Optimisation-based decision-making for complex networks in disastrous events

Camilo Gómez*, Jessica Buriticá and Mauricio Sánchez-Silva

Department of Civil and Environmental Engineering, Universidad de los Andes, Cra 1st East 19A-40, Bogotá, Colombia
E-mail: ch.gomez171@uniandes.edu.co
E-mail: ja.buritica57@uniandes.edu.co
E-mail: msanchez@uniandes.edu.co
*Corresponding author

Leonardo Dueñas-Osorio

Room 212, Ryon Laboratory, Department of Civil and Environmental Engineering, Rice University, 6100 Main Street, MS-318, Houston, Texas 77005-1827, USA
E-mail: leonardo.duenas-osorio@rice.edu

Abstract: Assistance needs after large catastrophes often exceed available resources. Effective resource allocation is paramount to support emergency management and recovery, particularly within infrastructure networks. However, network optimisation problems exhibit high computational complexity, becoming intractable at a global scale. This paper successfully handles complexity through a systems approach, which uses a description of networks at different levels of abstraction through a hierarchical structure. The community structure of networks is unravelled via clustering algorithms that successively partition them hierarchically. A resource allocation problem is formulated adding information from the hierarchy, leading to a reduced solution space. Besides computational improvement, decisions are enhanced due to the topological information provided by the hierarchy-based optimisation. An example regarding the allocation of support centres aims to maximise assistance, at minimum cost, in case of emergency events. Solutions that respond to the network topology are obtained in a fraction of the time required by standard formulations.

Keywords: decision-making; optimisation; clustering; graph theory; systems thinking; hierarchical decomposition; resource allocation; global catastrophe; infrastructure networks.


Copyright © 2011 Inderscience Enterprises Ltd.
Biographical notes: Camilo Gómez is an Electronics Engineer from Universidad de los Andes (Bogotá, Colombia), focused on computational intelligence and control theory. He obtained his Masters degree from the same university, in which he worked on organisational cybernetics and operational research. Currently, he is engaged in his PhD, aiming to bring his background into the socio-technical problem of modelling, assessment and intervention of complex infrastructure systems.

Jessica Buriticá is an Electronics Engineer from Universidad de los Andes (2010) and is currently engaged in his Masters degree at the same university. She is interested in optimisation and resource allocation in infrastructure networks; particularly, she works on telecommunication systems and algorithms for wireless sensor networks.

Mauricio Sánchez-Silva holds a BS in Civil Engineering from Universidad de Los Andes (1989), an MS in Civil Engineering from Universidad de Los Andes (1992), and a PhD in Civil Engineering from Bristol University, UK (1996). His main area of research is structural reliability and risk analysis. His work is directed mainly towards risk assessment and management of physical infrastructure related problems. His work includes stochastic modelling and optimisation. His research also includes problems where the socioeconomic context and the environment play a significant role, and consequently, traditional risk analysis models can only be used partially.

Leonardo Dueñas-Osorio holds a BS in Civil and Environmental Engineering from Universidad de La Salle, Bogotá, Colombia (1996), an MS in Structural Engineering from the Universidad de Los Andes, Bogotá, Colombia (1998), an MEng in High Performance Structures from the Massachusetts Institute of Technology (2001) and a PhD in Civil and Environmental Engineering from Georgia Institute of Technology (2005). His research interests include the development of theoretical and computational models for characterising the reliability and risk of complex lifeline networks under external and internal hazards.

This paper is a revised and expanded version of a paper entitled ‘Optimization-based decision-making for complex networks in disaster events’ presented at Fifth International Forum on Engineering Decision Making, Stoo, Switzerland, 7–10 December 2010.

1 Introduction

Among the many issues to address in catastrophic situations, a primary concern is the efficient assistance to the victims. The effectiveness of the assistance implies a resource allocation problem in which a demand (e.g., medical assistance, materials, equipment, etc.) is to be optimally supplied. The optimisation is clearly conditioned on the availability of support infrastructure (e.g., transportation network) and socioeconomic stability. Risk management and disaster mitigation activities have many aspects that need to be addressed, including the construction of resilient infrastructure and deployment of relief strategies after the event, so that primary needs are attended effectively (e.g., dispatch of rescue brigades, communications, and supplies).
Tasks related to disaster management require: first, designing and upgrading infrastructure systems according to the nature of the hazards and detecting points of primal importance (a priori vulnerability assessments); and second, developing response strategies that are consistent with the nature of the consequences (a posteriori diagnoses). It is also necessary to consider logistical issues such as costs, channels through which aid can be provided and the physical configuration of available infrastructure after a specific event. This paper deals with such risk-based decision-making problem within a framework of optimal resource allocation on network systems to address the problem of establishing a posteriori relief strategies by defining a support network to provide first aid supplies.

