The interplay of network structure and dispatch solutions in power grid cascading failures

Jose M. Reynolds-Barredo,[1] David E. Newman,[2] Benjamin A. Carreras,[1,2] and Ian Dobson[3]
[1]*Departamento de Física, Universidad Carlos III de Madrid, 28911 Leganes, Madrid, Spain*
[2]*Physics Department, University of Alaska, Fairbanks, Alaska 99775, USA*
[3]*Electrical and Computer Engineering Department, Iowa State University, Ames, Iowa 50011, USA*

(Received 19 July 2016; accepted 31 October 2016; published online 14 November 2016)

For a given minimum cost of the electricity dispatch, multiple equivalent dispatch solutions may exist. We explore the sensitivity of networks to these dispatch solutions and their impact on the vulnerability of the network to cascading failure blackouts. It is shown that, depending on the heterogeneity of the network structure, the blackout statistics can be sensitive to the dispatch solution chosen, with the clustering coefficient of the network being a key ingredient. We also investigate mechanisms or configurations that decrease discrepancies that can occur between the different dispatch solutions. *Published by AIP Publishing*. [http://dx.doi.org/10.1063/1.4967736]

The OPA (ORNL-PSerc-Alaska) model has been used to explore the power networks' robustness as characterized by the risk of large failures and temporal dynamics. In this model, the power demand is increased at a constant rate while the generation is periodically increased in response to the demand. In doing so, the system keeps the generation capacity margin above a given value. Each day power is dispatched, and generation is selected and optimized to exactly balance the load with minimum cost. This optimization problem is degenerate (there can be more than one optimal solution), so many different dispatch procedures are possible. As one might expect, even if the dispatch solutions are exactly equivalent from the optimization point of view (they reach the same optimal value), it is found that the long term statistics of the problem (power law tails, risk of failures, etc.) can be quite different. In this work, we try to understand how the structure of the network affects the sensitivity of the long term statistics to the dispatch solution chosen. We have found that linked networks are more sensitive to the dispatch chosen than homogeneous networks. In other words, in the case of linked networks, the simulation results can strongly differ when the dispatch is modified. Of particular note, the average clustering coefficient of the network is found to be one of the key measures that affect such sensitivity: the higher it is, the stronger the sensitivity of the long term statistics to the dispatch chosen. A possible explanation of such behavior is that a small clustering coefficient reduces the degeneracy of the problem. In addition, we study different ways of reducing the sensitivity to the dispatch in the linked network. From a practical standpoint, the results can be used to properly apply the dispatch solution that is superior to the others from the computational point of view.

I. INTRODUCTION

Power transmission networks as well as many other critical infrastructure networks come in a wide variety of shapes and sizes but share the characteristics that they are responding to ever-changing demands and are being pushed ever closer to their operational limits. This push toward their operational limit, combined with the operational and engineering responses when there are failures, gives the system characteristics of a critical point, heavy tails (failure size probability distribution functions that decay as a power law), and long time correlations. Understanding these characteristics is critical to both doing realistic risk analysis of cascading blackouts and to assessing the impact and risks of changes to the system and mitigation schemes.

We use the OPA (ORNL-PSerc-Alaska) model to explore the power networks' robustness as characterized by the risk of large failures and temporal dynamics. The OPA model[1–3] was developed to study the long-term patterns of blackout of a power transmission system under the complex system dynamics of an increasing power demand and the engineering responses to failure. In this model, the power demand is increased at a constant rate while also being modulated by random fluctuations. The generation capacity is automatically increased when the capacity margin is below a given critical level. From the numerical point of view, OPA solves at least one optimization problem for the calculation of the power generation-demand and transmission (the dispatch solution) per "day". Each optimization problem is solved using the Simplex algorithm.[4–6]

Using the OPA model, we have been able to study and characterize the mechanisms behind the power law tails in the distribution of the blackout size. These algebraic tails obtained in the numerical calculations are consistent with those observed in blackouts of real power systems.[7–13] Most importantly, the OPA model permits us to separate the underlying causes for cascading blackouts from the triggers that initiate them and therefore explore system characteristics that enhance or degrade resilience and reliability of the power transmission grid. One of these characteristics, the one investigated here, is the network structure and the heterogeneity of the network induced by linking homogeneous structures. With these different network structures, the issue of multiple different "optimal" dispatch solutions in systems of varying degrees of homogeneity can then be investigated.

This is both relevant and important because many of the real, large transmission grids are the result of linking smaller networks, and understanding risk vulnerability from both the network structure and dispatch operations is critical. A major example is the Western region of North America (the Western Electricity Coordinating Council or WECC) in which the population dense areas have denser grids that are interconnected by a smaller number of lines.[14] The resultant inhomogeneous grid looks like a loop of pearls (Fig. 1). The aim of this paper is to show that this type of grid characteristic may complicate the modeling of the dispatch of electricity in order to properly analyze the propagation of the cascading failures through the system. These issues will be examined with the OPA model.

In more detail, to test the sensitivity of different networks to the electricity dispatch, we use three variants of the Simplex algorithm in order to have different dispatch solutions for the same minimum of the cost. The difference between variants is in the pivot rule chosen. Under these different dispatch conditions, we will compare the dynamics of three sequences of networks. Each sequence is composed by a set of networks, each of them with a different number of nodes (from 100 to 1600). The sequences differ in the topological properties of the networks.

In a recent related work, the letter by Liu and Li[17] also describes the multiplicity of different dispatch solutions using OPA, some with similar generation cost, and shows that the dispatch solutions that maximize or minimize the number of overloaded lines during the cascade have different total costs that are the sum of the generation and outage costs. They suggest that the worst case dispatch should be considered in assessing cascading risk and that operators should be advised of the dispatch plans that best minimize cascading. In our work, the emphasis is on the relation between the structure of the network and the discrepancies between the different dispatch solutions. Also we study how such discrepancies can be minimized with a slight modification of the model or the dispatch. If the discrepancies between dispatches are small, there is an extra benefit: the computationally most efficient dispatch can be used, reducing the wallclock time required for the simulations.

The rest of the paper is organized as follows. Sec. II briefly describes the OPA model and Sec. III discusses the type of networks used in this paper. Section IV discusses the sensitivity to the dispatch solutions and looks for the origin of the underlying differences. Sec. V shows modifications in the dispatch in order to reduce the discrepancies between the different solutions. Finally, Sec. VI summarizes the work.

II. THE OPA MODEL

To study the long time complex systems dynamics of the power transmission system, we use the OPA model.[1–3] The OPA model calculates the long time behavior of a power transmission system under the forcing of an increasing power demand and the engineering/operational responses to failure in order to study the cascading failures in the system. In this model, the network is composed of a set of nodes $i = 1, \ldots, N$ that can be generators or loads (or a mix of both). The nodes are connected via a set of transmission lines $j = 1, \ldots, M$. The power demand P_i in a load node i increases at a constant rate λ plus daily random fluctuations with variance γ. There are two sorts of upgrades to meet the increase in demand. Transmission lines are upgraded as engineering responses to blackouts and maximum generator power is increased in response to the increasing demand. The transmission lines selected for upgrade are those overloaded transmission lines involved in a blackout. The transmission lines are upgraded by increasing their maximum flow F_j^{max} limits at rate μ. On generator nodes, the maximum generation power P_i^{max} increases automatically when the capacity margin, $\Delta P/P$, is below a given critical level. In the present studies, this is done by increasing the power limit on all generators so we keep the same generation profiles as in the existing situation.[18,19]

The OPA model for a given network represents transmission lines, loads, and generators with the usual DC load flow approximation. Starting from a solved base case, blackouts are initiated by random line outages. Whenever a line is outaged, the generation and load are redispatched using standard linear programming methods. Since the generation capacity of the grid is larger than the usual power demand, it is necessary to determine which generator should be used to balance the load in an optimal way that minimizes costs

