Paper accepted for presentation at 2009 IEEE Bucharest Power Tech Conference, June 28th - July 2nd, Bucharest, Romania

Complex Network Theory and Graph Partitioning: Application to large interconnected networks

B. Rozel, *Student Member, IEEE,* R. Caire, *Member, IEEE,*
N. Hadjsaid, *Senior Member, IEEE,* J-P. Rognon, C. Tranchita

***Abstract*—The research of weakness ties in graphs such as interconnected electrical transmission grids is a major concern in large infrastructures' studies. Some novel methodologies which are developed in advanced topics of complex systems studies and graph theory may be applied to the electrical grid. Indeed, in order to identify the potential cuts of large interconnected networks such as the interconnected European UCTE network first synchronous area, spectral partitioning and other graph weakness assessment have been studied. The use of the third eigenvalue is indeed unconventional, but it leads to useful results, allowing a new analysis of the last event which split the UCTE into three autonomous areas. Potentials and limits of this study are also presented.**

***Index Terms*—Complex Networks, Electrical Infrastructure, Spectral Partitioning, Interconnected power system.**

I. Introduction

The graph partitioning methods allows researchers to find the weakness arcs on a graph. Consequently, if the studied graph represents a real network, it becomes possible to evaluate the weak lines linking strong subgraphs. More specifically, when this infrastructure is an electrical grid, the knowledge of those tie lines can be useful in forecasting where a load cascading may happen in a case of a large unbalance between different network areas. In addition, it allows the resulting islanded areas to be determined. This study permits, thanks to the knowledge of the actual network, either some new lines or some countermeasures that may help the independent sub-networks to successfully resist to the disturbance to be planed. It can also be used for achieving some degraded operational modes..

The knowledge of a minimal cut for splitting the network is also useful information for avoiding the propagation of disturbances such as the cascading phenomena. In this case, this study allows the network to be preventively split into different areas and their autonomous operation to be planned. This situation is very similar to the previous one, but the splitting is planned.

When the graph is a social network, these methods are used to determine communities (or well connected groups). Then, it is possible to characterize every individual and whether they are plainly in a social group or at the border of different sub-groups. Some methods are also often used in order to do some load balancing for parallel computing or to design some telecommunication networks [1].

This paper presents several useful methods of graph partitioning applied to large networks such as a power grid. These splits should help, on the one hand, the operators to find operating measures to successfully operate asynchronous networks. On the other hand, they should help the planners to build essential additional lines to reinforce smartly the grid. The first section describes three different mathematical methods for splitting graphs. The second section shows the test example (benchmark) on which the developed methods were applied. This benchmark is the UCTE first synchronous area. The third section presents the obtained results. The fourth part shows the results analysis and discussions. A final section concludes on the advantages of the methods for our study and some perspectives.

II. Graph Partitioning

The problem of graph partitioning is considered as classical in computer science. It consists in finding a balanced partition of a graph so that the number of vertices in each part is nearly the same and the number of links between these different sub-parts (cut-edges) is minimized. This problem is known to be NP-complete [1]. Different methods were developed to solve this problem and three of them were studied and will be presented in this paper.

A. Spectral Partitioning

The spectral method leads generally to better results than heuristic methods that look only for solutions close to an initial partitioning and thus usually stick to a local minimum [1]. This method, in comparison with others methods, is fast and efficient to cut the graph into two parts. However, it is not adapted for splitting graphs in any number (different from two). Eigenvalues of different matrices can be used: adjacency matrix A, Laplacian matrix L or normal matrix N. The use of

Manuscript received October 31st, 2008; revised March 1st, 2009.

B. Rozel, R.Caire, N. Hadjsaid, J-P. Rognon and C. Tranchita are with *Grenoble Institute of Technology – G2Elab* (*Grenoble Electrical Engineering laboratory*), France (phone: +33-476826361; e-mail: raphael.caire@g2elab.grenoble-inp.fr).

those two last matrices is considered more efficient than the first one and for this study only the method with the Laplacian matrix was used.

This method, sometimes called min-cut theorem, was proposed at the early 1970s and popularized in the 1990s [1].

