HAL
open science

Large-scale power grid hierarchical segmentation based on power-flow affinities

Antoine Marot, Sami Tazi, Benjamin Donnot, Patrick Panciatici

To cite this version:

Antoine Marot, Sami Tazi, Benjamin Donnot, Patrick Panciatici. Large-scale power grid hierarchical segmentation based on power-flow affinities. 2017. hal-01633508v2

HAL Id: hal-01633508

https://hal.science/hal-01633508v2

Preprint submitted on 28 Nov 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Large-scale power grid hierarchical segmentation based on power-flow affinities

A. Marot, S. Tazi, B. Donnot, P. Panciatici (RTE R&D)[1]

***Abstract*— The segmentation of large scale power grids into zones allows a better understanding of its structure, as the control room operators will naturally but manually do for any study. In this paper we provide a new automatic hierarchical method based on the community detection algorithm *Infomap*. Our main contribution is to offer as input a new representation of the power grid, called the security analysis, that represents power flow affinities beyond the connectivity of the grid, a point that will become even more relevant for tomorrow's cyber-physical system. Indeed we already discover few relevant and important clusters that are not connected in the actual grid topology. To better describe and investigate the method, we apply it here on the well-studied IEEE-RTS-96 and IEEE-118. We further applied our method on the large-scale French Power Grid which showed promising results given its puzzling resemblance with the historical RTE regional segmentation.**

I. INTRODUCTION

Well-established power systems are starting to see a coming phase transition with a steep rise in complexity. This is due in part to the changing nature of the grid, with an end to the ever increasing overall consumption. This shifts the way we traditionally develop the grid, from expanding it with heavy investment relying on growth in revenues, to rather optimizing the existing one with every flexibilities at our disposal. We can also notice the revival of DC current, hybridizing the current AC grid with new dynamics. But this new complexity is also due to other external factors such as the changing energy mix with the massive integration of renewables, as well as the ever more fragmented set of actors at a more granular level like prosumers or at the supranational level with an interconnected European grid for instance. This new complexity will bring new dynamics with more varying flow amplitude but also changing flow directions, as opposed to what was the case in the past with centralized production from large power plants, "pushing" the flows to the loads in a very hierarchical and descendant way. New distributed controls will also be implemented, going toward an always more entangled cyber-physical system, not only physical as it is mainly today. Therefore, rethinking the way we operate the grid has become a necessity.

To handle the current complexity, our control room operators have built over time their own mental representations of the grid, cutting it into static zones to study them efficiently, that is being able to quickly identify remedial actions given security risks around. However we anticipate that these static views will be less and less relevant in the future to operate the grid, with moving electrical "frontiers" given this dynamic context, even along the course of a day. But we believe that a zonal segmentation will still be relevant to operate the grid by efficiently synthesizing the information to offer proper context awareness, helping in the decision making process. That is why an automatic segmentation built in a dynamic fashion to fit the specific context of a situation is needed. Hence, how can we build such contextual segmentation?

For segmentation, the overall goal is to cluster "similar" power lines or buses while avoiding clustering "dissimilar" ones. To achieve this, one needs to define a measure of similarity, that will represent the mutual influence between all the power lines or buses of the grid in our case, and then run a clustering algorithm on it. Previous works have relied on the one hand on gathering proper dynamical phasor measurements on the grid to compute disturbance-based coherency in the time-domain and find similarities between electrical nodes [1], [2], [3]. This supposes a massive deployment of PMUs or very accurate large-scale dynamic simulations. On the other hand, other approaches have investigated the electrical distance between buses and use different clustering methods [4] (hierarchical clustering), [5] (spectral clustering) [6] (hybrid K-means/evolutionary algorithm). This is much less costly but not grid state specific.

This paper however proposes a new method based on a weakly invasive perturbation to build an original influence graph which connectivity goes beyond the actual grid topology, an idea expressed by [7] and reminiscent of [8] when studying cascading failures. This approach will be even more relevant compared to topological ones as the system becomes more cyber-physical. To define similarity between power lines, we run with a simulator smooth topological sensibilities on a well-meshed grid to compute an influence matrix called the security analysis. We further run a suitable clustering algorithm, *Infomap*, on this influence graph to find electrically coherent zones. Our main objective is to demonstrate the relevance of this similarity for grid segmentation.

