Observability transition in real networks

Yang Yang[1] and Filippo Radicchi[2,*]

[1]*Department of Physics and Astronomy, Northwestern University, Evanston, Illinois 60208, USA*

[2]*Center for Complex Networks and Systems Research, School of Informatics and Computing, Indiana University, Bloomington, Indiana 47408, USA*

We consider the observability model in networks with arbitrary topologies. We introduce a system of coupled nonlinear equations, valid under the locally treelike ansatz, to describe the size of the largest observable cluster as a function of the fraction of directly observable nodes present in the network. We perform a systematic analysis on 95 real-world graphs and compare our theoretical predictions with numerical simulations of the observability model. Our method provides almost perfect predictions in the majority of the cases, even for networks with very large values of the clustering coefficient. Potential applications of our theory include the development of efficient and scalable algorithms for real-time surveillance of social networks, and monitoring of technological networks.

The state of an entire networked dynamical system can be determined by monitoring or dominating the states of a limited number of nodes in the network [1]. A power-grid network can be observed in real time by placing phasor measurement units to a selection of nodes in the network [2]. Routing tables in mobile *ad hoc* networks rely on gateway nodes to form connected dominating sets used as backbones for communication [3]. Disease outbreaks in urban environments can be efficiently detected by placing sensors on specific locations visited by potentially infected individuals [4].

Whereas all these examples markedly differ in their underlying dynamics, from the structural point of view, they can all be framed in terms of the so-called network observability model [2]. In this model, placing an observer on one node can make the node itself and all its nearest neighbors observable. Nodes in the network can therefore assume three different states: (i) directly observable, if hosting an observer; (ii) indirectly observable, if being the first neighbor of an observer; or (iii) not observable, otherwise. Observable, either directly or indirectly, nearest-neighbor nodes form clusters of connected observable nodes. Thus, structurally speaking, the network observability model can be thought as an extension of the more traditional, and much more studied, percolation model [2,5]. As in percolation, the question of interest in network observability is how to determine the macroscopic formation of observable clusters in the network on the basis of microscopic changes in the state of its individual nodes.

The observability model has been recently studied in its simplest formulation where directly observed nodes are randomly selected [2]. The model has been solved for both uncorrelated and correlated random network models in the limit of infinite size [2,6]. As real networks are not mere realizations of random network models, and their size is clearly not infinite, the methods deployed in Refs. [2,6] are not directly applicable to real-world networks. The present Rapid Communication introduces a theoretical approach able to describe the observability model in real graphs. We introduce a set of heuristic equations that takes as input the adjacency matrix of a network to draw its entire observability phase diagram. The mathematical framework consists in the formulation of a belief-propagation or message-passing algorithm [7] in a similar spirit as recent theoretical methods based on message-passing algorithms have been used to describe ordinary percolation transitions in real isolated and/or interdependent networks [8–12]. We show, through a systematic analysis of nearly 100 real networks, that the method is able to reproduce true phase diagrams with extraordinary accuracy, proving therefore its applicability to a wide range of real systems.

Here we consider an arbitrary network composed of N nodes and E edges. Without loss of generality, we assume that the network has one single connected component. Suppose that each node has a probability ϕ to host an observer, i.e., to be directly observable. Nodes that are connected to directly observable nodes are, in turn, indirectly observable. Observable nearest-neighbor nodes form clusters. For $\phi = 0$, no nodes are observable, hence there are no clusters. For $\phi = 1$, all nodes are directly observable, and thus they form a single cluster. At intermediate values of ϕ, the network can be found in two different phases: (i) the regime of nonobservability, where all clusters have microscopic size; and (ii) the phase of observability, where a single macroscopic cluster, comparable in size with the entire network, is present. To monitor the transition between these two phases, one usually relies on the order parameter P_∞, corresponding to the relative size of the largest observable cluster (LOC). In the limit of infinitely large networks, $P_\infty = 0$ for $\phi \leqslant \phi_c$, and $P_\infty > 0$ for $\phi > \phi_c$, with ϕ_c critical value of the probability ϕ. In the following, we describe a mathematical framework, deployed under the locally treelike approximation, to estimate the relative size of the LOC as a function of ϕ.

