

Identifying groups of critical edges in a realistic electrical network by multi-objective genetic algorithms

E. Zio^{a,b}, L.R. Golea^b, C.M. Rocco S.^{c,*}

^a Ecole Centrale Paris-Supelec, France

^b Politecnico di Milano, Italy

^c Universidad Central de Venezuela, Caracas, Venezuela

ARTICLE INFO

Article history:

Received 11 December 2009

Received in revised form

19 June 2011

Accepted 19 November 2011

Available online 13 December 2011

Keywords:

Genetic algorithm

Multi-objective optimization

Network performance measure

Power system vulnerability

ABSTRACT

In this paper, an analysis of the vulnerability of the Italian high-voltage (380 kV) electrical transmission network (HVIET) is carried out for the identification of the groups of links (or edges, or arcs) most critical considering the network structure and flow. Betweenness centrality and network connection efficiency variations are considered as measures of the importance of the network links. The search of the most critical ones is carried out within a multi-objective optimization problem aimed at the maximization of the importance of the groups and minimization of their dimension. The problem is solved using a genetic algorithm. The analysis is based only on information on the topology of the network and leads to the identification of the most important single component, couples of components, triplets and so forth. The comparison of the results obtained with those reported by previous analyses indicates that the proposed approach provides useful complementary information.

© 2011 Elsevier Ltd. All rights reserved.

1. Introduction

Societies are heavily dependent on many systems of distributed service, the so called infrastructures, e.g. computer and communication systems [8], electric power transmission and distribution systems [12] and rail and road transportation systems [23], constituted by networks of components. Such infrastructures are becoming more and more interconnected with each other, and therefore interdependent and potentially more vulnerable with respect to faults or malicious attacks: a random failure or an intentional attack in any of these complex networks may easily propagate across the network, increasing and multiplying its effects.

These systems need priority protection, as specified by a number of national and transnational directives [4,6]. The implementation of these directives and programs calls for the establishment of proper frameworks of vulnerability analysis for designing the adequate protections against failures and attacks [6,18,22].

In order to protect a large-scale critical infrastructure, the Decision-Maker (DM) must be able to assess what elements are critical, that is what elements affect the system performance the most. From a topological viewpoint, various measures of the importance of a network element (edge or node), i.e., of the relevance of its location in the network with respect to a given

topological indicator, can be introduced. The term ‘importance’ is then intended to qualify the role that the presence and location of the element plays with respect to the average global and local connection properties of the whole network.

In the case of electrical power systems, the existing literature on vulnerability analysis largely takes a topological approach to identify the critical components in the network [1,9,25]. Such analyses are capable of identifying elements of structural vulnerability, i.e. network edges and nodes whose failure can induce a severe structural damage to the network through the physical disconnection of its parts. Such analysis is very fast from a computational point of view and only requires the information of the topology of the network.

On the contrary, this kind of analysis is limited by the fact that it focuses only on the topological features of the network, thus neglecting its physical characteristics. In this respect, it is important to verify the extent of these limitations and possibly overcome them by additional more detailed physical analyses on critical parts of the network [11,27,3].

In this paper, an analysis of the vulnerability of the Italian high-voltage (380 kV) electrical transmission network (HVIET) is carried out for the identification of the groups of elements most critical with respect to the network connection topology. In [5], the authors proposed the use of betweenness centrality of groups of nodes as importance measures. In this paper, the approach is extended to consider the betweenness centrality of groups of edges and the variation in network connection efficiency [16] for identifying the critical groups. For the type of system under study

* Corresponding author.

E-mail address: croccouv@gmail.com (C.M. Rocco S.).

this point of view is more realistic since in power systems, transmission lines (edges) are more exposed to attacks than nodes (substations).

The identification of the most critical groups of edges of different sizes in a network is a NP-complete combinatorial problem, which may be effectively tackled by heuristic procedures, such as evolutionary search algorithms. In this paper, Genetic Algorithms (GAs) are used within a multi-objective formulation of the search problem, in which the decision variables are the groups of edges and the objectives are to maximize the importance of the groups, while minimizing their cardinality. This formulation guides the search towards the identification of the most important single components, couples of components, triplets and so forth. In other words, the problem is a multi-objective decision problem and the genetic algorithms are used to search for Pareto optimal solutions (i.e., non-dominated solutions): in this sense, several authors speak of multi-objective genetic algorithms, MOGA [17,7].

