

Evaluation of the robustness of critical infrastructures by Hierarchical Graph representation, clustering and Monte Carlo simulation

Elisa Ferrario, Nicola Pedroni, Enrico Zio

▶ **To cite this version:**

Elisa Ferrario, Nicola Pedroni, Enrico Zio. Evaluation of the robustness of critical infrastructures by Hierarchical Graph representation, clustering and Monte Carlo simulation. Reliability Engineering and System Safety, 2016, 155, pp.78-96. 10.1016/j.ress.2016.06.007 . hal-01346420

HAL Id: hal-01346420
https://hal.science/hal-01346420

Submitted on 24 Jul 2016

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.

Evaluation of the robustness of critical infrastructures by Hierarchical Graph representation, clustering and Monte Carlo simulation

E. Ferrario[a], N. Pedroni[a] and E. Zio[a,b]

[a]*Chair System Science and the Energy Challenge, Fondation Electricité de France (EDF), CentraleSupélec, Université Paris-Saclay, Grande Voie des Vignes, 92290 Chatenay-Malabry, France*

enrico.zio@ecp.fr, enrico.zio@supelec.fr

[b]*Department of Energy, Politecnico di Milano, Italy*

enrico.zio@polimi.it

Abstract

In this paper, we present a methodological work that adopts a system-of-systems (SoS) viewpoint for the evaluation of the robustness of interdependent critical infrastructures (CIs). We propose a Hierarchical Graph representation, where the product flow is dispatched to the demand nodes in consideration of different priorities. We use a multi-state model to describe different degrees of degradation of the individual components, where the transitions between the different states of degradation occur stochastically. The quantitative evaluation of the CIs robustness is performed by Monte Carlo simulation. The methodological approach proposed is illustrated by way of two case studies: the first one concerns small-sized gas and electricity networks and a supervisory control and data acquisition (SCADA) system; the second one considers a moderately large power distribution network, adapted from the IEEE 123 node test feeders. The large size of the second case study requires hierarchical clustering for performing the robustness analysis.

Keywords: Interdependent critical infrastructures, robustness, Hierarchical Graph, hierarchical clustering, multi-state model, system of systems, Monte Carlo simulation, Markov and semi-Markov processes.

1. Introduction

Critical infrastructures (CIs) are complex systems essential for providing goods (e.g., energy) and services (e.g., transportation) across local, regional and national boundaries [Kröger and Zio, 2011]. Typically, they present both structural and dynamic complexities. The former derive from i) heterogeneity of components across different technological domains, due to the integration among different systems and ii) scale and dimensionality of connectivity through a large number of components (nodes), highly interconnected by dependences (unidirectional relationships) and interdependences (bidirectional relationships). Dynamic complexity manifests through the emergence of (unexpected) system behavior in response to changes in the environmental and operational conditions of its components. Furthermore uncertainties exist in the failure behavior of CI components, interconnections and interactions, so that the prediction of CI failure behavior is difficult [Zio and Aven, 2011].
Engineered, physically networked CIs are considered in this paper. Examples are those providing: energy (electricity, oil and gas supply as subsectors); transportation (by rail, road, air, shipping); information and telecommunication (such as the internet); drinking water, including wastewater treatment, etc.

Due to the increasing complexity of CIs, random failures, natural events and malevolent attacks can have severe consequences on health, safety, security, economics and social well-being. In this respect, evaluating the robustness of CIs is fundamental to be able to improve their design and management so to reduce the impacts of disruptive events. There is no unique definition of robustness. Jensen defines it as the degree to which a system can function correctly in the presence of inputs different from those assumed [Jensen, 2001]; for [Carlson and Doyle, 2002] and [Jen, 2003], the robustness guarantees the maintenance of certain desired system characteristics, despite fluctuations in the behavior of its components or in its environment. [Jen, 2003] and [Ali et al., 2003] specify that the concept of robustness should be defined for a given set of system features, under a given set of perturbations applied to the system. According to Foster, robustness is the ability of a system to react to noisy input parameters with little performance degradation [Foster, 2008]. A recent definition of robustness is given in the glossary of the specialty group on "Foundations of Risk Analysis" of the Society for Risk Analysis, as the antonym of vulnerability [SRA, 2015]. In addition, a system is considered robust to uncertainty if specified goals are achieved, despite large information gaps (information gap is the disparity between what is known, and what needs to be known to ensure

specified goals) [SRA, 2015]. In this work, robustness is seen as the capability of the CIs to resist to failures or partial failures of the CIs components assuring the required level (or a high level) of supply of goods or services.

Traditionally, three steps are performed in the robustness assessment of CIs: (i) the system is represented to define the structural, logical and functional relations among its components; (ii) a mathematical model of the system is built to quantify its performance indicators; (iii) the model is solved, e.g., by simulating its behavior under different operational and accidental conditions.

With respect to system representation (i), several types of approaches exist in literature, many of which rely on a hierarchy or graph structure. Hierarchical modeling has been often adopted to represent and model complex systems, since many organizational and technology-based systems are hierarchical in nature [Courtois, 1985; Haimes, 2012]. Hierarchical functional models include Goal Tree Success Tree (GTST) [Kim and Modarres, 1987] – also combined with Master Logic Diagram (MLD) [Hu and Modarres, 2000] – and Multilevel Flow Models (MFM) [Lind, 2011a; Lind, 2011b]. In risk analysis, common representation techniques are hierarchical trees that are possibly used to identify i) the initiating causes of a pre-specified, undesired event or ii) the accident sequences that can generate from a single initiating event through the development of structured logic trees (i.e., fault and event trees, respectively) [Zio, 2007]. Recently, also networks have been represented by hierarchical modeling [Gómez et al., 2013; Buriticá Cortés et al., 2014].

In complex network theory approaches, instead, complex systems are represented by networks where the nodes stand for the components and the links describe the physical and relational connections among them [Gheorghe et al., 2006; Dueñas-Osorio and Vemuru, 2009; Hernandez-Fajardo and Dueñas-Osorio, 2013; Fang et al., 2014; Praks and Kopustinskas, 2014]. Network-based approaches model interdependent CIs on the basis of their topologies or flow patterns [Ouyang, 2014]. Physical and non-physical (heterogeneous) overlapping infrastructures have been represented as networks to identify risk scenarios and the impacts on connected networks in [Lambert and Sarda, 2005]. Also probabilistic methods (e.g., Petri nets [Laprie et al., 2007], Bayesian networks [Di Giorgio and Liberati, 2012] and flowgraphs [Huzurbazar, 2005]) are based on graph representations.

In this paper, we present a methodological work that embraces a system-of-systems (SoS) framework of analysis [Gheorghe and Vamanu, 2008; Eusgeld et al., 2011; Kröger and Zio,

2011; Tsilipanos et al., 2013] and propose a Hierarchical Graph representation to evaluate the robustness of interdependent CIs, measured by its capability to deliver the required amount of product (e.g., energy, water, etc.) to the demand nodes of the infrastructure. In this respect, the system can be considered robust if it can maintain the required level (or a high level) of delivery when it is affected by perturbations (failures and partial failures). In doing so, we take into account the fact that the demand nodes may have different importances, which leads to possibly different priorities in the distribution of the product flow through the connections to the elements of the CI. For example, hospitals may be considered more important than residential buildings given their role in the health-care system; as a consequence, in the case of a reduction of electric power that can be delivered in the network, hospitals may receive the priority with respect to houses. This ranking of priority should be fixed by the analyst and, then, criteria (hereafter also referred to as "importance criteria") for the partition of the flow (e.g., electric power) in the network can be defined. In this work, we assume three different importance criteria that depend on the geographic position of the demand nodes, the quantity of product required by each of them and the assumption of equality of the demand nodes.

The representation proposed consists of a graph structured in hierarchical levels that allows highlighting critical arcs and supporting the quantitative robustness evaluation by assigning different priorities to the demand nodes. Critical arcs are here defined as those links whose interruption or degradation affects several demand nodes. This concept of criticality can be related to that of "importance measure" used in reliability theory. Actually, importance measures quantify the contribution of a given component to a properly selected measure of system performance (e.g., robustness in this case): see, for example, the Birnbaum [Birnbaum, 1969], Fussell and Vesely [Fussell, 1975], criticality importance measures [Elsayed, 2012], etc. More specifically, with respect to network system analysis, other importance measures have been defined to measure component criticality [Borgatti, 2005; Kröger and Zio, 2011], like classical topological centrality measures including the degree of centrality [Nieminen, 1974; Freeman, 1979], the closeness centrality [Sabidussi, 1966; Freeman, 1979; Wasserman and Faust, 1994], the betweeness centrality [Freeman, 1979] and the information centrality [Latora and Marchiori, 2007].

For a more realistic representation, we adopt a multi-state model where different degrees of degradation of the individual components are contemplated [Ferrario and Zio, 2014] [Natvig, 2010]; the transitions between the different states of degradation occur stochastically and are modeled within Markov and semi-Markov processes.

For illustration purposes, we consider two case studies: the first one is characterized by small-sized interconnected gas and electricity networks and a supervisory control and data acquisition (SCADA) system [Nozick et al., 2005]; the second one is adapted from the IEEE 123 node test feeders [IEEE, 2000] and includes a large electricity distribution network. The first case study is chosen small enough to be able to clearly illustrate the Hierarchical Graph modeling of (three) connected systems, considering different priorities of the demand nodes. The second case study serves the purpose of showing how the approach can be extended when the size of the system increases.
As a measure of the robustness of the system, we evaluate the steady-state probability distributions of the product (e.g., gas and/or electricity) delivered to the demand nodes.

The quantitative evaluation of the system robustness is performed by Monte Carlo (MC) simulation [Kalos and Whitlock, 1986; Zio, 2013]; in the second case study of larger dimension, an unsupervised spectral clustering algorithm is employed to make the size of the CI manageable and reduce the computational burden related to the analysis [Fang and Zio, 2013].

The remainder of the paper is organized as follows. In Section 2, the Hierarchical Graph representation is introduced and the importance criteria considered are illustrated; in Section 3, the procedural steps to evaluate the robustness of interconnected CIs by Hierarchical Graph and MC simulation are provided, and the combination of Hierarchical Graph and clustering analysis is given, then the advantages and limitations of the approach are discussed; Section 4 contains the description of the two case studies, the representation of the corresponding systems and the results obtained; in Section 5, some conclusions are provided. Finally, in the Appendix the data related to the second case study are illustrated.

2. Hierarchical Graph representation of systems of systems

The proposed representation technique can be applied to *engineered, physically networked* CIs (energy, transportation, information and telecommunication) characterized by a radial structure, i.e., by unidirectional flows of "products" (power, water, gas, data). Actually, the representation requires that the CI of interest be first modeled by a directed graph of nodes and arcs without loops (in this case, the arcs may represent elements of an infrastructure or the connections between different infrastructures). Typical radial systems are the distribution networks.

To build the Hierarchical Graph representation, we then, need to distinguish between input, demand (load) and transmission arcs: the "input arcs" connect the production sources to the network, the "demand arcs" terminate with nodes that require a given amount of product, whereas the "transmission arcs" transfer the product to other components in the network. Notice that the transmission and the demand arcs may coincide: for example, an arc may be needed to supply the connected node and in addition it may be required to transmit the product to other arcs/nodes.

In the Hierarchical Graph representation, the adjective "hierarchical" does not imply a "decomposition of the system into different levels of details", as in other hierarchical models (e.g., Goal Tree Success Tree – Dynamic Master Logic Diagram [Hu and Modarres, 2000] and hierarchical clustering [Gómez et al., 2013]), but it simply means that the graph of interconnected CIs is structured in hierarchical levels. In extreme synthesis, the representation is built as follows: at the bottom of the graph, the inputs (i.e., the arcs through which the product is injected into the networks) are represented; at the top, the goals (i.e., the demand nodes that have to be satisfied) are reported; in the middle, all the other arcs (transmission and/or load arcs) that provide product to the demand nodes are organized in hierarchical levels. These levels are numbered on the basis of the number of demand nodes that are served by the corresponding arcs: the higher the number of demands supplied by an arc, the higher the hierarchical level of that arc. For example, all the arcs that are required to supply *LV* demand nodes are "placed" at hierarchical level *LV*.

Formally, let us consider A interconnected infrastructure systems $S^{(a)}$, $a = 1, \ldots, A$, constituting the overall SoS, numbered in order in such a way that the first q exchange physical product (e.g., energy or material) and the last $(A - q)$ exchange information, and are useful for the operation and control of the connected systems (e.g., a supervisory control and data acquisition - SCADA - system). The total number of components (arcs) transmitting physical flow is denoted as N.

For illustration purposes, refer to Figure 1 in which the graph of a SoS (top) and its corresponding Hierarchical Graph (bottom) are reported. The SoS in the example is composed by A = 4 systems, where the first two, i.e., $S^{(1)}$ and $S^{(2)}$, exchange physical product (solid links in Figure 1, top) and the last two, i.e., $S^{(3)}$ and $S^{(4)}$, support system $S^{(1)}$ (dotted links in Figure 1, top). The total number of components (arcs) is N = 8.

As described above, the Hierarchical Graph depicts the inputs at the bottom of the representation, i.e., in this case, arc $S_1^{(1)}_S_2^{(1)}$ in Figure 1 (bottom); also, it shows the goals (i.e.,

the demand nodes) at the top: in this case, the demand nodes are represented by all the nodes of systems $S^{(1)}$ and $S^{(2)}$, except $S_1^{(1)}$, which is the source of product. Finally, it organizes the arcs in different hierarchical levels according to the number of demand nodes they supply: for example, in this case arc $S_1^{(2)}_S_2^{(2)}$ is at hierarchical level 4 since it provides product to four demand nodes, i.e., $S_2^{(2)}$, $S_3^{(2)}$, $S_4^{(2)}$ and $S_5^{(2)}$. The quantity of product required by the demand nodes is referred to as D_{dem}, where the subscript '*dem*' is the indicator of the demand node among the N components.

Notice that the arcs referred to the (A – *q*) control and information systems (which are not contributing to the flow of product, but influence the state of the other arcs) do not appear in the hierarchical structure: instead, they are reported in a trapezoidal frame under the corresponding arc that they affect.

The squares located between the hierarchical levels mean that the product at that level has to be partitioned among the corresponding demand nodes.

This representation allows highlighting all the paths going from the input sources to the end nodes: for example, in Figure 2, the path from input $S_1^{(1)}$ to node $S_3^{(2)}$ is highlighted. In addition, the representation is able to put in evidence the critical arcs as those located at higher hierarchical levels, since their interruption or degradation affects more demand nodes: for example, in Figure 2 arc $S_2^{(1)}_ S_1^{(2)}$ is more critical than arc $S_2^{(1)}_ S_3^{(1)}$ since the first one is required to supply five demand nodes (i.e., $S_1^{(2)}$, $S_2^{(2)}$, $S_3^{(2)}$, $S_4^{(2)}$ and $S_5^{(2)}$), whereas the second one is necessary just for node $S_3^{(1)}$. As illustrated in Section 1, in this view the concept of critical arcs here adopted can be related to that of "importance measure" employed in classical reliability theory.