The paper is organized as followed. Section II is dedicated to the method, where we describe the security analysis matrix, justify its relevance compare to more classical distance matrices and talk about the suitability of the "*Infomap*" clustering algorithm. In section III we present the results on the 96-RTS and 118 IEEE grids on which we can compare it to other methods mentioned above. We eventually give some insights on how relevant this method is on large-scale power grids such has the French power grid with about 6 000 nodes. Finally, section IV provides conclusions on this work.

[1]Réseau de Transport d'Électricit (French TSO)

II. METHOD

A. Small simulated perturbations as an inspective tool

A power grid is a complex system with intricate power flows that have multiple long distance interactions. Only representing a grid state as a graph of electrical nodes, tied by power lines, doesn't highlight these complex interactions of flows, given productions, consumptions and topology. Simulators hopefully allow us to assess freely and quickly sensibilities on a grid state from which we can build an influence graph. But all sensibilities are not representative of either physical phenomena or expert knowledge for grid operations and hence don't properly sense its interactions.

For instance, a classical perturbation consist in playing a transaction of power between 2 electrical nodes as used by [6] for assessing their clustering quality. This shadows the influence of injections and their localizations, almost only capturing topological patterns. Indeed, since the topology is not modified, the superposition theorem allow us to linearly decompose the resulting grid state (currents and voltages) as the original grid state plus the perturbation, hence the sensibility results are nearly the same for power-flows whatever the injections and the power flows were in the original grid state. So it is not as state dependent as desired. Moreover the meaning of such a perturbation is not obvious. It assumes in essence that some power originally flows directly from one node to another, which cannot be clearly measured. Not to forget that electrical nodes are of different types: production, consumption, transmission. Some are active whereas other are passive which have consequences as explained by [9] who are actually working towards a better formalization under the use of those power injection perturbations.This is finally a component by component approach, not really taking into account that we are actually studying an interacting multi-component system. This kind of perturbations is closely related to the equivalent resistance computation between any two nodes, and as [6] mentioned it is limited to be a static segmentation not state dependent.

Instead we believe that the tripping of a line is a more meaningful perturbation to consider, smooth in a well-meshed grid, and grid state dependent: we are redistributing on the grid an existing and identified flow on the tripped line, which originally depends on the injection plan and topology. These multiple independent small perturbations over the grid can be viewed as many distributed sensors and local estimators of the interactions in place.

B. Security analysis influence graph: a new appropriate representation of the grid

In control rooms, dispatchers continually run security analysis on forecasted grid states to study contingencies. Recall that a contingency is a power line potentially turning out of service, resulting in power flow re-dispatch on other power lines. That way, they can anticipate possible overloads, somehow building mental representations for patterns of power-flow re-dispatch in the meantime, and study them to find remedial actions.

Hence, to have a representation of the power flow affinities of the grid power lines, we can play those perturbations on each line independently, one after the other. We can then construct the security analysis matrix where each row i is the result of active AC power-flow transfer on every line on the grid when tripping the line i ($i \in \mathcal{L}$, with $\mathcal{L}$ the lines ensemble), the element j ($j \in \mathcal{L}$) hence being the power-flow transfer on line j:

$$S_{ij} = \text{PowerFlow}(j)_{i \text{ tripped}} - \text{PowerFlow}(j)_{\text{originally}} \quad (1)$$

It's a square matrix of dimension $nL \times nL$ where nL is the total number of lines of our system : $nL = \text{card}(\mathcal{L})$. It can be represented as a *weighted* and *oriented* graph structure, in which the security analysis plays the role of the adjacency matrix. Each node in that graph is a power line in the grid and the edge weights are determined by the flow transfer after a power line disconnection. We will name it as the Influence Graph. It should be emphasized that its graph structure is different from the classical topological graph structure of the power grid which will further makes sense later. For our security analysis matrix, most lines transfer sensible active power flow to a limited number of other lines, by sensible we mean more than 1MW on the French Grid for instance. This still results in a relatively sparse matrix from which one wants to gather lines with high mutual dependency. Since we want to cluster lines with high mutual influence, we only care about the absolute value. Thus the matrix elements of our similarity matrix are:

