Rapid Communications

Counting the number of metastable states in the modularity landscape: Algorithmic detectability limit of greedy algorithms in community detection

Tatsuro Kawamoto[1] and Yoshiyuki Kabashima[2]

[1]*Artificial Intelligence Research Center, National Institute of Advanced Industrial Science and Technology, 2-3-26 Aomi, Koto-ku, Tokyo, Japan*

[2]*Department of Mathematical and Computing Science, Tokyo Institute of Technology, W8-45, 2-12-1 Ookayama, Meguro-ku, Tokyo, Japan*

(Received 23 August 2018; published 17 January 2019)

Modularity maximization using greedy algorithms continues to be a popular approach toward community detection in graphs, even after various better forming algorithms have been proposed. Apart from its clear mechanism and ease of implementation, this approach is persistently popular because, presumably, its risk of algorithmic failure is not well understood. This Rapid Communication provides insight into this issue by estimating the algorithmic performance limit of the stochastic block model inference using modularity maximization. This is achieved by counting the number of metastable states under a local update rule. Our results offer a quantitative insight into the level of sparsity at which a greedy algorithm typically fails.

DOI: 10.1103/PhysRevE.99.010301

Introduction. Since the proposal of the modularity function [1], a number of its maximization algorithms and related objective functions have been put forward, and some of them have been widely applied to the discovery of community structures in real-world networks [2]. Modularity maximization is also known to be equivalent to the maximum likelihood method of a statistical model [3,4]. The corresponding greedy algorithms, such as the Louvain algorithm [5], are commonly used for optimization. However, greedy algorithms have often been employed as baselines in benchmark tests and various better performing algorithms have been proposed. Moreover, from a Bayesian viewpoint [6,7], modularity maximization is known to be suboptimal when a graph is generated from an assumed statistical model, which implies the risk of overfitting [8]. Nevertheless, greedy algorithms remain very popular partly because, presumably, we do not know in which cases we should not expect greedy algorithms to work.

We conducted a theoretical performance analysis to provide insight into this issue. In this Rapid Communication, we considered a random graph model with a planted modular structure, called the *stochastic block model* [9–11], which is a canonical model for a theoretical investigation with regard to community detection. We derive the limit of the model parameters beyond the point at which a greedy algorithm completely loses the ability to identify the planted modular structure. Such a limit is termed as the *algorithmic detectability limit* [12]. The corresponding limits of other algorithms, e.g., spectral clusterings [13–17], and the expectation-maximization algorithm [18,19] have also been investigated. In contrast, the limit where all algorithms fail is known as the information-theoretic limit [20–23]. Such a limit exists because, when a planted modular structure is too weak, the corresponding graph instances can also be typically generated by a uniform random graph model.

Note that, in this Rapid Communication, the structure specified by the planted group assignments is the only community structure defined. Although we consider algorithms that aim to maximize modularity, we do not regard the group assignments that achieve the true maximum as the "real" community structure. (See the Supplemental Material for further discussions [24].)

Benchmark tests are an experimental approach toward investigating the detectability limits [25–28]. Although such tests have the advantage of being conducted in a straightforward manner, a definite conclusion can rarely be obtained. For example, it is usually unclear whether a better implementation of the adopted algorithm can significantly improve performance, or if there is no hope of improvement because the difficulty is inherent in the formulation. By contrast, although theoretical investigations [29,30] with regard to the detectability limit are available only for limited situations, they enable a more concrete understanding regarding the feasibilities and limitations of community detection.

In this study, we considered sparse undirected graphs without self-loops or multiedges. With regard to a planted modular structure, we focused on the community structure, i.e., the assortative structures of two groups. We define a graph as $G = (V, E)$, where V and E are the sets of vertices and edges, respectively. We let $N = |V|$ and the average degree be c. We denote $\sigma \in \{1, 2\}$ as a group label and σ_i as the group assignment of vertex i. We also denote the set of vertices in group σ as V_σ, i.e., $\cup_\sigma V_\sigma = V$ and $\gamma_\sigma \equiv |V_\sigma|/N$.

