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 tree-like 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\(^2\). 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; (iii) or 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 paper 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\(^8\)\(^9\).\(^10\). We show, through a systematic analysis of nearly one hundred 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 \(\phi\) 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 \(\phi\), the network can be found in two different phases: (i) the regime of non-observability, where all clusters have microscopic size; (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 \(P_\infty\) corresponding to the relative size of the largest observable cluster (LOC). In the limit of infinitely large networks, \(P_\infty = 0\), for \(\phi \leq \phi_c\), and \(P_\infty > 0\), for \(\phi > \phi_c\), with \(\phi_c\) critical value of the probability \(\phi\). In the following, we describe a mathematical framework, deployed under the locally tree-like approximation, to estimate the relative size of the LOC as a function of \(\phi\).

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. In particular, we consider three conditional versions of this probability. We denote them as \( u_{i \rightarrow j} \) if \( j \) is directly observable, as \( v_{i \rightarrow j} \) if \( j \) is not directly observable, and as \( z_{i \rightarrow j} \) if neither \( i \) nor \( j \) are directly observable. Working under the locally tree-like ansatz, we can write the following system of coupled equations (Fig. 1):

\[
u_{i \rightarrow j} = 1 - \prod_{q \in N_j \setminus \{i\}} [1 - \phi u_{j \rightarrow q} - (1 - \phi) v_{j \rightarrow q}], \quad (1)
\]

\[
v_{i \rightarrow j} = 1 - \prod_{q \in N_j \setminus \{i\}} [1 - \phi u_{j \rightarrow q} - (1 - \phi) z_{j \rightarrow q}], \quad (2)
\]

and

\[
z_{i \rightarrow j} = v_{i \rightarrow j} - (1 - \phi)^{k_j - 1} [1 - \prod_{q \in N_j \setminus \{i\}} (1 - z_{j \rightarrow q})]. \quad (3)
\]

In the above equations, \( N_j \) is the set of all neighbors of node \( j \), and \( k_j \) is the degree of node \( j \). We note that \( k_j = |N_j| \), where \( |X| \) indicates the size (i.e., number of elements) of the set \( X \). 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 \( \phi u_{j \rightarrow q} \), if node \( q \) is directly observable; (ii) with probability \( (1 - \phi) v_{j \rightarrow q} \), if node \( q \) is not directly observable. Thus, the probability that the connection \( j \rightarrow q \) brings to the LOC is \( \phi u_{j \rightarrow q} + (1 - \phi) v_{j \rightarrow q} \). The r.h.s. of Eq. (1) quantifies the probability that at least one of the connections \( j \rightarrow q \) leads to the LOC, where the tree-like ansatz allows us to consider probabilities associated with the individual edges as independent variables, hence their product appearing on the r.h.s. of Eq. (1).

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

\[
v_{i \rightarrow j} = 1 - \sum_{s_q \in N_j \setminus \{i\}} [\phi (1 - u_{j \rightarrow q})]^s_q \left[ (1 - \phi) (1 - z_{j \rightarrow q}) \right]^{1 - s_q}.
\]

For a proof of the equivalence between Eqs. (2) and (4), see SM. The sum on the r.h.s. of Eq. (4) runs over all \( 2^{k_j - 1} \) possible configurations \( \{s_q\} \) 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 \( \phi (1 - u_{j \rightarrow q}) \) is the probability that the connection \( j \rightarrow q \) does not bring node \( j \) to the LOC. Depending on whether node \( q \) is directly observable or not, this probability is either \( \phi (1 - u_{j \rightarrow q}) \) or \( (1 - \phi) (1 - z_{j \rightarrow q}) \), respectively.

