



HAL
open science

A multi-perspective framework of analysis of critical infrastructures with respect to supply service, controllability and topology

Fangyuan Han, Enrico Zio

► **To cite this version:**

Fangyuan Han, Enrico Zio. A multi-perspective framework of analysis of critical infrastructures with respect to supply service, controllability and topology. International journal of critical infrastructure protection, 2019, 24, pp.1-13. 10.1016/j.ijcip.2018.10.009 . hal-01988950

HAL Id: hal-01988950

<https://hal.science/hal-01988950>

Submitted on 8 Feb 2019

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.

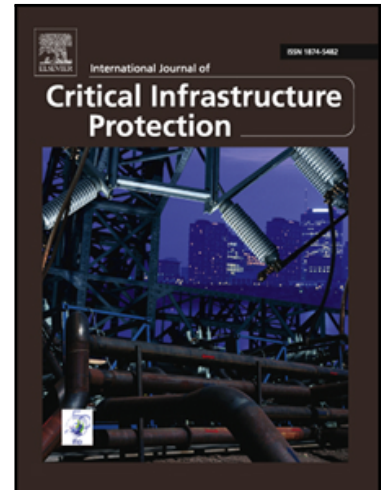
L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Accepted Manuscript

A multi-perspective framework of analysis of critical infrastructures with respect to supply service, controllability and topology

Fangyuan Han, Enrico Zio

PII: S1874-5482(17)30106-3
DOI: <https://doi.org/10.1016/j.ijcip.2018.10.009>
Reference: IJCIP 278



To appear in: *International Journal of Critical Infrastructure Protection*

Received date: 13 June 2017
Revised date: 10 August 2018
Accepted date: 21 October 2018

Please cite this article as: Fangyuan Han, Enrico Zio, A multi-perspective framework of analysis of critical infrastructures with respect to supply service, controllability and topology, *International Journal of Critical Infrastructure Protection* (2018), doi: <https://doi.org/10.1016/j.ijcip.2018.10.009>

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

A multi-perspective framework of analysis of critical
infrastructures with respect to supply service, controllability and
topology

Fangyuan Han¹ and Enrico Zio^{1,2}

¹Chair on Systems Science and the Energetic Challenge, Fondation EDF,
CentraleSupélec, France

²Department of Energy, Politecnico di Milano, Italy.

Abstract

In this work, we propose a multi-perspective framework of analysis of critical infrastructures (CIs) with respect to supply service, topology and controllability. The framework enables identifying the role of CI elements and quantifying the consequences of scenarios of multiple failures, with respect to the different perspectives considered. To present the analysis framework, a benchmark network representative of a real gas transmission network across several countries of the European Union (EU) is considered. The information extracted from such analysis can help us to identify the critical elements and how the properties of the network are affected by failures, and to propose corresponding improvements for CIs. The findings of this paper demonstrate the interest of considering several perspectives in the analysis of CIs for providing useful information for ensuring their safe and reliable operation.

Keywords: Critical infrastructures, Multi-perspective analysis, Complex networks, Supply, Controllability, Gas transmission network.

1 Introduction

Critical infrastructures (CIs), like power grids or gas transmission and distribution systems, rail and road transport or communication networks, are essential to the operation of modern society [1]. They need to be designed, maintained and protected to provide optimal performance, reliable operation and functional safety for long periods of time [2,3]. Hence, the great attention and priority given to the “care” of these systems by the EU, US and other national and transnational administrations [4–7], which calls for the risk assessment and resilience evaluation of CIs [8,9].

As CIs evolve and rely on information technologies more intensively, it is essential to understand their controllability and it is desirable to develop a control framework able to steer the network dynamics toward states with optimal performance, while avoiding undesired or unfavorable states. Control is a fundamental property for the safe and reliable operation of CIs, under a general control perspective, system safety can be framed as a control “problem” [10,11], whereby, accidents result from inadequate control or insufficient enforcement of safety-related constraints on the development, design, and operation of the system, leading to their violation and subsequently to accidents. According to Control Theory, a dynamical system is controllable if, by a suitable choice of inputs, it can be driven from any initial state to any desired final state within finite time [12,13]. However, the control of the complex network systems that make up a CI remains a challenging problem. Studying the controllability of complex networks requires an integration of classical control theory and network theory. In this perspective, the notion of structural controllability has been introduced in [14]. In [13], analytical tools have been developed to characterize the controllability of directed networks. In [15], an exact controllability measure has been proposed to generalize the determination of the set of driver nodes to arbitrary network structures and link weights. Several related topics can be considered under this framework, such as control centrality [16], achieving whole control by using only one controller [17], minimization of control inputs [18], control capacity [19], control mode [20], control of edge dynamics [21], structural controllability of temporal networks [22], control energy [23], etc.

Supply performance is the fundamental functional requirement of a CI and the security of supply is being addressed by an increasing number of researchers. Paper [24] presents a probabilistic model to study the security of supply in a gas network. The model is developed into a Monte Carlo simulation and graph-based tool aiming at the evaluation of CIs for different purposes, including reliability, vulnerability, bottleneck analysis, etc.

The fact that CIs are complex networks of interacting components raises the interest in

studying their topological characteristics [25–29]. A number of recent studies have proposed various measures to evaluate the structural properties of networks and addressed topological investigations to identify critical elements. Among these measures, topological centrality (including degree, closeness, betweenness and information centrality) [30, 31] and network efficiency [32] are two important and classical measures, quantifying the importance of individual network nodes and evaluating the connectivity of the whole network, respectively. Topological properties have also been studied in relation to vulnerability and risk analysis. For example, in [33] the authors analyze the structural vulnerability of the Italian GRTN power grid. In [34], electric power delivery networks are modelled as graphs and their topological characteristics are studied. In [35], centrality analysis is applied to identify the most important components of a railway infrastructure. Given the relationship between the topology of a network and its vulnerability and safety properties, the association between network topological features and system reliability is also of relevance. A common measure of network reliability is the so called K -terminal reliability, which calculates the probability that every two nodes in a specific subset of K nodes are connected by a path of operational edges [36]. Specifically, due to the requirement of reliable operation and the complex nature of CIs, the all-terminal reliability is of particular interest and often considered as a necessary condition for function-based reliability.

The complexity of CIs calls for approaches capable of viewing the problem from multiple perspectives [37–39]. System analysis, reliability engineering, graph theory and others have been propounded to study the behavior and performance of complex systems, also with respect to failure events, their protection and resilience [27, 40, 41]. Integration of the different perspectives and analysis of their relations is necessary. For example, in [42] an electrical transmission system is analyzed with the objective of identifying the most critical elements in terms of four different perspectives: topological, reliability, electrical and electrical-reliability. In [43], the correlation between connectivity reliability and controllability of network systems is studied. In [41], the authors perform network reliability analysis considering spatial constraints.

The context described above shows that the design and operation of CIs are multi-objective problems, in which the multiple objectives account for different functionality perspectives. Several multi-objective optimization methods have been developed, including Evolutionary Algorithms (EAs), which have proved very efficient thanks to their powerful meta-heuristics search ability and population-based framework. References [44–47] solve the optimization problems of network design considering all-terminal reliability and cost

as objectives. Reference [48] considers the optimal expansion of a power transmission network with respect to the transmission reliability efficiency and the cost. References [49,50] solve the multi-objective optimization of cascade-resilient design of electrical power infrastructures. Reference [51] treats a three-objective optimization of economic cost, hydraulic reliability and greenhouse gas emissions, and analyzes the nature of the tradeoffs among the objectives for achieving different solutions. Multi-objective evolutionary algorithms (MOEAs), including Strength Pareto Evolutionary Algorithm (SPEA) [52], SPEA2 [53], non-dominated sorting genetic algorithm (NSGA) [54], (NSGA-II) [55], have been successfully used to solve problems with two or three objectives: however, when encountering optimization problems with more than three objectives, so called many-objective problems, significant algorithmic difficulties and computational challenges arise [56,57], such as: (1) the increase in the number of objectives causes a large proportion of non-dominated individuals in a population and, thus, leads to the ineffectiveness of Pareto optimality: algorithms based on the ϵ -domination principle [58,59] have been proposed to address this problem; (2) the extremely large size of the search space of the optimization problem weakens the effects of the evolutionary operators of crossover and mutation: reference-point based algorithm NSGA-III [60] and decomposition methods such as MOEA/D [61] have been introduced; (3) the conflict between convergence and diversity of solutions is aggravated as the number of objectives increases. Given the level of integration and the number of functionalities of today's systems and CIs, many-objective optimization is attracting a lot of attentions for real-world applications [62–64], but approaches capable of simultaneously addressing all the related challenges mentioned above are still needed [57].

In this work, we develop a framework of analysis considering several perspectives (supply service, controllability and topology). Compared to previous works, which typically consider reliability and topology only, we include the control perspective into the safety and reliability analysis of CIs. The analysis is performed by simulation and the failure scenarios are generated by the software ProGasNetwork proposed in [65]. The complex network representative of a real EU gas transmission system supplying several countries is considered as case study to illustrate the analysis framework. The main contributions of this work are:

- Development of a multi-objective framework of analysis of CIs.
- Identification of the role of each component and quantification of the consequences of multiple failure scenarios, with respect to the different perspectives considered.