Particularly, disasters with environmental effects are considered, which have direct lethal impact on the population as well as indirect consequences such as contamination of water supplies and diseases. The case of radioactive effects and leaks of toxic gases exemplify such type of disasters, e.g., the Chernobyl accident in 1986 (Abbott et al., 2006), which spread 400 times more radioactive material than had been released by the atomic bombing of Hiroshima, and the Bhopal catastrophe in 1984 (Eckerman, 2005). In the specific case of Bhopal, a leak of methyl isocyanate at a pesticide plant occurred due to misbehaviour of safety systems. Estimates suggest that thousands of immediate victims and hundreds of thousands of people may have been affected by several temporary and permanent diseases, which led to many additional deaths; major side-effects were also perceived on animals and plants, causing difficulties related to food supplies (Eckerman, 2005).

Research efforts have been directed towards optimising resource allocation and infrastructure usage in order to guarantee effective response to emergency situations. A general framework including relevant elements and areas of study for disaster management is presented in Vijayan Iyer and Mastorakis (2006), in the context of developing a comprehensive software tool. The protection and enhancement of infrastructure systems supporting physical and virtual networks relevant for disaster management are essential; in the case of communications, the scheduling and traffic control of cellular networks is paramount (Zhou and Beard, 2010). Moreover, Huang et al. (2010) propose the use of web-based support for disaster management, highlighting the benefits of social networking for real-time communications behind logistics. Finally, Peng et al. (2010) provide different evidence by developing an economically-oriented optimisation model for the case of unconventional social emergencies, where the word unconventional is used in the original paper to refer to the consideration of collateral consequences resulting from the evolution of highly uncertain low-probability/large-consequence events.

This paper deals with the problem of optimising the allocation of resources to a region to respond to a major disaster event by deploying support centres, considering the nature of the demand and the available infrastructure; the objective of the contribution is to improve the efficiency for decision-making by taking into account the internal structure of a network following a systems approach (Gómez et al., 2011). Within this context, the term resources is used in a general manner and may account for medical supplies, food and shelter or rescue operations. The allocation of resources requires the construction of a support network, which commonly overlaps an existing infrastructure such as a transportation network. The optimum design and use of a network has been proved to be computationally expensive, leading commonly to NP-hard problems (Ahuja et al., 1993). NP-hard problems are a subset of NP problems, where NP stands for...
non-deterministic polynomial time and suggest that no solution is likely to be obtained in polynomial (rather than exponential) time, as a function of the number of operations relative to the size of the problem (e.g., number of nodes in a network). This curse of dimensionality is especially problematic when considering a large-scale scope with a large number of elements, as would be the case of managing a catastrophic event at an infrastructure system scale spanning multiple regions and political boundaries. Thus, most practical alternatives to solve the problem use heuristics and/or approximate methods to achieve ‘reasonable’ solutions (Bonabeau et al., 1999; Reeves, 1993; Lee and El-Sharkawi, 2008).

In order to deal with complexity, a systems approach is used to gain conceptual insight into the problem and reduce the computational cost of finding the solution (Gómez et al., 2011). In a systems approach, problems are structured hierarchically; this is, in the form of a tree where the description of the system varies from a general scope at the top to a detailed description at the bottom. In the case of networks, the decomposition is obtained by means of successive use of clustering methods (Scholkopf et al., 1998). Then, every level of the hierarchy represents a simplified network that carries along valuable information such as relevant subcomponents as well as critical links and special nodes (hubs) at specific levels of abstraction, enhancing evidence presentation for rapid decision-making as compared to a plain network definition.

The paper is organised as follows: the problem of system representation for decision-making in engineering is presented in Section 2, proposing clustering and systems thinking as a means to extract useful information out of networks. Section 3 discusses optimisation as a way to perform decision-making tasks such as resource allocation and presents a method based on the considered systems approach. An illustrative example is presented in Section 4, in the context of assistance to disasters involving environmental consequences on a large and populated area. The objective is to allocate different types of resources (support centres) throughout a complex network, which becomes highly expensive computationally as the size of the network increases; in this sense, a systems approach is brought into the optimisation problem in order to absorb its complexity by considering different levels of abstraction. Section 5 presents the conclusions and future research work.

2 Description and modelling of networks

As a first step for decision-making, the understanding and modelling of the problem at hand is presented in this research from a network theory perspective.

2.1 Network representation

Graph theory provides a natural way of modelling networks. A graph $G(V, E)$ consists of a set of nodes (vertices) $V = \{v_1, v_2, ..., v_n\}$ and a set of connecting edges $E = \{e_1, e_2, ..., e_m\}$ whose end-points are elements of $V$. The network structure is usually defined by the adjacency matrix, which describes how vertices are connected. It is a matrix $A_{n \times n}$ such that $a_{ij} = 1$ if nodes $v_i$ and $v_j$ are connected, and $a_{ij} = 0$ otherwise. The specific configuration of elements (i.e., topology) as well as the flow (e.g., energy, matter or information) of network systems is described by performance indices that might account for either local or global information (e.g., average shortest-path or degree
distribution); these indices take a probabilistic character when considering the failure of elements.