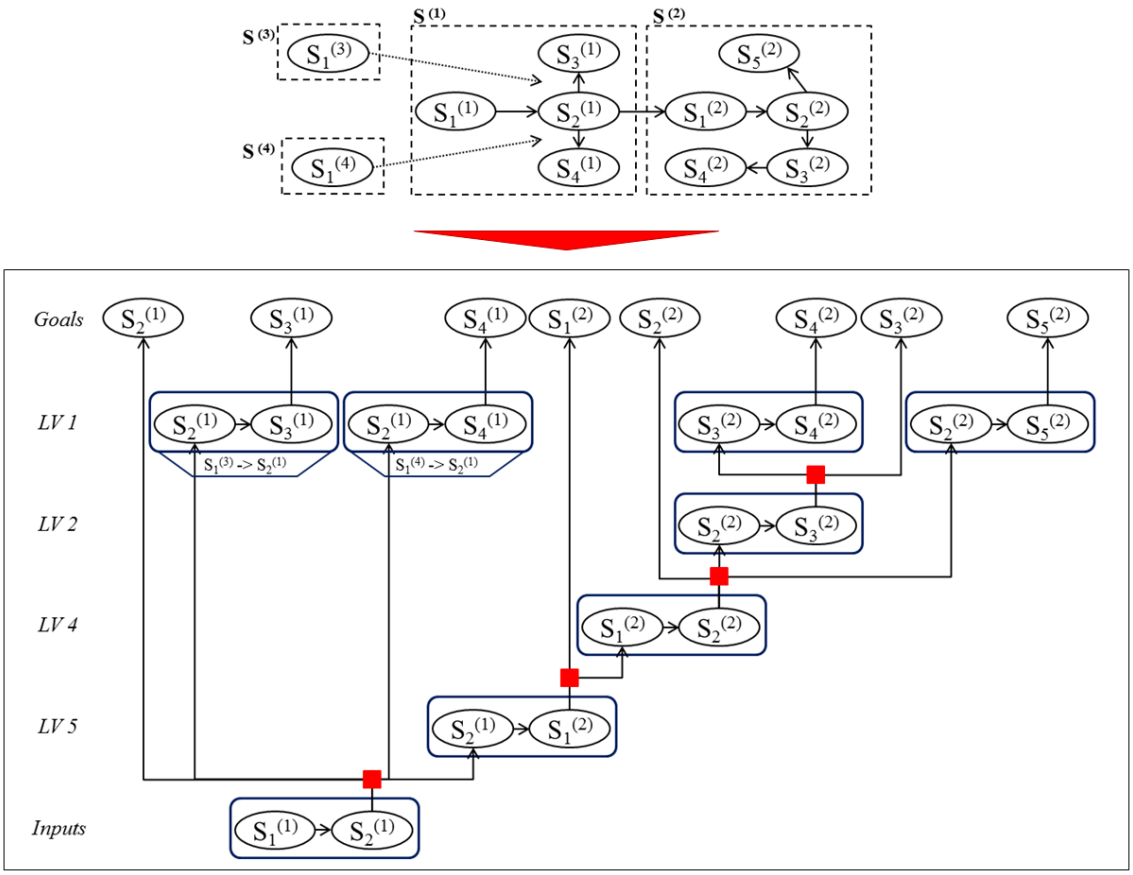


Figure 1: Top: graph of the components of the system of systems; the links represent the exchange of physical product (solid lines) and influence/support relationships (dotted lined). Bottom: corresponding Hierarchical Graph; LV: Level.

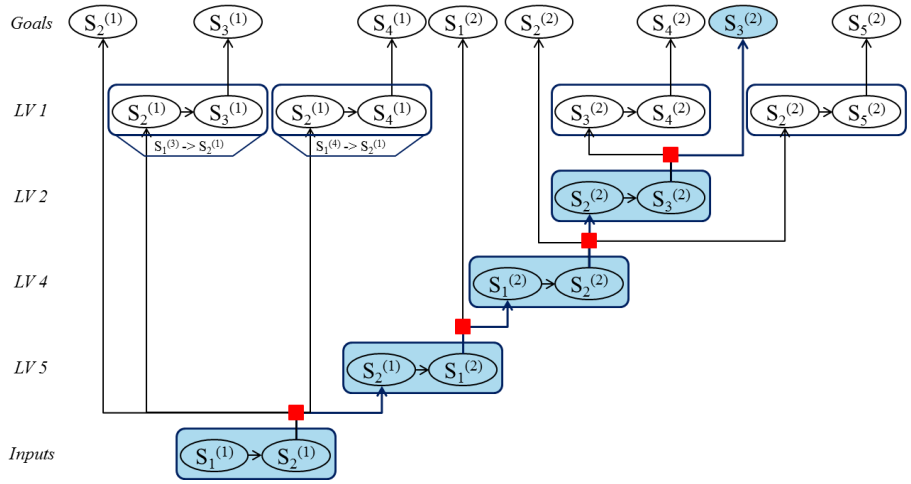


Figure 2: Hierarchical Graph of the system of systems in Figure 1, highlighting the path from the input to demand node $S_3^{(2)}$; LV: Level.

This representation is useful to analyze the robustness of interdependent CIs taking into account different priorities in the distribution of the product to the demand nodes according to their importance (relevance). In this context, Martinez et al. use path diagrams in representation of

the relative importance of network elements across infrastructure systems accounting for the nonquantifiable (qualitative) features of the allocation of resources among the network elements [Martinez et al., 2011]. In this work, we neglect the qualitative aspects and we focus only on the quantitative objectives. Notice that the word "importance" is not here related to the concept of "importance measure" used in reliability theory, but it is employed to indicate the relevance of the demand nodes. For example, hospitals may be considered more important than residential buildings, given their role in the health-care system; as a consequence, in the case of a malfunctioning in the electrical transmission line, hospitals may be a priority with respect to common residential areas. Regarding the water distribution systems, a fire-fighting consumption should have the priority with respect to daily water consumption. These priorities of the demand nodes are defined by the analyst and drive the definition of importance criteria needed for the optimal flow partition in the network. However, these importance criteria are not explicitly shown in the representation, which instead is more focused in highlighting the hard, physical constraints that affect the product distribution. Three different importance criteria are considered in this work, namely, sequential, proportional and equal; such criteria are explained hereafter with respect to a simple example consisting of an input of 50 units and two demand nodes S_1 and S_2 that require 40 and 100 units, respectively (i.e., $D_{S1} = 40$ and $D_{S2} = 100$).

Sequential importance consists in ranking the demand nodes sequentially on the basis of a chosen "ranking criterion" (e.g., according to their distance from the source node: the closer the node to the source, the higher the importance). In this case, the nodes classified as more important are given higher priority; with respect to the example above, if S_1 is more important than S_2, the input product is given first to S_1 until it is completely supplied and the rest (i.e., 50 – 40 = 10 units) is fed to S_2. Vice versa if S_2 is more important than S_1, the input is given entirely to S_2 and there is no product left to supply S_1.

Proportional importance orders the demand nodes on the basis of the quantity of product they need: the higher their demand, the higher their importance. Then, the product is partitioned into the network according to *ratios of importance* associated to each demand node, computed as the ratio between the quantity of product required by a node and the sum of all the demands of the entire system. With respect to the example above, S_2 is more important than S_1 because it requires a higher quantity of product. The ratios of importance are $D_{S2}/(D_{S1} + D_{S2}) = 0.7$ for S_2 and $D_{S1}/(D_{S1} + D_{S2}) = 0.3$ for S_1. Then, the input is partitioned as follows: 30% (i.e., 15 product units) is given to S_1 and 70% (i.e., 35 product units) to S_2.

Equal importance considers the demand nodes equal, even if their demands are different. Thus, the input is partitioned into equal parts on the basis of the number of the demand nodes. With

respect to the example above, the input is divided in two equal parts: 50% is given to S_1 and 50% to S_2.

The detailed operative steps of the algorithm for partitioning the product according to these criteria are given in Section 3.1.

3. Monte Carlo simulation and Hierarchical Graphs for critical infrastructures evaluation

Within a multi-state system-of-systems (SoS) analysis framework, we wish to evaluate the performance of critical infrastructures (CIs) in terms of *robustness*, measured by the steady-state probability distribution of the product delivered at the demand nodes of the system. The quantitative evaluation is carried out by combining the Hierarchical Graph representation of Section 2 with Monte Carlo (MC) simulation.

In Section 3.1, the operative steps of the basic procedure are presented. Then, a modification of the basic procedure is proposed in Section 3.2 to deal with CIs of large size: in particular, a clustering algorithm is adopted to pre-process the CI in order to make its size manageable and reduce the computational burden associated to the analysis; in Section 3.3 the advantages and limitations of the approach are discussed.

3.1. Operative steps combining Monte Carlo simulation and Hierarchical Graph

We generically denote the state of a component of the CI (i.e., the capacity of the arcs) as $\zeta_{comp,i}$, $i \in \{1, 2, \ldots, NS_{comp}\}$, $comp = 1, \ldots, N$, where N is the total number of components in the SoS, the subscript '*comp*' indicates the component of interest, identified by its name or by an integer number from 1 to N, NS_{comp} is the total number of states for component *comp*, and *i* is the state identification number (when $i = 1$, the component is in the worst state, whereas when $i = NS_{comp}$, it is in the best state). For example, supposing that component $S_1^{(2)}_S_2^{(2)}$ can enter three possible states, namely 0, 10 and 20, we denote the total number of states for the component as $NS_{S1(2)_S2(2)} = 3$, and the corresponding states as $\zeta_{S1(2)_S2(2),1} = 0$, $\zeta_{S1(2)_S2(2),2} = 10$, and $\zeta_{S1(2)_S2(2),3} = 20$.

The quantity of product requested by the demand nodes is indicated by the vector $\{D_{dem}\}$, $dem \in \{1, \ldots, N\}$, where the subscript '*dem*' indentifies the demand nodes.

In what follows, we describe an algorithm combining Monte Carlo simulation and Hierarchical Graph representation for the evaluation of the robustness of CIs within a multi-state SoS framework; as mentioned before, in this work the robustness can be seen as the capability of the CIs to resist to failures or partial failures of the arcs assuring the required level (or a high level) of supply of product to the demand nodes and it is quantified in terms of the steady-state probability distribution of the product delivered to the demand nodes.

In extreme synthesis, the algorithm requires as inputs:

- the Hierarchical Graph that allows representing the origin-destination paths and the corresponding arcs in the hierarchical levels (see Section 2);
- the steady state probabilities of transition between the different arc states (i.e., capacities) $\zeta_{comp,i}$, $i = \{1, 2, \ldots, NS_{comp}\}$, $comp = \{1, 2, \ldots, N\}$;
- the vector $\{D_{dem}\}$, $dem \in \{1, \ldots, N\}$, of product required by demand nodes;
- the importance (relevance) of the demand nodes (see Section 2).

The output of the algorithm is represented by the steady state probability distributions of the product delivered to the demand nodes.

The assumptions made are the following:

- only unidirectional relationships are considered;
- the flow in the network is limited by the arcs capacities (i.e., capacitated networks are considered);
- the flow in the network satisfies the flow conservation principle, i.e., no flow in arcs will decrease or increase during flow transmission [Ford and Fulkerson, 1962];
- the capacities of the arcs are statistically independent, i.e., the state of an arc does not depend on the state of other arcs.

For clarification purposes, we describe the procedure with reference to the simple example of Figure 3, where two interconnected systems, $S^{(1)}$ and $S^{(2)}$, are shown. The N = 5 components are: $S_1^{(1)}_S_2^{(1)}$, $S_2^{(1)}_S_3^{(1)}$, $S_2^{(1)}_S_1^{(2)}$, $S_1^{(2)}_S_2^{(2)}$ and $S_2^{(2)}_S_3^{(2)}$. The input component is arc $S_1^{(1)}_S_2^{(1)}$ that serves five demand nodes (i.e., the goals), $S_2^{(1)}$, $S_3^{(1)}$, $S_1^{(2)}$, $S_2^{(2)}$ and $S_3^{(2)}$, explicitly represented at the top of the diagram.

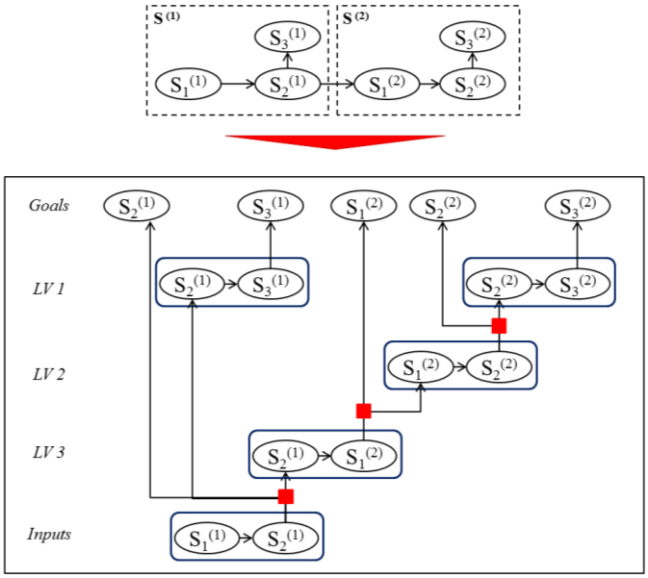


Figure 3: Hierarchical Graph of a generic example taken as reference to illustrate the algorithm; LV: Level.