The expression of \( z_{i \rightarrow j} \) in Eq. (3) can be quantified in almost the same way as \( v_{i \rightarrow j} \). We still need to consider the probabilities that the connection \( i \rightarrow j \) 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 \( s_q = 0 \), \( \forall q \in N_j \setminus \{i\} \), 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 \( z_{i \rightarrow j} \)), 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 \( \phi \). We consider the probability that the connection \( i \rightarrow j \) brings node \( i \) to the LOC. This probability is \( u_{i \rightarrow j} \), if node \( j \) is directly observable, and is \( v_{i \rightarrow j} \), if node \( j \) is not directly observable. Combining the contributions from all neighbors of node \( i \), and using again the locally tree-like ansatz, the probability \( r_i \) that node \( i \) is directly observable, but not part of the LOC is

\[
r_i = \phi \prod_{j \in N_i} [1 - \phi u_{i \rightarrow j} - (1 - \phi) v_{i \rightarrow j}].
\]
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 \( 2^k \) configurations for the neighbors of node \( i \). Again, we have to account for the special configuration \( s_j = 0, \forall j \in N_i \), 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

\[
t_i = (1 - \phi)^{k} \left\{ \sum_{s_j, r \in N_i} \prod_{j \in N_i} [\phi(1 - u_{i \rightarrow j})]^{s_j} \times [(1 - \phi)(1 - z_{i \rightarrow j})]^{1 - s_j} \right\} .
\]

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

\[
t_i = (1 - \phi) \left\{ \prod_{j \in N_i} [1 - \phi u_{i \rightarrow j} - (1 - \phi) z_{i \rightarrow j}] \right\} + (1 - \phi)^{k} \left\{ 1 - \prod_{j \in N_i} (1 - z_{i \rightarrow j}) \right\} . \tag{6}
\]

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

\[
p_i = 1 - \phi \left\{ \prod_{j \in N_i} [1 - \phi u_{i \rightarrow j} - (1 - \phi) z_{i \rightarrow j}] \right\} - (1 - \phi) \left\{ \prod_{j \in N_i} [1 - \phi u_{i \rightarrow j} - (1 - \phi) z_{i \rightarrow j}] \right\} + (1 - \phi)^{k} \left\{ 1 - \prod_{j \in N_i} (1 - z_{i \rightarrow j}) \right\} . \tag{7}
\]

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

\[
P_{\infty}^{(th)} = \frac{1}{N} \sum_{i=1}^{N} p_i . \tag{8}
\]

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

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 [1]. For every value of \( \phi \), we estimate the order parameter \( P_{\infty}^{(num)} \) 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 SM. 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 [16]

\[
\epsilon = \int_{0}^{1} |P_{\infty}^{(th)}(\phi) - P_{\infty}^{(num)}(\phi)| \, d\phi . \tag{9}
\]

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 \( \epsilon \) on the average clustering coefficient of the network \( C \) (Fig. 3h). Because the theoretical frame-
work is deployed under the locally tree-like ansatz, and $C$ can be interpreted as a good proxy for the degree of violation of this approximation, a positive correlation between $C$ and $\epsilon$ 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 found for the ordinary site percolation model, where high clustering implies large differences between ground truth and approaches based on the locally tree-like 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. 3b). This is also not a surprising result, given that our theoretical framework is expected become exact in the limit of (locally tree-like) infinite networks.

Given the continuous nature of the observability phase transition, in the vicinity of the critical point $\phi_c$, we can take a linear approximation of the system of Eqs. (1), (2) and (3), and rewrite them in matricial form as: 

$$\vec{z} = (I - M(1 - \phi) \delta_{i,r}) \vec{z}$$

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 $2E \times 2E$ non-backtracking matrix of the graph, whose generic element is defined as $M_{i \rightarrow j, \ell \rightarrow r}$, with $\delta$ Kronecker symbol [17, 18]. The generic element of the matrix $R^{(\phi)}$ is defined as $R^{(\phi)}_{i \rightarrow j, \ell \rightarrow r} = (1 - \phi)^{\delta_{i,j}} M_{i \rightarrow j, \ell \rightarrow r}$. Solving the previous system of linear equations (see SM), we arrive to the eigenvalue/eigenvector equation