Consider a graph G composed of n nodes. Its adjacency matrix A is defined as:

$$\text{if nodes i and j are connected } A_{ij} = 1 \quad \text{else } A_{ij} = 0 \qquad \text{Eq. 1}$$

Its diagonal matrix is composed of elements D_{ii}:

$$D_{ii} = k_i \qquad \text{Eq. 2}$$

with k_i the degree of the node i. The Laplacian matrix [3] of the graph L is defined as:

$$L = D - A \qquad \text{Eq. 3}$$

This matrix L has a size of nxn and therefore n eigenvalues w_i with n eigenvectors v_i associated. Theses eigenvalues have the property to be all real and positive:

$$\forall i \in 0...n-1, w_i \geq 0 \qquad \text{Eq. 4}$$

The smallest eigenvalue is zero ($w_0 = 0$) and have as eigenvector $v_0 = (1, 1, \ldots, 1)$.

Under the hypothesis that the graph is completely connected, this zero eigenvalue is unique; otherwise, the multiplicity of this zero eigenvalue is equal to the number of connected subgraphs. The eigenvalues can be sorted in ascending order. The eigenvector associated with the first strictly positive eigenvalue, i.e. the smallest non-zero eigenvalue is v_1. The partition is made according to the sign of the component of the eigenvalue associated with each node: one subgraph with all nodes corresponding to a positive component and the other subgraph with all nodes corresponding to a negative component. If the size of the network is very large, it can be convenient to compute only the second eigenvector with some special methods, for instance Lanczos algorithm because the used matrices are sparse. To summarize, the partitioning algorithm is the following:

1) Compute the Laplacian matrix L of the graph G

2) Compute pairs (w; v), respectively eigenvalues and eigenvectors of L

3) Sort pairs (w_i; v_i) with ascending order for w_i

4) Select v_1, the second element of v

5) Determinate iv_{1p}, components index of $v_1 > 0$ and iv_{1n}, component index of $v_1 < 0$

6) Build G_p, the subgraph of G composed with nodes iv_{1p} and G_n, the subgraph of G composed with nodes iv_{1n}

B. Girvan and Newman Algorithm

This algorithm was first described in reference [4] and improved in [5]. It is based on the hierarchical clustering. There are two classes of hierarchical clustering: agglomerative methods and divisive methods. The first ones consist in applying a merge sequence from the N nodes to lead to the complete graph. On the contrary, the divisive methods consist in splitting the complete graph to N different parts. The result can be represented with a hierarchical tree or dendogram that shows the subgraphs at each step of the agglomeration or division according to the method. Agglomerative methods are traditionally more used because of its fast computation. However, a problem of this method is that generally isolated nodes are merged only at the end of the process. Divisive methods are not subjected to this phenomenon.

Girvan and Newman algorithm is based on a divisive method. It is an iterative method which consists in removing gradually edges with the highest betweenness value until the graph is split in different subgraphs. The edge betweenness represents the part of smallest paths (geodesic paths) between all the nodes going through this specific edge [1] and [5]..

The partitioning algorithm is the following:

1) Compute the betweenness coefficient for all edges of the graph

2) Remove the edges with the highest coefficient

3) Repeat 1) and 2) until all edges are deleted

The betweenness coefficient calculation for all edges at each step is computational intensive but it appears that this step is very important to perform a high-quality partitioning. Because of the need to compute all geodesic paths between the n nodes at each loop l, this algorithm leads to important computing time. Indeed, in the worst case, computing the betweenness coefficient needs $O(ln)$ with l the number of links and n the numbers of nodes, so the total algorithm needs $O(l^2n)$ or $O(n^3)$ on sparse graphs. With currents CPU, this complexity leads to restrict to graphs below 10 000 nodes.

In order to choose the best division of the whole process, a measure of the partition's quality so-called modularity is used. This measure is computed from a symmetrical matrix E of size k, with k the number of current subgraphs. The element e_{ij} of this matrix is the fraction of all the edges in the graph which link a node in the subgraph i to a node in the subgraph j. For this calculation, all the edges of the initial graph are used, even those removed during the partitioning.

e_{ii} is the fraction of internal edges of the subgraph i.

$$\sum_i e_{ii} = Tr(E) \qquad \text{Eq. 5}$$

Equation 5 is the fraction of internal edges of subgraphs where $Tr(E)$ is the trace of matrix E.

$$a_j = \sum_j e_{ij} \qquad \text{Eq. 6}$$

Equation 6 is the fraction of edges connected to the

subgraph *i*. The modularity is then defined with equation 7:

$$Q = \sum_i \left(e_{ii} - a_i^2 \right) = Tr(E) - \left\| E^2 \right\| \qquad \text{Eq. 7}$$

with $\|E^2\|$, the sum of the element of the matrix E^2.