To proceed, we consider the probability that moving along the edge $i \rightarrow j$, we arrive to the LOC, irrespective of whether node i is in the LOC or not.[1] In particular, we consider three conditional versions of this probability. We denote them

---

*filiradi@indiana.edu

---

[1]Please note that the network is undirected, but, in our mathematical framework, every edge (i,j) is considered twice, as $i \rightarrow j$ and $j \rightarrow i$.
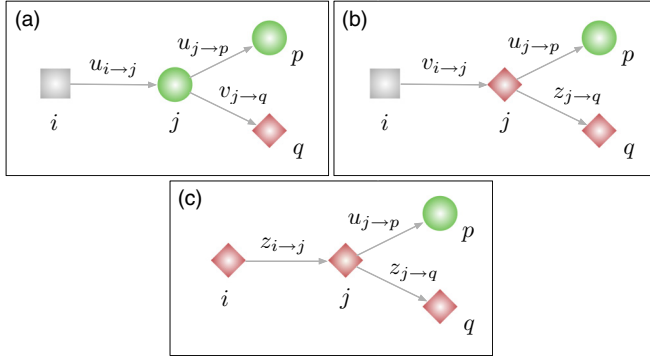
FIG. 1. Schematic illustration of the derivation of the system of Eqs. (1) [panel (a)], (2) [panel (b)], and (3) [panel (c)]. The different variables used in the equations are defined depending on the state of the nodes, here denoted by different shapes and colors (green circle = directly observable, red diamond = not directly observable, and gray square = arbitrary).

as $u_{i \to j}$ if j is directly observable, as $v_{i \to j}$ if j is not directly observable, and as $z_{i \to j}$ if neither i nor j are directly observable. Working under the locally treelike ansatz, we can write the following system of coupled equations (Fig. 1):

$$u_{i \to j} = 1 - \prod_{q \in \mathcal{N}_j \setminus \{i\}} [1 - \phi u_{j \to q} - (1 - \phi) v_{j \to q}], \quad (1)$$

$$v_{i \to j} = 1 - \prod_{q \in \mathcal{N}_j \setminus \{i\}} [1 - \phi u_{j \to q} - (1 - \phi) z_{j \to q}], \quad (2)$$

and

$$z_{i \to j} = v_{i \to j} - (1 - \phi)^{k_j - 1} \left[1 - \prod_{q \in \mathcal{N}_j \setminus \{i\}} (1 - z_{j \to q}) \right]. \quad (3)$$

In the above equations, $\mathcal{N}_j$ is the set of all neighbors of node j, and k_j is the degree of node j. We note that $k_j = |\mathcal{N}_j|$, where $|\mathcal{X}|$ indicates the size (i.e., number of elements) of the set $\mathcal{X}$. Equation (1) is derived as follows. If node j is directly observable, then node j is part of the LOC if at least one of its neighbors $q \neq i$ is part of the largest cluster. This fact can happen in two ways: (i) with probability $\phi u_{j \to q}$, if node q is directly observable; and (ii) with probability $(1 - \phi) v_{j \to q}$, if node q is not directly observable. Thus, the probability that the connection $j \to q$ brings to the LOC is $\phi u_{j \to q} + (1 - \phi) v_{j \to q}$. The right-hand side (rhs) of Eq. (1) quantifies the probability that at least one of the connections $j \to q$ leads to the LOC, where the treelike ansatz allows us to consider probabilities associated with the individual edges as independent variables, hence their product appearing on the rhs of Eq. (1).

The derivation of Eq. (2) is similar to the one just described for Eq. (1). We note that we can write

$$v_{i \to j} = 1 - \sum_{\{s_r\}, r \in \mathcal{N}_j \setminus \{i\}} \prod_{q \in \mathcal{N}_j \setminus \{i\}}$$
$$\times [\phi(1 - u_{j \to q})]^{s_q} [(1 - \phi)(1 - z_{j \to q})]^{1 - s_q}. \quad (4)$$