- Proposals for CIs reliable performance improvement.

The rest of the paper is organized as follows: Section 2 introduces the three system-level indexes considered in this paper. Section 3 describes the modelling of the considered gas transmission network. Section 4 presents the analysis of link importance and Section 5 presents the consequences of the failure scenarios from the three perspectives. Finally, conclusions and ideas for future work are provided in Section 6.

2 System-level indexes

In this work, we consider three perspectives of the CI assessment: supply, control and topology. For each perspective, we propose an index to evaluate the network performance.

2.1 Non-supplied demand

Supply performance is the fundamental functional requirement of a CI. Consider a CI network of N nodes and L links, which supplies service or products from N_s supply nodes (sources) to N_y user nodes (users) through a number of transmission nodes.

We introduce the non-supplied demand (NSD) as a measure of the network's capacity to satisfy its users' demands. The normalized NSD is introduced as a system-level index:

$$NSD = 1 - \frac{\sum_{i=1}^{N_y} \omega_i y_i}{\sum_{i=1}^{N_y} \omega_i D_i} \quad (1)$$

where ω_i is the weight of the i^{th} of the N_y users and varies according to its role or importance. y_i is the supply to user i and D_i is its demand, which is the target supply to user i and is considered as a constraint when determining the supply to it; and, thus, $y_i \leq D_i$. Then, the second term in Equation (1) represents the satisfied proportion of users' demand. The index NSD is normalized to take values in $[0, 1]$. NSD equals 0 when the users' demands are fully supplied.

2.2 Controllability

In control theory, a system is defined controllable if, by a suitable choice of inputs, it can be driven from any initial state to any desired final state within finite time [12, 13]. From a system safety perspective, controllability is the ability to guide the system's behavior towards a safe state through the appropriate choice of a few input variables [10].

Considering the network of N nodes, we describe its state dynamics by:

$$\mathbf{x}(t+1) = \mathbf{A}\mathbf{x}(t) + \mathbf{B}\mathbf{u}(t) \quad (2)$$

where $\mathbf{x}(t)$ is the system state vector, describing the state of each node in the network at time t ; \mathbf{A} is an $N \times N$ coupling matrix, in which a_{ij} represents the weight of the directed link from node i to node j (i.e. the interaction strength between node i and node j , for example, the flow in the pipeline of a gas transmission network); \mathbf{B} is an $N \times M$ input matrix ($M \leq N$), identifying the nodes that are controlled by the time-dependent input vector $\mathbf{u}(t)$, made of M independent control signals.

Based on dynamic control theory, the above system is controllable if and only if the $N \times NM$ controllability matrix $\mathbf{Ctrb} = (\mathbf{B} \ \mathbf{A}\mathbf{B} \ \dots \ \mathbf{A}^{n-1}\mathbf{B})$ has full rank (the so-called Kalman's rank condition) [66]:

$$\text{rank}(\mathbf{Ctrb}) = N \quad (3)$$

For complex network systems, the controllability problem can be formulated in terms of finding a suitable control matrix \mathbf{B} consisting of a minimum number of driver nodes (N_D) so as to satisfy the Kalman's rank condition (3). However, this requires the evaluation of the rank of \mathbf{C} for 2^N possible combinations of the driver nodes [67]: for real CI network systems, such a brute-force search is computationally prohibitive.

To overcome this problem, in [13], the authors have developed analytical methods to determine the minimum number of driver nodes (N_D) that are needed to fully control the network, by finding the maximum matching, i.e. the maximum set of links that do not share start or end nodes. Full control can be achieved if and only if each unmatched node is directly controlled and there are directed paths from the input signals to all matched nodes. The unmatched nodes determined by maximum matching are the so called driver nodes.

In [15], the exact controllability for arbitrary network structures and link weights (say arbitrary matrix \mathbf{A}) is introduced to calculate N_D :

$$N_D = \max_i \{\mu(\lambda_i)\} \quad (4)$$

and the minimum number of driver nodes N_D is determined by the maximum geometric multiplicity $\mu(\lambda_i)$ of the eigenvalue λ_i of \mathbf{A} . In fact, these are the nodes corresponding to the linearly-dependent rows: the controllers should be imposed on the linearly-dependent rows to eliminate all linear correlations and ensure the controllability condition.

The number of driver nodes N_D can be taken as a measure of the controllability of the network, indicating how many driver nodes are needed to control the network and directly relating to the cost of the resources needed to keep or bring the system under control. If $N_D = N$, i.e. the total number of nodes in the network (this means that the external control signal is applied to each node of the network), the likelihood of gaining full system control is high, but so is the associated cost [18]. A small N_D , instead, indicates a more controllable network system, in the sense that it requires less effort to obtain full control over the network.

To measure the structural controllability of the network system, we adopt the controllability index (C_{ind}) first introduced in [43]:

$$C_{ind} = \frac{N - N_D}{N} \quad (5)$$

Also, the index C_{ind} is normalized to take values in $[0, 1]$.

The occurrence of failures (represented as the removal of links) is likely to increase the number of the linearly-dependent rows in matrix \mathbf{A} and, thus, N_D would increase and C_{ind} decreases; when the current number of control nodes is insufficient to obtain full control over the whole system, there is no guarantee that the system can be brought back to the designed operation condition.

2.3 Network topological efficiency

Network topological efficiency is a measure of the connectivity of the whole network, i.e. of how well the nodes of a network exchange flow [32]. This measure is based on the assumption that the flow in a network travels along the shortest routes. In the case of gas transmission networks, a typical priority supply pattern, especially during gas crises, is distance-based [24]: nodes geographically closer to the gas source are served first, since gas pressure drops at the points far from the supply source: where the distance from the source is shorter, the demand is more likely to be served, thus, the efficiency of gas transmission is higher. Then, the efficiency in the communication between two nodes i and j , ε_{ij} , is inversely proportional to their shortest path length d_{ij} ; this latter is defined as the smallest sum of *physical distances* throughout all the possible paths in the weighted network: $\varepsilon_{ij} = 1/d_{ij}$. When there is no path between i and j , $d_{ij} = +\infty$, i.e. $\varepsilon_{ij} = 0$. The topological efficiency of the whole network is given by:

$$E[\mathbf{G}] = \frac{\sum_{i \neq j \in \mathbf{G}} \varepsilon_{ij}}{N(N-1)} = \frac{1}{N(N-1)} \sum_{i \neq j \in \mathbf{G}} \frac{1}{d_{ij}} \quad (6)$$

3 Case study

3.1 Network description and graph representation

We consider the case study from [65]. The system is visualized in Figure 1 and represents a real gas transmission network for supply across several countries in the EU. The gas transmission network includes 56 nodes and 74 links, where nodes represent sources or substations and links represent the gas transmission pipelines connecting the nodes. Among the 74 links, 10 links are virtual links representing the virtual connection of parallel pipelines, and their failure is not considered.

