SIXTH FRAMEWORK PROGRAMME

Project contract no. 043363

MANMADE
Diagnosing vulnerability, emergent phenomena and volatility in man-made networks

SPECIFIC TARGETED PROJECT

NEST PATHFINDER
Sub-Priority Tackling Complexity in Science

M36: Deliverable 6.4 A report on simulation of the dynamics (resilience and fragmentation) resulting from graph erosion of a realistic interconnected system

Revision [1]

Submission date: December, 2009

Start date of project: 1st of January 2007 Duration: 36 months

Lead authors for this Report: D.K. Arrowsmith (QMUL), L.Kocarev(MASA), Z. Farkas(COLB).

Project co-funded by the European Commission within the Sixth Framework Programme (2002-2006)		
Dissemination Level		
PU	Public	X
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	

PURPOSE OF THIS REPORT

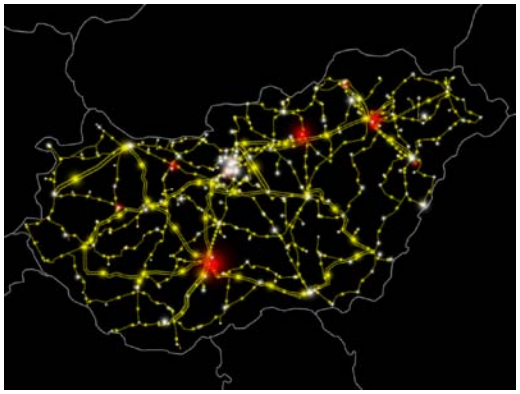
This deliverable is concerned with looking at realistic dynamical models which undergo disintegration according to various scenarios. A first report deals from MASA deals with vulnerability analysis (resilience and fragmentation) of realistic interconnected systems. Realistic interconnected systems are modelled as graphs with node annotations. Real-world networks which can be viewed as having labeled nodes are numerous and include EU power grid network, taxonomies of Autonomous systems in the Internet, the Wikipedia article graph where articles belong to various categories, sources and sinks in Europe's gas network, maps of university relationships from different countries. The second project is concerned with network evolution related to grid supplies. COLB has been concerned with the dynamics of electricity transmission and has developed a model involving constrained system of linear transmission differential equations using linear programming to develop network evolution scenarios. Basically the problem of overload in network connections is addressed as to how that can propagate through a network as over line take the excess capacity.

A. **A report on simulation of the dynamics (resilience and fragmentation) resulting from graph erosion of a realistic interconnected system** D. Trpevski, D. Smilkov, L. Kocarev, D. Trajanov, A. Bogoeska, M. Angelova , Macedonian Academy of Sciences, Skopje, Macedonia

B. Power grid dynamics and cascade breakdown Mark Szenes, Zeno Farkas, Gabor Papp, Collegium Budapest< Hungary.

Regular and breakdown dynamics of power grids is investigated. We have spent considerable time on formulating a model for this purpose. Such a model should be simple, yet capable of capturing the most fundamental properties of power flow. Initially, we used the DC load flow model. However, this turned out to be incapable of taking into account the maximum transmission capacity of power lines correctly. After that we switched to a linear programming model, which proved to be an excellent tool for our investigations. The technical details of this model are described in [1]. With this linear programming model we are able to:

1. Calculate the power flow under regular conditions.
2. Identify the power lines which typically operate close their maximum power capacity, hence becoming a bottleneck in the power flow.
3. When the power grid is unbalanced, we can identify the overloaded power lines. Using an iterative process we can model the breakdown of the power grid up to the point when it regains its balance. At this point the initially interconnected power grid may be fragmented into several independent areas or completely shut down.
4. Although initially it was not our goal, it turned out that this model is capable of calculating not only the power flow, but also the average and spot prices of power production.



As input data, we have used the Platts database provided by JRC, and also some data collected by ourselves. At this point we can present an animated picture showing the power flow in Hungary (for this country we could collect more detailed data - necessary for our model - than found in the Platts database) and a graphical representation of the UCTE network.

D6.4 A report on simulation of the dynamics (resilience and fragmentation) resulting from graph erosion of a realistic interconnected system

D. Trpevski, D. Smilkov, L. Kocarev,
D. Trajanov, A. Bogoeska, M. Angelova

I. SUMMARY

This report deals with the vulnerability analysis (resilience and fragmentation) of realistic interconnected systems. We model realistic interconnected systems as graphs with node annotations, where a node can have one of k distinct labels, or colors. Real-world networks which can be viewed as having labeled nodes are numerous and include EU power grid network, taxonomies of ASes in the Internet, the Wikipedia article graph where articles belong to various categories, sources and sinks in Europe's gas network, maps of university relationships from different countries, and so on.

We specifically consider the following tasks:

- T6.3 Theoretical analysis of vulnerability of interconnected grids of differing topologies.
- T6.4 Analysis of the effect of scaling (number of nodes and lines) on the vulnerability for given grid topology types.
- T6.5 Verify or otherwise the scale invariance network topology of real European electricity and gas grids.

II. INTRODUCTION

Many complex systems in the real world can be conceptually described as networks, where nodes represent the system constituents and edges depict the interaction between them. Often enough, the network representation of these systems is an undirected and unweighted graph which greatly simplifies the structure of complex systems. Attempts have been made to include network specific details by annotating graphs representing networks in various ways [1]. In this way heterogeneous networks where different entities may have diverse relationships are more appropriately described.

This report specifically deals with the vulnerability of graphs with node annotations, where a node can have one of k distinct labels, or colors. Real-world networks which can be viewed as having labeled nodes are numerous and include taxonomies of ASes in the Internet, the Wikipedia article graph where articles belong to various categories, sources and sinks in Europe's gas network, maps of university relationships from different countries [2–5], etc.

Vulnerability concerns the issue of network robustness to random failures or malicious attacks. Two aspects of vulnerability are met in the literature: one is the topological, or structural aspect which investigates the effect of removing nodes or edges on the network connectivity [6–10]; the other is the dynamical aspect where a model of entity flow (e.g. electrical current, gas, or Internet data) is presumed for the network, and the effect of removing nodes or edges on the network flow is studied [2, 11–14]. We deal with the structural aspect in this study. Additionally, on a real-world example we show how the measure of vulnerability used in this paper, the characteristic path length, relates to the interconnectedness between nodes of different labels.

The report proceeds as follows. In section III we define vulnerability for graphs with node annotations. Section IV investigates complex networks with node annotations. Section V deals with EU power grid network.

III. AVERAGE EDGE BETWEENNESS AS A MEASURE OF VULNERABILITY

A. Edge betweenness and characteristic path length

Several measures for network vulnerability have been proposed. It is easily shown [15] that a proper vulnerability measure should be derived from the link betweenness and hence, we follow this approach.

