Finite-size analysis of the detectability limit of the stochastic block model

Jean-Gabriel Young, Patrick Desrosiers, Laurent Hébert-Dufresne, Edward Laurence, and Louis J. Dubé

Phys. Rev. E **95**, 062304 — Published 19 June 2017

DOI: 10.1103/PhysRevE.95.062304

Finite size analysis of the detectability limit of the stochastic block model

Jean-Gabriel Young,[1, *] Patrick Desrosiers,[1, 2] Laurent Hébert-Dufresne,[3] Edward Laurence,[1] and Louis J. Dubé[1, †]

[1] *Département de Physique, de Génie Physique, et d'Optique,*
Université Laval, Québec (Québec), Canada G1V 0A6
[2] *Centre de recherche de l'Institut universitaire en santé mentale de Québec, Québec (Québec), Canada G1J 2G3*
[3] *Santa Fe Institute, Santa Fe, New Mexico, USA, 87501*

(Dated: May 1, 2017)

It has been shown in recent years that the stochastic block model is sometimes undetectable in the sparse limit, i.e., that no algorithm can identify a partition correlated with the partition used to generate an instance, if the instance is sparse enough and infinitely large. In this contribution, we treat the finite case explicitly, using arguments drawn from information–theory and statistics. We give a necessary condition for finite size detectability in the general SBM. We then distinguish the concept of average detectability from the concept of instance-by-instance detectability, and give explicit formulas for both definitions. Using these formulas, we prove that there exist large equivalence classes of parameters, where widely different network ensembles are equally detectable with respect to our definitions of detectability. In an extensive case study, we investigate the finite size detectability of a simplified variant of the SBM, which encompasses a number of important models as special cases. These models include the symmetric SBM, the planted coloring model, and more exotic SBMs not previously studied. We conclude with three Appendices, where we study the interplay of noise and detectability, establish a connection between our information-theoretic approach and Random Matrix Theory, and provide proofs of some of the more technical results.

PACS numbers: 64.60.aq, 64.60.an, 89.70.-a

I. INTRODUCTION

Mesoscopic analysis methods [1] are among the most valuable tools available to applied network scientists and theorists alike. Their aim is to identify regularities in the structure of complex networks, thereby allowing for a better understanding of their function [1–3], their structure [4, 5], their evolution [6, 7], and of the dynamics they support [8–10]. Community detection is perhaps the best-known method of all [1, 2], but it is certainly not the only one of its kind [3]. It has been shown, for example, that the separation of nodes in a core and a periphery occurs in many empirical networks [11], and that this separation gives rise to more exotic mesoscopic patterns such as overlapping communities [12]. This is but an example—there exist multitudes of decompositions in structures other than communities that explain the shape of networks both clearly and succinctly [13].

The stochastic block model (SBM) has proven to be versatile and principled in uncovering these patterns [14–16]. According to this simple generative model, the nodes of a network are partitioned in blocks (the *planted partition*), and an edge connects two nodes with a probability that depends on the partition. The SBM can be used in any of two directions: Either to generate random networks with a planted mesoscopic structure [8, 10], or to infer the hidden mesoscopic organization of real complex networks, by fitting the model to network datasets [13, 14, 17]—perhaps its most useful application.

Stochastic block models offer a number of advantages over other mesoscopic pattern detection methods [3]. One, there is no requirement that nodes in a block be densely connected, meaning that blocks are much more general objects than communities. Two, the sound statistical principles underlying the SBM naturally solve many hard problems that arise in network mesoscopic analysis; this includes the notoriously challenging problem of determining the optimal number of communities in a network [18–20], or of selecting among the many possible descriptions of a network [1, 20, 21].

Another advantage of the statistical formulation of the SBM is that one can rigorously investigate its limitations. It is now known, for example, that the SBM admits a *resolution limit* [18] akin to the limit that arises in modularity–based detection method [22]. The limitations that have attracted the most attention, however, are the *detectability limit* and the closely related concept of *consistency limit* [23]. The SBM is said to be detectable for some parameters if an algorithm can construct a partition correlated with the planted partition [24], using no information other than the structure of a single—infinitely large—instance of the model. It is said to be consistent if one can *exactly* recover the planted partition. Therefore, consistency begets detectability, but not the other way around. Understanding when and why consistency (or detectability) can be expected is important, since one cannot trust the partitions extracted by SBM if it operates in a regime where it is not consistent (or detectable) [23].

Due to rapid developments over the past few years, the locations of the boundaries between the different levels of detectability are now known for multiple variants of the

[*] jean-gabriel.young.1@ulaval.ca
[†] ljd@phy.ulaval.ca

SBM, in the limit of infinite network sizes. If the average degree scales at least logarithmically with the number of nodes, then the SBM is consistent [25, 26], unless the constant multiplicative factor is too small, in which cas the SBM is then detectable, but not consistent. If the average degree scales slower than logarithmically, then the SBM is at risk of entering an *undetectable* phase where no information on the planted partition can be recovered from the network structure [27, 28]. This happens if the average degree is a sufficiently small constant independent of the number of nodes.

These asymptotic results are, without a doubt, extremely useful. Many efficient algorithms have been developed to extract information out of hardly consistent infinite instances [28–31]. Striking connections between the SBM and other stochastic processes have been established in the quest to bound the undetectable regime from below [23, 26, 32, 33]. But real networks are not infinite objects. Thus, even though it has been observed that there is a good agreement between calculations carried out in the infinite size limit and empirical results obtained on small networks [31], it is not immediately clear that the phenomenology of the infinite case carries over, unscathed, to the finite case.

In this contribution, we explicitly investigate detectability in *finite* networks generated by the SBM. We understand detectability in the information-theoretic sense [33]; our analysis is therefore algorithm–independent, and yields the boundaries of the region of the parameter space where the planted partition is undetectable, even for an optimal algorithm (with possibly exponential running time).

The combination of this information–theoretic point of view with our finite size analysis leads to new insights and results, which we organize as follows. We begin by formally introducing the SBM and the necessary background in Sec. II. We use this section to briefly review important notions, including inference (Sec. II B), as well as the consistency and detectability of the infinite SBM (Sec. II C). In Sec. III, we present a necessary condition for detectability, and show that it is always met, on average, by finite instances of the SBM. We then establish the existence of a large equivalence class with respect to this notion of average detectability. In Sec. V, we introduce the related concept of η–detectability and investigate the complete detectability distribution, beyond its average. In Sec. VI, we apply the perfectly general framework of Secs. III–V to a constrained variant of the SBM: The general modular graph model of Ref. [34]. The results of this section hold for a broad range of models, since the general modular graphs encompass the symmetric SBM, the planted coloring model and many other models as special cases. We gather concluding remarks and open problems in Sec. VII. Three Appendices follow. In the first, we investigate the interplay between noise and our notion of average detectability (Appendix A); in the second, we establish a connection between our framework and Random Matrix Theory (Appendix B); in the third, we give the details of two technical proofs encountered in the main text (Appendix C).

II. STOCHASTIC BLOCK MODEL

A. Definition of the model

The stochastic block model is formally defined as follows: Begin by partitioning a set of n nodes in q blocks of fixed sizes $\boldsymbol{n} = (n_1, ..., n_q)$, with $n = \sum_{r=1}^q n_r$. Denote this partition by $\mathcal{B} = \{B_1, ..., B_q\}$, where B_r is the set of nodes in the r^{th} block. Then, connect the nodes in block B_r to the nodes in block B_s with probability p_{rs}. In other words, for each pair of nodes (v_i, v_j), set the element a_{ij} of the adjacency matrix $\boldsymbol{A}$ to 1 with probability $p_{\sigma(v_i)\sigma(v_j)}$ and to 0 otherwise, where $\sigma(v_i)$ is the block of v_i. Note that for the sake of clarity, we will obtain all of our results for simple graphs, where edges are undirected and self-loops (edges connecting a node to itself) are forbidden [35]. This implies that $p_{rs} = p_{sr}$ and that $a_{ii} = 0$.

We will think of this process as determining the outcome of a random variable, whose support is the set of all networks of n nodes. Due to the independence of edges, the probability (likelihood) of generating a particular network G is simply given by the product of $\binom{n}{2}$ Bernoulli random variables, i.e.,

$$\mathbb{P}(G|\mathcal{B}, \boldsymbol{P}) = \prod_{i<j} [1 - p_{\sigma(v_i)\sigma(v_j)}]^{1-a_{ij}} [p_{\sigma(v_i)\sigma(v_j)}]^{a_{ij}} \ , \quad (1)$$

where $\boldsymbol{P}$ is the $q \times q$ matrix of connection probabilities of element p_{rs} (sometimes called the affinity or density matrix), and $i < j$ is a shorthand for "$i, j : \ 1 \leq i < j \leq n$". It is easy to check that the probability (1) is properly normalized over the set of all networks of n distinguishable nodes.

A useful alternative to Eq. (1) expresses the likelihood in terms of the number of edges between each pair of blocks (B_r, B_s) rather than as a function of the adjacency matrix [17]. Notice how the number of edges m_{rs} appearing between the sets of nodes B_r and B_s is at most equal to

$$m_{rs}^{\max} = \begin{cases} \binom{n_r}{2} & \text{if } r = s, \\ n_r n_s & \text{otherwise.} \end{cases} \quad (2)$$

Each of these $m_{rs}^{\max}$ edges exists with probability p_{rs}. This implies that m_{rs} is determined by the sum of $m_{rs}^{\max}$ Bernoulli trials of probability p_{rs}, i.e., that m_{rs} is a binomial variable of parameter p_{rs} and maximum $m_{rs}^{\max}$. The probability of generating a particular instance G can therefore be written equivalently as

$$\mathbb{P}(G|\mathcal{B}, \boldsymbol{P}) = \prod_{r \leq s} (1 - p_{rs})^{m_{rs}^{\max} - m_{rs}} (p_{rs})^{m_{rs}} \ . \quad (3)$$

where $\{m_{rs}\}$ and $\{m_{rs}^{\max}\}$ are jointly determined by the partition $\mathcal{B}$ and the structure of G, and $r \leq s$ denotes "$r, s : \ 1 \leq r \leq s \leq q$".

Having a distribution over all networks of n nodes, one can then compute average values over the ensemble. For example, the average degree of node v_i is given by

$$\langle k_i\rangle = \sum_r p_{\sigma(v_i)r}(n_r - \delta_{\sigma(v_i)r}) \ , \tag{4}$$

where δ_{ij} is the Kronecker Delta. The expression correctly depends on the block $B_\sigma(v_i)$ of v_i; nodes in different blocks will in general have different average degree. Averaging over all nodes, one finds the average degree of the network

$$\langle k\rangle = \frac{2}{n}\sum_{r\leq s} m_{rs}^{\max} p_{rs} \ . \tag{5}$$

This global quantity determines the density of the SBM when $n\to\infty$. The SBM is said to be dense if $\langle k\rangle = \mathcal{O}(n)$, i.e., if p_{rs} is a constant independent of n. It is said to be sparse if $\langle k\rangle = \mathcal{O}(1)$, i.e., if $p_{rs} = c_{rs}/n$ goes to zero as n^{-1}. In the latter case, a node has a constant number of connections even in an infinitely large network—a feature found in most large scale real networks [36].

For finite instances, it will often be more useful to consider the average density directly. It is defined as the number of edges in G, normalized by the number of possible edges, i.e.,

$$\rho = \frac{\langle k\rangle}{n-1} = \sum_{r\leq s}(m_{rs}^{\max}/m^{\max})p_{rs} \equiv \sum_{r\leq s}\alpha_{rs}p_{rs} \ , \tag{6}$$

where $m^{\max} = \sum_{r\leq s} m_{rs}^{\max}$, and

$$\alpha_{rs} := m_{rs}^{\max}/m^{\max} \ . \tag{7}$$

The dense versus sparse terminology is then clearer: The density of sparse networks goes to zero as $\mathcal{O}(n^{-1})$, while dense networks have a nonvanishing density $\rho = \mathcal{O}(1)$.