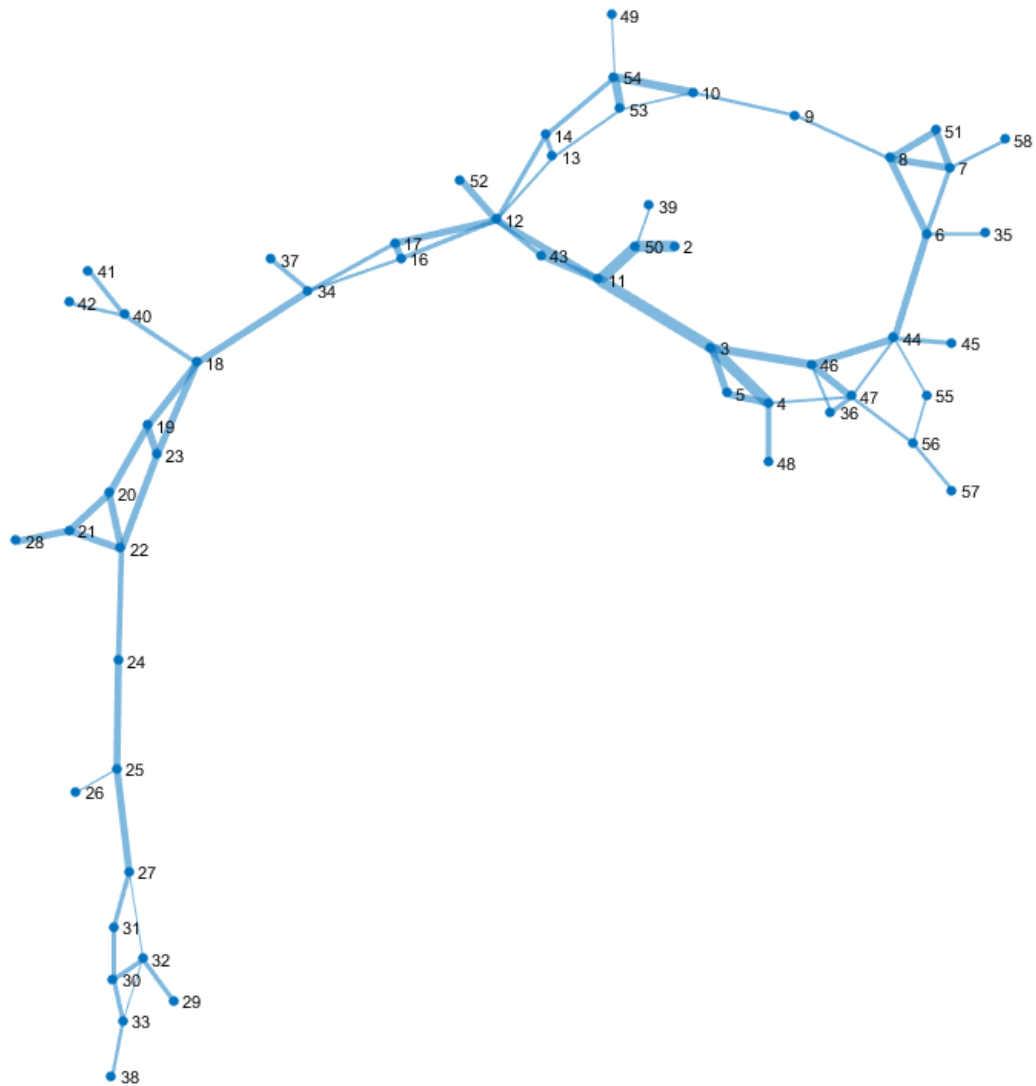


Figure 1: Gas transmission network [65]

The gas transmission network is modeled as an undirected graph \mathbf{G} . Its connectivity

structure can be defined by its $N \times N$ adjacency matrix \mathbf{Adj} , whose entries $[Adj_{ij}]$ are equal to 1 if there exists a link joining node i to node j and 0 otherwise.

Each link in the network is characterized by its capacity, i.e the maximum amount of flow that it is able to supply, and its length. The capacity matrix \mathbf{K} contains information about the capacity constraints of the network elements including source nodes, demand nodes and pipeline capacities. The length matrix \mathbf{Len} contains the lengths of the edges between nodes: entry Len_{ij} is the length of the pipeline connecting the i -th and j -th nodes; an entry of 0 indicates that the i -th and j -th nodes are not connected.

We distinguish: $N_y = 35$ demand nodes with deterministic daily demands for a total system daily demand of 45.9 millions of cubic meters (mcm), one LNG terminal (node 10), two compressor stations (nodes 11 and 12), two storage devices (nodes 18 and 19) and two pipeline source nodes (nodes 2 and 29).

The properties of the four nodes considered as supply nodes (sources), numbered 2, 10, 19 and 29, are presented in Table 1. The properties of the 35 demand nodes are shown in Table 2. The capacities and demands are expressed in millions of cubic meters per day (mcm/d). The data of supply and demand are realistic and they are expressed at a daily scale, in order to assume peak gas demand during one peak day (in winter) with extreme high gas demand [24]. These data are intended to represent the most stressed conditions for the gas transmission network. Depending on the purpose of the analysis, variable values of demand or supply can be considered, and the user demands satisfaction can be evaluated as the average over a simulation horizon, with associated uncertainty.

Given the capacities of the links connecting the nodes and the constraints on the sources and users, the supply to each user is used to calculate the non-supplied demand NSD defined in Equation 1 by maximum flow algorithms [24]. In the case study, for illustration purposes and with no loss of generality, we assume that all users have the same importance and ω_i is equal for the user nodes.

Node	Capacity	Type
2	31	Pipeline source
10	10.5	LNG terminal
19	25	Underground storage
29	4	Pipeline source

Table 1: Sources properties

Node	Demand	Node	Demand	Node	Demand
4	0.1	27	3.0	44	0.7
5	3.2	28	6.0	45	1.3
6	0.1	30	0.5	47	0.1
8	0.1	33	0.5	48	1.8
9	0.1	34	0.5	49	0.2
10	1.0	35	0.1	51	7.0
13	0.5	36	4.2	52	0.6
17	0.1	37	1.3	53	0.1
18	8.5	39	0.3	55	0.2
20	0.6	41	0.6	57	0.2
25	0.5	42	0.6	58	0.3
26	0.8	43	0.2		

Table 2: Demands of the 35 users

3.2 Failure modelling

We consider the failure of the LNG station, compressor stations, storages and 64 pipelines.

The failure of the LNG terminal and of the storage devices is modeled as the total capacity loss of each pipeline connected to it. According to [65], the monthly failure frequency of the LNG (node 10) is set to $f_{LNG} = 1.25E - 2$, and the monthly failure frequency of the storage (node 19) is $f_S = 8.33E - 3$.

In case of a compressor station failure, the capacity of each pipeline connected to it will reduce by 20%. The monthly failure frequency of the two compressor stations (nodes 11 and 12) are $f_{CS} = 2.08E - 2$.

According to the EGIG report [68], the average failure frequency of a European gas transmission pipeline is $3.5E - 4$ per kilometer-year. We consider the total rupture of a pipeline and we assume that 10% of the failures reported in a year cause such a rupture. Thus, the monthly failure frequency of a pipeline is $f_P = 2.92E - 6$ per kilometer [65].

4 Analysis of link importance

We focus on the importance of a link in terms of its influence on the three system properties considered.

Table 3 presents the values of the three indexes introduced in Section 2, calculated for the nominal network configuration \mathbf{G} . For the analysis of link importance, we systematically

NSD	C_{ind}	E
0	0.9107	0.6327

Table 3: Indexes values for the nominal configuration

disconnect one link at a time from the original network to obtain and compute the indexes of the new network configuration reached, \mathbf{G}' . We identify the most important nodes in terms of NSD , C_{ind} and E , respectively. Table 4 presents the ten most critical links in terms of NSD , the three most critical links in terms of C_{ind} , and the single most critical link in terms of E .

Link	NSD	C_{ind}	E
3-11	0.363	0.9107	0.6319
3-46	0.209	0.9107	0.6327
21-28	0.131	0.8929	0.6327
2-50	0.126	0.9107	0.6325
11-50	0.120	0.9107	0.6323
6-44	0.106	0.9107	0.6321
44-46	0.081	0.9107	0.6327
36-47	0.048	0.8929	0.6327
4-48	0.039	0.8929	0.6318
34-37	0.028	0.8929	0.6327
44-45	0.028	0.9107	0.6318
18-40	0.026	0.9107	0.6325
6-35	0.002	0.8750	0.6325
11-43	0	0.8750	0.6326
18-23	0	0.8750	0.6327
18-34	0	0.9107	0.6318

Table 4: Indexes values associated to the removal of the most critical links

With the removal of single links, the NSD value ranges from 0 to 0.363. Pipelines represented by links 3-11 and 3-46 are of large capacity, so they are essential to supplying the demand nodes in their neighborhood, and thus their importances are significant in terms of supply. A similar explanation applies for the removal of links 6-44 and 44-46.

Node 28 is a large demand node, and therefore, the removal of link 21-28, which is its only connection to the network, will affect the overall network NSD . The same explanation also applies for the impact of links 4-48, 34-37 and 44-45. Link 4-48 and link 44-45 are

critical in terms of efficiency E , because their removal disconnects the end nodes 48 and 45, respectively and, thus, decreases the network efficiency; moreover, considering that they are relatively short pipelines, the value of E drops much more than for the removal of links 34-37 and 21-28, which are long pipelines.

Links 2-50 and 11-50 connect the main source (node 2) to the rest of the network, and their removal leads to a deficit in supply capacity, since the remaining sources 10, 19 and 29 are not capable of fully supplying the total demands.

Link 18-34 is also a critical link in terms of topological efficiency: when it fails, the network will break into two separate parts and no gas flow can be transferred between them, so that, the topological efficiency E drops.

To rigorously quantify the importance of a link, we compute its Risk Achievement (RA) metric [69] with respect to NSD , C_{ind} and E :

$$IM_{ij}^{NSD} = NSD[\mathbf{G}(base)] - NSD[\mathbf{G}'(x_{ij} = 1)] \quad (7)$$

$$\begin{aligned} IM_{ij}^{C_{ind}} &= C_{ind}[\mathbf{G}(base)] - C_{ind}[\mathbf{G}'(x_{ij} = 1)] \\ &= \frac{N_D[\mathbf{G}'(x_i = 1)] - N_D[\mathbf{G}(base)]}{N} \end{aligned} \quad (8)$$

$$IM_{ij}^E = E[\mathbf{G}(base)] - E[\mathbf{G}'(x_{ij} = 1)] \quad (9)$$

where $\mathbf{G}'(x_{ij} = 1)$ is the graph of the network obtained by removing the link $i - j$ from the original network $\mathbf{G}(base)$.