For a proof of the equivalence between Eqs. (2) and (4), see Appendix A. The sum on the rhs of Eq. (4) runs over all $2^{k_j - 1}$

possible configurations $\{s_r\}$ for the state (that is directly or not directly observable) of the neighbors of node j, excluding node i. For every given configuration, the product appearing inside the sum is the probability that such a configuration appears, multiplied by the conditional probability that node j is not attached to the LOC in this configuration. To be more specific, the binary variable $s_q = 1$, if node q is directly observable, and $s_q = 0$, otherwise. The quantity $[\phi(1 - u_{j \to q})]^{s_q} [(1 - \phi)(1 - z_{j \to q})]^{1 - s_q}$ is the probability that the connection $j \to q$ does not bring node j to the LOC. Depending on whether node q is directly observable or not, this probability is either $\phi(1 - u_{j \to q})$ or $(1 - \phi)(1 - z_{j \to q})$, respectively.

The expression of $z_{i \to j}$ in Eq. (3) can be quantified in almost the same way as $v_{i \to j}$. We still need to consider the probabilities that the connection $i \to j$ does not bring node i to LOC, for all possible configurations of neighbors of node j. The probability associated with each configuration is the same as that appearing in Eq. (4). The only exception is the configuration $s_q = 0$, $\forall q \in \mathcal{N}_j \setminus \{i\}$, which happens with probability $(1 - \phi)^{k_j - 1}$, where all neighbors of node j are not directly observable (thanks to the underlying assumption that node i is not directly observable when we consider the conditional probability $z_{i \to j}$), hence node j is surely not observable and cannot be part of the LOC. Accounting for this exception, and using the equivalence between Eqs. (2) and (4), we finally derive Eq. (3).

We can now rely on Eqs. (1), (2), and (3) to compute the probability p_i that node i is part of the LOC. We start from the simpler case when node i is directly observable, which happens with probability ϕ. We consider the probability that the connection $i \to j$ brings node i to the LOC. This probability is $u_{i \to j}$, if node j is directly observable, and is $v_{i \to j}$, if node j is not directly observable. Combining the contributions from all neighbors of node i, and using again the locally treelike ansatz, the probability r_i that node i is directly observable, but not part of the LOC is

$$r_i = \phi \prod_{j \in \mathcal{N}_i} [1 - \phi u_{i \to j} - (1 - \phi) v_{i \to j}]. \quad (5)$$

If node i is not directly observable, which happens with probability $1 - \phi$, it is better to recast the approach used to compute Eq. (3). We need to consider all possible 2^{k_i} configurations for the neighbors of node i. Again, we have to account for the special configuration $s_j = 0$, $\forall j \in \mathcal{N}_i$, when node i is surely not observable. The probability t_i that node i is not directly observable, and none of its neighbors is attached to the LOC is given by

$$t_i = (1 - \phi) \left\{ \sum_{\{s_r\}, r \in \mathcal{N}_i} \prod_{j \in \mathcal{N}_i} [\phi(1 - u_{i \to j})]^{s_j} \right.$$
$$\times [(1 - \phi)(1 - z_{i \to j})]^{1 - s_j}$$
$$\left. + (1 - \phi)^{k_i} - (1 - \phi)^{k_i} \prod_{j \in \mathcal{N}_i} (1 - z_{i \to j}) \right\}.$$
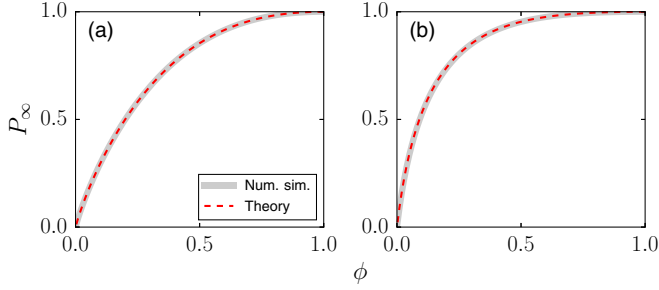
FIG. 2. Observability transition in real networks. We compare results from numerical simulations (gray lines) with the solution of our theoretical equations (red lines). (a) Analysis of the Internet at the autonomous system level, as of July 22, 2006 [9]. (b) Analysis of the scientific collaboration network derived from preprints posted in the section cond-mat of the arXiv between years 1993 and 2005 [14].

