote $\mathcal{C} = \{1, 2, \dots, n\}$. Let us suppose that the graph is connected (meaning in particular that $n \geq |\mathcal{N}| - 1$). As stated before, links are subject to failure, and at any time $t \geq 0$, they can be in one of two states, either operational (up) or not operational (down). At time $t = 0$, all components are operational. After random lifetimes in the up state, links fail, moving into the down state, in which they remain forever (that is, we consider non-repairable systems). For each link i , we define its lifetime T_i as the random time at which it fails. In the standard model of network reliability, the random variables T_1, \dots, T_n are assumed to be independent. Here, the model is much more general. Let \mathcal{P} be a collection of subsets of \mathcal{C} , which are not necessarily disjoint. We can assume that for each component i there exists at least some $V \in \mathcal{P}$ such that $i \in V$. For each subset $V \in \mathcal{P}$, we define a positive random variable W_V that represents the instant at which all links in V fail simultaneously. We refer to such an event as a *shock*. The result of a shock associated with subset V is that at time W_V , all links in V are in the down state (if they were already down, they remain down, that is, nothing happens). Hence, the lifetime of a link i corresponds to the earliest time at which a shock that affects i occurs, that is, $T_i = \min_{i \in V} \{W_V\}$. In the MO copula model, the random variables $\{W_V\}_{V \in \mathcal{P}}$ are mutually independent with exponential distributions. Hence, T_i also follows an exponential distribution for all $i \in \mathcal{C}$: if λ_V is the rate of the distribution of W_V and if λ_i is the corresponding rate of T_i , we have that $\lambda_i = \sum_{V: V \ni i} \lambda_V$. A MO model is then given by the triple $(\mathcal{C}, \mathcal{P}, \lambda)$, with $\mathcal{P} \subseteq 2^{\mathcal{C}}$ and where λ represents the map $\lambda: \mathcal{P} \rightarrow \mathbb{R}^+$.

At any time $t \geq 0$, the entire network is either up or down, depending on whether a certain connectivity property is satisfied. In this paper, we illustrate our approach using the basic s, t -connectivity: two nodes in \mathcal{N} , denoted by s and t , are marked, and the network is up if and only if there exists a path connecting them that is composed only of operational links. Assuming that the graph \mathcal{G} is connected, at time $t = 0$ the network is operational. Since the components are not repaired, there is a random time T at which

connectivity between s and t is lost (unless there is a path connecting s and t that never fails). At any time $t \geq T$ the network is then down.

The metric of interest here is the probability that the network will be up at time t , the reliability at time t , $R(t)$. In some cases, the main instant t of interest is the so-called ‘‘mission time’’, the maximal time value until which we wish the network to continue operating without problems. Because t is mathematically arbitrary here, we choose to develop our technique using a value of $t = 1$ for simplicity. Equivalently, we may suppose that we want the reliability at the mission time, and assume that we use precisely the mission time as time unit.

There is another reason for which a value of 1 is interesting. Consider the classical setting with independent components and a *static* model (that is, a model in which time is not an explicit variable). Components and systems are either up or down; we are given the graph and the individual probabilities for the components to be, say, down, denoted by p_i for component i , and we measure the network dependability using some measure of its capability to provide communication services, for instance, the probability that s and t are connected in the implicit random partial graph of \mathcal{G} . This measure is typically called the *network reliability*. Then, consider our initial dynamic model, with $\mathcal{P} = \{\{1\}, \dots, \{n\}\}$, where the random variable $W_{\{i\}}$ has rate $\lambda_i = -\ln(1 - p_i)$. The probability that link i will have failed by time t is $\Pr(T_i \leq t) = 1 - \exp(-\lambda_i t)$, and $\Pr(T_i \leq 1) = 1 - \exp(-\lambda_i) = 1 - (1 - p_i) = p_i$. Thus, in this case, the static s, t -network reliability is equal to $R(1)$. This is essentially the complement of the creation process of [27], which is extended for the general MO copula model in [15]. The state of each link (operational or non-operational) is given by the state of the link at a fixed time 1, that is link i is operational if and only if $T_i > 1$. Using the MO copula model, in this work we define the reliability R of the network as the s - t -connectivity, the probability that exists an operative path between two predefined nodes s and t .

As explained in the introduction, to capture the lack of information about communication network failures, we are interested in estimating the parameters of a MO copula model that represents the reliability behavior based only on the marginal failure probabilities and the linear correlations between failures, without considering higher-order dependences. The following lemma characterizes the parameters of a MO copula model that achieves the required marginal failures and their correlations.

Lemma 1. Let X_i denote the state of link i at time 1, that is, $X_i = 1(\text{link } i \text{ is up})$, where $1(P)$ is the indicator function of predicate P . Let p_i be the marginal failure probability of link i , that is, $p_i = \mathbb{P}(X_i = 0)$, and let $\rho_{i,j}$ be the correlation between the failures of links i and j , that is, $\rho_{i,j} = (\mathbb{E}(X_i X_j) - \mathbb{E}(X_i)\mathbb{E}(X_j)) / (\sqrt{\mathbb{V}(X_i)}\sqrt{\mathbb{V}(X_j)})$. Then, a MO copula model with subsets of components in \mathcal{P} and with these marginal distributions and correlations must satisfy the following set of equations, where the set of unknown variables is $\{\lambda_V : V \in \mathcal{P}\}$:

$$\sum_{V \in \mathcal{P}: i, j \in V} \lambda_V = \gamma_{i,j}, \quad \forall i, j \in \mathcal{C}, i \neq j, \quad (1a)$$

$$\sum_{V \in \mathcal{P}: i \in V} \lambda_V = -\ln(1 - p_i), \quad \forall i \in \mathcal{C}, \quad (1b)$$

$$\lambda_V \geq 0, \quad \forall V \in \mathcal{P}, \quad (1c)$$

where

$$\gamma_{i,j} = \ln \left(\frac{\rho_{i,j} \sqrt{p_i} \sqrt{p_j}}{\sqrt{(1 - p_i)(1 - p_j)}} + 1 \right). \quad (2)$$

We remark that although it is technically possible to consider all higher-order dependences (see [15, Sec. 2]), previous system of equations involves only the first two moments of the state of the network, to reflect the lack of information about failures. Also, note that the system of equations (1) does not

necessarily have a solution. For example, if any correlation $\rho_{i,j}$ is negative, then there is no MO copula model that captures this type of dependency.