Stochastic block model. We denote the adjacency matrix of a graph as $\boldsymbol{A}$, where $A_{ij} = 1$ when vertices i and j are connected, and $A_{ij} = 0$ otherwise. The stochastic block model defines the considered probability distribution of the graph configurations, i.e., the graph ensemble. In this model, the vertices of a graph have planted group assignments, and the edges are generated independently and randomly on the basis of these assignments. For example, the connection probability of the pair of vertices i and j being connected is given by $P_{ij}(A_{ij} = 1) = \rho_{\sigma_i\sigma_j}$. Note that, for the graphs to be sparse, we have $\rho_{\sigma\sigma'} = O(N^{-1})$. In community detection, $\boldsymbol{A}$ is the only input and the objective is to infer the hidden group

assignments. A particular case wherein the planted group sizes are equal and the connection probability is parametrized as $\rho_{11} = \rho_{22} = \rho_{\rm in}$ and $\rho_{12} = \rho_{21} = \rho_{\rm out}$ is often referred to as the symmetric stochastic block model. In this case, the strength of the community structure can be parametrized as $\epsilon \equiv \rho_{\rm out}/\rho_{\rm in}$.

Modularity maximization and its detectability. The objective function of modularity for bipartition can be expressed as

$$Q(\boldsymbol{s}) = \sum_{ij} s_i B_{ij} s_j = \text{const} + \sum_{i,j(i\neq j)} s_i B_{ij} s_j, \tag{1}$$

where $s_i \in \{-1, +1\}$ is a *spin* variable representing the group assignment of vertex $i \in V$, and matrix $\boldsymbol{B}$ is defined as

$$B_{ij} \equiv A_{ij} - \alpha c_i c_j, \tag{2}$$

where c_i is the degree of vertex i defined as $c_i = \sum_j A_{ij}$, α is an $O(N^{-1})$ scaling parameter given as the input, and α is called the resolution parameter. For a given adjacency matrix $\boldsymbol{A}$, the set of most plausible group assignments $\boldsymbol{s} = \{s_1, \ldots, s_N\}$ is obtained as that maximizing Eq. (1).

Here, we consider the update of group assignments $\boldsymbol{s}$ by a single spin flip, i.e., we may flip only one component s_i at each update. Note that the global maximum of $Q(\boldsymbol{s})$ may not be achieved by a single spin flip, owing to the existence of *metastable states*. We define the metastable state as a spin configuration $\boldsymbol{s}$ such that $Q(\boldsymbol{s})$ does not increase by any single spin flip.

An intuitive understanding of the algorithmic detectability limit is presented below. When Eq. (1) does not have metastable states, i.e., local maxima and saddle points, a local update algorithm is able to find its global maximum by starting from an arbitrary random initial state of group assignments. Even when metastable states exist, unless their number is sufficiently large, a local update algorithm can still achieve the global maximum of Eq. (1) by repeating the algorithm with various initial states. However, when the number of metastable states grows exponentially with respect to N, it is practically impossible to achieve the global maximum, because a repeated search from an extremely large number of initial states is required. Therefore, the detectability limit can be evaluated by counting the number of metastable states. To illustrate such a situation, the modularity landscape near the global optimum for the small stochastic block model is shown in Fig. 1 for $\epsilon = 0.04$ and $\epsilon = 0.4$, respectively. As the planted modular structure becomes less clear (larger ϵ), the landscape becomes more ragged. In fact, in many real-world networks, modularity landscapes are often very ragged near their global maximum [32].

Number of metastable states. The variation of the objective function $\Delta Q(s_i)$ caused by the spin flip with respect to s_i reads as

$$\begin{aligned}\Delta Q(s_i) &= (-s_i) \sum_{j(\neq i)} B_{ij} s_j - s_i \sum_{j(\neq i)} B_{ij} s_j \\ &= -2 s_i \sum_{j(\neq i)} B_{ij} s_j. \end{aligned} \tag{3}$$

Thus, the metastable state is the spin configuration $\boldsymbol{s}$ such that $\Delta Q(s_i) \leqslant 0$ for all i. In other words, it is either $s_i = \text{sgn}(\sum_{j(\neq i)} B_{ij} s_j)$ or $\sum_{j(\neq i)} B_{ij} s_j = 0$. This condition can also be expressed such that there exists a non-negative value of λ_i for each $i \in V$ such that

$$\lambda_i = s_i \sum_{j(\neq i)} B_{ij} s_j. \tag{4}$$

Based on the observation expressed in Eq. (4), the number of metastable states $\mathcal{N}_m$ can be counted as follows [33],

$$\mathcal{N}_m = \sum_{\{s_i\}} \prod_i \int_0^\infty d\lambda_i \, \delta\left(\lambda_i - s_i \sum_{j(\neq i)} B_{ij} s_j\right), \tag{5}$$

where $\delta(\cdot)$ is Dirac's delta function.

