Network Observability Transitions

Yang Yang,[1] Jianhui Wang,[2] and Adilson E. Motter[1,3]

[1]*Department of Physics and Astronomy, Northwestern University, Evanston, Illinois 60208, USA*
[2]*Decision and Information Sciences Division, Argonne National Laboratory, Argonne, Illinois 60439, USA*
[3]*Northwestern Institute on Complex Systems, Northwestern University, Evanston, Illinois 60208, USA*

(Received 26 July 2012; published 18 December 2012; publisher error corrected 21 December 2012)

In the modeling, monitoring, and control of complex networks, a fundamental problem concerns the comprehensive determination of the state of the system from limited measurements. Using power grids as example networks, we show that this problem leads to a new type of percolation transition, here termed a *network observability transition*, which we solve analytically for the configuration model. We also demonstrate a dual role of the network's community structure, which both facilitates optimal measurement placement and renders the networks substantially more sensitive to "observability attacks." Aside from their immediate implications for the development of smart grids, these results provide insights into decentralized biological, social, and technological networks.

DOI: 10.1103/PhysRevLett.109.258701 PACS numbers: 89.75.Hc, 05.10.−a, 64.60.ah

Like other dynamical systems, a network is observable if its state can be determined from the given set of measurements, with observability depending on both the number and the placement of the measurements [1]. This concept is important for a range of questions, including the identification of therapeutic interventions in intracellular networks, modeling and forecasting in social networks, web crawling, monitoring and management of ecological networks, and control of power-grid networks [2]. Because measurements are inherently limited by cost and physical considerations, a question of interest concerns the identification of the optimal set of measurement points—e.g., sensors—with adequate redundancy that allow complete or (prespecified) partial observability of the network.

In a power-grid network, the state of the system can be defined as the (complex) voltage at all nodes. Such a state can, in principle, be determined by phasor measurement units (PMUs) [3], which are sensor devices that measure the voltage and line currents at the corresponding node in real time. Therefore, a PMU placed on a node makes both the node and (given the relation between current and voltage) all of its first neighbors observable—i.e., the states of those nodes are completely determined. If any of the neighboring nodes is a zero-injection node (i.e., without consumption or generation of power), then a corresponding second neighbor may also be observable, and so on [4]. In either case, the problem of identifying the observable nodes and the observability of the network itself is thus reduced to a purely topological one.

The observability of power-grid networks is a timely and broadly significant problem, which is also representative of many others. Technologies that allow real-time wide-area monitoring of the network are an integral part of the next generation of power grids—the so-called smart grids [5,6]—and PMUs are a central aspect of these technologies. It is believed, for example, that PMU information along with appropriate response could have prevented major recent blackouts [7]. While the technology underlying PMUs is well established, the high cost of required infrastructure, installation, and operation continues to limit the number of such units that can be installed in a given power grid. Accordingly, significant recent research has been pursued in connection with PMU placement under various constraints for incomplete, complete, and redundant observability [3]. However, the fundamental question of how the observability of the network relates to its structure remains underexplored.

In this Letter, we show that the random placement of PMUs leads to a new type of percolation transition [8]—a network observability transition. This transition characterizes the emergence of macroscopic observable islands as the number of measurement nodes is increased. Using the generating function formalism [9], we derive the exact analytical solution describing the size of the network's largest observable component (LOC). We study its dependence on the network structure to show, in particular, how the transition threshold decreases as a function of both average degree and degree variance. We then consider the optimal placement of PMUs, a problem of practical interest that has been hindered because no fast, deterministic algorithms currently exist to address large networks. Taking advantage of the community structure of real systems, we introduce a community-based approach in which the network is judiciously partitioned into smaller, largely independent components that can be solved exactly. Our efficient approach allows us to address for the first time very large networks, including the largest interconnection of the North America power grid—a 56 892-node network. We show, however, that community structure can also make the network more sensitive to the deliberate disabling of PMUs, in that a surprisingly small attack can separate the system into very small observable islands. This adds a new dimension to existing research on the network vulnerability of power-grid systems [10,11].