B. Inference

Depending on the elements of $\boldsymbol{P}$, the SBM can generate instances reminiscent of real networks with, e.g., a community structure [3] ($p_{rr} > p_{rs}$) or a core-periphery organization [11] ($p_{11} > p_{12} > p_{22}$ and $p_{22} \sim 0$). However, the SBM really shines when it is used to infer the organization in blocks of the nodes of real complex networks—this was, after all, its original purpose [14].

To have inferred the mesoscopic structure of a network (with the SBM) essentially means that one has found the partition $\mathcal{B}^*$ and density matrix $\boldsymbol{P}^*$ that best describes it. In principle, it is a straightforward task, since one merely needs to (a) assign a likelihood $\mathbb{P}(\mathcal{B},\boldsymbol{P}|G)$ to each pair of partition and parameters [see Eqs. (1)–(3)], then (b) search for the most likely pair ($\mathcal{B}^*$, $\boldsymbol{P}^*$). Since there are exponentially many possible partitions, this sort of enumerative approach is of little practical use. Fortunately, multiple approximate and efficient inference tools have been proposed to circumvent this fundamental problem. They draw on ideas from various fields such as statistical physics [13, 28, 31], Bayesian statistics [17, 37], spectral theory [29, 30, 38, 39] and graph theory [40], to name a few, and they all produce accurate results in general.

C. Detectability and consistency

One could expect perfect recovery of the parameters and partition from most of these sophisticated algorithms. This is called the consistency property. It turns out, however, that all known inference algorithms for the SBM, as diverse as they might be, fail on this account. And their designs are not at fault, for there exists an explanation of this generalized failure.

Consider the density matrix of elements $p_{rs} = \rho\,\forall(r,s)$, It is clear that the block partition is irrelevant—the generated network cannot and will not encode the planted partition. Thus, no algorithm will be abe to differentiate the planted partition from other partitions. It is then natural to assume that inference will be hard or impossible if $p_{rs} = \rho + \epsilon_{rs}(n)$, where $\epsilon_{rs}(n)$ is a very small perturbation for networks of n nodes; there is little difference between the uniform case and this perturbed case. In contrast, if the elements of $\boldsymbol{P}$ are widely different from one another, e.g., if $p_{rr} = 1$ and $p_{rs} = 0$ for $r \neq s$, then easy recovery should be expected.

Understanding where lies the transition between these qualitatively different regimes has been the subject of much recent research (see Ref. [23] for a recent survey). As a result, the regimes have been clearly separated as follows: (i) the undetectable regime, (ii) the detectable (but not consistent) regime and (iii) the consistent regime (and detectable). It has further been established that the scaling of ρ with respect to n determines which regime is reached, in the limit $n\to\infty$.

The SBM is said to be *strongly consistent* if its planted partition can be inferred perfectly, with a probability that goes to 1 as $n\to\infty$ (it is also said to be in the *exact recovery* phase). Another close but weaker definition of consistency asks that the probability of misclassifying a node goes to zero with $n\to\infty$ (the *weakly consistent* or *almost exact recovery* phase). These regimes prevail when $\boldsymbol{P}$ scales at least as fast as $\boldsymbol{P} = \log(n)\boldsymbol{C}/n$, where $\boldsymbol{C}$ is a $q\times q$ matrix of constants [25, 26, 41]. Predictably, most algorithms (e.g., those of Refs. [17, 40, 41]) work well in the exact recovery phase regime, since it is the easiest of all .

In the *detectable* (but not consistent) regime, exact recovery is no longer possible (the *partial recovery* phase). The reason is simple: Through random fluctuations, some nodes that belong to, say, block B_1, end up connecting to other nodes as if they belonged to block B_2. They are thus systematically misclassified, no matter the choice of algorithms. This occurs whenever $\boldsymbol{P} = \boldsymbol{C}/n$, or $\boldsymbol{P} = f(n)\boldsymbol{C}/n$, with $f(n)$ a function of n that scales slower than $\log(n)$.

The discovery of the third regime—the *undetectable regime*—arguably rekindled the study of the fundamental limits of the SBM [27, 28]. In this regime, which occurs when $\boldsymbol{P} = \boldsymbol{C}/n$ and $\boldsymbol{C}$ is more or less uniform, it is impossible to detect a partition that is even correlated with the planted one. That is, one cannot classify nodes better than at random, and no information on the planted partition can be extracted. Thus, some parametrizations of the SBM are said to lie below the *detectability limit.* This limit was first investigated with informal arguments from statistical physics [27, 28, 31, 34, 42], and has since been rigorously formalized in Refs. [33, 45], among others.

There exist many efficient algorithms that are reliable close to the detectability limit; noteworthy examples include Belief-Propagation [28, 31, 46], and spectral algorithms based on the ordinary [29] and weighted [32] non-backtracking matrix, as well as matrices of self-avoiding walks [30]. But when the number of blocks is too large, most of these algorithms are known to fail well above the information–theoretic threshold, i.e., the point where it can be proven that the partition is detectable given arbitrary computational power. It has been therefore conjectured in Ref. [31], that there exists multiple troublesome phases for inference: A truly undetectable regime, and a regime where detection is not achievable *efficiently.* In the latter, it is thought that one *can* find a good partition, but only by enumerating all partitions—a task of exponential complexity.

In this contribution, however, we will not focus on this so-called hard regime. As far as we are concerned, detectability will be understood in terms of information, i.e., we will delimit the boundaries of the information-theoretically undetectable regime.

III. DETECTABILITY OF FINITE NETWORKS

Detectability and consistency are well separated phases of the infinite stochastic block model. A minute perturbation to the parameters may potentially translate into widely different qualitative behaviors. The picture changes completely when one turns to finite instances of the model. Random fluctuations are not smoothed out by limits, and transitions are much less abrupt. We argue that, as a result, one has to account for the complete distribution of networks to properly quantify detectability, i.e., define detectability for *network instances* rather than parameters. This, in turn, commands a different approach that we now introduce.

A. Hypothesis test and the detectability limit

Consider a single network G, generated by the SBM with some planted partition $\mathcal{B}$ and matrix $\boldsymbol{P} = r\mathbf{1}\mathbf{1}^{\intercal} + \boldsymbol{\epsilon}$, where $\mathbf{1}\mathbf{1}^{\intercal}$ is a matrix of ones, r a constant, and $\boldsymbol{\epsilon}$ a matrix of (small) fluctuations. Suppose that the average density equals ρ, and consider a second density matrix $\rho\mathbf{1}\mathbf{1}^{\intercal}$ for which the block structure has no effect on the generative process. If an observer with *complete knowledge* of the generative process and its parameters cannot tell which density matrix, $\boldsymbol{P}$ or $\rho\mathbf{1}\mathbf{1}^{\intercal}$, is the most likely to have generated G, then it is clear that *this particular instance* does not encode the planted partition. As a result, it will be impossible to detect a partition correlated with the planted partition.

This idea can be translated into a mathematical statement by way of a likelihood test. For a SBM of average density ρ, call the ensemble of Erdős-Rényi graphs of density ρ the ensemble of *equivalent random networks.* Much like the SBM (see Sec. II), its likelihood $Q(G|\rho)$ is given by the product of the density of $\binom{n}{2}$ independent and identically distributed Bernoulli variables, i.e.,

$$\mathbb{Q}(G|\rho) = \prod_{i<j} \rho^{a_{ij}}(1-\rho)^{a_{ij}} = \rho^{m}(1-\rho)^{m^{\max}-m} \ , \qquad (8)$$

where $m := \sum_{r\leq s} m_{rs}$ is the total number of edges in G.

The condition is then the following: Given a network G generated by the SBM of average density ρ and density matrix $\boldsymbol{P}$, one can detect the planted partition $\mathcal{B}$ if the SBM is more likely than its equivalent random ensemble of density ρ, i.e.,

$$\Lambda = \frac{\mathbb{P}(G|\mathcal{B},\boldsymbol{P})}{\mathbb{Q}(G|\rho)} > 1 \ . \qquad (9)$$

A similar condition has been used in Ref. [45] and [33] to pinpoint the location of the detectability limit in infinite and sparse instances of the SBM. Nothing forbids its application to the finite size problem; we will see shortly that it serves us well in the context of finite size detectability.

B. Normalized log-likelihood ratio

The (equivalent) normalized log-likelihood ratio

$$\mathcal{L} := \frac{\log\Lambda}{m^{\max}} \qquad (10)$$

will be more practical for our purpose. This simple transformation brings the line of separation between models from $\Lambda = 1$ to $\mathcal{L} = 0$, and prevents the resulting quantity from becoming too large. More importantly, it changes products into sums, and allows for a simpler expression

$$\mathcal{L}=\sum_{r\leq s}\left\{\frac{m_{rs}}{m^{\max}}\log\left[\frac{p_{rs}(1-\rho)}{\rho(1-p_{rs})}\right]+\alpha_{rs}\log\left[\frac{1-p_{rs}}{1-\rho}\right]\right\}. \qquad (11)$$

We will focus, for the remainder of this contribution, on the case where network instances G of n nodes are drawn from the SBM of parameters $(\mathcal{B},\boldsymbol{P})$. In this context, $\mathcal{L}$ can is a random variable whose support is the networks of n nodes with labeled nodes (see Fig. 1). Since $\boldsymbol{P},\rho,\alpha$ and $m^{\max}$ are all parameters, $\mathcal{L}$ can also be seen as a

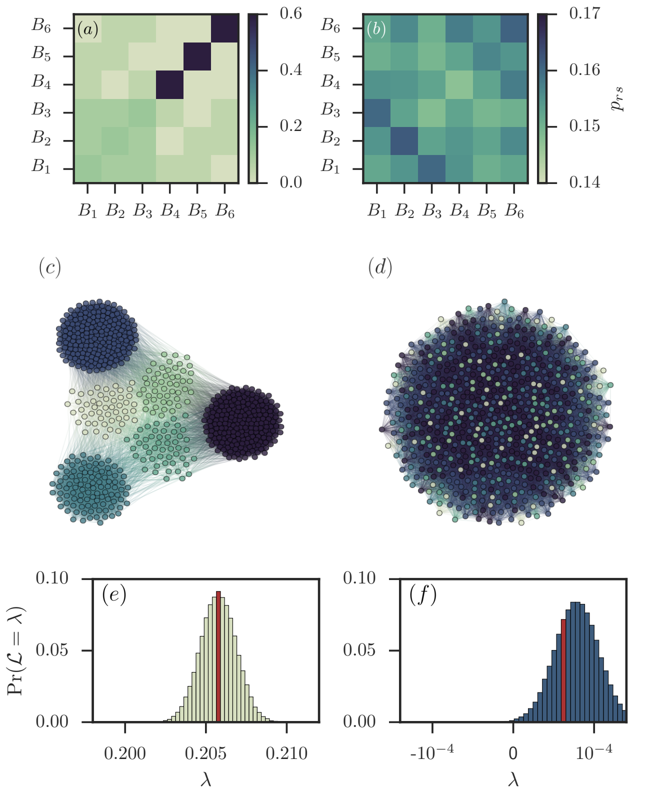

FIG. 1. (color online) Stochastic block model with (a,c,e) non-uniform density matrix and (b,d,f) nearly uniform density matrix. (a-b) Density matrix of the two ensembles. Notice the difference in scale. (c-d) One instance of each ensemble, with $\boldsymbol{n} = [50, 50, 50, 100, 200, 200]$. Each color denotes a block [47]. (e-f) Empirical distribution of the normalized log-likelihood obtained from 100 000 samples of $\mathcal{L}$. The bins in which the instances (c-d) fall are colored in red. Notice that a negative log-likelihood ratio is associated with some instances in (f).

weighted sum of binomial distributed random variables $m_{rs} \sim \mathrm{Bin}(m_{rs}^{\max}, p_{rs})$, with a constant offset. Its average will be a prediction of the detectability for the ensemble (Sec. IV), and the probability $\Pr(\mathcal{L} < 0; \boldsymbol{P}, \boldsymbol{\alpha}, m^{\max})$ will give the fraction of instances that are undetectable for the selected parameters (Sec. V).

