Network Controllability Is Determined by the Density of Low In-Degree and Out-Degree Nodes

Giulia Menichetti
Department of Physics and Astronomy and INFN Sezione di Bologna, Bologna University, Viale Berti Pichat 6/2, 40127 Bologna, Italy

Luca Dall'Asta
Department of Applied Science and Technology DISAT, Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129 Torino, Italy and Collegio Carlo Alberto, Via Real Collegio 30, 10024 Moncalieri, Italy

Ginestra Bianconi
School of Mathematical Sciences, Queen Mary University of London, London E1 4NS, United Kingdom
(Received 17 May 2014; revised manuscript received 7 July 2014; published 13 August 2014)

The problem of controllability of the dynamical state of a network is central in network theory and has wide applications ranging from network medicine to financial markets. The driver nodes of the network are the nodes that can bring the network to the desired dynamical state if an external signal is applied to them. Using the framework of structural controllability, here, we show that the density of nodes with in degree and out degree equal to one and two determines the number of driver nodes in the network. Moreover, we show that random networks with minimum in degree and out degree greater than two, are always fully controllable by an infinitesimal fraction of driver nodes, regardless of the other properties of the degree distribution. Finally, based on these results, we propose an algorithm to improve the controllability of networks.

DOI: 10.1103/PhysRevLett.113.078701 PACS numbers: 89.75.Hc, 05.70.Fh, 64.60.aq, 89.75.Fb

The controllability of a network [1–10] is a fundamental problem with wide applications ranging from medicine and drug discovery [11], to the characterization of dynamical processes in the brain [12–14], or the evaluation of risk in financial markets [15]. While the interplay between the structure of the network [16–19] and the dynamical processes defined on them has been an active subject of complex network research for more than ten years [20,21], only recently has the rich interplay between the controllability of a network and its structure started to be investigated. A pivotal role in this respect has been played by a paper by Liu *et al.* [6], in which the problem of finding the minimal set of driver nodes necessary to control a network was mapped into a maximum matching problem. Using a well established statistical mechanics approach [22–27], Liu *et al.* [6] characterize in detail the set of driver nodes for real networks and for ensembles of networks with given in-degree and out-degree distribution. By analyzing scale-free networks with minimum in degree and minimum out degree equal to one, they have found that the smaller the power-law exponent γ of the degree distribution, the larger is the fraction of driver nodes in the network. This result has prompted the authors of [6] to say that the higher the heterogeneity of the degree distribution, the less controllable is the network. Later, different papers addressed questions related to controllability of networks with similar tools [7,28].

In this Letter, we consider the network controllability and its mapping to the maximum matching problem, exploring the role of low in-degree and low out-degree nodes in the network. We show that by changing the fraction of nodes with in degree and out degree less than three, the number of driver nodes of a network can change in a dramatic way. In particular, if the minimum in degree and the minimum out degree of a network are both greater than two, then any network, independently on the level of heterogeneity of the degree distribution, is fully controllable by an infinitesimal fraction of nodes. Therefore, we show that the heterogeneity of the network is not the only element determining the number of driver nodes in the network and that this number is very sensible on the fraction of low in-degree and low out-degree nodes of the network. This result allows us to propose a method to improve the controllability of networks by decreasing the density of nodes with in degree and out degree less than three, adding links to the network.

The structural controllability of a network.—Given a graph $G = (V, E)$ of N nodes, we consider a continuous-time linear dynamical system