Consider an undirected, unweighted graph $G = (V, E)$, where V represents the node set and E is the edge set. The betweenness of edge e is defined as:

$$b_e = \sum_{i,j \in V, i \neq j} \frac{n_{i,j}(e)}{n_{i,j}}. \tag{1}$$

$n_{i,j}$ denotes the number of shortest paths between nodes i and j, and $n_{i,j}(e)$ is the number of those paths containing e. Thus, it measures the extent to which edge e is involved in the shortest paths between nodes. The average edge betweenness of the graph is:

$$b(G) = \frac{1}{|E|} \sum_{e \in E} b_e. \tag{2}$$

The average edge betweenness is related to the characteristic path length of a graph, which represents an average of all the geodesic distances (shortest path lengths). In the case of multiple shortest paths between any two nodes $i, j \in V$, the characteristic path length is defined as

$$L(G) = \frac{1}{|V|(|V|-1)} \sum_{i,j \in V, i \neq j} \frac{\sum_{g \in P_{i,j}} d_{i,j}}{n_{i,j}}, \tag{3}$$

where $P_{i,j}$ is the set of all geodesics between nodes i and j, and $d_{i,j}$ is the length of each such geodesic distance. $L(G)$ is related to $b(G)$ by

$$b(G) = \frac{|V|(|V|-1)}{|E|} L(G) \tag{4}$$

Essentially the two quantities express the same information about the graph G. Since for a graph with a fixed number of nodes $b(G)$ and $L(G)$ decrease as the number of edges in the graph increases, it can be said that they represent how "well connected" a graph is. The higher the values of $b(G)$ and $L(G)$, the more vulnerable G is to loss of edges. In the next section we ask whether this statement holds for labeled graphs as well.

B. Labeled graphs

Further, let G be a labeled graph where each node can have one of k distinct labels (classes), i.e. $V = \cup_{i=1}^{k} V_i$ and $V_m \cap V_n$ is not necessarily empty. In such a setting a valid question concerns the vulnerability of communication between nodes of specific labels. For example, consider the graphs on Fig. 1. Although all of them represent cycles, the difference between them is obvious. While the leftmost graph is unlabeled, nodes of the same label are two hops away in the middle graph, and one hop away in the rightmost graph. Clearly, communicating with a node of the same label requires more edges in the middle graph G_2 than in the rightmost graph G_3, which intuitively makes this communication in G_2 more vulnerable to loss of edges.

To formalize the difference between such graphs, the first step is choosing which of the previously defined metrics is suitable for the matter. We require that a metric for vulnerability of labeled graphs has the following properties:

- Considering a labeled graph as an unlabeled one (i.e. having only one label), it reduces to the corresponding metric for an unlabeled graph.
- It has the same value for all possible labellings in a fully connected graph, reflecting the notion that any arbitrarily labeled fully connected topology is equally vulnerable as any other.

Consider the edge betweenness in (1) first, which can be decomposed as:

$$b_e = \sum_{m,n \in \{1,\ldots,k\}} b_e^{V_m, V_n}, \tag{5}$$

where

$$b_e^{V_m,V_n} = \sum_{i\in V_m, j\in V_n} \frac{n_{i,j}(e)}{n_{i,j}}, i \neq j \tag{6}$$

indicates the importance of the edge in the shortest path communication between nodes of sets V_m and V_n, $m, n \in \{1, \ldots, k\}$. Consequently, (2) can be rewritten as:

$$b(G) = \sum_{m,n\in\{1,\ldots,k\}} b(G)^{V_m,V_n}, \tag{7}$$

where

$$b^{V_m,V_n}(G) = \frac{1}{|E|} \sum_{e\in E} b_e^{V_m,V_n}, \tag{8}$$

$m = 1, 2$ and $n = 1, 2$. This stresses the relative importance of communication between particular labels of nodes to the overall betweenness of the graph. In the case of two labels, such as in the example on Fig. 1, (5) and (7) become:

$$b_e = b_e^{V_1,V_1} + b_e^{V_1,V_2} + b_e^{V_2,V_1} + b_e^{V_2,V_2} \tag{9}$$

and

$$b(G) = b^{V_1,V_1}(G) + b^{V_1,V_2}(G) + b^{V_2,V_1}(G) + b^{V_2,V_2}(G), \tag{10}$$

respectively. For the graphs on Fig. 1 one obtains:

$$\begin{aligned} b(G_1) &= b(G_2) = b(G_3) = 2, \\ b^{V_1,V_1}(G_2) &= b^{V_2,V_2}(G_2) = \frac{1}{2}, \\ b^{V_1,V_2}(G_2) &= b^{V_2,V_1}(G_2) = \frac{1}{2}, \\ b^{V_1,V_1}(G_3) &= b^{V_2,V_2}(G_3) = \frac{1}{4}, \\ b^{V_1,V_2}(G_3) &= b^{V_2,V_1}(G_3) = \frac{3}{4}. \end{aligned}$$

The results would indicate that in G_2, on average, an edge participates equally in all kinds of shortest path communication, which would also suggest that all communication is of equal vulnerability. Clearly, when $k = 1$, (10) reduces to (2). However, for an arbitrarily labeled fully connected graph the values of $b^{V_m,V_n}(G)$ depend on the number of nodes in V_m and V_n, which gives an inaccurate result of non-equal vulnerability.

Next, consider the second metric, $L(G)$. For a graph with k distinct labels, the characteristic path length between nodes of class m and n, $m, n \in \{1, \ldots, k\}$ is:

$$L^{V_m,V_m}(G) = \frac{1}{|V_m|(|V_m|-1)} \sum_{i,j\in V_m, i\neq j} \frac{\sum_{g\in P_{i,j}} d_{i,j}}{n_{i,j}} \tag{11}$$

and

$$L^{V_m,V_n}(G) = \frac{1}{|V_m||V_n|} \sum_{i,j\in V_m, i\neq j} \frac{\sum_{g\in P_{i,j}} d_{i,j}}{n_{i,j}} \tag{12}$$

when $m \neq n$.

Calculating these for G_2 and G_3 on Fig. 1 gives:

$$\begin{aligned} L^{V_1,V_1}(G_2) &= L^{V_2,V_2}(G_2) = 2, \\ L^{V_1,V_2}(G_2) &= L^{V_2,V_1}(G_2) = 1, \\ L^{V_1,V_1}(G_3) &= L^{V_2,V_2}(G_3) = 1, \\ L^{V_1,V_2}(G_3) &= L^{V_2,V_1}(G_3) = 1.5, \end{aligned}$$

revealing the average distance between nodes of respective classes and giving some indication as to the topology of the graph. In the case of G_2 it suggests the more intuitive result that communication between nodes of the same label is more vulnerable than that of nodes in different labels.

When there is only one label in the graph (11) and (12) clearly reduce to (3). Note that the characteristic path length as defined with (11) and (12) is 1 for an arbitrarily labeled fully connected graph of any size, reflecting the notion that any labeled fully connected topology is equally vulnerable as any unlabeled one. Also note that

$$b^{V_m,V_m}(G) = \frac{|V_m|(|V_m|-1)}{|E|} L^{V_m,V_m}(G),$$

$$b^{V_m,V_n}(G) = \frac{|V_m||V_n|}{|E|} L^{V_m,V_n}(G), m \neq n$$

i.e. the characteristic path length is the normalized average betweenness which is invariant to the number of nodes in V_m and V_n in a fully connected topology. Therefore we propose inspecting the characteristic path lengths between nodes of specific labels to gain insight how well these nodes are interconnected in the sense of shortest path communication. A noteworthy remark is that characteristic path length has been used to inspect a network's efficiency [8, 9]

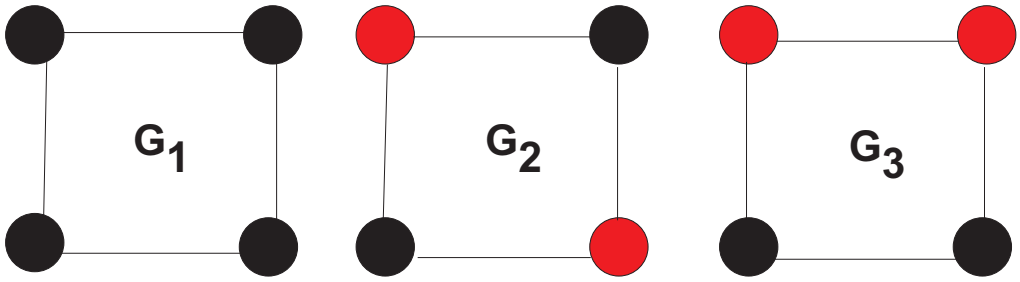


FIG. 1: A graph with three different annotations.