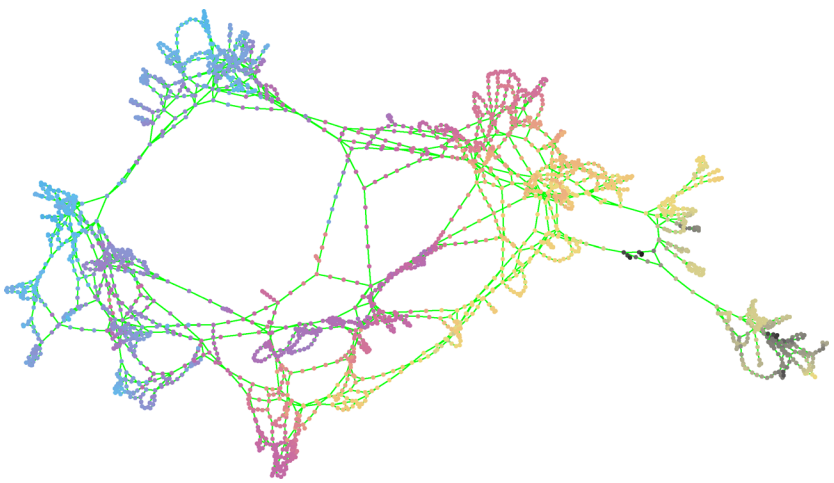

FIG. 1. A 2504 node model of the WECC. Color indicates the base case voltage phasor angle.

and satisfies the transmission constraints. If any lines were overloaded in the minimization solution, then these lines are outaged with probability p_1. The process of redispatch and testing for outages is iterated until there are no more outages. Then, the total load shed is the power lost in the blackout.

Because of its importance, the minimization process is now described in detail. The cost function to be minimized

$$Z(p_1, \ldots, p_N) = \sum_{generators} \alpha_i p_i + \sum_{loads} 100 p_i, \qquad (1)$$

where $\alpha_i = 1$ (except in Sec. V) and p_i is the power injected/extracted on node i (depending if node i is a generator or load). Note that $p_i > 0$ for generator nodes and $p_i < 0$ for load nodes. The objective is to find the set of values for p_i that minimizes Z subject to the following constrains:

- Overall power balance:

$$\sum_{i=1}^{N} p_i = 0$$

- Line flow limits:

$$-F_j^{max} \leq f_j \leq F_j^{max}; \quad j = 1, \ldots, M$$

- Load limits:

$$-P_i \leq p_i \leq 0; \quad i\ load$$

- Generator limits:

$$0 < p_i < P_i^{max}; \quad i\ generator$$

Note that, because the weighting of the loads (100) is larger than the weighting of the generators (α_i), the priority in the minimization process is to satisfy the loads demand whenever possible (this is, $-p_i = P_i$ for $i = 1, N$).

To solve the proposed minimization problem, we use the Simplex method.[4–6] There are two stages in the Simplex algorithm. In the first stage, a "feasible" solution is found that solves the problem but is not necessarily the optimal solution. The second stage is a search procedure to find the optimal solution by following paths in the solution space. This search procedure requires formulating the problem in the matrix form and carrying out a standard iterative algorithm. In each iteration, a process called pivoting swaps one of the variables associated with a column in the matrix with one variable associated with a row in the matrix. We will not go into more detail of pivoting here but rather refer the reader to the specialized bibliography.[15,16] The interesting point is that the pivoting procedure can be understood as the way in which the solver looks for the optimal solution. This is in a sense the dispatch method so that different pivoting rules can be thought of as different dispatch rules. The better the solver is, the faster the optimal solution is found. There are many ways of doing the pivoting called "pivot rules". In this work, we have chosen three different pivoting rules for the three solvers that we use:

- Solver 1 uses one of the first pivot rules developed, the Dantzig rule.[4,15] Roughly speaking, in each iteration, the solver just looks for the variable whose maximization coefficient is larger.
- Solver 2 uses the Devex pivot rule,[16] an approximation of Steepest-edge rule (see solver 3).
- Solver 3 uses the Steepest-edge rule.[20] This rule chooses the variable for pivoting that, when modified, gives the largest gradient of the maximization function in the solution space. This rule is generally thought to work very well for many optimization problems.

From the computational point of view, Solver 1 is the slowest of the three; it is about 10 times slower than the Solver 3. Solver 2 is the fastest, being about a factor of four faster than the Solver 3. One could expect that applying the three solvers to the same network problem, the same optimal solution would be found. However, in our problem, depending on the network, sometimes the three solvers find different solutions.[17] Despite the different solutions, the cost function reached is the same for the three solvers. This indicates degeneracy in the problem, something that might be expected because many configurations of the generators can supply the power to the loads and thus have the same cost. This degeneracy is even more noticeable for the present calculations because we have the same cost for all generators in the networks.

We can interpret the three solvers as three different dispatch policies that have different effects depending on the network structures. Thus, it is interesting to study the sensitivity of our results to the "dispatch policy" used. It should be noted that this study does not aspire to systematically study the possible and potentially more intelligent ways of doing a correct power dispatch. However as an initial investigation of possible effects of different dispatch rules, interesting results are obtained, even with the simple selection of solvers just described. Furthermore, techniques to reduce the discrepancies between the solutions while preserving the complex dynamics will allow us to use the computationally most efficient solver.

Finally, the parameters that we use in the OPA model are shown in Table I. Of the six basic parameters that control the slow time evolution of the system in OPA,[1,2] four (the demand growth rate, the generation margin, the load variance, and the upgrade rate) have been estimated from the data available for the US power transmission grid[14,21] and are shown in Table I. The other two model parameters, which are very important in the determination of the dynamics, are the probability p_0 of failure of a component by a daily random event and the probability p_1 of a transmission

TABLE I. OPA parameters used.

Variable name	Symbol	Value
Daily rate of increase of the demand	λ	1.00005
Critical generation margin	$\Delta p/p$	0.2
Variance of loads	γ	1.15
Upgrade rate	μ	1.07

line overload becoming an outage. p_0 represents the chances of random accidental failures while p_1 is a measure of the reliability of system components and their interactions, which impacts the propagation of failures through the system. Ranges for p_0 and p_1 can be estimated from data though with less certainty. Therefore, several values of p_0 and p_1 will be considered in what follows.

III. NETWORKS USED IN THE CALCULATIONS

Using the parameters discussed, the main additional input to OPA is a model network. The OPA model has been validated against the real WECC network using different size network models and real grid models.[14] In this work, we use three sequences of artificial networks with realistic network characteristics, each of them composed of networks with 100, 200, 400, 800, and 1600 nodes.

- The first sequence is composed of homogeneous artificial networks generated using the method of Wang *et al.*[21] In a first stage of the method, the N nodes of the network are placed with a uniform random distribution in a square of normalized area $S = 1$. In a second stage, a set of links are found such that they follow a prescribed distribution for the average node degree and the average line length. The impedance of each line is obtained as a function of its length. Let us name this sequence of networks H. Thus, H_{100} will refer to the homogeneous network of $N = 100$ nodes, H_{200} to the homogeneous network of $N = 200$ nodes, and so on. Here, when we use the word homogeneous we refer to the topological and engineering properties of the network, not to the possible space distribution of the network. We use the term homogeneous to contrast with the linked networks that will be introduced later. By increasing/decreasing the area while keeping the average node degree and average line length constant, the sparsity of the nodes increases/decreases and mean distance (in hops) between nodes increases/decreases.
- The second sequence of networks is constructed by linking a number of the homogeneous networks, each of which contain a prescribed number of nodes. Each of the homogeneous networks, after the linkage, will be referred to as a *zone*. Here, we use balanced linked networks, which are networks with similar levels of power generation and demand in each of the homogeneous zones. The resistivity of the linking lines has been chosen following the distribution of resistivity in the homogeneous networks. The case H_{100} is used as base for each zone. Let us name this sequence of networks as L. Thus, a linked network composed of 4 clusters of 100 nodes will be named as $L_{4\times100}$. Two examples are shown in Fig. 2. For the whole L sequence, the different zones are connected by a ring topology. It is important to avoid confusions and make clear that these networks are in reality networks formed from loosely interconnected sub-networks, but we use the denomination *linked networks* to be brief when referring to them.
- Finally, a third sequence will be used (named H^{big}) also composed with the same algorithm as sequence H. The reason for introducing this sequence is to obtain a homogeneous sequence similar to L (the linked one) from the point of view of the structure of the network. In particular, we try to obtain a similar curve for the average node path length and the average clustering coefficient to that of the L sequence. These quantities are important to characterize the structure of a network and will be introduced in detail later in this section. As shown in Fig. 3, both the average node path length and the average clustering coefficients have a strong dependence on S, the area of the square used for building the networks. Thus, this will be the parameter used as a degree of freedom to build the sequence H^{big}. In particular, the networks in the sequence are built with $S = N/100$, this is, the area of the network grows linearly with the number of nodes. For visual comparison, in Fig. 4, the standard H_{800} homogeneous network is shown in the left panel and H^{big}_{800} is shown in the right panel.