The maximal value of Q is 1 and high values of modularity indicate that more accurate is the partition into subgraphs, nevertheless this bound of 1 is never reached [5]. Others measures of the partition's quality exist, but this one is the most used.

C. Improved Spectral Algorithm

A possible extension of the spectral bi-partitioning presented in section *A.* was introduced in [6]. The principle is to use not only the first non-zero eigenvalue of the Laplacian matrix, but the *D* first ones and their associated eigenvectors. Then, each node of the graph is represented by a point in a D-dimensional space.

The coordinates of this point correspond to the components of this node for each eigenvalue. The association of the points is made with a measure of distance. In reference [6] is shown that the angular distance, with the angle between the two vectors, is better than the Euclidian distance. When distances between each point are computed, they are grouped together.

Various possibilities are available in order to define distance between two points groups. It can be defined as the minimal value of the distance set between two points belonging to each group (*single linkage clustering*), the maximal value (*complete linkage clustering*) or the mean value (*group average clustering*).

Reference [6] claims that none of these methods has been proved better than the others. Nevertheless, the first one has the drawback to lead to the clustering of distant points, but linked together through an intermediary's chain. It is called the *chaining property*.

During the whole process, the modularity, as defined for the previous method, is used to measure the partition's quality and therefore to choose the best value for the dimension *D* and the optimal cut number.

III. Description of the Benchmark

The benchmark is a test system corresponding to a part of the European transmission network. It is, more especially, the first synchronous area of the UCTE (Union for the Co-ordination of Transmission of Electricity).

It includes 18 countries: Portugal, Spain, France, Belgium, Luxembourg, Germany, Netherlands, Switzerland, Italy, Denmark (continental part), Czech Republic, Austria, Slovenia, Poland, Slovakia, Hungary, Croatia and a part of the Bosnia and Herzegovina. It was established by Zhou Qiong and Janusz Bialek in [2].

The test system is composed of 1254 nodes, 1944 lines and 378 generators. This model is approximate due to the unavailability of the exact data. Indeed, electrical transmission companies do not publish complete information for commercial or security reasons. For establishing this model, only public available data where used. The hypotheses are:

- only lines with a voltage above 220 kV are considered;
- series resistance, shunt admittance and series capacitors, when they exists, are ignored. Only series reactance is considered;
- series reactance is computed from lines length with an impedance of 0.31 Ω/km for the 220 kV lines and 0.28 Ω/km for the 380 kV lines (at 50 Hz);
- all breakers are considered in their closed states.

The resulting data are available at the reference [9].

IV. Results

The three algorithms presented in section II were applied on the UCTE first synchronous area graph. For this task, the software NetworkX [7], developed in the Los Alamos National Laboratory, was used and also Matlab for the verification.

A. Spectral Partitioning

When the network is split into only two parts, the proposed min-cut follows roughly the east border of France with his neighbor countries. It begins with the French-Belgian border, continues towards the south to the French-Swiss and French-Italian borders (see Fig. 1).

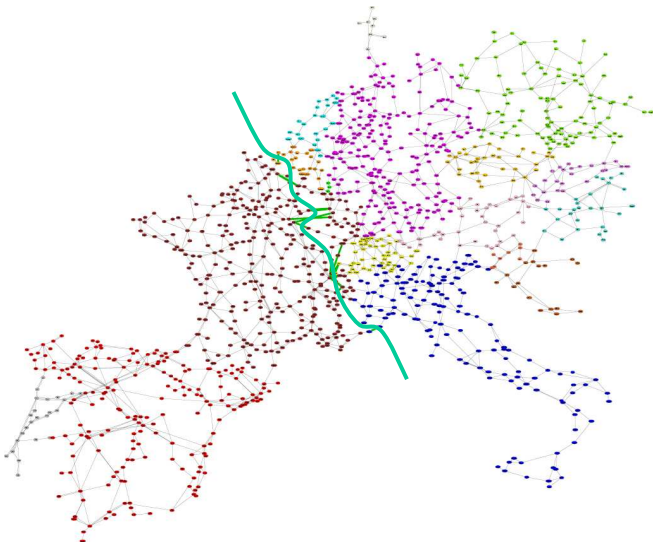

Fig 1. First cut of the UCTE network using v_1 eingenvector

This weakness of the UCTE network is well known and corresponds almost to the interconnection lines between France and its East neighbors.