$$\vec{z} = (\vec{z} - M(1 - \phi) \delta_{i,r}) \vec{z}$$

Eq. (10) serves to study the linear stability of the trivial solution $\vec{z}^2 = (0, \ldots, 0)$. The critical value $\phi_c$ of the transition equals the value of $\phi$ for which the trivial solution becomes unstable, and corresponds to the $\phi$ value for which the operator appearing on the r.h.s. of Eq. (10) has principal eigenvalue/eigenvector equal to one. Eq. (10) is useful only in a limited number of cases, as for example regular graphs (see SM). For general networks instead, solving Eq. (10) is not computationally efficient. This operation requires 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 $\phi_c$ 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 tree-like infinite networks, its performances are almost perfect regardless of the size and/or the average clustering coefficient of the network. In this paper, 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 [19–21]. Considering that the optimal solution of the observability model is formally equivalent to the minimum (partial) dominating set of a graph [22], such an extension could represent a very important contribution for research in several domains, including, among others, biology [23, 24] and social sciences [25, 26].

The authors thank C. Castellano and A. Motter for critical comments on the early stages of this research. FR acknowledges support from the National Science Foundation (CMMI-1552487) and the US Army Research Office (W911NF-16-1-0104). YY was supported by the National Science Foundation (DMS-1057128).

---

* Electronic address: filiradi@indiana.edu

[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, 1999), pp. 7–14.
[4] S. Eubank, H. Guclu, V. A. Kumar, M. V. Marathe, A. Srinivasan, Z. Toroczkai, and N. Wang, Nature 429, 180 (2004).
[5] D. Stauffer and A. Aharony, Introduction to percolation theory (Taylor and Francis, 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, 2009).
[8] K. E. Hamilton and L. P. Pryadko, Phys. Rev. Lett. 113, 208702 (2014).
[9] B. Karrer, M. E. J. Newman, and L. Zdeborová, Phys. Rev. Lett. 113, 208702 (2014).
[10] F. Radicchi and C. Castellano, Nat. Comm. 6, 10196 (2015).
[11] F. Radicchi and C. Castellano, Phys. Rev. E 93, 030302 (2016).
[12] F. Radicchi, Nature Phys. 11, 597 (2015).
[13] M. E. J. Newman and R. Ziff, Phys. Rev. Lett. 85, 4104 (2000).
[14] M. E. Newman, Proc. Natl. Acad. Sci. USA 98, 404 (2001).
[15] F. Radicchi, Physical Review E 91, 010801 (2015).
[16] S. Mehnil, A. Hackett, M. A. Porter, P. J. Mucha, and
J. P. Gleeson, Phys. Rev. E 83, 036112 (2011).
[17] K.-i. Hashimoto, Automorphic forms and geometry of arithmetic varieties. pp. 211–280 (1989).
[18] F. Krzakala, C. Moore, E. Mossel, J. Neeman, A. Sly, L. Zdeborová, and P. Zhang, Proc. Natl. Acad. Sci. USA 110, 20935 (2013).
[19] F. Morone and H. A. Makse, Nature 524, 65 (2015).
[20] A. Braunstein, L. Dall’Asta, G. Semerjian, and L. Zdeborová, arXiv:1603.08883 (2016).
[21] P. Clusella, P. Grassberger, F. J. Perez-Reche, and A. Politi, arXiv:1604.00073 (2016).
[22] D.-Z. Du and P.-J. Wan, Connected dominating set: theory and applications, vol. 77 (Springer Science & Business Media, 2012).
[23] S. Wuchty, Proc. Natl. Acad. Sci. USA 111, 7156 (2014).
[24] T. Milenković, V. Menišević, A. Bonato, and N. Pržulj, PloS one 6, e23016 (2011).
[25] L. L. Kelleher and M. B. Cozzens, Mathematical Social Sciences 16, 267 (1988).
[26] F. Wang, H. Du, E. Camacho, K. Xu, W. Lee, Y. Shi, and S. Shan, Theoretical Computer Science 412, 265 (2011).
Appendix