There is still an open issue. When building the sequence L, we limited the number of links between zones to 2 (to keep the non homogeneity of the network), but there are many ways one can link the zones. One important criterion we use in the linking process is to minimize the cost of transferring energy from one zone to another. If we choose the nodes to be connected in such a way that the average path length between nodes in the network $\langle l \rangle$ has the lowest possible value, the transmission costs are minimized. In practice, the minimization of $\langle l \rangle$ or the average resistance $\langle \eta \rangle$ of the lines between nodes leads to the same linked solutions because the impedence is proportional to the line length. The length $\langle l \rangle$ is defined as the average value of l_i over all nodes:

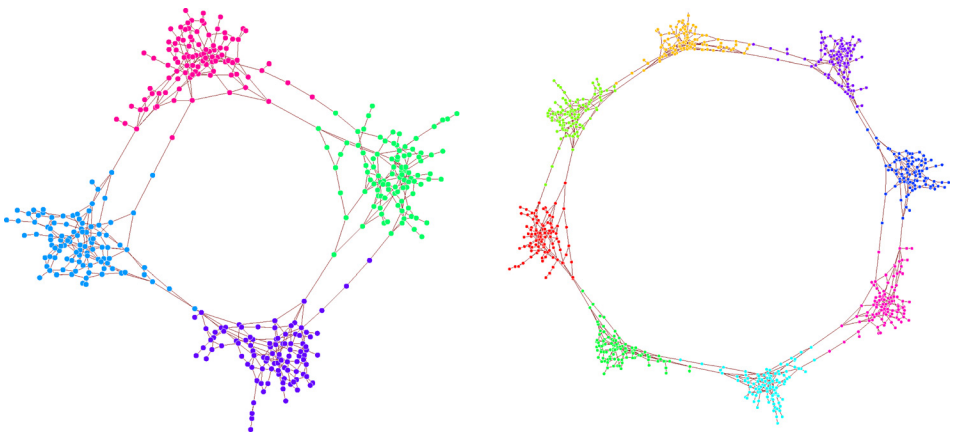

FIG. 2. Examples of four (left) and eight (right) 100 node linked networks: $L_{4\times100}$ and $L_{8\times100}$, respectively. Each zone is colored differently.

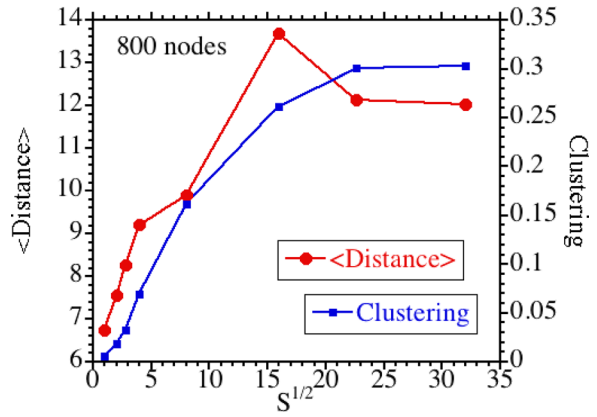

FIG. 3. Average path length $\langle l \rangle$ and averaged clustering coefficient $\bar{C}$ for networks of size $N = 800$ as a function of network area S. The algorithm used to compose the networks is the same used when building sequences H and H^{big} but keeping N constant and modifying the value of S.

$$\langle l \rangle = \frac{1}{N}\sum_{i=1}^{N} l_i, \tag{2}$$

where l_i is the average path length of node i to all other nodes:

$$l_i = \frac{1}{N}\sum_{j=1}^{N} d_{ij}, \tag{3}$$

where N is the number of nodes in the network and the distance d_{ij} between the node i and node j is defined as the minimum number of lines traversed in going from one node to another. A similar definition to that of $\langle l \rangle$ can be used to define $\langle \eta \rangle$.

In the case of the sequences of the homogeneous networks H and H^{big}, the average path length $\langle l \rangle$ increases very slowly with the number of nodes of the network, as shown in Fig. 5. However, in the case of a linked network the value of $\langle l \rangle$ depends on how the network has been constructed. If we just choose the nodes to be linked at random, the average path length between each node and all the other nodes can be very large compared with the homogeneous case, as shown in Fig. 5. This is because in going from one node to any other we have to move through the linking lines when we go from one zone to the other and the position of the linking lines

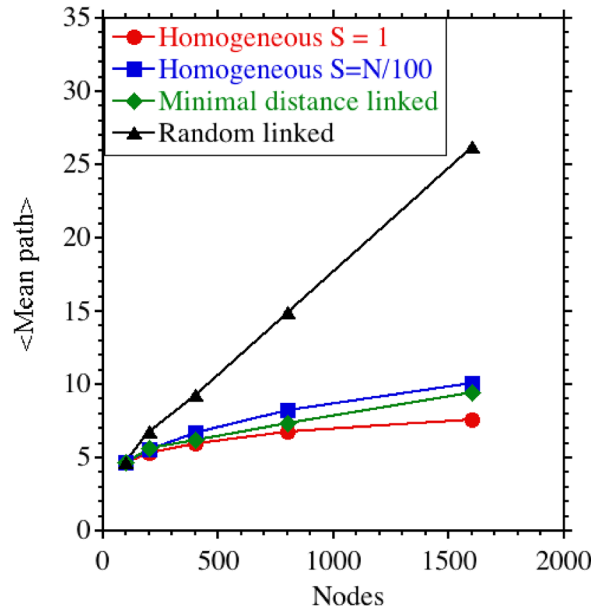

FIG. 5. Average path length in the network as a function of size for the different sequences of networks analyzed (H, L, H^{big}) and random linked.

may not be optimal to minimize this distance. On the other hand if, as previously mentioned, $\langle l \rangle$ is minimized in the linking (and thus the cost), then the $\langle l \rangle$ scaling is similar to those of the homogeneous cases. This last option is the criteria for the construction of the network sequences L used in the rest of the paper. As a side note, it is interesting to notice that using this linking criterion for the networks not only minimizes transmission costs but also reduces the vulnerability of the network to large size failures as we have shown in Ref. 22.

The average path length between nodes is one of the measures that are typically used when analyzing networks. There is another important measure used less frequently, the average clustering coefficient, which measures the degree to which nodes in a network tend to cluster together. The average clustering coefficient $\bar{C}$ can be defined as[23]

$$\bar{C} = \frac{1}{N}\sum_{n=1}^{N} C_n, \tag{4}$$

where C_n is the local clustering coefficient of the node n:

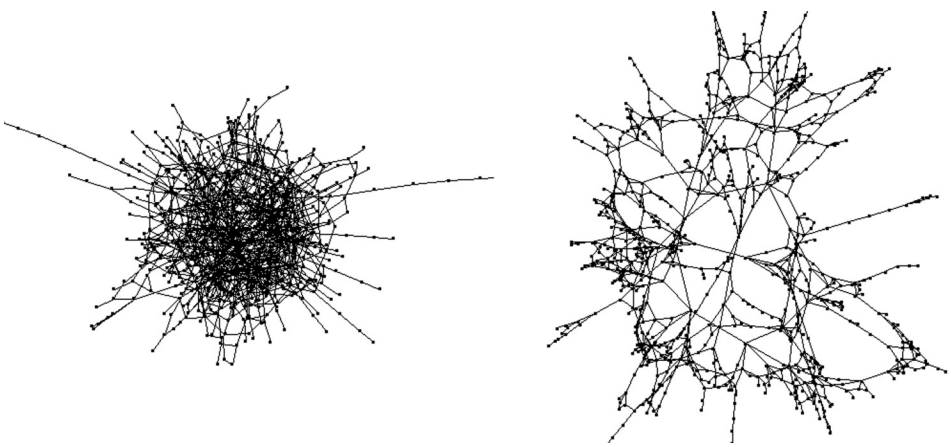

FIG. 4. Comparison of the homogeneous networks H_{800} (left) and H^{big}_{800} (right).

$$C_n = \frac{e_n}{k_n(k_n - 1)}, \tag{5}$$

where k_n is the number of neighbors of n and e_n is the number of connected pairs between all neighbors of n.