On the following, we will assume that the map of a MO model λ allows the rates to take the value 0. A rate $\lambda_V = 0$ implies that the shock arise in $W_V = \infty$. When we consider all subsets of \mathcal{C} in the MO model, that is, when $\mathcal{P} = 2^{\mathcal{C}}$, linear system (1) in λ contains $n + \binom{n}{2} = n(n+1)/2$ equations and $2^n - 1$ variables, and $2^n - 1 > n(n+1)/2$ if $n \geq 3$, so in general, it will have an infinite number of solutions. The reader can check that already for $n = 3$, the space of solutions has dimension 1. Note than an infinite number of these solutions could have many variables λ_V with value 0.

Henceforth, we will denote the network reliability at $t = 1$ simply by R , that is, $R = R(1)$.

III. AN OPTIMIZATION PROBLEM FOR ESTIMATING MO COPULA PARAMETERS

In this section, we discuss how to obtain a set of copula parameters that satisfies the marginal failure probabilities (1b) and approximates the correlation between failures (1a). In the worst case, when we choose all possible subsets of V as copula subsets, the system of equations in (1) has $n(n+1)/2$ equalities and $2^n - 1$ positive variables. So, solving a system with this exponential number of variables cannot be done using traditional methods for even moderate values of n . Moreover, we are interested in comparing different additional criteria that are relevant in many practical situations. This leads to the conclusion that an optimization-based methodology is appropriate to solve our problem.

A. Approximating the correlations between failures

Assuming independence between failures is equivalent to consider only copula subsets of size 1 (i.e., single component failures), so (1b) gives the copula rates of the model. We could consider other copula subsets such that $|\mathcal{P}| \ll 2^n - 1$, for example, copula subsets of a size smaller than k for some moderate integer k , where there is no guarantee of having a solution to the system (1). In this case, we require the solutions to satisfy at least the marginal probabilities of failures, and we seek to approximate correlations using the L_1 -norm of the difference between the desired and obtained values of (1a), in order to still have a linear system of equations.

To find this approximated MO copula model, similarly to the classic “phase-1”-model of the Simplex method, we add positive slack variables t^+ and t^- to equations (1a), and we minimize the sum of these slack values. The resulting linear optimization model is the following:

$$\text{minimize } \sum_{i,j \in \mathcal{C}} (t_{ij}^+ + t_{ij}^-) \quad (3a)$$

subject to

$$\sum_{V \in \mathcal{P}: i,j \in V} \lambda_V + t_{i,j}^+ - t_{i,j}^- = \gamma_{i,j}, \forall i, j, \in \mathcal{C}, i \neq j, \quad (3b)$$

$$\sum_{V \in \mathcal{P}: i \in V} \lambda_V = -\ln(1 - p_i), \quad \forall i \in \mathcal{C}, \quad (3c)$$

$$\lambda_V \geq 0, \quad \forall V \in \mathcal{P}, \quad (3d)$$

$$t_{i,j}^+, t_{i,j}^- \geq 0, \quad \forall i, j \in \mathcal{C}. \quad (3e)$$

For a case where \mathcal{P} contains an exponential number of copula subsets, a solution to this linear optimization problem with an exponential number of variables can be obtained using a technique known as column generation (see [28], [29]). The process consists of the following steps, repeated iteratively.

First, we solve problem (3) with a reduced set of variables λ_V , with V contained in a subset $\hat{\mathcal{P}} \subset \mathcal{P}$. We call this problem with a reduced set of variables the “master” problem. Solving a problem with this reduced set is equivalent to finding a solution to (3) assuming that all $\lambda_V \in \mathcal{P} \setminus \hat{\mathcal{P}}$ have value zero. The solution to this master problem with a reduced set of variables is optimal for the “full” problem if and only if the reduced cost of the all λ_V variables is non-negative, including those λ_V with $V \in \mathcal{P} \setminus \hat{\mathcal{P}}$. Hence, we need to verify if there is a set V with a negative reduced cost. In our second step, we solve a “pricing” problem to find a set $V \subseteq \mathcal{C}$ such that the reduced cost of its variable λ_V is minimized. If the optimal objective value of this pricing problem is non-negative (i.e. all reduced costs are non-negative), then we are finished, and the solution of the master problem is an optimal solution of (3). If not, then there is a V with a negative reduce cost that should be included in the master problem, so we include V in $\hat{\mathcal{P}}$, and iterate solving the master problem and repeating this process.

For the second step, we need to compute the minimal reduced cost of λ_V for an arbitrary set $V \subseteq \mathcal{C}$. The reduced cost associated with variable λ_V is given by

$$\bar{c}_V = - \sum_{i \in V} \mu_i - \sum_{i,j \in V} \nu_{i,j},$$

where μ_i and $\nu_{i,j}$ are the dual variables of equations (3c) and (3b), respectively, obtained by solving the master problem. Hence, to solve our pricing problem, we need to select a subset of links in V that minimizes \bar{c}_V . In general, as stated in the following lemma, this problem can be hard to solve.

Lemma 2. The problem of finding a subset $V \subseteq \mathcal{C}$ that minimizes \bar{c}_V is NP-hard.

Even if the problem of finding a subset V minimizing \bar{c}_V can be difficult, we can formulate an integer programming model to solve this pricing problem as follows:

$$\text{minimize} \quad - \sum_{i \in \mathcal{C}} \mu_i x_i - \sum_{i,j \in \mathcal{C}} \nu_{i,j} y_{i,j} \tag{4a}$$

$$\text{subject to} \tag{4b}$$

$$y_{i,j} \leq x_i \quad \forall i, j \in \mathcal{C}, \tag{4c}$$

$$y_{i,j} \leq x_j \quad \forall i, j \in \mathcal{C}, \tag{4d}$$

$$x_i + x_j \leq y_{i,j} + 1 \quad \forall i, j \in \mathcal{C}, \tag{4e}$$

$$x_i, y_{i,j} \in \{0, 1\} \quad \forall i, j \in \mathcal{C}. \tag{4f}$$

The binary variable x_i indicates whether link i should be included in the copula subset V . The variable $y_{i,j}$ takes value 1 if and only if x_i and x_j take value 1 (i.e., links i and j are included in the copula subset V). The objective function is equal to the reduced cost associated with V . Additional constraints could be included to ensure that $V \in \mathcal{P}$. An optimal value of the objective function smaller than 0 indicates that the new copula subset defined by $\{i \in \mathcal{C} : x_i = 1\}$ should be included in $\hat{\mathcal{P}}$, and resolves the master problem. Note that even if the pricing problem is NP-hard, our implementation shows that optimization software can solve large-scale instances of this pricing problem in a few seconds by virtue of the structure of the pricing problem that is exploited by such solvers. Similar column generation techniques, even with NP-hard pricing problems, are widely utilized to successfully solve large-scale problems in many areas (see [29]).