Using the same trick as the one considered to pass from Eq. (4) to Eq. (2), we rewrite t_i as

$$t_i = (1 - \phi) \left\{ \prod_{j \in \mathcal{N}_i} [1 - \phi u_{i \to j} - (1 - \phi) z_{i \to j}] \right.$$
$$\left. + (1 - \phi)^{k_i} [1 - \prod_{j \in \mathcal{N}_i} (1 - z_{i \to j})] \right\}. \quad (6)$$

Combining the two cases, we derive the probability p_i that node i is part of the LOC as

$$p_i = 1 - \phi \prod_{j \in \mathcal{N}_i} [1 - \phi u_{i \to j} - (1 - \phi) v_{i \to j}]$$
$$- (1 - \phi) \left\{ \prod_{j \in \mathcal{N}_i} [1 - \phi u_{i \to j} - (1 - \phi) z_{i \to j}] \right.$$
$$\left. + (1 - \phi)^{k_i} [1 - \prod_{j \in \mathcal{N}_i} (1 - z_{i \to j})] \right\}. \quad (7)$$

The relative size of the LOC, predicted in the locally treelike ansatz, can be finally calculated as

$$P_\infty^{(\text{th})} = \frac{1}{N} \sum_{i=1}^N p_i. \quad (8)$$

For every value of ϕ, $P_\infty^{(\text{th})}$ can be numerically estimated by first solving by iteration the system of Eqs. (1), (2), and (3) for every directed edge $i \to j$. We can then plug the solution in the system of Eqs. (7), and estimate every p_i. These values can be finally inserted into Eq. (8) to compute $P_\infty^{(\text{th})}$.

In Fig. 2, we present results from the analysis of two real-world networks. The plots show a comparison of the observability phase diagram obtained from the solution of our framework, and the one computed from numerical simulations of the model. Simulations are performed using a modified version of the Newman-Ziff algorithm, originally introduced to simulate ordinary percolation processes in arbitrary topologies [13]. For every value of ϕ, we estimate the order parameter
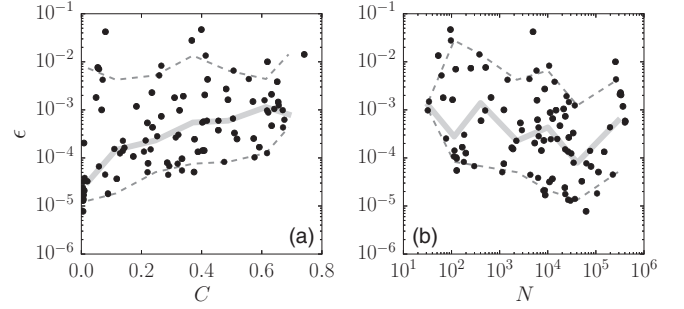

FIG. 3. Analysis of real networks. We consider 95 real-world graphs [15,16]. For every network, we compute the discrepancy ϵ between the theoretical and numerical estimates of the relative size of the LOC [Eq. (9)]. (a) For every network, we plot ϵ as a function of the average clustering coefficient C. To construct the lines, we consider seven equally spaced bins for the range of C values. For all networks falling in a given bin, we compute the median value of ϵ (full line), and the lower and upper ends of 90% confidence intervals (dashed lines). (b) Scatter plot of ϵ versus the network size N. Lines are constructed in a similar way as those appearing in panel (a). The only difference is that we divide the range of N values into six equally spaced bins on the logarithmic scale.

$P_\infty^{(\text{num})}$ as the average value over 10 000 independent realizations of the algorithm. The analysis of Fig. 2 reveals an almost perfect match between theoretical predictions and results of numerical simulations.