The evaluation is carried out from the bottom to the top of the hierarchy and consists of the following steps:

1) Determine one possible system configuration by sampling the capacity of the arcs, $\zeta_{comp,i}$, $i \in \{1, 2, \ldots, NS_{comp}\}$, $comp = \{1, \ldots, N\}$, from the corresponding steady state probability distributions;
2) Identify the minimum arc capacity ($mpath_{dem}$, $dem \in \{1, \ldots, N\}$) for each origin-destination path: this capacity corresponds to the maximum product that can be delivered to the corresponding demand node *dem*, $dem \in \{1, \ldots, N\}$; for example, in Figure 4 the minimum arc capacity for the path from $S_1^{(1)}$ to $S_3^{(1)}$ is the minimum among the capacities of arcs $S_1^{(1)}_S_2^{(1)}$ and $S_2^{(1)}_S_3^{(1)}$, connecting $S_1^{(1)}$ and $S_3^{(1)}$;
3) Set the input (*inp*) to the network equal to the capacity of the input arc, i.e., $inp = \zeta_{comp,i}$, where $i \in \{1, 2, \ldots, NS_{comp}\}$ and *comp* is the index of the input arc (in the example of Figure 3, the input arc is $S_1^{(1)}_S_2^{(1)}$);
4) If the input is zero ($inp = 0$), no product can be delivered to the demand nodes: $EP_{dem} = 0$ for all $dem \in \{1, \ldots, N\}$; otherwise, estimate the optimal flows $\{EP_{dem}\}$ that can be delivered to the demand nodes by the following steps:
    a. Estimate the vector $\{EP_{dem}\}$ of optimal flows to the demand nodes taking into account i) the importance of the demand nodes and ii) the minimum capacity of each path ($mpath_{dem}$, $dem \in \{1, \ldots, N\}$) that limits the quantity of product that can be delivered to the demand nodes (Figure 5, top). For example, referring to

Figure 3, let us consider a proportional importance of the demand nodes (see Section 2) and assume that the ratio of importance of $S_2^{(2)}$ and $S_3^{(2)}$ is 0.2 and 0.3, respectively. According to the importance criterion considered and assuming a total input of 100 units, we assign 20 units to $S_2^{(2)}$ and 30 units to $S_3^{(2)}$. On the contrary, if the minimum capacity of the path to $S_3^{(2)}$ were lower than 30 units (say, $mpath_{S3(2)}$ = 10 units), $S_3^{(2)}$ would receive at most a quantity equal to $mpath_{S3(2)}$ (i.e., 10 units) and the surplus quantity in exceedance (i.e., 20 units) would be distributed to other nodes (see the following steps 4 b. and 4 c.).

b. Initialize to zero an auxiliary variable *surp* (i.e., *surp* = 0). This variable is used to quantify the surplus, i.e., the amount of product that cannot be allocated in the network due to arc capacity constraints (i.e., due to the bottlenecks of the infrastructure).

c. Check if the capacities of the links, $\zeta_{comp,i}$, $i \in \{1, 2, \ldots, NS_{comp}\}$, can support the sum of the estimated optimal products to the corresponding demand nodes, (*dem*) computed at the previous step 4 a. Such evaluation is performed from the bottom to the top of the diagram. If the sum of the estimated optimal product to the nodes served by a link is higher than its capacity, save the exceeding amount (Δ) in the auxiliary variable *surp* (i.e., *surp* = *surp* + Δ) and compute the optimal partition just for the nodes that are supplied by that link, considering as input the corresponding arc capacity $inp = \zeta_{comp,i}$, where $i \in \{1, 2, \ldots, NS_{comp}\}$ and *comp* is the link under analysis (Figure 5, middle).

d. Create a "new" graph, where the "new" capacities of all the arcs are updated on the basis of the quantity of product, $\{EP_{dem}\}$, that has been effectively allocated at step 4 c. In particular, the arc capacities are reduced by the total quantity of product that they have already supplied to the corresponding demand nodes (Figure 5, bottom).

e. Compute again the minimum arc capacity for each path of the "new" graph (as in step 2) to evaluate the new maximum product that can reach the corresponding demand nodes (Figure 5, bottom).

f. Update the demands $\{D_{dem}\}$, $dem \in \{1, \ldots, N\}$, reducing them by the quantity $\{EP_{dem}\}$ that has been already allocated at step 4.c (Figure 5, bottom).

g. Set the input *inp* equal to the auxiliary variable *surp* (*inp* = *surp*) and repeat step 4 until *surp* > 0 and the minimum ("new") arc capacity is not zero for at least one path. When one of these conditions is verified, the final vector $\{EP_{dem}\}$ of the optimal product that can be delivered to the demand nodes is determined (Figure 5, bottom).

The procedure above is repeated a large number of times (e.g., 10000) for many different MC-sampled values of the arc capacities and the probability distribution of the product delivered at steady state to each demand node is obtained.

A final remark is in order with respect to the use of Monte Carlo simulation for the propagation of uncertainty of the arc capacities. Real-world complex infrastructure systems include a large number of components, many of which have uncertain (i.e., aleatory, stochastic) performance. This high-dimensionality constitutes a major challenge for the extensive propagation of (aleatory) uncertainty through the corresponding mathematical models. Classical analytical or numerical schemes are not suitable. On the other hand, stochastic simulation methods (e.g. Monte Carlo simulation), based on the repeated random sampling of possible component states and the evaluation of the system model for the different states sampled, offer a feasible means: such approaches are, therefore, adopted in the present work.

Concerning, instead, the issue of uncertainty representation, the classical way to address the uncertainty due to randomness is to collect data about the random phenomenon of interest and perform a statistical analysis to identify the probability distribution that best captures the variability of the available data. In a frequentist view, the available data may be interpreted as observable random realizations of an underlying, repeatable probabilistic model (e.g., a probability distribution) representing the stochastic (aleatory) phenomenon of interest which can be approximated with increasing precision by the analyst as the size of the available data set increases [Apostolakis, 1990].

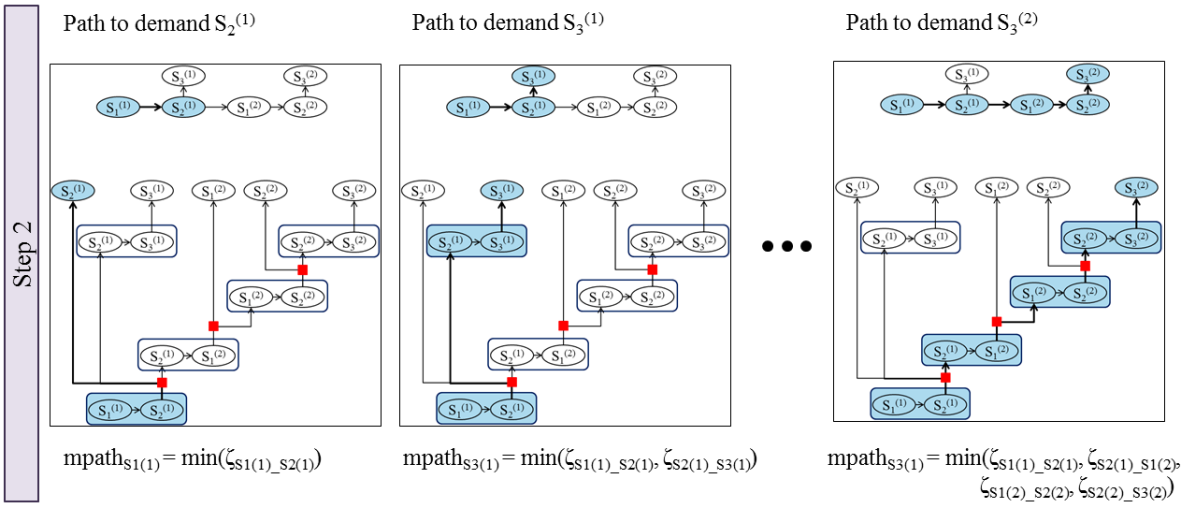


Figure 4: Exemplification of step 2 of the algorithm with respect to the example proposed in Figure 3.

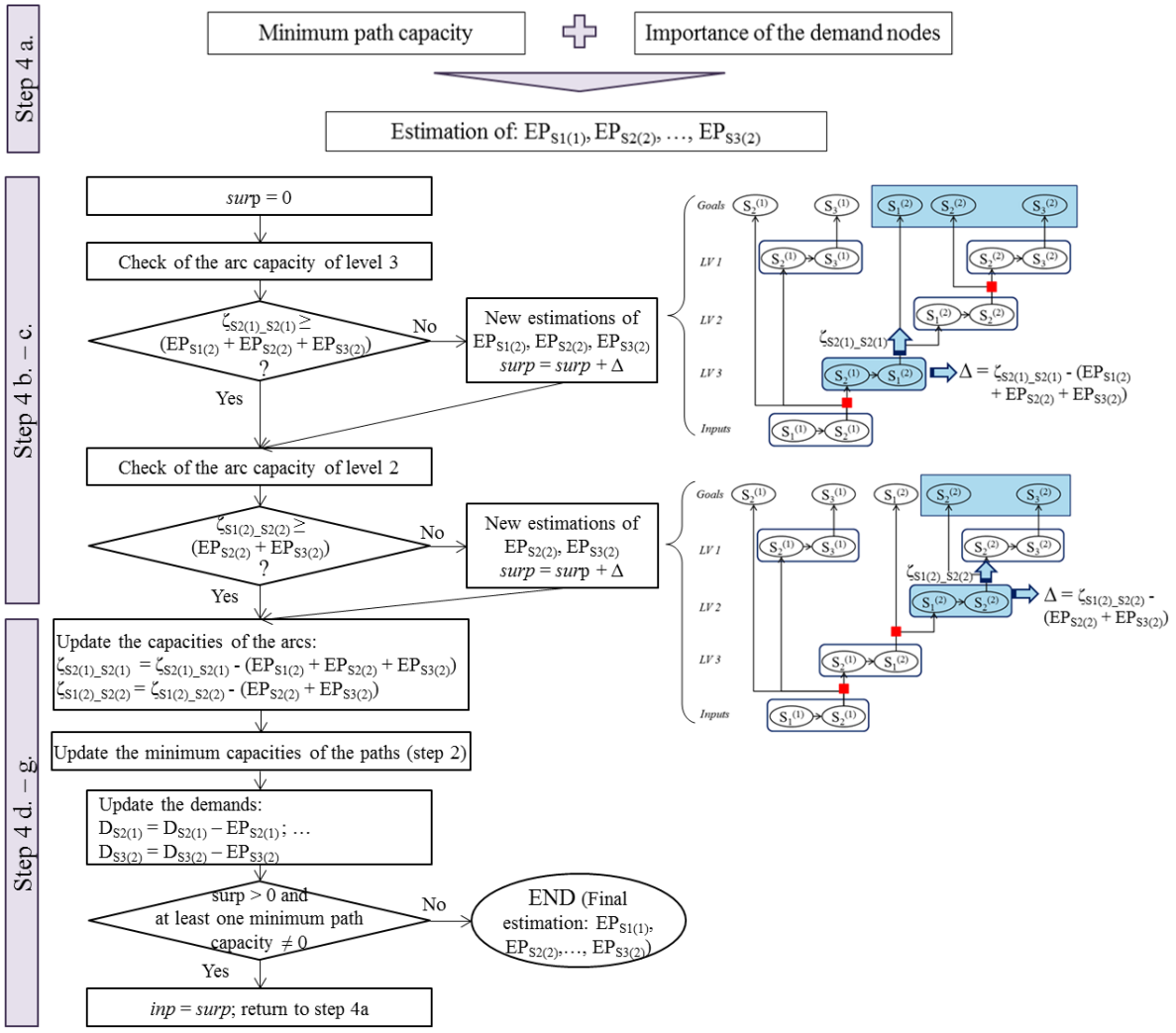


Figure 5: Exemplification of step 4 of the algorithm with respect to the example proposed in Figure 3.

It is worth noting that the procedure proposed is based on several iterative estimations of the vector $\{EP_{dem}\}$, obtained by repeating steps 4 a. – g. from the bottom to the top of the hierarchy: in the very first iteration, the system configuration is the one sampled at step 1. and the input product corresponds to the capacity of the input arc; then, at each loop a "new" graph is considered where (i) the new product input value is represented by the surplus (*surp*), i.e., the amount of product that has not been allocated in the network at the previous iteration, (ii) the "new arc capacities are reduced by the total amount of product they have already supplied at the previous iteration and (iii) the "new" demands are scaled by the quantity already allocated in the previous iteration.

Finally, it is worth mentioning that a drawback of the Hierarchical Graph representation proposed may be represented by its difficult applicability to large networks since *all* the origin-destination paths have to be identified and the bottlenecks of each path have to be spotted out. To overcome this limitation, we propose to pre-process the infrastructure system by means of a clustering algorithm to reduce the systems dimension by "collapsing" many components in few representative clusters and, then, apply the Hierarchical Graph to the "clustered" infrastructure. The general concepts underlying the pre-processing phase based on clustering is discussed in the following Section 3.2; a discussion on the advantages and limitations of the approach is given in Section 3.3.

3.2. Combination of Hierarchical Graph representation and clustering for large-sized critical infrastructures

In order to manage large-sized CIs, it is useful to resort to clustering techniques to reduce the complexity and dimension of the system. Actually, complex systems are characterized by modularity that allows identifying groups of elements highly interconnected within them and sparsely linked to other dense groups in the network [Sales-Pardo et al., 2007; Porter et al., 2009]. In addition, several studies show that networks often exhibit hierarchical organization [Ravasz and Barabasi, 2003; Clauset et al., 2008]. These features lead to combine hierarchical modeling and clustering analysis to represent complex networks.

Cluster analysis aims at identifying groups of "similar behavior" in their data. For illustration purposes, refer to the simple example of Figure 6, left, where the original components of a network, namely $S_1^{(1)}$, $S_2^{(1)}$, ..., $S_{16}^{(1)}$, are shown. According to some features of interest (e.g., proximity), the components can be clustered in groups of "similar characteristics": in the

example proposed, four clusters, C_1, …, C_4, are identified (dotted oval shape in Figure 6, left). Then, a less complex analysis can be performed on the new fictitious, artificial (i.e., clustered) network, composed just by the identified clusters (Figure 6, right).

The cluster analysis can be carried out at different levels of details: an artificial network with a high number of clusters is closer to the original one and, thus, it is more detailed (i.e., it carries more information) than one with a small number of clusters. The system can be clustered at different levels of details, which allows building a hierarchical[1] clustering representation where the different hierarchical levels correspond to the different levels of detail of the analysis [Fang and Zio, 2013].

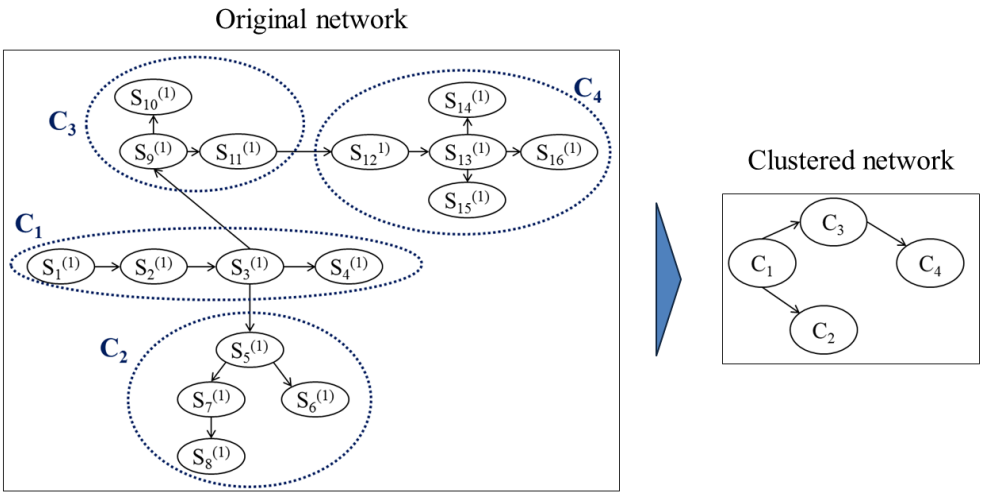


Figure 6: Exemplification of the clustering procedure.