A try with the third eigenvalue, i.e. the second non zero, was also made. Hence, the graph is cut in tree parts. The first cut is along the Pyrenees, weakness well-known of the UCTE network for its numerous congestions. The second cut begins in North at the border between the Netherlands and the Germany, goes south east to cross Austria, goes through Hungary and Slovenia and finally Croatia. The min-cuts are shown in bold green on Fig. 2.

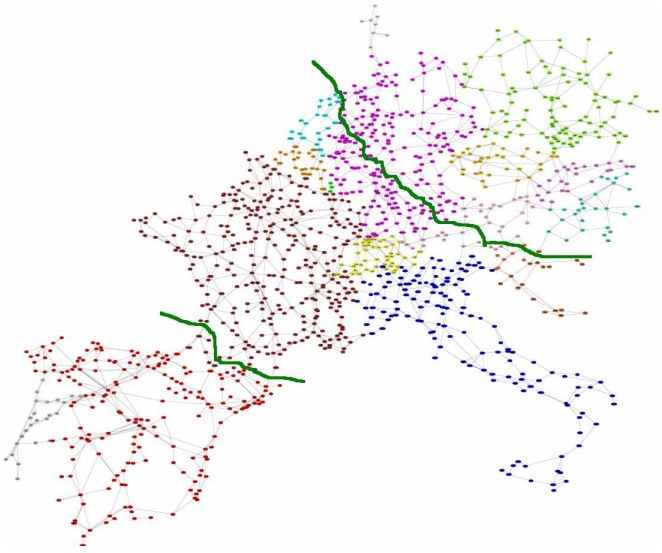

Fig 2. Second cut of the UCTE network using v_2 eigenvector

It appears, on this specific case and even if the UCTE network is larger than our benchmark, that this partition corresponds to the real cut that happens during the evening of the 4 November 2006 [8] as shown on Fig. 3. The second cut does not appear between France and Spain, maybe due to the fast and deep response of the Spanish operator. The third area which occurs during the event is not modeled in this study. The use of this eigenvalue is absolutely unconventional, but it leads to interesting results.

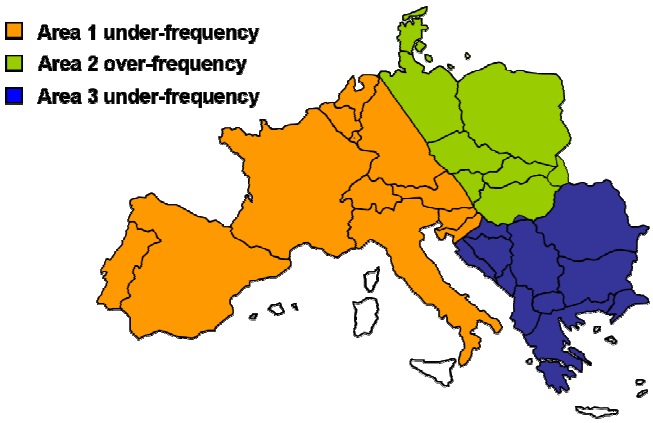

Fig 3. Splitting of the UCTE network during the 4th November 2006 event (www.ucte.org credits)

B. *Girvan and Newman Algorithm*

The algorithm was applied on the UCTE network until obtaining 182 subgraphs, as shown in figure 4.

Indeed, the modularity is globally decreasing after about twenty cuts, it is useless to do the calculations until the complete disintegration of the graph in 1254 one-sized parts. Nevertheless, it is useful not to stop at the first maximum, because it can happen that this one is only local, followed by a second peak which is the global maximum.