The results are compared with those reported in [20,19]; in the first reference, the authors adopt a DC power flow to assess the behavior of the power system under a limited set of edges outages whereas in the second reference the most critical groups of edges are identified by evaluating the maximum flow in the power network, within a multiobjective formulation. Both papers try to model the network considering the physical phenomenon. In the case of [20], they used a DC power flow, a simplified model that take into consideration the impedances of the network, the transmission line capacities and the load and generation nodes. In the approach proposed by [19], a proxy of the power flow is used, by evaluating the maximum flow in the network: its value corresponds to the total load of the system. This approach requires in addition to the topology of the network, the capacity of the transmission lines, the load and generation nodes.

The paper is organized as follows. The next Section 2 presents the measures used to quantify the importance of groups of edges. Section 3 presents the proposed optimization approach. In Section 4, the approach is applied to the Italian high-voltage (380 kV) electrical transmission network. Conclusions on the outcomes of the analysis are drawn in Section 5.

2. Centrality measures

The network system under analysis is modeled as a graph $G(V,E)$ where V represents the set of vertexes (or nodes) ($N = \dim(V)$ is the number of nodes) and E represents the set of edges (i, j) ($K = \dim(E)$ is the number of edges). The graph is represented by its $N \times N$ adjacency (connection) matrix $\{a_{ij}\}$ whose entries are 1 if there is an edge joining nodes i and j , 0 otherwise. Associated with each link (i, j) is a $K \times K$ matrix $\{\gamma_{ij}\}$, describing the capacities of the links (i, j) . Additionally, a node is classified as a generating source (i.e., power source) or a load (i.e., power demand) or simply a junction node (neither generating nor demanding power).

To evaluate the role played by the edges of the network with respect to its connectivity, two quantitative measures are considered in this paper.

2.1. Group betweenness centrality measure

Edge betweenness was first proposed by Girvan [14] as a measure of centrality used to detect community structure in networks of various kinds. This measure is based on the idea that an edge is central if it is traversed by many of the shortest paths connecting pairs of nodes, from generation nodes to load nodes. Edge betweenness is a measure of the influence of an edge over

the flow of information in the network, especially in cases where information flow follows the shortest available path. Note that this measure does not evaluate the magnitude of the flow through the shortest path.

By generalizing this concept, the betweenness centrality of a group of edges, denoted by $C^B(g)$, can be defined as follows:

$$C^B(g) = \frac{\sum_{i,j \in G, i < j} (s_{ij}(g)/s_{ij})}{(N - \dim(g) - 1)(N - \dim(g))} \quad (1)$$

where g is the subset of edges of the graph, and $s_{ij}(g)$ represents the number of geodesics (i.e., the shortest paths) connecting i to j that pass through g . In other words, the group betweenness centrality measure indicates the proportion of geodesics connecting pairs of non-group members that pass through the group. In the above equation, the sum is taken over all pair of nodes. The measure is normalized by dividing by the theoretical maximum value. If there is more than one shortest path between a pair of nodes, each path is given equal weight such that the total weight of all of the paths is unity. Note that all the information required to evaluate this measure is contained in the adjacency matrix of the network.

2.2. Global efficiency relative variation importance

The global efficiency of the graph $E_f(G)$ representing the network is defined as [16]:

$$E_f(G) = \frac{1}{N(N-1)} \sum_{i,j \in G, i \neq j} \frac{1}{d_{ij}} \quad (2)$$

where $1/d_{ij}$ is the efficiency of the connection between nodes i and j in terms of the number of edges on the shortest path linking the two nodes. It relates the importance of an edge to the impact on the network transmission performance of losing to failure the edges of a group. The relative variation of the global efficiency due to the removal of a group g of edges is computed as the difference between the global efficiency of the network with all the edges of the group removed and the global efficiency of the original network, normalized to the latter value: $(E_f(G) - E_f(G'))/E_f(G)$; $G' = G - g$. This value can be interpreted as a measure of importance of the group of edges removed [9].

As for the previous measure, all the information required to evaluate $E_f(G)$ is contained in the adjacency matrix of the network.

3. Multi-objective genetic algorithm approach

Genetic Algorithms (GAs) are powerful heuristics that have been successfully used in optimization and search problems. The terminology adopted in GAs contains many terms borrowed from biology, suitably redefined to fit the algorithmic context.

Thus, GAs operate on a set of (artificial) chromosomes, which are strings of numbers, generally sequences of binary digits (bits) 0 and 1, coding the values of the decision variables. The values of one or more objective functions in correspondence of the values of the decision variables of a chromosome, define the fitness of that chromosome. The GA search is performed by constructing a sequence of populations of chromosomes, the individuals of each population being the children of those of the previous population and the parents of those of the successive population.