In order to reduce the size of the infrastructure under analysis, in this work we have applied the unsupervised spectral clustering algorithm (USCA) and the Fuzzy c-means (FCM) clustering as in [Fang and Zio, 2013], for its simplicity of implementation and its effectiveness in providing satisfactory results. The operative steps of the procedure are not reported here for brevity sake; the interested reader is referred to [von Luxburg, 2007; Fang and Zio, 2013]. The result of clustering is sensitive to the similarity function which defines the proximity of the nodes in the network; in this respect, the not trivial task is to ensure "meaningful" local neighborhoods [von Luxburg, 2007; Fang and Zio, 2013]. As similarity measure to perform the clustering, we have adopted the adjacency matrix, i.e., two nodes are similar if they are linked

[1] Notice that in this case the term "hierarchical" refers to the level of detail of the clustering and not to the levels of the Hierarchical Graph representation that instead correspond to the number of demands served by a given arc of the network.

directly, otherwise they are not. Different similarity definitions can lead to different cluster partitioning of the network. In this case, the use of the adjacency affinity, which considers only local direct connection information, is justified by the fact that it allows obtaining radial networks for each (clustering) hierarchical level, a condition needed to employ the Hierarchical Graph (see Section 2).

By recursively operating the USCA and the FCM a hierarchical structure of the system can be obtained, where, at the top of the hierarchy, the system is represented by just one node and, at the bottom, by the whole original network. In the middle, each hierarchical level represents a different degree of resolution of the systems (from the top to the bottom it can be seen an increasing quantity of information about the local connectivity) and corresponds to artificial networks that include artificial nodes and links (these last ones are composed by those original network links connecting – in parallel – the original nodes in the clusters forming the artificial nodes).

By applying the USCA and the FCM, an artificial (fictitious) network composed by k_L clusters, $C_1^{(L)}$, …, $C_{k_L}^{(L)}$, is produced at each (clustering) hierarchical level L. Notice that the last level of the clustering hierarchy coincides with the real SoS, i.e., the corresponding clusters coincide with the actual/original/real nodes of the SoS. The clustering is performed on the entire network except for the input nodes that are left out (only one generation node is considered in the application of the present work). For illustration purposes, Figure 7 depicts a sketch of the decomposition in five (clustering) hierarchical levels of a SoS with one input node, $S_1^{(1)}$; level 1 of the hierarchy is, then, composed by two nodes: the input, $S_1^{(1)}$, and the rest of the system "condensed" in cluster $C_1^{(1)}$. The clustering algorithm allows a new analysis at hierarchical level 2 and it decomposes cluster $C_1^{(1)}$ of hierarchical level 1 into two clusters $C_1^{(2)}$ and $C_2^{(2)}$. At this point, if we want to increase the level of refinement of the analysis we can use the algorithm to further split clusters $C_1^{(2)}$ and $C_2^{(2)}$. In the example of Figure 7, this results in the decomposition of cluster $C_1^{(2)}$ into three clusters ($C_1^{(3)}$, $C_2^{(3)}$ and $C_3^{(3)}$) and cluster $C_2^{(2)}$ into five clusters ($C_4^{(3)}$, $C_5^{(3)}$, $C_6^{(3)}$, $C_7^{(3)}$ and $C_8^{(3)}$). The Hierarchical Graph representation of the decomposed system at level 3 is also shown on the right.

A cluster k is characterized by its demand D_k that is the sum of the demands of the real nodes it contains: for example, cluster $C_1^{(4)}$ of Figure 7 has demand equal to the sum of the demands of nodes $S_3^{(1)}$, $S_6^{(1)}$ and $S_7^{(1)}$.

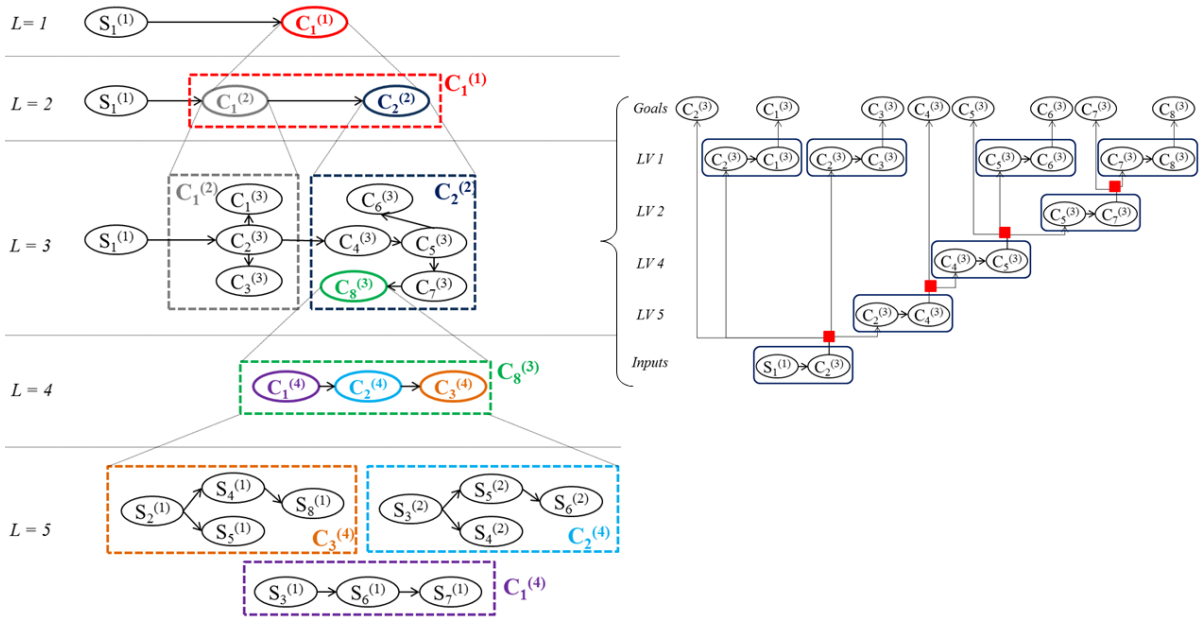


Figure 7: Left: sketch of the decomposition of a system in five hierarchical levels (L), where the last one (L = 5) coincides with the actual nodes of the system; right: Hierarchical Graph of the corresponding hierarchical level 3; LV: Level of the Hierarchical Graph.

For a given clustering hierarchical level L, the quantitative evaluation of the performance of the "artificial" clustered system is carried out as illustrated in Section 3.1 with reference to an indicator that represents the "global" state of the clusters $C_1^{(L)}$, …, $C_{k_L}^{(L)}$ of level L as a "synthesis" of the real capacity $\zeta_{comp,i}$, $i \in \{1, 2, \ldots, NS_{comp}\}$, of the arcs, $comp = 1, \ldots, N$, contained in the cluster itself. Actually, a measure of the cluster state is needed to approximately estimate the quantity of product that a cluster can receive and deliver to other clusters.

To represent the state of a cluster k (i.e., its performance), we consider an indicator id_k based on the ratio of the expected capacity of cluster k at current and at nominal (optimal) conditions as follows:

$$id_k = \frac{\sum_{comp=1}^{n_k} w_{comp} * \zeta_{comp,i}}{\sum_{comp=1}^{n_k} w_{comp} * \zeta_{comp,NS_{comp}}}, \quad (1)$$

where $comp$ indicates the component (arc) of the original network, n_k is the number of arcs inside cluster k, $\zeta_{comp,i}$, $i \in \{1, 2, \ldots, NS_{comp}\}$, is the current (i.e., actual / sampled) state of the component $comp$, $\zeta_{comp,NS_{comp}}$ is the maximum capacity of the arc $comp$ and w_i is the weight associated to the capacity of the arc $comp$. The weight w_{comp} is computed as the ratio between the capacity of the arc $comp$ and the sum of the maximum capacities of all the arcs of the network, i.e., $w_{comp} = \zeta_{comp,i} / \sum_{comp=1}^{N} \zeta_{comp,NS_{comp}}$, $i \in \{1, 2, \ldots, NS_{comp}\}$ and gives an idea of the weight of the arc in the entire network. The index of the cluster state (id_k), takes value between 0 and 1.

Notice that the state of a cluster affects the cluster itself and the connected clusters, since the cluster is both a *fictitious load node* (which should provide itself with the required amount of product) and a *fictitious transmission node* (which should transmit the product to the other connected clusters). The top of Figure 8 shows two clusters, C_1 and C_2, supplied by the input source $S_1^{(1)}$: cluster C_2 is both a load and a transmission node, since on one side it contains five demand nodes ($S_2^{(1)}$, $S_3^{(1)}$, $S_4^{(1)}$, $S_5^{(1)}$ and $S_6^{(1)}$ in Figure 8, bottom) and on the other side it is required to transmit the product to cluster C_2. In particular, the product from input source $S_1^{(1)}$ has to pass through two arcs ($S_2^{(1)}_S_5^{(1)}$ and $S_5^{(1)}_S_6^{(1)}$) contained in C_1 to reach cluster C_2: if their capacities decrease, then the flow to nodes $S_5^{(1)}$ and $S_6^{(1)}$ (i.e., to the cluster C_1) and to nodes $S_7^{(1)}$, $S_8^{(1)}$ and $S_9^{(1)}$ (i.e., to the cluster C_2) is reduced.

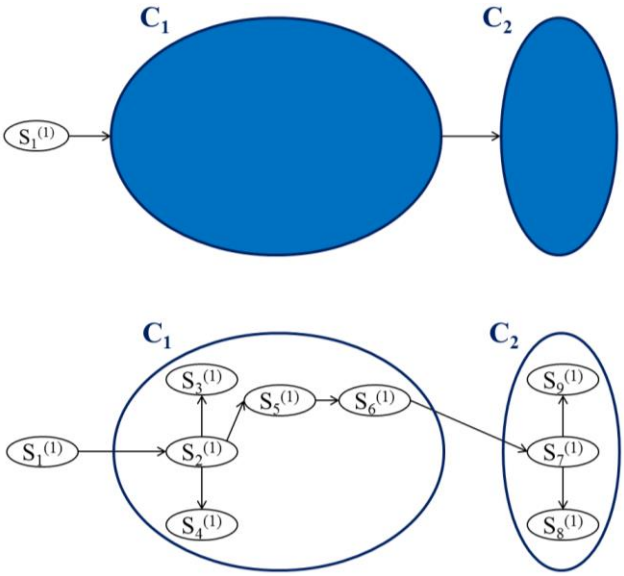


Figure 8: Top: artificial system composed by two clusters C_1 and C_2 supplied by one input node $S_1^{(1)}$. Bottom: illustration of the real nodes inside the fictitious clusters: two arcs of C_1 are needed to supply C_2.

Thus, when the "capacity" of a cluster decreases, the consequence is twofold: the cluster cannot satisfy itself (i.e., the demands of the demand nodes it contains) and also the connected clusters. In order to take into account the “twofold” reduction of performance, we “artificially reduce” the amount of product that can be given to the cluster itself and that can be delivered to the connected clusters, by multiplying by the indicator of the state of the cluster, id_k i) the maximum demand D_k that it requires and ii) the maximum capacities, $\zeta_{comp,NS_{comp}}$, of the arcs (*comp*) that link the output clusters, respectively.

3.3. Discussion on the Monte Carlo simulation and Hierarchical Graph modeling for critical infrastructures analysis

The Hierarchical Graph representation supports both a qualitative and quantitative analysis of CIs. The former allows identifying critical arcs; the latter allows determining the partitioning of the flow in the network and computing a measure of system performance, e.g., the steady state probability distribution of the product delivered to demand nodes. However, the representation does not make evident how the partition of the flow is carried out, since the algorithm used for its computation is recursive.

The Hierarchical Graph representation can aid capturing the structural and dynamic complexities of CIs as follows. It can represent components of different technological domains (e.g., physical, hard components – like pipelines – and soft components – like information and communication systems) and their interconnections by dependences, distributed over large geographic areas. Different types of service and product demands can be included, e.g., supply of power to critical buildings (hospitals, industrial plant) and residential areas, supply of water for daily use or for emergency situations (fire-fighting), etc. However, a limitation is that the interdependences (bidirectional relationships) cannot be represented, since the representation is based only on unidirectional flow. Also, the Hierarchical Graph becomes large when the scale and dimensionality of connectivity is large, and the identification of all the origin-destination paths may become computationally prohibitive. To overcome this limitation, the clustering technique is employed to reduce the size of the system, obtaining a smaller radial artificial (clustered) network, i.e., without loops. This allows a preliminary analysis of system robustness to be performed in short time, followed by a more detailed analysis on the most relevant parts of the system.

With respect to the dynamic complexity, the quantitative analysis of the system robustness supported by the Hierarchical Graph gives indications on how much the system is able to adapt to failure scenarios for continuing to efficiently deliver the products and services through the network, according to different criteria of importance. It is to be noted that the Hierarchical Graph is not well suited to represent growing systems, since the addition of a new component to the system methods may change its entire structure (i.e., it lacks of flexibility).

The overall approach is based on the assumption that capacitated networks (i.e., networks where the flows of product in the links are limited by arcs capacities) provide an appropriate representation of infrastructure performance, e.g., robustness. This assumption does not represent a limitation for the analysis and, indeed, several works in the scientific literature are based on the same hypothesis. Actually, in many real flow networks such as electric power networks, gas distribution networks, computer networks, urban traffic networks, logistics networks, each arc has several possible capacities or states due to failures, partial failures, maintenance, etc. Works related to capacitated networks can be found in [Lin et al., 1995; Ramirez-Marquez and Coit, 2007; Yeh, 2008; Ramirez-Marquez and Rocco, 2009; Lin, 2010; Lin and Yeh, 2011] [Nozick et al., 2005].

Table 1 summarizes the advantages and limitations of the approach."

Table 1: Advantages and limitations of the Hierarchical Graph representation

Advantages
• Both qualitative analysis (identification of critical elements) and quantitative analysis (computation of a measure of system performance) can be performed.
• The complexity of CIs can be handled.
• Different types of demands (daily water consumption, fire-fighting consumption, power consumption, etc.), with different importance (priority), can be taken into account.
• A clustering algorithm can be employed to reduce the model dimension.
• A multi-state model can be adopted for a more realistic representation of the system states.
Limitations
• Interdependences (bidirectional relationships) cannot be included in the representation.
• Difficult applicability to large networks (but clustering can reduce this problem).
• The representation itself shows the physical constraints that affect the product and service distribution, but it does not illustrate how the partition of the products and service is carried out, since the algorithm used for its computation is recursive.
• The representation is not flexible, i.e., the addition of a new component may change the entire structure.