Proof by induction of Equation (4) of the main text

In the main text, we used the fact that

\[ \prod_{q \in \mathcal{Q}} [1 - \phi a_q - (1 - \phi) b_q] = \sum_{\{s, r\} \in \mathcal{Q} \setminus \{p\}} \prod_{q \in \mathcal{Q} \setminus \{p\}} \left[ (1 - \phi)(1 - b_q) \right]^{-s_q} \cdot (1 - \phi) \cdot \prod_{q \in \mathcal{Q} \setminus \{p\}} \left[ (1 - \phi)(1 - b_q) \right]^{-s_q}. \]  

(SM1)

Note that Eq. (SM1) is a more general version of Eq. (4). We recall that sum on the r.h.s. 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 \( \phi \) 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. (SM1). To this end, we first note that if \( \mathcal{Q} = \emptyset \), then Eq. (SM1) is automatically satisfied, being both sides equal to one. If \(|\mathcal{Q}| > 0\), we hypothesize that

\[ \prod_{q \in \mathcal{Q} \setminus \{p\}} [1 - \phi a_q - (1 - \phi) b_q] = \sum_{\{s, r\} \in \mathcal{Q} \setminus \{p\}} \prod_{q \in \mathcal{Q} \setminus \{p\}} \left[ (1 - \phi)(1 - b_q) \right]^{-s_q} \cdot (1 - \phi) \cdot \prod_{q \in \mathcal{Q} \setminus \{p\}} \left[ (1 - \phi)(1 - b_q) \right]^{-s_q}. \]  

(SM2)

The latter equation is the analogue of Eq. (SM1) 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. (SM1), we have

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

By virtue of the hypothesis of Eq. (SM2), the validity of Eq. (SM1) 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

\[ \phi(1 - a_p) + (1 - \phi)(1 - b_p) = 1 - \phi a_p - (1 - \phi) b_p. \]

Linear approximation

Using the 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

\[ \vec{u} = M[\phi \vec{u} + (1 - \phi) \vec{v}], \quad (SM3) \]

\[ \vec{v} = M[\phi \vec{u} + (1 - \phi) \vec{z}], \quad (SM4) \]

and

\[ \vec{z} = \vec{v} - R^{(\phi)} \vec{z}, \quad (SM5) \]

where \( R^{(\phi)}_{i \to j, l \to r} = (1 - \phi)^{l-1} M_{i \to j, l \to r} \), and \( M \) is the non-backtracking matrix of the graph. From Eq. (SM3), we obtain

\[ \vec{u} = (1 - \phi)(I - \phi M)^{-1} M \vec{v}. \]
Inserting this expression into Eq. (SM4), we have

\[ \vec{v} = \left[ (1 - \phi) (1 - \phi M)^{-1} M \right]^{-1} (1 - \phi) M \vec{z} \, . \]