The initial population is generated by randomly sampling the bits of all the strings; at each step in the search sequence, the new population is obtained by probabilistically manipulating the strings of the old population with fitness-improving rules, which mimic genetic evolution. The search sequence continues until a pre-established optimality termination criterion is reached.

The string manipulation consists in selecting and mating pairs of chromosomes in order to groom chromosomes of the next population. This is done by repeatedly performing on the strings the four fundamental operations of reproduction, crossover, replacement and mutation, all based on random sampling: the parents selection step determines the individuals which participate in the reproduction phase; reproduction itself allows the exchange of already existing genes whereas mutation introduces new genetic material; the substitution defines the individuals for the next population. This way of proceeding enables to efficiently arrive at optimal or near-optimal solutions. Since the initial population is generated at random, it is important to perform several runs in order to assess the convergence of the heuristic. As with any other heuristic, the approach cannot be claimed to converge for every problem setting.

In the case of a multiobjective optimization problem, two or more possibly conflicting objective functions $f_i(\cdot)$, $i=1, 2, \dots, n_f$, must be evaluated in correspondence with each decision variable vector U in the search space. In this case, the GA search proceeds by comparing the solutions in terms of the concepts of Pareto optimality and dominance [15]. The decision variable vectors which are not dominated by any other of a given set are called non-dominated with respect to this set while the decision

variable vectors that are non-dominated within the entire search space are said to be Pareto optimal and constitute the so called Pareto optimal front, which is the object of the optimization.

3.1. Genetic algorithm for identifying critical groups

Consider a network $G(V,E)$ and one of the measure defined by (1) or (2). Let

$$C(x_1, x_2, \dots, x_K) \tag{3}$$

represent the importance measure selected when different edges are grouped, with $x_i = 1$ if edge i belongs to the group, $x_i = 0$ otherwise. For example, in a network with $K=3$ edges and using (1), $C(1,0,0)$ indicates the betweenness importance of the first edge alone; $C(0,0,1)$ indicates the importance of the third edge alone; $C(1,0,1)$ indicates the betweenness importance of the group made of the first and third edges.

In a network with K edges, the number of groups (single edges, pairs, triplets and so forth) that in principle can be formed is 2^K . A complete analysis of all groups to find the most critical is therefore impractical for large networks [26]. This fact suggests devising appropriate heuristic procedures to solve the problem, as the one based on GA here proposed.

To overcome this obstacle, the task of determining the most critical groups of edges can be framed as a multiobjective (MO) optimization problem with respect to the two following objectives:

$$\text{Max}_x f_1(x) = C(x_1, x_2, \dots, x_K) \tag{4}$$

$$\text{Min}_x f_2(x) = \sum_{i=1}^K x_i \tag{5}$$

Table 1
Examples of chromosome-coding of groups of edges in a network of 10 edges.

Chromosome coding (10 bits)										Corresponding edges in the group
1	2	3	4	5	6	7	8	9	10	
0	0	0	1	0	0	1	0	1	0	4,7,9
1	0	0	0	0	1	0	1	0	1	1,6,8,10