4. Applications

In this Section, we apply the proposed Hierarchical Graph representation to evaluate the robustness of two case studies (hereafter referred to as "A" and "B"): case study A (Section 4.1) consists of two interdependent infrastructures (gas and electric power networks) and a supervisory control and data acquisition (SCADA) system connected to the gas network; case study B (Section 4.2) considers an electric power distribution network adapted from the IEEE 123 node test feeders. As mentioned before, the robustness addressed in this paper refers to the capability of the CIs to resist to failures or partial failures of the CIs components (arcs) assuring

the required level (or a high level) of supply of goods or services. It is measured as the steady-state probability distributions of the product (e.g., gas and/or electricity) delivered to the demand nodes.

In both cases studies, we adopt a multi-state model to account for the different degrees of degradation of the components and we describe the state transition processes by Markov and semi-Markov modeling.

4.1. Case study A

The case study is taken from [Nozick et al., 2005] and deals with two interconnected infrastructures, i.e., a natural gas distribution network and an electricity generation/distribution network (Figure 9, solid and dashed lines, respectively). The gas distribution network is supported by a SCADA system (Figure 9, dotted lines). The objective of this interconnected system of systems (SoS) is to provide the necessary amount of gas and electricity (hereafter also called "product") to four demand nodes (end-nodes), namely *D1* and *D2* (gas), and *L1* and *L2* (electricity).

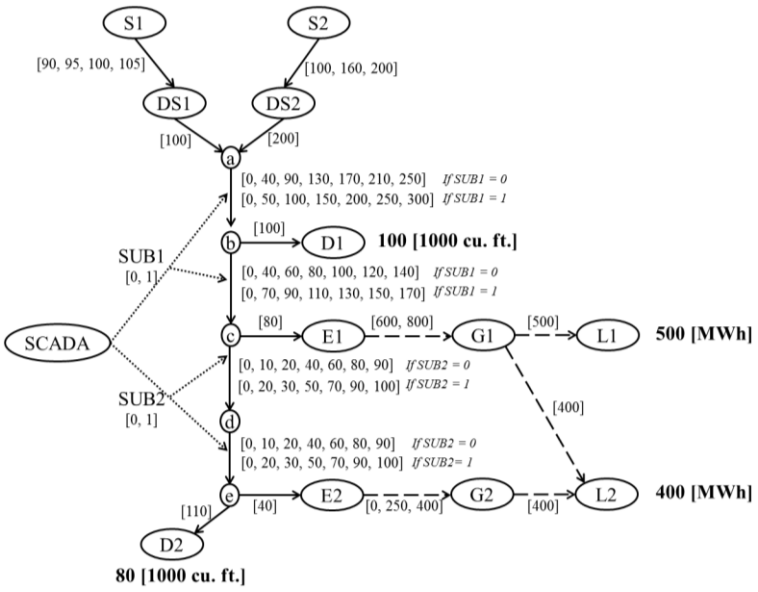


Figure 9: Interdependent gas (solid lines) and electric (dashed lines) infrastructures and SCADA system (dotted lines) [Nozick et al., 2005]. The possible states (i.e., capacities) of the arcs are given in square brackets; the quantities of product demanded by end-nodes D1, D2, L1, L2 are reported in bold.

The gas distribution network, supplied by two sources of gas (namely, *S1* and *S2*, connected to the network by arcs *S1_DS1* and *S2_DS2*, respectively), provides gas to the end-nodes *D1* and *D2* and to two nodes of the electricity network (*E1* and *E2*). Once the gas enters into nodes *E1* and *E2*, it is transformed into electrical energy that flows through arcs *E1_G1* and *E2_G2*

(representing the electric power generation stations) to supply the end-nodes of electricity (*L1* and *L2*); notice that demand *L2* can be supplied by both electrical generations *E1_G1* and *E2_G2*. The assumption is made that the gas-electricity transformation occurs with a constant coefficient, i.e., 100 cu. ft. of natural gas produces 1 MWh of electricity [Nozick et al., 2005]. A SCADA system controls the gas flow through arcs *a_b*, *b_c*, *c_d* and *d_e*. It is assumed that: i) the SCADA has two core subsystems controlling different sets of arcs (in particular, the first one – SUB1 – refers to links *a_b* and *b_c*, whereas the second one – SUB2 – controls arcs *c_d* and *d_e*); ii) the SCADA is always provided with electric power [Nozick et al., 2005].

The capacities of the arcs of the gas and electricity networks (determining the maximum flows of gas or electricity supported by the arcs) can be deterministic (i.e., fixed constant values) or stochastic (i.e., randomly evolving in time) (Figure 9, values in the square brackets). The stochastic capacities give rise to a multi-state model that reflects the possibly different degrees of degradation of the arcs. On the contrary, the SCADA system state is defined by a binary random variable, whose values 1 and 0 represent its complete and partial functioning, respectively. For example, when the state of the SCADA subsystem SUB1 (controlling arcs *a_b* and *b_c*) is 0, the capacity of these arcs decreases because of the incorrect information provided by the SCADA subsystem (even if the arcs are not subject to a direct damage). On the basis of the two states of the SCADA subsystems, two different vectors of capacities are identified for each arc *a_b*, *b_c*, *c_d* and *d_e*: as illustrated in Figure 9, the first vector is used when the corresponding SCADA subsystem is in state 0, whereas the second one is employed when the SCADA subsystem is in state 1.

Changes in the arc capacities are due to random degradations or recovery actions. The state transitions over time are modeled by Markov and semi-Markov processes as in [Nozick et al., 2005]. Semi-Markov processes are adopted to model the evolution of the capacities of the gas supply links (*S1_DS1* and *S2_DS2*), whereas Markov processes are used for all the others arcs. Both Markov and semi-Markov processes for a generic component '*comp*' are defined by a transition probability matrix $\underline{\underline{P}}_{comp} = \{p_{ij} : i, j = 1, 2, ..., S_{comp}\}$, where p_{ij} is the one-step probability of transition from state *i* to state *j*. In addition, the semi-Markov processes are characterized by continuous probability distributions for the holding time T^{ij}_{comp}, i.e., for the time of residence of a component in state *i* before performing a transition to state *j*.

The steady-state probability vectors for a generic component '*comp*' described by a Markov process is computed as $\underline{\Pi}^{comp} = \underline{\Pi}^{comp} \cdot \underline{\underline{P}}^{comp}$ [Zio, 2009]. For a semi-Markov process, this equation is weighted by the expected time of residence τ^i in a given state *i* before performing a transition, i.e., as $\xi^{comp,i} = \Pi^{comp,i} \cdot \tau^i / \sum_{j=1}^{S_{comp}} \Pi^{comp,j} \cdot \tau^j$, $i = 1, \ldots, NS_{comp}$ [Barry, 1995].

Figure 10 reports the steady-state probability vectors of the arcs of the system of Figure 9 assuming the state transition probabilities given in [Nozick et al., 2005].

$\underline{\Pi}^{S1_DS1}$ =

1	0.0001
2	0.0002
3	0.5001
4	0.4996

$\underline{\Pi}^{S2_DS2}$ =

1	0.0033
2	0.1703
3	0.8264

$\underline{\Pi}^{SCADA}$ =

1	0.0042
2	0.0012
3	0.0012
4	0.9934

$\underline{\Pi}^{a_b,\ b_c,\ c_d,\ d_e}$ =

1	0.0022
2	0.0045
3	0.0022
4	0.0065
5	0.0038
6	0.0052
7	0.9756

$\underline{\Pi}^{E2_G2}$ =

1	0.0006
2	0.0007
3	0.9987

$\underline{\Pi}^{E1_G1}$ =

1	0.0011
2	0.9989

Figure 10: Steady state probability vectors for the states of the components of the case study A (comp = S1_DS1, S2_DS2, a_b, b_c, c_d, d_e, E2_G2, E1_G1, SCADA).

4.1.1. Hierarchical Graph representation for the system of case study A

In Figure 11, the Hierarchical Graph of the system of case study A (Section 4.1) is illustrated. The injection of product (i.e., gas) in the SoS is made through arcs *S1_DS1* and *S2_DS2*, located at the bottom of the diagram (Section 2). Since both arcs enter in node *a*, also the following links *DS1_a* and *DS2_a* are considered part of the inputs and reported at the bottom of the hierarchy. Four demand nodes, i.e., *D1*, *D2* (gas) and *L1*, *L2* (electricity), represent the goals of the analysis and they are explicitly located at the top of the diagram. The graph presents four hierarchical levels: in level 4, arc *a_b*, is reported since it supplies all the four demand nodes; in level 3, arc *b_c* is depicted, since it serves three demand nodes (i.e., *L1*, *L2* and *D2*); in level 2, arcs *c_E1*, *E1_G1*, *c_d* and *d_e* are considered, since they supply two demand nodes: in particular, arcs *c_E1* and *E1_G1* supply *L1* and *L2*, whereas arcs *c_d* and *d_e* serve *L2* and *D2*; in level 1, there are the remaining arcs that are related just to one demand node: for example, *e_E2* serves only node *L2*. The influence of the SCADA subsystem SUB1 on the arcs *a_b* and *b_c* and of the SCADA subsystem SUB2 on the arcs *c_d* and *d_e* is illustrated in the trapezoidal frames under the corresponding arcs.

For illustration purposes, three different importance criteria for the demand nodes are considered (see Section 2):

- sequential importance: the demand nodes are ranked on the basis of the proximity to the sources: 1) *D1* (the most important), 2) *L1*, 3) *L2* and 4) *D2* (the least important).
- proportional importance: the demand nodes are satisfied on the basis of their demands (the nodes that require more product are given higher priority). Since *D1* and *D2* require 100 and 80 cu. ft. of gas and *L1* and *L2* need 500 and 400 MWh (equivalent to 50 and 40 cu. ft.), the resulting importance ranking of the demand nodes is: 1) *D1* (the most important), 2) *D2*, 3) *L1* and 4) *L2* (the least important).
- equal importance: the product is divided equally among four demand nodes.

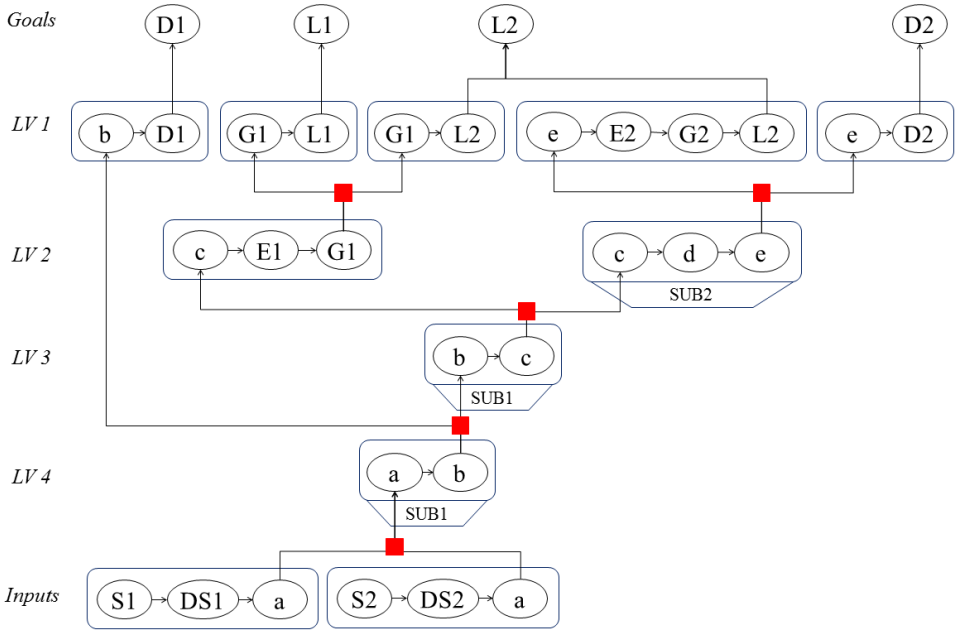


Figure 11: Hierarchical Graph of the SoS depicted in Figure 9 (case study A); LV: Level.

4.1.2. Results of case study A

The robustness of the case study A is given by steady state probability distributions of the product delivered to the four demand nodes *D1*, *L1*, *L2* and *D2*, by applying the operative steps illustrated in Section 3.1. From these distributions two quantities can be analyzed as illustrated in Table 2:

(i) Steady state probabilities of delivering the (maximum, optimal) required product to the demand nodes (top);

(ii) Steady state probability of delivering a quantity of product exceeding 90% of the corresponding demands (bottom).

Table 2: Steady state probabilities of (i) delivering the (maximum, optimal) required product to the demand nodes (top) and (ii) delivering a quantity of product exceeding 90% of the corresponding demands (bottom).

Importance criterion	**P(D1 = 100 [1000 cu. ft.])**	**P(L1 = 50 [10 MWh])**	**P(L2 = 40 [10 MWh])**	**P(D2 = 80 [1000 cu. ft.])**
Sequential	0.9927	0.9867	0.9723	0.7526
Proportional	0.8195	0.7563	0.7563	0.7568
Equal	0.8205	0.9306	0.9678	0.9063

Importance criterion	**P(D1 > 90 [1000 cu. ft.])**	**P(L1 > 45 [10 MWh])**	**P(L2 > 36 [10 MWh])**	**P(D2 > 72 [1000 cu. ft.])**
Sequential	0.9927	0.9867	0.9723	0.7526
Proportional	0.9778	0.9155	0.9507	0.9160
Equal	0.9717	0.9482	0.9816	0.9063

As expected, in the case of *sequential importance* of the demand nodes, *D1* is the demand node most satisfied, whereas *D2* is the least served: the corresponding probabilities of being completely satisfied are 0.9927 and 0.7526, respectively (Table 2, top). Differently, in the ranking produced by the *proportional importance* criterion, *D2* is more important than *L1* and *L2*. Thus, the probability of delivering the required product to *D2* should increase, whereas the probability of satisfying the last two demand nodes should decrease. In fact, the steady state probabilities of delivering the required maximum product to *L1* and *L2* decrease (Table 2, top). On the contrary, the probability P(D2 = 80 [1000 cu. ft.]) of delivering the maximum product to *D2* remains almost the same, since (i) the path needed to reach *D2* is affected by the uncertainty on the capacity of many arcs and (ii) *D2* is not the most important demand node. However, an effect of the proportional importance criterion on *D2* can be seen by analyzing the steady state probability P(D2 > 72 [1000 cu. ft.]) that the product given to *D2* exceeds 90% of its demand: actually, this value increases considerably from 0.7526 of the previous case (sequential importance, Table 2, top) to 0.9160 (proportional importance, Table 2, bottom). Finally, the criterion of *equal importance*, turns out to give preference to the nodes that have the lowest demands (this is expected since a lower demand has higher probability to be satisfied by an equal partition of the product). Actually, in this case the steady state probabilities of supplying *L1*, *L2* and *D2* at the optimum level increase with respect to the case of proportional importance (from 0.7563 to 0.9306 for *L1*, from 0.7563 to 0.9678 for *L2* and from 0.7568 to 0.9063 for *D2*, see Table 2 top); the probabilities to serve *D1* remain almost the same (around 0.82, see Table 2 top).