C. Interpretation of $\mathcal{L}$: Information–theoretic bound and inference difficulty

Because likelihood ratio tests can be understood as quantifying the amount of evidence for a hypothesis (compared to a null hypothesis), there will be two interpretations of $\mathcal{L}$.

On the one hand, the condition $\mathcal{L} > 0$ will provide a lower bound on detectability; if $\mathcal{L}(G, \mathcal{B}, \boldsymbol{P}) < 0$, then we can safely say that the instance G is information–theoretically undetectable. However, $\mathcal{L}(G, \mathcal{B}, \boldsymbol{P}) > 0$ does not necessarily mean that the instance is information–theoretically detectable. This is due to the fact that the condition $\mathcal{L} > 0$ is necessary but not sufficient, since we assume a complete knowledge of the generative process in calculating $\mathcal{L}$.

On the other hand, we will interpret $\mathcal{L}$ operationally as a measure of the difficulty of the inference problem (not in the computational sense). A large ratio of a hypothesis $\mathbb{H}$ to its null model $\mathbb{H}_0$ implies that the hypothesis is a much better explanation of the data than $\mathbb{H}_0$; therefore $\mathcal{L}$ measures how easy it is to select between $\mathbb{P}$ and $\mathbb{Q}$, given full knowledge of the generative process, and inference algorithms will perform better when the ratio is larger. Many empirical results will validate this interpretation (see Sec. VI).

IV. AVERAGE DETECTABILITY

A. Average normalized log-likelihood

The average of a log-likelihood ratio is also known as the Kullback-Leibler (KL) divergence $D(\cdot||\cdot)$ of two hypotheses [48], i.e.

$$\begin{aligned}\langle \mathcal{L}(\boldsymbol{\alpha}, \boldsymbol{P}) \rangle &= \sum_{\{G\}} \frac{\mathbb{P}(G|\mathcal{B}, \boldsymbol{P})}{m^{\max}} \log \frac{\mathbb{P}(G|\mathcal{B}, \boldsymbol{P})}{\mathbb{Q}(G|\rho)} \\ &= \frac{D(\mathbb{P}||\mathbb{Q})}{m^{\max}} \;, \end{aligned} \tag{12}$$

where the sum runs over all n nodes networks. Since the KL divergence is always greater or equal to zero, with equality if and only if $\mathbb{P} = \mathbb{Q}$, and since $\mathcal{L} > 0$ is only a necessary condition for detectability, the average $\langle \mathcal{L} \rangle$ will not be enough to conclude on detectability of the SBM, except for the case $\mathbb{P} = \mathbb{Q}$ [49]. Results pertaining to $\langle \mathcal{L} \rangle$ will therefore be best interpreted in terms of inference difficulty.

However, even if the average log-likelihood ratio is always positive (assuming $\mathbb{P} \neq \mathbb{Q}$), it can be extremely close to zero for density matrix $\boldsymbol{P}$ "close" to $\rho \mathbf{1}\mathbf{1}^\intercal$ [Fig. 1 (f)]. In fact, as we will see in Sec. V, $\langle \mathcal{L}(\boldsymbol{\alpha}, \boldsymbol{P}) \rangle \approx 0$ implies that there are instances for which $\mathcal{L} < 0$. Therefore, whenever the average is small, we may also take it as a sign that a significant that the planted partition of some instances are truly undetectable.

B. Compact form

While Eq. (12) has a precise information–theoretic interpretation, there exists an equivalent form, both more compact and easier to handle analytically. It is given by

$$\langle \mathcal{L}(\boldsymbol{\alpha}, \boldsymbol{P}) \rangle = h(\rho) - \sum_{r \leq s} \alpha_{rs} h(p_{rs}) \;, \tag{13}$$

where

$$h(p) = -(1-p)\log(1-p) - p\log(p) \tag{14}$$

is the binary entropy of $p \in [0,1]$. This expression can be obtained in a number of ways, the most direct of which is to take the average of Eq. (11) over all symmetric matrices $\boldsymbol{m} = (m_{11}, m_{12}, \dots, m_{qq})$ with entries in $\mathbb{N}$ and upper bounds given by $\boldsymbol{m}^{\max} = (m_{11}^{\max}, m_{12}^{\max}, \dots, m_{qq}^{\max})$. That is to say, we use the interpretation where $\mathcal{L}$ is a weighted sum of binomial distributed random variable, instead of the interpretation where it is a random variable over the networks of n nodes (see Sec. III B). The probability mass function associated to $\boldsymbol{m}$ is then $\Pr[\boldsymbol{m}] = \prod_{r\leq s} \Pr[m_{rs}]$, where $\Pr[m_{rs}]$ is the binomial distribution of parameter p_{rs} and upper bound $m_{rs}^{\max}$. Due to the linearity of expectations, it is straightforward to check that the average of the first sum of Eq. (11) equals

$$\sum_{\boldsymbol{m}} \Pr \boldsymbol{m} \sum_{r\leq s} \frac{m_{rs}}{m^{\max}} \log\left[\frac{p_{rs}}{\rho}\frac{1-\rho}{1-p_{rs}}\right] = \sum_{r\leq s} \log\left[\frac{p_{rs}}{\rho}\frac{1-\rho}{1-p_{rs}}\right]\frac{m_{rs}^{\max} p_{rs}}{m^{\max}} \ .$$

Recalling expression (6), one then finds

$$\begin{aligned}\langle \mathcal{L}(\boldsymbol{\alpha}, \boldsymbol{P})\rangle &= -\sum_{r\leq s} \alpha_{rs}\big[(1-p_{rs})\log(1-\rho) + p_{rs}\log\rho\big] \\ &\quad + \sum_{r\leq s} \alpha_{rs}[(1-p_{rs})\log(1-p_{rs}) + p_{rs}\log p_{rs}] \\ &= h(\rho) - \sum_{r\leq s} \alpha_{rs}\, h(p_{rs}) \ .\end{aligned}$$

where α_{rs} is defined in Eq. (7) with the normalization $\sum_{r\leq s} \alpha_{rs} = 1$. Notice how this expression does not depend on $\mathcal{B}$ anymore. In this context, the only role of the planted partition is to fix the relative block sizes $\boldsymbol{\alpha}$. Thus, the average log-likelihood $\langle\mathcal{L}\rangle$ of two models with different planted partitions but identical $\boldsymbol{\alpha}$ is the same (up to a size-dependent constant).

With these two expressions for $\langle\mathcal{L}\rangle$ in hand [Eqs. (12) and (13)], we can now build an intuition for what the easiest and most difficult detectability problems might look like. The KL divergence is never negative, and Eq. (13) shows that the maximum of $\langle\mathcal{L}\rangle$ is $h(1/2)$; the average of the normalized log-likelihood is thus confined to the interval

$$0 \leq \langle \mathcal{L}(\boldsymbol{\alpha}, \boldsymbol{P})\rangle \leq h(1/2) \ . \tag{15}$$

An example of parameters that achieves the upper bound would be the SBM of density matrix $p_{11} = p_{22} = 1$, $p_{12} = 0$, with $\boldsymbol{n} = [n/2, n/2]$, i.e., the "ensemble" of disconnected $n/2$–cliques (which contains a single instance). An example of parameters that achieves the lower bound would be $\mathbb{P} = \mathbb{Q}$, but also $\rho \to 0$ [see Eq. (13)].

C. Equivalent stochastic block models

We now use Eq. (13) to uncover hidden connections between different regimes of the SBM. Notice how this expression induces equivalence classes in the parameter space of the model, with respect to $\langle\mathcal{L}\rangle$, i.e., subsets of parameters which all satisfy

$$\lambda = \langle \mathcal{L}(\boldsymbol{P}, \boldsymbol{\alpha})\rangle \ , \tag{16}$$

where λ is a constant that defines the equivalence class.

In the next paragraphs, we will characterize these equivalence classes in two complementary ways. First, we will look for global transformations that preserve λ and map parameters $(\boldsymbol{\alpha}, \boldsymbol{P})$ to some other—not necessarily close—pair of parameters $(\boldsymbol{\alpha}', \boldsymbol{P}')$. Provided that they satisfy a number of standard constraints, these transformations will be shown to correspond to the symmetry group of the set of *hypersurfaces* $\langle\mathcal{L}(\boldsymbol{\alpha}, \boldsymbol{P})\rangle = \lambda$. Second, we will consider Eq. (16) explicitly and use it to obtain an approximate hypersurface equation. This equation will be used in later sections to determine the location of the hypersurfaces that separate the parameter space of the SBM in different detectability phases.

1. *Global transformations: the symmetry group of the SBM*

We first look for the set of λ–preserving global transformations, i.e., all transformations $T(f_1, f_2)$ of the form

$$\boldsymbol{\alpha}' = f_1(\boldsymbol{\alpha}), \quad \boldsymbol{P}' = f_2(\boldsymbol{P}) \tag{17}$$

valid at every point of the parameter space. This is a broad definition and it must be restricted if we are to get anything useful out of it. Intuitively, we do not want these transformations to change the space on which they operate, so it is natural to ask that they be space-preserving. Under the (reasonable) constraint that these transformations are invertible as well, we can limit our search for λ–preserving transformations to the symmetry group of the parameter space.

We will be able to harness known results of geometry and algebra once the parameter space of the SBM is properly defined. This space is in fact the Cartesian product of two parameter spaces: The parameter space of $\boldsymbol{\alpha}$ and that of $\boldsymbol{P}$. Since there is $q^* = q(q+1)/2$ free parameters in both $\boldsymbol{\alpha}$ and $\boldsymbol{P}$, the complete space is of dimension $2q^* - 1$. It is the product of the q^*–dimensional hypercube—in which every point corresponds to a choice of $\boldsymbol{P}$—, and the (q^*-1)–dimensional simplex—in which every point corresponds to a choice of $\boldsymbol{\alpha}$. The latter is a simplex due to the normalization $\sum_{r\leq s}\alpha_{rs} = (m^{\max})^{-1}\sum_{r\leq s} m_{rs}^{\max} = 1$.

Now, the symmetry group of the q^*–dimensional hypercube and that of the (q^*-1)–dimensional regular simplex are well-known [50]: They are respectively the hyperoctahedral group B_{q^*} and the symmetric group S_{q^*}. Their action on $\boldsymbol{\alpha}$ and $\boldsymbol{P}$ can be described as

$$\begin{aligned}\alpha_{rs} &\mapsto \alpha'_{rs} = \alpha_{\pi(r,s)} \ , \\ p_{rs} &\mapsto p'_{rs} = \gamma_{rs} + (1 - 2\gamma_{rs})p_{\omega(r,s)} \ ,\end{aligned}$$

where $\gamma_{rs} = \{0,1\}$, and where both $\pi(r,s)$ and $\omega(r,s)$ are permutations of the indexes (r,s). While the symmetries of $\mathcal{L}(\boldsymbol{\alpha},\boldsymbol{P})$ are automatically symmetries of the parameters, the converse is not true. We therefore look for transformations T that satisfy

$$\langle\mathcal{L}(\boldsymbol{\alpha},\boldsymbol{P})\rangle = \langle\mathcal{L}\big(f_1(\boldsymbol{\alpha}),f_2(\boldsymbol{P})\big)\rangle \ . \tag{18}$$

It turns out that this constraint is satisfied if and only if $\pi=\omega$ and $\gamma_{rs}=\gamma\ \forall(r,s)$, i.e., for transformations of the form

$$\alpha_{rs}\mapsto\alpha'_{rs} = \alpha_{\pi(r,s)} \ , \tag{19a}$$

$$p_{rs}\mapsto p'_{rs} = \gamma + (1-2\gamma)p_{\pi(r,s)} \ , \tag{19b}$$

with $\gamma=\{0,1\}$ (see Appendix C 1 for a detailed proof). The permutation component of the symmetry is not to be confused with the symmetries generated by relabeling blocks: The latter only leads to $q!$ different symmetries, whereas the former correctly generates $q^*!\gg q!$ symmetries, or a total of $2q^*!$ symmetries once they are compounded with $p_{rs}\mapsto 1-p_{rs}$. The symmetries come about because the ordering of summation of the terms $\alpha_{rs}h(p_{rs})$ in Eq. (13) does not matter, and both $h(\rho)$ and $h(p_{rs})$ are preserved when $p_{rs}\mapsto 1-p_{rs}$.