To test how good our theoretical predictions are, we perform a systematic comparison between theory and numerical simulations on 95 real-world graphs [15]. For a list of all networks analyzed see [16]. We consider networks of very different nature (e.g., technological, social, biological), and heterogeneous in terms of their topological properties (e.g., clustering coefficient, size, degree distribution). To quantify the discrepancy between theoretical predictions and the ground truth offered by numerical simulations, we use the following expression [17]:

$$\epsilon = \int_0^1 |P_\infty^{(\text{th})}(\phi) - P_\infty^{(\text{num})}(\phi)| d\phi. \quad (9)$$

As the results of our analysis reveal, the discrepancy between theory and numerical simulations is generally very small (Fig. 3). Besides, we observe only a very weak dependence of ϵ on the average clustering coefficient of the network C [Fig. 3(a)]. Because the theoretical framework is deployed under the locally treelike ansatz, and C can be interpreted as a good proxy for the degree of violation of this approximation, a positive correlation between C and ϵ is expected. However, the error committed by our framework to estimate the true observability diagram is very small even for networks with extremely large values of the clustering coefficient. This result is in stark contrast with what was found for the ordinary site percolation model, where high clustering implies large differences between ground truth and approaches based on the locally treelike approximation [11]. We further note that even for extremely small networks composed of tens of nodes, theoretical predictions are very accurate. Moreover, the discrepancy between theory and simulations tends to decrease

as the size of the network increases [Fig. 3(b)]. This is also not a surprising result, given that our theoretical framework is expected to become exact in the limit of (locally treelike) infinite networks.

Given the continuous nature of the observability phase transition, in the vicinity of the critical point ϕ_c, we can take a linear approximation of the system of Eqs. (1), (2), and (3), and rewrite them in matricial form as $\vec{u} = M[\phi\vec{u} + (1 - \phi)\vec{v}]$, $\vec{v} = M[\phi\vec{u} + (1 - \phi)\vec{z}]$, and $\vec{z} = \vec{v} - R^{(\phi)}\vec{z}$. In the above expressions, $\vec{u}$, $\vec{v}$, and $\vec{z}$ are column vectors composed of $2E$ components, each corresponding to a directed edge of the graph. Matrix M is the $2E \times 2E$ nonbacktracking matrix of the graph, whose generic element is defined as $M_{i\to j,\ell\to r} = \delta_{j,\ell}(1 - \delta_{i,r})$, with δ Kronecker symbol [18,19]. The generic element of the matrix $R^{(\phi)}$ is defined as $R^{(\phi)}_{i\to j,\ell\to r} = (1 - \phi)^{k_j - 1} M_{i\to j,\ell\to r}$. Solving the previous system of linear equations (see Appendix B), we arrive at the eigenvalue/eigenvector equation

$$\vec{z} = \{[\mathbb{1} - M\phi(1 - \phi)(\mathbb{1} - \phi M)^{-1}M]^{-1}$$
$$\times (1 - \phi)M - R^{(\phi)}\}\vec{z}. \tag{10}$$

Equation (10) serves to study the linear stability of the trivial solution $\vec{z}^T = (0, \ldots, 0)$. The critical value ϕ_c of the transition equals the value of ϕ for which the trivial solution becomes unstable, and corresponds to the ϕ value for which the operator appearing on the rhs of Eq. (10) has a principal eigenvalue equal to 1. Equation (10) is useful only in a limited number of cases, as, for example, regular graphs (see Appendix B). For general networks instead, solving Eq. (10) is not computationally efficient. This operation requires one to determine the inverse of several matrices. From a numerical point of view, it is thus better to rely on a binary search combined with the numerical solution of the system of nonlinear Eqs. (1), (2), and (3). We further stress that the determination of the critical point in the observability transition is not as meaningful as in the case of percolation. The critical point ϕ_c is in fact very close to zero for almost all networks. Thus, the emergence of the LOC happens as soon as a very small number of observers are randomly placed in the system.

Our method to estimate observability phase diagrams is the first theoretical framework that can be applied to arbitrary network topologies. Although the method is exact only for locally treelike infinite networks, its performances are almost perfect regardless of the size and/or the average clustering coefficient of the network. In this Rapid Communication, we considered and solved the ordinary version of the observability model, where observers are randomly placed on nodes of the network. We believe, however, that the framework has the potential to be generalized to arbitrary strategies for the placement of observers. In this sense, a very important extension of our formalism could be to study optimal observability, on the same footing as recent work on optimal percolation [20–22]. Considering that the optimal solution of the observability model is formally equivalent to the minimum (partial) dominating set of a graph [23], such an extension could represent a very important contribution for research in several domains, including, among others, biology [24,25] and social sciences [26,27].