Changing the importance criterion induces changes in the ranking of the demand nodes that receive less product. However, it can be noticed that the demand node *D2* is the most difficult to supply to, even in the case of proportional importance where *D2* is the second most important demand node. The highest probability of supply of *D2* is 0.9160 while the others demand nodes can reach a probability of supply higher than 0.98. Then, to improve the robustness of the system the path from the sources to the demand node *D2* should be strengthened.

4.2. Case study B

Figure 12 shows the electric power distribution network here considered, as adapted from the IEEE 123 nodes test feeder [IEEE, 2000] in the sense that regulators, capacitors, switches and feeders with length equals to zero are neglected. With these simplifications, the network is composed of 114 nodes: 1 generation point (node 115) and 113 load/transmission nodes. Node 61 of the original IEEE 123 node test feeders is missing here, since after the removal of switches and transformers it turns out to be an end node with load equal to zero. The arcs (i.e., the feeders) connect different nodes and distribute the power through the network. In the analysis by Hierarchical Graph, we focus on the arcs (and not on the nodes), hereafter also called "components"; thus, the total number N of components is 113.

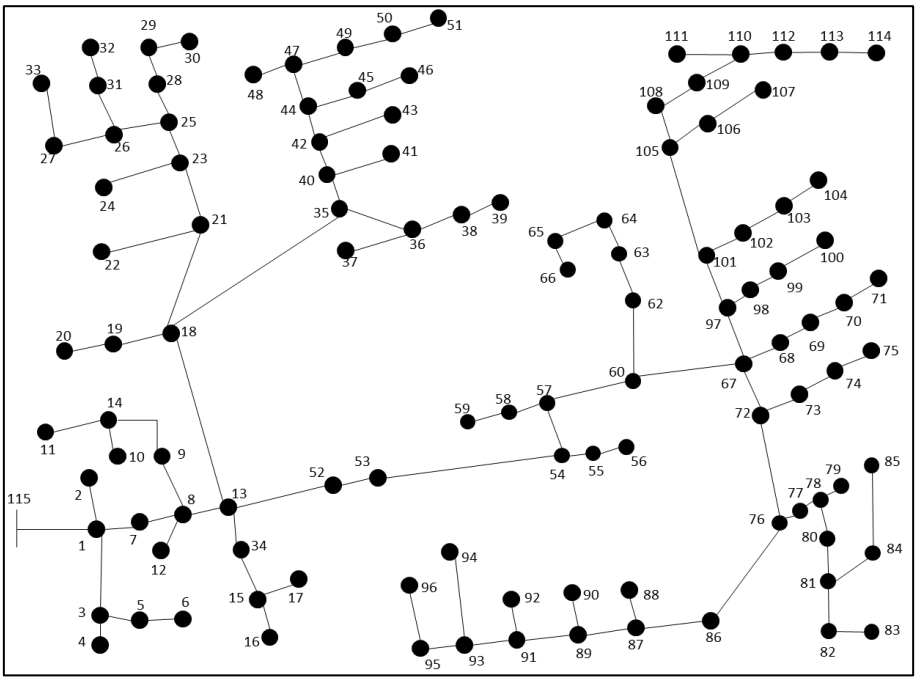


Figure 12: IEEE 123 node test feeders adapted to the purposes of the present analysis.

Four states, $i = 1, \ldots, 4$, characterize the capacity ($\zeta_{comp,i}$, $comp = 1, \ldots, N$) of the arcs: the first one ($i = 1$) represents the worst state and corresponds to the complete failure of the link, i.e., no

product can flow through it ($\zeta_{comp,1} = 0$, $comp = 1, \ldots, N$); the fourth one ($i = 4$) corresponds to the best state, i.e., to its nominal designed capacity. To obtain a reasonable value for the nominal capacity of the arcs useful for the purposes of the present analysis, we have solved the DC power flow equations (DCPF) that provide the physical flows of electricity through the lines given i) the total power injected in the network (assumed equal to the sum of all the powers required by the demand nodes), ii) the network topology (i.e., the adjacency matrix) and iii) the reactance of all the arcs [McCalley, 2012]; the nominal capacity of a link is set equal to the optimal power flow through that link.

In the Appendix, the demands of the network nodes provided by the IEEE database and the maximal capacity $\zeta_{comp,4}$ of the arcs ($comp = 1, \ldots, N$) obtained by the DC power flow equations, are given.

Changes in the arc capacities are due to random failures (as in the previous case study A): the state transitions over time are modeled as Markov processes.

4.2.1. Hierarchical Graph representation and clustering for case study B

An unsupervised spectral clustering algorithm (Section 3.2) is applied to the IEEE electric power distribution network of interest and identifies five hierarchical clustering levels: level 1 is composed by the generation node and one cluster representative of the whole network, whereas level 5 coincides with the original network. In Figures 13 – 15, the clusters generated at levels 2, 3 and 4 are reported, respectively: the corresponding Hierarchical Graph representation to levels 2 and 3 is also given for illustration purposes. Notice that level 2 is characterized by two clusters, level 3 by 8 clusters and level 4 by 29 clusters (besides the generation point): these clustered representations correspond to different possible “levels of detail” that the analyst may choose to study the network.

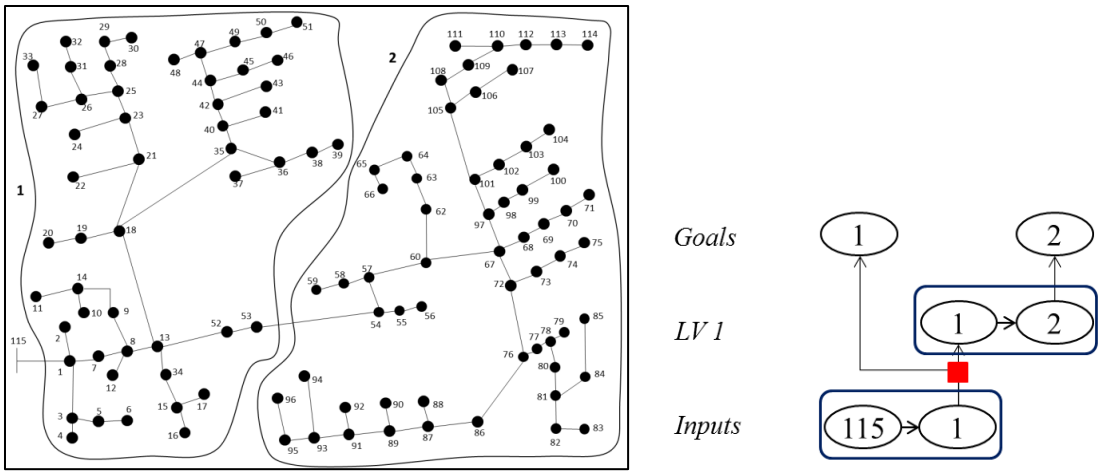


Figure 13: IEEE network clustered at hierarchical level 2 (left) together with the corresponding Hierarchical Graph representation (right); LV: Level.

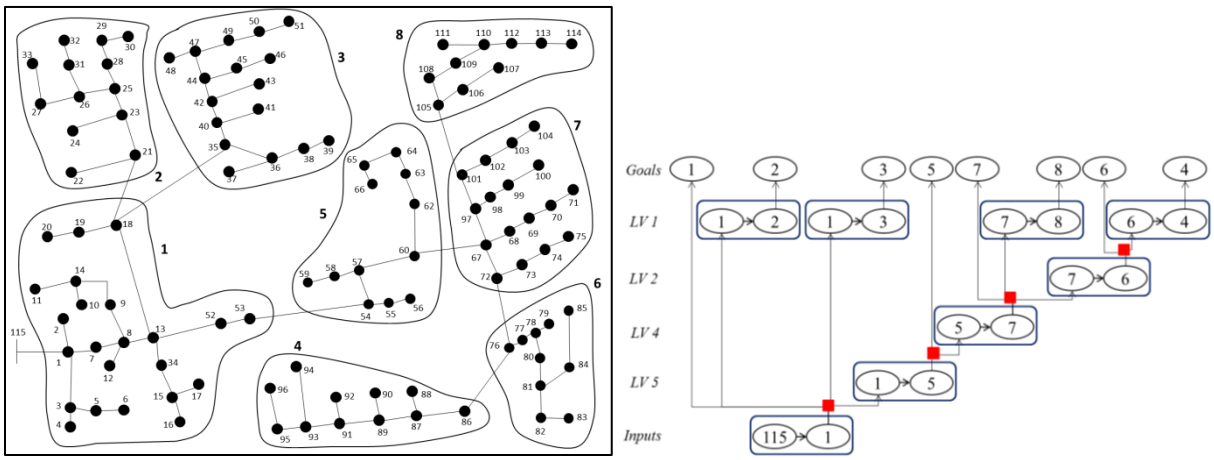


Figure 14: IEEE network clustered at hierarchical level 3 (left) together with the corresponding Hierarchical Graph representation (right); LV: Level.

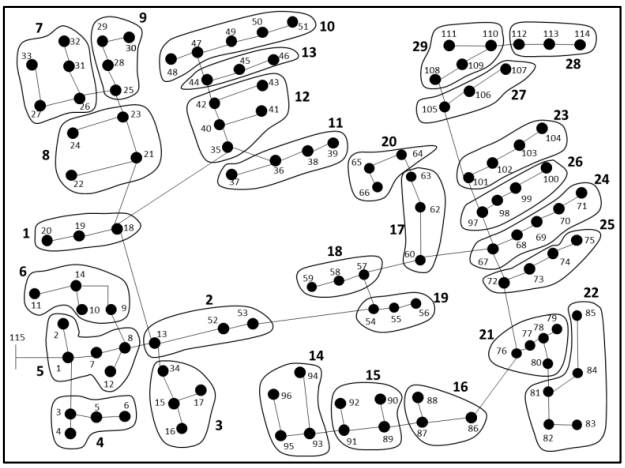


Figure 15: IEEE network clustered at hierarchical level 4.

4.2.2. Results of case study B

In order to evaluate the robustness of the electric power distribution network of Section 4.2, an unsupervised spectral clustering algorithm (Section 3.2) is first applied to reduce the size of the system and, then, Monte Carlo simulation is performed at a given clustering level in combination with the Hierarchical Graph representation (Section 3.1). The outcome of the procedure is represented by the steady state probability distribution of the electricity delivered at the demand nodes (or clusters). In this case, we consider proportional importance of the clusters/demand nodes (see Section 2).

As illustrated in Section 4.2.1, five hierarchical clustering levels are identified by the unsupervised spectral clustering algorithm. In the following, for the sake of brevity the results obtained from the analysis of the "fictitious" clustered networks are mainly given for the levels 3 and 4 of the clustering hierarchy and are compared to those obtained from the analysis of the original (i.e., not clustered) system.

In order to perform a fair comparison, the values of the product delivered to the nodes obtained analyzing the real network are "grouped" on the basis of the clustered structure of the corresponding "fictitious" network: for example, at hierarchical clustering level 4 the products delivered to nodes 9, 10, 11 and 14 have to be summed to allow the comparison with the corresponding cluster 6 of the "fictitious" network (Figure 15). The comparison is, then, made by means of the expected product not supplied $idNS_k$ to cluster k, computed as:

$$idNS_k = \frac{D_k - \mu_k}{D_k}, \quad (2)$$

where D_k is the demand of cluster k (equal to the sum of the demands of the nodes of the original network contained in cluster k) and μ_k is the mean value of the product delivered to the cluster k at steady state.

We have performed the comparison between the performances of the clustered and real networks in three cases:

- *Case 1:* all the arcs are characterized by the steady state probability vector $\underline{\Pi}^1$ of Figure 16 left; in this case, the system is expected to have a satisfactory performance as all the arcs remain in their best state (i.e., state 4) with very high probability (i.e., 0.9982).
- *Case 2:* all the arcs are characterized by the steady state probability vector $\underline{\Pi}^2$ of Figure 16, middle; in this case the arcs turn into state 3 (i.e., a state of partially reduced functionality) with high probability (i.e., around 0.5) so that the global performance of the system is not expected to be very high;

- *Case 3:* same as case 2, except for the fact that the arcs contained in clusters 8 and 12 at hierarchical clustering level 4 are assumed to be more prone to degradation than the other arcs and then characterized by the steady state probability vector $\underline{\Pi}^3$ of Figure 16, right; in particular, they degrade to their worse states (e.g., state 2) with high probability.

$\underline{\Pi}^1$ =

	Steady state probability
1	0.0006
2	0.0003
3	0.0009
4	0.9982

$\underline{\Pi}^2$ =

	Steady state probability
1	0.0005
2	0.0010
3	0.4995
4	0.4990

$\underline{\Pi}^3$ =

	Steady state probability
1	0.0291
2	0.0971
3	0.8651
4	0.0087

Figure 16: Steady state probabilities of the four levels of capacity of the arcs of the network in Figure 12 for cases 1 (left), 2 (middle), 3 (right).

In the following, comments on the results obtained for the three cases above are given with reference to Tables 3 – 4 where the clusters at hierarchical levels 3 and 4, respectively, are ranked in ascending order based on the product not supplied ($idNS_k$) (i.e., the mean value of the product supplied to clusters at the top of the ranking is closer to their demands than that of the clusters at the bottom). Tables 3 – 4, right, illustrate also the ranking of the clusters based on their distance (measured in terms of number of arcs) from the source point: for example, at level 3 cluster 8 is connected to the input source by four arcs (i.e., 115_1, 53_54, 60_67 and 101_105) (see Figure 14). The reference to the distance of a demand node from the source is motivated by the fact that in general the higher the distance from the source, the higher the probability of reduced product supply.

Table 3: Product not supplied, $idNS_k$, to the cluster k for the real and clustered networks at hierarchical level 3, with respect to cases 1, 2 and 3. Right: ranking (in ascending order) of the clusters based on their distance from the input source.

Level 3

Case 1			
REAL		**CLUSTERED**	
$idNS_k$	k	$idNS_k$	k
0.005	1	0.002	1
0.011	2	0.004	2
0.011	3	0.004	3
0.013	5	0.004	5
0.013	7	0.004	7
0.014	6	0.008	4
0.018	8	0.008	6
0.019	4	0.009	8

Case 2			
REAL		**CLUSTERED**	
$idNS_k$	k	$idNS_k$	k
0.298	1	0.250	1
0.338	5	0.298	5
0.338	6	0.304	2
0.339	2	0.305	3
0.339	3	0.322	7
0.341	7	0.332	6
0.342	4	0.332	8
0.345	8	0.338	4

Case 3			
REAL		**CLUSTERED**	
$idNS_k$	k	$idNS_k$	k
0.264	1	0.245	1
0.338	5	0.296	5
0.340	6	0.324	7
0.341	7	0.325	2
0.345	8	0.332	3
0.346	4	0.332	6
0.396	2	0.332	8
0.412	3	0.335	4

Distance - ranking	
$dist_k$	k
1	1
2	2
3	2
5	2
7	3
6	4
8	4
4	5

Table 4: Product not supplied, $idNS_k$, to the cluster k for the real and clustered networks at hierarchical level 4, with respect to cases 1, 2 and 3. Right: ranking (in ascending order) of the clusters based on their distance from the input source.