$$\frac{d\mathbf{x}(t)}{dt} = A\mathbf{x} + B\mathbf{u}, \tag{1}$$

in which the vector $\mathbf{x}(t)$, of elements $x_i(t)$ with $i = 1, 2, \ldots, N$, represents the dynamical state of the network, A is an $N \times N$ (asymmetric) matrix describing the directed weighted interactions within the network, and B is an $N \times M$ matrix describing the interaction between the nodes of the graph and $M \leq N$ external signals, indicated by the vector $\mathbf{u}(t)$ of elements u_α and $\alpha = 1, 2 \ldots M$. For any given realization of A and B, the dynamical system is controllable if it satisfies Kalman's controllability rank condition; i.e., the matrix $C = (B, AB, A^2B, \ldots, A^{N-1}B)$ is full rank. In addition to the fact that the verification of Kalman's condition can be computationally very demanding for large systems, in most real systems, the notion of exact controllability is unusable since the entries of A and B are not perfectly known. As an alternative, if we assume that the nonzero matrix elements of A and B are free parameters, we can consider the concept of structural controllability [5]. The system is structurally controllable if, for any choice of the free parameters in A and B, except for a variety of zero Lebesgue measure in the parameter space, C is full rank [5]. Since structural controllability only distinguishes between zero and nonzero entries of the matrices A and B, a given directed network is structurally controllable if it is possible to determine the input nodes (i.e., the position of the nonzero entries of matrix B) in a way to control the dynamics described by any realization of matrix A with the same nonzero elements, except for atypical realizations of zero measure. In practice, a network can be structurally controlled by identifying a minimum number of driver nodes, that are controlled nodes which do not share input vertices. In their seminal paper [6], Liu and co-workers showed that this control theoretic problem can be reduced to a well-known optimization problem: their minimum input theorem states that the minimum set of driver nodes that guarantees the full structural controllability of a network is the set of unmatched nodes in a maximum matching of the same directed network.

The maximum matching problem.—A matching M of a directed graph is a set of directed edges without common start or end vertices, and it is maximum when it contains the maximum possible number of edges. The problem of finding a maximum matching of a directed graph can be cast on a statistical mechanics problem, by introducing variables $s_{ij} \in \{1, 0\}$ on each directed link from node i to node j, indicating whether the directed link is in M ($s_{ij} = 1$) or not ($s_{ij} = 0$). The configurations of variables $\{s_{ij}\}$ have to satisfy the following matching condition:

$$\sum_{j \in \partial_+ i} s_{ij} \leq 1, \qquad \sum_{j \in \partial_- i} s_{ji} \leq 1, \tag{2}$$

where $\partial_- i$ indicates the set of nodes j that point to node i in the directed network, and $\partial_+ i$ indicates the set of nodes j that are pointed by node i. Moreover, the variables $\{s_{ij}\}$ should minimize the energy function

$$E = 2\sum_{i=1}^{N}\left(1 - \sum_{j \in \partial_- i} s_{ji}\right). \tag{3}$$

Note that a vertex is matched if it is the endpoint of one of the edges in the matching, otherwise, the vertex is unmatched. It follows that $E = 2N_D$, where N_D is the number of unmatched nodes in the network, and this number also determines the minimum number of driver nodes required to fully control the network. Following Refs. [6,22], we use the cavity method in the zero-temperature limit to study the statistical properties of maximum matchings on directed random graphs for which the locally tree-like approximation holds. Under the decorrelation (replica-symmetric) assumption, the energy of a maximum matching can be written in terms of the cavity fields (or messages) $h_{i\to j}$ or $\hat{h}_{i\to j}$ sent from a node i to the linked node j. The fields are sent in the same direction $h_{i\to j}$ or in the opposite direction $\hat{h}_{i\to j}$ of the links and indicate the following messages [22]: $h_{i\to j} = \hat{h}_{i\to j} = 1$ indicates "match me," $h_{i\to j} = \hat{h}_{i\to j} = -1$ indicates "do not match me," finally, $h_{i\to j} = \hat{h}_{i\to j} = 0$ indicates "do what you want." In fact, the energy E follows (see Supplemental Material [29] for details):

$$E = -\sum_{i=1}^{N}\max\left[-1, \max_{k\in\partial_+ i}\hat{h}_{k\to i}\right] - \sum_{i=1}^{N}\max\left[-1, \max_{k\in\partial_- i}h_{k\to i}\right] + \sum_{\langle i,j\rangle}\max\left[0, h_{i\to j} + \hat{h}_{j\to i}\right] \tag{4}$$

in which, for each directed link (i, j), the cavity fields $\{h_{i\to j}, \hat{h}_{i\to j}\}$ satisfy the following zero-temperature version of the belief propagation (BP) equations, also known as max-sum (MS) equations,

$$h_{i\to j} = -\max\left[-1, \max_{k\in\partial_+ i\setminus j}\hat{h}_{k\to i}\right], \tag{5a}$$

$$\hat{h}_{i\to j} = -\max\left[-1, \max_{k\in\partial_- i\setminus j}h_{k\to i}\right], \tag{5b}$$

with the assumption that the maximum over an empty set is equal to -1. In the infinite size limit, the MS equations are closed for cavity fields with support on $\{-1, 0, 1\}$ [6,22,23]. These equations can be solved by iteration using the BP/MS algorithm.