We first consider random PMU placement on networks generated using the configuration model for a given degree distribution $P(k)$ [12]. All nodes in the network are assumed to have a common probability ϕ of hosting a PMU. The observable nodes are classified into directly observable (hosting a PMU) and indirectly observable (neighboring a node with a PMU) [13]. To determine the LOC size as ϕ increases, we calculate the probability that a randomly selected observable node i is not connected to the LOC via a randomly selected edge e_{ij}. This probability is denoted u if node i is directly observable and s if neither i nor j is directly observable. To proceed, we use the generating function $G_0(x) = \sum_k P(k)x^k$, associated with the degree distribution, and the generating function $G_1(x) = G_0'(x)/G_0'(1)$, which describes the probability q_ℓ that, by following a randomly selected edge, one reaches a node with ℓ other edges (i.e., ℓ excess edges) [14].

Self-consistent equations for the probabilities s and u can be derived as follows. Starting with s, there are two independent cases in which node i is not connected to the LOC via a randomly selected edge e_{ij} [Fig. 1(a)]. In the first case, node j is not observable, which occurs with probability $G_1(1-\phi)$. In the second case, node j is observable (i.e., $n \geq 1$ excess neighbors of j are directly observable) and hence the probability that an excess neighbor of node j will not be connected to the LOC is $\phi G_1(u)$ if this neighbor is directly observable and $(1-\phi)s$ if it is not [Fig. 1(a), orange and green sub-boxes]. Accounting for all possible degrees of node j and values of n, the latter case occurs with probability $\sum_{\ell=1}^{\infty} q_\ell \sum_{n=1}^{\ell} \binom{\ell}{n} [\phi G_1(u)]^n [(1-\phi)s]^{\ell-n}$. Combining both cases, we obtain the final expression for s:

$$s = G_1(1-\phi) + G_1[\Psi(s,u;\phi)] - G_1[(1-\phi)s], \quad (1)$$

where $\Psi(s,u;\phi) = \phi G_1(u) + (1-\phi)s$ corresponds to the probability that one indirectly observable node is not connected to the LOC via a specific edge.

To derive a corresponding equation for u, we again split the problem into two cases [Fig. 1(b)]. In the first case, node j is directly observable but not part of the LOC, which occurs with probability $\phi G_1(u)$. In the second case, node j is indirectly observable but not connected to the LOC via any of its excess edges, and this occurs with probability $(1-\phi)G_1[\Psi(s,u;\phi)]$. Combining these two cases, we arrive at the final expression for u:

$$u = \phi G_1(u) + (1-\phi)G_1[\Psi(s,u;\phi)]. \quad (2)$$

Together, the self-consistent equations [(1) and (2)] provide all the information needed to determine s and u.

With u and s in hand, we now calculate the probability that a randomly selected node i is part of the LOC. If this node is directly observable, which occurs with probability ϕ, this probability is $\sum_{k=1}^{\infty} P(k)(1-u^k) = [1-G_0(u)]$. This expression has the same form as for ordinary site percolation [15], but here u is functionally different. On the other hand, we also have to account for indirectly observable nodes. If node i is not directly observable, which occurs with probability $1-\phi$, this node is observable only if $m \geq 1$ of its neighbors are directly observable. Thus, the probability that node i is part of the LOC is $1 - G_1(u)^m s^{k-m}$, where the term $G_1(u)^m$ accounts for the m directly observable neighbors and s^{k-m} accounts for the $k-m$ other neighbors of i. Considering all possible degrees k of node i and all possible values of m, the probability that this node is part of the LOC is $\sum_{k=1}^{\infty} P(k) \sum_{m=1}^{k} \binom{k}{m} \phi^m (1-\phi)^{k-m} [1 - G_1(u)^m s^{k-m}]$, which can be rewritten as $1 - G_0[\Psi(s,u;\phi)] - G_0(1-\phi) + G_0[(1-\phi)s]$. Combining the two cases, we obtain that the normalized size of the LOC is

$$S = 1 - \phi G_0(u) - (1-\phi)\{G_0[\Psi(s,u;\phi)] + G_0(1-\phi) - G_0[(1-\phi)s]\}. \quad (3)$$

This result is in excellent agreement with numerical simulations, as shown in Fig. 2(a) for configuration-model networks with the degree distributions from a selection of real power grids.