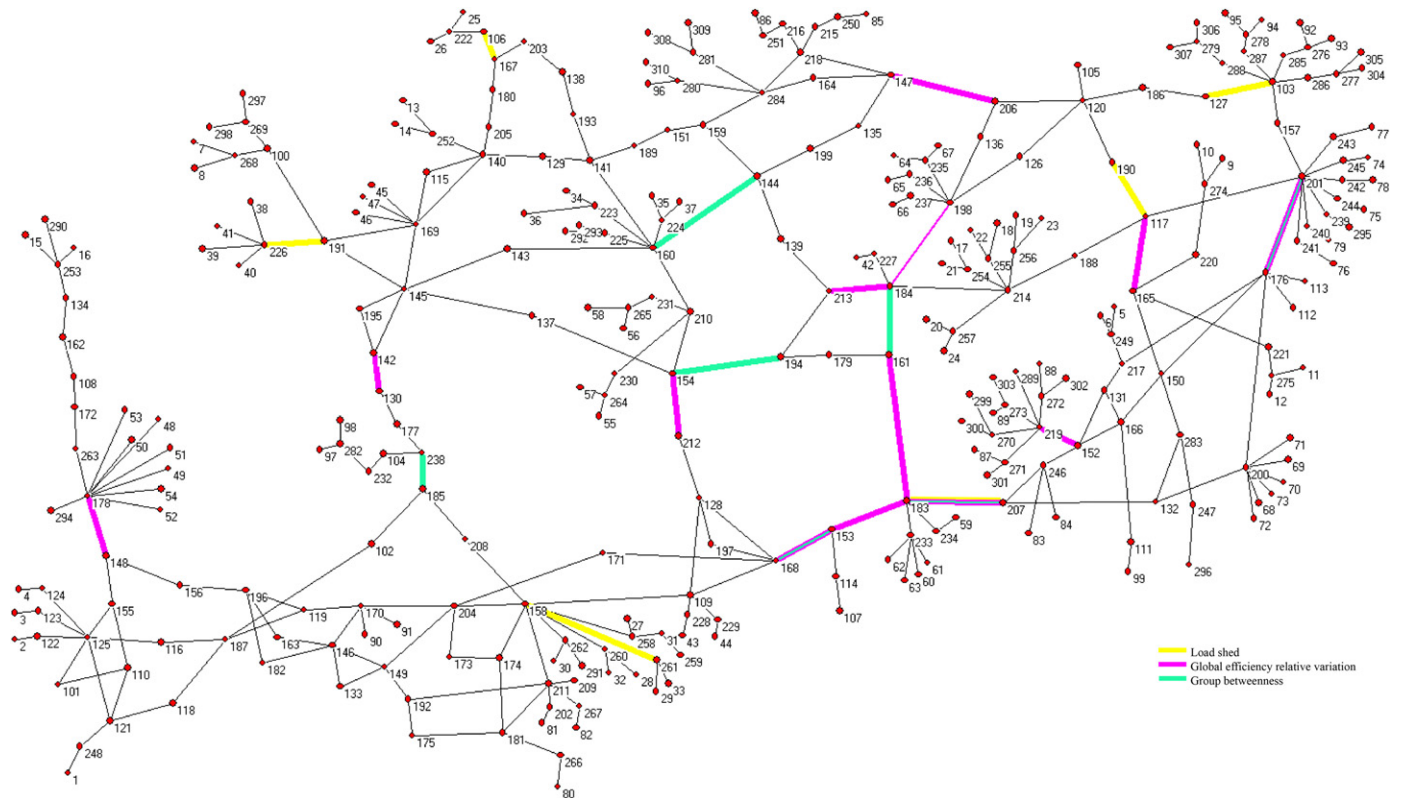


Fig. 1. Italian high-voltage transmission network, drawn using Pajek [2].

As a term of comparison, the detailed approach presented in [20] recognizes that the edges critical for electrical transport are those connecting nodes 103, 117, 127, 184, 190 and 214; these edges also form a subset of the screening results obtained in [19] (Table 5, columns 1–3). The results summarized in Table 5 show that the most critical groups identified in [19] in terms of load shed are assembled around the group of edges 184–214, 183–207, 103–127 and 117–190. These edges are concentrated on the north–east side of the network in Fig. 1, with edges 184–214 and 183–207 in the center of this area which has turned out to be the most vulnerable part of the network system according to the group measures considered in the multiobjective GA search. A significant drop in efficiency (–32.2%, Table 5, column 5, of a maximum of –34.7%, Table 3, column 1) was obtained in the case of interdiction of the edges composing the most important group of size four in terms of load shed, leading to the conclusion that the failure of the edges affects the network both from a structural and functional point of view.

The global efficiency relative variation results of Table 5 show a significant drop of the efficiency of the network for the most critical groups identified, of size four to seven. This leads to the conclusion that the removal of specific edges redirects the flow towards the shortest topological paths.

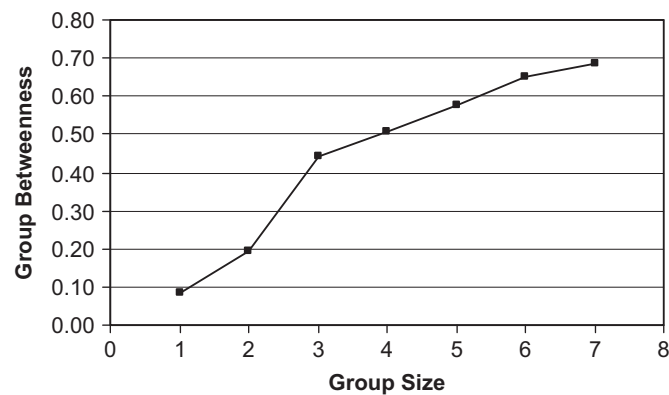


Fig. 2. Pareto front approximation: group size vs betweenness.