Sufficient condition for the full controllability of networks.—Let us, now, show that, for any network topology, if the in degree and the out degree of the network is greater than two, the fraction of driver nodes is zero. First, we observe that the configuration in which all fields are zero, i.e., $h_{i\to j} = \hat{h}_{i\to j} = 0$, is an allowed solution of Eqs. (5a)–(5b) as soon as the minimum in degree and minimum out degree equal one. In fact, if a node has in-degree one, this link must be matched, and a similar situation occurs for the nodes with out degree one, generating a set of hard constraints incompatible with the configuration in which all the fields are zero, while if the minimum in degree or out degree of the network is greater than one, all the nodes can be matched in a variety of ways; therefore, all the fields can be equal to zero. This solution corresponds to a fraction of driver nodes $n_D = 0$ if the minimum in degree and the minimum out degree are greater than one. This solution is also stable if, when we change a single field from zero to a value different from zero, the perturbation does not propagate in the network. Suppose that $\hat{h}_{k\to i}$ is changed, say, from 0 to 1, meaning that the message is "match me," then all the nodes $j \in \partial_+ i$ neighboring i and different from k receive a message "do not match me." But, if all the nodes j have more than two incoming links, also, if the link (j, k) is not matched, they can still send to their incoming neighbors the message "do what you want" since there are different ways in which the matching can be achieved, and they do not have to impose on any of their other links to be matched. Therefore, the perturbation does not propagate in the network. A similar argument holds for a change of the field $h_{k\to i}$ to one which does not propagate if the out degree of the network is greater than two. This stability argument shows that for every tree-like network for which the BP/MS equations are valid, if the in degree and the out degree of the network is greater than two, then the density of driver nodes is $n_D = 0$. Note that this a sufficient condition for the stability of the $n_D = 0$ solution but more stringent conditions are discussed in the following for networks with given degree distribution.

Conditions for the full controllability of random networks.—In the following, we focus on ensembles of random networks with given in-degree and out-degree distribution $P^{\mathrm{in}}(k)$ and $P^{\mathrm{out}}(k)$. In this case (see the Supplemental Material [29]), it is possible to write the BP/MS equations and the energy in terms of the probabilities $w_i \in [0, 1]$ and $\hat{w}_i \in [0, 1]$ with $i = 1, 2, 3$ that the cavity fields $h_{i\to j}$ and $\hat{h}_{i\to j}$ are, respectively, given by $\{1, -1, 0\}$. From the BP/MS equations of the matching problem on random networks with given degree distribution, we found that the solution $n_D = 0$ is allowed if and only if $P^{\mathrm{in/out}}(0) = P^{\mathrm{in/out}}(1) = 0$. The replica-symmetric cavity equations are supposed to give the correct solution to the maximum matching problem if no instabilities take place [22,33,34]. By analyzing the stability condition of the BP/MS equations [29], we find that the stability conditions for this solution in an ensemble of networks with given in-degree and out-degree sequences, are

$$P^{\mathrm{out}}(2) < \frac{\langle k\rangle_{\mathrm{in}}^2}{2\langle k(k-1)\rangle_{\mathrm{in}}}, \quad P^{\mathrm{in}}(2) < \frac{\langle k\rangle_{\mathrm{in}}^2}{2\langle k(k-1)\rangle_{\mathrm{out}}}. \qquad (6)$$

In particular, when the minimum in degree and the minimum out degree of scale-free networks are both greater than two, i.e., $P^{\mathrm{in/out}}(0) = P^{\mathrm{in/out}}(1) = P^{\mathrm{in/out}}(2) = 0$, the fraction of driver nodes is zero in the thermodynamic limit, for any choice of the degree distribution with this property. By changing the minimum in degree and minimum out degree of the network, the number of driver nodes can change dramatically, independently of the tail of the degree distribution and the level of degree heterogeneity.