The topology and dynamics of networks define special structures and lead to different network models; these characteristics provide relevant evidence to decision-making, especially in the case of complex networks. For instance, regarding structure, subcomponents might be found according to connectivity, including cliques and clusters as well as hubs or centroids which are of special importance for the network behaviour and specifically for the resource allocation problem under consideration. Similarly, two main different models can be distinguished: small-world networks (SWN) and scale-free networks (SFN) (Watts, 2004), which exhibit particular dynamics that influence the overall network performance and behaviours of interest such as damage propagation. Effective decision-making must rely on topological features in addition to the risk-analysis (e.g., consequences of lack of reliability) of its individual elements.

In the case of complex systems, such as infrastructure networks, the amount of data representing relevant information may be extremely large. Additionally, the calculation of performance measures and the solution of flow/cost problems in graphs, related to resource allocation, are computationally expensive. Therefore, data analysis and pattern recognition techniques can be used advantageously (e.g., through pre-processing information) to tackle computational demands.

2.2 Managing complexity by network decomposition

Systems thinking (Checkland, 1981) is built upon the idea that a system can be described at different levels of detail through a hierarchical structure (Blockley and Godfrey, 2000). Each level in the hierarchy provides different evidence on network features to the decision-making process. Therefore, the level of detail of the analysis may change depending upon the nature of the decision problem at hand. Focusing on the decision-maker needs, the problem can be significantly simplified; in fact, the system should not be taken to a level of detail beyond which the relevance of the decision is more valuable than its specificity.

A hierarchical description of the network can be obtained by recursive clustering. In this process, the internal structure of a network is unravelled by decomposing it into subnetworks until a full resolution description is obtained. Thus, the network is initially considered as a single unit (a whole) and successively subdivided into communities (i.e., subsystems) until subsystems correspond to actual network components at the bottom of the hierarchy.

The strategies used to identify patterns within the network around which communities of elements can be grouped are commonly known as clustering methods. The objective of clustering methods is to generate a partition of the network into δ subgroups. Each of these subsets includes a centroid (i.e., a representative of the group) and constitutes a zone referred to as a Voronoi region, which is determined by a convex set containing the nodes for which such centroid is the nearest.

Thus, the network representation is given by the union of the Voronoi regions, each including a centroid and the nodes associated to it (Filippone et al., 2008). A centroid is representative of elements within its cluster, with higher centrality and connectivity properties than any other node in the cluster.

Conceptually, most clustering approaches are based on developing a similarity measure \( m_{ij} \) between pairs of vertices \((v_i, v_j)\) and an iterative process of vertex grouping
up to a point where a minimum or maximum similarity value is achieved. A review of clustering concepts and sophisticated algorithms can be found elsewhere (Filippone et al., 2008; Xu and Wunsch, 2005, 2008).

2.3 Hierarchical network representation

Hierarchical representations of networks provide information that can be used to define a *fictitious network* for each of its levels. Thus, the hierarchical representation of a network system leads to a collection of graphs (one for each level \( k \)) \( G^{(k)}(\Lambda^{(k)}, E^{(k)}) \) where \( \Lambda^{(k)} \) and \( E^{(k)} \) are the sets of fictitious nodes and links, respectively. In a fictitious network, fictitious nodes correspond to clusters and fictitious edges are parallel arrangements of actual connecting edges between clusters. Then, in the fictitious network at hierarchical level \( k \), node \( i \) is defined as \( V_i^{(k)} \) and fictitious link connecting fictitious nodes \( i \) and \( j \) is defined as \( E_{ij}^{(k)} \), which is composed by all links travelling between \( V_i^{(k)} \) and \( V_j^{(k)} \). The union of the subsystems at a specific level constitutes the complete system’s representation (fictitious network) at the \( k \)-th level of abstraction. Then, at the top of the hierarchy (i.e., \( k = 1 \)), the network is interpreted as a single unit (i.e., one fictitious node), which consists of all actual nodes \( v \). The set of vertices at level 1 is given by a single fictitious node, i.e., \( \Lambda^{(1)} = \{V_1^{(1)}\} = \{v_1, v_2, ..., v_n\} \). In the second level \( (k = 2) \), the network is described by \( d \) fictitious nodes: \( \Lambda^{(2)} = \{V_1^{(2)}, V_2^{(2)}, ..., V_d^{(2)}\} \), with \( V_1^{(2)}, V_2^{(2)}, ..., V_d^{(2)} \subseteq V_1^{(1)} \) (Figure 1). Note that at the bottom level both the fictitious and the real networks are the same (Gómez et al., 2011).

A significant amount of information is extracted by using hierarchical representations; this is, centroids (or hubs), as well as fictitious networks, fictitious nodes (clusters) and fictitious links (link arrangements) at different levels of abstraction. These elements constitute the *community structure* of the network (at different levels) in terms of the location and connectivity of nodes.

The evidence for decision-making is enhanced, passing from a raw set of nodes to a set of strategic nodes at which resources can be placed to generate benefits for a specific community, which is key for the resource allocation problem at hand. For instance, considering only the \( \delta \) centroids instead of the \( n \geq \delta \) nodes reduces the possibilities (i.e., having fewer candidates) where to allocate resources, based on the enhanced properties of centroids relative to other nodes; this can be thought of as taking samples from the network intelligently under the criteria of centrality and connectivity. In this sense, a variety of resources of different characteristics (cost, capacity, accessibility) can be assigned coherently with the multi-scale intrinsic structure of the network. Taking advantage of the network structure can result in improved scalability and robustness of the resource allocation solution due to exploiting features such as connectivity and redundancy.