Finally, consider graph G_2 on Fig. 1. Note that there are *two* paths of length 2 connecting nodes of the same label. The characteristic path length is oblivious of this fact. Indeed, imagine deleting an edge form G_2: again $L^{V_1,V_1}(G_2) = L^{V_2,V_2}(G_2) = 2$, but only one path exists between nodes of the same label, intuitively making G_2 with a missing edge more vulnerable than the original G_2. Therefore, in order to determine the vulnerability of a network to edge deletions, one needs to simulate edge loss and calculate how the characteristic path length is affected.

IV. VULNERABILITY OF COMPLEX NETWORKS

A. Erdos-Renyi graph

In this section we investigate 2-labeled Erdos-Renyi (ER) graphs. The model proposed by Erdos and Renyi describes random graphs with N nodes in which every link exists with probability p. The degree distribution of these networks is Poisson, hence the homogeneous structure in the sense that all nodes have degree close to the average degree. Also, in this model there is a critical probability value $p_c = \frac{1}{N}$ under which the resulting network consists of small disconnected components, and above which there is a giant component in the network containing $O(N)$ nodes. All the networks used in the simulations are generated with $p > p_c$, and only the giant component is considered.

Fig. 2 depicts the characteristic path length between classes of nodes when $|V_1|$ is changed. V_1 contains $f|V|$ rich-club nodes in the graph, $0.2 \leq f \leq 0.98$.

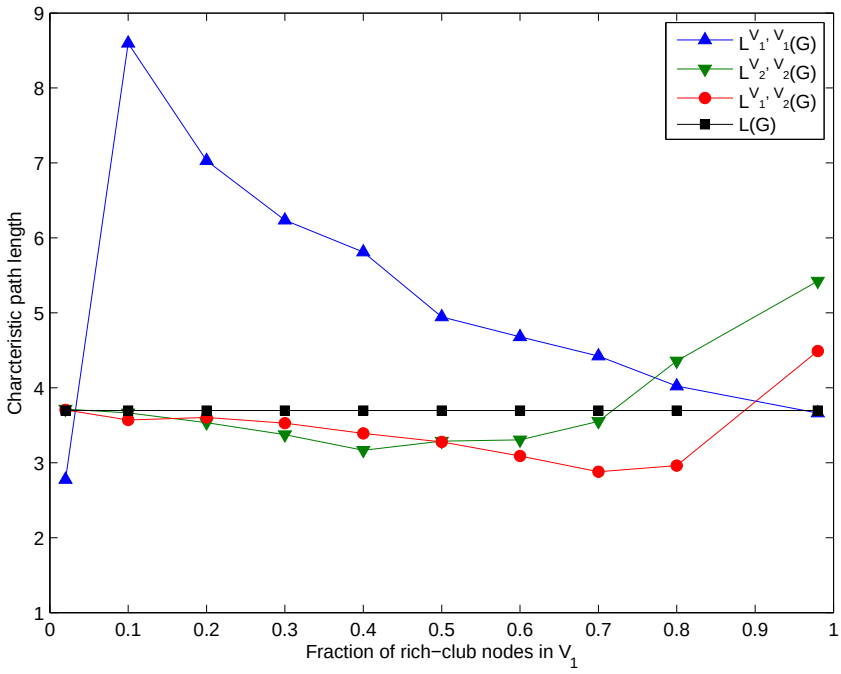


FIG. 2: ER networks are of size $|V| = 500$, generated with $p = 0.01$. Results are averages of 5 network realizations.

Fig. 3 depicts the results when V_1 contains $f|V|$ randomly chosen nodes, $0.2 \leq f \leq 0.98$.

B. Barabasi-Albert power-law graph

This section investigates the vulnerability of 2-labeled power-law graphs generated by the Barabasi-Albert (BA) model. The original BA algorithm as given in [18] is used to construct scale-free networks. One starts from a seed of m_0 connected nodes and adds a new node with $m \leq m_0$ links at each step according to the preferential attachment rule.

Fig. 4 depicts the characteristic path length between classes of nodes when $|V_1|$ is changed. V_1 contains $f|V|$ rich-club nodes in the graph, $0.1 \leq f \leq 0.8$.

Fig. 5 depicts the results when V_1 contains $f|V|$ randomly chosen nodes, $0.2 \leq f \leq 0.98$.

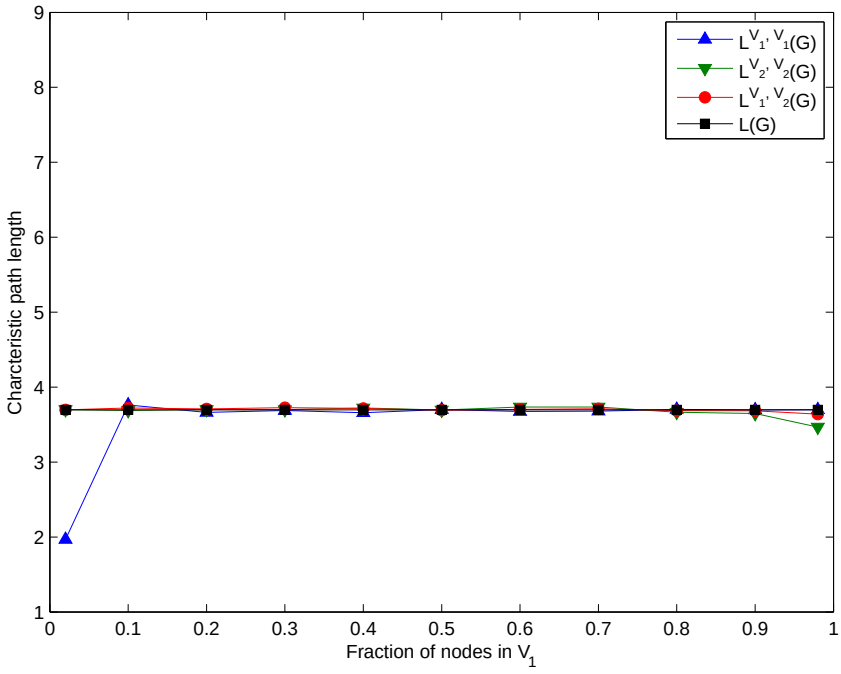


FIG. 3: ER networks are of size $|V| = 500$, generated with $p = 0.01$. Results are averages of 5 network realizations.