$$a_{i,j} = |S_{ij}| \quad (2)$$

C. Infomap: a generic hierarchical clustering algorithm

There are several algorithms for graph segmentation, known in literature as community detection algorithms [10], [11], [12]. One can refer to the following article for a review on community detection algorithms [13]. The algorithm developed by Rosvall *et al.* [14] known as "*Infomap*", has the advantage of being particularly suitable for oriented, weighted graphs, and able to identify flow patterns inside the graph. It is recursively hierarchical [15] and can automatically find the proper number of levels and clusters. By level on a power grid, we mean that in the case of the European grid we should discover aggregates of National Grid on top, then regions like the ones for RTE French power grid and then more focused areas of production, transmission or consumption. In addition, "*Infomap*" can handle overlapping [16] which could be of interest for future works. Indeed electrical frontiers are fuzzy and it could make sense that some lines interconnect some clusters. We will here briefly describe the main ideas of this method, the reader can refer to the original article for a complete description.

The idea is to use the duality between the problem of minimizing the description length of places visited along a path given the influence graph, and the problem of how we should best partition the network with respect to flow.

Aiming at minimizing the description length L of a likely path, we simulate a random walk on the graph where

each node are identified by a codeword and weighted-edges that represent the random walker likely direction. To better minimize the average length of codewords, one can take advantage of the graph regional cycling structure that highlights modules M, and define a "module codebook" for each area that contains all the nodes codewords of this area. Thereby it is possible to use the same codeword for different nodes since we can specify which module codebook to use. We then need an "index codebook" containing a codeword for each "module codebook". Going from one node to another in the same region, one only needs to refer to a short codewords to identify it, knowing the region codebook. An easy analogy is the case of maps with streets, cities and countries. For instance, in different cities you will find the same street names, and you can do so because you can name the city as well, to better identify this street in a country. But being in a given city, you don't even need to name the city again to refer to a street in it: you compressed the information while still being able to communicate it.

Thus to calculate the shortest description length one can apply the Shannon's source coding theorem from information theory. It establishes the limits to possible data compression, for n codewords describing n states of a random variable X that occur with frequencies p_i, the average length of a codeword can be no less than: $H(X) = -\sum_1^n p_i log(p_i)$. We can then apply it for the average length of codewords from the index codebook and the module codebooks, weighting them by their rates of use. This leads to the map equation:

$$L(M) = q_{\curvearrowright} H(Q) + \sum_{i=1}^{m} p^i_{\circlearrowright} H(P^i) \tag{3}$$

with $H(Q)$ the weighted average length of codewords in the codebook index and $H(P_i)$, the weighted average length of codewords in the module codebook i. The codeword index is used at frequency $q_{\curvearrowright}$ the probability to change module at every step of the random walk. The i module codebook is used at frequency $p^i_{\circlearrowright}$, which is the number of moves continually spent in module i plus the probability to leave.

D. Beyond the grid graph

Very little works for grid segmentation have tried to use representations of the grid that go beyond its connectivity, a hard constraint which seems at first natural and intuitive. Nevertheless, this might overlook that a power grid is a very entangled system with complex interactions, sometimes counter-intuitive as [7] explained. Here we use the security analysis whose graph has a different connectivity than the grid as shown in figure 1 for RTS-96 system [17]: it is actually more connected. But in the segmentation process, some links will appear stronger and relevant while other will be weak and ignored. As a consequence, the results will most generally lead to connected elements in clusters from the grid topological graph. But some might not be connected, which could highlight complex interactions between flows and potential areas where the grid needs to be reinforced. This is one of the striking facts we retrieve when applying our method.

We argue that the security analysis Influence Graph should *a priori* be a better and smoother representation of the grid for our purpose. Indeed, it actually represents that every lines are interacting with one another with distributed strengths as one would expect for a system, in contrast of the binary topological grid adjacency representation. In particular, we can see a quite colored row and filled rectangle for the 7th and last detected cluster in our security analysis Influence Graph matrix on figure 1, which is totally ignored by the other representations. Let's now display and describe our findings.

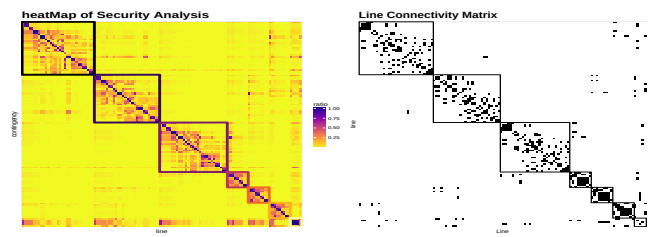

Fig. 1. HeatMaps to illustrate graph connectivity given different representations for RTS-96 system, ordered by the clusters found by our method: a) security analysis Influence Graph Matrix b) topological adjacency Matrix.