As an example of symmetry, let us focus on the special transformation $p_{rs}\mapsto 1-p_{rs}\ \forall(r,s)$ with $\pi(r,s)=(r,s)$, i.e., the only transformation that does not change the block structure of the model. Since networks generated by these parameters can be seen as complement of one another (i.e., an edge present in G is absent from G', and vice-versa), we may call this transformation the *graph complement* transformation. The fact that it preserves detectability can be understood on a more intuitive level with the following argument. Suppose that we are given an unlabeled network G generated by the SBM and that we are asked to confirm or infirm the hypothesis that it was, in fact, generated by the SBM. Even if nothing is known about the generative process, we can take the complement of the network—a trivial (and reversible) transformation. But this should not help our cause. After all, this transformation cannot enhance or worsen detectability since no information is added to or removed from G in the process. So we expect that λ be preserved, and it is. Because all other symmetries affect the block structure through a change of $\boldsymbol{\alpha}$, what the above result shows is that there is no other "information-preserving" transformation that can be applied to G without a prior knowledge of its planted partition.

2. *Hypersurfaces and detectability regions*

We now turn to the problem of finding compact and explicit formulas that describe the hypersurfaces of constant $\langle\mathcal{L}\rangle$ in the parameter space [see Eq. (16)]. In so doing we will have to be mindful of the fact that the scale $m^{\max}$ intervenes in the calculation, even though it is absent from our expression for $\langle\mathcal{L}\rangle$. This can be made explicit by rewriting Eq. (16) as $\langle\log\Lambda\rangle/m^{\max}=\widetilde{\lambda}$; it is easy to see that any given hypersurface will be comparatively closer to the region $\langle\mathcal{L}\rangle=0$ in larger networks. We focus on the universal behavior of the hypersurfaces and remove all references to the scale of the problem by defining $\lambda := m^{\max}\widetilde{\lambda}$—predictions for real systems can be recovered by reverting to the correct scale.

While Eq. (16) is easily stated, it is not easily solved for, say, $\{p_{rs}\}$. The average normalized log-likelihood ratio involves a sum of logarithmic terms; the hypersurface equation is thus transcendental. To further complicate matters, there are $2q^*-1=q(q-1)-1$ degrees of freedom and the number of free parameters grow quadratically with q. As a result, little can be said of truly general instances of the SBM—at least analytically. All is not hopeless, however, because there are approximation methods that work well when the number of free parameters is not too large. We sketch the idea here, and apply it to a simpler variant of the SBM in the case study of Sec. VI.

Expanding the binary entropy functions $h(p_{rs})$ around $p_{rs}=\rho\ \forall r\leq s$ drastically simplifies the hypersurface equation. Leaving the term $h(\rho)$ untouched, we find from Eq. (16)

$$\lambda = h(\rho) - \sum_{r\leq s}\alpha_{rs}\left[h(\rho)+\sum_{k=1}^{\infty}\frac{1}{k!}\frac{\partial^k h(x)}{\partial x^k}\bigg|_{x=\rho}(p_{rs}-\rho)^k\right] \ .$$

Due to the normalization of $\{\alpha_{rs}\}_{r\leq s}$, all terms in $h(\rho)$ cancel out, and the definition $\sum_{r\leq s}\alpha_{rs}p_{rs}=\rho$ allows us to eliminate the first order terms as well. We are therefore left with

$$2\lambda\rho(1-\rho) = \sum_{r\leq s}\alpha_{rs}(p_{rs}-\rho)^2 + \mathcal{O}[(p_{rs}-\rho)^3] \ , \tag{20}$$

where ρ is fixed and $(\boldsymbol{\alpha},\boldsymbol{P})$ take on values constrained by both Eqs. (6) and (20). We then resort to a change of parameters and choose $\rho(\boldsymbol{P},\boldsymbol{\alpha})$ as one of the parameters. Selecting the q^*-1 other parameters Δ_{rs} such that $p_{rs}=\rho(\boldsymbol{P},\boldsymbol{\alpha})+\Delta_{rs}(\boldsymbol{P},\boldsymbol{\alpha})$, we obtain the form

$$2\lambda\rho(1-\rho) = \sum_{r\leq s}\alpha_{rs}(\Delta_{rs})^2 \ . \tag{21}$$

Hypersurfaces are therefore ellipsoids when $p_{rs}\approx\rho\ \forall(r,s)$.

Besides the simplicity of Eq. (21), there are two additional arguments for dropping the higher order terms in Eq. (20). One, the series is invariant under the symmetry $p_{rs}\mapsto 1-p_{rs}\ \forall(r,s)$ only if we limit ourselves to the second order expression: It easily verified that

$$\frac{\partial^k h(x)}{\partial x^k}\bigg|_{x=\rho}(p_{rs}-\rho)^k =$$

$$(-1)^k(k-2)!\left[\frac{1}{(\rho-1)^{k-1}}-\frac{1}{(\rho)^{k-1}}\right](p_{rs}-\rho)^k$$

is off by a sign for odd powers of k under the mapping $p_{rs} \mapsto 1 - p_{rs}$, which also implies $\rho \mapsto 1 - \rho$. Two, the true hypersurfaces enclose sets of parameters which are convex with respect to $\boldsymbol{P}$, and so does the hypersurface implicitly defined in Eq. (20). The convexity of the hypersurface follows from the fact that the sublevel set of a convex function encloses a convex set [51], and from the observation that $\langle\mathcal{L}\rangle$ is convex with respect to $\boldsymbol{P}$ [this is easy to show with Eq. (13) and the log-sum inequality, see Appendix C 2]. The convexity of the approximate level set is trivial to the second order, since it is an ellipsoid [Eq. (21)]. However, the approximate level set need not be convex when higher order terms are included. Together, these two observations tell us that while not exact, Eq. (20) captures the important *qualitative* features of the problem, and that it is not necessarily true of approximate solutions with only a few extra terms.

V. DETECTABILITY DISTRIBUTION

In the previous section, we have computed the average $\langle\mathcal{L}\rangle$ and used it to obtain equivalence among the parameters, with respect to detectability. We have also shown that $\langle\mathcal{L}\rangle > 0$ for most parameters, i.e., that we could not use the necessary condition $\mathcal{L} > 0$ to conclude on the *undetectability* of the finite SBM, on average. As we will now argue, this conclusion must be further refined; the full distribution of $\mathcal{L}$ leads to a more accurate picture of detectability.

A. The whole picture: η–detectability

Consider a parametrization $(\mathcal{B}, \rho\mathbf{1}\mathbf{1}^\intercal + \boldsymbol{\epsilon})$ of the SBM which yields $\langle\mathcal{L}\rangle \approx 0$. Turning to the distribution of $\mathcal{L}$ for this parametrization, one expect to find $\mathcal{L} < 0$ with non-zero probability (unless the distribution of $\mathcal{L}$ concentrates on $\mathcal{L} = 0$). Therefore, $\langle\mathcal{L}\rangle$ could be indicative of detectability for some *fraction* of the networks generated by the SBM.

Let us formalize this notion and introduce the concept of η–detectability. We will say that the ensemble of networks generated with the SBM of parameters $(\mathcal{B}, \boldsymbol{P})$ is η–detectable if

$$\Pr(\mathcal{L} < 0; \mathcal{B}, \boldsymbol{P}) = 1 - \eta\ . \tag{22}$$

That is, η gives the fraction of networks in the ensemble which evades the necessary condition for undetectability. If $\eta \to 0$, then detection is impossible, in the sense that most instances are best described by the null hypothesis $\mathbb{Q}$. If $\eta \to 1$, then most instances contain statistical evidence for $\mathcal{B}$; detection cannot be ruled out on the basis of the log-likelihood test.

We must compute the complete distribution or the cumulative distribution function of $\mathcal{L}$ to determine η. An exact result is out of reach since the outcome of $\mathcal{L}$ is determined by a weighted sum of independent binomial variables with non-identical distributions. In the following paragraphs, we give an approximate derivation based on the central limit theorem—it agrees extremely well with empirical results for all but the smallest networks.

B. Approximate equation for η

Equation (11) gives the normalized log-likelihood ratio as a sum of independent binomial random variables; it can be written as

$$\mathcal{L} = \sum_{r\leq s} \frac{m_{rs}}{m^{\max}} x_{rs} + C \tag{23a}$$

where the constants x_{rs} and C are given by

$$x_{rs} = \log\left[\frac{p_{rs}}{\rho}\frac{1-\rho}{1-p_{rs}}\right]\ , \tag{23b}$$

$$C = \sum_{r\leq s} \alpha_{rs} \log\left[\frac{1-p_{rs}}{1-\rho}\right]\ , \tag{23c}$$

and where $m_{rs} \sim \mathrm{Bin}(p_{rs}, m_{rs}^{\max})$.

The central limit theorem (CLT) predicts that the distribution of an appropriately rescaled and centered transformation of $\mathcal{L}$ will converge to the normal distribution $\mathcal{N}(0,1)$ if the number of summed random variables $q^* = q(q+1)/2$ goes to infinity. In the finite case, q^* obviously violates the conditions of the CLT, but it nonetheless offers a good approximation of the distribution of $\mathcal{L}$ (see Fig. 2).

To apply the CLT, we first define the centered and normalized variable $Z = (\mathcal{L} - C - \mu_{q^*})/S_{q^*}$ where

$$\begin{aligned} S_{q^*}^2 &= \sum_{r\leq s}\left[\left\langle \left(\frac{x_{rs}m_{rs}}{m^{\max}}\right)^2\right\rangle - \left\langle \left(\frac{x_{rs}m_{rs}}{m^{\max}}\right)\right\rangle^2\right] \\ &= \sum_{r\leq s} \frac{\alpha_{rs}}{m^{\max}}\ p_{rs}(1-p_{rs})x_{rs}^2 \end{aligned} \tag{23d}$$

is the sum of the variances of the q^* scaled binomial variables $x_{rs}m_{rs}/m_{rs}^{\max}$, and where

$$\begin{aligned} \mu_{q^*} &= \sum_{r\leq s}\left\langle \frac{x_{rs}}{m^{\max}} m_{rs}\right\rangle = \sum_{r\leq s} \alpha_{rs}p_{rs}x_{rs} \\ &\equiv h(\rho) - \sum_{r\leq s}\alpha_{rs}h(p_{rs}) - C \end{aligned} \tag{23e}$$

is the sum of their means [we have used Eq. (13) in the last step]. The CLT then tells us that $Z \sim \mathcal{N}(0,1)$, approximately.

