

Parsimonious Module Inference in Large Networks

Tiago P. Peixoto*

Institut für Theoretische Physik, Universität Bremen, Hochschulring 18, D-28359 Bremen, Germany
(Received 19 December 2012; published 5 April 2013; publisher error corrected 5 April 2013)

We investigate the detectability of modules in large networks when the number of modules is not known in advance. We employ the minimum description length principle which seeks to minimize the total amount of information required to describe the network, and avoid overfitting. According to this criterion, we obtain general bounds on the detectability of any prescribed block structure, given the number of nodes and edges in the sampled network. We also obtain that the maximum number of detectable blocks scales as \sqrt{N} , where N is the number of nodes in the network, for a fixed average degree $\langle k \rangle$. We also show that the simplicity of the minimum description length approach yields an efficient multilevel Monte Carlo inference algorithm with a complexity of $O(\tau N \log N)$, if the number of blocks is unknown, and $O(\tau N)$ if it is known, where τ is the mixing time of the Markov chain. We illustrate the application of the method on a large network of actors and films with over 10^6 edges, and a dissortative, bipartite block structure.

DOI: [10.1103/PhysRevLett.110.148701](https://doi.org/10.1103/PhysRevLett.110.148701)

PACS numbers: 89.75.Hc, 02.50.Tt, 89.70.Cf

The detection of modules—or communities—is one of the most intensely studied problems in the recent literature of network systems [1,2]. The use of generative models for this purpose, such as the stochastic blockmodel family [3–20], has been gaining increasing attention. This approach contrasts drastically with the majority of other methods thus far employed in the field (such as modularity maximization [21]), since not only is it derived from first principles, but also it is not restricted to purely assortative and undirected community structures. However, most inference methods used to obtain the most likely blockmodel assume that the number of communities is known in advance [14,18,22–25]. Unfortunately, in most practical cases this quantity is completely unknown, and one would like to infer it from the data as well. Here we explore a very efficient way of obtaining this information from the data, known as the minimum description length principle (MDL) [26,27], which predicates that the best choice of model which fits given data is the one which most compresses it, i.e., minimizes the total amount of information required to describe it. This approach has been introduced in the task of blockmodel inference in Ref. [28]. Here, we generalize it to accommodate an arbitrarily large number of communities, and to obtain general bounds on the detectability of arbitrary community structures. We also show that, according to this criterion, the maximum number of detectable blocks scales as \sqrt{N} , where N is the number of nodes in the network. Since the MDL approach results in a simple penalty on the log-likelihood, we use it to implement an efficient multilevel Monte Carlo algorithm with an overall complexity of $O(\tau N \log N)$, where τ is the average mixing time of the Markov chain, which can be used to infer arbitrary block structures on very large networks.

The model.—The stochastic blockmodel ensemble is composed of graphs with N nodes, each belonging to one

of B blocks, and the number of edges between nodes of blocks r and s is given by the matrix e_{rs} (or twice that number if $r = s$). The degree-corrected variant [14] further imposes that each node i has a degree given by k_i , where the set $\{k_i\}$ is an additional parameter set of the model. The directed version of both models is analogously defined, with e_{rs} becoming asymmetric, and $\{k_i^-\}$ together with $\{k_i^+\}$ fixing the in- and out-degrees of the nodes, respectively. These ensembles are characterized by their microcanonical entropy $S = \ln \Omega$, where Ω is the total number of network realizations [29]. The entropy can be computed analytically in both cases [30],

$$S_t \cong E - \frac{1}{2} \sum_{rs} e_{rs} \ln \left(\frac{e_{rs}}{n_r n_s} \right), \quad (1)$$

for the traditional blockmodel ensemble and,

$$S_c \cong -E - \sum_k N_k \ln k! - \frac{1}{2} \sum_{rs} e_{rs} \ln \left(\frac{e_{rs}}{e_r e_s} \right), \quad (2)$$

for the degree corrected variant, where in both cases $E = \sum_{rs} e_{rs}/2$ is the total number of edges, n_r is the number of nodes which belong to block r , and N_k is the total number of nodes with degree k , and $e_r = \sum_s e_{rs}$ is the number of half-edges incident on block r . The directed case is analogous [30] (see Supplemental Material [31] for an overview).