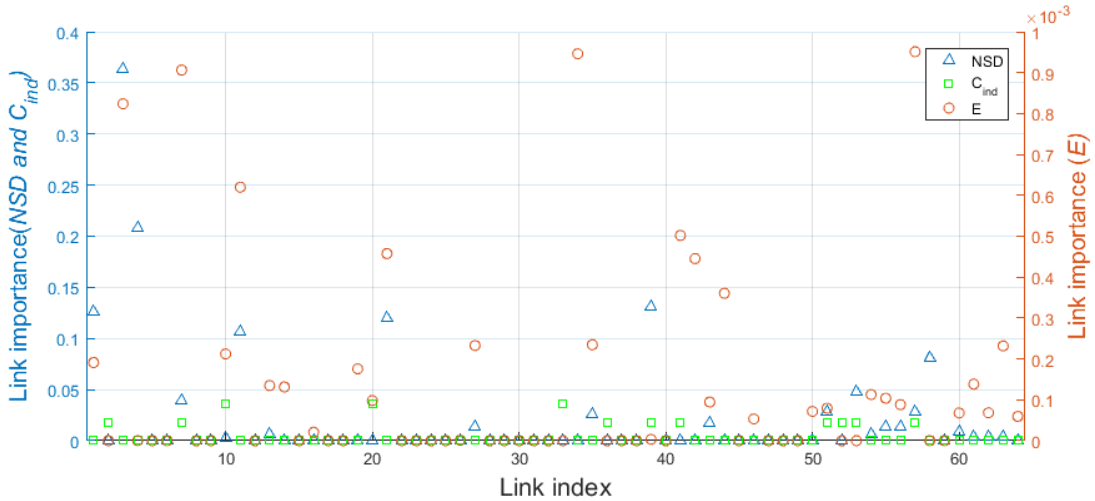


Figure 2: Link importance in terms of NSD , C_{ind} and E

Figure 2 shows the links importance values in terms of the three indexes: the left vertical axis is the values for NSD (triangles) and C_{ind} (squares), while the right vertical axis is

the topological efficiency E (circles). It is seen that NSD presents the largest range, while a single link disconnection has little impact on controllability and E has the smallest range.

Among the 64 links, only 23 links have impact on NSD and most of them also have an impact on E , but only five links influence the controllability index.

The influence of a link is not the same for the three perspectives, which confirms the need of a multi-perspective framework of analysis.

5 Simulation and analysis

We have run 10^6 dynamic simulations by ProGasNet [65], sampling nodes and links according to their occurrence probabilities, as introduced in Section 3.2. A total of 335 different gas transmission states (cases) emerge from the sampled configurations. The most frequent state sample is the one with no link failures.

We classify the 334 failure cases into different categories by their combination of failures. We consider the thirty most frequent states and investigate how these affect the three system-level indexes considered. For each of the indexes, we quantify their consequences and analyze the impacts of different types of failures.

5.1 Classification by failure types

Both links and nodes of the gas transmission network may fail and multiple failures may occur. In order to understand the influence of different types of failures and of their combinations, we classify the 334 failure cases into seven classes as:

- Single link failure (SL)
- Single node failure (SN)
- Single link failure and single node failure (SL-SN)
- Single link failure and multiple node failures (SL-MN)
- Multiple link failures (ML)
- Multiple link failures and single node failure (ML-SN)
- Multiple node failures (MN)

Single node failure (SN) includes 4 cases (only four nodes may fail according to our definition), but they cover 83.23% of the failure configurations. Single link failure (SL)

includes 61 cases and is the second most frequent class. The cases of SL-MN, ML and ML-SN only occur once in all simulations performed. Figure 3 shows the number of cases for each class and their counts.

As we are analyzing an existing gas transmission network, it is reasonable to have low probabilities for multiple failures scenarios; however, the failures of low frequency of occurrence may have an important impact on the properties of the system, which is one of our interests in this study. Therefore, although the probability of their occurrence is small, it is interesting to consider such multiple-element failures and understand the corresponding consequences, which provides additional information for CIs design.

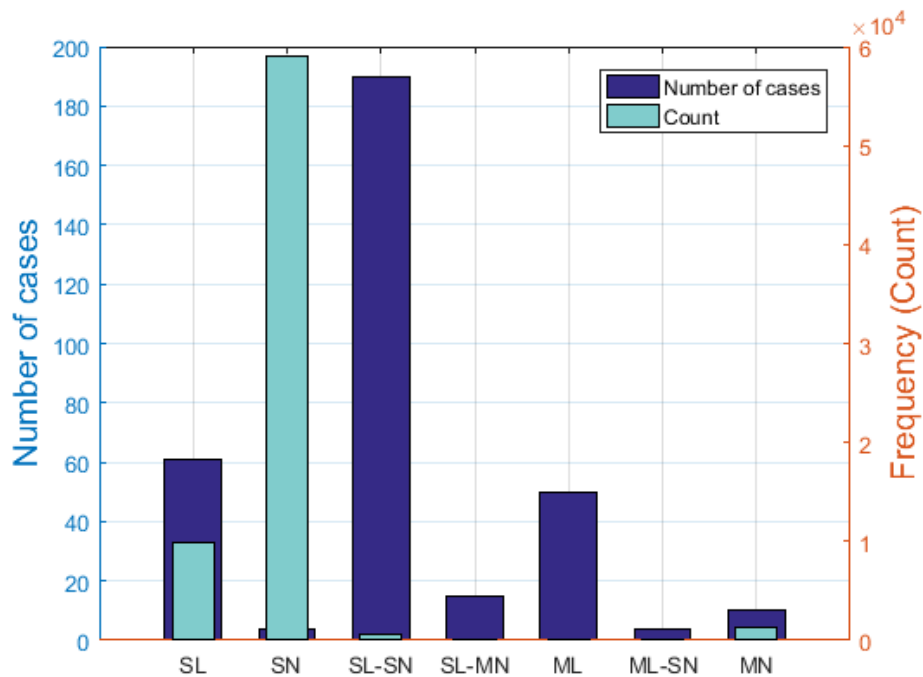


Figure 3: Histogram: number of cases and frequency

5.2 30 most frequent cases

We consider the 30 most frequent cases and apply the analysis framework. Table 5 summarizes the failure types and frequencies of the 30 most frequent cases. Node failure is the most common failure type, the four most frequent failure cases being the four single node failure (SN) cases.

Type	Cases	Frequency (over 10^6 simulations)
Failure free	1	929 013
SN	4	59 040
SL	21	6 238
MN	4	1 098

Table 5: 30 most frequent cases

5.3 The three indexes

We analyze the three indexes separately, with the objectives of identifying the failures affecting each index, quantifying their consequences in terms of loss in the properties considered and calculating their frequency.

5.3.1 Non-supplied demand

Figure 4 shows the NSD index value for the 335 cases, where the abscissa axis is the frequency rank of the 335 cases. The non-supplied demand NSD ranges from 0 to 0.64. In the presence of multiple failures, the network may reach a much higher level of non-supplied demand NSD . The highest value 0.64 corresponds to the SN-SL case where both node 19 and link 2-50 are failed. Node 19 is the second largest source and its failure alone results in $NSD = 0.2261$, since without it, some demand nodes far from the main source (e.g. node 2) are not fully supplied due to the limited capacity of pipelines connecting different areas (e.g. link 18-34), even though the total supply capacity of the sources is able to cover all demands. Combining with the failure of link 2-50, which cuts the supply from the main source (node 2) to other nodes, the supply of the whole network drops even more.

$NSD = 0$ is the most frequent value. The case ranked 5 is the failure of node 19 and has $NSD = 0.2261$. It occurs 7 888 times out of the one million simulations. Generally speaking, high values of NSD ($NSD > 0.3$) tend to have low frequency. For 146 out of 335 cases (43.6%), the demand can not be fully supplied, i.e. $NSD > 0$.

Figure 5 shows the cumulative distribution function (CDF) of NSD for the failure configurations. The mean value of NSD is 0.0285 over the 70 987 configurations with failures and 0.0020 over all 10^6 configurations simulated.

Figure 6 shows the NSD of the original network (Ori) and the mean values for all sampled configurations (All), the failure configurations only (Failure), the 30 most frequent cases (Top 30) and the seven classes of failures (SL, SN, SL-SN, SL-MN, ML, ML-SN, MN).