Level 4

Case 1				Case 2				Case 3				Distance - ranking	
REAL		CLUSTERED		REAL		CLUSTERED		REAL		CLUSTERED			
$idNS_k$	k	$idNS_k$	k	$idNS_k$	k	$idNS_k$	k	$idNS_k$	k	$idNS_k$	k	$dist_k$	k
0	4	0	1	0.220	5	0.250	5	0.160	5	0.220	5	1	5
0	5	0	2	0.300	4	0.300	2	0.263	2	0.250	2	2	2
0	13	0	4	0.313	2	0.310	4	0.288	1	0.275	1	2	4
0	18	0	5	0.320	3	0.310	6	0.290	3	0.300	4	2	6
0	19	0	6	0.320	6	0.320	3	0.290	4	0.300	6	3	1
0.008	9	0	13	0.325	1	0.325	1	0.290	6	0.300	19	3	3
0.008	12	0	18	0.325	8	0.325	8	0.313	11	0.310	3	3	19
0.010	3	0	19	0.325	18	0.325	12	0.325	18	0.325	18	4	8
0.010	6	0.008	9	0.325	19	0.325	13	0.325	19	0.330	17	4	12
0.010	10	0.008	12	0.330	17	0.325	18	0.330	17	0.333	24	4	18
0.010	17	0.008	24	0.333	12	0.325	19	0.337	21	0.333	25	5	9
0.010	23	0.008	25	0.337	21	0.330	17	0.338	27	0.333	26	5	11
0.013	1	0.008	26	0.338	7	0.333	9	0.340	16	0.337	21	5	13
0.013	2	0.010	3	0.338	11	0.333	24	0.340	23	0.338	15	5	17
0.013	7	0.010	10	0.338	15	0.333	25	0.341	20	0.338	20	6	7
0.013	8	0.010	16	0.338	27	0.338	7	0.342	24	0.338	27	6	10
0.013	11	0.010	17	0.340	10	0.338	10	0.342	25	0.340	16	6	20
0.013	15	0.010	20	0.340	16	0.338	11	0.342	26	0.340	23	6	24
0.013	27	0.010	23	0.340	23	0.338	15	0.350	14	0.342	22	7	25
0.014	20	0.011	21	0.341	20	0.338	20	0.350	15	0.350	14	7	26
0.014	21	0.013	7	0.342	9	0.338	27	0.350	22	0.350	28	8	21
0.017	22	0.013	8	0.342	22	0.340	16	0.350	28	0.350	29	8	23
0.017	24	0.013	11	0.342	24	0.340	21	0.350	29	0.375	13	9	16
0.017	25	0.013	14	0.342	25	0.340	23	0.375	12	0.383	12	9	22
0.017	26	0.013	15	0.342	26	0.342	22	0.388	7	0.388	7	9	27
0.017	29	0.013	27	0.350	13	0.342	26	0.392	9	0.388	8	10	15
0.020	16	0.013	28	0.350	14	0.350	14	0.413	8	0.388	10	10	29
0.025	14	0.017	22	0.350	28	0.350	28	0.425	13	0.388	11	11	14
0.025	28	0.017	29	0.350	29	0.350	29	0.433	10	0.392	9	11	28

Case 1

The results obtained from simulations on the clustered and real networks are very similar at all hierarchical clustering levels. In particular, the higher is the hierarchical level (i.e., the finer is the clustering), the closer are the results of the fictitious network to those of the real network.

The probability distribution functions (pdfs) of the product delivered to the clusters of the fictitious and real networks at steady state present the same modal values (i.e., the same peaks) at hierarchical levels 2, 3 and 4 and mean values very close to each other at all hierarchical clustering levels. For illustration purposes, the mean and modal values of the pdf of the product delivered at steady state to cluster 1 at hierarchical levels 2, 3 and 4 obtained from the analysis of the real and clustered networks is given in Table 5, together with their percentage difference.

Table 5: Modal and mean values [kW] of the steady state probability distributions of the product delivered to cluster 1 considering the original and artificial networks clustered at hierarchical levels 2, 3 and 4. Their percentage difference is also given.

		Hierarchical levels		
		2	**3**	**4**
Modal values [kW]	**Real network**	1550	565	85
	Clustered network	1550	565	85
	Difference %	0	0	0
Mean values [kW]	**Real network**	1582	557	79
	Clustered network	1592	559	80
	Difference %	0.63	0.36	1.27

The values of the product not supplied $idNS_k$ are in general very low (i.e., lower than 0.025), so that the presence of less supplied cluster(s) is not evident. In general, the ranking produced by the clustered and real networks are similar (Tables 3 – 4, case 1). Differences in the ranking can be found at clustering level 4, but they can be neglected given the very low values of the corresponding product not supplied index. Also a correspondence between the rankings of "groups" of clusters that have similar product not supplied index can be found: for example, the clusters closer to the generation source, e.g., 4, 5, 13, 18, 19, are slightly more supplied than those farther, e.g., 14, 15, 16, 22, 27, 28, 29 for both the fictitious and real systems, confirming the physical coherence of the approach.

Case 2

In general, the results obtained from simulations on the fictitious and real networks are less similar to each other than in case 1, where the arc state capacities present less variability. However, the trends in the network behavior that have been pointed out in the previous case can still be identified.

The pdfs of the product delivered to the clusters of the fictitious and real networks at steady state present (almost) the same peak values at hierarchical levels 2, 3 and 4. Also the mean values are very close to each other at hierarchical levels 3 and 4 (actually, the maximum percentage difference is 6.9 % for cluster 1 at level 3 and 3.9 % for cluster 5 at level 4); instead, at level 2 the difference is larger, i.e., 10.7 % for cluster 1. Thus, as expected, the higher the hierarchical level (i.e., the finer the clustering), the closer the results of the artificial network to those of the real network. For illustration purposes, the mean and modal values of the pdf of the product delivered at steady state to cluster 1 at hierarchical levels 2, 3 and 4 obtained from the

analysis of the real and clustered networks is given in Table 6, together with their percentage difference.

Table 6: Modal and mean values [kW] of the steady state probability distributions of the product delivered to cluster 1 considering the original and artificial networks clustered at hierarchical levels 2, 3 and 4. Their percentage difference is also given.

		Hierarchical levels		
		2	**3**	**4**
Modal values [kW]	**Real network**	1050	375	55
	Clustered network	1050	375	55
	Difference %	0	0	0
Mean values [kW]	**Real network**	1078	393	54
	Clustered network	1193	420	54
	Difference %	10.67	6.87	0

With respect to the previous case 1, the probability distributions present higher variability; in particular, this effect is more evident in the results obtained on the artificial networks. For example, referring to cluster 1, the variances of the distributions of the product delivered are 3514, 725 and 25 at levels 2, 3, 4, respectively, for the original network, whereas they are 25283, 3199 and 35, respectively, for the clustered network. This can be due to the presence of the clusters and to the definition of their performance (Section 3.2). Actually, the state of a cluster is represented by indicator id_k (eq. 1) that tries to capture and synthetize the main features of the nodes inside the cluster itself. By so doing, the (detrimental) effect of the degradation of an arc is "spread" through the entire cluster instead of having a contained local impact only on the physically connected nodes. This leads to an increase in the variability of the performance of the network: as expected, such variability decreases with the level of detail of the analysis, i.e., with the reduction of the cluster size.

The ranking based on the product not supplied (Tables 3 – 4, case 2) is similar for the groups of clusters at the top and at the bottom (i.e., at the tails) of the ranking, so that it is possible to highlight those that are more supplied (e.g., clusters 1 and 5 at level 3, and clusters 1, 2, 3, 4, 5, 6 at level 4) with respect to those who are less supplied (e.g., clusters 4 and 8 at level 3, and 14, 28, 29 at level 4, see Table 4, case 2). As for case 1, the tails of the ranking reflect the distance of the clusters from the generation point: at the top we find the elements closer to the input source, whereas at the bottom those that are farther (see Tables 3 – 4, right).

Case 3

The pdfs of the product delivered to the clusters of the artificial and real networks at steady state present almost the same modal (peak) values at hierarchical levels 2, 3 and 4. Instead, the mean values are farther from each other than in the previous case 2: for example, at hierarchical level 3 the maximum percentage difference is 14.5 % for cluster 3, whereas at level 4 it is 7.9 % for cluster 10. At level 2 the maximal percentage difference is larger, i.e., 16 % for cluster 1. Again, the higher the hierarchical level, the higher the similarity of the results produced by the artificial network and by the real network. For illustration purposes, the mean and modal values of the pdf of the product delivered at steady state to cluster 1 at hierarchical levels 2, 3 and 4 obtained from the analysis of the real and clustered networks is given in Table 7, together with their percentage difference.

Table 7: Modal and mean values [kW] of the steady state probability distributions of the product delivered to cluster 1 considering the original and artificial networks clustered at hierarchical levels 2, 3 and 4. Their percentage difference is also given.

		Hierarchical levels		
		2	**3**	**4**
Modal values [kW]	**Real network**	1050	375	55
	Clustered network	1050	375	55
	Difference %	0	0	0
Mean values [kW]	**Real network**	1026	412	57
	Clustered network	1190	423	58
	Difference %	15.98	2.67	1.75

In this case, the reduction of performance of clusters 8 and 12 at level 4 strongly influences the ranking performed according to the product not supplied ($idNS_k$) indicator.

The simulations carried out on the artificial networks at hierarchical level 3 cannot capture the reduction in the performances of the arcs inside clusters 8 and 12 at level 4. Actually, according to the simulation performed on the original network, at level 3 clusters 2 and 3 should be the most impacted, since they are those that contain the arcs with the worst capacities; those are followed by clusters 4 and 8 that are the farthest from the input source. On the contrary, the simulations of the fictitious networks at level 3 identifies clusters 4 and 8 as the least supplied (Table 3, case 3). At level 4, the simulations of the fictitious and original networks produce the same results (Table 4, case 3): the clusters most supplied are clusters 1, 2, 3, 4, 5, 6 (i.e., those closer to the input source) and those less served are 7, 8, 9, 10, 12, 13 (i.e., those characterized

by a reduction of performance and those connected to these last ones), followed by clusters 14, 15, 22, 28, 29 (i.e., those that are farther from the generation point). Thus, it can be concluded that in those cases where the performance of some arcs is biased towards very low values, a finer level of analysis is needed to capture the global behavior of the network.

Finally, the computational cost of the simulations depends on the configuration of the system: the higher the capacities of the arcs, the faster the process to allocate the product by Hierarchical Graph. However, in all the three cases explored, it is evident a considerable reduction in the computational time when the level of detail of the analysis is reduced (i.e., when the network is clustered). In Table 8, the computational times needed to perform the simulations on the clustered networks at hierarchical levels 2, 3 and 4 and on the original network are given for cases 1, 2 and 3 detailed above.

Table 8: Computational time [s] of 10000 Monte Carlo simulations for the analysis of the real and clustered networks at hierarchical levels 2, 3, 4 for cases 1, 2 and 3 considered.

	Case 1		**Case 2**		**Case 3**	
Level	**REAL**	**CLUSTERED**	**REAL**	**CLUSTERED**	**REAL**	**CLUSTERED**
2	454	11	1304	11	1392	50
3	454	33	1304	40	1392	78
4	454	110	1304	201	1392	256

In conclusion, a robustness analysis has been carried out at different (hierarchical) clustering levels by Hierarchical Graph and Monte Carlo simulation for an electric power distribution network adapted from the IEEE 123 node test feeders. The results have shown that it is possible to apply the approach resorting to clustering techniques to reduce the computational time of the simulations since:

- The modal and mean values of the distribution of the pdfs of the product delivered at steady state to the clusters are similar for the analysis of the real and clustered networks;
- The simulations of the fictitious and original networks produce a similar rankings of the product-not-supplied index: it can be noticed a strong correspondence of “groups” of clusters at the tails of the ranking.

However, attention should be paid when some arcs inside the clusters are more prone to degradation than other arcs. In this case, the simulation carried out on a clustered network characterized by a low level of detail (i.e., by a low number of clusters) may not always capture this reduction of performance: a finer clustering level is, then, necessary.

These analyses carried out on clustered networks provide approximate results that are useful in a decision making process for a preliminary design step, reducing considerably the computational cost of the simulation.

5. Conclusions

In this paper, we have proposed a new representation technique, i.e., the Hierarchical Graph, to analyze the performance of interconnected critical infrastructures (CIs) under a multi-state system-of-systems (SoS) framework. In particular, the robustness of the SoS, i.e., the capability of the CIs to resist to failures or partial failures assuring the required level (or a high level) of supply of goods or services, has been evaluated in terms of the product delivered at steady state to the demand nodes of the CIs.

The Hierarchical Graph representation supports both a qualitative and quantitative analysis to evaluate the robustness of CIs. Actually, critical arcs, which interruption or degradation affects more demand nodes, can be visualized and the product delivered at steady state to the demand nodes of the CIs can be quantified, while accounting for:

- the importance (relevance, priority) of the demand nodes;
- the quantity of product required by the demand nodes;
- the state of the arcs of the networks (possibly affected by failures and partial failures).

The quantification algorithm based on the Hierarchical Graph follows a logical analysis, while not resorting to a flow model.

In the paper, we have applied the approach on two case studies. First, we have analyzed a small-sized SoS composed by two interconnected CIs (gas and electricity networks) and a supervisory control and data acquisition (SCADA) system, and we have evaluated its robustness by Monte Carlo (MC) simulation considering different importance criteria (sequential, proportional, and equal) for the demand nodes. We have shown that the Hierarchical Graph representation can support this kind of analyses that are useful for decision makers to understand margins of improvement of the SoS to optimize the delivery of product to the demand nodes by changing their importance. The small size of this case study has allowed understanding the details of Hierarchical Graph modeling.

Then, we have evaluated a moderately large-sized power distribution network by adopting a hierarchical clustering algorithm to analyze the SoS at different levels of detail and simplify the

Hierarchical Graph representation. In this case, only a proportional importance of the demand nodes has been considered. The results have shown that the Hierarchical Graph can be adopted together with hierarchical clustering algorithms to provide approximate results by analyzing clustered networks instead of the entire large-sized, real network. This can be useful in a first preliminary phase of design of SoS, in order to have satisfactory, physically coherent results with relatively low computational cost.