As explained, the homogeneous sequence H^{big} was built to emulate the structure of the linked sequence L. As seen in Fig. 6, both network sequences show a large very constant $\bar{C}$, when compared with the sequence H that decays with the size of the grid. The constant $\bar{C}$ associated with the sequence H^{big} makes sense when we realize that the area of the networks grows linearly with N as $S = N/100$: if for example, we duplicate the number of nodes of the network at the same time as the area, the connectivity between nodes found by the algorithm that built the networks will be, on average, the same. And the value of $S = 1$ for the case $N = 100$ makes that network equivalent to the H_{100}, and this is because both sequences converge in Fig. 6 at the point $N = 100$. Of course at that point also the linked sequence will converge because $L_{1\times100}$ is also equivalent to H_{100}.

In Sec. IV, using the three sequences of networks just described, we will analyze how important the differences are between the simulations of the different solvers (dispatch solutions), the underlying reasons for them in the model, and what effect the network structure has on the differences.

IV. OPA RESULTS FOR THE NETWORK SIZE SCAN: SENSITIVITY TO THE DETAILS OF THE DISPATCH

A. Homogeneous sequence *H*

Now we can start to investigate the effect of different dispatch procedures on different network structures. First, we use the OPA model with each of the three solvers introduced in Sec. II to study the dynamical evolution of the sequence of homogeneous networks H_{100}, H_{200}, H_{400}, H_{800}, and H_{1600}. Once again, the three solvers give different solutions for dispatch, although the minimum cost is the same for the three solvers. For a fixed set of parameters, we can compare the failure statistics of the model systems with the different dispatch procedures (really the different solvers). For this sequence of networks, the statistical results are remarkably virtually the same for the three solvers within the expected statistical uncertainty due to the finite size of the samples, as shown in Fig. 7. In this work, we will start by focusing on two measures of the failure statistics important for complex system dynamics. In Fig. 7(a), we compare the frequency of blackouts for the solutions of the three solvers. These results are for $p_1 = 0.037$ and $p_0 = 0.00025$, although similar results are obtained for other values of the parameters. We can see the similarity of the frequency curves for the 3 solvers, although there is a systematic deviation, still within error bars, for solver 2, giving slightly higher values of the blackout frequency. It should be noted that the increase in frequency with size in this figure comes from the fixed p_0 which causes p_0N, the total probability of a trigger, to grow with size. In Fig. 7(b), for the H_{1600} network we have plotted the Rank function of the load shed during each blackout normalized to the total power demand. The Rank function measures the complementary cumulative distribution function (ccdf) and is a method for measuring the tail of the distribution and therefore the probability of the large rare events that can dominate the risk. Again, we can see that the results obtained by the three solvers are very close. These results from the OPA model seem to indicate that although the dispatch solution is different for the three solvers, it does not matter for the statistical results of the model for the homogeneous networks of varying size.

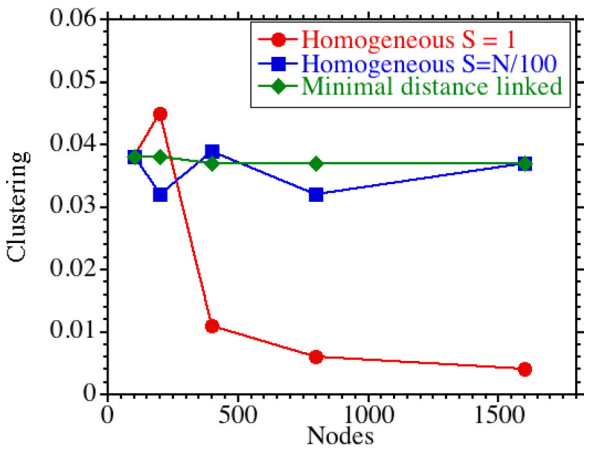

FIG. 6. Average clustering coefficient $\bar{C}$ in the network as a function of size for the three types of network sequences H and H^{big} and L.

B. Linked sequence *L*

Following the same reasoning, we now apply the same analysis with the same OPA parameters to a sequence of linked networks. Remember that each linked network is built by linking several 100-node homogeneous networks. The particular networks used are $L_{1\times100}$, $L_{2\times100}$, $L_{4\times100}$, $L_{8\times100}$, and $L_{16\times100}$, and the results for this sequence are quite different from the sequence of homogeneous networks. In Fig. 8(a), we have again plotted the frequency of blackouts. The minimum costs for the three solvers are, as with the homogeneous networks, relatively close. However, the blackout size distributions are completely different, as is shown in Fig. 8(b). The Rank functions have very different tails for the three cases considered. The tail for the first dispatch method, solver 1, is somewhat heavier than that for the third, solver 3, while the distribution for the second method, solver 2, has a very heavy tail.

Differences can also be studied as a function of one of the most important metrics of the problem from a practical point of view: the risk of blackout. There are many ways to calculate Risk. One measure for the risk is proposed in Ref. 24. In brief, the measure is an integrated value for the risk taking into account the size of the blackout into the cost. The risk associated with failure i is then defined as:

$$Risk(i) = Probability(i) \cdot Cost(i). \tag{6}$$

The probability of an event $Probability(i)$ is obtained from the OPA simulations. The cost associated with the event $Cost(i)$ is more difficult to accurately determine, as discussed in Ref. 24.

However, based on that reference and as a simple first approach, the cost is set proportional to the product of energy lost during the blackout,[25] that is, the power lost times the

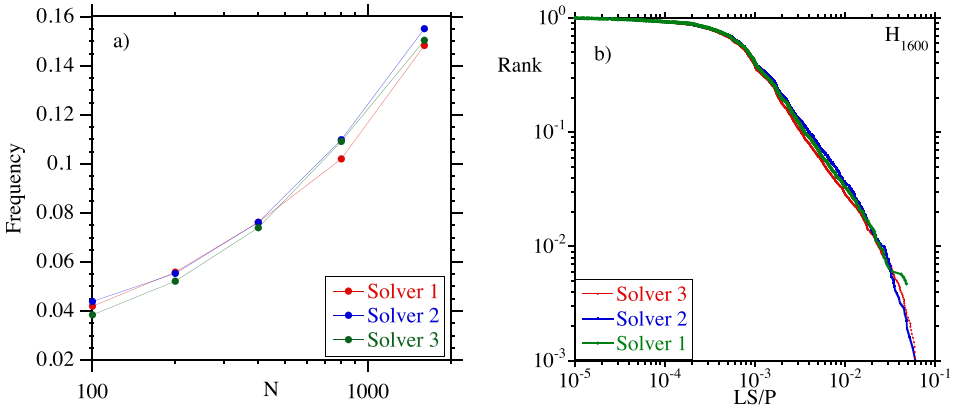

FIG. 7. (a) Frequency of blackouts as a function of the size of the homogeneous network and (b) rank function of the load shed for the particular case H_{1600}. They are obtained with the OPA model for the three solvers.