In particular, for a given degree distribution, and hence G_0 and G_1, there is a threshold ϕ_c at which S becomes nonzero. This percolation threshold is given by

$$\frac{(1-\phi)G_1'(1-\phi)}{G_1'(1)-1}[1 - \phi G_1'(1) - \phi(1-\phi)G_1'(1)^2] = 1, \quad (4)$$

which can be derived directly as the smallest ϕ at which Eqs. (1) and (2) hold for s and u smaller than 1. It follows immediately from this expression that the threshold ϕ_c is strictly positive unless $G_1'(1)$ (and hence the second moment of the degree distribution) diverges. In power grids, however, the degree distribution is relatively homogeneous, meaning that an observability transition will occur at a nonzero value of the threshold ϕ_c; nevertheless, $\phi_c \ll 1$ for the degree distributions we consider. This is

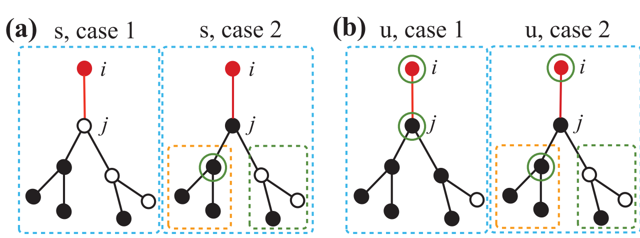

FIG. 1 (color). Diagram for the self-consistent equations of (a) the probability s and (b) the probability u that an observable node i (red circle) is not connected to the LOC through a specific edge e_{ij} (red line). The nodes are either not observable (open circles) or observable (solid circles), where green rings mark directly observable nodes.

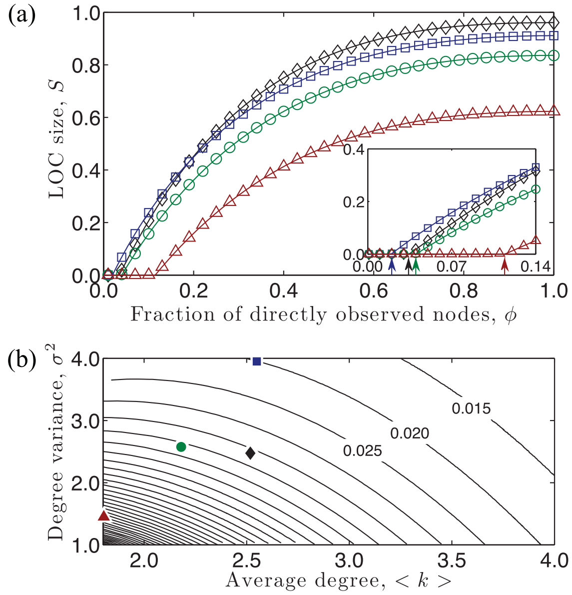

FIG. 2 (color). Network observability transitions. (a) LOC size as a function of ϕ in networks with the degree distributions of the power grids of Germany (red), Europe (green), Spain (blue), and Eastern North America (black) [11,21]. The continuous lines correspond to our analytical predictions, and the symbols to an average over ten 10^6-node random networks for ten independent random PMU placements each. The inset shows a magnification around the transitions, with the predicted thresholds ϕ_c indicated by arrows. (b) Dependence of ϕ_c on $\langle k\rangle$ and σ^2 for networks (in the thermodynamic limit) with Gamma degree distributions, where the curves indicate equispaced isolines of ϕ_c and the symbols indicate the $(\langle k\rangle, \sigma^2)$ positions of the corresponding networks in (a).

emphasized in the inset of Fig. 2(a), where the values of ϕ_c predicted in Eq. (4) are indicated by the arrows.