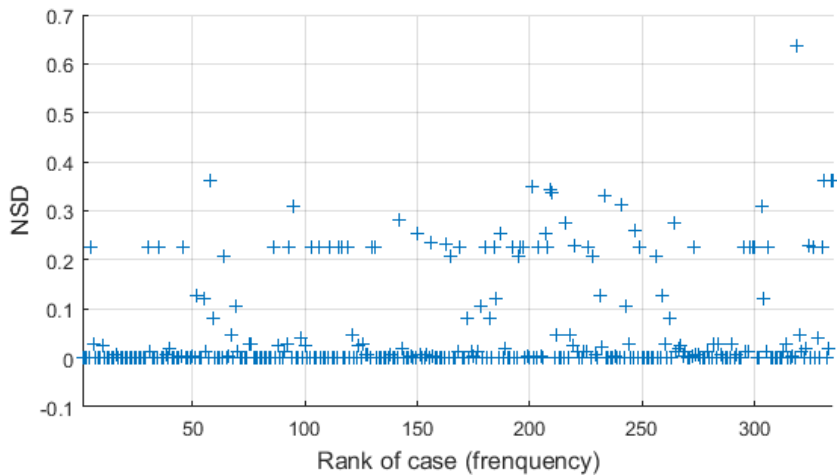


Figure 4: Non-supplied demand for the 335 cases

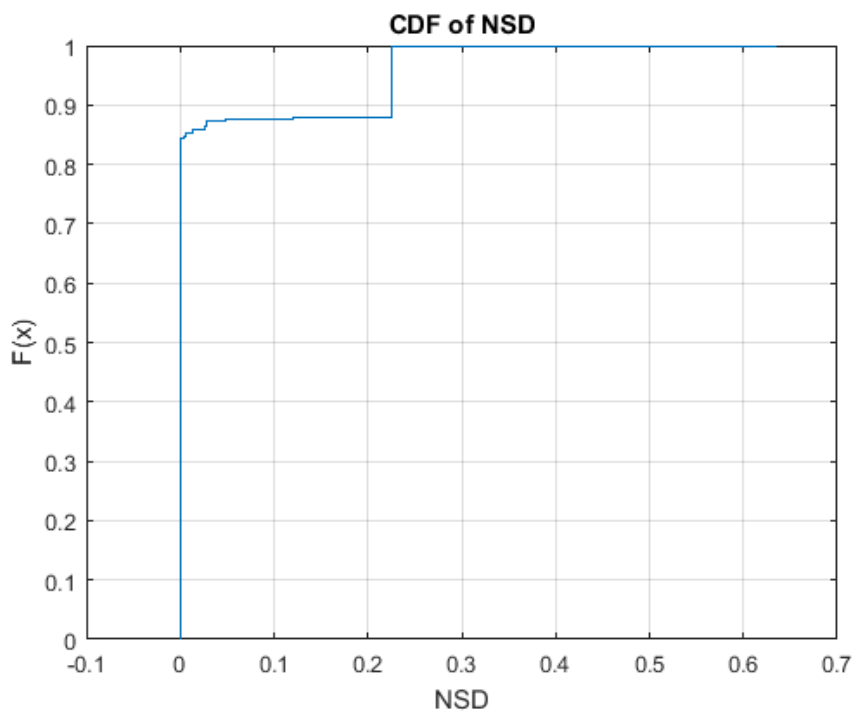


Figure 5: CDF of non-supplied demand for the failure configurations

The NSD for the top 30 cases is higher than that for the nominal configuration and comparable to that of all configurations. This indicates that, the most frequent cases have non-negligible impact on the demand supply.

If we compare the seven failure classes, we see that, as a whole, node failures have a more important impact in terms of non-supplied demand. In presence of node failure (SN, SL-SN and SL-MN), the non-supplied demand NSD is higher than for the cases with single

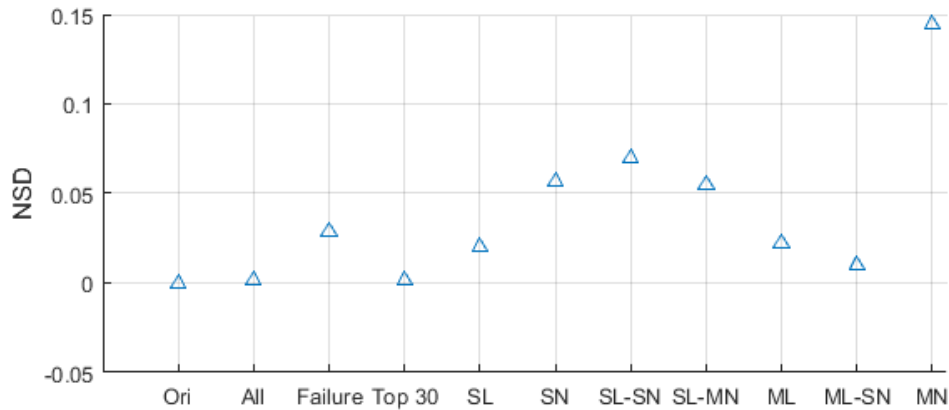


Figure 6: Non-supplied demand for different categories

link failure (SL). The combination of link failures only increases slightly the NSD : the mean NSD of ML is slightly higher than that of SL. As for ML-SN, among all the possible combinations of failures, only four cases occur once each, and they happen to have relatively small influence in terms of NSD , the mean being 0.010. The class of multiple node failures (MN) has a significantly high value of non-supplied demand. All MN cases without failure of node 19 have $NSD = 0$. The failure of node 19 alone would lead to $NSD = 0.2261$, and for the case where node 10, 11 and 19 are all failed at the same time, $NSD = 0.3111$. In fact, the link 3-11 represents a large capacity pipeline and in the absence of gas supply from the storages represented by nodes 10 and 19, the reduction of its capacity due to the failure of node 11 would result in the non-supply to demand nodes in the vicinity, depending mainly on the main source (node 2). Considering the relatively high failure probability of node 19, the MN class has a high value of NSD .

5.3.2 Controllability index

Figure 7 shows the index value of C_{ind} for the 335 cases and the abscissa axis is the frequency rank of the 335 cases. The controllability index takes three values 0.9107, 0.8928 and 0.875 for the 335 cases, as for the removal of single links. $C_{ind} = 0.9107$ is the most frequent value. For 58 out of 335 cases (17.3%), the controllability index C_{ind} is lower than that of the failure-free network configuration. The lowest value 0.875 is more present for the less frequent cases (i.e. Rank > 150).

Figure 8 shows the cumulative distribution function (CDF) of C_{ind} . Only 1 559 out of the 70 987 failure configurations (2%) have C_{ind} lower than 0.9107. The mean value of C_{ind} is 0.9102 over the 70 987 failure configurations and very close to 0.9107 over the 10^6

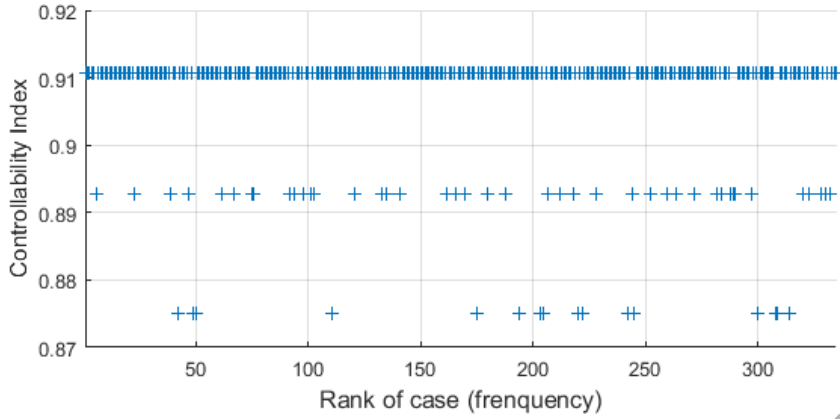


Figure 7: Controllability index for the 335 cases

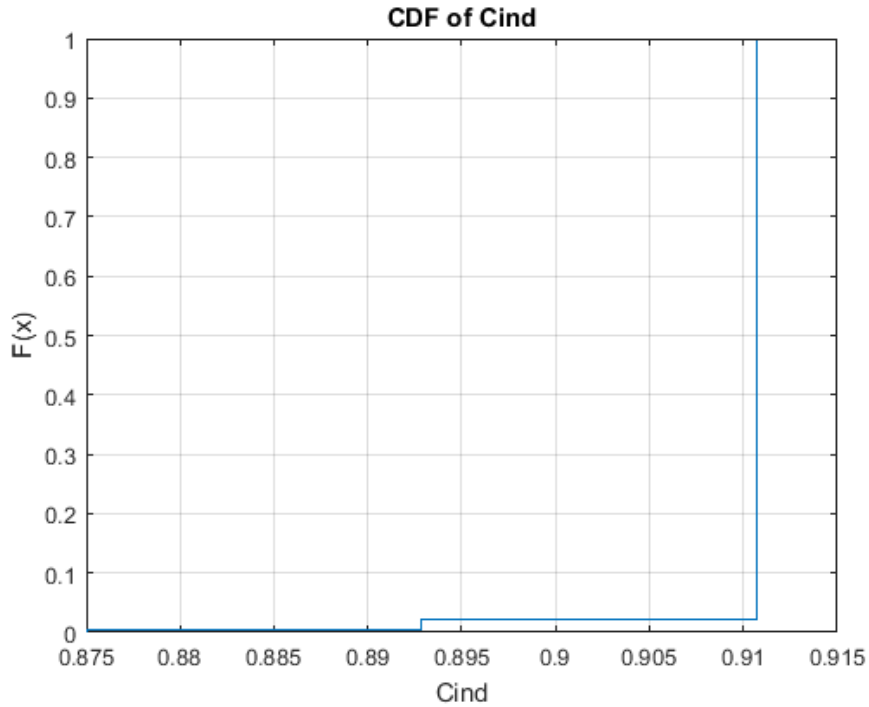


Figure 8: CDF of the controllability index for the failure configurations simulations.