In order to use the above calculation to estimate the role of low-degree nodes on the fate of the zero-energy solution in finite networks, we consider uncorrelated random graphs with the following power-law degree distribution:

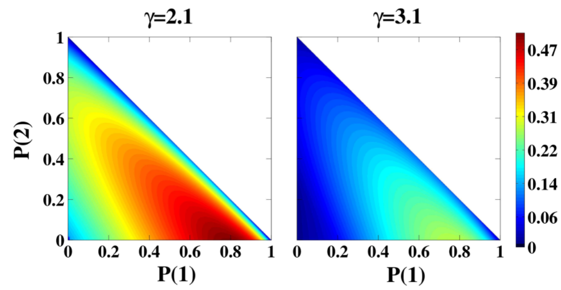

FIG. 1 (color online). Heat map representing the density of driver nodes n_D as a function of the parameters $P(1)$ and $P(2)$ for networks of $N = 10^6$ nodes with degree distribution given by Eq. (7) and $\gamma = 2.1$ (left), 3.1 (right). The density n_D is obtained by numerically solving the BP/MS equations for an ensemble of networks with given degree distribution. The region in which $P(1) + P(2) > 1$ is nonphysical.

$$P^{\mathrm{in}}(k) = P^{\mathrm{out}}(k) = \begin{cases} P(1) & \text{if } k = 1 \\ P(2) & \text{if } k = 2 \\ Ck^{-\gamma} & \text{if } k \in [3, K] \end{cases} \qquad (7)$$

with C a constant determined by normalization and maximum degree $K = \min(\sqrt{N}, \{[1-P(1)-P(2)]N\}^{1/(\gamma-1)})$ for $\gamma > 2$ and $K = \min(N^{1/\gamma}, \{[1-P(1)-P(2)]N\}^{1/(\gamma-1)})$ for $\gamma \in (1, 2]$, that is the minimum between the structural cutoff [35,36] of the network and the natural cutoff of the degree distribution. These networks can be generated numerically using the configuration model. As long as $P(1) = P(2) = 0$, the density of driver nodes goes to zero ($n_D \to 0$) for any exponent $\gamma > 1$. More generally, the density n_D of driver nodes changes dramatically as a function of $P(1)$ and $P(2)$ as shown by the heat map in Fig. 1 for $\gamma = 2.1, 3.1$. Moreover, in Fig. 2, we plot the

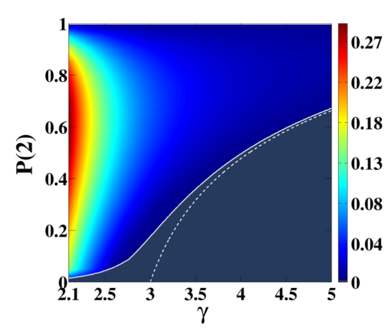

FIG. 2 (color online). Phase diagram of the density of driver nodes n_D as a function of the parameters γ and $P(2)$ for networks of $N = 10^6$ nodes with degree distribution given by Eq. (7) and $P(1) = 0$. The density n_D is obtained by numerically solving the BP/MS equations for an ensemble of networks with given degree distribution. The solid lines indicate the stability lines for $N = 10^6$, the dotted lines indicate the stability lines in the limit $N \to \infty$.

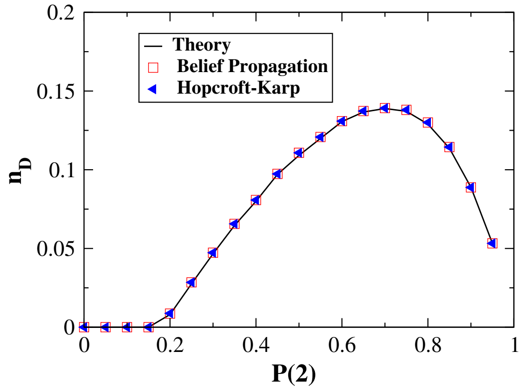

FIG. 3 (color online). Density of driver nodes n_D as a function of $P(2)$ for in-degree and out-degree distributions as in Eq. (7) with $P(1) = 0$ and $\gamma = 2.3$. The fraction of driver nodes computed with the BP/MS algorithm on a network of $N = 10^4$ nodes (averaged over 50 network realizations) is compared with the exact results obtained using the Hopcroft-Karp algorithm for maximum matching [32] and with the theoretical expectation for the density n_D in an ensemble of random networks with the same degree distribution.