III. Results

We applied our method on 3 different kinds of systems to test the method genericity: on the 96-RTS toy system which has been studied in the past as an interesting baseline for segmentation purposes, on the 118-bus system which is a realistic and readable middle-sized one, and finally over the large-scale French power grid to demonstrate how it scales.

A. IEEE-RTS-96

Taking the reliability test system 1996 we obtained the clustering showed in figure 2. It highlights one level and 7 clusters, 6 agreeing with the power grid connectivity and 1 not. We argue that this surprising non-connected cluster comes from the system and is not an artifact of our method. As for the 6 others, since they represent the same IEEE-24 case[18], it is consistent to segment them in the same way.

About the non-connected one, it gathers high voltage interconnecting power lines close to productions that link the three same sub-grids corresponding, but leave aside the low voltage one close to loads, which can be understood. We rediscovered that that these lines were artificial additions for the purpose of this toy grid, and not the result of a coherent grid development. Cutting one of those power lines leads to significant changes in flows over the whole grid, as illustrated in the security analysis heatmap in figure 1. Hence these interconnecting lines play the same roles in the power grid, even if they are non-connected, and it then makes sense to cluster them together. This is a first example of one possible other application of our method: identifying weakly meshed interconnecting areas that are strongly interacting over long distance.

It could not have been possible to identify this role with other classical representations. To prove so, we run a naive

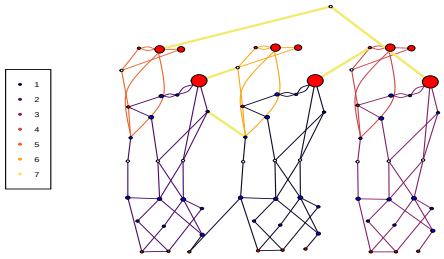

Fig. 2. IEEE-RTS-96 segmentation. Node sized by injection volume, production in red and load in blue. Thick edges for non-connected cluster.

comparison with the adjacency matrices, the pure power line connectivity one (clustering lines) and the electrical conductance-weighted one (clustering nodes). Both resulted in a similar 3-clustering. But being only topological, it misses a better clustering by subgrids which are balanced systems given the injections.

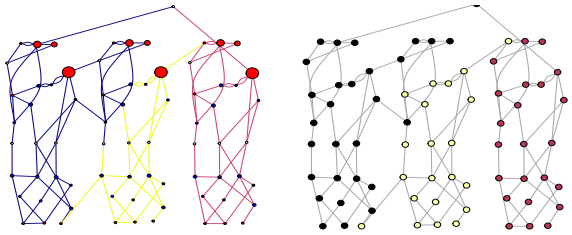

Fig. 3. Naive IEEE-RTS-96 segmentation with: a) connectivity matrix b) conductance matrix.

To test the robustness of our security analysis representation and show that the information can be captured beyond the grid connectivity, we set to 0 the adjacent coefficients that are represented in black in the connectivity matrix figure 1. Those $a_{i,j}$ coefficients are usually non-zero and the ones with the highest flow transfers, being neighbors of the line tripped. Doing so, we reduce the total available information by 40%, if we consider the amount of information as the sum of matrix coefficients. Running the *Infomap* algorithm on this even sparser matrix with missing information, it is still able to recover the previous clustering. This further validates that our method is a good system representation: the whole system is actually shaking under a perturbation, not only the supposed neighboring lines which we can actually overlook.

Finally, we compare the results of our method to other works in figure 4, [6] who uses electrical distance and [1] who uses time-domain measurements. Overall, the clusterings are very similar while being computed by 3 different methods which might indicate we are close to an interesting clustering. There are however slight differences we can comment. In addition to this non-connected cluster we highlighted, we can notice 2 differences as circled on the figure 4, besides that we are actually clustering lines and they are clustering nodes. We argue that our method properly circumscribes the upper cluster to a meshed clique whereas the other clustering has slight less obvious unmeshed extensions.