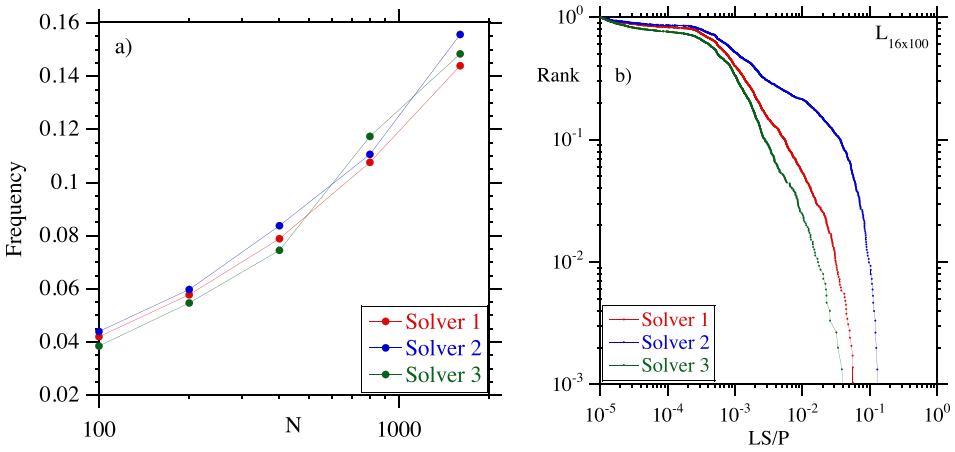

FIG. 8. (a) Frequency of blackouts as a function of the size of the linked networks and (b) rank function of the load shed for the $L_{16\times100}$ network. They are obtained with the OPA model for the three solvers.

duration of the blackout. The duration of the blackout we assume to be proportional to the size of the blackout and therefore to the energy lost is proportional to the square of the blackout size. Then, for an event with load shed L_{shed} the risk is

$$Risk(L_{shed}) = BP^2 \cdot probability(L_{shed}) \cdot \left(\frac{L_{shed}}{P}\right)^2, \qquad (7)$$

where B is a constant and P is the total power demand and L_{shed}/P is therefore the normalized load shed. Finally, the total risk value is obtained by integrating Eq. (7) over all possible load shed and using the load shed distribution obtained from the OPA simulations. The interested reader can find more details of the implementation on the reference. Here, we only use the risk as an integrated measure of the differences between dispatch solutions. In Fig. 9, we show the risk as a function of the network size for the linked sequence. The differences between solvers grow with the system size N, probably due to the extra degrees of freedom that allow the solvers to increase the divergence.

As previously discussed, the three solvers give the same value for the minimum cost for the same initial condition of the solver; however, the solutions are clearly not the same and the probability of a large failure, and the tail, is greatly different in the three cases. We can find some of the causes of these differences by looking at the distribution of the fractional loading of the lines, M_i, coming from the different solvers with the same conditions of the network. The fractional line loading for line i, M_i, is the ratio of the power flow in line i, F_i, to the maximum power flow allowed in this line, F_i^{max}. Using identical initial conditions for the different solvers after calculating the dispatch on a single iteration, we can plot the distribution of the M_i for the two solvers that give the maximum discrepancy. The results for solvers 2 and 3 are shown in Fig. 10(a) for the H_{1600} network and in Fig. 10(b) for the $L_{16\times100}$ network. The distribution has a peak

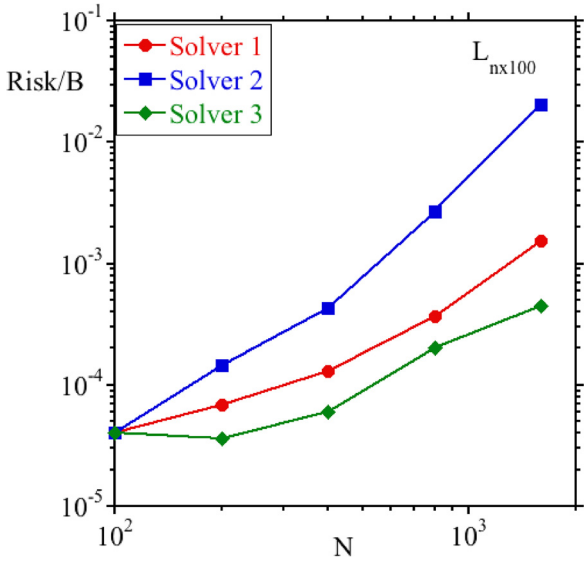

FIG. 9. Risk measure as a function of the nodes number N for the L sequence of networks using the different solvers. The three sequences converge on the case N = 100 because there is only one zone, so this case is equivalent to the homogeneous H_{100}.

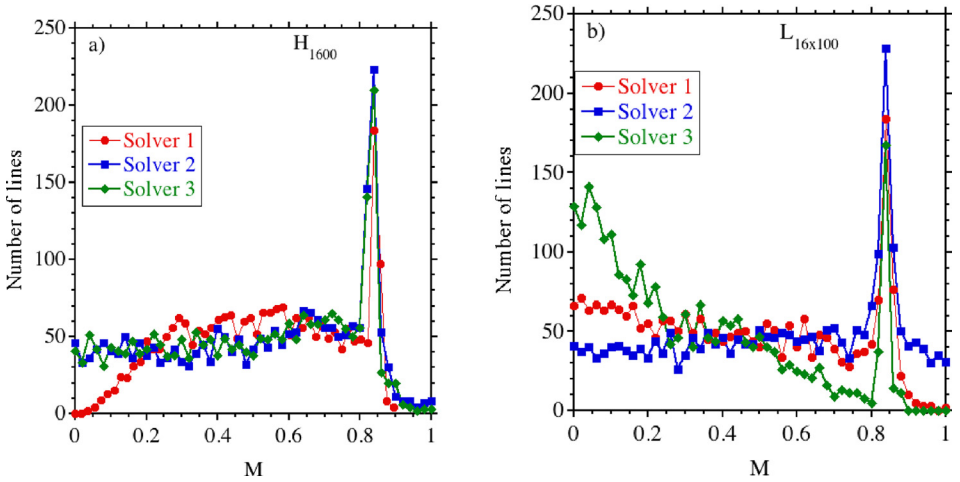

FIG. 10. Distribution of the fraction of load of the lines after a single step using solvers 1, 2 and 3 for: (a) H_{1600} network and (b) $L_{16\times100}$ network.

above $M_i = 0.8$ because of the combination of two effects. On one hand, the self-organization of the system drives the lines to be close to their operational limits so lines move M_i to be above 0.8. On the other hand, for $M_i \geq 0.9$, lines fail with probability p_1 so there are few lines with M_i above 0.9.

For the homogeneous network, the fractional line loading given by the solvers to each line is not the same for all lines; however, the overall distribution is very similar. Note that solver 2 has a few more lines over the value of $M = 0.9$, so this solution is slightly more vulnerable. For the linked network, the difference between distributions is dramatic; solver 3 gives a very much more conservative solution than solver 2's solution with lower values of M. Because the p_1 gives the probability of failure of lines with $M > 0.9$, for the same value of p_1, the solution from solver 2 has a larger chance of failure and the subsequent cascading process will be a great deal longer. This is an indication that the linked systems are more sensitive to a dispatch than the homogeneous ones, and the vulnerability of the system varies with the details of the dispatch solutions.

The increased overloading of lines by solver 3 in the linked network case may be a consequence of the selection of generators from which the dispatched power is chosen. For the same case shown in Fig. 10(b), we show in Fig. 11 the distribution of the generation power by the zones that have been linked. The demand is practically the same in the 16 zones; however, the power generation dispatch varies depending on the solvers. Solver 3 gives a practically uniform dispatch by zones; this is the solver with lower values of M_i and a less heavy tail. However, the generation varies a great deal from zone to zone for solver 2; this dispatch also has most of the higher values of M_i and the heaviest tail. Solver 1 gives more variation in the generation distribution than solver 3, but is not as bad as solver 2 and hence has the tail which is in the middle in terms of heaviness. Finally, note that any trend in Fig. 11 is purely algorithmic: each node has associated an ordering number when the network is build (with a ring topology). The dispatch solver tends to pick up this order of the nodes in beginning of the dispatch. For example, solver 2 tries to dispatch starting from the lowest node number generator, which is one of the roots of its problems (visible in Fig. 11 as an unbalance in the generation).

Naturally, in real systems, the operators do the most intelligent dispatch they can and some of the problems mentioned here might not be relevant in a real network; however, they are important in the building of models for the networks' dynamics in order to do the optimal dispatch. The present results indicate that the Solver 3 does the best job. In Sec. V, we will discuss methods to reduce the vulnerability of the networks to the type of dispatch.

C. Homogeneous sequence H^{big}

An important issue still left open is whether the sensitivity of the networks to the dispatch is only due to the fact that they were linked networks or if there are other possible parameters that could also be affecting their sensitivity to dispatch. In this section, we investigate homogeneous networks where the parameter S (the area where the nodes of the network are placed) is varied as S = N/100. Remember that the effect of this variation is to obtain similar structural properties to the linked sequence, while keeping the homogeneous character. The corresponding sequence is composed of the networks H^{big}_{100}, H^{big}_{200}, H^{big}_{400}, H^{big}_{800}, and H^{big}_{1600}.