We are interested in the typical number of metastable states within the graph ensemble, rather than a single graph instance. To this end, we estimate the configuration average of graphs generated by the stochastic block model in the limit of $N \to \infty$, which we denote as $[\mathcal{N}_m]_A$. However, its exact calculation is technically difficult. Therefore, we adopt the *rotating-wave approximation* [34] or low-frequency approximation for the contribution from the delta functions in Eq. (5). By conducting the calculation described in the Supplemental Material [24],

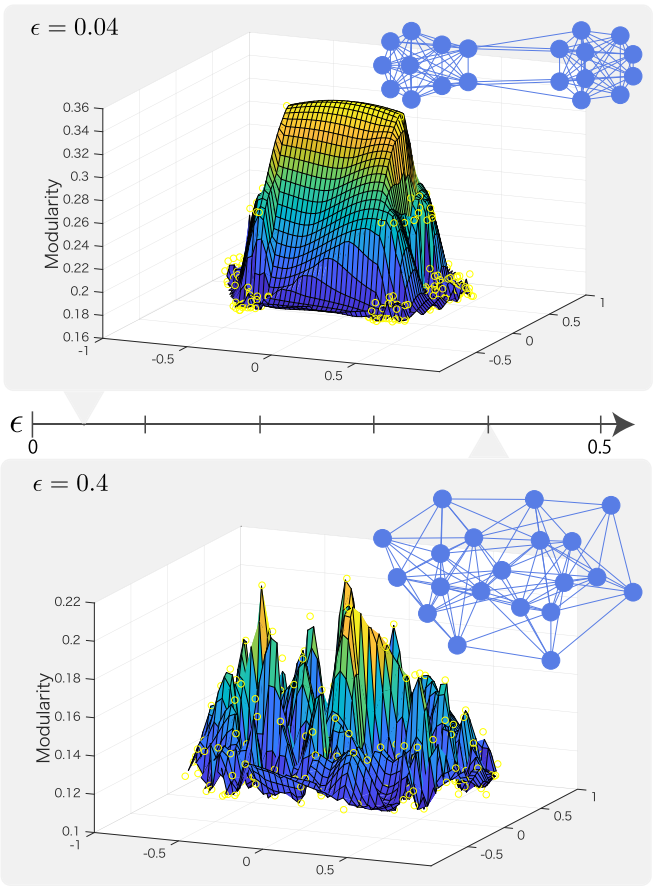

FIG. 1. Modularity landscapes and network figures for instances of the small stochastic block model with equal group sizes ($N = 20$). The average degrees and strengths of the modular structures are ($c = 9.6, \epsilon = 0.04$) (top) and ($c = 9.2, \epsilon = 0.4$) (bottom), respectively. The landscapes were drawn using the code of the curvilinear component analysis distributed by Ref. [31] (see Ref. [32] for a detailed description of the visualization).

we arrive at the following expression,

$$[\mathcal{N}_m]_A \sim e^{Nf}. \tag{6}$$

Here, instead of the result obtained for a general case (found in the Supplemental Material [24]), we show a compact expression obtained by considering the symmetric stochastic block model and adopting an approximation such that the graph is regular, i.e., the degree is constant for all vertices. Additionally, we let the resolution parameter be $\alpha N = 1/c$, which is often employed as the "standard value." In such a case, the following relationship holds,

$$f = -\frac{1}{2c}\hat{E}_1^2 - \frac{1}{c}\frac{1+\epsilon}{1-\epsilon}\tilde{G}_1\tilde{F}_1 + \log Z_1. \tag{7}$$