Table 4

Pareto optimal results of the multi-objective search for betweenness centrality groups and values of the efficiency variation of the corresponding edges.

Group betweenness—Pareto	Group size	Edges out (from-to-node)							Global efficiency relative variation (%)
0.758	7	153–168	144–160	154–194	128–212	117–201	185–238	183–207	–35.8
0.707	6	153–168	144–160	154–194	128–212	176–201	185–238		–33.9
0.65	5	153–168	144–160	154–194	128–212	176–201			–16.8
0.584	4	153–168	144–160	154–194	128–212				–15
0.504	3	153–168	144–160	154–194					–12.7
0.384	2	153–168	144–160						–4.9
0.258	1	153–168							–3.7

Table 5

Pareto optimal results of the multiobjective search, and corresponding group betweenness and global efficiency relative variation.

Load shed—Pareto	Group size	Edges out (from-to-node)							Group betweenness	Global efficiency relative variation (%)
26	7	106–167	103–127	184–214	183–207	117–190	158–261	191–226	0.52	–36.9
23	6	106–167	103–127	184–214	183–207	117–190	158–261		0.49	–34.9
21	5	106–167	103–127	184–214	183–207	117–190			0.47	–33.4
16	4		103–127	184–214	183–207	117–190			0.45	–32.2
12	3		103–127	184–214	183–207				0.42	–8.4
8	2		103–127	184–214					0.16	–2.8
4	1	106–167							0.025	–1.5

Finally, the outcomes of the analysis of the critical groups of edges are qualitatively similar to those obtained in [20]: a limited number of “central” edges with “functional” relevance for the network is discovered through the topological analysis of its graph. Also, the most critical groups of edges identified in [19] in terms of the load shed and those identified in this paper with respect to the edge betweenness expand around a core of edges, which can be subject to further analysis by simulation of AC/DC load flow.

On the other hand, differences in the results exist due to the different model assumptions driving the two models. In [1], the maximum flow reduction is used as performance function for carrying out an analysis which is a proxy of the physical analysis with additional information requirements. The search is done aiming at the minimum set of edges to interdict for maximizing the load shed. Note that the group of size four could be considered a fourth-order cut set for the north–east side of the network. In this view, the reason why the critical group search performed in this paper does not identify edges 117–190 as critical is explained by the fact that these nodes are high-generation capacity nodes and as such they are identified as critical in [19]: on the contrary, the measures (1) and (2) driving the criticality search performed in this paper do not consider capacity.

5. Conclusions

In this paper, a GA-based procedure for the vulnerability analysis of a network system has been devised to identify the most critical groups of edges of different sizes in the network. The GA search for the most important groups has been framed as a multi-objective optimization problem whose decision variables are the edge group compositions and the objectives are the maximization of the importance of the groups and the minimization of group cardinality.

The Italian high-voltage electrical transmission network (HVIET) has been taken as case study for the analysis of the importance of groups of edges considering the network structure and flow. The group betweenness centrality measure and the global efficiency relative variation have been used as importance indicators to underline the structural vulnerabilities. The results

have been compared with the results of two approaches previously presented in the literature, obtained from a more realistic functional vulnerability analysis of the network: one based on the evaluation of the load flow through a DC load flow model; the other using the maximum flow in the network, as a proxy of the real load flow. It has been shown that the proposed approach requires minimum information of the network (only the topology of the network) and leads to the identification of a number of critical edges that have 'functional' relevance and provides complementary information, which could be used as a preliminary screening analysis.

As a general remark, the proposed procedure allows for a preliminary analysis for identifying groups of critical edges by analyzing only the topology of the network. This requires a minimum of information. The analysis produces a set of critical elements that must be analyzed by appropriate procedures that model the physical phenomenon of the network. The results suggest that a preliminary set of critical edges can be provided by the proposed approach. For obtaining realistic insights on the robustness and vulnerability of realistic electrical transmission systems to faults and attacks, the analysis focused on the topological features of the network must be expanded to include also additional physical characteristics. In this respect, work is currently undergoing in establishing effective ways of bringing different physical characteristics into the topological analysis, e.g. the reliability and electrical characteristics of the transmission edges in a power network.