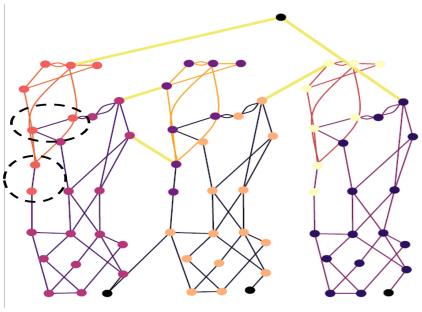

Fig. 4. Comparison of our IEEE-RTS-96 segmentation (line colored) with two other methods (node colored)[6][1]. Recurrent dissemblances highlighted on 1 subgrid, plus the thick edges non-connected cluster

B. Middle-sized grid: IEEE-118

One remaining important consideration in a power grid are unmeshed areas with one-way in or out for power flow. In these cases, the removal of one electrical node or power line, the kind of perturbations played in the security analysis, can split the grid into two systems which need to be balanced on their own: this is islanding. Taking it into account leads to non-connected cluster areas shown on the 2-clustering figure 5 for the black lines and black area. This is actually an undesired artifact from our N-1 security analysis that the simulator naively compensate for. Indeed, power grids are rarely islanded in operations except for exceptional situations. To overcome that undesired effect, we ignore the related perturbations, removing them from the security analysis as a pre-processing. For post-processing, we stick them back to the closest cluster afterwards, using in that case the grid connectivity. We do so for both the IEEE-118 and the French Grid.

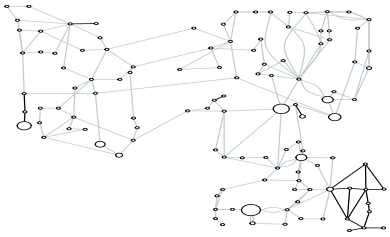

Fig. 5. IEEE-118 segmentation under islanding artifact: in black th unmeshed area cluster.

The IEEE-118 bus test case is a reduced model of the Midwestern US power grid in 1962 [19]. In figure 6, one can see the IEEE-118 case segmentation without islanding effect. We can distinguish at the top level 9 clusters, 8 agreeing with the grid connectivity and 1 which does not. This non-connected cluster 7 in the grid topology plays the same role as its counterpart in the IEEE-96-RTS case: important weakly-meshed interconnections between two well-meshed East and West grids, with some unbalance here, the East grid having more production and the West too much consumption. The remaining connected clusters seem reasonable with a proper clique segmentation and localized injections. We

could actually expect some overlapping for some lines in-between clusters so that they each have proper cliques. This is something we actually observe when running *Infomap* with that option and will be further studied in future works, as well as the level 2 of the hierarchical clustering in which there might be few clusters that could make sense.

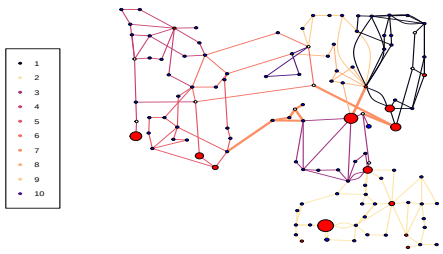

Fig. 6. IEEE-118 segmentation. The non-connected cluster 7 with East-West interconnections is highlighted with thick edges.

C. Large grids: the French power grid

Finally, we discuss our segmentation over the French power grid composed of 6 000 nodes, 10 000 lines and 10 000 injections. On level 1 of the hierarchical clustering, the 8-clustering results, 8 being an output Infomap discovered and not a parameter, makes already a lot of sense qualitatively given its resemblance with RTE historical regional segmentation on figure 7: identical segmentation for the West and few differences with Northern and Eastern parts which were explained qualitatively by our operators. The historical RTE partition is not a pure electrical one, human resources, workload, maintenance teams and their localization were taken into account by that time. Nonetheless, this resemblance is a good matter for discussion on grid management. Further validation on the quality of lower level clusters such as level 2, that have sizes close to the ones usually drawn by our dispatchers, could lead to redesigning some of our study tools and offering some long-awaited context awareness in a more cyber-physical system.

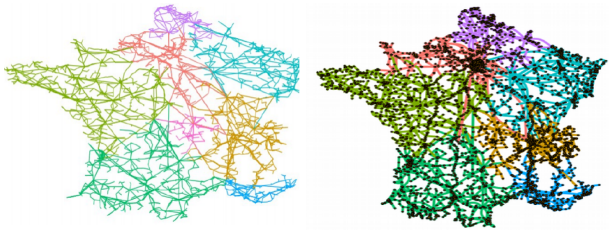

Fig. 7. Comparison of a) our French power grid segmentation with b) historical RTE regional segmentation.