The hierarchical representation is then used as a complexity attenuator for the resource allocation problem. It is worth noting that the proposed model is actually more complex; this is, a hierarchy composed of a set of graphs is used instead of the original network. However, the increased size of data structures is compensated by the additional information obtained, which allows to simplify the optimisation process by filtering potential solutions in terms of the community structure, as will be further discussed in
Section 4. The latter is linked to the fact that, in computer science, the complexity of algorithms is inverse to that of data structures (Tarjan, 1983).

Figure 1  Fictitious networks and hierarchical representation of infrastructure systems

3 Resource allocation by clustering-based optimisation

Most problems in engineering imply resource allocation at least from the economics perspective. This section deals with the specific problem of allocating physical resources to satisfy a demand that is geographically distributed. The considered approach is that of a rational decision-maker who seeks to perform a function at a minimum cost considering a risk-neutral attitude.

3.1 Optimisation-based decision-making

Optimisation methods seek to find the variable values that optimise a multivariate objective function under a set of constraints. Constraints define a search space also known as feasible region within which the solution must be enclosed. Among optimisation methods (Luenberger, 2003), linear programming is a widely used optimisation paradigm in many applications because of its ease for implementation and because it exhibits greater stability and convergence properties than other methods (e.g., non-linear gradient methods).

For the case of large-scale problems, such as resource allocation at a global scale, a variety of techniques have been developed based on either exact or heuristic methods.
The column generation method (Gondzio and Sarkissian, 1997) and Dantzig-Wolfe decomposition (Vanderbeck and Savelsbergh, 2006) are useful exact approaches for problems with a huge number of variables or constraints, respectively. There also exist powerful heuristics for many different types of problems, within which evolutionary and other bio-inspired techniques have gained momentum in the last decade (Bonabeau et al., 1999; Reeves, 1993; Lee and El-Sharkawi, 2008); particularly, evolutionary algorithms were used in Chang (2010) for optimisation of supply chains when taken to a global level.

From a decision-making perspective, the definition of the objective(s) and evaluation criteria is a concern that adds up to the complexity of the problem itself. Most practical problems have several (usually) conflicting objectives: in the case under consideration, although assistance would increase as more money is invested, the former is to be maximised whereas the latter is to be minimised. Multi-objective optimisation (MOO) (Coello et al., 2006) deals with such kind of problems by finding a Pareto front, which defines a region in which neither of the objectives can be improved without degrading another. An alternative option is to fix one of the objectives as a constraint, i.e., attend at least $P$ people or spend at most $D$ dollars and then perform sensitivity analyses on $P$ and/or $D$ to unravel and evaluate trade-offs. This way, stakeholders get to know a set of equivalent choices and their consequences on the objective.

Most applications can be fitted to any of a set of well-known problem archetypes which encapsulate common topological features. In practice, specific optimisation strategies have been developed for problems such as the set-covering problem (Caprara et al., 1998), the assignment problem, and the capacitated facility location problem (CFLP) (Levi and Shmoys, 2004). For instance, a location set covering problem (LSCP) approach is used in Bell et al. (2011) to optimise alert-site allocation for homeland defence. In the particular case of the CFLP a set of facilities/service providers are to be installed to satisfy the demand from a set of users/clients, under a restricted capacity. This classic resource allocation problem complies, with some minor adjustments, with the type of problem discussed in this paper.

In the case under consideration, an optimal supply network is to be designed, considering:

1. whether a facility of a specified type $h$ is located at each node $i$ [hard capacity (Levi and Shmoys, 2004)]

2. the capacity of a facility is neither fixed nor free, in contrast, there is a discrete set of alternatives with different characteristics that can be installed at any node in order to optimally satisfy the demand

3. the total number of facilities is not fixed, so it might take values between 1 (i.e., totally centralised) and the number of nodes $n$ (i.e., totally decentralised) for each alternative according to feasibility and optimality.

With the exposed conditions, the basic assignment problem has proved to be NP-hard (Khan, 2003) due to the fact that the number of possible solutions is highly combinatorial. Therefore, crude exact optimisation methods may not suffice for practical applications.
3.2 Proposed approach

This paper proposes that the hierarchical representation of networks enhances the approach to the optimum resource allocation problem: first, it forces the quality and suitability of the solution to the network’s topology; and second, it improves the computational efficiency of the optimisation programme. This strategy is defined so that the hierarchy-based resource allocation problem suits the characteristics of linear programming, making the solution very efficient.

The hierarchical description of the network can be used to reduce the space of feasible solutions by imposing additional constraints that force the solution to fit the multi-level community structure of the network. Thus, special parameters and functions that encapsulate relevant information from the hierarchy are necessary. In this sense, the mathematical model not only takes into account conventional technical/physical constraints but also considers the system’s internal structure, which is given by the centroids and fictitious elements.