The threshold ϕ_c depends dominantly on the average degree $\langle k\rangle$ and the variance of the degree distribution σ^2. The transition occurs earlier in denser and more heterogeneous networks, as illustrated in Fig. 2(b). This diagram provides a close approximation to the positions of the transitions for the power grids shown in Fig. 2(a), even though it was generated using Gamma degree distributions, which deviate from the approximately exponential distributions of the power-grid networks. This occurs because, even though ϕ_c can in principle depend on higher moments of the degree distribution through the term $G_1'(1-\phi)$, this dependence is weak for systems with small ϕ_c. We can show that for any degree distribution ϕ_c is upper bounded by a function ϕ_B of $G_1'(1) = \frac{\langle k^2\rangle - \langle k\rangle}{\langle k\rangle}$ that approaches zero rapidly as $G_1'(1)$ increases and, for fixed $G_1'(1)$, is lower bounded by a function ϕ_b of $\langle k^3\rangle/\langle k\rangle$ that decreases as this ratio increases (see the Supplemental Material [16]).

These results provide insights relevant for real systems, but also point to other practical considerations concerning the observability of (necessarily finite-size and structured) real power-grid networks. For instance, what is the minimum number (and corresponding optimal placement) of PMUs needed for complete observability of an entire network?

This optimization question can be formulated as a binary integer programming problem [4], which is nevertheless NP-complete and hence not solvable in large networks. Metaheuristic optimization methods can be relatively efficient [17], but the reliability of the solutions remains to be demonstrated. Greedy algorithms [18], on the other hand, are effective but provide only conservative estimates. A common feature of these approaches is that they do not take advantage of the internal organization of real power grids. To proceed, we introduce an approach that is both efficient and effective. Specifically, we use modularity maximization [19] to split the network into communities, so that the placement problem within each community can be solved using binary integer programming. We solve the placement problem within one community, then we update the set of observable nodes on the whole network and move to the neighboring community most connected with the previously solved communities, and so on. This procedure is repeated by starting from each of the communities; we select the minimum-PMU solution, although for the systems considered here we verified that the community sequence has very small impact on the number of PMUs (e.g., relative standard deviation $< 2\times 10^{-4}$ for the Eastern North America power grid).

As shown in Table I, benchmarking of this approach using small networks that can be solved exactly shows that it offers very good approximations of the optimal solutions. As a comparison, the application of the greedy algorithm maximizing at each step the increase in the fraction of observable nodes (FON) results in a solution that requires 1000 additional PMUs in the Eastern North America power grid. For both this system and the PJM (Pennsylvania–New Jersey–Maryland) power grid, our approach shows that the resulting minimum number of PMUs for complete observability corresponds to approximately 30% of the nodes in the network (Table I), which is

TABLE I. Optimal PMU placement based on community splitting, where N is the number of nodes, $\langle k\rangle$ is the average degree, Q is the modularity, N_C is the number of communities, N_P is the minimum number of PMUs estimated from the community structure, N_P^{opt} is the exact minimum number (only computable for the small IEEE test systems), and N_P^g is the greedy optimal solution.

Power grid	N	$\langle k\rangle$	Q	N_C	N_P	N_P^{opt}	N_P^g
IEEE118	118	3.16	0.72	8	32	32	36
IEEE300	300	2.73	0.83	14	89	87	96
PJM	14 077	2.60	0.95	52	4246		4493
Eastern	56 892	2.52	0.97	96	17 216		18 216

comparable to previous estimates and exact calculations on small networks available in the literature [20].

Interestingly, an abrupt (albeit smooth) transition of the LOC size also occurs for real networks and even if we limit the random PMU placement to the optimal set (i.e., the solution set of the optimal placement problem), as illustrated in Fig. 3(a) (continuous red line). The FON, in contrast, grows approximately linearly as the fraction of directly observed nodes increases from zero (dot-dashed red line). However, we can cause both the LOC size and the FON to grow sharply from the beginning by changing the placement sequence [Fig. 3(a), green and blue lines, respectively]. Using optimal PMU placement on the 2010 Eastern North America power grid and data on the planned upgrades of the network until 2020 [21], we also demonstrate that both the LOC size and the FON are rather robust against the evolution of the network [Fig. 3(b)]. Even after nearly 10% of the nodes have been removed, added, or rewired, neither the LOC size nor the FON decreases (and they in fact increase) relative to the number of nodes in the initial network. (See the Supplemental Material [16] for an analogous conclusion when considering the impact of random edge rewiring). This suggests that an initially optimal (hence, minimally redundant) placement of PMUs remains effective as the network evolves.