IV. CONCLUSIONS

In this study we derived a new method to efficiently cluster power grids. The method relies on a suitable algorithm for hierarchical graph clustering, "Infomap", and a new representation of power-flow affinities, the "security analysis" influence graph based on a small-perturbation approach. In particular, this representation allows us to highlight relevant non-connected clusters, is robust to missing information since it is distributed, and seems overall a good representation of the interconnected power systems. It could be extended to take into account more cyber-physical behaviors, even further away from the grid connectivity, given that our simulator has an internal model for those automatic regulations and that we identify meaningful small perturbations to run. Future works will be of many kinds: defining more quantitative measures beside our interesting analytical results, analyzing lower level clusters, running overlapping, extracting a backbone clustering over time by clustering aggregation, and eventually studying the importance of grid state context.

REFERENCES

[1] I. Kamwa, A. K. Pradhan, and G. Joos. Automatic segmentation of large power systems into fuzzy coherent areas for dynamic vulnerability assessment. *IEEE Transactions on Power Systems*, 22(4):1974–1985, Nov 2007.

[2] Mang-Hui Wang and Hong-Chan Chang. Novel clustering method for coherency identification using an artificial neural network. *IEEE Transactions on Power Systems*, 9(4):2056–2062, Nov 1994.

[3] C. Juarez, A. R. Messina, R. Castellanos, and G. Espinosa-Perez. Characterization of multimachine system behavior using a hierarchical trajectory cluster analysis. *IEEE Transactions on Power Systems*, 26(3):972–981, Aug 2011.

[4] S. Blumsack, P. Hines, M. Patel, C. Barrows, and E. C. Sanchez. Defining power network zones from measures of electrical distance. In *2009 IEEE Power Energy Society General Meeting*, pages 1–8, July 2009.

[5] R. J. Snchez-Garca, M. Fennelly, S. Norris, N. Wright, G. Niblo, J. Brodzki, and J. W. Bialek. Hierarchical spectral clustering of power grids. *IEEE Transactions on Power Systems*, 29(5):2229–2237, Sept 2014.

[6] E. Cotilla-Sanchez, P. D. H. Hines, C. Barrows, S. Blumsack, and M. Patel. Multi-attribute partitioning of power networks based on electrical distance. *IEEE Transactions on Power Systems*, 28(4):4979–4987, Nov 2013.

[7] P.Rezaei P.D.H.Hines, I.Dobson. Cascading power outages propagate locally in an influence graph that is not the actual grid topology. *IEEE Transactions on Power Systems*, 2015.

[8] B.Lesieutre C.Asavathiratham, S.Roy and G.Verghese. The influence model. *IEEE Control Systems*, 2001.

[9] E. Pons E. Bompard and D. Wu. Extended topological metrics for the analysis of power grid vulnerability. *IEEE Syst. J.*, 2012.

[10] M. E. J. Newman and Girvan M. Finding and evaluating community structure in networks. *Phys. Rev. E 69, 026113*, 2004.

[11] Guimera R., M. Sales-Pardo, and L. A. N. Amaral. Modularity from fluctuations in random graphs and complex networks. *Phys. Rev. E 70(2), 025101*, 2004.

[12] V. D. Blondel, J-L Guillaume, R. Lambiotte, and E. Lefebvre. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment (10), P10008 (12pp)*, 2008.

[13] Santo Fortunato. Community detection in graphs. *Physics Reports*, 486(3):75 – 174, 2010.

[14] Rosvall M.and Axelsson D.and Bergstrom C. T. The map equation. *The European Physical Journal Special Topics*, 178(1):13–23, Nov 2009.

[15] Martin Rosvall and Carl T. Bergstrom. Multilevel compression of random walks on networks reveals hierarchical organization in large integrated systems. *PLOS ONE*, 6(4):1–10, 04 2011.

[16] Alcides Viamontes Esquivel and Martin Rosvall. Compression of flow can reveal overlapping-module organization in networks. *Phys. Rev. X*, 1:021025, Dec 2011.

[17] C. Grigg et. al. The ieee reliability test system-1996. *IEEE Transactions on Power Systems*, 14(3):1010–1020, Aug 1999.

[18] ICSEG. Illinois center for a smarter electric grid.

[19] PSTCA. Power systems test case archive, university of washington, electrical engineering, 2007.