The work presented is methodological in nature and intended to explore the feasibility of the proposed approach, while understanding its functioning and potential benefits. To this aim, simplifying assumptions have been made in the numerical case studies analyzed. In spite of some limitations due to the assumptions made (e.g., the use of exclusively unidirectional relationships, the prediction of possibly imprecise results when the clustering algorithm is employed), the proposed approach has shown some interesting benefits that make it potentially attractive for the analysis of CIs robustness, e.g., the fact that qualitative and quantitative analyses can be combined and multistate modeling can be adopted for more refined insights on system behavior.

References

Ali, S., Maciejewski, A. A., Siegel, H. J., and Kim, J. K. (2003). "Definition of a Robustness Metric for Resource Allocation." Proceedings of the 17th International Symposium on Parallel and Distributed Processing, IEEE Computer Society, 42.1.

Apostolakis, G. (1990). "The Concept of Probability in Safety Assessments of Technological Systems." *Science*, 250(4986), 1359-1364.

Barry, L. N. (1995). *Stochastic modeling: analysis and simulation*, McGraw-Hill, New York.

Birnbaum Z. W. (1969) "On the importance of different components in a multicomponent system," in Multivariate Analysis 2, ed: New York, Academic Press.

Borgatti, S. P. (2005). “Centrality and network flow.” *Social networks*, 27(1), 55-71.

Buriticá Cortés, J. A. M., Sánchez-Silva, M., and Tesfamariam, S. (2014). "A hierarchy-based approach to seismic vulnerability assessment of bulk power systems." *Structure and Infrastructure Engineering*, 1-17.

Carlson, J. M., and Doyle, J. (2002). "Complexity and robustness." *Proceedings of the National Academy of Sciences (PNAS)*, 99(1), 2538-2545.

Clauset, A., Moore, C., and Newman, M. E. J. (2008). "Hierarchical structure and the prediction of missing links in networks." *Nature*, 453(7191), 98-101.

Courtois, P. J. (1985). "On Time and Space Decomposition of Complex Structures." *Communications of the Acm*, 28(6), 590-603.

Di Giorgio, A., and Liberati, F. (2012). "A Bayesian Network-Based Approach to the Critical Infrastructure Interdependencies Analysis." *Systems Journal, IEEE*, 6(3), 510-519.

Dueñas-Osorio, L., and Vemuru, S. M. (2009). "Cascading failures in complex infrastructure systems." *Structural Safety*, 31(2), 157-167.

Elsayed, E. A. (2012). Reliability Engineering, John Wiley & Sons.
Eusgeld, I., Nan, C., and Dietz, S. (2011). "System-of-systems" approach for interdependent critical infrastructures." *Reliability Engineering & System Safety*, 96(6), 679-686.
Fang, Y. P., Pedroni, N., and Zio, E. (2014). "Comparing Network-Centric and Power Flow Models for the Optimal Allocation of Link Capacities in a Cascade-Resilient Power Transmission Network." *Systems Journal, IEEE*, PP(99), 1-12.
Fang, Y. P., and Zio, E. (2013). "Unsupervised spectral clustering for hierarchical modelling and criticality analysis of complex networks." *Reliability Engineering & System Safety*, 116, 64-74.
Ferrario, E., and Zio, E. (2014). "Goal Tree Success Tree-Dynamic Master Logic Diagram and Monte Carlo simulation for the safety and resilience assessment of a multistate system of systems." *Engineering Structures*, 59, 411-433.
Ford, L. R., and Fulkerson, D. R. (1962). *Flows in networks*, Princeton University Press, New Jersey.
Foster, C. (2008). "Estimating Complex System Robustness from Dual System Architectures." Unifying Themes in Complex Systems, A. Minai, D. Braha, and Y. Bar-Yam, eds., Springer Berlin Heidelberg, 527-534.
Freeman, L. C. (1979). "Centrality in social networks conceptual clarification." *Social networks*, 1(3), 215-239.
Fussell, J. (1975). "How to calculate system reliability and safety characteristics." *IEEE Transactions on Reliability*, 24(3), 169-174.
Gheorghe, A. V., Masera, M., Weijnen, M., and De Vries, L. J. (2006). *Critical Infrastructures at Risk. Securing the European Electric Power System.*, Springer, Dordrecht.
Gheorghe, A. V., and Vamanu, D. (2008). "Mining intelligence data in the benefit of critical infrastructures security: vulnerability modelling, simulation and assessment, system of systems engineering." *International Journal of System of Systems Engineering*, 1(1), 189-221.
Gómez, C., Sanchez-Silva, M., Dueñas-Osorio, L., and Rosowsky, D. (2013). "Hierarchical infrastructure network representation methods for risk-based decision-making." *Structure and Infrastructure Engineering*, 9(3), 260-274.
Haimes, Y. Y. (2012). "Modeling complex systems of systems with Phantom System Models." *Systems Engineering*, 15(3), 333-346.
Hernandez-Fajardo, I., and Dueñas-Osorio, L. (2013). "Probabilistic study of cascading failures in complex interdependent lifeline systems." *Reliability Engineering & System Safety*, 111, 260-272.
Hu, Y. S., and Modarres, M. (2000). "Logic-based hierarchies for modeling behavior of complex dynamic systems with applications." Fuzzy systems and soft computing in nuclear engineering, D. Ruan, ed., Springer-Verlag, Berlin Heidelberg.
Huzurbazar, A. V. (2005). "Flowgraph models: a Bayesian case study in construction engineering." *Journal of Statistical Planning and Inference*, 129(1-2), 181-193.
IEEE. (2000). "IEEE power and energy society. Distribution test feeders." http://ewh.ieee.org/soc/pes/dsacom/testfeeders/index.html.
Jen, E. (2003). "Stable or robust? What's the difference?" *Complexity*, 8(3), 12-18.
Jensen, M. T. (2001). "Improving Robustness and Flexibility of Tardiness and Total Flowtime Job Shops Using Robustness Measures." *Journal of Applied Soft Computing*, 1(1), 35-52.
Kalos, M. H., and Whitlock, P. A. (1986). *Monte Carlo Methods Volume 1: Basics*, John Wiley & Sons, Inc., New York.

Kim, I. S., and Modarres, M. (1987). "Application of Goal Tree-Success Tree Model as the Knowledge-Base of Operator Advisory Systems." *Nuclear Engineering and Design*, 104(1), 67-81.

Kröger, W., and Zio, E. (2011). *Vulnerable Systems*, Springer, London.

Latora, V., & Marchiori, M. (2007). "A measure of centrality based on network efficiency." *New Journal of Physics*, 9(6), 188.

Lambert, J. H., and Sarda, P. (2005). "Terrorism Scenario Identification by Superposition of Infrastructure Networks." *Journal of Infrastructure Systems*, 11(4), 211-220.

Laprie, J. C., Kanoun, K., and Kaaniche, M. (2007). "Modelling interdependencies between the electricity and information infrastructures." *Computer Safety, Reliability, and Security, Proceedings*, 4680, 54-67.

Lin, J.-S., Jane, C.-C., and Yuan, J. (1995). "On reliability evaluation of a capacitated-flow network in terms of minimal pathsets." *Networks*, 25(3), 131-138.

Lin, Y.-K. (2010). "On transmission time through k minimal paths of a capacitated-flow network." *Applied Mathematical Modelling*, 34(2), 245-253.

Lin, Y.-K., and Yeh, C.-T. (2011). "Maximal network reliability for a stochastic power transmission network." *Reliability Engineering & System Safety*, 96(10), 1332-1339.

Lind, M. (2011a). "An introduction to multilevel flow modeling." *Nuclear safety and simulation*, 2(1), 22-32.

Lind, M. (2011b). "Reasoning about causes and consequences in Multilevel Flow Models." Advances in Safety, Reliability and Risk Management, C. Guedes Soares, ed., CRC Press, 2359-2367.

Martinez, L. J., Joshi, N. N., and Lambert, J. H. (2011). "Diagramming qualitative goals for multiobjective project selection in large-scale systems." *Systems Engineering*, 14(1), 73-86.

McCalley, J. D. (2012). "The DC Power Flow Equations 1.0 Introduction." Department of Electrical and Computer Engineering, Iowa State University.

Natvig, B. (2010). *Multistate Systems Reliability Theory with Applications*, John Wiley & Sons.

Nieminen, J. (1974). "On the centrality in a graph." *Scandinavian Journal of Psychology*, 15(1), 332-336.

Nozick, L. K., Turnquist, M. A., Jones, D. A., Davis, J. R., and Lawton, C. R. (2005). "Assessing the performance of interdependent infrastructures and optimising investments " *International Journal of Critical Infrastructures*, 1(2-3), 144-154.

Ouyang, M. (2014). "Review on modeling and simulation of interdependent critical infrastructure systems." *Reliability Engineering & System Safety*, 121, 43-60.

Porter, M. A., Onnela, J. P., and Mucha, P. J. (2009). "Communities in Networks." *Notices of the American Mathematical Society*, 56(9), 1082-1097.

Praks, P., and Kopustinskas, V. (2014). "Monte-Carlo Based Reliability Modelling of a Gas Network Using Graph Theory Approach." Availability, Reliability and Security (ARES), 2014 Ninth International Conference on, 380-386.

Ramirez-Marquez, J. E., and Coit, D. W. (2007). "Multi-state component criticality analysis for reliability improvement in multi-state systems." *Reliability Engineering & System Safety*, 92(12), 1608-1619.

Ramirez-Marquez, J. E., and Rocco, C. M. (2009). "Stochastic network interdiction optimization via capacitated network reliability modeling and probabilistic solution discovery." *Reliability Engineering & System Safety*, 94(5), 913-921.

Ravasz, E., and Barabasi, A. L. (2003). "Hierarchical organization in complex networks." *Physical Review E*, 67(2).

Sabidussi, G. (1966). "The centrality index of a graph." *Psychometrika*, 31(4), 581-603.

Sales-Pardo, M., Guimera, R., Moreira, A. A., and Amaral, L. A. N. (2007). "Extracting the hierarchical organization of complex systems." *Proceedings of the National Academy of Sciences of the United States of America*, 104(39), 15224-15229.

SRA. (2015). "SRA Develops Glossary of Risk-Related Terms." Society for Risk Analysis, http://www.sra.org/news/sra-develops-glossary-risk-related-terms

Tsilipanos, K., Neokosmidis, I., and Varoutas, D. (2013). "A System of Systems Framework for the Reliability Assessment of Telecommunications Networks." *IEEE Systems Journal*, 7(1), 114-124.

von Luxburg, U. (2007). "A tutorial on spectral clustering." *Statistics and Computing*, 17(4), 395-416.

Wasserman, S., and Faust, K. (1994). "Social Network Analysis: Methods and Applications." Cambridge, ENG and New York: Cambridge University Press.

Yeh, W. C. (2008). "A simple minimal path method for estimating the weighted multi-commodity multistate unreliable networks reliability." *Reliability Engineering & System Safety*, 93(1), 125-136.

Zio, E. (2007). *An Introduction to the Basics of Reliability and Risk Analysis*, World Scientific Publishing Co. Pte. Ltd.

Zio, E. (2009). *Computational methods for reliability and risk analysis*, World Scientific Publishing Co. Pte. Ltd., Singapore.

Zio, E. (2013). *The Monte Carlo Simulation Method for System Reliability and Risk Analysis*, Springer, London.

Zio, E., and Aven, T. (2011). "Uncertainties in smart grids behavior and modeling: What are the risks and vulnerabilities? How to analyze them?" *Energy Policy*, 39(10), 6308-6320.

Appendix: data of case study B

Table A1 reports the demands of the network nodes provided by the IEEE database and the maximal capacity $\zeta_{comp,4}$ of the arcs ($comp$ = 1, …, N) obtained by the DC power flow equations.

Table A1: Left: loads [KW] for all the nodes of the network (except the generation node whose load is zero). Right: maximal capacity $\zeta_{comp,4}$ of the arcs.

NODES	LOADS [kW]	NODES	LOADS [kW]
1	40	58	20
2	20	59	20
3	0	60	20
4	40	62	40
5	20	63	40
6	40	64	75
7	20	65	140
8	0	66	75
9	40	67	0
10	20	68	20
11	40	69	40
12	20	70	20
13	0	71	40
14	0	72	0
15	0	73	40
16	40	74	40
17	20	75	40
18	0	76	245
19	40	77	40
20	40	78	0
21	0	79	40
22	40	80	40
23	0	81	0
24	40	82	40
25	0	83	20
26	0	84	20
27	0	85	40
28	40	86	20
29	40	87	40
30	40	88	40
31	20	89	0
32	20	90	40
33	40	91	0
34	40	92	40
35	40	93	0
36	0	94	40
37	40	95	20
38	20	96	20
39	20	97	0
40	0	98	40
41	20	99	40
42	20	100	40
43	40	101	0
44	0	102	20
45	20	103	40
46	20	104	40
47	105	105	0
48	210	106	40
49	140	107	40
50	40	108	0
51	20	109	40
52	40	110	0
53	40	111	20
54	0	112	20
55	20	113	40
56	20	114	20
57	0	115	0

ARCS *(comp)*		$\zeta_{comp,4}$ [kW]	ARCS *(comp)*		$\zeta_{comp,4}$ [kW]
1	2	20	57	60	1815
1	3	100	58	59	20
1	7	3330	60	62	370
3	4	40	60	67	1425
3	5	60	62	63	330
5	6	40	63	64	290
7	8	3310	64	65	215
8	12	20	65	66	75
8	9	100	67	68	120
8	13	3190	67	72	865
9	14	60	67	97	440
13	34	100	68	69	100
13	18	1115	69	70	60
13	52	1975	70	71	40
14	11	40	72	73	120
14	10	20	72	76	745
15	16	40	73	74	80
15	17	20	74	75	40
18	19	80	76	77	240
18	21	280	76	86	260
18	35	755	77	78	200
19	20	40	78	79	40
21	22	40	78	80	160
21	23	240	80	81	120
23	24	40	81	82	60
23	25	200	81	84	60
25	26	80	82	83	20
25	28	120	84	85	40
26	27	40	86	87	240
26	31	40	87	88	40
27	33	40	87	89	160
28	29	80	89	90	40
29	30	40	89	91	120
31	32	20	91	92	40
34	15	60	91	93	80
35	36	80	93	94	40
35	40	635	93	95	40
36	37	40	95	96	20
36	38	40	97	98	120
38	39	20	97	101	320
40	41	20	98	99	80
40	42	615	99	100	40
42	43	40	101	102	100
42	44	555	101	105	220
44	45	40	102	103	80
44	47	515	103	104	40
45	46	20	105	106	80
47	48	210	105	108	140
47	49	200	106	107	40
49	50	60	108	109	140
50	51	20	109	110	100
52	53	1935	110	111	20
53	54	1895	110	112	80
54	55	40	112	113	60
54	57	1855	113	114	20
55	56	20	115	1	3490
57	58	40			