APPENDIX A: PROOF BY INDUCTION OF EQUATION (4)

In the main text, we used the fact that

$$\prod_{q\in\mathcal{Q}}[1 - \phi a_q - (1 - \phi)b_q]$$
$$= \sum_{\{s_r\}, r\in\mathcal{Q}}\prod_{q\in\mathcal{Q}}[\phi(1 - a_q)]^{s_q}[(1 - \phi)(1 - b_q)]^{1-s_q}. \tag{A1}$$

Note that Eq. (A1) is a more general version of Eq. (4). We recall that the sum on the rhs runs over all $2^{|\mathcal{Q}|}$ configurations, with $|\mathcal{Q}|$ number of elements in $\mathcal{Q}$, where the element q in the set $\mathcal{Q}$ can be in an active state, i.e., $s_q = 1$, or in an inactive state, i.e., $s_q = 0$. These events happen with probability ϕ and $1 - \phi$, respectively, providing the proper way to weight the probability of appearance of every configuration. We provide here a proof by induction of Eq. (A1). To this end, we first note that if $\mathcal{Q} = \emptyset$, then Eq. (A1) is automatically satisfied, being both sides equal to 1. If $|\mathcal{Q}| > 0$, we hypothesize that

$$\prod_{q\in\mathcal{Q}\setminus\{p\}}[1 - \phi a_q - (1 - \phi)b_q]$$
$$= \sum_{\{s_r\}, r\in\mathcal{Q}\setminus\{p\}}\prod_{q\in\mathcal{Q}\setminus\{p\}}[\phi(1 - a_q)]^{s_q}[(1 - \phi)(1 - b_q)]^{1-s_q}. \tag{A2}$$

The latter equation is the analog of Eq. (A1) for the set $\mathcal{Q}\setminus\{p\}$. We are thus supposing that the equation is valid not for the entire set $\mathcal{Q}$, but the set minus one its elements. If we factorize out the contribution of the element p in Eq. (A1), we have

$$[1 - \phi a_p - (1 - \phi)b_p]\prod_{q\in\mathcal{Q}\setminus\{p\}}[1 - \phi a_q - (1 - \phi)b_q]$$
$$= [\phi(1 - a_p) + (1 - \phi)(1 - b_p)]\sum_{\{s_r\}, r\in\mathcal{Q}\setminus\{p\}}$$
$$\times \prod_{q\in\mathcal{Q}\setminus\{p\}}[\phi(1 - a_q)]^{s_q}[(1 - \phi)(1 - b_q)]^{1-s_q}.$$

By virtue of the hypothesis of Eq. (A2), the validity of Eq. (A1) is obtained by proving that the two extra factors due to the element p that appear on both sides of the previous equation are equal. This fact can be trivially shown by rewriting

$$\phi(1 - a_p) + (1 - \phi)(1 - b_p) = 1 - \phi a_p - (1 - \phi)b_p.$$

APPENDIX B: LINEAR APPROXIMATION

Using the linear approximation

$$\prod_q (1 - x_q) \simeq 1 - \sum_q x_q,$$

we can rewrite Eqs. (1), (2), and (3) of the main text,

respectively, as

$$\vec{u} = M[\phi\vec{u} + (1-\phi)\vec{v}], \tag{B1}$$

$$\vec{v} = M[\phi\vec{u} + (1-\phi)\vec{z}], \tag{B2}$$

and

$$\vec{z} = \vec{v} - R^{(\phi)}\vec{z}, \tag{B3}$$

where $R^{(\phi)}_{i \to j, \ell \to r} = (1-\phi)^{k_j-1} M_{i \to j, \ell \to r}$, and M is the non-backtracking matrix of the graph.

From Eq. (B1), we obtain

$$\vec{u} = (1-\phi)(\mathbb{1} - \phi M)^{-1}M\vec{v}.$$

Inserting this expression into Eq. (B2), we have

$$\vec{v} = [\mathbb{1} - M\phi(1-\phi)(\mathbb{1} - \phi M)^{-1}M]^{-1}(1-\phi)M\vec{z}.$$

Finally, using this expression in Eq. (B3), we obtain Eq. (10).