After solving problem (3), if the optimal solution has an objective value equal to 0, then all t^+ and t^- variables have value 0, and the value of the λ_V variables is a solution to (1) (i.e., the MO copula model attains the required correlations). On the contrary, if the optimal value of problem (3) is greater than 0, then there is no solution to the system of equations (1) for the available copula subsets in \mathcal{P} , and

the optimal solution obtained represents a set of copula parameters that minimizes the L_1 -norm of the difference between the required and obtained values in equation (1a).

Other criteria to the copula subsets in \mathcal{P} could be imposed by including more constraints in the pricing problem. For example, we can limit the size of the copula subset V by adding a cardinality constraint ($\sum_{i \in \mathcal{C}} x_i \leq k$) to problem (4). Additionally, if a solution to equation (1) is obtained, then probably we would have an infinite number of solutions, because the number of variables could be considerably larger than the number of constraints. In that case, supplementary criteria can be imposed to the solution, as we explain in the following subsection.

B. Additional criteria based on empirical failure behaviors in telecommunication networks

As mentioned in the introduction, empirical analyses of networks have shown that correlation is significant for links that are close enough, and although the probability of simultaneous failures in a network decreases as the number of components involved in the simultaneous failure increases, it is still significant. For example, for the data center studied in [5, §4.6], the authors showed that failures involving one link are around 50% of the total and failures that involve between two and four links accounted for 40% of all failures. In [4, §IV.B] the authors showed that the number of simultaneous failure events decreases when the distance between the components involved increases. In [6, §4.2.2] the study showed that simultaneous failures occur mainly in links that share a node, or more precisely, a router. A similar behavior was reported by [30] for IP backbone networks. Therefore, MO copula models satisfying (1) with copula subsets containing nearby components, should replicate the behavior reported in these studies. Moreover, additional arbitrary criteria could be naturally added to the copula model. In the following, we discuss how to modify the optimization model (3) when an optimal solution satisfying (1) has been found, to incorporate additional requirements to the solution.

In a general framework, we can assume that there exists a performance measure $f(\lambda)$ of the desired properties that the set of copula parameters should have. In that case, we can formulate the following optimization problem:

$$\min f(\lambda) \text{ subject to (1).}$$

To solve this exponentially-sized problem, it is nice if $f(\lambda)$ is linear, to maintain the linear structure of the formulation and be able to apply a column generation approach as before. Hence, we assume that each copula subset V has a parameter w_V associated with, such that our interest is to find copula parameters that minimize (or maximize) $\sum_{V \in \mathcal{P}} w_V \lambda_V$. For example, w_V can represent the size of the copula subset, the maximum/minimum/average distance between the components of the copula subset, or other criteria. Therefore, we are now interested in solving the following linear optimization problem:

$$\text{minimize } \sum_{V \in \mathcal{P}} w_V \lambda_V \tag{5a}$$

subject to

$$\sum_{V \in \mathcal{P}: i, j \in V} \lambda_V = \gamma_{i, j} \quad \forall i, j \in \mathcal{C}, i \neq j, \tag{5b}$$

$$\sum_{V \in \mathcal{P}: i \in V} \lambda_V = -\ln(1 - p_i), \quad \forall i \in \mathcal{C}, \tag{5c}$$

$$\lambda_V \geq 0, \quad \forall V \in \mathcal{P}. \tag{5d}$$

As before, we can apply a column generation technique to solve this problem. In this case, the reduced cost of a variable is given by

$$\bar{c}_V = w_V - \sum_{i \in V} \mu_i - \sum_{i,j \in V} \nu_{ij}. \quad (6)$$

It is then important to see which types of parameters w_V can be utilized in such a way that a pricing problem can find a subset V with minimum reduced cost. We discuss next some criteria that can be applied following the preceding remarks and maintaining the structure of the problem. A natural criteria to choose weights w_V could be the size of the copula subset ($w_V = |V|$), or the average distance between components of V . Surprisingly, all these criteria do not let us distinguish better copula parameters, because any solution satisfying (1) will have the same objective value, as stated in the following lemma.

Lemma 3. Given fixed $\alpha \in \mathbb{R}^{|\mathcal{C}|}$ and $\beta \in \mathbb{R}^{|\mathcal{C}|^2}$, let $w_V = \sum_{i \in V} \alpha_i + \sum_{i,j \in V} \beta_{i,j}$. Then, the objective value $\sum_{V \in \mathcal{P}} w_V \lambda_V$ is constant for any feasible solution λ of (1).

Proof. Note that

$$\begin{aligned} \sum_{V \in \mathcal{P}} w_V \lambda_V &= \sum_{V \in \mathcal{P}} \lambda_V \sum_{i \in V} \alpha_i + \sum_{V \in \mathcal{P}} \lambda_V \sum_{i,j \in V} \beta_{i,j} \\ &= \sum_{i \in \mathcal{C}} \alpha_i \sum_{V \in \mathcal{P}: i \in V} \lambda_V + \\ &\quad \sum_{i,j \in \mathcal{C}} \beta_{i,j} \sum_{V \in \mathcal{P}: i,j \in V} \lambda_V. \end{aligned}$$

We conclude by observing that the terms $\sum_{V \in \mathcal{P}: i \in V} \lambda_V$ and $\sum_{V \in \mathcal{P}: i,j \in V} \lambda_V$ have a constant value for any λ satisfying (1), because the p_i and $\gamma_{i,j}$ are given. \square

Therefore, we propose two criteria that allow us to distinguish between the different solutions to (1):

a) *Lasso (Least Absolute Shrinkage and Selection Operator):* A natural criteria that can be applied is to reduce the number of copula subsets V with $\lambda_V > 0$. This cannot be done without introducing binary variables into our master problem, in which case our column generation approach cannot be applied. However, following the ideas of the Lasso regression, we can obtain a similar behavior by minimizing the sum of the L_1 -norm of the variables. That is, this corresponds to the case of $w_V = 1$ for all $V \in \mathcal{P}$.

b) *Diameter:* As mentioned before, the literature suggests that components in a copula subset should be “close” to each other. Due to previous lemma, we cannot use the average distance between elements in the copula subset, but we can use the maximum distance between them. Formally, we set $w_V = \max_{i,j \in V} d_{i,j}$ where $d_{i,j}$ is the distance between components i and j . In order to use this criteria, we need to modify our pricing problem accordingly. We add a new continuous variable z representing the maximum distance between components in V . Hence, our modified objective in the pricing problem (4) is

$$\text{minimize } z - \sum_{i \in \mathcal{C}} \mu_i x_i - \sum_{i,j \in \mathcal{C}} \nu_{ij} y_{ij}$$

and we add the constraint

$$z \geq d_{i,j} y_{i,j} \quad \forall i, j \in \mathcal{C}, i \neq j.$$

These are examples of different criteria that can be implemented. Other criteria could be natural in other contexts, and implemented in a similar way.