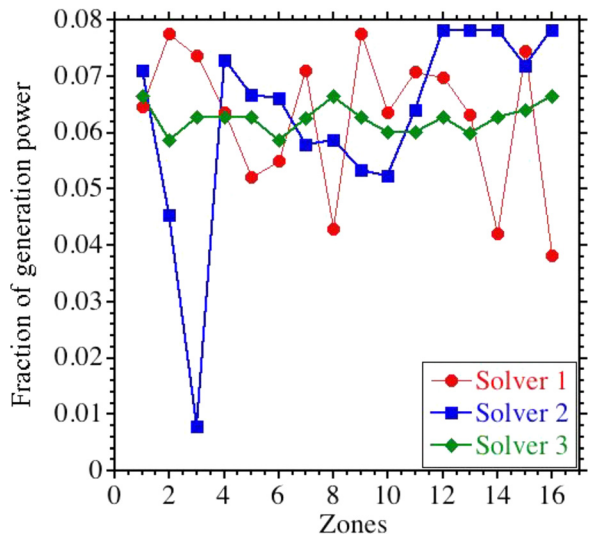

FIG. 11. Distribution of the generation power by the zones for the $L_{16\times100}$ network using the three solvers.

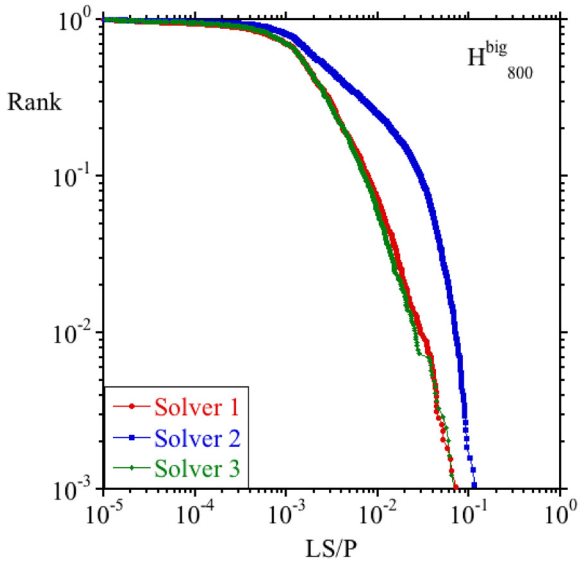

FIG. 12. Rank function of the load shed for the H^{big}_{800} network for solvers 1, 2 and 3.

We find that the H^{big} sequence shows a similar sensitivity to the dispatch to that of the linked one. An example of the size distributions is shown in Fig. 12. It is clear that the big network with dispatch solver 2 has the heavier tail, very similar to the linked networks. This result is not surprising from the structural point of view, because both sequences are similar. But from a modeling point of view, both sequences have a very different origin, so it is important to be careful and not suggest that the linking property is the only property responsible for the enhanced discrepancies between the different solvers. The parameter S (area of the network) can also have a similar effect as perhaps could other parameters.

D. Effect of $\langle l \rangle$ and $\bar{C}$ on the discrepancies

Up to this point, we have found that, due to the degeneracy of the problem, the different solvers (the different dispatch solutions) can cause different results in the long term complex system dynamics. The differences between the results of the simulations can be enhanced or reduced depending on the structure of the network. It is important to emphasize here that the dispatch solutions applied to each network simulation are exactly the same, but the difference between the simulations final results depends on the network structure. In this section, we try to determine the correlations between this difference and the structure of the network interpreted as an undirected graph. Namely, is there any graph measure that correlates with the discrepancy? There are many related works studying the interplay between the power grid and its topology (see, for example, Refs. 26 and 27) but here we particularize the analysis to the OPA model results. We do not find large changes in the average path length $\langle l \rangle$ curves between the three sequences studied (Fig. 5), which suggests that it is not a critical measure in distinguishing between the networks sensitive to the solver and those which are insensitive. However, the average clustering coefficient $\bar{C}$ curves of the different sequences (Fig. 6) show interesting differences. Both network sequences L and H^{big} that present the largest differences between the different dispatch methods (especially for large N) also show a large $\bar{C}$, when compared with the sequence H which is rather insensitive to the solver used. This suggests a correlation between a high value of $\bar{C}$ and large differences between the different dispatch solutions. The fact that the sequence H^{big} is completely different from L but is similar from the structural point of view and shows a similar behaviour to that in the L cases also adds support to this argument.

An interesting question opens up: why is a large value of $\bar{C}$ correlated with the differences? A possible explanation is that a large clustering coefficient allows for many more ways of dispatching the power to different nodes and regions (see Fig. 13), giving extra degrees of freedom that increase the degeneracy of the minimization problem. Thus, the differences in the solutions between the different Simplex solvers can be large. On the other hand, a small clustering coefficient does not allow for many options of dispatch, keeping the solutions of the different Simplex solvers close.

Finally, it is important to add that the parameter $\bar{C}$ is not the only ingredient, even if it is an important one. For example, the linked sequence L shows a similar value for the $\bar{C}$ in all the network sizes. The discrepancy between solvers increases with the number of nodes, so N is also an important parameter.

V. OPA RESULTS FOR THE NETWORK SIZE SCAN: REDUCING THE DISPATCH VARIABILITY

In this section, we will explore modifications in the dispatch solution required to reduce the discrepancy between the different Simplex solvers. This helps in understanding the underlying dynamics, understanding what makes a good versus bad dispatch and also results in more robust dispatch

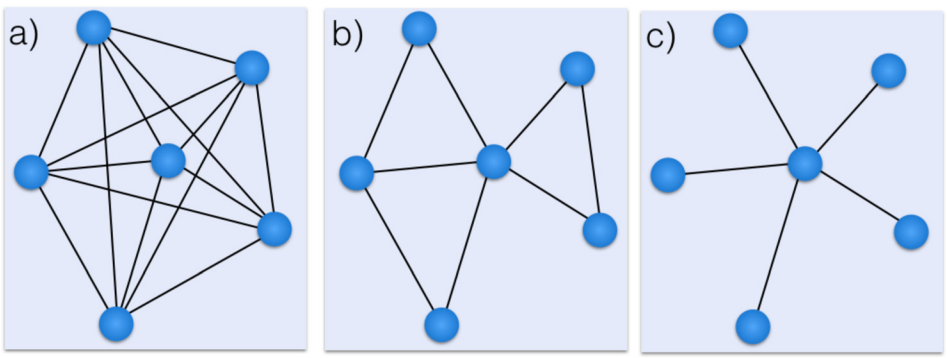

FIG. 13. Example of simple networks with different clustering coefficients: (a) $\bar{C} = 1$, (b) $\bar{C} = 0.3$, and (c) $\bar{C} = 0$. Clearly the network (c) will show less possible dispatch solutions and thus less degeneracy.

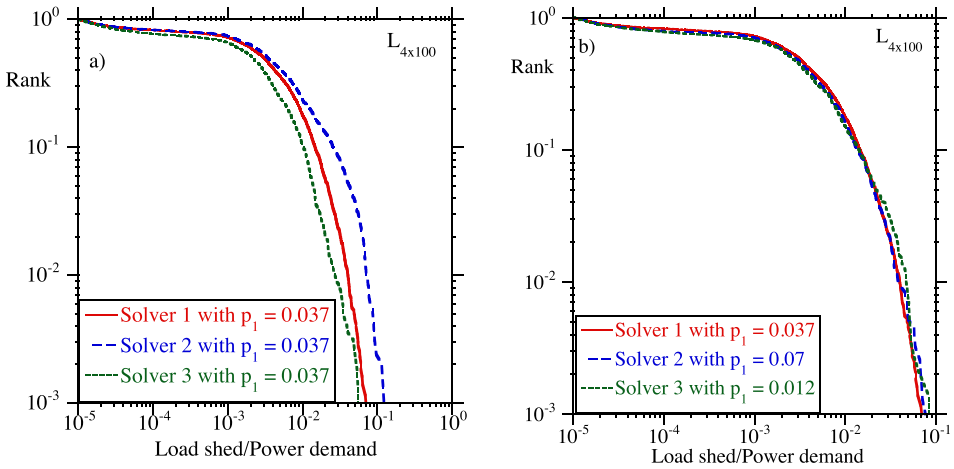

FIG. 14. Rank function of the normalized load shed for the $L_{4\times100}$ network using the three solvers, (a) for a fixed value of p_1 and (b) for values of p_1 leading to the same rank function.

solutions that are independent of the algorithm used for the minimization. Also it allows the use of the Simplex solver that is computationally more efficient. The results of Sec. IV suggest three approaches to solving the discrepancies between the solvers' dispatches.