Finally, using this expression in Eq. (SM5), 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 \( \phi_c \) 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 \( \mu = k - 1 \).
| network                  | N    | $\epsilon$ | $C$   | Refs. | Url |
|-------------------------|------|------------|-------|-------|-----|
| Social 3                | 32   | 0.000978   | 0.3266 | 9     | url |
| Karate club             | 34   | 0.001490   | 0.5706 | 1     | url |
| Protein 2               | 35   | 0.013476   | 0.4135 | 3     | url |
| Dolphins                | 62   | 0.005209   | 0.2590 | 5     | url |
| Social 1                | 67   | 0.001794   | 0.3099 | 9     | url |
| Les Miserables          | 77   | 0.000253   | 0.5731 | 6     | url |
| Protein 1               | 95   | 0.046827   | 0.3991 | 3     | url |
| E. Coli, transcription  | 97   | 0.027824   | 0.3675 | 7     | url |
| Political books         | 105  | 0.001623   | 0.4875 | 8     | url |
| David Copperfield       | 112  | 0.000107   | 0.1728 | 9     | url |
| College football        | 115  | 0.000142   | 0.4032 | 10    | url |
| S 208                   | 122  | 0.007029   | 0.0591 | 3     | url |
| High school, 2011       | 126  | 0.000103   | 0.5759 | 11    | url |
| Bay Wet                 | 128  | 0.000096   | 0.3346 | 12    | url |
| Bay Dry                 | 128  | 0.000055   | 0.3346 | 12    | url |
| Radoslaw Email          | 167  | 0.000168   | 0.5919 | 12    | url |
| High school, 2012       | 180  | 0.000083   | 0.4752 | 11    | url |
| Little Rock Lake        | 183  | 0.000303   | 0.3226 | 12    | url |
| Jazz                    | 198  | 0.000126   | 0.6175 | 16    | url |
| S 420                   | 252  | 0.007366   | 0.0561 | 3     | url |
| C. Elegans, neural      | 297  | 0.000068   | 0.2924 | 17    | url |
| Network Science         | 379  | 0.014257   | 0.7412 | 9     | url |
| Dublin                  | 410  | 0.000595   | 0.4558 | 12    | url |
| US Air Traffic          | 500  | 0.000970   | 0.6175 | 19    | url |
| S 838                   | 512  | 0.007634   | 0.0547 | 3     | url |
| Yeast, transcription    | 662  | 0.001809   | 0.0490 | 19    | url |
| URV email               | 1,133| 0.000050   | 0.2202 | 21    | url |
| Political blogs         | 1,222| 0.000076   | 0.3203 | 5     | url |
| Air traffic             | 1,226| 0.001000   | 0.0675 | 12    | url |
| Yeast, protein          | 1,458| 0.004232   | 0.0708 | 22    | url |
| Petster, hamster        | 1,788| 0.000163   | 0.1433 | 12    | url |
| UC Irvine               | 1,893| 0.000154   | 0.1097 | 12    | url |
| Yeast, protein          | 2,224| 0.000228   | 0.1381 | 24    | url |
| Japanese                | 2,698| 0.000536   | 0.2196 | 3     | url |
| Open flights            | 2,905| 0.004378   | 0.5569 | 26    | url |
| GR-QC, 1993-2003        | 4,158| 0.004387   | 0.5569 | 26    | url |
| Tennis                  | 4,338| 0.000045   | 0.2888 | 12    | url |
| US Power grid           | 4,941| 0.042156   | 0.0801 | 17    | url |
| HT09                    | 5,352| 0.009204   | 0.0807 | 15    | url |
| Hep-Th, 1995-1999       | 5,835| 0.005651   | 0.5062 | 2     | url |