The detection problem consists in obtaining the block partition $\{b_i\}$ which is the most likely, when given an unlabeled network G , where b_i is the block label of node i . This is done by maximizing the log-likelihood $\ln \mathcal{P}$ that the network G is observed, given the model compatible with a chosen block partition. Since we have simply $\mathcal{P} = 1/\Omega$, maximizing $\ln \mathcal{P}$ is equivalent to minimize the entropy $S_{t/c}$, which is the language we will use henceforth. Entropy minimization is well defined, but only

as long as the total number of blocks B is known beforehand. Otherwise, the optimal value of $\mathcal{S}_{t/c}$ becomes a strictly decreasing function of B . Thus, simply minimizing the entropy will lead to the trivial $B = N$ partition, and the block matrix e_{rs} becomes simply the adjacency matrix. A principled way of avoiding such overfitting is to consider the total amount of information necessary to describe the data, which includes not only the entropy of the fitted model, but also the information necessary to describe the model itself. This quantity is called the description length, and for the stochastic blockmodel ensemble it is given by

$$\Sigma_{t/c} = \mathcal{S}_{t/c} + \mathcal{L}_{t/c}, \quad (3)$$

where $\mathcal{L}_{t/c}$ is the information necessary to describe the model via the e_{rs} matrix and the block assignments $\{b_i\}$. The minimum value of $\Sigma_{t/c}$ is an upper bound on the total amount of information necessary to describe a given network to an observer lacking any *a priori* information [28]. Therefore, the best model chosen is the one which best compresses the data, which amounts to an implementation of Occam's razor. For the specific problem at hand, it is easy to compute $\mathcal{L}_{t/c}$. The e_{rs} matrix can be viewed as the adjacency matrix of a multigraph with B nodes and E edges, where the blocks are the nodes and self-loops are allowed. The total number of e_{rs} matrices is then simply $\binom{B+E}{E}$ [32]. The total number of block partitions is B^N . Assuming no prior information on the model, we obtain \mathcal{L}_t by multiplying these numbers and taking the logarithm,

$$\mathcal{L}_t \cong Eh\left(\frac{B(B+1)}{2E}\right) + N \ln B, \quad (4)$$

where $h(x) = (1+x)\ln(1+x) - x \ln x$, and $E \gg 1$ was assumed. Note that Eq. (4) is not the same as the expression derived in Ref. [28], which is obtained by taking the limit $E \gg B^2$, in which case we have $\mathcal{L}_t \approx \frac{B(B+1)}{2} \ln E + N \ln B$ [33]. We do not take this limit *a priori*, since, as we show below, block sizes up to $B_{\max} \sim \sqrt{E}$ can in principle be detected from empirical data. For the degree-corrected variant, we still need to describe the degree sequence of the network; hence,

$$\mathcal{L}_c = \mathcal{L}_t - N \sum_k p_k \ln p_k, \quad (5)$$

where p_k is the fraction of nodes with degree k . Note that for the directed case we need simply to replace $B(B+1)/2 \rightarrow B^2$ and $k \rightarrow (k^-, k^+)$ in the equations above.