Within the context of this paper, the objective function takes into consideration the costs associated to both the construction of facilities \( C_{\text{build}} \) and the cost users must assume to gain access to a specific service \( C_{\text{attend}} \); the latter is measured in terms of the distance from each user to its corresponding facility. Two sets of constraints (operational and hierarchical ones) limit the solution space. The formulation of the objective function is then stated as a cost minimisation problem as:

\[
\min \left\{ \sum_{i=1}^{M} C_{bi} + \sum_{j=1}^{n} C_{aj} \right\}
\]

where \( C_{bi} \) is the cost of building the \( i^{th} \) of \( M \) facilities and \( C_{aj} \) is the cost of attending the \( j^{th} \) of \( n \) users.

Subject to:

\[
\Gamma = \left\{ \gamma_{G} \in \mathbb{R}^{n} : \gamma_{G} = f(G(V, E), \Theta) \right\}
\]

\[
\kappa = \left\{ \kappa_{k} \in \mathbb{R}^{n} : \kappa_{k} = f(T(G), R(G)) \right\}
\]

which are the sets of operational (\( \Gamma \)) and hierarchical (\( \kappa \)) constraints, respectively. The first set is associated to issues such as specific costs, and maximum allowed distances and capacities, whereas the second one deals with topological features such as centroid mapping and conditions within the hierarchy. For this reason, the constraints \( \gamma_{i} \) are (linear) functions of the graph itself \( G \) and a set of attributes \( \Theta \) (e.g., costs, distances, capacities), whilst the constraints \( \kappa_{i} \) depend on the tree representing the hierarchical structure of the graph \( T(G) \) and the distribution of centroids \( R(G) \); such hierarchical constraints reduce the search space for the algorithm by bringing information about community structures within the network. Notice that an essential constraint is that users are assigned only to installed support centres.

In particular, hierarchical constraints include:

- **Facilities must be located only at centroids of any hierarchical level**: This condition acts directly on reducing the complexity (number of possible permutations); it is
supported by the fact that cluster centroids condense relevant information about fictitious nodes, working as structure-based samples of the network at different levels.

- **Network nodes must reflect resources allocated at fictitious nodes**: The ultimate decision is made in terms of fictitious nodes (i.e., there is a decision variable for each fictitious node), aiming to exploit the hierarchical structure of the network. However, when an alternative is placed at a fictitious node in the hierarchy, its centroid must reflect such decision in the actual network. For this purpose, a mapping function is designed which returns an actual network node for every fictitious node, based on the centroids provided by the clustering method. This way, an auxiliary decision variable is created using the mapping function, so that a solution in terms of the actual nodes can be derived from the solution obtained for the fictitious nodes.

- **Alternatives must be guaranteed along every line of heritage**: Let us define a dominant alternative as one within the set of alternatives having outperforming characteristics (i.e., greater capacity and radius of attention). Also, let us define a line of heritage as the path from the leaves (single nodes at the bottom) to the root (whole network at the top) of the hierarchy tree. Therefore, in order to guarantee user $i$ access to alternative $a$, at least one alternative $a$ must be placed within the line of heritage associated to $i$. Aiming to provide integral attention, a condition is imposed such that all users have access to a dominant alternative, either directly or by means of a connection through another alternative without passing through nodes without assigned facilities, which enhances efficiency (e.g., reduces delays in case of emergencies). Notice that such condition could be guaranteed by placing a dominant alternative at the top of the hierarchy (total centralisation), which would be either unfeasible or suboptimal with great likelihood; the same would occur if $n$ dominant alternatives were placed at the bottom level of the hierarchy (total decentralisation). Therefore, the optimiser is to provide the appropriate level at which different alternatives must be placed, based on the structured representation of the network; this is, it defines a topologically consistent degree of centralisation.

- **Fictitious nodes cannot be assigned more than one alternative**: It might occur for (eventual) concentric clusters that, at different hierarchy levels, the same actual node is centroid for different clusters; this implies building several different size alternatives at the same location. Thus, a condition is introduced such that there can be only one alternative per node. In this sense, the optimisation algorithm is forced to ‘decide’ an optimal scope for the alternative.

The clustering-based optimisation approach that results from adding these constraints is not only very efficient computationally (it reduces the number of possible solutions) but it also leads to a solution that takes advantage of the system’s intrinsic structure, enhancing properties such as scalability of the designed network. For instance, when new nodes are added to a network, it is more likely that they attach to those with greater degree; consequently, centroids will remain a good option to allocate resources.
4 Illustrative example

An example is presented aiming to provide insight into the proposed methodology, by means of a generic resource allocation problem [i.e., a variation of the CFLP (Levi and Shmoys, 2004)]. Some context features are provided intending to demonstrate the potential and applicability of the method in civil engineering, specifically in the case of disaster assistance. It is important to notice that specific assumptions on the example are only included for demonstrative purposes, for the method is presented in a more general decision-making perspective.

4.1 Problem description

A transportation network with 64 nodes (urban regions) and a diameter of approximately 1,500 km is generated by using the Delaunay triangulation algorithm (Khanban et al., 2002). The population associated to every node was arbitrarily assigned based on a set of Gaussian bells, centred at a subset of network centroids (Figure 2); the latter, under the assumption that large cities, exhibiting large populations, are highly connected to neighbouring cities with smaller populations.