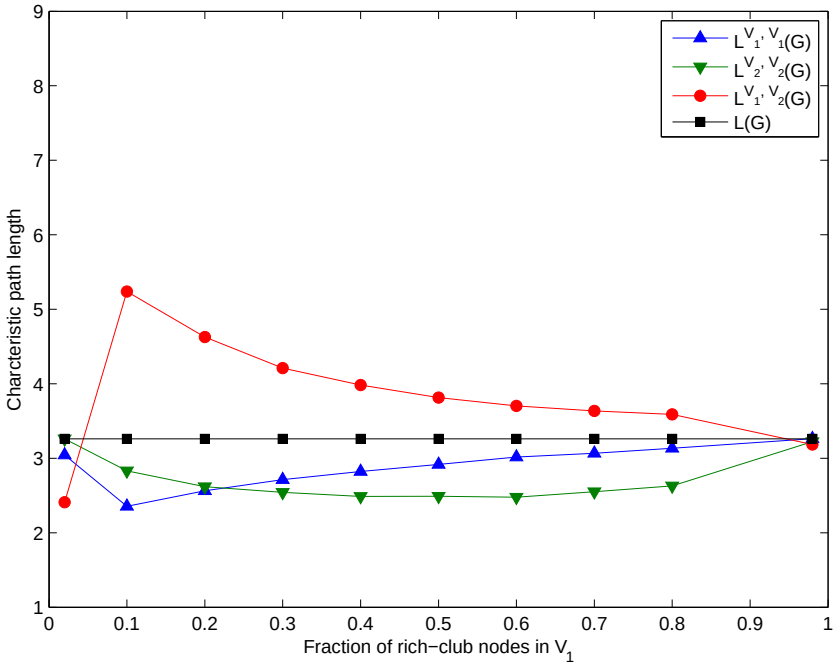


FIG. 4: BA networks are of size $|V| = 500$, generated from a starting fully connected seed of 4 nodes. Results are averages of 5 network realizations.

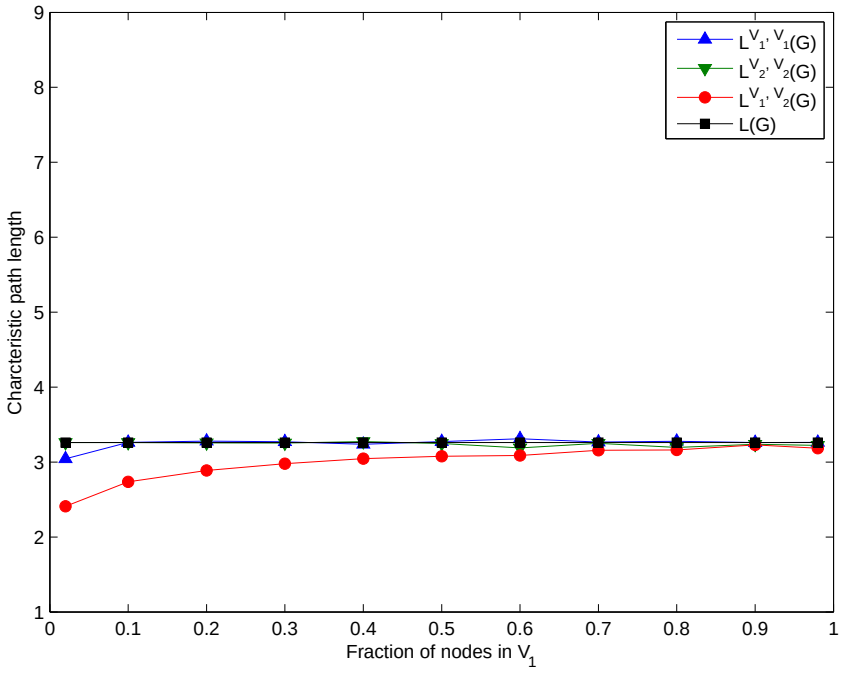


FIG. 5: BA networks are of size $|V| = 500$, generated from a starting fully connected seed of 4 nodes. Results are averages of 5 network realizations.

C. Watts-Strogats small-world graph

We now calculate the vulnerability metric for 2-labeled Watts-Strogatz (WS) small-world graphs. We use the WS model as defined in [17] for generating the networks. The algorithm uses a starting ring lattice to construct a small-world network. In a ring lattice each node has $2K$ neighbors, K in the clockwise and K in the anti-clockwise direction. Each edge is rewired with probability β, not allowing self-loops or multiple edges between nodes.

Fig. 6 depicts the characteristic path length between classes of nodes when $|V_1|$ is changed. V_1 contains $f|V|$ rich-club nodes in the graph, $0.1 \leq f \leq 0.8$.

Fig. 7 depicts the results when V_1 contains $f|V|$ randomly chosen nodes, $0.2 \leq f \leq 0.98$.

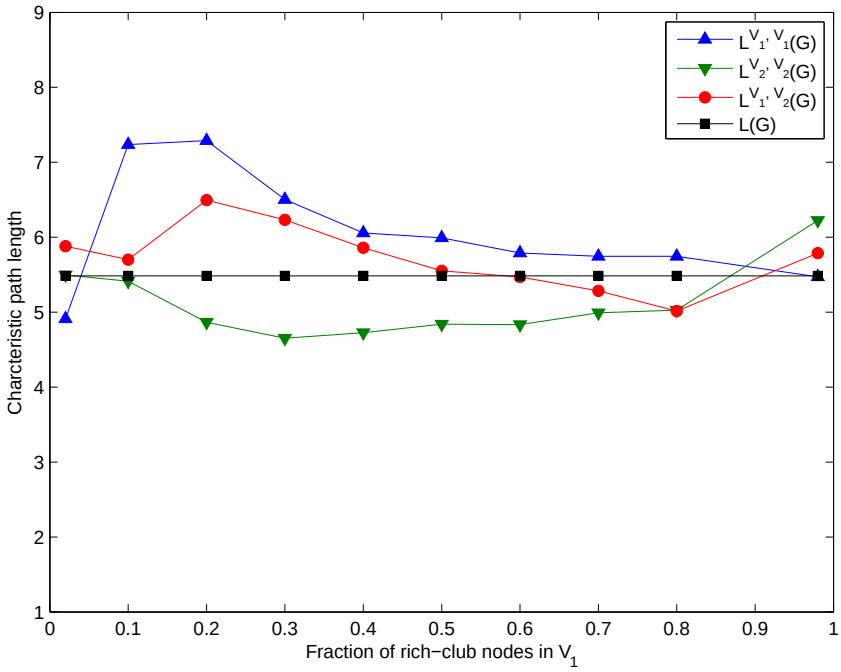


FIG. 6: WS networks are of size $|V| = 500$. The starting ring lattice has $K = 3$, and the rewiring probability is $\beta = 0.1$. Results are averages of 5 network realizations.

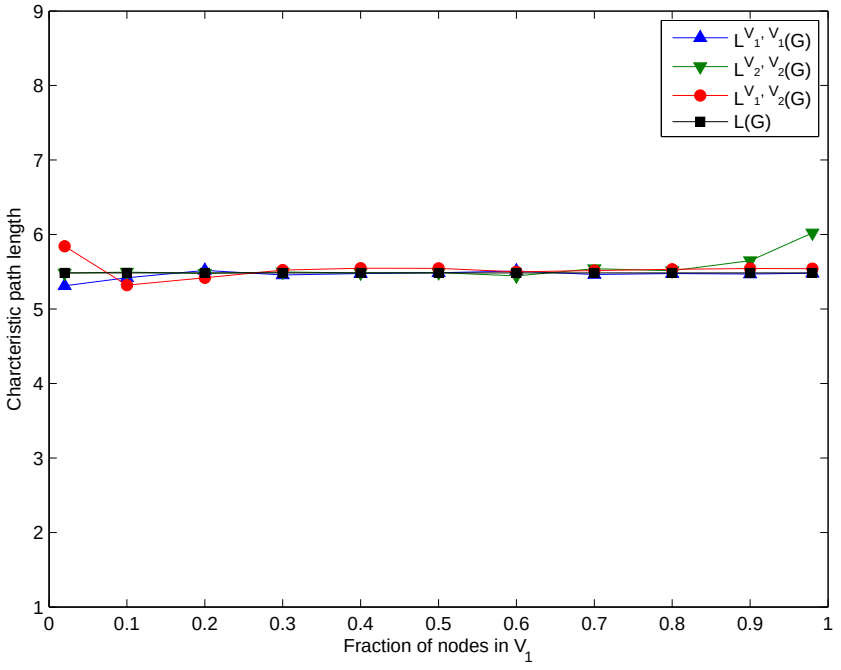


FIG. 7: WS networks are of size $|V| = 500$. The starting ring lattice has $K = 3$, and the rewiring probability is $\beta = 0.1$. Results are averages of 5 network realizations.