MDL bound on detectability.—The difference $\Sigma_b \equiv \Sigma_{t/c} - \Sigma_{t/c}|_{B=1}$ of the description length of a graph with some block structure and a random graph with $B = 1$ can be written as

$$\Sigma_b = Eh\left(\frac{B(B+1)}{2E}\right) + N \ln B - EI_{t/c}, \quad (6)$$

with $I_t = \sum_{rs} m_{rs} \ln(m_{rs}/w_r w_s)$ and $I_c = \sum_{rs} m_{rs} \times \ln(m_{rs}/m_r m_s)$, where $m_{rs} = e_{rs}/2E$ and $w_r = n_r/N$ [and equivalently for directed graphs, with $B(B+1)/2 \rightarrow B^2$]. We note that $I_{t/c} \in [0, \ln B]$. If for any given graph we have $\Sigma_b > 0$, the inferred block structure will be discarded in favor of the simpler fully random $B = 1$ model. Therefore, the condition $\Sigma_b < 0$ yields a limit on the detectability of prescribed block structures according to the MDL criterion. For the special case where $E \gg B^2$, this inequality translates to a more convenient form,

$$\langle k \rangle > \frac{2 \ln B}{I_{t/c}}. \quad (7)$$

The directed case is analogous, with $2 \ln B \rightarrow \ln B$ replaced in the equation above.

Partial detectability and parsimony.—The condition $\Sigma_b < 0$ is not a statement on the absolute detectability of a given model, only to what extent the extracted information (if any) can be used to compress the data. Although these are intimately related, the MDL criterion is based on the idea of perfect (or lossless) compression, and thus corresponds simply to a condition necessary (but not sufficient) for the perfect recoverability of the model parameters from the data. Perfect inference, however, is only possible in the asymptotically dense case $\langle k \rangle \rightarrow \infty$ [18], and in practice one always has some amount of uncertainty. Therefore, it remains to be determined how practical is the parsimony limit derived from MDL to establish a noise threshold on empirical data. In Fig. 1, is shown an example of a block structure with $B = 10$ and $I_t = \ln B/6$. In Fig. 1(b), is shown the minimum of Σ_b/E as a function of B , for sampled networks with different $\langle k \rangle$, obtained with the Monte Carlo algorithm described below. If $\langle k \rangle$ is large enough [$\langle k \rangle > 6$, according to Eq. (7)], the minimum of Σ_b is clearly at the correct $B = 10$ value, and, as is shown in Fig. 1(b), this is exactly where the normalized mutual information (NMI) [34] between the known and inferred partition is the largest. However, for $\langle k \rangle < 6$ the minimum of Σ_b is no longer at $B = 10$, and instead it is at $B = 1$. Nevertheless, the overlap with the correct partition is overall positive and is still the largest at $B = 10$, so the correct partition is to some extent detectable, but the MDL criterion rejects it. By experimenting with different planted block structures [see Fig. 1(d)], one observes that the MDL threshold lies very close to the parameter region where inferred partition is no longer well correlated with the true partition. This comparison can be made in more detail by considering the special case known as the planted partition model (PP) [35], which imposes a diagonal block structure given by $m_{rr} = c/B$, $m_{rs} = (1-c)/B(B-1)$ for $r \neq s$, and $w_r = 1/B$, and $c \in [0, 1]$ is a free parameter. In this case, it can be shown that even partial inference is only possible if $\langle k \rangle > [(B-1)/(cB-1)]^2$ [17, 18, 36, 37], otherwise no information at all on the original model can be extracted [38]. For smaller values of B , this bound is higher than Eq. (7) for