Disasters consist of a release of toxic material (i.e., gas/radioactive). A discrete set of eight sources is considered (e.g., plants) and their locations are consistently attached to the considered urban centres (i.e., centroid). This assumption is based on the fact that major industrial centres are usually in, or nearby, metropolitan areas. Each source fails (or not) independently, with probability $\gamma = 0.15$, and random intensity $\beta$ according to a uniform distribution $[0, 1]$. This pair (occurrence/intensity) produces a Gaussian pattern of decay with which the demand of population to be attended computed for every node; the Gaussian function is normalised to values between $(0, 1)$, which are multiplied by the population at each point to determine the demand. A more complete analysis may include modelling the drifting vapour cloud of the toxic release, but such approach is beyond the current scope of the research.
4.2 Problem formulation

The objective is to allocate different types of support centres throughout the network in order to satisfy the demand generated by the accidents, at a minimum cost. Additionally, the designed support network must cover a percentage \( \lambda \) (e.g., around 20\%) of the population, even if the demand is below that level; this is enforced so that a minimum level of assistance is guaranteed everywhere, which might not be achieved if based only on the random demand.

Then, a solution is designed for a specific disaster, whose performance is measured in terms of the assistance deficit (AD) when faced to new disasters. AD is computed as the difference between the demand (population affected) and the availability of services provided (i.e., available capacity of the support designed network).

It is assumed that there are basic support centres located at every node but they can only cover a small fraction of the demand. Therefore, the AD is evaluated only in terms of the population that cannot be attended by the local centre. In this sense, the decision-making problem focuses on allocating two types of centres: type 1 centres, which are the dominant alternative and have greater capacity than type 2 centres. Additionally, in order to complete the support network, type 2 centres must be connected to a type 1 centre within a distance \( d_1 \), and users must be connected to any centre within a distance \( d_2 \); these distances are fixed around 350 km so that large clusters are not neglected due to distance constraints.

The cost is divided into the cost of building and operating support centres, and the cost it takes for a person to get access to the service from his/her location; the latter is quantified proportionally to the distance travelled by a person and reflects both the actual transportation cost and the impact of travelling (i.e., time delay for the service). Therefore, the objective function from equation (1) becomes:

\[
\min \sum_{h \in H} \sum_{q \in Q} c_h x_{hq} + \sum_{i \in I} \sum_{j \in J} p_i \sum_{q \in Q} w_j d_{ij}
\]

where \( c_h \) is the cost of installing a support centre of type \( h \); \( p_i \) denotes the amount of people to be attended; and \( d_{ij} \) is the distance between such points. The terms in equation (4) account for the aforementioned parts of the cost:

- installation/operation cost, which is attached to the decision variable \( x_{hq} \) that denotes whether alternative \( h \) is placed at cluster (fictitious node) \( q \)
- user-access cost, which is linked to the decision variable \( w_j \) that represents whether user \( i \) is attended at a centre in \( j \).

Previous to the optimisation process, a recursive clustering process is carried out in order to find the intrinsic structure of the network. A recursive \( k \)-medoids clustering algorithm (Chu et al., 2002) is used for this purpose; the reason for this choice is that unlike \( k \)-means-like methods, \( k \)-medoids provides (by exhaustive searching) a set of centroids that do belong to the original set of nodes, which is necessary for the application under consideration. The algorithm leads to a five-level hierarchy, with the whole network at the top and all single nodes at the bottom. To illustrate this result, levels two and four of the hierarchy are shown in Figure 3.
Based on the resulting hierarchical structure, the formulation of the linear programme condensed in equations (1), (2) and (3) [where equation (1) is expanded as shown in equation (4)] is carried out. The target is to find the optimal location of support centres (type 1 and type 2) as well as their associated connections. It is assumed that the type of disaster under consideration is such that its main effect is on nodes (i.e., population) rather than on links (e.g., road infrastructure).
4.3 Robustness approach

A robustness-oriented simulation is carried out as shown in the flow-chart in Figure 4, aiming to estimate the performance of the designed network under the uncertainty associated to the damaging events; this is, quantifying the response of a designed support network when faced to unforeseen damaging events. For this purpose, a set of $S$ disasters are simulated and then $S$ solutions for the resource allocation are obtained by running the optimisation algorithm. For each solution, $P$ additional disastrous events are generated independently and the $AD$ of the network is computed as the amount of people unattended for each event:

$$AD = \begin{cases} D - C & \text{if } C \leq D \\ 0 & \text{if } C > D \end{cases}$$

(5)

where $D$ represents the generated demand and $C$ is the designed network capacity. Then, a robustness measure for the $i^{\text{th}}$ solution (out of $S$ cases considered) is computed as:

$$R(i) = \frac{1}{\text{mean}_j(AD_{ij}) + \text{std}_j(AD_{ij})}$$

(6)

where $i = \{1, ..., S\}, j = \{1, ..., P\}$ and $AD_{ij}$ denotes the AD for the $i^{\text{th}}$ solution when event $j$ occurs, while mean and std represent the mean and standard deviation operators. Therefore, maximum robustness is obtained for the solution with less amount of population unattended throughout the $P$ events, as calculated in the denominator of equation (6). Note that equation (6) is not an optimisation objective function itself; conversely, it allows to select a solution with enhanced robustness relative to the $S$ solutions provided by the actual objective function stated in equation (4).