The subscript 1 indicates that the variables are values for $\sigma = 1$. Because of symmetry, the magnitudes of the variables for $\sigma = 2$ are equal to those of $\sigma = 1$. In Eq. (7), Z_σ is a function of $\hat{E}_\sigma$, $\tilde{F}_\sigma$, and $\tilde{G}_\sigma$,

$$Z_\sigma = \sum_s e^{s\tilde{F}_\sigma}\Phi\left(\frac{1}{\sqrt{c}}(\hat{E}_\sigma + s\tilde{G}_\sigma)\right), \tag{8}$$

where $\Phi(\cdot)$ is the standard normal cumulative distribution function,

$$\Phi(y) \equiv \int_{-\infty}^{y}\frac{dx}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}. \tag{9}$$

Therefore, the number of metastable states can be evaluated if $\hat{E}_1$, $\tilde{F}_1$, and $\tilde{G}_1$ are determined. According to the saddle-point conditions, these variables are evaluated by the following self-consistent equations,

$$\hat{E}_1 = \frac{2c}{Z_1\sqrt{2\pi c}}\exp\left[-\frac{1}{2c}\left(\hat{E}_1^2 + \tilde{G}_1^2\right)\right]\cosh(\tilde{F}_1 - c^{-1}\hat{E}_1\tilde{G}_1), \tag{10}$$

$$\tilde{F}_1 = \frac{\mathcal{G}}{Z_1\sqrt{2\pi c}}\exp\left[-\frac{1}{2c}\left(\hat{E}_1^2 + \tilde{G}_1^2\right)\right]\sinh(\tilde{F}_1 - c^{-1}\hat{E}_1\tilde{G}_1), \tag{11}$$

$$\tilde{G}_1 = \frac{\mathcal{G}}{2Z_1}\sum_s s\, e^{s\tilde{F}_1}\Phi\left(\frac{1}{\sqrt{c}}(\hat{E}_1 + s\tilde{G}_1)\right), \tag{12}$$

where $\mathcal{G} \equiv 2c(1-\epsilon)/(1+\epsilon) = (\rho_{\text{in}} - \rho_{\text{out}})N$.

Detectability limit of a simple greedy algorithm. From Eq. (7), it is evident that the graphs have an exponentially large number of metastable states as long as $f > 0$. Otherwise, they only have a subexponential number of metastable states. Thus, the detectability limit is located at the value of ϵ^* where

$$c\log Z_1 = \frac{1}{2}\hat{E}_1^2 + \frac{1+\epsilon^*}{1-\epsilon^*}\tilde{G}_1\tilde{F}_1 \tag{13}$$

is satisfied.

The accuracy of our estimate is shown in Fig. 2. Here, we consider a simple greedy algorithm, wherein the vertex to be updated is chosen randomly and its spin s_i variable is flipped if $\Delta Q(s_i) > 0$. This algorithm is exactly the process considered in metastable state counting. The detectability phase diagram of this algorithm is shown in Fig. 2(a) as a density plot,