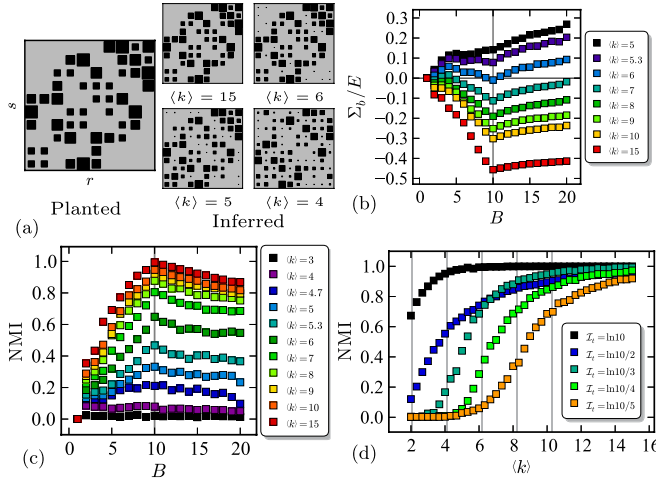


FIG. 1 (color online). (a) Prescribed block structure with $B = 10$ and $I_t = \ln B/6$, together with inferred parameters for different $\langle k \rangle$; (b) description length Σ_b/E for different B and $\langle k \rangle$, for networks sampled from (a). The vertical line marks the position of the global minimum; (c) NMI between the true and inferred partitions, for the same networks as in (b); (d) the same as (c), but for different $\langle k \rangle$ and prescribed block structures. The grey lines correspond to the threshold of Eq. (7). In all cases we have $N = 10^4$. The legend ordering matches the curve order in (b) and (d), and the inverse curve order in (c).

this model [where we have $I_{t/c} = c \ln(Bc) + (1 - c) \times \ln(B(1 - c)/(B - 1))$], which means that there is a region of parameters where the MDL criterion discards potentially useful (albeit clearly noisy) information [see Fig. 2(a)]. Interestingly, however, for larger values of B , the MDL criterion will most often result in lower bounds [see Fig. 2(b)], meaning that whatever partial information which can be recovered from the model will not be discarded. For $B \rightarrow \infty$ we have $c_{\text{MDL}}^* \approx 2/\langle k \rangle$ and $c^* \approx 1/\sqrt{\langle k \rangle}$, and thus, $c_{\text{MDL}}^* < c^*$ for $\langle k \rangle > 4$ [41]. Therefore, so far as the PP model serves as a good representation of more general block structures, one should not expect excessive parsimony from MDL, at least for sufficiently large values of B .

The largest detectable value of B .—The MDL approach imposes an intrinsic constraint on the maximum value of B

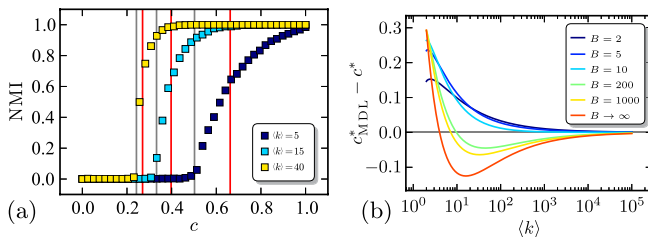


FIG. 2 (color online). (a) NMI between the true and inferred partitions for PP samples with $B = 10$ as a function of c for different $\langle k \rangle$. The grey (red) lines correspond to the threshold c^* of Ref. [17] [c_{MDL}^* given by Eq. (7)]; (b) difference between c_{MDL}^* and c^* , for different $\langle k \rangle$ and B .

which can be detected, B_{max} , given a network size and density. This can be obtained by minimizing Σ_b over all possible block structures with a given B , which is obtained simply by replacing $I_{t/c}$ by its maximum value $\ln B$ in Eq. (6),

$$\Sigma'_b = Eh\left(\frac{B(B+1)}{2E}\right) - (E - N) \ln B. \quad (8)$$

Equation (8) is a strictly convex function on B . This means there is a global minimum $\Sigma'_b|_{B=B_{\text{max}}}$ given uniquely by N and E . It is easy to see that even if the prescribed block structure with some $B > B_{\text{max}}$ has minimal entropy (i.e., $I_{t/c} = \ln B$), alternative partitions with $B' < B$ blocks (obtained by merging blocks such that $I'_{t/c} = \ln B'$) will necessarily possess a smaller Σ'_b . Imposing $\partial \Sigma'_b / \partial B = 0$, one obtains $B_{\text{max}} \cong \mu(\langle k \rangle) \sqrt{E}$, with $\mu(\langle k \rangle)$ being the solution of $\mu \ln(2/\mu^2 + 1) - (1 - 1/\langle k \rangle)/\mu = 0$ [for the directed case we make $2/\mu^2 \rightarrow 1/\mu^2$ and $1/\langle k \rangle \rightarrow 2/\langle k \rangle$]. Therefore, according to the MDL criterion, the maximum number of blocks which is detectable scales as $B_{\text{max}} \sim \sqrt{N}$ for a fixed value of $\langle k \rangle$. This is consistent with the detectability analysis in Ref. [42] for the traditional blockmodel variant, which showed by other means that the model parameters can only be recovered if B does not scale faster than \sqrt{N} . Note that this means that the limit $E \gg B^2$ cannot be taken *a priori* when inferring from empirical data, and hence, the value of \mathcal{L}_t computed in Ref. [28] needs to be replaced with Eq. (4) in the general case.