The proposed robustness approach seeks to be able to respond to unforeseen events, and sacrifices optimality of cost in order to maintain acceptable performance under extreme scenarios. Moreover, it relies on two components: simulation (to provide a comprehensive set of scenarios) and optimisation (to generate solutions throughout such space). The performance of the approach depends on the quality of simulation so that the problem of ‘bias’ is overcome; i.e., avoid making decisions based on samples that are not representative. The proposed method uses Monte Carlo simulation to illustrate the process. However, more powerful simulation alternatives may be required for practical problems [e.g., Abbass et al. (2011) for computational red teaming, Hurtado (2004) for margin based importance sampling and other statistical-learning-based methods, and Au and Beck (2001) for subset simulation].

Finally, expected cost minimisation (ECM) is also attractive for the purpose of introducing uncertainty in the optimisation. Nevertheless, it requires establishing a set of events for which the probability of occurrence is known, which implies two difficulties: first, due to high uncertainty, it is difficult to obtain accurate probabilities for actual events and to cover a comprehensive set, so that minimising the expected cost is useful; and second, the philosophy of robustness is about being able to perform well under unforeseen and extreme events to some extent, rather than performing well in average (as ECM may do), because it might mean performing extremely well only in trivial cases.
4.4 Results

The optimisation is structured as a linear problem solved by FICO\textsuperscript{TM}-Xpress software package. For the purpose of comparing performance, a built-in heuristic method from the software’s database was also executed to solve the problem, including the robustness approach; the heuristic method performs a fast exploration of the *branch and bound* (Martin et al., 1985) tree necessary to solve integer (in this case, binary) problems. The resulting support networks for the proposed approach and the heuristic method are shown in Figure 5; i.e., those with maximum robustness for the hierarchical approach and the heuristic method. The nodes marked with squares denote type 1 support centres, whereas diamonds are type 2 centres and circles represent users.

The best out of the $S$ solutions are chosen for both the hierarchical and the heuristic approach, respectively (i.e., those with better robustness with respect to the $P$ disasters). The solutions obtained through the hierarchical approach exhibit higher robustness and similar costs as compared to the results by the heuristic method: using the robustness measure from equation (6), which is inverse to the average unattended population, the hierarchical approach leads to a robustness of $10^{-6}$ whereas the heuristic obtains about half that value.

The hierarchy-based solution in Figure 5 is divided into five parts (A, B, C, D and E). Part B corresponds exactly to fictitious node 1 at level 2, i.e., $V^{(2)}_1$ (Figure 3). Parts A and C are two fictitious nodes from level 3, which in turn belong to $V^{(2)}_2$; also, note that parts B and C contain nodes $V^{(4)}_{3,6,10}$ and $V^{(4)}_{11,12,15,16}$, respectively (Figure 3). Part D corresponds exactly to $V^{(4)}_6$. Finally, part E is a fictitious node from level 3 which belongs to $V^{(2)}_5$ and contains $V^{(4)}_{14,14,18}$. 

---

**Figure 4** Flow diagram of the robustness-oriented process for simulation and optimisation

![Flow diagram](image-url)
In the example, parameters such as distances and costs mainly affect the escalation of the overall cost. Contrarily, the capacity of support centres is very important because it defines whether it is feasible to assign a resource to a certain fictitious node in the hierarchy; capacities must be such that relevant hierarchical levels are feasible, so that the proposed approach has an influence on performance.

Varying capacities of support centres (within a range in which fictitious nodes are feasible) affects the frequency of choice of the two types of alternatives. Type 1 alternatives are chosen more frequently when type 2 alternatives do not suffice demands, or vice versa when the extra capacity of the type 1 alternative is not worth its cost. Such rearrangements only modify the overall cost accordingly. The response to variations in capacity is shown in Table 1, for both hierarchical and heuristic approaches for the cases in which type 1 alternatives have five times, ten times and 20 times the capacity of type 2 alternatives.

The efficiency of installed facilities and transportation-distances remain higher for the proposed approach for the robustness approach; however, it is important to recall that outside the limits where fictitious nodes are feasible, the additional information used by the method does not provide much value. Consequently, adequate clustering methods are a primary condition for performance.

Qualitatively, the solution for the proposed approach exhibits greater decentralisation; i.e., the community structure of the network is reflected and groups with different scales appear in the solution according to optimality. This fact is important because it enhances properties such as robustness and scalability in the solution.