Recall that the cumulative distribution function of a normal random variable can be expressed in terms of the error function as

$$\Pr(Z < z) = \frac{1}{2}\left[1 + \mathrm{erf}\left(\frac{z}{\sqrt{2}}\right)\right]\ . \tag{24}$$

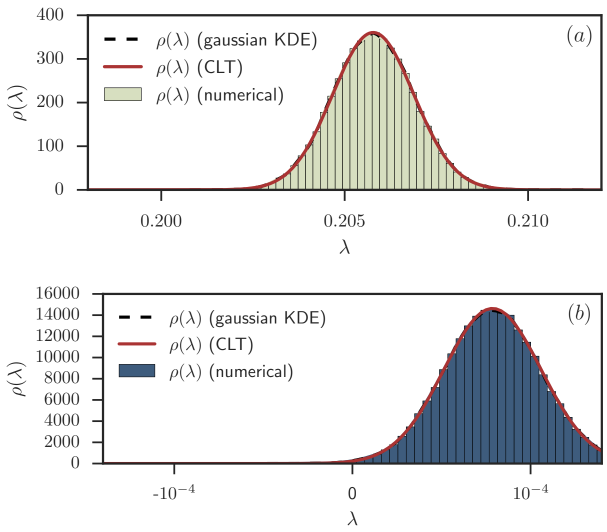

FIG. 2. (color online) Accuracy of the CLT approximation for the (a) non-uniform and (b) nearly uniform SBM of Fig. 1. Both histograms aggregate 100 000 samples of $\mathcal{L}$. The prediction of the CLT is shown in red [see Eqs. (23b)–(23e)]. We plot for comparison the Gaussian kernel density estimate (KDE) of $\rho(\lambda)$ (dashed black line, hidden by the CLT curve). Equation (25) predicts $\eta_{(a)} = 1$ and $\eta_{(b)} = 0.981(2)$; for the sample shown, numerical estimates yield $\hat{\eta}_{(a)} = 1$ and $\hat{\eta}_{(b)} = 0.980(7)$.

Now, assuming that Z is indeed normally distributed we can use the fact that $\Pr(\mathcal{L} < 0)$ is equivalent to $\Pr[Z < -(C + \mu_{q^*})/S_{q^*}]$ to compute η. Writing $\mu_{q^*} + C$ as $\langle\mathcal{L}\rangle$ [see Eq. (23e)], we find

$$\eta \approx \frac{1}{2}\left[1 + \operatorname{erf}\left(\frac{\langle\mathcal{L}\rangle}{\sqrt{2}S_{q^*}}\right)\right] , \tag{25}$$

i.e., a (approximate) equation in closed form for η.

Crucially, Eq. (25) predicts that η can never be smaller than 1/2. This comes about because (i) $\langle\mathcal{L}\rangle > 0$ and (ii) S_{q^*} is a sum of variances, i.e., a positive quantity. There are therefore two possible limits which will yield $\langle\mathcal{L}\rangle/S_{q^*} \approx 0$ and $\eta = 1/2$: Either $\langle\mathcal{L}\rangle = 0$ or $S_{q^*} \gg 0$. Some care must be exerted in analyzing the case $\langle\mathcal{L}\rangle = 0$; equations (11)–(12) tell us that the distribution of $\mathcal{L}$ is concentrated on 0 when its average is exactly equal to 0. We conclude that $\eta = 1/2$ is never reached but only approached asymptotically, for parameters that yield $\langle\mathcal{L}\rangle = \varepsilon$, with ε small but different from zero. The consequence of $\eta \geq 1/2$ is that at most half of the instances of the SBM can be declared undetectable on the account of the condition $\mathcal{L} < 0$.

C. Relation between average detectability and η–detectability

We can immediately reach a few conclusions on the interplay between the notions of average and η–detectability. First, the symmetries of $\langle\mathcal{L}\rangle$, (see Sec. IV C 1) translates into symmetries for η. To see this, first notice that S_{q^*} is conserved under the mapping $p_{rs} \mapsto 1 - p_{rs}$

$$\begin{aligned}[x_{rs}(p_{rs},\rho)]^2 &\mapsto [-x_{rs}(1-p_{rs},1-\rho)]^2 \ , \\ p_{rs}(1-p_{rs}) &\mapsto (1-p_{rs})p_{rs} \ .\end{aligned}$$

and that a permutation of the indexes $\pi(r,s)$ only changes the order of summation of the terms of S_{q^*}. Second, hypersurfaces of constant average detectability need not be hypersurfaces of constant η–detectability.

To investigate this second important aspect of the connection between average detectability and η–detectability, let us further approximate Eq. (25). The MacLaurin series of the error function is, to the first order,

$$\begin{aligned}\eta &= \frac{1}{2}\left\{1 + \frac{2}{\sqrt{\pi}}\left[\frac{\langle\mathcal{L}\rangle}{S_{q^*}} - \mathcal{O}\big(\langle\mathcal{L}\rangle^3/S_{q^*}^3\big)\right]\right\} \ , \\ &\approx \frac{1}{\sqrt{2\pi}}\frac{\langle\mathcal{L}\rangle}{S_{q^*}} + \frac{1}{2} \ .\end{aligned} \tag{26}$$

This is a reasonably accurate calculation of η when $\langle\mathcal{L}\rangle$ is small, i.e., close to the *average* undetectable regime. (Recall that we do not allow diverging S_{q^*} for the reasons stated in Sec. V B). It then becomes clear that on the hypersurfaces where $\langle\mathcal{L}\rangle = \lambda$ is constant (and close to 0),

$$\sqrt{2\pi}\left(\eta - \frac{1}{2}\right)S_{q^*} = \lambda \ , \tag{27}$$

is conserved rather than η itself. Equation (27) embodies a trade-off between accuracy (η) and variance (S_{q^*}): In the regions of the hypersurface of constant $\langle\mathcal{L}\rangle$ where the variance is large, η must be comparatively small, and vice-versa.

D. 1–detectability

Now, turning to the complementary case where $\langle\mathcal{L}\rangle$—and consequently η—is close to its maximum, we obtain a simple criterion for 1–detectability based the asymptotic behavior of $\operatorname{erf}(x)$. It is reasonable to define a (small) threshold T beyond which $\operatorname{erf}(x > T) = 1$ for all practical purposes. The error function goes asymptotically to 1 with large values of its argument, but reaches its maximum of $\operatorname{erf}(x) = 1$ very quickly, so quickly, in fact, that $\operatorname{erf}(5)$ is numerically equal to 1 to the 10^{th} decimal place.

Asking that the argument of $\operatorname{erf}(x)$ in Eq. (25) be greater than this practical threshold, we obtain the inequality

$$\langle\mathcal{L}\rangle \geq \sqrt{2}TS_{q^*} \tag{28}$$

for 1–detectability. Whenever the inequality holds, the associated ensemble is 1–detectable with a tolerance threshold T, i.e., we can say that for all practical purposes, there are no instances of the SBM which are necessarily [52] undetectable.

VI. CASE STUDY: GENERAL MODULAR GRAPHS

The stochastic block model encompasses quite a few well-known models as special cases; noteworthy examples include the *planted partition model* [40, 53], the closely related *symmetric SBM* (SSBM) [26, 28, 43], the *core-periphery model* [11], and many more. These simplified models are important for two reasons. One, they are good abstractions of structural patterns found in real networks, and a complete understanding of their behavior with respect to detectability is therefore crucial. Two, they are simple enough to lend themselves to a thorough analysis; this contrasts with the general case, where simple analytical expressions are hard to come by.

In the paragraphs that follow, we investigate the *general modular graph model* (GMGM) [34], a mathematically simple, yet phenomenologically rich simplified model. Thanks to its simpler parametrization, we obtain easily interpretable versions of the expressions and results derived in Secs. III–V.

A. Parameterization of general modular graphs

The GMGM can be seen as constrained version of the SBM, in which *pairs* of blocks assume one of two roles: Inner or outer pairs. If a pair of blocks (B_r, B_s) is of the "inner type", then one sets $p_{rs} = p_{\text{in}}$. If a pair of blocks (B_r, B_s) is of the "outer type", then one sets $p_{rs} = p_{\text{out}}$. The resulting density matrices can therefore be expressed as

$$\boldsymbol{P} = (p_{\text{in}} - p_{\text{out}})\boldsymbol{W} + p_{\text{out}}\mathbf{1}\mathbf{1}^\intercal \ , \tag{29}$$

where $\boldsymbol{W}$ is a $q \times q$ indicator matrix [$w_{rs} = 1$ if (B_r, B_s) is an inner pair], and where $\mathbf{1}$ is a length q vector of ones. A non-trivial example of a density matrix of this form is shown in Fig. 3 (a). The figure is meant to illustrate just how diverse the networks generated by the GMGM may be, but it is also important to note that the results of this section apply to *any* ensemble whose density matrix can be written as in Eq. (29). This includes, for example, the q–block SSBM, a special case of the GMGM obtained by setting $\boldsymbol{W} = \boldsymbol{I}_q$ and $\{n_r = n/q\}_{r=1,..,q}$ (see Ref. [23] for a discussion of the SSBM).

Whilst the parametrization in terms of p_{in} and p_{out} is simple, we will prefer an arguably more convoluted parameterization which is also more revealing of the natural symmetries of the GMGM (in line with the transformation proposed in Sec. IV C 2). The first natural parameter is the average density, which can be computed from Eqs. (6) and (29) and which equals

$$\begin{aligned}\rho &= \sum_{r\leq s} \alpha_{rs}[w_{rs}p_{\text{in}} + (1 - w_{rs})p_{\text{out}}] \ , \\ &= \beta p_{\text{in}} + (1 - \beta)p_{\text{out}} \ ,\end{aligned} \tag{30a}$$

where $\beta := \sum_{r\leq s} \alpha_{rs}w_{rs}$ is the fraction of *potential* edges that falls between block pairs of the inner type. The second natural parameter is simply the difference

$$\Delta = p_{\text{in}} - p_{\text{out}} \ . \tag{30b}$$

The absolute value of Δ quantifies the distance between the parameters of the GMGM and that of the equivalent random ensemble; its sign tells us which type of pairs is more densely connected. In this natural parametrization the density matrix takes on the form $\boldsymbol{P} = \rho\mathbf{1}\mathbf{1}^\intercal + \Delta(1 - \beta)\boldsymbol{W}$, i.e., a uniform matrix of ρ with perturbation proportional to $\Delta(1 - \beta)$ for the inner pairs. It might appear that we have increased the complexity of the model description, since the additional parameter β now appears in the definition of the density matrix. It is, however, not the case, because we could consider the combined parameter $\widetilde{\Delta} = \Delta(1 - \beta)$. Therefore, Eqs. (30a)–(30b), together with $\boldsymbol{W}$ and $\boldsymbol{n}$, suffice to unambiguously parametrize the model.

B. Average detectability of general modular graphs

The average normalized log-likelihood ratio $\langle\mathcal{L}\rangle$ is tremendously simplified in the natural parametrization of the GMGM; it is straightforward to show that the ratio takes on the compact (and symmetric) form

$$\begin{aligned}\langle\mathcal{L}(\rho,\Delta;\beta)\rangle = \beta&\Big\{h(\rho) - h\big[\rho + (1-\beta)\Delta\big]\Big\} \\ &+ (1-\beta)\Big\{h(\rho) - h\big[\rho - \beta\Delta\big]\Big\} \ ,\end{aligned} \tag{31}$$

by using $p_{rs} = w_{rs}p_{\text{in}} + (1 - w_{rs})p_{\text{out}}$ together with the inverse of Eqs. (30a)–(30b):

$$p_{\text{in}} = \rho + (1 - \beta)\Delta \ , \tag{32a}$$

$$p_{\text{out}} = \rho - \beta\Delta \ . \tag{32b}$$

In Fig. 3 (b), we plot $\langle\mathcal{L}(\rho,\Delta;\beta)\rangle$ in the (ρ,Δ) space—hereafter the density space—for the indicator matrix $\boldsymbol{W}$ shown in Fig. 3 (a) (and unequal block sizes, see caption). Unsurprisingly, $\langle\mathcal{L}\rangle$ is largest when the block types are clearly separated from one another, i.e., when $|\Delta|$ is the largest. Notice, however, how large separations are *not* achievable for dense or sparse networks. This is due to the fact that not all (ρ,Δ) pairs map to probabilities $(p_{\text{in}}, p_{\text{out}})$ in $[0,1]$. The region of the density space which *does* yield probabilities is the interior of the quadrilateral whose vertices are, in (ρ,Δ) coordinates: $(0,0), (\beta,1), (1,0), (1-\beta,-1)$. Changing the value of β skews this accessible region and, presumably, the functions that are defined on it, such as $\langle\mathcal{L}(\rho,\Delta;\beta)\rangle$.