The limit $B_{\text{max}} \propto \sqrt{E}$ is very similar to the so-called “resolution limit” of community detection via modularity optimization [43], which is $B_{\text{max}}^Q = \sqrt{E}$. These two limits, however, have different interpretations: the value of B_{max}^Q arises simply from the definition of modularity, which can be to some extent alleviated (but not entirely avoided) by properly modifying the modularity function with scale parameters [44–49]. On the other hand, the value of B_{max} has a more fundamental character, and corresponds to the situation where knowledge of the complete block structure is no longer the best option to compress the data. This value can be improved only if any *a priori* information is known which leads to a smaller class of models to be inferred, and hence, smaller \mathcal{L}_t . In general, if we have $\mathcal{L}_t = Ef(B^\alpha/E) + N \ln B$, where $f(x)$ is any (differentiable) function, performing the same analysis as above leads to $B_{\text{max}} = [\mu(k)E]^{1/\alpha}$, with $\alpha f'(\mu)\mu + 2/\langle k \rangle - 1 = 0$. However, it should also be noted that if the existing block structure is locally dense (i.e., $e_{rs} \sim n_r n_s$), as the union of B complete graphs considered in Ref. [43], the expressions in Eqs. (1) and (2) are no longer valid, and will overestimate the entropy. Using the correct entropy {Eqs. (5) and (9) in Ref. [30]} will lead to an improved resolution. Unfortunately, for the dense case, the entropy for the degree-corrected variant cannot be computed in a closed form [30].

Detection algorithm.—For a fixed B , the best partition can be found by minimizing $S_{t/c}$, via well-established

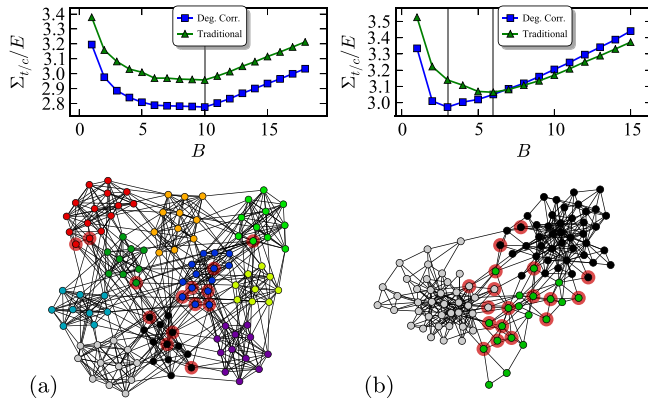


FIG. 3 (color online). Top: Value of Σ_b/E for both blockmodel variants as a function of B for (a) the American football network of Ref. [54] (with the corrections described in Refs. [57,58]) and (b) the political books network of Ref. [55]. Bottom: Inferred partitions with the smallest Σ_b . Nodes circled in red do not match the known partitions.