One simple approach considers the vulnerability of the lines as arising from the distribution of the lines with $M > 0.9$ combined with the value of p_1. The value of p_1 is obtained by fitting the rank function curve for the load shed in the real systems data. If our objective is modeling a realistic network, we can use any of the three solvers and by adjusting the value of p_1, we obtain the desired overall vulnerability (load shed curve). This can be seen in Fig. 14 for the $L_{4\times100}$ network. With systematic adjustment of p_1, the rank functions can be modified from showing a large disparity (the left panel) to very good agreement (right panel). Of course, because the dispatch is different (each solver makes different decisions), the obtained value for p_1 will differ. That is, since the dispatch has not been modified, the system dynamics has to be modified in order to force the solutions to converge. A possible application of this approach could be to obtain the required p_1 value for each of the dispatch solutions, based on a prescribed load shed rank function. Based on the results, because p_1 is a measure of the reliability of the lines in the network, the more reliable are the lines the more robust is the system (load shed rank function pushed down). In particular, solver 2 requires a smaller value of p_1 in order to show the same performance as the other solvers, which precludes it from the engineering point of view. The preferred dispatch would be from solver 3, the least restrictive in terms of p_1. Finally, these kinds of analyses have to be done with care, because the fitting process accounts for the risk discrepancy, but not necessarily the dynamical differences.

A second approach consists of keeping the original p_1, but introducing modifications to the conditions of the dispatch to make them agree better, in particular, adding a small random perturbation in the generator costs (values of α_i in the cost function Z, Eq. (1)). Thus, from the many possible solutions due to the degeneracy of the problem, the random perturbation in the generation costs favors one of them. We cannot prove that there is only one solution so it could be cases where degeneracy still persists even with the proposed modification. However, the case studied shows an almost complete destruction of the degeneracy. This is, the three solvers converge to the same solution. The result is shown in Fig. 15 for the $L_{4\times100}$ network (the same case studied in the first approach). In that case, a perturbation of one part in a thousand is applied to the generators costs and, as can be noted, the three solvers' results match almost perfectly. However, even if this seems a good approach, there is some risk behind its use. It could happen in some cases that the results are very sensitive to the particular random perturbation chosen. This can happen because the random cost perturbation forces the three solvers to make the same choice between all the possible degenerated states, but there is no guarantee that this choice models properly the desired dispatch. The solution of course is not to do a random perturbation but an intelligent one, as we discuss in the third approach.

In the third approach, useful for the case of linked networks, we ensure a more efficient distribution of the generation of electricity in each of the subsystems that we have

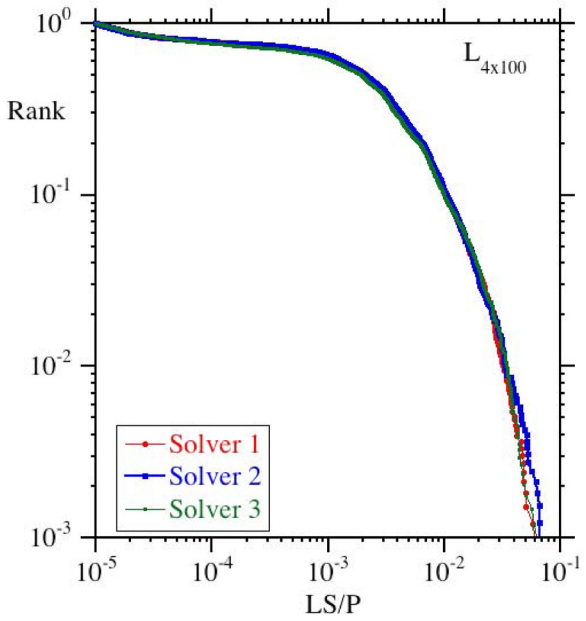

FIG. 15. Rank function of the normalized load shed for the $L_{4\times100}$ network using the three solvers but with a slight random perturbation (on part in a thousand) on the generator costs.

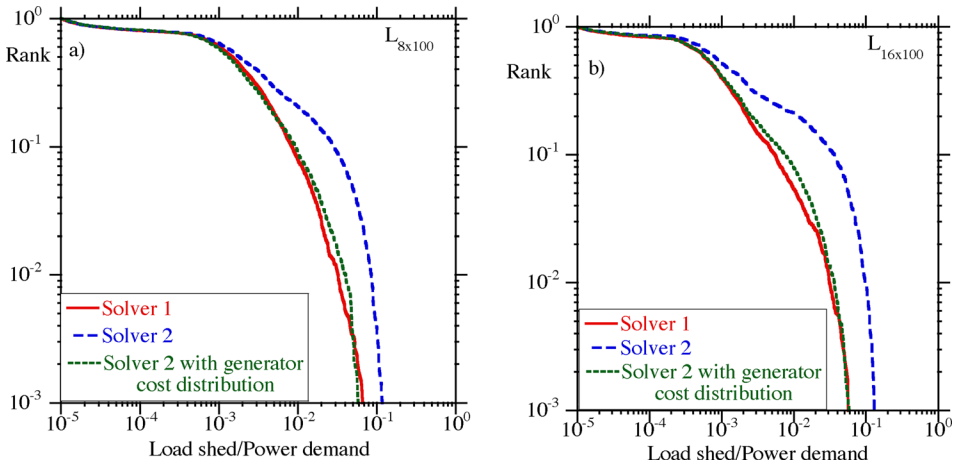

FIG. 16. Rank function of the normalized load shed using solvers 1 and 2 and solver 2 with modified generation cost distribution as a function of the number of nodes N, (a) for the $L_{8\times100}$ node network and (b) for the $L_{16\times100}$ node network.

linked to avoid the kind of dispatch shown by solver 2 in Fig. 11. In the OPA model, this can be achieved by lowering some of the generator costs in each of the zones (modifying their α_i), which reduces transmission of electricity across zones and forces more of the generation to be local. For the sequence of linked networks L, we set half of the generators in each 100 node subsystem to a lower cost and then recalculate the dynamical evolution of the sequence of networks. In Fig. 16, the results of using the solver 2 with this modification of the cost function against the solver 1 are compared to the original cost function. The analysis is done for the $L_{8\times100}$ and $L_{16\times100}$ networks. Fig. 16 shows a relatively good agreement between both solvers on the blackout size distribution, when compared with the results of the solver 2 without modification. Also the risk function shows a better convergence with the modified solver, as shown in Fig. 17. The reason for the convergence of the solvers in this approach can also be understood in terms of what we learned on the second approach: the variation in the generator costs reduces the degeneracy of the problem. Even if this third approach does not allow so good agreement between solvers as the second one, it will be preferred because of the physical arguments behind it.

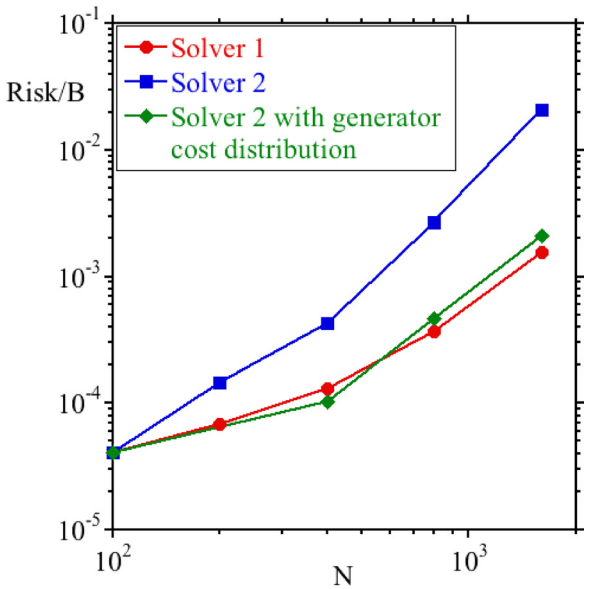

FIG. 17. Risk using solvers 1 and 2 and solver2 with modified generation cost distribution as a function of the number of nodes N.