Figure 9 shows the mean value of C_{ind} of the original network (Ori) and the mean values for all sampled configurations (All), the failure configurations only (Failure), the 30 most frequent cases (Top 30) and the seven classes of failures (SL, SN, SL-SN, SL-MN, ML, ML-SN, MN). The mean of C_{ind} of the top 30 cases is slightly lower than but still close to that of the nominal network configuration.

Node failures have no impact on the controllability index, since they only concern the

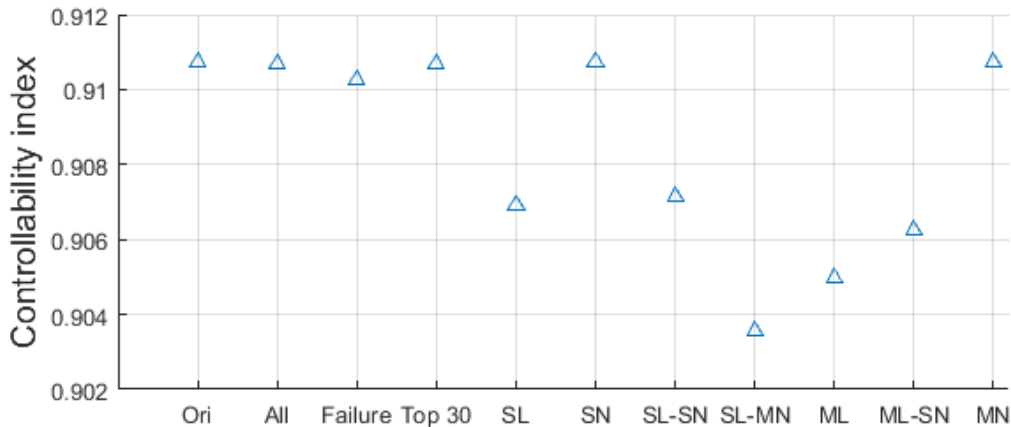


Figure 9: Controllability index for different categories

reduction of pipelines capacity but not their disconnection. Thus, link failures are the only contribution to the loss in controllability. The mean of all cases containing single link failures (i.e. SL, SL-SN and SL-MN) is 0.9069, slightly higher than that of the cases containing multiple link failures (i.e. ML and ML-SN) cases, which equals to 0.9051. This indicates that multiple failures have a more important impact on the controllability index, with C_{ind} reaching values no lower than 0.8750. For the SL, SL-SN and SL-MN cases, this lowest value results from the failures of links 6-35, 11-43 or 18-23, the removal of each of which decreases C_{ind} to 0.8750. As for ML and ML-SN cases, the lowest value comes from the combination of two link failures with no separate impact on C_{ind} (links 9-10 and 10-53), or the combination of two links whose removal decreases C_{ind} to 0.8929 (links 22-24 and 34-37).

5.3.3 Network topological efficiency

Figure 10 shows the values of E for the 335 cases, and the abscissa axis is the frequency rank of the 335 cases. The network topological efficiency ranges from 0.6317 to 0.6327. The most critical links are link 44-45 and link 18-34, whose removal decreases the efficiency to 0.6318. Multiple failures decrease the lowest value of E slightly. In fact, all the seven values below 0.6318 are related to the failure of link 44-45 or link 18-34: the five first cases include single link failures alone or together with single node failure, while the last two cases with low values are multiple link failures.

Figure 11 shows the cumulative distribution function of E . The topological efficiency E stays close to the value of the failure-free configuration. However, 10 581 out of the 70 987 failure configurations (14.91%) have E lower than the value of the failure-free configuration,

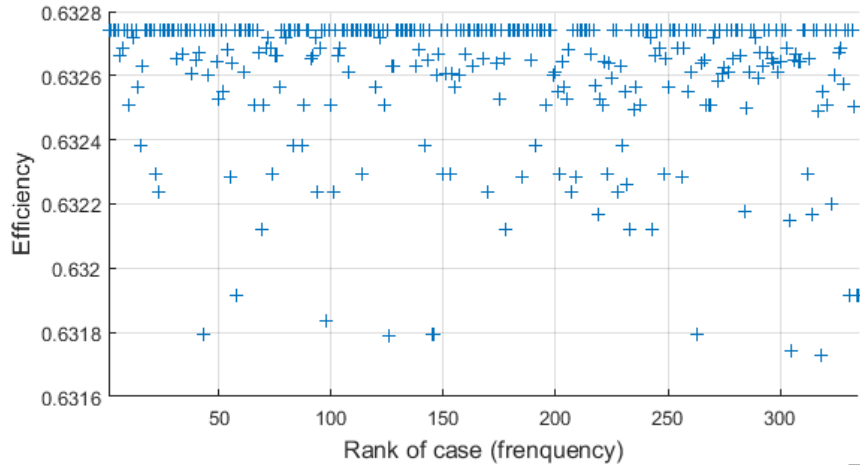


Figure 10: Network topological efficiency for the 335 cases

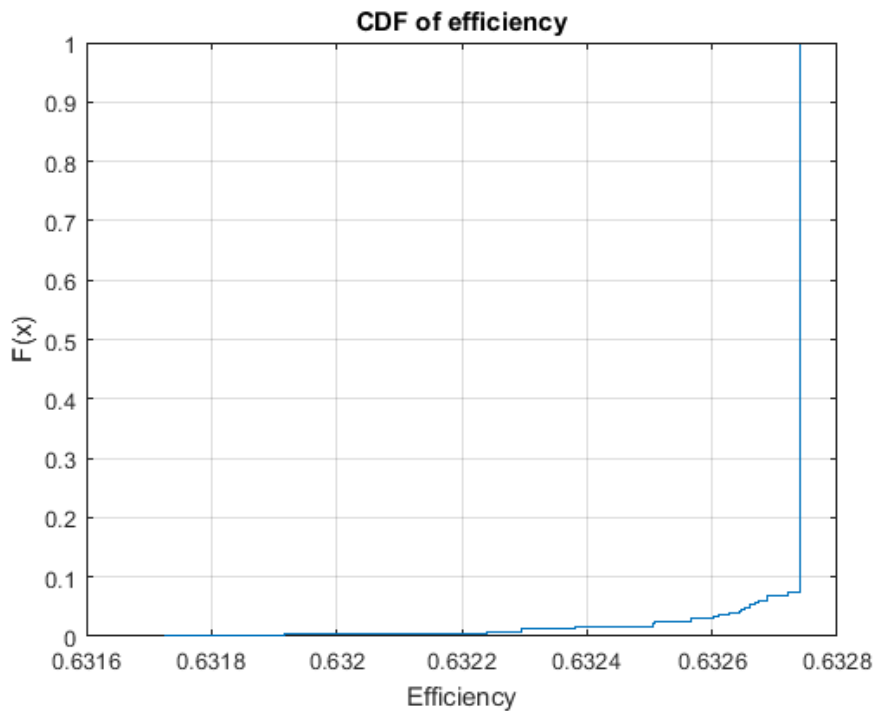


Figure 11: CDF of the network topological efficiency for the failure configurations

and 10581 is the number of configurations with at least one link failure. This means that failure of any link will influence the efficiency.

Figure 12 shows the mean value of E of the original network (Ori) and the mean values for all sampled configurations (All), the failure configurations only (Failure), the 30 most frequent cases (Top 30) and the seven classes of failures (SL, SN, SL-SN, SL-MN, ML, ML-SN, MN).

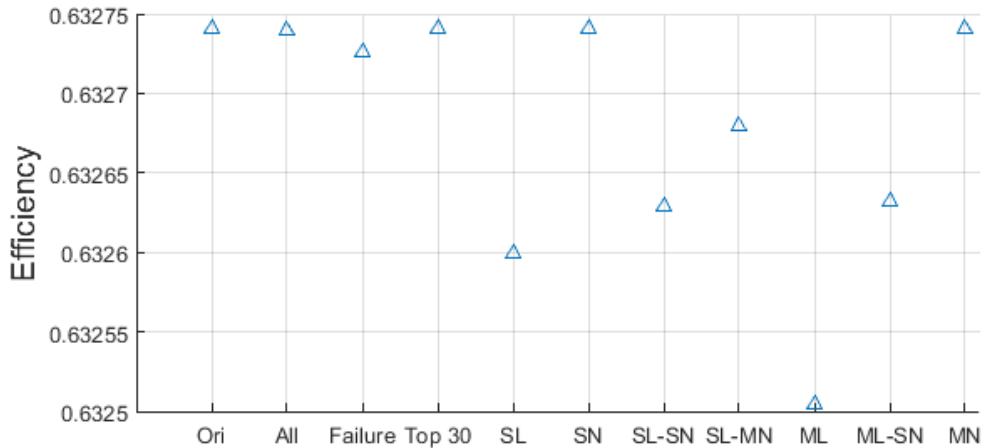


Figure 12: Network topological efficiency for different categories

Similar to the controllability index, node failures alone have no influence on topological efficiency and multiple link failures have a more important influence than single link failures. The mean of E over all cases of single link failure (i.e. SL, SL-SN and SL-MN) is 0.6326 and for the cases containing multiple link failures (i.e. ML and ML-SN) it is 0.6325.