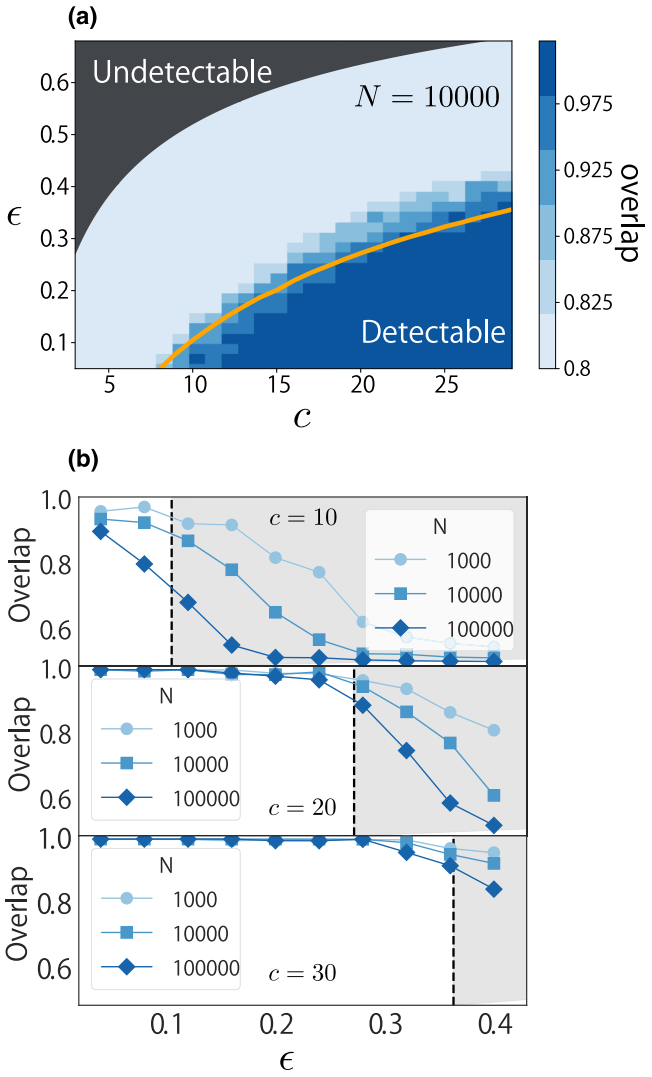

FIG. 2. (a) Detectability phase diagram of a simple greedy algorithm for the symmetric stochastic block model with $N = 10\,000$. The density plot represents the overlaps, while the solid yellow line represents our detectability limit estimate. The shaded region at the upper-left corner represents the region where the detection is information-theoretically impossible. (b) Overlaps of $c = 10$ (top), $c = 20$ (middle), and $c = 30$ (bottom) as functions of ϵ for different graph sizes N. The shaded region with the dashed border line represents our undetectable region estimate. In all plots, the average overlap value of 100 graph instances was determined for each pair of c and ϵ values.

and is obtained by executing the algorithm for the graphs generated by the stochastic block model with various values of the average degree c and strength of community structure ϵ. The color depth represents the *overlap*, which is defined as the fraction of vertices correctly assigned to the planted groups, i.e., $\max\{\sum_i(1+s_it_i)/2N, 1-\sum_i(1+s_it_i)/2N\}$, where t_i is the planted group assignment such that $\{t_i = +1|\sigma_i = 1\}$ and $\{t_i = -1|\sigma_i = 2\}$. The minimum overlap is 0.5 and is achieved when the group assignments are determined in a completely random manner. Owing to the finite size effect, the overlap only gradually decreases around the estimate of the detectability limit (solid yellow line). However, as shown

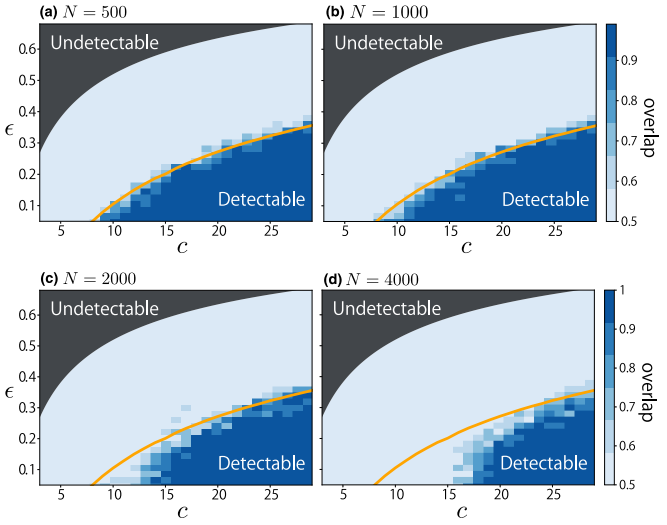

FIG. 3. Detectability phase diagrams of Louvain algorithm for a symmetric stochastic block model with (a) $N = 500$, (b) $N = 1000$, (c) $N = 2000$, and (d) $N = 4000$, respectively. Plotting was carried out in the same manner as that shown in Fig. 2. The overlap is set to 0.5 whenever the graph is partitioned into more than two groups. In all plots, the average overlap value of ten graph instances was determined for each c and ϵ pair.

in Fig. 2(b), when the average degree is sufficiently high, the overlap decreases more sharply as N increases, which implies that our estimate is accurate in the limit of $N \to \infty$. In the case of low average degrees, our result appears overestimated, likely because of the adopted approximations.