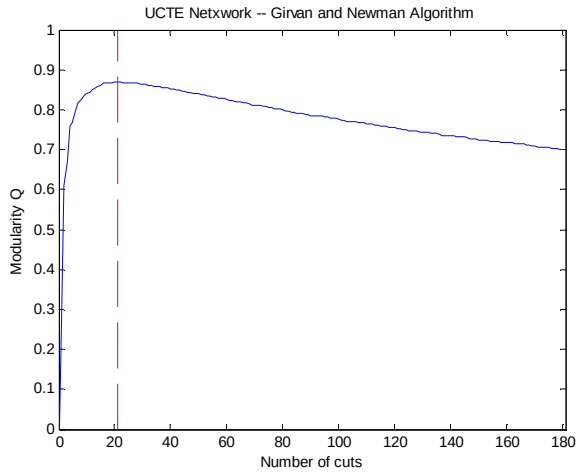

Fig 4. Variation of the modularity

The modularity variation is presented in Fig. 4. Modularity's maximum is reached for 21 cuts and its value is $Q_{max} = 0.8692$ represented with a vertical line on the figure.

The resulting graph partitioning is depicted on Fig. 5.

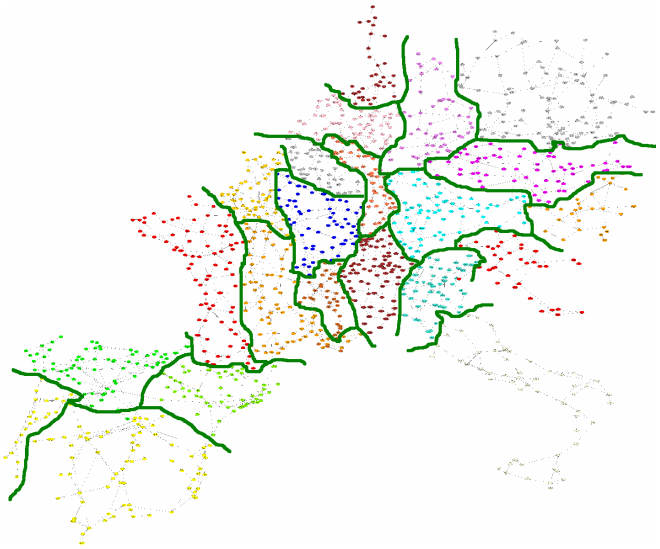

Fig 5. UCTE network partitioning with the Girvan and Newman algorithm

C. *Improved Spectral Algorithm*

The Euclidian distance was tested until 20 eigenvectors and the angular one with only two eigenvectors. For each case, all the clustering methods previously presented were studied.

In order to better visualize the representation of graph's nodes projected in the eigenvectors space, the plot of the points corresponding to the eigenvectors components has been made. The Fig. 6 shows the plane with the second eigenvector in abscissa and the third in ordinate.

When only two eigenvectors are considered, the Euclidian distance provides better results than with the angular distance.

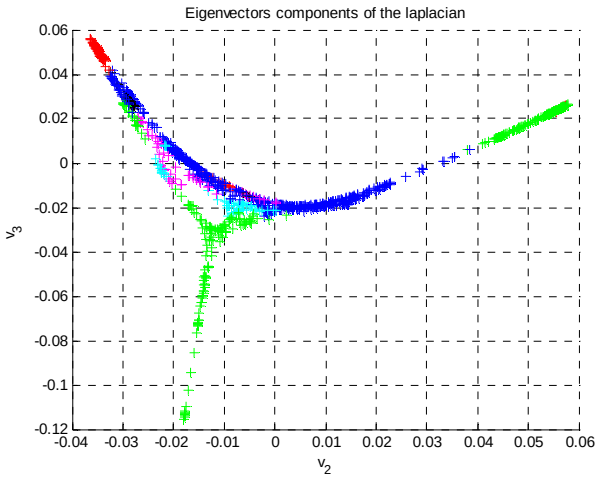

Fig 6. Projection on the v2/v3 plane

The table I presents the results for all the studied cases in Euclidian distance computation. The bold values represent the maximum value for the modularity index.