Generally speaking, the variation of E is not significant, much less than the other two indexes. This is reasonable, because the network is not a sequential one and multiple paths exist between any two nodes: when one link fails, the gas flow can still be transferred via an alternative path, although of longer distance.

5.4 Protective actions

From the above analysis, we understand that node failures have significant importance in terms of supply, but do not affect other properties, that the link failures influence on NSD is less important than that of node failures and that the consequences of failures on controllability and topological efficiency are limited compared to NSD . This means that, supply is the primary concern with respect to protection from failures, whereas network connections and control are not so sensitive and more fault tolerant.

Node 19 is found to be an important node which may require protection priority, since it has a relatively high failure probability and at the same time has a significant impact on supply. Figure 13 compares the cumulative distribution function of NSD over all simulated configurations and the 991 570 configurations in which node 19 is not failed: the mean value of NSD of all configurations drops from 0.0020 to 0.0001, which is much lower.

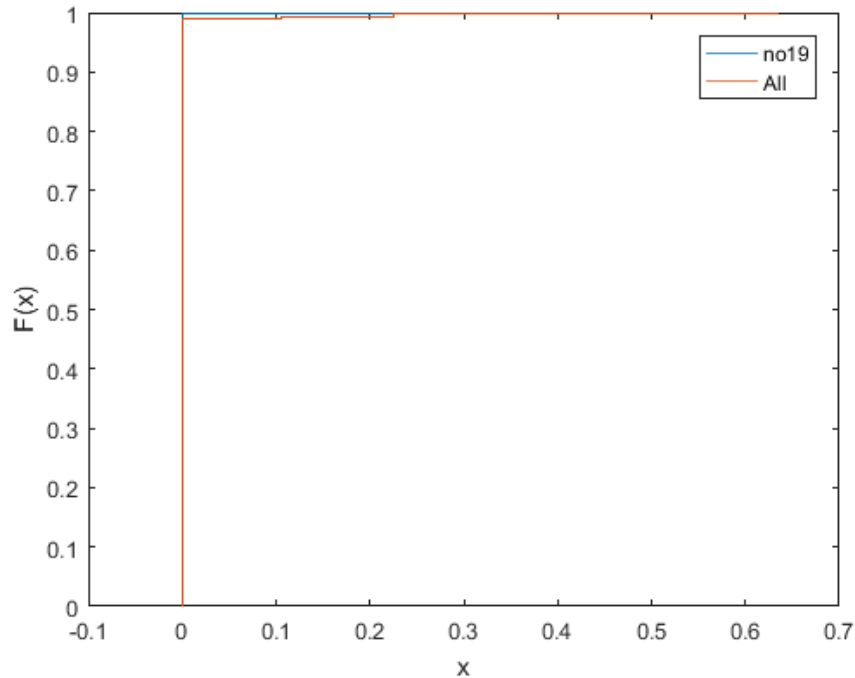


Figure 13: CDF of NSD for all configurations and for the configurations without failure of node 19

6 Conclusions

Critical infrastructures, such as power grids, gas or water distribution networks, etc., are complex networks designed and operated to supply the service demanded. The increasing threats to the safety and security of their functions make it crucial to ensure safe and reliable performance.

In this paper, we have developed an analysis framework considering three perspectives: supply service, controllability and topology. By performing such an analysis, we are able to identify the most critical elements within the network and quantify the consequences of failure scenarios. The analysis framework has been applied to a gas transmission network. In the current work, repair is not taken into account and only one failure level is simulated. Even with such simplifications, the presented framework is able to identify the critical elements in terms of different perspectives considered. In future work, a multi-state approach and repair will be considered.

The results of the analysis show that the influence of a single link is not the same for the three perspectives and neither are the various failure scenarios. Supply turns out to be the most affected by failures, and can, thus, be used as the objective for the prioritization of

investments for CI protection. However, other key performance indicators should be considered, including flexibility, recovery capacity, etc., which the other perspectives are sensitive to; then, the integration of the three perspectives should be considered within a multi-objective optimization for the multi-KPI design of CIs. This will be the object of future work. The findings of this work show the importance of considering several perspectives of analysis for CIs.

In the current paper, we emphasize the identification of the elements critical to the perspectives considered and the stochastic approach is adopted. Consideration of cascading effects and targeted attacks will complete the study in the continuation of this work, since this can provide more information for understanding the correlation among the different properties. Also, as CIs are more and more interconnected and automated, interdependencies and multi-CIs modeling become of great interest. The framework will be expanded and improved by taking into consideration the interdependencies in future works.

Acknowledgement

The authors would like to thank Dr. Vytis Kopustinskas and Dr. Pavel Praks of the Joint Research Center of the European Union for providing the data of the case study.

References

- [1] Wolfgang Kröger and Enrico Zio. *Vulnerable systems*. Springer Science & Business Media, 2011.
- [2] William B. Rouse. Engineering complex systems: Implications for research in systems engineering. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 2(33):154–156, 2003.
- [3] Julio M Ottino. Engineering complex systems. *Nature*, 427(6973):399–399, 2004.
- [4] EU. European Energy Security Strategy. *Communication from the Commission to the European Parliament and the Council. COM (2014) 330 final. Brussels, 2014*, 2014.
- [5] Madelene Lindström and Stefan Olsson. The European programme for critical infrastructure protection. In *Crisis Management in the European Union*, pages 37–59. Springer, 2009.

- [6] Ted G Lewis. *Critical infrastructure protection in homeland security: defending a networked nation*. John Wiley & Sons, 2014.
- [7] William Clinton. Presidential Decision Directive PDD-63, Protecting America's Critical Infrastructure. *The White House, Washington, DC*, 1998.
- [8] EU. Regulation no. 994/2010 of the european parliament and of the council of 20 october 2010 concerning measures to safeguard security of gas supply and repealing council directive 2004/67/ec. *Official Journal of the European Union*, 2010.
- [9] Eric Rigaud and Franck Guarnieri. Proposition of a conceptual and a methodological modelling framework for resilience engineering. In *2nd Symposium on Resilience Engineering*, pages 8–pages, 2006.
- [10] Efstathios Bakolas and Joseph H Saleh. Augmenting defense-in-depth with the concepts of observability and diagnosability from control theory and discrete event systems. *Reliability Engineering & System Safety*, 96(1):184–193, 2011.
- [11] Nancy Leveson. A new accident model for engineering safer systems. *Safety science*, 42(4):237–270, 2004.
- [12] Rudolf Kalman. On the general theory of control systems. *IRE Transactions on Automatic Control*, 4(3):110–110, 1959.
- [13] Yang-Yu Liu, Jean-Jacques Slotine, and Albert-László Barabási. Controllability of complex networks. *Nature*, 473(7346):167–173, 2011.
- [14] Ching-Tai Lin. Structural controllability. *Automatic Control, IEEE Transactions on*, 19(3):201–208, 1974.
- [15] Zhengzhong Yuan, Chen Zhao, Zengru Di, Wen-Xu Wang, and Ying-Cheng Lai. Exact controllability of complex networks. *Nature communications*, 4, 2013.
- [16] Yang-Yu Liu, Jean-Jacques Slotine, and Albert-László Barabási. Control centrality and hierarchical structure in complex networks. *Plos one*, 7(9):e44459, 2012.
- [17] Noah J Cowan, Erick J Chastain, Daril A Vilhena, James S Freudenberg, and Carl T Bergstrom. Nodal dynamics, not degree distributions, determine the structural controllability of complex networks. *PloS one*, 7(6):e38398, 2012.
- [18] Wen-Xu Wang, Xuan Ni, Ying-Cheng Lai, and Celso Grebogi. Optimizing controllability of complex networks by minimum structural perturbations. *Physical Review E*, 85(2):026115, 2012.

- [19] Tao Jia and Albert-László Barabási. Control capacity and a random sampling method in exploring controllability of complex networks. *Scientific reports*, 3, 2013.
- [20] Tao Jia, Yang-Yu Liu, Endre Csóka, Márton Pósfai, Jean-Jacques Slotine, and Albert-László Barabási. Emergence of bimodality in controlling complex networks. *Nature communications*, 4, 2013.
- [21] Tamás Nepusz and Tamás Vicsek. Controlling edge dynamics in complex networks. *Nature Physics*, 8(7):568–573, 2012.
- [22] Yujian Pan and Xiang Li. Structural controllability and controlling centrality of temporal networks. *PloS one*, 9(4):e94998, 2014.
- [23] Gang Yan, Jie Ren, Ying-Cheng Lai, Choy-Heng Lai, and Baowen Li. Controlling complex networks: How much energy is needed? *Physical review letters*, 108(21):218703, 2012.
- [24] Pavel Praks, Vytis Kopustinskas, and Marcelo Masera. Probabilistic modelling of security of supply in gas networks and evaluation of new infrastructure. *Reliability Engineering & System Safety*, 144:254–264, 2015.
- [25] Min Ouyang. Review on modeling and simulation of interdependent critical infrastructure systems. *Reliability engineering & System safety*, 121:43–60, 2014.
- [26] Adam M Lewis, David Ward, Lukasz Cyra, and Naouma Kourti. European reference network for critical infrastructure protection. *International journal of critical infrastructure protection*, 6(1):51–60, 2013.
- [27] Yi-Ping Fang and Enrico Zio. Hierarchical modeling by recursive unsupervised spectral clustering and network extended importance measures to analyze the reliability characteristics of complex network systems. *American Journal of Operations Research*, 3(1A):101–112, 2013.
- [28] Yi-Ping Fang and Enrico Zio. Unsupervised spectral clustering for hierarchical modelling and criticality analysis of complex networks. *Reliability Engineering & System Safety*, 116:64–74, 2013.
- [29] Weiwei Lu, Meirong Su, Brian D Fath, Mingqi Zhang, and Yan Hao. A systematic method of evaluation of the Chinese natural gas supply security. *Applied Energy*, 165:858–867, 2016.
- [30] Juhani Nieminen. On the centrality in a graph. *Scandinavian journal of psychology*, 15(1):332–336, 1974.