IV. COMPUTATIONAL EXPERIMENTS

Recall that our main goal is to obtain a MO copula model that by capturing the correlations between the failures of the components of a network leads to a more accurate evaluation of its s, t -reliability than assuming the standard independence between those components. To evaluate the capacity of our approach to provide a better estimation of the global network reliability in presence of dependencies, we must obviously use an “external” model, typically a “lower level” one, a model that can represent the specific way in which dependencies between the states of the components appear, in a given technological and/or physical context, and that allows to observe and then, *to measure*, both the correlations between those states, and also the chosen global metric that captures the way the system is globally affected by failures. That is, (i) individual (marginal) reliabilities, (ii) correlations, and (iii) system’s reliability must be outputs of this model, that we call the *reference* one. Then, (i) and (ii) are used as inputs to our technique that will evaluate the global system reliability, allowing thus to check that its output is close enough to the values appearing in (iii).

For instance, suppose that dealing with telecommunication networks, we are interested in their reliabilities face to problems associated with errors made by technicians working, say, for energy or water distribution companies, that lead to fiber cuts and thus, simultaneous failures of several logical connections. A simulator allowing to represent all these phenomena and to measure the induced correlations between the states of the components (plus a global system’s reliability value) will allow to see if our technique, that deals with partial information only, that is, the graph, the marginal probabilities of failure and the correlations between them, can produce a good approximation to the chosen global reliability metric.

In the paper, we chose a specific context (failures produced by a seismic event) and used a pretty complex failure model that naturally induces correlations between failures. Using this model, we obtained marginal probabilities of failures, correlations and reference s, t -reliabilities for all possible pairs (s, t) of nodes. Using the first two, we constructed a MO copula model following the procedure described in Section III, and we compared our obtained s, t -reliabilities estimated by means of a standard Monte Carlo procedure with the s, t -reliabilities coming out of the reference failure model.

To test our methodology in a natural correlated failures situation, we used as our reference model the seismic failure one recommended by the US Federal Emergency Management Agency [31]. The details of this failure model and how to compute the corresponding parameters are presented in Appendix A.

We performed the experiment on three network instances called *cross*, *wheel* and *NSFnet*, presented in Figure 1. The figure shows the topology of the network, and the s, t -network reliabilities obtained using the reference failure model, presented as a colored grid where the color of the cell (s, t) corresponds to the reliability between nodes s and t , with a color range going from yellow to red, where yellow color represents reliability 1. Table I resumes the resulting marginal failure probabilities and correlations obtained by means of the reference model. For the reference s, t -reliabilities, the values range from 0.894 to 0.988 for the cross instance, from 0.943 to 0.952 for the wheel instance, and for the NSFnet instance from 0.919 to 0.946. For all these values, the estimation was obtained using 10^5 simulations, which gives a variation coefficient smaller than 1.1×10^{-3} for all instances. The characteristics of the failures induced by the seismic model, presented in Table I, suggest that a MO copula model is suitable to study the network reliabilities. The following experiments confirm this hypothesis, and discuss different criteria that can be used to construct an appropriate MO copula model.

A. Effect of improving the approximation of the correlations

In this part, we show that to obtain good approximations of the reliabilities of the network, it is important to have good approximations of the reference failure correlations. To show this, we solve

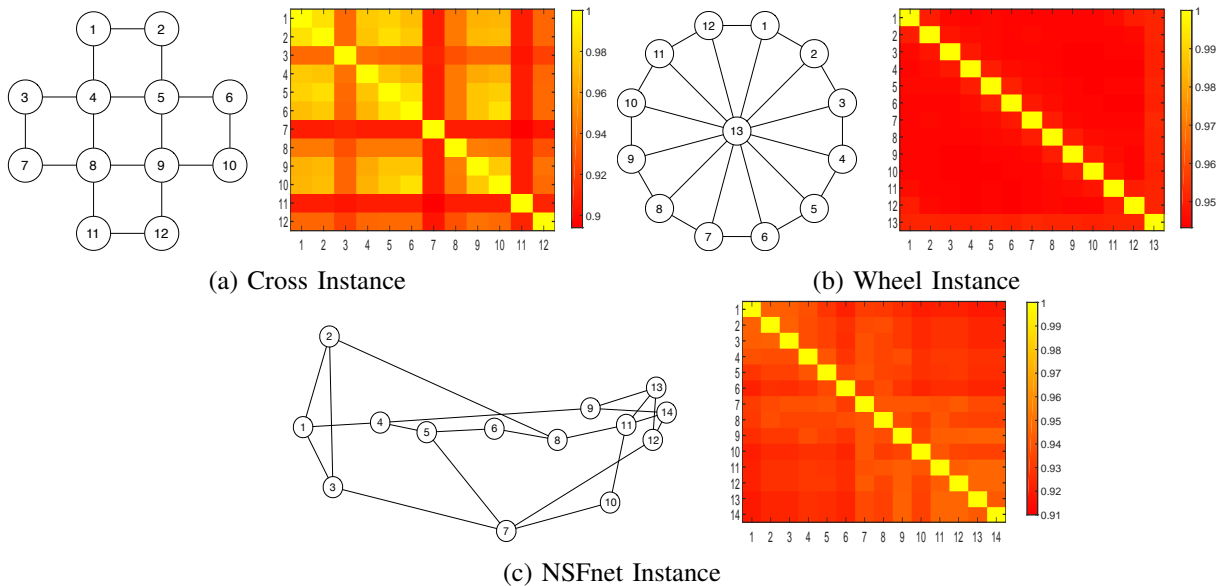


Fig. 1: Network instances for computational experiments and its corresponding s, t -reliabilities obtained from the reference model.

TABLE I: Marginal probabilities and correlation between failures of the reference model for each instance.

Instance	Marginal Prob.			Correlation		
	Min	Avg.	Max	Min	Avg.	Max
Cross	0.045	0.066	0.097	0.660	0.820	0.905
Wheel	0.274	0.278	0.282	0.666	0.706	0.730
NSFnet	0.071	0.077	0.082	0.491	0.609	0.712

model (3) for \mathcal{P} the set of all copula subsets of at most k components. For $k = 1$ (copula subsets considering only a single component), we obtain the classic “Independent model”, that is, the resulting model in which we ignore the correlations among failures. By increasing k , we improve the obtained approximation of the correlations. For a large number k , this can be solved using our column generation approach adding a cardinality constraint on the pricing problem. We solve this model for $k = 2, 4, 8, 16$ and 24 . Note that the cross, wheel and NSFnet instances have 16, 24 and 21 components, respectively. Hence, the case $k = 24$ is only necessary for the wheel and NSFnet instances.