methods such as the Markov chain Monte Carlo method, using the Metropolis-Hastings algorithm [50,51]. However, a naïve implementation based on fully random block membership moves can be very slow. We found that the performance can be drastically improved by using local information and current knowledge of the partially inferred block structure, simply by proposing moves $r \rightarrow s$ with a probability $p(r \rightarrow s|t) \propto e_{ts} + 1$, where t is the block label of a randomly chosen neighbor of the node being moved. Each sweep of this algorithm can be performed in $O(E)$ time, independent of B (see Supplemental Material [31]). Having obtained the minimum of $\mathcal{S}_{t/c}$, the best value of B is obtained via an independent one-dimensional minimization of Σ_b , using a Fibonacci search [52], based on subsequent bisections of an initial interval which brackets the minimum. This method finds a local minimum in $O(\ln B_{\max})$ time. The overall number of steps necessary for the entire algorithm is $O(\tau E \ln B_{\max})$, where τ is the average mixing time of the Markov chain. If we have no prior information on B_{\max} , we need to assume $B_{\max} \sim \sqrt{E}$, in which case the complexity becomes $O(\tau E \ln E)$, or $O(\tau N \ln N)$ for sparse graphs. This compares favorably to minimization strategies which require the computation of the full marginal probability π_r^i that node i belongs to block r , such as belief propagation [17,18,53], which results in a larger complexity of $O(NB^2)$ per sweep (or $O(NB^2 l)$ for the degree-corrected variant, with l being the number of distinct degrees [53]), or $O(N^2)$ for $B \sim B_{\max}$.

Empirical networks.—The MDL approach yields convincing results for many empirical networks, as can be seen in Fig. 3, which shows results for the college football network of Ref. [54] and the political books network of Ref. [55]. In both cases the correct number blocks is inferred, and the best partition matches reasonably well the known true values, at least for the degree-corrected

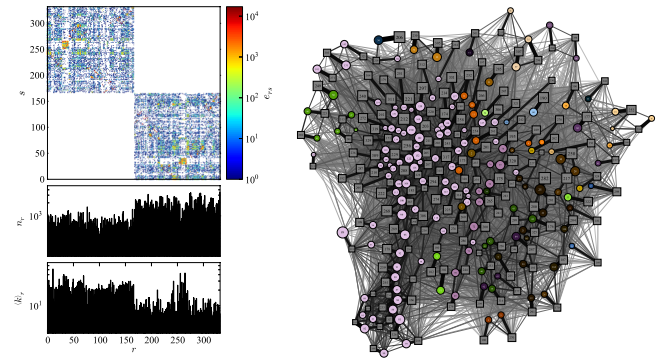


FIG. 4 (color online). Left: Inferred block structure for the IMDB network, with $N = 372, 787$, $E = 1, 812, 657$, and $B = 332$, according to the MDL criterion, and the degree-corrected stochastic blockmodel. Right: Circles correspond to film blocks, and squares to actors. The node colors correspond to the countries of production. See Supplemental Material [31] for more details.

variant. Employing the Monte Carlo algorithm above, results may be obtained for much larger networks. We show in Fig. 4 the obtained block partition with the degree-corrected variant for the Internet Movie Database (IMDB) network of actors and films [56], where a film node is connected to all its cast members. The bipartite nature of the network is fully reflected in the inferred block partition, where films and actors always belong to different blocks, although this has not been imposed *a priori* (something which would be impossible to obtain with, e.g., modularity optimization). Besides this role separation, the film blocks are divided sharply along spatial, temporal and genre lines, and the actor blocks are closely correlated with such film classes (see Supplemental Material [31] for a more detailed analysis).

In summary, we showed how minimizing the full description length of empirical network data enables simple, efficient, unbiased, and fully nonparametric analysis of the large-scale properties of large networks, for which no *a priori* information is available, while at the same time providing general bounds on the detectability of arbitrary block structures from empirical data.

I would like to thank Tim S. Evans for pointing out some corrections to the American football data, and Laerte B.P. de Andrade for useful conversations about the IMDB data.

*tiago@itp.uni-bremen.de

- [1] S. Fortunato, *Phys. Rep.* **486**, 75 (2010).
- [2] M. E. J. Newman, *Nat. Phys.* **8**, 25 (2011).
- [3] P. W. Holland, K. B. Laskey, and S. Leinhardt, *Soc. Netw.* **5**, 109 (1983).
- [4] S. E. Fienberg, M. M. Meyer, and S. S. Wasserman, *J. Am. Stat. Assoc.* **80**, 51 (1985).