The formulation of the linear programme with the proposed approach helps reduce the computational burden for the solver, i.e., 33 fictitious nodes instead of 64 nodes, which leads to 31 less decision variables $x_{hq}$ (whether a centre $h$ is built at $q$) plus $64^2 - 33^2 = 3,007$ less decision variables $w_{ij}$ (whether user $i$ is attended at $j$). Moreover, the additional hierarchical information leads to a much narrower solution space, whose quantification is object of further research. Therefore, a good solution for the original problem, which is optimal for the modified problem (i.e., the one that considers only centroids), can be achieved within a reasonable time limit; a solution can be obtained in less than a minute using either the heuristic or the hierarchical approach, whereas the crude linear programming formulation does not converge to an at least acceptable solution after several running days.

Finally, in order to demonstrate the potential for reducing complexity as the network size increases, the proposed approach was compared with the crude linear programming formulation, considering running time and objective function (Table 2); these results correspond to an implementation of the CFLP in a similar fashion to that implemented in this paper. Subindex $F$ for time and cost denote the formal optimisation problem, whereas subindex $H$ denotes the hierarchical approach. Since the formal approach does not converge within reasonable limits, it was run until a gap of 55% was achieved; the hierarchical approach converges to optimum. The gap in mixed integer programming (MIP) is a measure of convergence and is based on the complementary slackness theorem from optimisation theory (Luenberger, 2003); it means that, in the worst case, the solution is 55% away from the optimum.
Figure 5  Resulting support networks, (a) hierarchical approach and (b) heuristic approach (see online version for colours)

Table 1  Sensitivity analysis on the capacity of support centres

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>S.Ctr. T-1</td>
<td>5</td>
<td>6</td>
<td>8</td>
<td>11</td>
<td>13</td>
<td>16</td>
</tr>
<tr>
<td>S.Ctr. T-2</td>
<td>16</td>
<td>25</td>
<td>12</td>
<td>15</td>
<td>4</td>
<td>8</td>
</tr>
<tr>
<td>Distance</td>
<td>15.2</td>
<td>20.4</td>
<td>14.5</td>
<td>18.63</td>
<td>13.2</td>
<td>19.3</td>
</tr>
<tr>
<td>Cost × 10⁶</td>
<td>1.4</td>
<td>1.5</td>
<td>1.7</td>
<td>1.8</td>
<td>2</td>
<td>2.3</td>
</tr>
<tr>
<td>Time</td>
<td>14</td>
<td>10</td>
<td>11</td>
<td>3.2</td>
<td>7.1</td>
<td>2.6</td>
</tr>
</tbody>
</table>

Table 2  Performance comparison of heuristic and hierarchical methods for different network sizes

<table>
<thead>
<tr>
<th>Size</th>
<th>Time_F</th>
<th>Time_H</th>
<th>Cost_F</th>
<th>Cost_H</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>3 s</td>
<td>0.3 s</td>
<td>76</td>
<td>69</td>
</tr>
<tr>
<td>45</td>
<td>492 s</td>
<td>1.3 s</td>
<td>102</td>
<td>90</td>
</tr>
<tr>
<td>85</td>
<td>440,000 s</td>
<td>15.7 s</td>
<td>142</td>
<td>136</td>
</tr>
<tr>
<td>100</td>
<td>450,000 s</td>
<td>1,000 s</td>
<td>154</td>
<td>154</td>
</tr>
</tbody>
</table>
5 Conclusions

A novel approach to resource allocation problems in infrastructure networks was presented in the context of global catastrophes. The approach uses information about the internal structure of a network (i.e., centroids and fictitious nodes at several levels) in order to carry out an efficient search within the optimisation problem that finally generates the deployment of different types of resources throughout the network.

The amount of decision variables is reduced by focusing only on centroids as representatives of the network exhibiting higher connectivity and centrality. Furthermore, the feasible region (search space) of the optimisation problem is reduced intelligently by considering topological features of the network. Such an approach not only provides a computational benefit by reducing the number of possible solutions but also forces a solution that is related to the system’s intrinsic structure. An illustrative example is presented regarding the distribution of support centres after a disaster event related to the impact of radioactive/toxic gas leaks.

A robustness approach was proposed to deal with uncertainty; specifically, regarding high-consequence unforeseen events. Such an approach relies on the interaction between optimisation and simulation. Sensitivity analyses showed that the proposed optimisation strategy for resource allocation is useful as long as resources are compatible with the hierarchical representation of the network; this is, either the hierarchy or the available resources must be shaped to make the mentioned enhancements feasible. Furthermore, the effective management of uncertainty achieved with the robustness approach relies on the quality of simulation methods; powerful methods were recommended aiming at this goal.

A key advantage of the approach is that solutions are suitable to the network structure and can adapt to its dynamics. The example shows a solution with twice the robustness of the heuristic method, using a cheaper configuration of the network: the costs of building centres and transporting victims added to USD$2 \times 10^6$, as compared to USD $2.3 \times 10^6$ by the heuristic approach, for which the number of facilities and overall distance were always higher.

Future work considers more complex assumptions on the use of the hierarchical model, MOO, and the use of more powerful methods and heuristics for both the clustering and optimisation processes.

Acknowledgements

The authors gratefully acknowledge the financial support by Centro de Estudios Interdisciplinarios Básicos y Aplicados (CEIBA) and the US National Science Foundation (Grant CMMI-0748231).

References


Optimisation-based decision-making for complex networks