V. EU POWER GRID

The main vulnerability analysis is done for a network representing the power grid of the European Union and results are presented in this section. It contains a total of 12741 nodes of the following 6 types:

1. Gas node (a junction, intersection, or endnode) – 2212 nodes
2. Electricity substation – 5096 nodes
3. Natural gas power plant – 998 nodes
4. Generic power plant – 4383 nodes
5. DC line connected substation – 31 nodes
6. Liquefied natural gas (LNG) terminal – 21 nodes

connected by 17798 edges which have been treated as undirected. Figure (8) depicts node degree distribution for EU power grid network. It clearly demonstrates power law distribution for the whole graph (considered as unannotated network) and several annotated subnetworks.

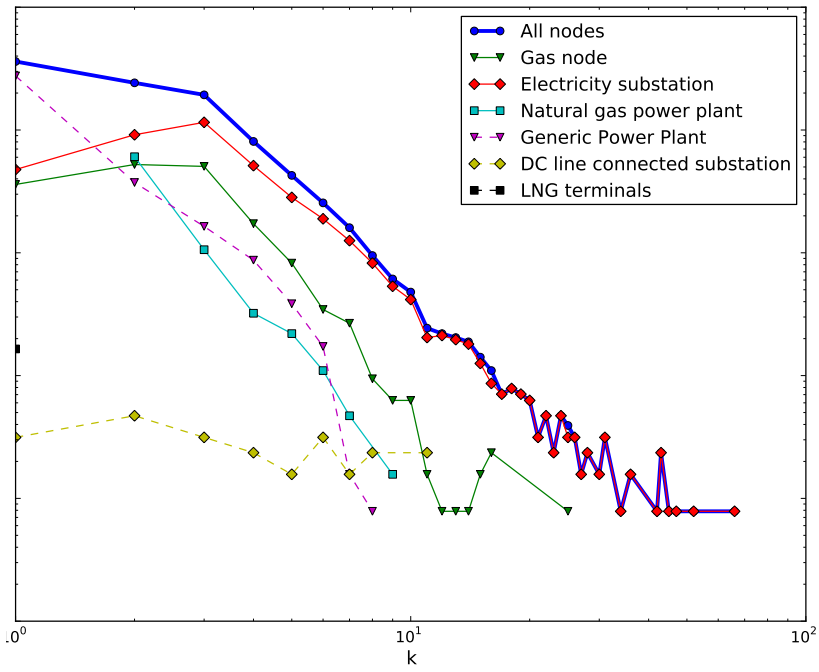


FIG. 8: Node degree distribution for EU power grid.

As the results show, Table I, the characteristic path length between types of nodes ranges approximately from 26 to 39. The most efficiently connected nodes in terms of distance are natural gas power plants, gas nodes, and substations connected by DC lines. In addition, electricity substations and generic power plants are similarly distanced to natural gas power plants. On the other hand, the longest characteristic path lengths are observed between LNG terminals and every other type of node in the network, which might be due to the peripheral geographical position of these terminals in the network as well as to the late appearance of these nodes to the power grid network, which is why communication with these nodes is more vulnerable.

We investigate how the characteristic path length and the size of the giant component change when a certain number of nodes are removed from the network. In the chosen scenario we remove the nodes with the highest load in the network, i.e. the nodes that has the highest value of node betweenness. Tables II and III show the relative increase of the characteristic path length between different types of nodes when gas nodes are removed from the network,

5 lowest $L^{V_m,V_n}(G)$			5 highest $L^{V_m,V_n}(G)$		
$L^{V_m,V_n}(G)$	Label m	Label n	$L^{V_m,V_n}(G)$	Label m	Label n
26.7148	3	3	38.3702	5	6
27.1145	1	1	37.7245	1	6
28.0086	5	5	34.0184	4	6
28.0409	2	3	33.3845	2	6
28.9582	4	3	32.4937	3	6

TABLE I: The 5 most and least interconnected pairs of labels in the EU power grid in the sense of shortest path communication.

considering the initial network. Tables IV and V show the relative increase of the characteristic path length between different types of nodes when electricity substations (nodes with label 2) are removed from the network, considering the initial network. Tables VI, VII show the relative increase of the characteristic path length between different types of nodes when natural gas power plants (nodes with label 3) are removed from the network, considering the initial network. In addition, the same analysis was conducted when generic power plant (nodes with label 4) were removed from the network, but the results showed that the characteristic path length between the different groups of nodes was not change dramatically.

		(a)	(b)
Label m	Label n	Relative increase of $L^{V_m,V_n}(G)$	Relative increase of $L^{V_m,V_n}(G)$
5	6	14.0453	15.4023
3	5	9.1498	10.7676
1	6	9.1098	13.0704
2	5	9.0098	9.5134
1	1	5.8370	12.7163
3	4	5.1850	6.9883
2	2	5.6458	6.6673
3	3	3.5481	6.8931
4	4	6.5332	7.4215

TABLE II: Relative change of the characteristic path length of different groups of nodes when certain number of gas nodes are removed from the network. Column (a): 30 gas nodes with the highest load were removed from the initial network, the size of the giant component is 11641 nodes. Column (b) 70 gas nodes with the highest load were removed from the initial network, the size of the giant component is 11606 nodes.

		(a)	(b)
Label m	Label n	Relative increase of $L^{V_m,V_n}(G)$	Relative increase of $L^{V_m,V_n}(G)$
5	6	16.9635	18.5055
3	5	13.1664	14.5129
1	6	14.7560	13.0704
2	5	10.8980	11.6744
1	1	16.8339	27.3655
3	4	9.2390	10.4122
2	2	8.0257	8.6875
3	3	10.2627	12.1860
4	4	8.8420	9.3912

TABLE III: Relative change of the characteristic path length of different groups of nodes when certain number of gas nodes are removed from the network. Column (a): 200 gas nodes with the highest load were removed from the initial network, the size of the giant component is 11415 nodes. Column (b) 400 gas nodes with the highest load were removed from the initial network, the size of the giant component is 11053 nodes.

		(a)	(b)
Label m	Label n	Relative increase of $L^{V_m,V_n}(G)$	Relative increase of $L^{V_m,V_n}(G)$
5	6	11.5177	12.5906
3	5	10.3707	13.2054
1	6	5.2222	6.0387
2	5	12.2502	16.0107
1	1	1.4399	2.9946
3	4	13.0285	17.4513
2	2	15.1562	20.1681
3	3	7.8608	11.0024
4	4	16.0698	22.0071

TABLE IV: Relative change of the characteristic path length of different groups of nodes when certain number of electricity substations are removed from the network. Column (a): 30 electricity substations with the highest load were removed from the initial network, the size of the giant component is 12662 nodes. Column (b) 70 electricity substations with the highest load were removed from the initial network, the size of the giant component is 12522 nodes.