Table SM1: Summary table for real-world networks. The columns of the table respectively report: the name of the network, the number of nodes in the giant component, the value $\epsilon$ as defined in Eq. (9) of the main text, the average clustering coefficient of the network, reference(s) of the paper where the network has been first analyzed, and url of where the network data have been downloaded (to open the web page in your browser, just click on the word url).
| network               | $N$    | $\epsilon$ | $C$    | Refs. | Url |
|----------------------|--------|------------|--------|-------|-----|
| Reactome             | 5,973  | 0.001586   | 0.6091 | 12,28 | url |
| Jung                 | 6,120  | 0.000623   | 0.6752 | 12,29 | url |
| Gnutella, Aug. 8, 2002 | 6,299  | 0.000038   | 0.0109 | 20,20 | url |
| JDK                  | 6,434  | 0.000431   | 0.6707 | 12    | url |
| AS Oregon            | 6,474  | 0.000283   | 0.2522 | 31    | url |
| English              | 7,377  | 0.000142   | 0.4085 | 3     | url |
| Gnutella, Aug. 9, 2002 | 8,104  | 0.000021   | 0.0095 | 26,30 | url |
| French               | 8,308  | 0.000231   | 0.2138 | 3     | url |
| Hep-TH, 1993-2003    | 8,638  | 0.002722   | 0.4816 | 26    | url |
| Gnutella, Aug. 6, 2002 | 8,717  | 0.000021   | 0.0067 | 26,30 | url |
| Gnutella, Aug. 5, 2002 | 8,842  | 0.000017   | 0.0072 | 26,30 | url |
| PGP                  | 10,680 | 0.008295   | 0.2659 | 32    | url |
| Gnutella, August 4 2002 | 10,876 | 0.000032   | 0.0062 | 26,30 | url |
| Hep-Ph, 1993-2003    | 11,204 | 0.000879   | 0.6216 | 26    | url |
| Spanish              | 11,558 | 0.000245   | 0.3764 | 3     | url |
| DBLP, citations      | 12,495 | 0.000307   | 0.1178 | 12,33 | url |
| Spanish              | 12,643 | 0.000658   | 0.5042 | 12    | url |
| Cond-Mat, 1995-1999  | 13,861 | 0.003846   | 0.6514 | 2     | url |
| Astrophysics         | 14,845 | 0.000986   | 0.6696 | 2     | url |
| Google               | 15,763 | 0.000248   | 0.5176 | 31    | url |
| AstroPhys, 1993-2003 | 17,903 | 0.000468   | 0.6328 | 26    | url |
| Cond-Mat, 1993-2003  | 21,363 | 0.001047   | 0.6417 | 26    | url |
| Gnutella, Aug 25, 2002 | 22,663 | 0.000024   | 0.0053 | 26,30 | url |
| Internet             | 22,963 | 0.000151   | 0.2304 | -     | url |
| Thesaurus            | 23,132 | 0.000018   | 0.0888 | 12,35 | url |
| Cora                 | 23,166 | 0.000734   | 0.2660 | 12,36 | url |
| Linux, mailing list  | 24,567 | 0.001928   | 0.3391 | 12    | url |
| AS Caida             | 26,475 | 0.000135   | 0.2082 | 31    | url |
| Gnutella, Aug 24, 2002 | 26,498 | 0.000013   | 0.0055 | 26,30 | url |
| Hep-TH, citations    | 27,400 | 0.000267   | 0.3139 | 12,26 | url |
| Cond-Mat, 1995-2003  | 27,519 | 0.001457   | 0.6546 | 2     | url |
| Digg                 | 29,652 | 0.000013   | 0.0054 | 12,37 | url |
| Linux, soft.         | 30,817 | 0.000124   | 0.1286 | 12    | url |
| Enron                | 33,696 | 0.000330   | 0.5092 | 38    | url |
| Hep-Ph, citations    | 34,401 | 0.000079   | 0.2856 | 12,26 | url |
| Cond-Mat, 1995-2005  | 36,458 | 0.001240   | 0.6566 | 2     | url |
| Gnutella, Aug 30, 2002 | 36,640 | 0.000014   | 0.0063 | 26,30 | url |
| Slashdot             | 51,083 | 0.000033   | 0.0201 | 12,39 | url |
| Gnutella, Aug 31, 2002 | 62,561 | 0.000008   | 0.0055 | 26,30 | url |
| Facebook             | 63,392 | 0.000077   | 0.2218 | 30    | url |

Table SM2: Continuation of Table SM1
| network         | N     | \(\epsilon\) | C     | Refs | Url          |
|-----------------|-------|---------------|-------|------|--------------|
| Epinions        | 75,877| 0.000140      | 0.1378| [12] | [11] url     |
| Slashdot zoo    | 79,116| 0.000067      | 0.0584| [12] | [12] url     |
| Flickr          | 105,722| 0.000018      | 0.0884| [12] | [13] url     |
| Wikipedia, edits| 113,123| 0.000051      | 0.3748| [12] | [14] url     |
| Petster, cats   | 148,826| 0.000134      | 0.3877| [12] | [15] url     |
| Gowalla         | 196,591| 0.000434      | 0.2367| [12] | [15] url     |
| EU email        | 224,832| 0.000045      | 0.0791| [12] | [16] url     |
| Web Stanford    | 255,265| 0.010009      | 0.6189| [17] | [17] url     |
| Amazon, Mar. 2, 2003 | 262,111| 0.004310      | 0.4198| [16] | [18] url     |
| DBLP, collaborations | 317,080| 0.002107      | 0.6324| [12] | [19] url     |
| Web Notre Dame  | 325,729| 0.002287      | 0.2346| [16] | [19] url     |
| MathSciNet      | 332,689| 0.002048      | 0.4104| [16] | [19] url     |
| CiteSeer        | 365,154| 0.001183      | 0.1832| [12] | [19] url     |
| Amazon, Mar. 12, 2003 | 400,727| 0.000545      | 0.4022| [16] | [19] url     |
| Amazon, Jun. 6, 2003 | 403,364| 0.000595      | 0.4177| [16] | [19] url     |
| Amazon, May 5, 2003 | 410,236| 0.000573      | 0.4064| [16] | [19] url     |