We also show on Fig. 3 (b) two members of the level set defined by $\langle\mathcal{L}(\rho,\Delta;\beta)\rangle = \lambda$. As mentioned previously, the exact functional form of this family of hypersurfaces (here simply curves) seems elusive, but an approximate

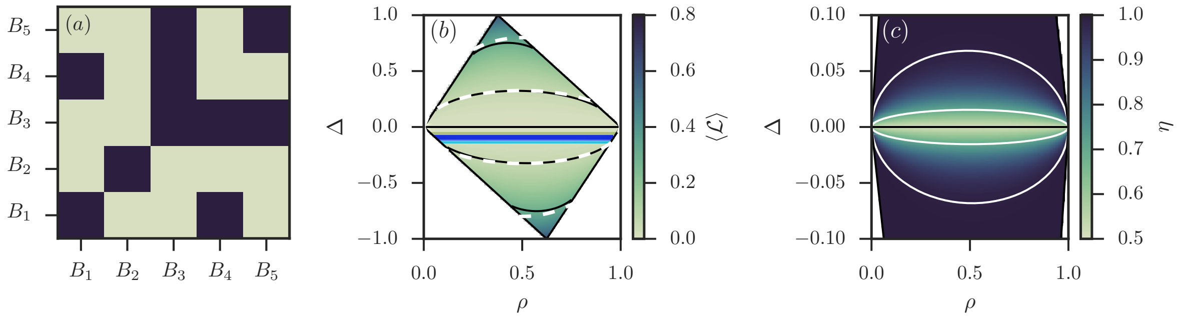

FIG. 3. (color online) Detectability in the general modular graph model. All figures use the same indicator matrix $\boldsymbol{W}$ [panel (a)] and the size vector $\boldsymbol{n} = [10, 30, 20, 20, 20]$ ($n = 100$ nodes). (a) Example of density matrix $\boldsymbol{P}$ allowed in the GMGM. Dark squares indicate block pairs of the inner type and light squares indicate pairs of the outer type. (b) Average detectability in the density space of the GMGM. Both the numerical solution of $\langle\mathcal{L}\rangle = \lambda$ (solid black line) and the prediction of Eq. (34) (dashed white line) are shown, for $\lambda = 0.05$ and 0.3. (c) $\eta(\rho, \Delta; \beta)$ in the density space of the GMGM; notice the change of Δ–axis. Solid white lines are curves where $\eta(\rho, \Delta; \beta) = \eta^*$, with $\eta^* = 0.7$ (central curve) and $\eta^* = 0.99$ (outer curve). Equation (25) is used to compute both η and η^*.

solution is available. Using the method highlighted in Sec. IV, we find, to the second order,

$$\begin{aligned} 2\lambda\rho(1-\rho) &\approx \sum_{r\leq s} \alpha_{rs}(p_{rs} - \rho)^2 \\ &= \beta[(1-\beta)\Delta]^2 + (1-\beta)(\beta\Delta)^2 \ . \end{aligned} \tag{33}$$

Equation (33) fixes the relative value of all parameters on the line where $\langle\mathcal{L}\rangle = \lambda$. Solving for Δ, we find

$$\Delta^*(\rho; \lambda, \beta) = \pm\sqrt{2\lambda\frac{\rho(1-\rho)}{\beta(1-\beta)}} \ , \tag{34}$$

also shown on Fig. 3 (b) for comparison.

Figure 3 highlights the accuracy of our approximation when λ is small. But it also highlights its inaccuracy when λ is large; $\lambda \gg 1$ forces $\Delta^*(\rho; \lambda, \beta)$ to pass through a region where $\Delta^* \approx 1$, i.e., a region where the omitted terms on the RHS of Eq. (33) contribute heavily. Fortunately, this is not so problematic, since most detectability related phenomena—phase transitions, undetectable instances, etc.—arise near $\Delta = 0$, i.e., where the approximation works.

C. η–detectability of general modular graphs

While $\langle\mathcal{L}(\rho, \Delta; \beta)\rangle$ takes on a particularly compact form once we substitute $\{p_{rs}\}$ by the natural parameters of the GMGM, the same cannot be said of $\eta(\rho, \Delta; \beta, n)$. Some analytical progress can be made by, e.g., noticing that only two types of terms are involved in the calculation of S_{q^*}, but, ultimately, the resulting expression is no more useful than the simple Eq. (25) and Eq. (26). We will therefore omit the calculation of η.

In Fig. 3 (c) we plot $\eta(\rho, \Delta; \beta, n)$ in the density space [using Eq. (25)]. We also display the numerical solutions of $\eta(\rho, \Delta; \beta, n) = \eta^*$ for two values of η^*. The figure highlights just how quickly η goes to 1 as a function of Δ, even for the fairly small system sizes considered: We find that $\eta \geq 0.99$ for *any* value of ρ, as soon as $\Delta > 0.06$. The condition (9) is therefore a *weak* one. It allows us to determine that some parameters are overwhelmingly undetectable, but only when Δ is very close to 0.

Figure 3 also shows how increases in variance translate into decreases in accuracy [see Eq. (27)]: Following a line of constant (and relatively small) Δ, one can see that η is minimized close to $\rho = 1/2$, i.e., near the maximum of variance. This is characteristic of many parameterizations of the SBM and GMGM; it turns out that, for fixed n, impossible detection problems are not confined to vanishing densities. In fact, values of ρ closer to $1/2$ are associated with a comparatively larger interval of Δ for which detection is impossible.

D. Symmetries of general modular graphs

In Secs. IV–V, we have proved that there are $2q^*!$ transformations that preserve $\langle\mathcal{L}(\rho, \Delta; \beta)\rangle$ and $\eta(\rho, \Delta; \beta, n)$. We could therefore go about computing the symmetries of the GMGM by listing all of these transformations in terms of (ρ, Δ, β). But since there are only 3 free parameters in the GMGM, we can also choose an alternative route and directly solve $\langle\mathcal{L}(\rho, \Delta; \beta)\rangle = \langle\mathcal{L}(a_1\rho + b_1, a_2\Delta + b_2; a_3\beta + b_3)\rangle$ by, e.g., obtaining a linear system from the Taylor series of $\langle\mathcal{L}(\rho, \Delta; \beta)\rangle$. This simpler approach yields the following set of λ–preserving transformations for the model:

$$(\rho, \Delta, \beta) \mapsto (\rho, \Delta, \beta) \ , \tag{35a}$$
$$(\rho, \Delta, \beta) \mapsto (\rho, -\Delta, 1-\beta) \ , \tag{35b}$$
$$(\rho, \Delta, \beta) \mapsto (1-\rho, \Delta, 1-\beta) \ , \tag{35c}$$
$$(\rho, \Delta, \beta) \mapsto (1-\rho, -\Delta, \beta) \ . \tag{35d}$$

It is straightforward to check that these transformations form a group, whose product is the composition of two transformations. A Cayley table reveals that the group is isomorphic to the Klein four-group $Z_2 \times Z_2$.

One immediately notices a large gap between the number of symmetries predicted by the calculations of Sec. IV C 1 ($2q^*!$) and the number of symmetries appearing in Eq. (35) (4, independent of q). The gap is explained by the fact that every symmetry of the general SBM maps onto one of the four transformations listed in Eq. (35) [54] A sizable fraction of the symmetries reduce to Eq (35a), since permutations $\pi(r,s)$ cannot modify the natural parameters of the GMGM: The type of block pair (B_r, B_s)—characterized by p_{rs}—is permuted alongside its share of potential edges α_{rs}. Another important fraction of the symmetries is accounted for by the "graph complement transformation": Any transformation $\boldsymbol{P} = \boldsymbol{1}\boldsymbol{1}^\intercal - \boldsymbol{P}$ plus a permutation reduces to Eq. (35d). This leaves two symmetries, which happen to be artifacts of our choice of nomenclature. To see this, *rename* pair types, i.e., call inner pairs "outer pairs" and vice-versa. Neither the density ρ nor $|\Delta|$ will change. But both the sign of Δ and the value of β will be altered. With this in mind, it becomes clear that Eq. (35b) corresponds to the permutation symmetry, and that Eq. (35c) corresponds to the graph complement symmetry, both up to a renaming of the types.

E. Where the framework is put to the test: Inference

1. Procedure

It will be instructive to put our framework to the test and compare its predictions with numerical experiments that involve inference, i.e., the detection of the planted partition of actual instances of the GMGM. We will use the following procedure: (i) generate an instance of the model, (ii) run an inference algorithm on the instance, and (iii) compute the correlation of the inferred and planted partition (see below for a precise definition). The average detectability $\langle\mathcal{L}\rangle$ should bound the point where the average correlation becomes significant, and η–detectability should give an upper bound on the fraction of correlated instances.

Even for the small size considered, it is impossible to compute all quantities involved in the process exactly; we therefore resort to sub-sampling. We use an efficient algorithm [55] based on the Metropolis-Hastings algorithm of Ref. [17], which, unlike Belief Propagation [28], works well for dense networks with many short loops. The principle of the algorithm is to construct an ergodic chain of partitions $\mathcal{B}_0, ..., \mathcal{B}_T$, and to sample from the chain to approximate the probability

$$\mu_i^r(G) = \sum_{\{\mathcal{B}_\sigma\}} \Pr(\mathcal{B}_\sigma | G, \boldsymbol{P}, \boldsymbol{n}) \delta(\sigma(v_i) = r) \tag{36}$$

that node v_i is in block B_r, given a network G and some parameter $\boldsymbol{P}$ and $\boldsymbol{n}$. It is easy to see that one can then maximize the probability of guessing the partition correctly by assigning nodes according to [31]

$$\hat{\sigma}(v_i) = \mathrm{argmax}_r(\mu_i^r) \ . \tag{37}$$

We choose a simple set of moves that yields an ergodic chain over all $\{\mathcal{B}\}$: at each step, we change the block of a randomly selected node v_i from $\sigma(v_i) = B_r$ to a randomly and uniformly selected block B_s, with probability $\min\{1, \mathcal{A}\}$, where

$$\begin{aligned}\mathcal{A} = &\left[\frac{p_{rs}(1-p_{rr})}{p_{rr}(1-p_{rs})}\right]^{k_r^{(i)}}\left[\frac{p_{ss}(1-p_{rs})}{p_{rs}(1-p_{ss})}\right]^{k_s^{(i)}}\\ &\times\left[\frac{1-p_{rs}}{1-p_{rr}}\right]^{n_r-1}\left[\frac{1-p_{ss}}{1-p_{rs}}\right]^{n_s}\\ &\times\prod_{l\neq r,s}\left[\frac{p_{ls}(1-p_{rl})}{p_{rl}(1-p_{ls})}\right]^{k_l^{(i)}}\left[\frac{1-p_{ls}}{1-p_{rl}}\right]^{n_l} \quad ,\end{aligned} \tag{38}$$

and $k_l^{(i)}$ the number of neighbors of node v_i in block B_l [17]. The space of all partitions is obviously connected by this move set, and the possibility of re-sampling a configuration ensures that the chain is aperiodic. Furthermore, since transition probabilities are constructed according to the prescription Metropolis-Hastings, the chain is ergodic and samples from $\mathbb{P}(\mathcal{B}|G, \boldsymbol{P}, \boldsymbol{n})$. Note that we assume that $\boldsymbol{P}$ is known when we compute (36). Learning the parameters can be done separately, see Ref. [31] for example.