However, this does not mean that the network is robust against intentional "observability attacks," which we define as the deliberate disabling of PMUs (rather than of power-grid nodes themselves). In fact, while the FON remains large upon a sequential inactivation of PMUs that maximizes reduction of the FON at each step, the LOC size decreases rapidly [Fig. 3(c), blue lines]. Moreover, this decrease is significantly faster if we attack the LOC by targeting intercommunity PMUs, effectively breaking the LOC into observable islands defined by the network community structure [Fig. 3(c), green lines]. Ironically, the same network property that facilitates identification of optimal PMU placement—community structure—makes the network vulnerable to observability attacks.

We suggest that similar analysis can also be useful for other networked systems, such as traffic monitoring in diverse networks and network discovery. For example, in content-based network crawling, the initial nodes in the crawling problem play the role of directly observable nodes, and the emergence of a LOC indicates that a fraction of the nodes will be visited from multiple initial nodes. These problems invoke the notion of depth-L observability, in which the direct observation of a node can make all neighbors within distance L indirectly observable. While here we have focused on depth-1 observability, which is the most relevant for power-grid networks, our analysis can be extended to higher observability depths (see the Supplemental Material [16] for a depth-2 example). These concepts can also be extended to systems in which observability depends on additional network structural properties, as in the case of metabolic networks (see the Supplemental Material [16]). Therefore, like other percolation processes studied previously [8,9,12] and recently [22–24] on networks, network observability transitions have implications for a wide range of systems.

Network observability is challenging in part because networks represent distributed dynamical systems, whose states cannot be assessed from single measurement points. However, randomness, long-range connections, and the consequent small node-to-node distances common to many real networks facilitate observability as they significantly reduce the necessary number of directly observable nodes. This underlies the finite but surprisingly small threshold for the observability transitions identified here even for fairly sparse and homogeneous networks. In infrastructure networks, wide-area observability and monitoring is necessary for modernization of the systems' operation [20]. Yet, reliance on observability comes at the risk of making the network vulnerable to a new form of attack, in which the deliberate disabling of a relatively small number of sensors may render the network

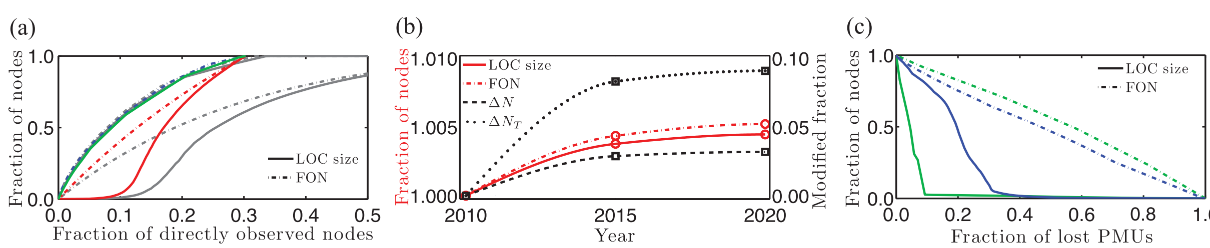

FIG. 3 (color). Observability on the largest available power grid (Eastern North America). (a) Complete and incomplete observability: LOC size and FON for random PMU placement (red), greedy LOC size optimization (green), and greedy FON optimization (blue) on the optimal set. The corresponding curves for placement on the full network are shown in gray. (b) Network evolution: LOC size and FON on the planned networks for the years 2015 and 2020 given the optimal PMU placement on the 2010 network (red). The black lines represent the net increase in the number of nodes (ΔN) and the total number of nodes modified by node additions, node removals, or edge rewires (ΔN_T). To facilitate comparison, all curves are plotted relative to the initial number of nodes. (c) Observability attack: LOC size and FON for both FON attack (blue) and LOC size attack (green), where the latter takes advantage of the community structure of the network.

unobservable, hence potentially nonoperational, even when it is robust against conventional attacks [25].