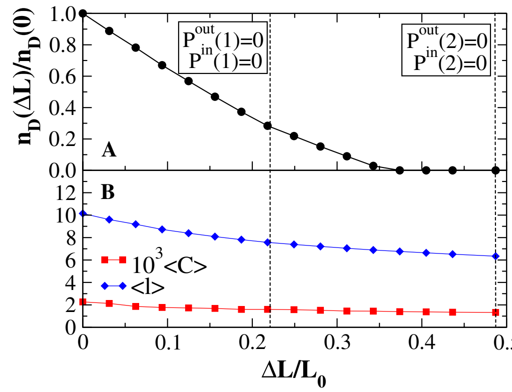

FIG. 4 (color online). Fraction of driver nodes $n_D(\Delta L)/n_D(0)$ (a), average clustering coefficient $\langle C \rangle$ and average distance $\langle l \rangle$ (b) of the network as a function of the fraction of added links to low degree nodes. The results are obtained from the BP/MS algorithm. The initial network is a power-law network with in-degree distribution equal to the out-degree distribution, $N = 10^4$ nodes, and power-law exponent $\gamma = 2.3$. The symbol ΔL indicates the number of added links to the network, whereas L_0 indicates the initial number of links of the network.

phase diagram for $P(1) = 0$ indicating the region where the solution $n_D = 0$ is stable both for a finite network of $N = 10^6$ nodes (white solid line) and for $N \to \infty$ (white dotted line). Note that, for $\gamma \in (2, 3]$, stability line converges quite slowly to zero in the infinite size limit.

A confirmation of the validity of this scenario is reported in Fig. 3 from a direct comparison of the theoretical results in the ensemble of networks with a given degree distribution, with those obtained by the BP algorithm, or by computing explicitly the maximum matching using the Hopcroft-Karp algorithm [32] finding very good agreement. Figure 3 also shows that n_D vanishes by decreasing $P(2)$. From our numerical results (reported in the Supplemental Material [29]), in the region in which the solution $n_D = 0$ is stable and we are far from the stability transition, both algorithms give a zero number of driver nodes $N_D = 0$, meaning that all the nodes are matched, and therefore, a single external input can be used to control the network.

Improving the controllability of a network.—These results suggest a simple and very effective way to improve the controllability of a network, by decreasing the fraction of nodes with in degree and out degree equal to zero, one, and two. Starting from a network with given degree distribution, we first add links starting from any node of out degree equal to zero (if present in the network) and randomly attached to any other node of the network, or starting from any random node of the network and ending to nodes of in degree zero. When there are no more nodes with in degree or out degree equal to zero, we repeat the process of random addition of links to nodes with in degree or out degree equal to one and two. At the end of the process, the minimum in degree of the network and the minimum out degree is equal to three.

Figure 4(a) shows the reduction in the fraction of driver nodes $n_D(\Delta L)$ compared to the original one $n_D(0)$ due to the addition of a fraction $\Delta L/L_0$ of directed links to a network with pure power-law degree distribution and structural cutoff. It is clear that, by lowering the ratio of low in-degree and low out-degree nodes, it is possible to reach full controllability of the network. However, this can be costly, since for a given network the number of links that need to be added can be a significant fraction of the initial number of links. Nevertheless, by means of this link-addition process, the number of driver nodes decreases steadily, and for example, in the case considered in Fig. 4, the number of driver nodes is decreased by 50% just by adding 12% of links. Finally, we have measured how other properties of the network change during this procedure, observing that the clustering coefficient does not change significantly while the average distance decreases. Note that this procedure can also be applied to networks with other degree distributions as Poisson networks (see the Supplemental Material [29]).

Conclusions.—We have shown that the structural controllability of a network depends strongly on the fraction of low in-degree and low out-degree nodes. For any uncorrelated directed network with given in-degree and out-degree distribution, the minimum fraction of driver nodes is zero, i.e., $n_D = 0$, if the in degrees and the out degrees of all nodes are both greater than two. For the relevant class of networks with power-law degree distribution, the number of driver nodes can change dramatically by changing the fraction of nodes with in degree and out degree equal to one or two. Finally, we have proposed a strategy for improving the structural controllability of networks by adding links to low degree nodes. Since studying the controllability of real networks is essential for drug design, business applications, and studying the stability of financial markets, we believe that our results will improve the understanding of controllability in such systems.