TABLE I
MODULARITY INDEX AND TOTAL AREA NUMBER DEPENDING ON THE EIGENVECTOR NUMBER AND THE CLUSTERING METHOD

D dimension	single linkage clustering		complete linkage clustering		group average clustering	
	modularity index	zone amount	modularity index	zone amount	modularity index	zone amount
2	0,70124	135	0,8275	29	0,83064	28
3	0,69747	116	0,84102	27	0,84909	27
4	0,71774	127	0,83879	31	0,84538	28
5	**0,72935**	**84**	0,84409	29	0,84796	41
6	0,70761	126	0,83841	21	0,84548	44
7	0,68136	126	0,85075	29	0,85702	39
8	0,69223	120	**0,85205**	**29**	0,84955	44
9	0,56685	105	0,84821	36	0,84926	31
10	0,46805	76	0,85087	24	0,85737	23
11	0,45163	105	0,85038	37	0,85805	31
12	0,62103	126	0,84621	22	0,86244	25
13	0,61041	119	0,84509	27	0,86159	26
14	0,48208	121	0,84182	28	**0,86563**	**25**
15	0,45831	96	0,85034	35	0,86192	30
16	0,44793	112	0,84871	36	0,85766	23
20	0,6057	126	0,85788	33	0,86165	31

Single linkage clustering results are not so good compared to the two others methods. The best result with the complete linkage clustering is obtained with 8 eigenvectors, the resulting graph is composed of 29 subgraphs and the modularity value is $Q_{max} = 0.85205$. Nevertheless, the best result for the improved spectral algorithm was obtained with the group average clustering and 14 eigenvectors. There are then 25 subgraphs and $Q_{max} = 0.86563$. The corresponding partitioning is illustrated in Fig. 7. In all cases, this method's results are not so good compared to those of the Girvan and Newman algorithm with $Q_{max} = 0.8692$.

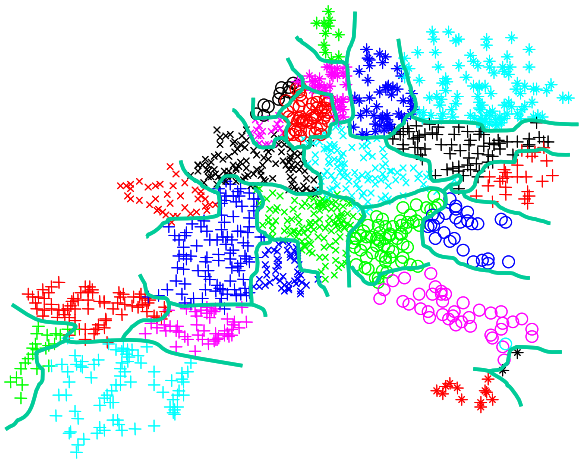

Fig 7. Improved spectral partitioning with 14 eigenvectors and group average clustering for the UCTE network

V. ANALYSIS AND DISCUSSION

There are many limitations to the above study. The first one comes from the scale and the border of the network. Indeed, only the first synchronous area was studied while the real infrastructure is, in fact, larger. As the problem consists in splitting the graph into parts of equivalent size, the choice of the limits of the network provokes the result.

In a regular use of the first method, only the first non zero eigenvalue is computed. For partitioning in more than two parts, the method is usually applied again on the subgraph in a recursive way. Thus, finally there are 2^n subgraphs, with n the number of times the method is applied. Indeed, regular structures generally split into only two parts at each step. For instance a sheet of paper submitted to n forces will firstly tear in two distinct parts. However, if the number of parts expected to be obtained finally is not known in advance, this method does not allow the number of needed iterative steps to be determined. Moreover, although the partitioning is made in an optimal way at each step, there is no guarantee that the final partitioning is globally optimal.

The two other methods, while following the modularity index, the amount of independent communities is very large. There are very few data to validate such results. If the previous disturbance (4th November 2006) is only considered, the targeted number of sub-areas pointed out the first method as a very promising one. These were only premises of such studies but contingency analysis researchers should be able to evaluate more deeply such interesting research path. The betweenness evaluation gives result which can be closer to a load flow evaluation in specific test case. More information may be found in [5] about such comparison.

VI. Conclusion

The study of graph partitioning applied to a large interconnected system allows obtaining some useful information in order to forecast splitting phenomena of the network or also to establish some countermeasures that could avoid larger disturbances or cascading propagation.

Three different methods that originally come from computer science and sociology were presented and evaluated in the context of electrical power systems. These methods show that using state of the art algorithms applied to large systems such as power grids may obtain unexpected and promising results.