Figure 2 and Table II show the results of these experiments. In the figure, we present the absolute difference between the obtained reliabilities and the original s, t -reliability of the reference model, for each pair of nodes s and t . Additionally, Table II shows the geometric mean absolute difference of the correlations and the reliabilities between the MO copula model and the reference failure model. We also present the relative reliability difference, computed as the geometric mean over all (s, t) pairs of the absolute difference of the s, t -reliabilities divided by the s, t -unreliability of the reference model.

As it can be seen, ignoring the correlation (column INDEPM in Table II) produces a mean absolute reliability error of 0.04 or 0.06, which correspond to a relative error in the computation of the reliability in the range 85% to 98%. Even including a reduced number of copula subsets of small sizes to capture the correlation between failures, allows us to improve considerably the reliability estimation. For the cross instance, using copula subsets of size up to 16 components allows us to replicate the correlations of the reference model, reducing the relative reliability estimation error to 4.05%. For the wheel instance, the

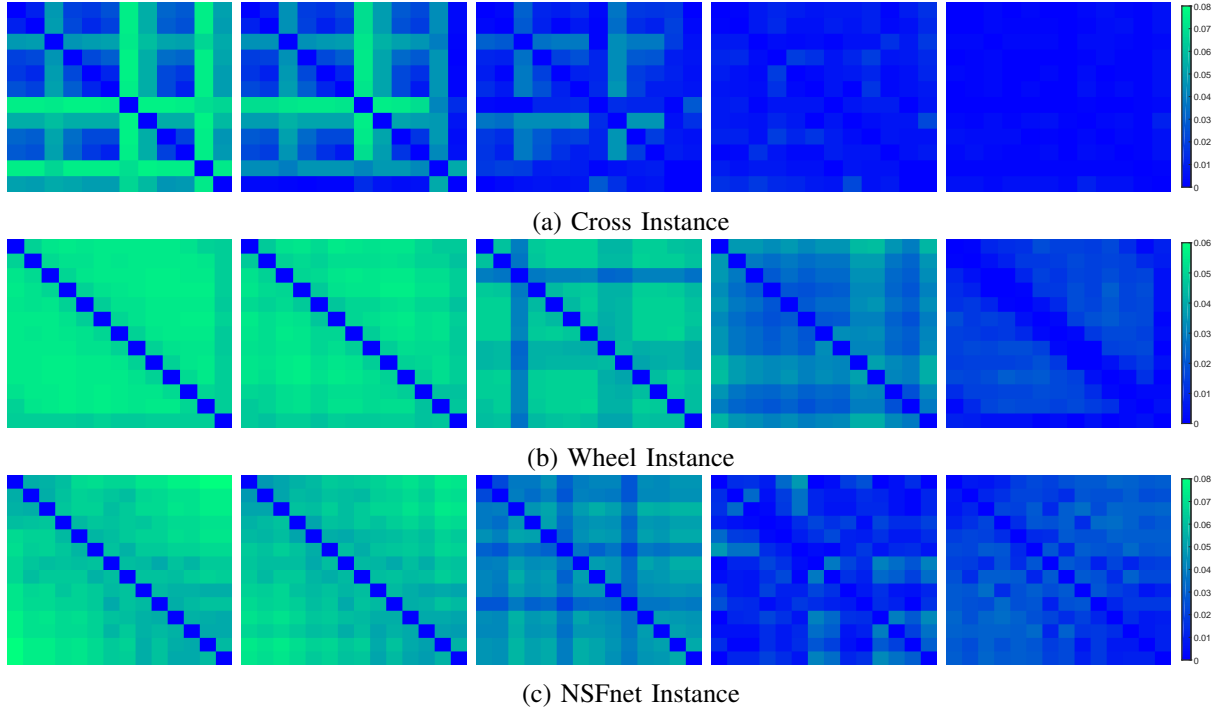


Fig. 2: Absolute reliability difference between the reference model and the MO copula model for Problem (3) for \mathcal{P} including only copula subsets of sizes up to 1, 2, 4, 8 and 16 components (from left to right).

TABLE II: Reliability differences between the reference model and MO copula model from Problem (3) for \mathcal{P} including only copula subsets of sizes up to 1, 2, 4, 8, 16 and 24 components.

Cross instance	INDEPM	2 COMP	4 COMP	8 COMP	16 COMP	
Mean correlation difference	0.4844	0.2227	0.185	0.0038	$< 10^{-6}$	
Mean reliability difference	0.0437	0.0243	0.0119	0.0065	0.0021	
Relative reliability difference	85.31%	47.31%	23.28%	12.77%	4.05%	
Wheel instance	INDEPM	2 COMP	4 COMP	8 COMP	16 COMP	24 COMP
Mean correlation difference	0.5983	0.383	0.3733	0.2501	2.13E-06	$< 10^{-6}$
Mean reliability difference	0.0528	0.0509	0.0423	0.0287	0.0082	0.0040
Relative reliability difference	98.84%	95.41%	79.23%	53.68%	15.39%	7.55%
NSFnet instance	INDEPM	2 COMP	4 COMP	8 COMP	16 COMP	21 COMP
Mean correlation difference	0.6062	0.4102	0.3691	0.1868	$< 10^{-6}$	$< 10^{-6}$
Mean reliability difference	0.0635	0.0604	0.0424	0.0111	0.0208	0.0143
Relative reliability difference	95.22%	90.50%	63.54%	16.68%	31.17%	21.49%

reliability difference do not decrease with the maximum sizes of copula subsets as fast as in the analysis of the cross instance, probably due to the topology of this network which has several components close to each other. Nevertheless, the relative estimation error decreases from value 98.84% when we ignore correlations, to 7.55% when we use all possible copula subsets. As in the previous cases, for the NSFnet the reliability differences also decrease, but the final relative error is higher than the obtained in the other instances. Note that using copula subsets of sizes up to 16 components, the desired correlations are achieved, suggesting that several MO copula models attain the imposed constraints. Hence, adding larger copula subsets will not necessarily improve the reliability differences, unless additional criteria are used

to choose a MO copula model. In fact, the mean relative reliability difference decreases up to a 9.21% when we impose these criteria.

Our experiments clearly confirm our hypothesis, showing that a better approximation of the correlation between failures leads to a better approximation of the s, t -reliabilities. Ignoring the correlation between failures could lead to reliability estimations with relative errors higher than 80%; however, allowing certain dependencies through the MO copula model diminishes these errors considerably.