- [31] Linton C Freeman. Centrality in social networks conceptual clarification. *Social networks*, 1(3):215–239, 1978.
- [32] Vito Latora and Massimo Marchiori. Efficient behavior of small-world networks. *Physical review letters*, 87(19):198701, 2001.
- [33] Paolo Crucitti, Vito Latora, and Massimo Marchiori. A topological analysis of the Italian electric power grid. *Physica A: Statistical Mechanics and its Applications*, 338(1):92–97, 2004.
- [34] Åke J Holmgren. Using graph models to analyze the vulnerability of electric power networks. *Risk analysis*, 26(4):955–969, 2006.
- [35] Zili Zhang, Xiangyang Li, and Hengyun Li. A quantitative approach for assessing the critical nodal and linear elements of a railway infrastructure. *International Journal of Critical Infrastructure Protection*, 8:3–15, 2015.
- [36] James J Kelleher. Tactical communications network modelling and reliability analysis: An overview. Technical report, DTIC Document, 1991.
- [37] E. Zio. Reliability analysis of systems of systems. *IEEE Reliability Magazine*, pages 1–6, February 2016.
- [38] Enrico Zio. Challenges in the vulnerability and risk analysis of critical infrastructures. *Reliability Engineering & System Safety*, 152:137–150, 2016.
- [39] Enrico Zio. Critical infrastructures vulnerability and risk analysis. *European Journal for Security Research*, (DOI 10.1007/s41125-016-0004-2):1–18, 2016.
- [40] Enrico Zio. Reliability engineering: Old problems and new challenges. *Reliability Engineering & System Safety*, 94(2):125–141, 2009.
- [41] Zhang Limiao, Li Daqing, Qin Pengju, Fu Bowen, Jiang Yinan, Enrico Zio, and Kang Rui. Reliability analysis of interdependent lattices. *Physica A: Statistical Mechanics and its Applications*, 452:120–125, 2016.
- [42] Enrico Zio and Lucia R Golea. Analyzing the topological, electrical and reliability characteristics of a power transmission system for identifying its critical elements. *Reliability Engineering & System Safety*, 101:67–74, 2012.
- [43] Jian Li, Leonardo Dueñas-Osorio, Changkun Chen, and Congling Shi. Connectivity reliability and topological controllability of infrastructure networks: A comparative assessment. *Reliability Engineering & System Safety*, 156:24–33, 2016.

- [44] José Emmanuel Ramirez-Marquez and Claudio M Rocco. All-terminal network reliability optimization via probabilistic solution discovery. *Reliability Engineering & System Safety*, 93(11):1689–1697, 2008.
- [45] Darren L Deeter and Alice E Smith. Heuristic optimization of network design considering all-terminal reliability. In *Reliability and Maintainability Symposium. 1997 Proceedings, Annual*, pages 194–199. IEEE, 1997.
- [46] Berna Dengiz, Fulya Altiparmak, and Alice E Smith. Efficient optimization of all-terminal reliable networks, using an evolutionary approach. *IEEE transactions on Reliability*, 46(1):18–26, 1997.
- [47] Manolis Papadrakakis and Nikos D Lagaros. Reliability-based structural optimization using neural networks and monte carlo simulation. *Computer methods in applied mechanics and engineering*, 191(32):3491–3507, 2002.
- [48] Francesco Cadini, Enrico Zio, and Cristina-Andreea Petrescu. Optimal expansion of an existing electrical power transmission network by multi-objective genetic algorithms. *Reliability Engineering & System Safety*, 95(3):173–181, 2010.
- [49] Yiping Fang, Nicola Pedroni, and Enrico Zio. Optimization of cascade-resilient electrical infrastructures and its validation by power flow modeling. *Risk Analysis*, 35(4):594–607, 2015.
- [50] Yan-Fu Li, Giovanni Sansavini, and Enrico Zio. Non-dominated sorting binary differential evolution for the multi-objective optimization of cascading failures protection in complex networks. *Reliability Engineering & System Safety*, 111:195–205, 2013.
- [51] Wenyuan Wu, Holger R Maier, and Angus R Simpson. Multiobjective optimization of water distribution systems accounting for economic cost, hydraulic reliability, and greenhouse gas emissions. *Water Resources Research*, 49(3):1211–1225, 2013.
- [52] Eckart Zitzler. Evolutionary algorithms for multiobjective optimization: Methods and applications. 1999.
- [53] Eckart Zitzler, Marco Laumanns, and Lothar Thiele. Spea2: Improving the strength pareto evolutionary algorithm. 2001.
- [54] Nidamarthi Srinivas and Kalyanmoy Deb. Multiobjective optimization using nondominated sorting in genetic algorithms. *Evolutionary computation*, 2(3):221–248, 1994.

- [55] Kalyanmoy Deb, Amrit Pratap, Sameer Agarwal, and TAMT Meyarivan. A fast and elitist multiobjective genetic algorithm: Nsga-ii. *IEEE transactions on evolutionary computation*, 6(2):182–197, 2002.
- [56] Hisao Ishibuchi, Noritaka Tsukamoto, and Yusuke Nojima. Evolutionary many-objective optimization: A short review. In *Evolutionary Computation, 2008. CEC 2008. (IEEE World Congress on Computational Intelligence). IEEE Congress on*, pages 2419–2426. IEEE, 2008.
- [57] Zhenan He and Gary G Yen. Many-objective evolutionary algorithms based on coordinated selection strategy. *IEEE Transactions on Evolutionary Computation*, 21(2):220–233, 2017.
- [58] Kalyanmoy Deb, Manikanth Mohan, and Sikhar Mishra. A fast multi-objective evolutionary algorithm for finding well-spread pareto-optimal solutions. *KanGAL report*, 2003002:1–18, 2003.
- [59] Kalyanmoy Deb, Manikanth Mohan, and Shikhar Mishra. Evaluating the ε -domination based multi-objective evolutionary algorithm for a quick computation of pareto-optimal solutions. *Evolutionary computation*, 13(4):501–525, 2005.
- [60] Kalyanmoy Deb and Himanshu Jain. An evolutionary many-objective optimization algorithm using reference-point-based nondominated sorting approach, part i: Solving problems with box constraints. *IEEE Trans. Evolutionary Computation*, 18(4):577–601, 2014.
- [61] Qingfu Zhang and Hui Li. Moea/d: A multiobjective evolutionary algorithm based on decomposition. *IEEE Transactions on evolutionary computation*, 11(6):712–731, 2007.
- [62] Robert J Lygoe, Mark Cary, and Peter J Fleming. A real-world application of a many-objective optimisation complexity reduction process. In *International Conference on Evolutionary Multi-Criterion Optimization*, pages 641–655. Springer, 2013.
- [63] Oliver Chikumbo, Erik Goodman, and Kalyanmoy Deb. Approximating a multi-dimensional pareto front for a land use management problem: A modified moea with an epigenetic silencing metaphor. In *Evolutionary Computation (CEC), 2012 IEEE Congress on*, pages 1–9. IEEE, 2012.
- [64] Kaname Narukawa and Tobias Rodemann. Examining the performance of evolutionary many-objective optimization algorithms on a real-world application. In *Genetic and Evolutionary Computing (ICGEC), 2012 Sixth International Conference on*, pages 316–319. IEEE, 2012.

- [65] P Praks and V Kopustinskas. Identification and ranking of important elements in a gas transmission network by using progasnet. In *Risk, Reliability and Safety: Innovating Theory and Practice*, pages 1573–1579. CRC Press, 2016.
- [66] Rudolf Emil Kalman. Mathematical description of linear dynamical systems. *Journal of the Society for Industrial & Applied Mathematics, Series A: Control*, 1(2):152–192, 1963.
- [67] Anna Lombardi and Michael Hörnquist. Controllability analysis of networks. *Physical Review E*, 75(5):056110, 2007.
- [68] EGIG. 8th report of the european gas pipeline incident data group. December 2011.
- [69] M Van der Borst and H Schoonakker. An overview of psa importance measures. *Reliability Engineering & System Safety*, 72(3):241–245, 2001.