The notion of a metastable state is algorithm dependent because it is defined with respect to a single spin flip. However, it is doubtful whether other update rules, such as cluster updates (i.e., multispin flips), may significantly improve performance in the case where the single-spin-flip algorithm (simple greedy algorithm) has a highly ragged modularity landscape. Therefore, it is worth comparing our estimated detectability limit with more sophisticated greedy algorithms.

Detectability limit of Louvain algorithm. The Louvain algorithm [5] is a widely used greedy heuristic for modularity maximization (see Ref. [5] for details regarding this algorithm). For the specific implementation, we used the code distributed at Ref. [35]. The Louvain algorithm does not exactly correspond to the situation that we considered in the metastable state counting. First, the number of groups is determined automatically during the optimization process. Second, the Louvain algorithm contains multispin updates or cluster updates.

The detectability phase diagrams of the Louvain algorithm are shown in Fig. 3 as density plots. When the algorithm identifies more than two groups, we set the overlap to 0.5.

Interestingly, when the graph size N is not very large, the detectability limit estimated by Eq. (13) (solid yellow line) coincides with the phase boundary of the region where the overlap is greater than 0.5, although the detectable region of the lower average degrees decreases as N increases. To the extent of our investigation, the detectable region did not exceed Eq. (13). This experimental observation implies that our estimate of the detectability limit is an intrinsic upper bound of modularity maximization, which holds more generally for greedy algorithms than for the single-spin-flip algorithm. The same analysis was carried out for the so-called *fast greedy algorithm*, as presented in the Supplemental Material [24].

Discussion. Greedy algorithms have simple mechanisms and are relatively easy to implement. However, it is known that modularity maximization using a greedy algorithm is not optimal for inferring the stochastic block model. Here, we conducted a quantitative investigation with regard to this algorithm's feasibility and limitations. Our result indicates that the algorithm fails for a considerably large region in the parameter space of the stochastic block model, even when the corresponding graphs have statistically significant structures. Note that we never focused on the true maximum of modularity; whether the partition with the maximum modularity is correlated to the planted partition is a very different problem and is not of our interest at all.

Most importantly, our result indicates that greedy algorithms are expected to fail when a graph has a sufficiently low average degree, *regardless of the modular structure's strength*. In the case of the symmetric stochastic block model, our approximated estimation predicted that this happens when $c \lesssim 7$. Although this value is not very accurate, our analysis successfully explains the experimentally observed limitations of the greedy algorithms in a qualitative manner. Thus, a quantitative insight into the limited utility of greedy algorithms is provided in terms of sparsity level. We also note that this limitation will be relaxed for the stochastic block models with different group sizes. This is because the symmetric stochastic block model, in which the group size is uninformative to the inference, is a relatively difficult problem.

When our objective is to extract meaningful structures from real-world networks, we should carefully investigate the behaviors of the considered algorithm. For example, while modularity maximization entails the risk of underfitting [36], it tends to overfit [37,38] in comparison with other model selection criteria for various real-world networks. However, without quantitative knowledge, one might falsely expect a greedy algorithm to work well in a certain case, although there is very little chance that it will work appropriately.

Note that the overfit and underfit concepts depend on the assumed graph ensemble, and many modern algorithms are formulated on the basis of the graph ensemble defined by the stochastic block model [8,39–43]. Therefore, the present result can be used as a practical reference to perform modularity maximization.

Acknowledgments. This study was funded by the New Energy and Industrial Technology Development Organization (NEDO) and JSPS KAKENHI [No. 18K18127 (T.K.) and No. 17H00764 (Y.K.)].

[1] M. E. J. Newman and M. Girvan, Phys. Rev. E **69**, 026113 (2004).
[2] S. Fortunato, Phys. Rep. **486**, 75 (2010).
[3] M. E. J. Newman, Phys. Rev. E **88**, 042822 (2013).