In the spirit of Refs. [28, 31], we initialize the algorithm with the planted partition itself. This ensure that we will achieve the information–theoretic threshold, even if efficient inference is impossible [31]. To see this, first consider the case where the planted partition is information-theoretically detectable. In this case, the chain will concentrate around the initial configuration, and the marginal distribution [Eq. (36)] will yield a distribution correlated with the planted partition. We will have to proceed with care, however, since two scenarios may occur in the information-theoretically undetectable phase. If there is no hard-phase—e.g., when $q = 2$ [32]—, the algorithm will show no particular preference for the initial configuration and wander away towards partitions uncorrelated with the planted partition. But if there is a hard-phase, one will have to wait for a period that diverges exponentially in the system size before the sampler becomes uncorrelated with its initial state [31]. This complicates convergence diagnosis and can lead one to conclude that correlated inference is possible even though it's not. To avoid these difficulties, we will simply restrict ourselves to the cases where the hard-phase does not exist [23].

Once the estimated partition $\hat{\mathcal{B}}$ is obtained via Eq. (37), we compute its correlation with $\mathcal{B}$—the planted partition—using a measure that accounts for finite size effects. The so-called renormalized normalized mutual information (rNMI) of Ref. [56] appears a good choice. Much like the well-known NMI [57, 58], the rNMI is bounded to the [0, 1] interval, and $\mathrm{rNMI}(\mathcal{B}_p, \hat{\mathcal{B}}) = 1$ means that the planted partition $\mathcal{B}_p$ and the inferred partition $\hat{\mathcal{B}}$ are identical. Unlike the NMI, $\mathrm{rNMI}(\mathcal{B}_p, \hat{\mathcal{B}}) = 0$ signals the absence of correlation between the two partitions, even in finite networks.

2. *Results*

In Fig. 4 (a), we plot $\langle \mathrm{rNMI}(\mathcal{B}_p, \hat{\mathcal{B}})\rangle$ in the density space of the GMGM. We use the parameters $\boldsymbol{W} = \boldsymbol{I}$, and $\boldsymbol{n} = [n/2, n/2]$ (i.e., the SSBM), since the resulting ensemble is conjectured to be the hardest of all, with respect to detectability [31]. Two important parallels can be drawn between the results shown in Fig. 4 (a) and the functional form of $\langle \mathcal{L}(\rho, \Delta; \beta)\rangle$ and $\eta(\rho, \Delta; \beta, n)$ [shown in Fig. 3 (b)–(c) for a different GMGM]. First, notice how the boundary that marks the onset of the (theoretically) 1–detectable region partitions the density space in two qualitative regimes: A regime where perfect detection is possible *for all instances*, and a region where it is not. There is, of course, some level of arbitrariness involved in selecting the threshold T [see Eq. (28)]. But the fact that a line of constant η partitions the space is a hint that while $\mathcal{L} < 0$ is not sufficient for undetectability, there exists a level of significant λ^* for which $\mathcal{L}$ properly separates detectable and undetectable instances.

The second important parallel concerns hypersurfaces of constant $\langle \mathcal{L}\rangle$ and their connection with $\langle \mathrm{rNMI}\rangle$. We have argued in Sec. IV that $\langle \mathcal{L}\rangle$ is a good predictor of the accuracy of an optimal inference algorithms (with potentially exponential complexity). It should therefore not be surprising that there is an hypersurface of constant $\langle \mathcal{L}\rangle$ which *also* partitions the density space in two qualitative regions [59]: One where $\langle \mathrm{rNMI}\rangle \approx 0$ and one where $\langle \mathrm{rNMI}\rangle$ is clearly greater than zero. On this hypersurface, the average level of significance is the same for all parameterizations of the GMGM; our results show that the inference algorithm achieves correspondingly uniform accuracy for all parameters on the surface.

One could argue that these parallels are not so obvious in Fig. 4 (a); we therefore focus on a subset of the density space in Fig. 4 (b)-(c) to make our case clearer. In these figures, we plot the same information, but only for networks of constant density $\rho = 0.25$ and size $n = 100$ (b) and $n = 500$ (c). We also show the probability $\Pr(\mathrm{rNMI}(\mathcal{B}_p, \hat{\mathcal{B}}) > 0)$ that the inferred partition is correlated with the planted partition. This a direct measurement of the fraction of detectable instances, which we compare against $\eta(\Delta; \rho, \beta, n)$. It never reaches 0, because random fluctuations produce correlated partitions even when $\mathbb{P} = \mathbb{Q}$ (the rNMI corrects for the *average* correlation). If $\mathcal{L} > 0$ were a necessary and sufficient condition for detectability, then $\eta(\Delta; \rho, \beta, n)$ and $\Pr(\mathrm{rNMI} > 0|\Delta, \rho, \beta, n)$ would correspond perfectly. But since $\mathcal{L} > 0$ is only a *necessary* condition, $\eta(\Delta)$ acts as an upper bound rather than an exact expression, i.e., $\Pr(\mathrm{rNMI} > 0; \eta)$ can never be greater than $\eta(\Delta)$.

Two further observations must be made. First, it is

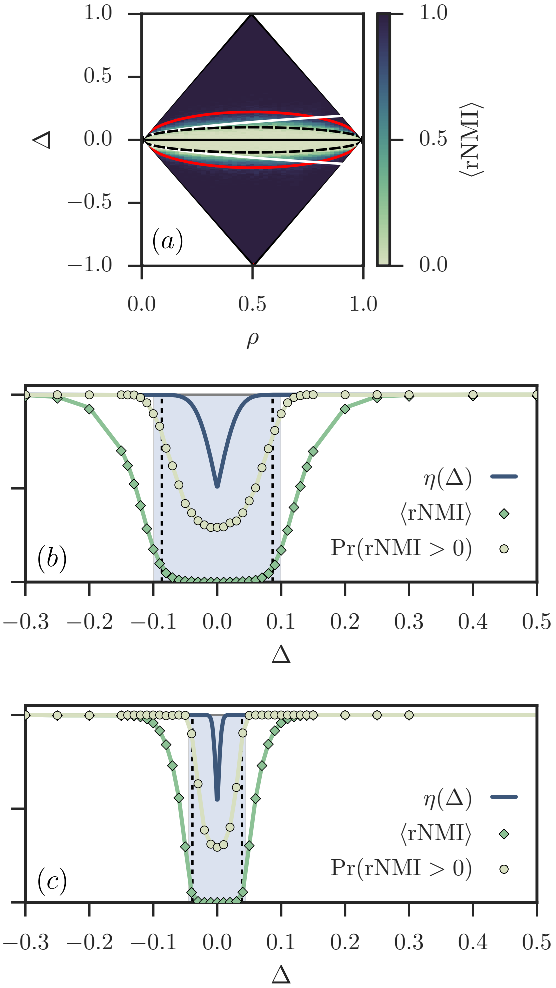

FIG. 4. (color online) Inference of the GMGM. All figures show results for the special case $q = 2$, $\boldsymbol{W} = \boldsymbol{I}_2$ and $\boldsymbol{n} = [n/2, n/2]$, corresponding to the $q = 2$ SSBM [23]. All empirical results are averaged over 10^4 independent instances of the SBM. (a) Average rNMI of the planted and the inferred partition in the density space of the model of size $n = 100$. Solid red lines mark the boundaries of the 1–detectability region, with tolerance threshold $T = 4\sqrt{2}$, see Eq. (28). Dotted black lines show the two solutions of $\Delta^*(\rho; \lambda = 1/2n, \beta)$, see Eq. (34). White lines show the finite size KS-bound, before it is adjusted for the symmetries of the problem. (b)–(c) Phase transition at constant $\rho = 0.25$ for networks of $n = 100$ nodes (b) and $n = 500$ nodes (c). Circles indicate the fraction of instances for which a correlated partition could be identified, while diamonds show the average of the rNMI (lines are added to guide the eye). Blue solid curves show $\eta(\Delta; \rho, \beta, n)$, see Eq. (25). The shaded region lies below the finite size Kesten-Stigum bound $\Delta = \pm q\sqrt{\rho/n}$ (here with $q = 2$). The dotted lines show the two solutions of $\Delta^*(\rho; \lambda = 1/2n, \beta = 1/2)$.

known that in the sparse two-blocks SSBM, the transition between the information-theoretically undetectable and detectable regions occurs on the so-called Kesten-Stigum (KS) bound—located at $\Delta = \pm q\sqrt{\rho/n}$ for finite size instances (this is not generally true, but the equivalence holds when $q = 2$ [32]). Despite the fact that this bound was derived for infinite ensembles, it holds very well in the finite case, as shown in Fig. 4 (b)–(c). But the finite size approach has the potential to be more precise. Building upon the interpretation of $\langle\mathcal{L}\rangle$ as a measure of the average difficulty of the inference problem, we set a threshold $\langle\mathcal{L}\rangle = 1/2n$ on the average detectability. For this choice of threshold, the approximate hypersurface equation predicts a transition at

$$\Delta^* = \pm 2\sqrt{\rho(1-\rho)/n}\ ,$$

very close to the KS bound, but with a correction for nonvanishing densities. Interestingly, one can motivate this choice of threshold with Random Matrix Theory [43, 60, 61] (see Appendix B for details) or the theory of low-rank matrix estimation [62]. The uncorrected and corrected bounds are shown on Fig. 4 (a). The corrected bound is qualitatively accurate in all density regimes, unlike the KS bound.

Second, in asymptotic theories, the SBM is either said to be undetectable with overwhelming probability, or the converse. The finite size approach is more nuanced in the sense that it accounts for random fluctuations, which are also manifest in empirical results [see the curves $\Pr(\mathrm{rNMI}(\mathcal{B}_p, \hat{\mathcal{B}}) > 0)$]. While η–detectability is not perfect, as is argued above, it nonetheless goes through a smooth transition instead of an abrupt one. This reflects the continuous nature of the finite size transition.

VII. CONCLUSION

Building upon ideas from statistical theory, we have developed a framework to study the information-theoretic detectability threshold of the finite size SBM. Our analysis relies on two different interpretations of the log-likelihood ratio $\mathcal{L}$ of the SBM and its equivalent random ensemble. We have used the rigorous interpretation of $\mathcal{L}$ to put a necessary condition on detectability. We have then computed the distribution of $\mathcal{L}$, and proved that up to half of the instances of the finite size SBM could be declared undetectable on the basis of this simple test alone. We have further argued that the average of $\mathcal{L}$ could be interpreted as a proxy for the performance of an optimal inference algorithm (with possibly exponential running time). This interpretation has proved to be fruitful; starting with a compact form for $\langle\mathcal{L}\rangle$, we have established the existence of a large equivalence class with respect to average detectability. In Appendix A, we have shown that $\mathcal{L}$ can also be used to prove that, quite naturally, detectability decreases when the datasets are noisy. Using a correspondence with the finite size information–theoretic-threshold (as well as with Random Matrix Theory, see Appendix B), we have presented numerical evidence that the hypersurface $\langle\mathcal{L}\rangle = 1/2n$ separates detectable from undetectable instances in a special case of the SBM.

The unifying theme of this contribution has been the idea that $\langle\mathcal{L}\rangle$ quantifies both detectability and consistency in the finite size SBM. This interpretation leaves many questions open for future works. Perhaps the most important of all: Can one determine the threshold within the framework of the theory itself, for general SBM?

A second important question pertains to sufficiency: Can one modify the condition to make it necessary *and* sufficient? Or is a completely different approach needed? In asymptotic analyses of the limit, one can use different conditions to bound the limit from above and below, as is done in Ref. [33]. Can a similar approach be fruitfully applied to finite instances?

In closing, let us mention a few of the many possible generalizations of the methods introduced. First, it will be important to verify how our approach behaves in the limit $n \to \infty$. How this limit is taken will matter. In particular, we believe that our framework has much to say about the limit where $q \to \infty$, since it does not assume Poisson distributed degree, unlike other asymptotic theories of the limit. Second, we see no major obstacle to a generalization of our methods to other generative models of networks with a mesoscopic structure. This includes, for example, the consistency of graphons, a subject whose study has been recently undertaken [63]. Changing the null model from the equivalent random network ensemble to the configuration model [64, 65] could even allow an extension to degree-corrected SBM [66].