A special case where Eq. (10) can be simplified is for regular graphs with valency k, so that M and $R^{(\phi)}$ have the same eigenvectors. We can write the condition for the critical probability ϕ_c as

$$[1 - \phi_c(1-\phi_c)(1-\phi_c\mu)^{-1}\mu^2]^{-1}(1-\phi_c)\mu - (1-\phi_c)^{k-1}\mu = 1,$$

with $\mu = k - 1$.

---

[1] Y.-Y. Liu, J.-J. Slotine, and A.-L. Barabási, Proc. Natl. Acad. Sci. USA **110**, 2460 (2013).

[2] Y. Yang, J. Wang, and A. E. Motter, Phys. Rev. Lett. **109**, 258701 (2012).

[3] J. Wu and H. Li, in *Proceedings of the 3rd International Workshop on Discrete Algorithms and Methods for Mobile Computing and Communications* (ACM, Seattle, 1999), pp. 7–14.

[4] S. Eubank, H. Guclu, V. A. Kumar, M. V. Marathe, A. Srinivasan, Z. Toroczkai, and N. Wang, Nature (London) **429**, 180 (2004).

[5] D. Stauffer and A. Aharony, *Introduction to Percolation Theory* (Taylor & Francis, London, 1991).

[6] T. Hasegawa, T. Takaguchi, and N. Masuda, Phys. Rev. E **88**, 042809 (2013).

[7] M. Mézard and A. Montanari, *Information, Physics, and Computation* (Oxford University Press, New York, 2009).

[8] K. E. Hamilton and L. P. Pryadko, Phys. Rev. Lett. **113**, 208701 (2014).

[9] B. Karrer, M. E. J. Newman, and L. Zdeborová, Phys. Rev. Lett. **113**, 208702 (2014).

[10] F. Radicchi and C. Castellano, Nat. Commun. **6**, 10196 (2015).

[11] F. Radicchi and C. Castellano, Phys. Rev. E **93**, 030302 (2016).

[12] F. Radicchi, Nat. Phys. **11**, 597 (2015).

[13] M. E. J. Newman and R. M. Ziff, Phys. Rev. Lett. **85**, 4104 (2000).

[14] M. E. Newman, Proc. Natl. Acad. Sci. USA **98**, 404 (2001).

[15] F. Radicchi, Phys. Rev. E **91**, 010801 (2015).

[16] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevE.94.030301 for the list of all real-world networks included in our analysis.

[17] S. Melnik, A. Hackett, M. A. Porter, P. J. Mucha, and J. P. Gleeson, Phys. Rev. E **83**, 036112 (2011).

[18] K.-i. Hashimoto, *Automorphic Forms and Geometry of Arithmetic Varieties* (Kinokuniya Company Ltd., Tokyo, Japan, 1989), p. 211.

[19] F. Krzakala, C. Moore, E. Mossel, J. Neeman, A. Sly, L. Zdeborová, and P. Zhang, Proc. Natl. Acad. Sci. USA **110**, 20935 (2013).

[20] F. Morone and H. A. Makse, Nature (London) **524**, 65 (2015).

[21] A. Braunstein, L. Dall'Asta, G. Semerjian, and L. Zdeborová, arXiv:1603.08883.

[22] P. Clusella, P. Grassberger, F. J. Perez-Reche, and A. Politi, arXiv:1604.00073.

[23] D.-Z. Du and P.-J. Wan, *Connected Dominating Set: Theory and Applications* (Springer Science & Business Media, New York, 2012), Vol. 77.

[24] S. Wuchty, Proc. Natl. Acad. Sci. USA **111**, 7156 (2014).

[25] T. Milenković, V. Memišević, A. Bonato, and N. Pržulj, PloS One **6**, e23016 (2011).

[26] L. L. Kelleher and M. B. Cozzens, Math. Soc. Sci. **16**, 267 (1988).

[27] F. Wang, H. Du, E. Camacho, K. Xu, W. Lee, Y. Shi, and S. Shan, Theor. Comput. Sci. **412**, 265 (2011).