[4] M. E. J. Newman, Phys. Rev. E **94**, 052315 (2016).
[5] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, J. Stat. Mech.: Theory Exp. (2008) P10008.
[6] H. Nishimori, *Statistical Physics of Spin Glasses and Information Processing: An Introduction* (Oxford University Press, New York, 2001).
[7] Y. Iba, J. Phys. A **32**, 3875 (1999).
[8] P. Zhang and C. Moore, Proc. Natl. Acad. Sci. U.S.A. **111**, 18144 (2014).
[9] P. W. Holland, K. B. Laskey, and S. Leinhardt, Soc. Networks **5**, 109 (1983).
[10] Y. J. Wang and G. Y. Wong, J. Am. Stat. Assoc. **82**, 8 (1987).
[11] B. Karrer and M. E. J. Newman, Phys. Rev. E **83**, 016107 (2011).
[12] The detectability limit is also termed the *detectability threshold*.
[13] R. R. Nadakuditi and M. E. J. Newman, Phys. Rev. Lett. **108**, 188701 (2012).
[14] X. Zhang, R. R. Nadakuditi, and M. E. J. Newman, Phys. Rev. E **89**, 042816 (2014).
[15] F. Radicchi, Phys. Rev. E **88**, 010801 (2013).
[16] T. Kawamoto and Y. Kabashima, Phys. Rev. E **91**, 062803 (2015).
[17] T. Kawamoto and Y. Kabashima, Eur. Phys. Lett. **112**, 40007 (2015).
[18] T. Kawamoto, arXiv:1710.08816.
[19] T. Kawamoto, Phys. Rev. E **97**, 032301 (2018).
[20] A. Decelle, F. Krzakala, C. Moore, and L. Zdeborová, Phys. Rev. Lett. **107**, 065701 (2011).
[21] A. Decelle, F. Krzakala, C. Moore, and L. Zdeborová, Phys. Rev. E **84**, 066106 (2011).
[22] E. Mossel, J. Neeman, and A. Sly, Probab. Theory Relat. Fields **162**, 431 (2015).
[23] L. Massoulié, in *Proceedings of the 46th Annual ACM Symposium on Theory of Computing* (ACM, New York, 2014), STOC '14, pp. 694–703.
[24] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevE.99.010301 for a detailed derivation of the main result, miscellaneous numerical experiments, and minor comments on the problem setting.
[25] A. Lancichinetti and S. Fortunato, Phys. Rev. E **80**, 056117 (2009).
[26] P. Ronhovde, D. Hu, and Z. Nussinov, Eur. Phys. Lett. **99**, 38006 (2012).
[27] D. Hu, P. Ronhovde, and Z. Nussinov, Philos. Mag. **92**, 406 (2012).
[28] R. K. Darst, D. R. Reichman, P. Ronhovde, and Z. Nussinov, J. Complex Netw. **3**, 333 (2015).
[29] C. Moore, arXiv:1702.00467.
[30] E. Abbe, J. Mach. Learn. Res. **18**, 1 (2018).
[31] http://tuvalu.santafe.edu/∼aaronc/modularity/.
[32] B. H. Good, Y.-A. de Montjoye, and A. Clauset, Phys. Rev. E **81**, 046106 (2010).
[33] F. Tanaka and S. F. Edwards, J. Phys. F: Metall. Phys. **10**, 2769 (1980).
[34] K. Fujii, J. Mod. Phys. **8**, 2042 (2017).
[35] https://github.com/vtraag/louvain-igraph.
[36] S. Fortunato and M. Barthélemy, Proc. Natl. Acad. Sci. U.S.A. **104**, 36 (2007).
[37] T. Kawamoto and Y. Kabashima, Phys. Rev. E **97**, 022315 (2018).
[38] A. Ghasemian, H. Hosseinmardi, and A. Clauset, arXiv:1802.10582.
[39] P. Latouche, E. Birmelé, and C. Ambroise, Stat. Modell. **12**, 93 (2012).
[40] T. P. Peixoto, Phys. Rev. X **4**, 011047 (2014).
[41] M. E. Newman and A. Clauset, Nat. Commun. **7**, 11863 (2016).
[42] M. E. J. Newman and G. Reinert, Phys. Rev. Lett. **117**, 078301 (2016).
[43] T. P. Peixoto, arXiv:1705.10225.