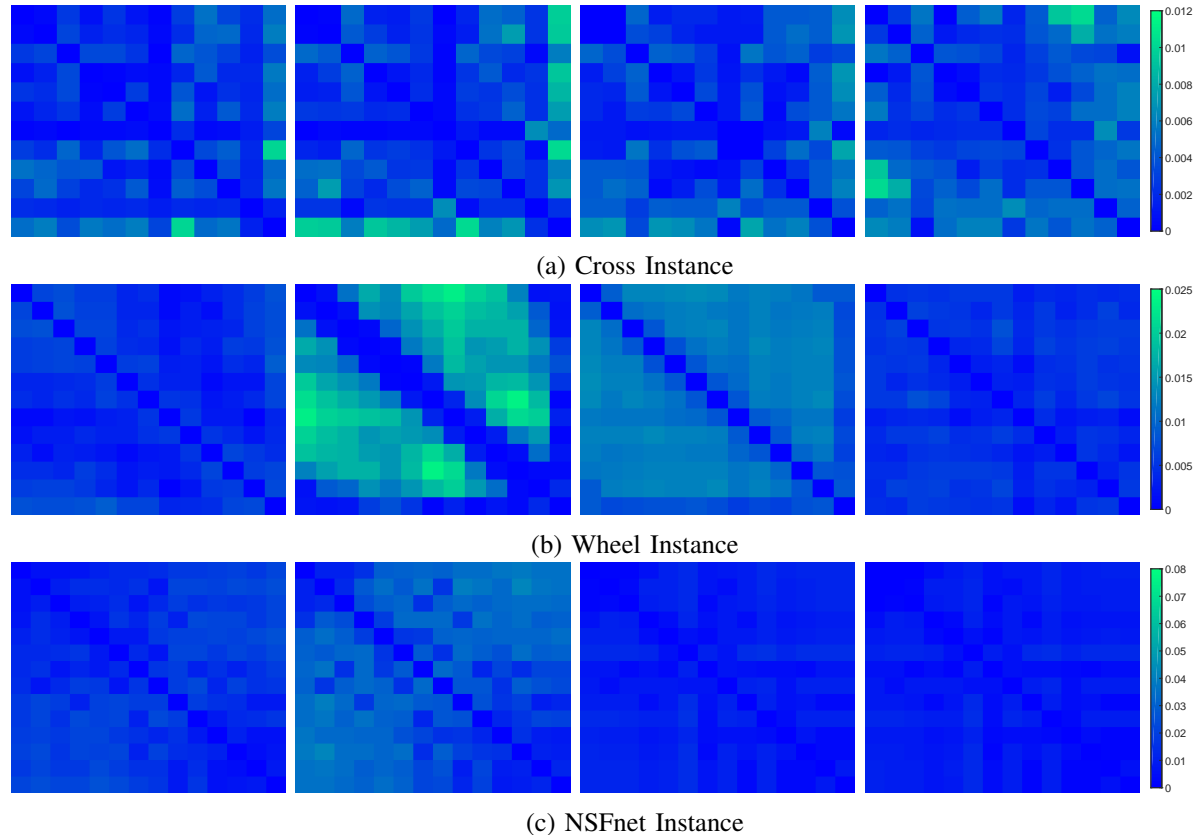


Fig. 3: Absolute reliability difference between the reference model and MO copula model from Problem (3) for $\mathcal{P} = 2^C$ and Problem (5) for $\mathcal{P} = 2^C$ using *Lasso*, *MaxLambda* and *Diameter* criteria (from left to right).

TABLE III: Differences in reliability between the reference model and the other models.

Cross Instance	INDEPM	MODEL (3)	LASSO	MAXLAMBDA	DIAMETER
Mean absolute diff.	0.0437	0.0021	0.0023	0.0024	0.0034
Mean relative diff.	85.31%	4.05%	4.54%	4.75%	6.63%
Wheel instance	INDEPM	MODEL (3)	LASSO	MAXLAMBDA	DIAMETER
Mean absolute diff.	0.0528	0.0040	0.0082	0.0108	0.0045
Mean relative diff.	98.84%	7.55%	15.41%	20.27%	8.43%
NSFnet instance	INDEPM	MODEL (3)	LASSO	MAXLAMBDA	DIAMETER
Mean absolute diff.	0.0635	0.0143	0.0244	0.0078	0.0061
Mean relative diff.	95.22%	21.49%	36.51%	11.73%	9.21%

B. Imposing additional criteria for MO copula

As mentioned before, model (3) allows to obtain copula subsets that replicate the correlation between failures of the reference model. However, in this case, there is an infinite number of combinations of copula parameters that allow to match these correlations. As mentioned in [22], in extreme topologies it is possible to have very different reliabilities for two MO copula models that have the same correlations. Hence, we investigate including additional criteria to the copula subset, as mentioned earlier, to see if they produce differences in the resulting reliabilities. For this purpose, we solve the optimization model (5) using the objective function corresponding to the Lasso and Diameter criteria discussed in Section III-B. Additionally, in contrast with these criteria, we also consider an additional case denoted by MaxLambda, which, at the opposite side of the Lasso approach, maximizes the sum of the copula parameters, instead of minimizing it. As in the previous experiments, we compare the resulting difference in the s, t -reliability using each criteria with the reference case, also including the results obtained by model (3) for the full copula subsets. The results of these experiments are presented in Figure 3 and Table III.

Considering that the *Diameter* criterion agrees with the telecommunication literature on dependent failures discussed in Section III-B, it could be expected that it exhibits a better performance than the other criteria. However, our experiments show that all criteria produce MO copula models that obtain good and similar estimated reliabilities. Even opposite criteria as Lasso and MaxLambda show a similar behavior. These experiments illustrate the fact that, to obtain good approximations of the system reliabilities, the critical issue is to accurately approximate the correlations between failures and not the other criteria discussed in the paper.

V. CONCLUSIONS

As remarked in the literature (see [4]–[8]), ignoring the dependencies between component failures in networks can considerably affect computations of network reliability. Our experiments not only show this fact. They also confirm the hypothesis that approximating the correlation of a reference failure model results in a better approximation of the s, t -reliability of the network. In fact, the relative difference between the reference and estimated reliabilities decreases from $\sim 90\%$ when correlations are ignored, to $\sim 7\%$ when we consider MO copula subsets that correctly capture these dependencies. To find this MO copula model, we proposed an optimization approach that matches the marginal failure probabilities and approximates the correlation of failures in a reference model, that can be solved following a column generation method. We modified the precedent model to incorporate additional criteria, showing that their contribution is minor compared to accurately matching the failure correlations. We think that these findings should remain valid for other definitions of reliability (e.g., reliabilities that consider all-to-all connectivity, or existence of k -disjoint paths), but further experiments should be conducted to confirm this hypothesis. Finally, we expect that these ideas can be applied to other copula models, like the different generalizations of the MO copula model that have arise recently in the literature.

ACKNOWLEDGMENT

The authors acknowledge the financial support from Fondecyt grants 1130681 and 1161064, INRIA-CIRIC and Programa Iniciativa Científica Milenio NC120062.