		(a)	(b)
Label m	Label n	Relative increase of $L^{V_m,V_n}(G)$	Relative increase of $L^{V_m,V_n}(G)$
5	6	20.6173	24.3690
3	5	25.4596	29.4085
1	6	10.2621	11.4672
2	5	31.6198	41.9406
1	1	7.4704	8.5331
3	4	26.1185	39.2006
2	2	30.2494	45.2913
3	3	17.5893	19.5212
4	4	33.7233	55.7777

TABLE V: Relative change of the characteristic path length of different groups of nodes when certain number of electricity substations are removed from the network. Column (a): 200 electricity substations with the highest load were removed from the initial network, the size of the giant component is 11818 nodes. Column (b) 400 electricity substations with the highest load were removed from the initial network, the size of the giant component is 11065 nodes.

		(a)	(b)
Label m	Label n	Relative increase of $L^{V_m,V_n}(G)$	Relative increase of $L^{V_m,V_n}(G)$
5	6	10.0244	11.7339
3	5	9.3797	11.0790
1	6	3.3970	4.6156
2	5	9.2048	10.5484
1	1	1.7394	3.1449
3	4	7.8835	8.9314
2	2	7.6605	8.8304
3	3	5.4568	6.1162
4	4	8.9586	10.1070

TABLE VI: Relative change of the characteristic path length of different groups of nodes when certain number of natural gas power plants are removed from the network. Column (a): 30 gas power plants with the highest load were removed from the initial network, the size of the giant component is 12711 nodes. Column (b) 70 gas power plants with the highest load were removed from the initial network, the size of the giant component is 12671 nodes.

		(a)	(b)
Label m	Label n	Relative increase of $L^{V_m,V_n}(G)$	Relative increase of $L^{V_m,V_n}(G)$
5	6	14.2039	15.5611
3	5	13.3995	16.0155
1	6	5.7340	7.3607
2	5	12.8874	14.3377
1	1	5.0067	6.9847
3	4	10.4365	11.5732
2	2	10.5978	11.8090
3	3	6.5512	6.6192
4	4	11.7993	13.1244

TABLE VII: Relative change of the characteristic path length of different groups of nodes when certain number of natural gas power plants are removed from the network. Column (a): 200 natural gas power plants with the highest load were removed from the initial network, the size of the giant component is 12541 nodes. Column (b) 400 natural gas power plants with the highest load were removed from the initial network, the size of the giant component is 12341 nodes.

Next, we investigate how the characteristic path length and the size of the giant component change when a certain number of edges are removed from the network. In the chosen scenario we remove the edges with the highest load in the network, i.e. the nodes that has the highest value of edge betweenness. Table VIII shows the relative increase of the characteristic path length between different types of nodes when the communication links (edges) between the gas nodes (label 1) are removed from the network, considering the initial network. Table IX shows the relative increase of the characteristic path length between different types of nodes when the communication links (edges) between the electricity substations (label 2) are removed from the network, considering the initial network.

		(a)	(b)
Label m	Label n	Relative increase of $L^{V_m,V_n}(G)$	Relative increase of $L^{V_m,V_n}(G)$
5	6	17.1482	18.8795
3	5	12.3805	13.8508
1	6	12.3179	17.6531
2	5	10.5588	11.5519
1	1	13.1183	23.4674
3	4	8.3258	9.4886
2	2	7.7087	8.4251
3	3	8.6617	10.4416
4	4	8.4820	9.1166

TABLE VIII: Relative change of the characteristic path length of different groups of nodes when certain number of communication links between the gas nodes are removed from the network. Column (a): 200 communication links were removed from the initial network, the size of the giant component is 11570 nodes. Column (b): 400 communication links were removed from the initial network, the size of the giant component is 11489 nodes.

Finally, we measure the size of the largest cluster (as a fraction of the total system size), when a certain percentage of the nodes and edges are removed from the networks. The results are shown in Fig. 9 and Fig. 10.

[1] X. Dimitropoulos, D. Krioukov, A. Vahdat, G. Riley, *Graph Annotations in Modeling Complex Network Topologies*, ACM Transactions on Modeling and Computer Simulation, vol. 19, no. 4, pp. 17, Oct 2009

[2] R. Carvalho, L. Buzna, F. Bono, E. Gutierrez, W. Just, D. Arrowsmith, *Robustness of trans-European gas networks*, Physical Review E 80, 016106 (2009)

[3] X. Dimitropoulos, D. Krioukov, G. Riley, KC Claffy, *Revealing the Autonomous System Taxonomy: The Machine Learning Approach*, Passive and Active Measurements Workshop (PAM), Mar. 2006

		(a)	(b)
Label m	Label n	Relative increase of $L^{V_m,V_n}(G)$	Relative increase of $L^{V_m,V_n}(G)$
5	6	8.4471	16.1736
3	5	9.4004	18.3469
1	6	6.9323	7.6844
2	5	11.5814	21.6968
1	1	3.7647	4.8784
3	4	13.4359	20.0220
2	2	15.2086	23.4121
3	3	10.4193	14.4123
4	4	15.1178	24.2397

TABLE IX: Relative change of the characteristic path length of different groups of nodes when certain number of communication links between the electricity substations are removed from the network. Column (a): 200 communication links were removed from the initial network, the size of the giant component is 12661 nodes. Column (b): 400 communication links were removed from the initial network, the size of the giant component is 12558 nodes.

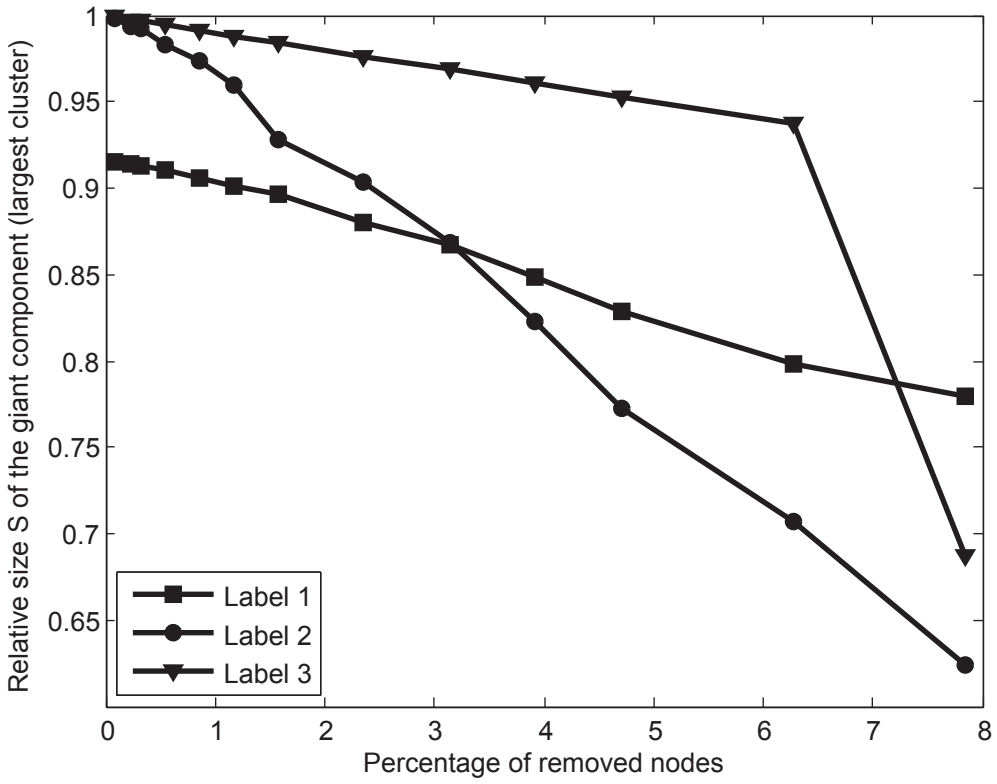


FIG. 9: The relative size S of the giant component (largest cluster) as a function of the percentage of removed nodes for different type of nodes (i.e. labels)