Table SM3: Continuation of Tables SM1 and SM2

* Electronic address: filiradi@indiana.edu

[1] M. E. J. Newman and R. Ziff, Physical Review Letters 85, 4104 (2000).
[2] M. E. Newman, Proceedings of the National Academy of Sciences 98, 404 (2001).
[3] R. Milo, S. Itzkovitz, N. Kashtan, R. Levitt, S. Shen-Orr, I. Ayzenshtat, M. Sheffer, and U. Alon, Science 303, 1538 (2004).
[4] W. W. Zachary, Journal of anthropological research pp. 452–473 (1977).
[5] D. Lusseau, K. Schneider, O. J. Boisseau, P. Haase, E. Slooten, and S. M. Dawson, Behavioral Ecology and Sociobiology 54, 396 (2003).
[6] D. E. Knuth, D. E. Knuth, and D. E. Knuth, The Stanford GraphBase: a platform for combinatorial computing, vol. 37 (Addison-Wesley Reading, 1993).
[7] S. Mangan and U. Alon, Proceedings of the National Academy of Sciences 100, 11980 (2003).
[8] L. A. Adamic and N. Glance, in Proceedings of the 3rd international workshop on Link discovery (ACM, 2005), pp. 36–43.
[9] M. E. Newman, Physical review E 74, 036104 (2006).
[10] M. Girvan and M. E. Newman, Proceedings of the National Academy of Sciences 99, 7821 (2002).
[11] J. Fournet and A. Barrat, PloS one 9, e107878 (2014).
[12] J. Kunegis, in Proc. Int. Conf. on World Wide Web Companion (2013), pp. 1343-1350, URL http://userpages.uni-koblenz.de/~kunegis/paper/kunegis-koblenz-network-collection.pdf
[13] R. Ulanowicz, C. Bondavalli, and M. Egnotovich, Annual Report to the United States Geological Service Biological Resources Division Ref. No.[UMCES] CBL pp. 98–123 (1998).
[14] R. Michalski, S. Palus, and P. Kazienko, in Lecture Notes in Business Information Processing (Springer Berlin Heidelberg, 2011), vol. 87, pp. 197–206.
[15] N. D. Martinez, Ecological Monographs pp. 367–392 (1991).
[16] P. M. Gleiser and L. Danon, Advances in complex systems 6, 565 (2003).
[17] D. J. Watts and S. H. Strogatz, nature 393, 440 (1998).
[18] L. Isella, J. Stehlik, A. Barrat, C. Cattuto, J.-F. Pinton, and W. Van den Broeck, Journal of theoretical biology 271, 166 (2011).
[19] V. Colizza, R. Pastor-Satorras, and A. Vespignani, Nature Physics 3, 276 (2007).
[20] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon, Science 298, 824 (2002).
[21] R. Guimerà, L. Danon, A. Diaz-Guilera, F. Giralt, and A. Arenas, Physical review E 68, 065103 (2003).
[22] H. Jeong, S. P. Mason, A.-L. Barabási, and Z. N. Oltvai, Nature 411, 41 (2001).
[23] T. Opsahl and P. Panzarasa, Social networks 31, 155 (2009).
[24] D. Bu, Y. Zhao, L. Cai, H. Xue, X. Zhu, H. Lu, J. Zhang, S. Sun, L. Ling, N. Zhang, et al., Nucleic acids research 31, 2443 (2003).
[25] T. Opsahl, F. Agneessens, and J. Skvoretz, Social Networks 32, 245 (2010).
[26] J. Leskovec, J. Kleinberg, and C. Faloutsos, ACM Transactions on Knowledge Discovery from Data (TKDD) 1, 2 (2007).