REFERENCES

- [1] J. S. Provan and M. O. Ball, “The complexity of counting cuts and of computing the probability that a graph is connected,” *SIAM Journal on Computing*, vol. 12, no. 4, pp. 777–788, 1983.
- [2] G. Rubino, “Network reliability evaluation,” in *State-of-the art in performance modeling and simulation*, K. Bagchi and J. Walrand, Eds. Gordon and Breach Books, 1998, pp. 275–302.

- [3] D. P. Kroese, T. Taimre, and Z. I. Botev, *Handbook of Monte Carlo Methods*. John Wiley & Sons, 2013, vol. 706.
- [4] A. J. Gonzalez, B. E. Helvik, J. K. Hellan, and P. Kuusela, "Analysis of dependencies between failures in the UNINETT IP backbone network," in *2010 IEEE 16th Pacific Rim International Symposium on Dependable Computing*. Institute of Electrical and Electronics Engineers, Inc., 2010, pp. 149–156.
- [5] P. Gill, N. Jain, and N. Nagappan, "Understanding network failures in data centers," in *Proceedings of the ACM SIGCOMM 2011 Conference*, vol. 41. Association for Computing Machinery, 2011, pp. 350–361.
- [6] D. Turner, K. Levchenko, A. C. Snoeren, and S. Savage, "California fault lines: understanding the causes and impact of network failures," in *Proceedings of the ACM SIGCOMM 2010 Conference*, vol. 40. Association for Computing Machinery, 2010, pp. 315–326.
- [7] P. Cholda, E. L. Folstad, B. E. Helvik, P. Kuusela, M. Naldi, and I. Norros, "Towards risk-aware communications networking," *Reliability Engineering and System Safety*, vol. 109, pp. 160–174, 2013.
- [8] Z. I. Botev, P. L'Ecuyer, and B. Tuffin, "Dependent failures in highly reliable static networks," in *Proceedings of the 2012 Winter Simulation Conference (WSC)*, Dec 2012, pp. 1–12.
- [9] A. W. Marshall and I. Olkin, "A multivariate exponential distribution," *Journal of the American Statistical Association*, vol. 62, no. 317, pp. 30–44, 1967.
- [10] M. Rausand and H. Arnljot, *System reliability theory: models, statistical methods, and applications*. John Wiley & Sons, 2004, vol. 396.
- [11] T. Yuge, M. Maruyama, and S. Yanagi, "Reliability of a k -out-of- n system with common-cause failures using multivariate exponential distribution," *Procedia Computer Science*, vol. 96, pp. 968–976, 2016.
- [12] I. Bayramoglu and M. Ozkut, "The reliability of coherent systems subjected to Marshall–Olkin type shocks," *IEEE Transactions on Reliability*, vol. 64, no. 1, pp. 435–443, 2015.
- [13] —, "Mean residual life and inactivity time of a coherent system subjected to Marshall–Olkin type shocks," *Journal of Computational and Applied Mathematics*, vol. 298, pp. 190–200, 2016.
- [14] J.-F. Mai and M. Scherer, "Sampling exchangeable and hierarchical Marshall–Olkin distributions," *Communications in Statistics-Theory and Methods*, vol. 42, no. 4, pp. 619–632, 2013.
- [15] Z. Botev, P. L'Ecuyer, R. Simard, and B. Tuffin, "Static network reliability estimation under the Marshall–Olkin copula," *ACM Transactions on Modeling and Computer Simulation*, vol. 26, no. 2, p. 14, 2016.
- [16] J. Barrera, H. Cancela, and E. Moreno, "Topological optimization of reliable networks under dependent failures," *Operations Research Letters*, vol. 32, no. 2, pp. 132–136, 2014.
- [17] U. Cherubini, F. Durante, and S. Mulinacci, *Marshall - Olkin Distributions-Advances in Theory and Applications: Bologna, Italy, October 2013*. Springer, 2015, vol. 141.
- [18] N. D. Singpurwalla, "Dependence in network reliability," in *Proceedings of the 5th International Conference on Information Fusion, FUSION 2002*, M. Morelli and A. J. DeSimone, Eds., vol. 2. International Society of Information Fusion, 2002, pp. 981–985.
- [19] J.-F. Mai and M. Scherer, *Simulating copulas: stochastic models, sampling algorithms and applications*. World Scientific, 2012, vol. 4.
- [20] J. Engel, M. Scherer, and L. Spiegelberg, "One-factor Lévy-Frailty copulas with inhomogeneous trigger rates," in *Soft Methods for Data Science*. Springer, 2017, pp. 205–212.
- [21] J. N. Hagstrom and S. Ross, "Component state dependence and error in reliability computation," 2001, available in <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.20.8460>.
- [22] O. Matus, J. Barrera, E. Moreno, and G. Rubino, "Calibrating a dependent failure model for computing reliabilities on telecommunication networks," in *Proceedings of the Winter Simulation Conference 2016*, 2016, pp. 490–500.
- [23] R. Billinton and W. Li, *Reliability assessment of electric power systems using Monte Carlo methods*. Springer Science & Business Media, 2013.
- [24] M. El-Khadiri and W.-C. Yeh, "An efficient alternative to the exact evaluation of the quickest path flow network reliability problem," *Computers & Operations Research*, vol. 76, pp. 22–32, 2016.
- [25] M. E. Bertinat, H. Cancela, M. F. González, F. Robledo, and P. Romero, "Statistical methods for diameter constrained reliability estimation in rare event scenarios," in *Reliable Networks Design and Modeling (RNDM), 2015 7th International Workshop on*. IEEE, 2015, pp. 243–250.
- [26] U. Cherubini, E. Luciano, and W. Vecchiato, *Copula methods in finance*. John Wiley & Sons, 2004.
- [27] T. Elperin, I. Gertsbakh, and M. Lomonosov, "Estimation of network reliability using graph evolution models," *IEEE Transactions on Reliability*, vol. 40, no. 5, pp. 572–581, 1991.
- [28] G. B. Dantzig and P. Wolfe, "Decomposition principle for linear programs," *Operations Research*, vol. 8, no. 1, pp. 101–111, 1960.
- [29] G. Desaulniers, J. Desrosiers, and M. M. Solomon, Eds., *Column Generation*, ser. GERAD 25th Anniversary Series. Springer Science & Business Media, 2006, vol. 5.
- [30] A. Markopoulou, G. Iannaccone, S. Bhattacharyya, C.-N. Chuah, Y. Ganjali, and C. Diot, "Characterization of failures in an operational IP backbone network," *IEEE/ACM Transactions on Networking (TON)*, vol. 16, no. 4, pp. 749–762, 2008.