[1] J.-J. Slotine and W. Li, *Applied Nonlinear Control* (Prentice-Hall, Englewood Cliffs, NJ, 1991).
[2] X. F. Wang and G. Chen, Physica (Amsterdam) **310A**, 521 (2002).
[3] F. Sorrentino, M. di Bernardo, F. Garofalo, and G. Chen, Phys. Rev. E **75**, 046103 (2007).
[4] R. Gutiérrez, I. Sendiña-Nadal, M. Zanin, D. Papo, and S. Boccaletti, Sci. Rep. **2**, 396 (2007).
[5] C.-T. Lin, IEEE Trans. Autom. Control **19**, 201 (1974).
[6] Y.-Y. Liu, J.-J. Slotine, and A.-L. Barabási, Nature (London) **473**, 167 (2011).
[7] M. Pósfai, Y.-Y. Liu, J.-J. Slotine, and A.-L. Barabási, Sci. Rep. **3**, 1067 (2013).
[8] W.-X. Wang, X. Ni, Y.-C. Lai, and C. Grebogi, Phys. Rev. E **85**, 026115 (2012).
[9] G. Yan, J. Ren, Y.-C. Lai, C. H. Lai, and B. Li, Phys. Rev. Lett. **108**, 218703 (2012).
[10] T. Nepusz and T. Vicsek, Nat. Phys. **8**, 568 (2012).
[11] P. Csermely, T. Korcsmáros, H. J. M. Kiss, G. London, and R. Nussinov, Pharmacol. Therapeut. **138**, 333 (2013).
[12] E. Bullmore and O. Sporns, Nat. Rev. Neurosci. **10**, 186 (2009).
[13] P. Bonifazi, M. Goldin, M. A. Picardo, I. Jorquera, A. Cattani, G. Bianconi, A. Represa, Y. Ben-Ari, and R. Cossart, Science **326**, 1419 (2009).
[14] J. D. Power *et al.*, Neuron **72**, 665 (2011).
[15] D. Delpini, S. Battiston, M. Riccaboni, G. Gabbi, F. Pammolli, and G. Caldarelli, Sci. Rep. **3**, 1626 (2012).
[16] R. Albert and A.-L. Barabási, Rev. Mod. Phys. **74**, 47 (2002).
[17] M. E. J. Newman, SIAM Rev. **45**, 167 (2003).
[18] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D.-U. Hwang, Phys. Rep. **424**, 175 (2006).
[19] G. Caldarelli, *Scale-Free Networks: Complex Webs in Nature and Technology* (Oxford University Press, Oxford, 2007).
[20] S. N. Dorogovtsev, A. V. Goltsev, and J. F. F. Mendes, Rev. Mod. Phys. **80**, 1275 (2008).
[21] A. Barrat, M. Barthélemy, and A. Vespignani, *Dynamical Processes on Complex Networks* (Cambridge University Press, Cambridge, England, 2008).
[22] L. Zdeborová and M. Mézard, J. Stat. Mech. (2006) P05003.
[23] F. Altarelli, A. Braunstein, A. Ramezanpour, and R. Zecchina, Phys. Rev. Lett. **106**, 190601 (2011).
[24] M. Mézard and, G. Parisi, Eur. Phys. J. B **20**, 217 (2001).
[25] O. C. Martin, R. Monasson, and R. Zecchina, Theor. Comput. Sci. **265**, 3 (2001).
[26] A. K. Hartmann and M. Weigt, *Phase Transitions in Combinatorial Optimization Problems* (Wiley-VCH, Weinheim, 2005).
[27] M. Mézard and A. Montanari, *Information, Physics, and Computation* (Oxford University Press, Oxford, 2009).
[28] T. Jia, Y.-Y. Liu, E. Csóka, M. Pósfai, J.-J. Slotine, and A.-L. Barabási, Nat. Commun. **4**, 1 (2013).
[29] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevLett.113.078701, which includes Refs. [30–32], for the theoretical derivation of the BP/MS equations, the stability criterion given by Eq. (6), further details on the controllability of networks with degree distributions different from Eq. (7), and details on the numerical results.
[30] C. Lucibello and F. Ricci-Tersenghi, J. Stat. Mech. (2014) 136829.
[31] T. Castellani, F. Krzakala, and F. Ricci-Tersenghi, Eur. Phys. J. B **47**, 99 (2005).
[32] J. E. Hopcroft and R. M. Karp, SIAM J. Comput. **2**, 225231 (1973).
[33] A. Montanari and F. Ricci-Tersenghi, Eur. Phys. J. B **33**, 339 (2003).
[34] O. Rivoire, G. Biroli, O. C. Martin, and M. Mézard, Eur. Phys. J. B **37**, 55 (2003).
[35] M. Boguñá, R. Pastor-Satorras, and A. Vespignani, Eur. Phys. J. B **38**, 205 (2004).
[36] H. Seyed-allaei, G. Bianconi, and M. Marsili, Phys. Rev. E **73**, 046113 (2006).