- [5] K. Faust and S. Wasserman, *Soc. Netw.* **14**, 5 (1992).
- [6] C.J. Anderson, S. Wasserman, and K. Faust, *Soc. Netw.* **14**, 137 (1992).
- [7] M.B. Hastings, *Phys. Rev. E* **74**, 035102 (2006).
- [8] D. Garlaschelli and M.I. Loffredo, *Phys. Rev. E* **78**, 015101 (2008).
- [9] M.E.J. Newman and E.A. Leicht, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 9564 (2007).
- [10] J. Reichardt and D.R. White, *Eur. Phys. J. B* **60**, 217 (2007).
- [11] J.M. Hofman and C.H. Wiggins, *Phys. Rev. Lett.* **100**, 258701 (2008).
- [12] P.J. Bickel and A. Chen, *Proc. Natl. Acad. Sci. U.S.A.* **106**, 21068 (2009).
- [13] R. Guimerà and M. Sales-Pardo, *Proc. Natl. Acad. Sci. U.S.A.* **106**, 22073 (2009).
- [14] B. Karrer and M.E.J. Newman, *Phys. Rev. E* **83**, 016107 (2011).
- [15] B. Ball, B. Karrer, and M.E.J. Newman, *Phys. Rev. E* **84**, 036103 (2011).
- [16] J. Reichardt, R. Alamino, and D. Saad, *PLoS ONE* **6**, e21282 (2011).
- [17] A. Decelle, F. Krzakala, C. Moore, and L. Zdeborová, *Phys. Rev. Lett.* **107**, 065701 (2011).
- [18] A. Decelle, F. Krzakala, C. Moore, and L. Zdeborová, *Phys. Rev. E* **84**, 066106 (2011).
- [19] Y. Zhu, X. Yan, and C. Moore, [arXiv:1205.7009](https://arxiv.org/abs/1205.7009).
- [20] E.B. Baskerville, A.P. Dobson, T. Bedford, S. Allesina, T.M. Anderson, and M. Pascual, *PLoS Comput. Biol.* **7**, e1002321 (2011).
- [21] M.E.J. Newman and M. Girvan, *Phys. Rev. E* **69**, 026113 (2004).
- [22] Y. Zhao, E. Levina, and J. Zhu, *Ann. Stat.* **40**, 2266 (2012).
- [23] C. Moore, X. Yan, Y. Zhu, J.-B. Rouquier, and T. Lane, in *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '11* (Association of Computing Machinery, New York, 2011), p. 841.
- [24] A. Chen, A.A. Amini, P.J. Bickel, and E. Levina, [arXiv:1207.2340](https://arxiv.org/abs/1207.2340).
- [25] P. Zhang, F. Krzakala, J. Reichardt, and L. Zdeborová, *J. Stat. Mech.* (2012), P12021.
- [26] P.D. Grünwald, *The Minimum Description Length Principle* (MIT, Cambridge, MA, 2007).
- [27] J. Rissanen, *Information and Complexity in Statistical Modeling* (Springer, New York, 2010), 1st ed.
- [28] M. Rosvall and C.T. Bergstrom, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 7327 (2007).
- [29] G. Bianconi, *Phys. Rev. E* **79**, 036114 (2009).
- [30] T.P. Peixoto, *Phys. Rev. E* **85**, 056122 (2012).
- [31] See Supplemental Material at <http://link.aps.org/supplemental/10.1103/PhysRevLett.110.148701> for an overview of the entropy expressions for the different blockmodel ensembles, a full description of the Markov chain Monte Carlo inference algorithm used, and more details on the analysis of the IMDB network.
- [32] Where $\binom{n}{k} = \binom{n+k-1}{k}$ is the number of k -combinations with repetitions from a set of size n .