- [31] Federal Emergency Management Agency, “Multi-hazard loss estimation methodology: Earthquake model, Hazus – MH MR5: Technical manual,” Washington DC. US, 2015, available in <https://www.hsd.org/?view&did=12760>.
- [32] C. Crouse, “Ground-motion attenuation equations for earthquakes on the cascadia subduction zone,” *Earthquake Spectra*, vol. 7, no. 2, pp. 201–236, 1991.
- [33] S. Espinoza, R. Sacaan, H. Rudnick, A. Poulos, J. C. de la Llera, M. Panteli, P. Mancarella, A. Navarro, and R. Moreno, “Seismic resilience assessment and adaptation of the northern chilean power system,” in *2016 IEEE Power and Energy Society General Meeting (PESGM)*, 2017.
- [34] B. Gutenberg and C. Richter, *Seismicity of the Earth and Associated Phenomena*, 2nd ed. Princeton University Press, 1954.

APPENDIX

PROOFS OF LEMMATA

Proof of Lemma 1. This result is obtained directly by expressing the covariances and marginal failure probabilities in terms of the links’ lifetimes,

$$\begin{aligned}\mathbb{E}(X_i) &= \Pr(T_i > 1) = 1 - p_i \\ &= \exp\left(-\sum_{V \subseteq \mathcal{C}: i \in V} \lambda_V\right), \\ \mathbb{V}(X_i) &= (1 - p_i)p_i, \quad \text{and} \\ \mathbb{E}(X_i X_j) &= \Pr(T_i > 1 \wedge T_j > 1) \\ &= (1 - p_i)(1 - p_j) \exp\left(-\sum_{V \subseteq \mathcal{C}: i, j \in V} \lambda_V\right),\end{aligned}$$

and isolating the corresponding terms. □

Proof of Lemma 2. This can be proved by reduction of MAXIMUM INDEPENDENT SET problem to our problem. Given an arbitrary graph $\hat{G} = (\hat{V}, \hat{E})$, we can construct an instance of our problem with components \hat{V} . Let $\mu_i = 1$ for all $i \in \hat{V}$, $\nu_{i,j} = -|\hat{V}|$ for all $(i, j) \in \hat{E}$ and $\nu_{i,j} = 0$ for all $(i, j) \notin \hat{E}$. Then, a solution to the problem cannot include two adjacent nodes in \hat{G} , because in that case $\bar{c}_V > 0$. Hence, the subset V that minimizes \bar{c}_V corresponds to an independent set of maximum cardinality in \hat{G} . □

REFERENCE MODEL

The reference model is the seismic failure model recommended by the US Federal Emergency Management Agency (FEMA). The model simulate an earthquake depending on several variables (position, depth and magnitude), and using a *fragility curve* of the components it is assigned a failure probability conditional to the earthquake’s variables.

TABLE IV: Parameters of the model used in the reference model

States	Truncated Exponential		Position	Depth	Fragility curve
	Rate	Truncated point			
Cross	0.03	9.5	[0,10]x[0,10]	[24,59]	$1 - 1.001E - \text{PGA}^{2.5}$
Wheel	0.03	9.5	[-30,30]x[-30,30]	[24,59]	$1 - 1.0001E - \text{PGA}^2$
NSFnet	0.5	9.5	[0,80]x[5,55]	[24,59]	$1 - 1.00005E - \text{PGA}^2$

According to FEMA procedure, we simulate an earthquake according to the model described in [32]. This model determines the peak ground acceleration (PGA) to which a component is exposed, given

the position, depth and intensity of the earthquake. This PGA is obtained according to the following equation:

$$\text{PGA} = \exp(6.36 + 1.76M - 2.73 \ln(R + 1.58 \exp(0.608M)) + 0.00916h), \quad (7)$$

where M is the magnitude of the earthquake, R is the distance between the component and the epicenter of the earthquake, and h is the depth of it.

To define the position and the depth of an earthquake, we use a uniform distribution, following the idea of [33]. For the intensity (or magnitude) of the earthquake, we use the model described in [34] in which, a truncated exponential distribution is used simulate this parameter. Finally, this PGA value is evaluated in a fragility curve to assign a probability of failure of the component.

In our experiments, we do 10^7 simulations of earthquakes and the resulting state of each component in the network. The parameters used for *Cross* and *Wheel* network instances are given in Table IV.

Finally, we compute the s, t -reliability of the reference model evaluating for each simulation if it exists a s, t -path in the simulated state of the network after the earthquake. In an analogous way, we empirically obtain the marginal probability of failure for each component and the correlation between them using the simulated scenarios, to provide the input required to construct a MO copula model of correlated failures. Notice that this estimator of the correlation, in some settings, could show a bias. The potential impact of a biased correlation estimation in our proposed approach is out of the scope of this paper.

Omar Matus is a Project Engineer at Universidad Adolfo Ibáñez and a part-time lecturer. He holds a Master of Science degree in Industrial Engineering from Universidad Adolfo Ibáñez. His master's thesis studied dependent failures in telecommunication networks, which is the basis of this paper. His email address is omar.matus@edu.uai.cl.

Javiera Barrera is an Assistant Professor with the Faculty of Engineering and Sciences, Universidad Adolfo Ibáñez, Chile. She received her Ph.D. from Université René Descartes-Paris 5 and Universidad de Chile in 2006. Her research interests lie in the asymptotic behavior of stochastic models and, more recently, in stochastic optimization. Applications in telecommunications, computer science, physics and finance motivate the models addressed in her research. Her email address is javiera.barrera@uai.cl.

Eduardo Moreno is an Associate Professor with the Faculty of Engineering and Sciences, Universidad Adolfo Ibáñez, Chile. He received his Ph.D. from Université Paris-Est Marne-la-Vallée and from Universidad de Chile in 2005. His research focuses on combinatorial optimization, algorithms and computational methods, with applications in network design, telecommunications, transportation and mine planning. His email address is eduardo.moreno@uai.cl.

Gerardo Rubino is a senior researcher at INRIA (the French National Institute for Research in Computer Science and Control), where he is the leader of the DIONYSOS (Dependability, Interoperability and perFormaNce anAlYsiS of netwOrKS) team. His research interests lie in the quantitative analysis of computer and communication systems, mainly using probabilistic models. He also works on the quantitative evaluation of perceptual quality of multimedia communications over the Internet. With Bruno Tuffin, he co-edited the book *Rare event simulation using Monte Carlo methods*, published by John Wiley & Sons in 2009, and co-authored several of its chapters. He has published more than two hundred papers in journals and conference proceedings, in several fields of applied mathematics and computer science, and has performed various editorial tasks and managing activities in research. He is a member of the IFIP WG 7.3. His email address is gerardo.rubino@inria.fr.