The authors thank Cong Liu for providing data and Yu Cheng for inputs on the data processing. This work was supported by a Northwestern-Argonne Early Career Investigator Award for Energy Research.

[1] A. Monticelli and F. F. Wu, IEEE Transactions on Power Apparatus and Systems **PAS-104**, 1035 (1985).
[2] For a related recent review, see A. E. Motter and R. Albert, Phys. Today **65**, No. 4, 43 (2012).
[3] A. G. Phadke and J. S. Thorp, *Synchronized Phasor Measurements and Their Applications* (Springer, Berlin, 2010).
[4] B. Xu and A. Abur, in *Proceedings of the IEEE PES Power Systems Conference and Exposition, 2004* (IEEE, New York, 2004), Vol. 2, p. 943.
[5] V. C. Güngör, D. Sahin, T. Kocak, S. Ergüt, C. Buccella, C. Cecati, and G. P. Hancke, IEEE Trans. Inf. Theory **7**, 529 (2011).
[6] C. W. Gellings and K. E. Yeager, Phys. Today **57**, No. 12, 45 (2004).
[7] U.S.-Canada Power System Outage Task Force, Final Report on the August 14, 2003 Blackout in the United States and Canada, 2004.
[8] For a review on percolation transitions, see S. N. Dorogovtsev, A. V. Goltsev, and J. F. F. Mendes, Rev. Mod. Phys. **80**, 1275 (2008).
[9] M. E. J. Newman, *Networks: An Introduction* (Oxford University Press, New York, 2010).
[10] R. Albert, I. Albert, and G. L. Nakarado, Phys. Rev. E **69**, 025103 (2004).
[11] R. V. Solé, M. Rosas-Casals, B. Corominas-Murtra, and S. Valverde, Phys. Rev. E **77**, 026102 (2008).
[12] R. Cohen and S. Havlin, *Complex Networks: Structure, Robustness and Function* (Cambridge University Press, Cambridge, England, 2010).
[13] We do not discriminate zero-injection nodes as this makes the results more broadly applicable.
[14] M. E. J. Newman, S. H. Strogatz, and D. J. Watts, Phys. Rev. E **64**, 026118 (2001).
[15] D. S. Callaway, M. E. J. Newman, S. H. Strogatz, and D. J. Watts, Phys. Rev. Lett. **85**, 5468 (2000).
[16] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevLett.109.258701 for further analysis of the observability transition threshold, observability robustness to edge rewiring, depth-2 observability, and observability in metabolic networks.
[17] F. Aminifar, C. Lucas, A. Khodaei, and M. Fotuhi-Firuzabad, IEEE Transactions on Power Delivery **24**, 1014 (2009).
[18] M. Zhou *et al.*, in *Proceedings of the 3rd International Conference on Electric Utility Deregulation and Restructuring and Power Technologies, 2008* (IEEE, Nanjing, China, 2008), p. 2862.
[19] A. Clauset, M. E. J. Newman, and C. Moore, Phys. Rev. E **70**, 066111 (2004).
[20] V. Terzija, G. Valverde, D. Cai, P. Regulski, V. Madani, J. Fitch, S. Skok, M. M. Begovic, and A. Phadke, Proc. IEEE **99**, 80 (2011).
[21] Annual Transmission Planning and Evaluation Report, Federal Energy Regulatory Commission Form 715, 2010.
[22] D. Achlioptas, R. M. D'Souza, and J. Spencer, Science **323**, 1453 (2009).
[23] R. A. da Costa, S. N. Dorogovtsev, A. V. Goltsev, and J. F. F. Mendes, Phys. Rev. Lett. **105**, 255701 (2010).
[24] P. Grassberger, C. Christensen, G. Bizhani, S.-W. Son, and M. Paczuski, Phys. Rev. Lett. **106**, 225701 (2011).
[25] A. Annibale, A. C. C. Coolen, and G. Bianconi, J. Phys. A **43**, 395001 (2010).