[4] T. Zesch, I. Gurevych, *Analysis of the Wikipedia Category Graph for NLP Applications*,
[5] J. L. Ortega, I. Aguillo, *Mapping World-class universities on the Web*, Information Processing & Management, 2009. Vol 2 (March), 45: 272-279
[6] R. Albert and A.-L. Barabasi, *Statistical mechanics of complex networks*, Rev. Mod. Phys. 74, 47 (2002)
[7] R. Albert, A. L. H. Jeong, and A.-L. Barabasi, Nature (London) 406, 378(2000)
[8] P. Holme, B. J. Kim, C. N. Yoon, and S. K. Han, *Attack vulnerability of complex networks*, Phys. Rev. E 65, 056109 (2002)
[9] P. Crucitti, V. Latora, M. Marchiori, A. Rapisarda, *Efficiency of scale-free networks: error and attack tolerance*, Physica A, vol. 320 pp. 622 642 (2003)
[10] R. Albert, I. Albert, and G. L. Nakarado, *Structural vulnerability of the North American power grid*, Phys. Rev. E 69,

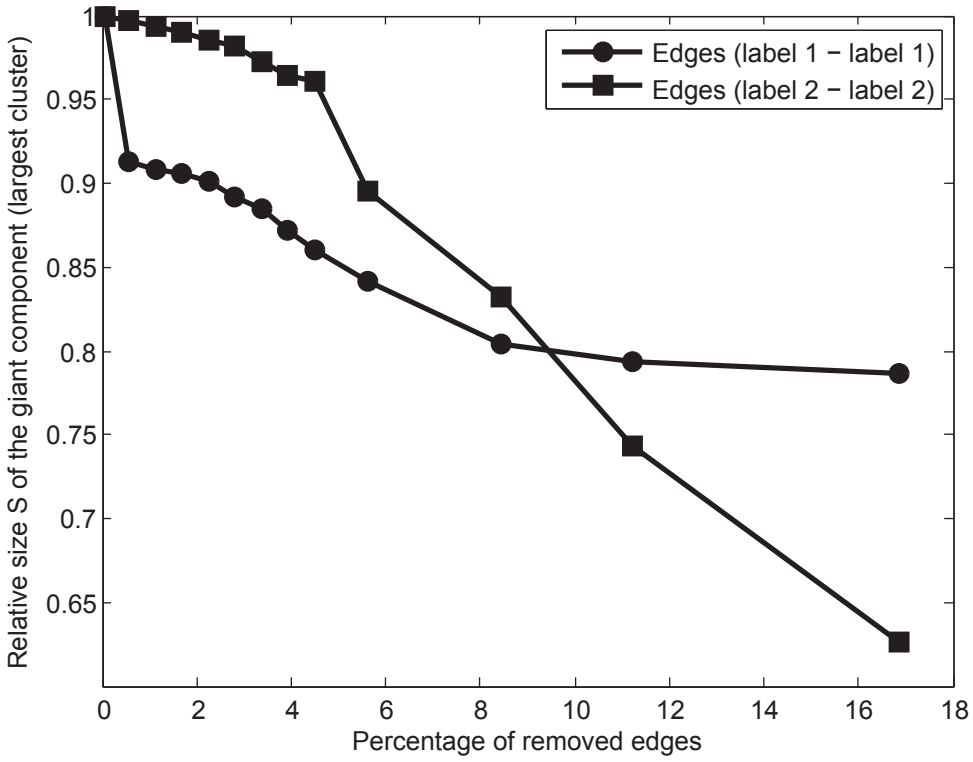


FIG. 10: The relative size S of the giant component (largest cluster) as a function of the percentage of removed edge which are connecting nodes with same labels

025103(R) (2004)
[11] P. Crucitti, V. Latora, and M. Marchiori, *Model for cascading failures in complex networks* Phys. Rev. E 69, 045104(R) (2004)
[12] L. Huang, L. Yang, and K. Yang, Phys. Rev. E 73, *Geographical effects on cascading breakdowns of scale-free networks* 036102 (2006)
[13] A. E. Motter and Y.-C. Lai, Phys. Rev. E 66, *Cascade-based attacks on complex networks* 065102(R) (2002).
[14] A. E. Motter, Phys. Rev. Lett. 93, *Cascade Control and Defense in Complex Networks* 098701 (2004)
[15] S. Boccaletti, J. Buld, R. Criado, J. Flores, V. Latora, J. Pello, M. Romance, *Multiscale vulnerability of complex networks*, Chaos 17, 043110 (2007)
[16] http://www.ece.gatech.edu/research/labs/MANIACS/as_taxonomy
[17] D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small-world' networks", Nature 393, pp. 440–442 (1998)
[18] A.-L. Barabasi and R. Albert, "Emergence of scaling in random networks", Science, Vol. 286, No. 5439, pp. 509–512 (1999)

B Modelling the dynamics and fragmentation of electric power systems

Márk Szenes, Zénó Farkas, Gábor Papp
Collegium Budapest, H-1014 Budapest, Szentháromság u. 2.
Eötvös Loránd University, H-1117 Budapest, Pázmány P. stny. 1/A

December 2009

We have analysed the regular and breakdown dynamics of power grids. In the MANMADE project, we have spent considerable time on formulating a model for this purpose. Such a model should be simple, yet capable of capturing the most fundamental properties of power flow. For the better part of the project, we used the DC load flow model [1]. However, this model turned out to be incapable of taking into account the maximum transmission capacity of power lines correctly. At that point switched to a linear programming model, which proved to be an excellent tool for our investigations. The details of the model is described below in detail.

1 Optimal power flow

We analyzed the load flow problem and studied the cascading breakdown phenomena with linear programming method. The objective function represents the generation and transmission costs that has to be minimized. The set of equality constraints represents the inflow–outflow balances, and the set of inequality constraints represents the generating capacities of power plants and the loadability limits of transmission lines. We can formulate the above

mentioned linear programming problem:

$$\begin{aligned}
\text{min: } (\{P_i \mid i \in V^g\}, \{F_{ij}\}) &\longmapsto \sum_{i \in V^g} K_i^g \cdot P_i \cdot \Delta t + \sum_{\langle ij \rangle} K_{ij}^t \cdot F_{ij} \cdot \Delta t \quad (1) \\
P_i &= \sum_{j \in \{\langle ij \rangle\}} F_{ij} \quad (2) \\
P_i &\leq C_i^g \qquad (i \in V^g) \quad (3) \\
|F_{ij}| &\leq C_{ij}^t \quad (4)
\end{aligned}$$

1. P_i denotes the actual power produced/consumed by the power plant/consumer i, measured in MW.
2. F_{ij} is the power flows from i to j, measured in MW.
3. K_i^g is the generation cost of the power plant i (which is element of the set of the power plants V^g), measured in EUR/MWh.
4. K_{ij}^t is the transmission cost. For the sake of simplicity we choose it to depend only the transmission line length l_{ij}, so $K_{ij}^t = K^t \cdot l_{ij}$ (the unit of K^t is EUR/MWh·km).
5. C_i^g is the nominal capacity of power plant i, measured in MW.
6. C_{ij}^t is the line loadability, measured in MW.