- [33] The value of \mathcal{L}_t was computed in Ref. [28] as the number of symmetric matrices with entry values from 0 to E , without the restriction that the sum must be exactly $2E$, which is accounted for in Eq. (4). If $E \gg B^2$ this restriction can be neglected, but not otherwise.
- [34] NMI here is defined as $2I(X, Y)/(H(X) + H(Y))$, where $I(X, Y)$ is the mutual information between X and Y , and $H(X)$ is the entropy of X .
- [35] A. Condon and R.M. Karp, *Random Struct. Algorithms* **18**, 116 (2001).
- [36] E. Mossel, J. Neeman, and A. Sly, [arXiv:1202.1499](https://arxiv.org/abs/1202.1499).
- [37] R.R. Nadakuditi and M.E.J. Newman, *Phys. Rev. Lett.* **108**, 188701 (2012).
- [38] In Refs. [17,37] this threshold was expressed with a different notation, as a function of $c_{\text{in}} = \langle k \rangle c B$ and $c_{\text{out}} = \langle k \rangle (1 - c) B / (B - 1)$, instead of c . It is also common to express such transitions as a function of $k_{\text{in}} = c \langle k \rangle$ and $k_{\text{out}} = (1 - c) \langle k \rangle$, or the mixing parameter $\mu = k_i^{\text{out}} / (k_i^{\text{in}} + k_i^{\text{out}})$ [39,40]. For the PP model, we have simply $\mu \approx 1 - c$, for sufficiently large degrees.
- [39] A. Lancichinetti, S. Fortunato, and F. Radicchi, *Phys. Rev. E* **78**, 046110 (2008).
- [40] A. Lancichinetti and S. Fortunato, *Phys. Rev. E* **80**, 056117 (2009).
- [41] Here we impose $E \gg B^2$ first, and $B \rightarrow \infty$ later.
- [42] D.S. Choi, P.J. Wolfe, and E.M. Airolidi, *Biometrika* **99**, 273 (2012).
- [43] S. Fortunato and M. Barthélemy, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 36 (2007).
- [44] J. Reichardt and S. Bornholdt, *Phys. Rev. E* **74**, 016110 (2006).
- [45] J.M. Kumpula, J. Saramaki, K. Kaski, and J. Kertesz, *Fluct. Noise Lett.* **7**, 209 (2007).
- [46] A. Arenas, A. Fernández, and S. Gómez, *New J. Phys.* **10**, 053039 (2008).
- [47] A. Lancichinetti and S. Fortunato, *Phys. Rev. E* **84**, 066122 (2011).
- [48] J. Xiang and K. Hu, [arXiv:1108.4244](https://arxiv.org/abs/1108.4244).
- [49] P. Ronhovde and Z. Nussinov, [arXiv:1208.5052](https://arxiv.org/abs/1208.5052).
- [50] N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, and E. Teller, *J. Chem. Phys.* **21**, 1087 (1953).
- [51] W.K. Hastings, *Biometrika* **57**, 97 (1970).
- [52] W.H. Press, S.A. Teukolsky, W.T. Vetterling, and B.P. Flannery, *Numerical Recipes: The Art of Scientific Computing*, (Cambridge University Press, Cambridge, England, 2007), 3rd ed.
- [53] X. Yan, J.E. Jensen, F. Krzakala, C. Moore, C.R. Shalizi, L. Zdeborova, P. Zhang, and Y. Zhu, [arXiv:1207.3994](https://arxiv.org/abs/1207.3994).
- [54] M. Girvan and M.E.J. Newman, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 7821 (2002).
- [55] V. Krebs, <http://www.orgnet.com>.
- [56] Retrieved from <http://www.imdb.com/interfaces>.
- [57] T.S. Evans, *J. Stat. Mech.* (2010), P12037.
- [58] T.S. Evans, “FigShare” (2012), doi:10.6084/m9.figshare.93179.