The main limitation of this kind of methods is that they are based on purely topological criterion and that they are purely static, in the sense that dynamic phenomena are not considered. Nevertheless, complex network theory is an active research area and other methods can overcome this limitation.

VII. References

[1] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, D.-U. Hwanga, Complex networks: Structure and dynamics, Physics reports, vol.424, no.4–5, pp.175–308, 2006.

[2] Z. Qiong, J.W. Bialek, Approximate model of European interconnected system as a benchmark system to study effects of cross-border trades, IEEE Transactions on Power Systems, vol.20, no.2, pp.782–788, May 2005.

[3] V. Rosato, S. Bologna, F. Tiriticco, Topological properties of highvoltage electrical transmission networks, Electric Power Systems Research, vol.77, pp.99–105, 2007.

[4] M. Girvan, M. E. J. Newman, Community structure in social and biological networks, Proceedings of the National Academy of Sciences, 2002.

[5] M. E. J. Newman, M. Girvan, Finding and evaluating community structure in networks, Physical Review E, 2004.

[6] L. Donetti, M. A. Muñoz, Detecting Network Communities: a new systematic and efficient algorithm, J. Stat. Mech, 2004.

[7] A. Hagberg, D. Schult, P. Swart, NetworkX: Python Software for the Analysis of Networks, http://networkx.lanl.gov/, Mathematical Modeling and Analysis, Los Alamos National Laboratory, 2005.

[8] UCTE, Final Report – System Disturbance on 4 November 2006, Union for the Co-ordination of Transmission of Electricity, January 2007.

[9] http://www.see.ed.ac.uk/~jbialek/Europe_load_flow/

VIII. Biographies

Benoit Rozel (S'06) received his Master in electrical engineering in 2006 from the Grenoble Institute of Technology, France. He is currently preparing a Ph.D. at the Grenoble Electrical Engineering laboratory (G2Elab) about the security of critical infrastructures and more especially about modeling the interdependencies between them.

Raphael Caire (M'04) received his Diplôme d'Etudes Approfondies (DEA) and Doctorat de l'INPG degrees from the Institut National Polytechnique de Grenoble (INPG) in 2000 and 2004. He had been working in Power Electronic field, in USA at the Center of Power Electronic System (CPES) in 2000 and within several EDF research centers in Germany and in France from 2004 to 2006. He is now associate professor at Grenoble Institute of Technology (Grenoble-InP) at the Ecole d'ingénieurs en Energie Eau et Environnement (ENSE3) in the Grenoble Electrical Engineering laboratory (G2Elab). His research is centered on the impacts, production control of dispersed generation on distribution system, distribution network architectures and critical infrastructures.

Nouredine Hadjsaid (SM'05) received his Diplôme d'Etudes Approfondies (DEA) and Doctorat de l'INPG degrees from the Institut National Polytechnique de Grenoble (INPG) in 1988 and 1992. From 1988 to 1993, he served as a research and teaching assistant at the Ecole Nationale Supérieure d'Ingénieurs Electriciens de Grenoble (ENSIEG) and at the Laboratory d'Electrotechnique de Grenoble (LEG). He is now a full time professor professor at Grenoble InP at the Ecole Nationale Supérieure de l'Energie, Eau et Environnement (ENSE3) of Grenoble Institute of Technology and with the Grenoble Electrical Engineering laboratory (G2Elab). His is also the Director of a common research center between EDF, Grenoble Institute of Technology and Schneider Electric (IDEA). His research interests are power system operation and security.

Jean-Pierre Rognon is a Professor at the "Institut National Polytechnique de Grenoble (INPG)" and in the Ecole Centrale de Lyon. He is the former head of the Laboratoire d'Electrotechnique de Grenoble (LEG). His researches deal with electrical system control and fault detection and isolation.

Carolina Tranchita received her MSc. degree in electrical engineering from Universidad de los Andes in Colombia. She received a Doctorat degree in 2008 from the Institut National Polytechnique de Grenoble in France and the Universidad de los Andes. She is now a temporary lecturer at Grenoble InP at the Ecole d'Ingénieurs pour l'Energie, l'Eau et l'Environnement de Grenoble (ENSE3) in the Grenoble Electrical Engineering laboratory (G2Elab). Her research interests are modeling and securing critical infrastructures and expert systems applications in power systems.