The advantage of the application of linear programming technique described above is that it takes into account economic considerations under given physical constraints. Thus this method fairly reproduced the operation of the power transmission system operator that is responsible for the most effective and reliable distribution of power.

2 Parameterizing the model

Our electricity network database [4] contains the information about:

- the network topology – the set of edges $\{\langle ij \rangle\}$ and nodes $\{i\}$ (with country information)
- the length (l_{ij}) and the voltage level (U_{ij}) of transmission lines
- the fuel type (nuclear, coal, natural gas, fuel oil, lignite, wind, biomass, hydro, etc.) and the nominal capacity of power plants (C_i^g)

- the population belongs to the nearest substation

Hourly consumption data of European countries available from the page of the European Network of Transmission System Operators [2]. Combined with the population information we assigned the consumption values (P_i) to the consumer nodes in the ratio of the corresponding populations.

The determination of the generation and transmission costs (K^g and K^t) is based only on expert estimation which doesn't take into account political and geographical and such specific factors that can affect the real costs. In our model the generation cost depends only on the fuel type of the power plant.

3 Line loadability

For the purpose of determining the line loadability, we apply the method originally proposed by St. Clair [6], and which later on analytically derived by Dunlop et al. [5]. Although the method has some limitations, and is based on several assumptions (like the neglect of resistance, the terminal system impedance and the effect of series or shunt compensation etc.), it is a good approximation for quickly estimating the line loading limit. The papers cited above showed that the loadability characteristics for uncompensated high voltage transmission lines is universal, the maximal power in units of surge impedance loading (SIL) is independent of voltage levels, and depends only the line length. Three factors influence the maximal power that can be transmitted, these are:

1. the thermal limitation;
2. the line-voltage-drop limitation;
3. the steady-state-stability limitation.

The thermal limitation is relevant only for lines shorter than 80 km, and within this range the maximal power is approximately $3 \cdot \mathrm{SIL}$. The maximum allowable voltage drop along the line is 5%, and relevant in the 80–320 km region. The steady-state-stability limitation is important for lines longer than 320 km. The steady-state-stability margin is defined as $100\% \cdot (P_{\max} - P_{\mathrm{limit}})/P_{\max}$ and it is assumed to be 35% (corresponds to $\delta = 40.5^\circ$ power angle). Fig. 1 shows the loadability curve, which we used to determine the power transmission capability of the transmission lines in our load flow simulation. This is the so called St. Clair curve which gives the load carrying capability in the units of SIL. We applied the following typical SIL values (Tab. 6.1 in [3]):

B4

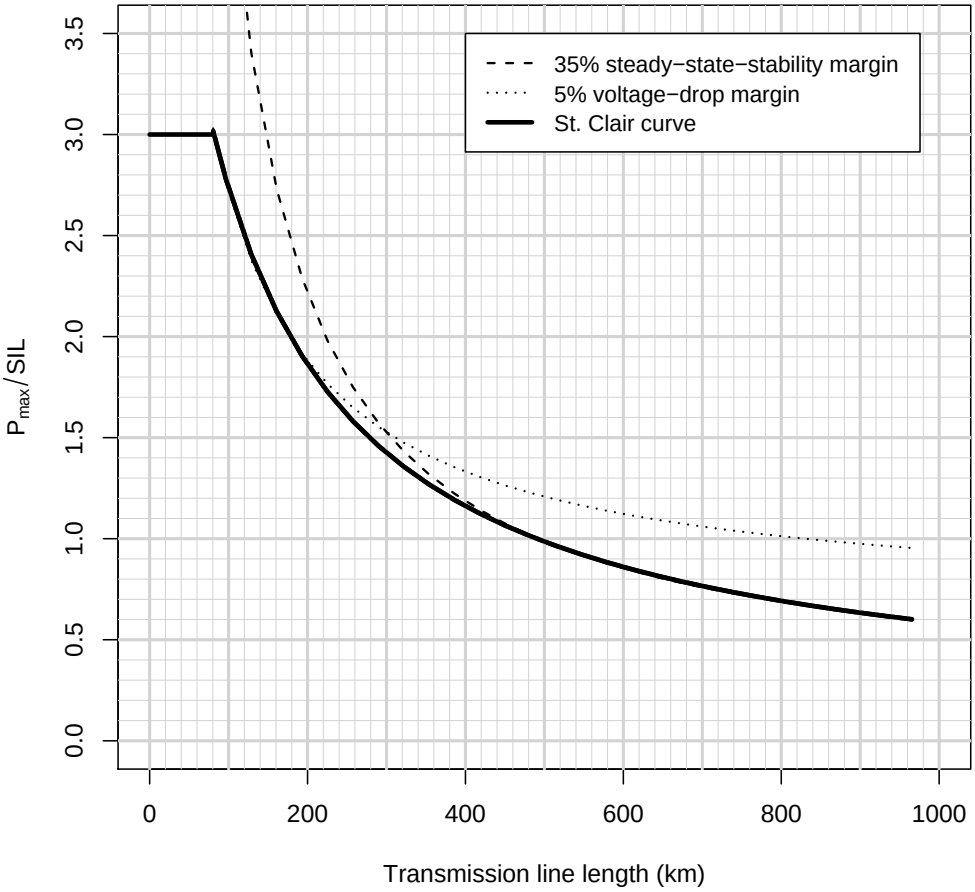


Figure 1: Loadability curve for uncompensated overhead transmission lines

Voltage levels	230 kV	345 kV	500 kV	765 kV	1100 kV
SIL	140 MW	420 MW	1000 MW	2280 MW	5260 MW

For other voltages we interpolated the SIL values correspond to the nearest voltage levels.

4 Breakdown process

A failure in the power network can cause successive failures which can propagate to the whole system. For example, when a transmission line (power

plant) fails, the other lines (power plants) have to supply the missing carrying (generating) capacity that may lead to overload of some network components. This cascading failure mechanism causes the large blackouts, like the disturbance of the European power system on 4 November 2006, when the European interconnected power system was split into three independent parts [7].

In our linear programming approach (with linear equality and inequality constraints) there is no information about which line become overloaded, because if a constraint can't be satisfied the linear programming problem became infeasible. To find the weak lines in the system and to perform the cascading breakdown process we followed the next steps:

1. Decreasing the total consumption P to the limit P_{lim}, where the solution of the problem still doesn't exist, but for appropriately small further decrease of P_{lim} the problem becomes feasible
2. Finding the lines which the problem should become feasible with at P_{lim} if they have infinitely large capacity
3. Removing the lines found in the previous step. Updating the total consumption P if some node(s) dropped out.
4. Repeating these steps unless the electricity network becomes operable again

References

[1] B. F. Wollenberg A. J. Wood. *Power Generation, Operation and Control*. John Wiley & Sons, New York, 1996.

[2] ENTSO-E. `www.entsoe.eu`.

[3] P. Kundur. *Power System Stability and Control*. McGraw-Hill, New York, NY, USA, 1994.

[4] Platts. `www.platts.com`.

[5] P. P. Marchenko R. D. Dunlop, R. Gutman. Analytical development of loadability characteristics for ehv and uhv transmission lines. *IEEE Transactions on Power Apparatus and Systems*, PAS-98:606–617, 1979.

[6] H. P. St.Clair. Practical concepts in capability and performance of transmission lines. *Power Apparatus and Systems, Part III. Transactions of the American Institute of Electrical Engineers*, 72:1152–1157, 1953.

[7] UCTE. Final report - system disturbance on 4 november 2006. http://www.entsoe.eu/fileadmin/user_upload/_library/publications/ce/otherreports/Final-Report-20070130.pdf, 2007.