References

- [1] Albert R, Albert I, Nakarano GL. Physical Review E 2004;69:025103(R).
- [2] Batagelj V, Mrvar A. Pajek—program for large network analysis. *Connections* 1998;21:47–57.
- [3] Bier VM, Gratz ER, Haphuriwat NJ, Magua W, Wierzbicki KR. Methodology for identifying near-optimal interdiction strategies for a power transmission system. *Reliability Engineering and System Safety* 2007;92(9):1155–61.
- [4] Birchmeier J. Systematic Assessment of the Degree of Criticality of Infrastructures. *Proceedings of the ESREL 2007, Stavanger, Norway*, vol. 1, 25–27 June 2007, p. 859–64.
- [5] Cadini F, Zio E, Petrescu CA. Using centrality measures to rank the importance of the components of a complex network infrastructure. *Lecture Notes in Computer Science* 2009;5508/2009:155–67.
- [6] CNIP'06. *Proceedings of the International Workshop on Complex Network and Infrastructure Protection*, Rome, Italy, 28–29 March 2006.
- [7] Coello Coello CA, Lamont GA, Van Veldhuizen DA. *Evolutionary algorithms for solving multi-objective problems*. 2nd ed. Springer; 2007.
- [8] Cohen R, Erez K, ben-Avraham D, Havlin S. Resilience of the Internet to random breakdowns. *Physical Review Letters* 2000;85(21):4626–8.
- [9] Crucitti P, Latora V, Marchiori M. Locating critical lines in high-voltage electrical power grids. *Fluctuation Noise Letters* 2005;5:L201–8.
- [10] Crucitti P, Latora V, Marchiori M, Topological A. Analysis of the Italian Electric Power Grid. *Physica A* 2004;338:92–7.
- [11] Eusgeld I, Kroger W, Sansavini G, Schlapfer M, Zio E. The role of network theory and object-oriented modeling within a framework for the vulnerability analysis of critical infrastructures. *Reliability Engineering and System Safety* 2009;94(5):954–63.
- [12] Dobson I, Carreras BA, Lynch VE, Newman DE. Complex systems analysis of series of blackouts: cascading failure, critical points, and self-organization. *Chaos: An Interdisciplinary Journal of Nonlinear Science* 2007;17(2):026103.
- [14] Girvan M, Newman MEJ. Community structure in social and biological networks. *Proceedings of the National Academic Science, USA* 2002;99(5890).
- [15] Goldberg DE. *Genetic algorithms in search, optimization, and machine learning*. Addison-Wesley Publ. Co; 1989.
- [16] Latora V, Marchiori M. Efficient behavior of small-world networks. *Physical Review Letters* 2001;87(19).
- [17] Murata T, Ishibuchi H. MOGA: multi-objective genetic algorithms. *IEEE International Conference on Evolutionary Computation*, 1995. p. 289.
- [18] Rocco CM, Zio E, Salazar DE. Multi-objective Evolutionary Optimisation of the Protection of Complex Networks Exposed to Terrorist Hazard. *Proceedings of ESREL 2007, Stavanger, Norway*, vol. 1, 25–27 June 2007. p. 899–905.
- [19] Rocco CM, Ramirez-Marquez JE, Salazar DE, Zio E. A Flow Importance Measure via Multiple-Objective Optimization and its Application to an Italian Transmission Power System, accepted for publication in *International Journal of Performability Engineering*, 2009.
- [20] Rosato V, Issacharoff L, Bologna S. Influence of the topology on the power flux of the Italian high-voltage electrical network. *Europhysics Letters* 2006.
- [21] Tiriticco F, Bologna S, Rosato V. *Electric Power Systems Research* 2006.
- [22] ESREL, 2007 Vulnerability, Reliability and safety of Complex Networks and Critical Infrastructures, Special Sessions I and II. *Proceedings of ESREL 2007, Stavanger, Norway*, vol. 1, 25–27 June 2007.
- [23] Zheng J-F, Gao Z-Y, Zhao X-M. Clustering and congestion effects on cascading failures of scale-free networks. *EPL (Europhysics Letters)* 2007(5):58002.
- [25] Zio E, Petrescu CA, Sansavini G. Vulnerability analysis of a power transmission network. *Proceedings of PSAM9—International Probabilistic Safety Assessment and Management Conference, Hong Kong, China*, 18–23 May 2008.
- [26] Zio E, Golea LR. Identification of betweenness-central groups of components in a complex network infrastructure by genetic algorithms, accepted, *ESREL Safety and Reliability for Managing Risk, Prague, Czech Republic*, 7–10 September 2009.
- [27] Bompard E, Napoli R, Xue F. Analysis of structural vulnerabilities in power transmission grids. *International Journal of Critical Infrastructure Protection* 2009;2(1–2):5–12.