Therefore, by having at least partial generation dispatch from each of the regions of the linked networks, many of the problems due to the sensitivity to dispatch solutions may disappear. A similar analysis could be done for the sequence of H^{big} networks as Fig. 4(b) suggests (some clusters can be found visually), but a more complicated procedure would be required.

VI. CONCLUSIONS

Using the OPA model, we have been able to study and characterize the mechanisms behind the power law tails in the distribution of the blackout size. These algebraic tails obtained in the numerical calculations are consistent with those observed in the study of the blackouts for real power systems. The OPA code uses the Simplex algorithm to calculate the dispatch solutions for the same minimum of the cost function. For this minimum cost, there may exist several dispatch solutions. We use three different solvers to model some possible dispatch solutions and in this way evaluate the sensitivity of the networks to the chosen dispatch solution. These different solutions do not affect the minimum found in the optimization problem through the Simplex algorithm (because the solution is not unique) but when they are incorporated into the full dynamics of the system they may lead to different levels of vulnerability.

To explore these effects, we have constructed different sequences of networks of varying sizes. One is homogeneous keeping the same distribution of distances between nodes as we increase the size of the network; another is the same type (homogeneous) but with the surface of the network increased in a way proportional to the area. Finally, a third sequence is obtained by linking smaller networks, using the connection criteria of minimizing the transmission costs.

The results indicate that for a given network size, the sensitivity to the dispatch increases when the network is built from a linked set of smaller networks and also when the area is increased. From the graph analysis point of view, there is a strong suggestion that a high averaged cluster coefficient together with a large size of the network causes increased sensitivity to the dispatch method used.

We have also explored how to decrease the sensitivity to the dispatch method. Adding small discrepancies between generation costs almost completely destroys the degeneracy of the problem and thus eliminates the sensitivity. A disadvantage of that method is the lack of control about the particular solution at which the degeneracy converges, but an intelligent non-random choice of the costs could constitute a good approach. Based on this, another possibility for decreasing the sensitivity consists of forcing the generation to be partly local by distributing some low cost generation to the clusters. With either of the two modifications both the rank function of the load shed and the risk significantly reduce their sensitivity to the dispatch solution used. This allows the use of a computationally more efficient solver. In particular, Solver 2 is computationally more efficient and, in those cases where its solution is good enough, it should be the preferred solver.

In future work, a more detailed analysis of the dispatch solution and risk function in quantitative measures of the network structure (radius of the network, average path length, averaged clustering coefficient, and/or others) will be done.

ACKNOWLEDGMENTS

The authors want to thank the reviewer 1 of the manuscript for proposing the second choice in Section V: randomly perturb slightly the generator costs in order to destroy the degeneracy. This research was sponsored by Ministerio de Economía y Competitividad of Spain under Project Nos. ENE2012-31753, ENE2012-33219, and ENE2015-68265-P. Simulations have been run in the supercomputer cluster Uranus located at Universidad Carlos III de Madrid (Spain), funded by the Spanish Government via the National Project Nos. UNC313-4E-2361, ENE2012-33219 and ENE2012-31753. One of the authors (J.M.R.B.) acknowledges useful interactions with members of the research network Avalanchas en biofsica, geofsica, materiales y plasmas, funded by the Spanish Project No. MAT2015-69777-REDT.

[1]B. A. Carreras, V. E. Lynch, I. Dobson, and D. E. Newman, "Complex dynamics of blackouts in power transmission systems," Chaos **14**, 643–652 (2004).
[2]I. Dobson, B. A. Carreras, V. E. Lynch, and D. E. Newman, "Complex systems analysis of series of blackouts: Cascading failure, critical points, and self-organization," Chaos **17**, 026103 (2007).
[3]S. Mei, X. Zhang, and M. Cao, *Power Grid Complexity* (Tsinghua University Press, published by Springer, 2011).
[4]G. B. Dantzig, "Programming in a linear structure," Econometrica **17**, 200–211 (1949).
[5]G. B. Dantzig, *Linear Programming and Extensions* (Princeton University Press, Princeton, NJ, 1963).
[6]J. Matoušek and B. Gärtner, *Understanding and Using Linear Programming* (Springer, 2007).
[7]B. A. Carreras, D. E. Newman, I. Dobson, and A. B. Poole, "Evidence for self-organized criticality in a time series of electric power system blackouts," IEEE Trans. Circuits Syst. I **51**, 1733–1740 (2004).
[8]A. J. Holmgren and S. Molin, "Using disturbance data to assess vulnerability of electric power delivery systems," J. Infrastruct. Syst. **12**, 243–251 (2006).
[9]J. Ø. H. Bakke, A. Hansen, and J. Kertesz, "Failures and avalanches in complex networks," Europhys. Lett. **76**, 717–723 (2006).
[10]G. Ancell, C. Edwards, and V. Krichtal, *Electricity Engineers Association 2005 Conference Implementing New Zealands Energy Options, Aukland, New Zealand* (2005).
[11]X. Weng, Y. Hong, A. Xue, and S. Mei, "Failure analysis on China power grid based on power law," J. Control Theory Appl. **4**, 235–238 (2006).
[12]Y. Qun and G. Jianbo, "Self-organized criticality and its application in power system collapse prevention," in *International Conference on Power System Technology, Chongqing, China* (IEEE, 2006).
[13]X. Zhao, X. Zhang, and B. He, "Study on self organized criticality of China power grid blackouts," Energy Convers. Manage. **50**(3), 658–661 (2009).
[14]B. A. Carreras, B. A. Newman, I. Dobson, and N. S. Degala, "Validating OPA with WECC data," in *Proceedings of the 46th Hawaii International Conference on System Sciences* (IEEE, 2013).
[15]G. B. Dantzig, "Maximization of a linear function of variables subject to linear inequalities," in *Activity Analysis of Production and Allocation*, Cowles Commission Monograph 13, edited by T. C. Koopmans (Wiley, New York, 1950).
[16]P. M. J. Harris, "Pivot selection methods of the Devex LP code," Math. Program. **5**, 1–28 (1973).
[17]X. Liu and Z. Li, "Revealing the impact of multiple solutions in DCOPF on the risk assessment of line cascading failure in OPA model," IEEE Trans. Power Syst. **31**(5), 4159–4160 (2016).
[18]B. A. Carreras, D. E. Newman, M. Zeidenberg, and I. Dobson, "Dynamics of an economics model for generation coupled to the OPA power transmission model," Forty-third Hawaii International Conference on System Sciences, Hawaii, IEEE, January 2010.
[19]B. A. Carreras, D. E. Newman, I. Dobson, and M. Zeidenberg, "The impact of risk-averse operation on the likelihood of extreme events in a simple model of infrastructure," Chaos **19**, 43107 (2009).
[20]D. Goldfarb and J. Reid, "A practicable steepest edge simplex algorithm," Math. Program. **12**, 361–371 (1977).
[21]Z. Wang, R. J. Thomas, and A. Scaglione, "Generating random topology power grids," in *Proceedings of the 41st Hawaii International Conference on System Sciences* (IEEE, 2008).
[22]B. A. Carreras, D. E. Newman, I. Dobson, and J. M. Reynolds-Barredo, "The impact of local power balance and link reliability on blackout risk in heterogeneous power transmission grids," in *Proceedings of the 49th Hawaii International Conference on System Sciences* (IEEE, 2016).
[23]D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small-world' networks," Nature **393**, 440–442 (1998).
[24]B. A. Carreras, D. E. Newman, and I. Dobson, "Does size matter?," Chaos **24**, 023104 (2014).
[25]R. Billinton, J. Otengadjei, and R. Ghajar, "Comparison of two alternate methods to establish an interrupted energy assessment rate," IEEE Trans. Power Syst. **2**, 751–757 (1987).
[26]E. Cotilla-Sanchez, P. D. H. Hines, C. Barrows, and S. Blumsack, "Comparing the topological and electrical structure of the North American electric power infrastructure," IEEE Syst. J. **6**, 616–626 (2012).
[27]P. Hines and S. Blumsack, "A centrality measure for electrical networks," in *Proceedings of the 41th Hawaii International Conference on System Sciences* (IEEE, 2008).