[27] F. Radicchi, PloS one 6, e17249 (2011).
[28] G. Joshi-Tope, M. Gillespie, I. Vastrik, P. D’Eustachio, E. Schmidt, B. de Bono, B. Jassal, G. Gopinath, G. Wu, L. Matthews, et al., Nucleic acids research 33, D428 (2005).
[29] L. Šubelj and M. Bajec, in Proceedings of the First International Workshop on Software Mining (ACM, 2012), pp. 9–16.
[30] M. Ripeanu, I. Foster, and A. Iamnitchi, arXiv preprint cs/0209028 (2002).
[31] J. Leskovec, J. Kleinberg, and C. Faloutsos, in Proceedings of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining (ACM, 2005), pp. 177–187.
[32] M. Boguñá, R. Pastor-Satorras, A. Díaz-Guilera, and A. Arenas, Physical Review E 70, 056122 (2004).
[33] M. Ley, in String Processing and Information Retrieval (Springer, 2002), pp. 1–10.
[34] G. Palla, I. J. Farkas, P. Pollner, I. Derenyi, and T. Vicsek, New Journal of Physics 9, 186 (2007).
[35] G. R. Kiss, C. Armstrong, R. Milroy, and J. Piper, The computer and literary studies pp. 153–165 (1973).
[36] L. Šubelj and M. Bajec, in Proceedings of the 22nd international conference on World Wide Web companion (International World Wide Web Conferences Steering Committee, 2013), pp. 527–530.
[37] M. De Choudhury, H. Sundaram, A. John, and D. D. Seligmann, in Computational Science and Engineering, 2009. CSE’09. International Conference on (IEEE, 2009), vol. 4, pp. 151–158.
[38] J. Leskovec, K. J. Lang, A. Dasgupta, and M. W. Mahoney, Internet Mathematics 6, 29 (2009).
[39] V. Gómez, A. Kaltenbrunner, and V. López, in Proceedings of the 17th international conference on World Wide Web (ACM, 2008), pp. 645–654.
[40] B. Viswanath, A. Mislove, M. Cha, and K. P. Gummadi, in Proceedings of the 2nd ACM workshop on Online social networks (ACM, 2009), pp. 37–42.
[41] M. Richardson, R. Agrawal, and P. Domingos, in The Semantic Web- ISWC 2003 (Springer, 2003), pp. 351–368.
[42] J. Kunegis, A. Lommatzsch, and C. Bauckhage, in Proceedings of the 18th international conference on World wide web (ACM, 2009), pp. 741–750.
[43] J. McAuley and J. Leskovec, in Advances in Neural Information Processing Systems (2012), pp. 548–556.
[44] U. Brandes and J. Lerner, Journal of classification 27, 279 (2010).
[45] E. Cho, S. A. Myers, and J. Leskovec, in Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining (ACM, 2011), pp. 1082–1090.
[46] J. Leskovec, L. A. Adamic, and B. A. Huberman, ACM Transactions on the Web (TWEB) 1, 5 (2007).
[47] R. Albert, H. Jeong, and A.-L. Barabási, Nature 401, 130 (1999).
[48] G. Palla, I. J. Farkas, P. Pollner, I. Derényi, and T. Vicsek, New Journal of Physics 10, 123026 (2008).
[49] K. D. Bollacker, S. Lawrence, and C. L. Giles, in Proceedings of the second international conference on Autonomous agents (ACM, 1998), pp. 116–123.