ACKNOWLEDGMENTS

We thank Charles Murphy and Guillaume St-Onge for useful discussions and comments. This work has been supported by the Fonds de recherche du Québec-Nature et technologies (J.-G.Y., P.D), the Conseil de recherches en sciences naturelles et en génie du Canada (L.J.D.) and the James S. McDonnell Foundation Postdoctoral Fellowship (L.H.-D.), and Grant DMS-1622390 from the National Science Foundation (L.H.-D.). P.D. and E.L. are grateful to D. Côté (P.D., E.L) and P. Mathieu (E.L.) for financial support.

Appendix A: Detectability and noise

One almost never has a perfect knowledge of the structure of real networks. The culprit can lie at the level of data collection, storage, transmission—or a combination of the above—, but the outcome is the same: Some edges are spurious and others are omitted [67]. To model imperfect knowledge, we will suppose that instances of the SBM first go through a noisy channel where

T modifications—random edge removals or additions—are applied to the structure. Only then are we asked to tell which of hypotheses $\mathbb{P}$ and $\mathbb{Q}$ is the most likely. It should be clear that it will be more difficult to separate the two hypotheses, since noise is not necessarily aligned with the planted partition.

We will approach the problem with the following *universal perturbation process* (UPP): At each step t of this process, a new random edge is added with probability c; otherwise, a random edge is removed. If a new edge must be added, then it is selected uniformly from the set of non-edges. If an edge must be removed, then it is selected uniformly from the set of edges already present in the network. This randomization step is then repeated T times. We call this process universal because one can map arbitrary perturbation patterns onto one or successive UPPs with different parameters c.

To prove that $\langle\mathcal{L}\rangle$ decreases as a result of any sufficiently long UPP, we will show that the total derivative

$$\frac{d}{dt}\langle\mathcal{L}\rangle = \sum_{r\leq s}\frac{\partial\langle\mathcal{L}\rangle}{\partial p_{rs}}\frac{dp_{rs}(t)}{dt} \tag{A1}$$

is negative everywhere. In so doing, we assume that the process can be approximated as a continuous one (both with regards to "time" t and discrete quantities such as m_{rs}). Admittedly, a more rigorous approach would be needed to settle the matter unequivocally, but we argue that the method presented in this Appendix give a good intuition for the problem.

Without specifying the dynamics, and using Eq. (13), one can compute

$$\frac{\partial\langle\mathcal{L}\rangle}{\partial p_{rs}} = \alpha_{rs}\log\left[\frac{p_{rs}}{\rho}\frac{1-\rho}{1-p_{rs}}\right] = \alpha_{rs}x_{rs}\ , \tag{A2}$$

where x_{rs} is identical to Eq. (23b). This leaves the $\dot{p}_{rs}(t)$ terms, whose expressions are determined by the perturbation dynamics. For the UPP, the evolution of $\{m_{rs}(t)\}_{r\leq s}$ is determined by the set of differential equations

$$\dot{m}_{rs}(t) = -\frac{(1-c)[m_{rs}(t)]}{\sum_{r\leq s}m_{rs}(t)} + \frac{c\ [m_{rs}^{\max}-m_{rs}(t)]}{m^{\max}-\sum_{r\leq s}m_{rs}(t)}. \tag{A3}$$

The first term accounts for edge removal events, which occur with probability $(1-c)$ and involve edges that connect nodes in blocks (B_r, B_s) with probability $m_{rs}/\sum m_{rs}(t)$. A similar argument leads to the second term, which accounts for edge creation events.

Equation (A3) can be transformed into an equation for $\dot{p}_{rs}(t)$ by dividing through by $m_{rs}^{\max}$, and then using the definitions $p_{rs}(t) = m_{rs}(t)/m_{rs}^{\max}$ and $\rho(t) = \sum_{r\leq s}m_{rs}(t)/m^{\max}$. We find

$$\dot{p}_{rs}(t) = \binom{n}{2}^{-1}\left[c\,\frac{1-p_{rs}(t)}{1-\rho(t)} - (1-c)\,\frac{p_{rs}(t)}{\rho(t)}\right]\ , \tag{A4}$$

which, upon substitution in Eq. (A1), yields

$$\frac{d\langle\mathcal{L}\rangle}{dt} = \Theta\sum_{r\leq s}\alpha_{rs}\log\left[\frac{f(p_{rs})}{f(\rho)}\right]\left[\frac{f(c)f(\rho)}{f(p_{rs})}-1\right]\ , \tag{A5}$$

where $\Theta = [2(1-c)p_{rs}]/[\rho n(n-1)]$ is a non-negative factor, and where we have defined $f(x) = x/(1-x)$. It turns out that the sum is not only globally negative, but that each term is also individually negative, i.e.,

$$-\log\left[\frac{f(\rho)}{f(p_{rs})}\right]\left[\frac{f(c)f(\rho)}{f(p_{rs})}-1\right]\leq 0 \qquad \forall r\leq s. \tag{A6}$$

This comes about because the sign of the logarithm always matches that of the bracket.

To prove this statement, we treat 5 different cases and use the following identities repeatedly:

$$\frac{f(x)}{f(y)} < 1 \qquad \Longrightarrow x < y\ , \tag{A7}$$

$$\frac{f(c)f(\rho)}{f(p_{rs})} > 1 \Longrightarrow c > \frac{p_{rs}(1-\rho)}{\rho(1-p_{rs})+p_{rs}(1-\rho)}. \tag{A8}$$

The cases are:

1. If $\rho = p_{rs}$: The logarithm equals 0 and the upper bound of Eq. (A6) holds.

2. If $p_{rs} < \rho$ and $c < 1/2$: The logarithm is positive [see Eq. (A7)]. The bracket is also positive, since inequality (A8) can be rewritten as $(1-\rho)p_{rs}\leq \rho(1-p_{rs})$ using the fact that $c<1/2$. This simplifies to $p_{rs}\leq\rho$, in line with our premise.

3. If $p_{rs} < \rho$ and $c \geq 1/2$: The logarithm is positive. Using our premise, we conclude that $f(\rho)/f(p_{rs}) > 1$ and $f(c)\geq 1$. Therefore, $f(c)f(\rho)/f(p_{rs}) > 1$, i.e., the bracket is positive.

4. If $p_{rs} > \rho$ and $c \leq 1/2$: The logarithm is negative. Using our premise, we conclude that $f(\rho)/f(p_{rs}) < 1$ and $f(c)\leq 1$. Therefore, $f(c)f(\rho)/f(p_{rs}) < 1$, i.e., the bracket is negative.

5. If $p_{rs} > \rho$ and $c > 1/2$: The logarithm is negative. The bracket is also negative, since the converse of inequality (A8) can be rewritten as $(1-\rho)p_{rs}\geq \rho(1-p_{rs})$ using the fact that $c>1/2$. This simplifies to $p_{rs}\geq\rho$, in line with our premise.

This list covers all cases, and therefore completes the proof that $d\langle\mathcal{L}\rangle/dt\leq 0$, i.e., that average detectability decreases as a result of the application of a UPP.

Appendix B: Connection with Random Matrix Theory

In Refs. [43, 61] it is argued that SBM is not efficiently detectable when the extremal eigenvalues of the modularity matrix of its instances merge with the so-called "continuous eigenvalue band". It is proved in Ref. [43] that this occurs when

$$n(p_{\text{in}}-p_{\text{out}}) = \pm\frac{1}{n}\sqrt{2n(p_{\text{in}}+p_{\text{out}})}\ , \tag{B1}$$

for the 2 block SSBM with Poisson distributed degrees. Furthermore, in this case, there is no so-called hard-phase [32], meaning that the above limit affords a comparison with the prediction if our information theoretic framework.

Since we are concerned with the finite case, let us first modify this result to account for binomial distributed degrees instead. It turns out that the corrected condition is found by substituting the expectations of Poisson variables [in the RHS of Eq. (B1)] by that of binomial variables. This leads to

$$(p_{\mathrm{in}} - p_{\mathrm{out}}) = \pm\frac{1}{n}\sqrt{2n[p_{\mathrm{in}}(1-p_{\mathrm{in}}) + p_{\mathrm{out}}(1-p_{\mathrm{out}})]}\ , \tag{B2}$$

or, in terms of the natural parameters of the GMGM,

$$\Delta^* = \pm\sqrt{\frac{4}{n-1}\rho(1-\rho)}\ . \tag{B3}$$

This equation bears striking similarity with Eq. (34), our approximate equation for curves of constant $\langle\mathcal{L}\rangle$. In fact, for the 2 block SSBM ($\beta \approx 1/2$), the latter reads

$$\Delta^* = \pm\sqrt{8\lambda\rho(1-\rho)}\ . \tag{B4}$$

One obtains an exact equivalence between the two expressions by setting $\lambda = 1/2(n-1) \approx 1/2n$. The fact that modularity based spectral methods cannot infer a correlated partition if $\Delta \leq \Delta^*$ [Eq. (B3)] can thus be understood as stemming from a lack of statistical evidence for the SBM.

Appendix C: Detailed proofs

1. Symmetries of the average detectability

Theorem 1 (λ–preserving symmetries)**.** *All transformations $T(\boldsymbol{\alpha}, \boldsymbol{P})$ of the parameter space of the SBM that are (i) reversible, (ii) space-preserving, and (iii) valid at every point of the parameter space can be written as*

$$p_{rs} \mapsto p'_{rs} = \gamma_{rs} + (1-2\gamma_{rs})p_{\omega(r,s)}\ , \tag{C1a}$$

$$\alpha_{rs} \mapsto \alpha'_{rs} = \alpha_{\pi(r,s)}\ , \tag{C1b}$$

where $\gamma_{rs} \in \{0,1\}$ and where π and ω are permutations that acts on the set $\{(r,s)\,|\,1 \leq r, \leq s \leq g\}$. Under the additional constraint that $\langle\mathcal{L}(\boldsymbol{\alpha},\boldsymbol{P})\rangle$ be preserved by $\{T\}$ and equal to λ, one must have

$$\pi = \omega \quad \textit{and} \quad \gamma_{rs} = \gamma \quad \forall (r,s)\ .$$

Let us first introduce new notations to clarify the proof of Theorem 1. First, we define vectors $|p\rangle$ and $|\alpha\rangle$ whose entries are the $q^* = \binom{q}{2} + q$ entries of the upper triangle (and diagonal) of $\boldsymbol{P}$ and $\boldsymbol{\alpha}$. In this notation, we write the average density as $\langle\alpha|p\rangle$ and the average detectability as

$$\langle\mathcal{L}(\boldsymbol{\alpha},\boldsymbol{P})\rangle = \langle\alpha|u(\boldsymbol{\alpha},\boldsymbol{P})\rangle\ , \tag{C2}$$

where $|u(\boldsymbol{\alpha},\boldsymbol{P})\rangle$ is q^*–dimensional vector parametrized by $(\boldsymbol{\alpha},\boldsymbol{P})$, whose entries are given by

$$u_{rs}(\boldsymbol{\alpha},\boldsymbol{P}) = p_{rs}\log\frac{p_{rs}}{\langle\alpha|p\rangle} + (1-p_{rs})\log\frac{1-p_{rs}}{1-\langle\alpha|p\rangle}\ .$$

We also introduce $\boldsymbol{\Pi}$ and $\boldsymbol{\Omega}$, two $q^* \times q^*$ permutation matrices such that $\boldsymbol{\Pi}\,|\alpha\rangle_{rs} = \alpha_{\pi(r,s)}$ and $\boldsymbol{\Omega}\,|p\rangle_{rs} = p_{\omega(r,s)}$, where $|a\rangle_{ij}$ is the element (i,j) of vector $|a\rangle$. In this notation, Eqs. (C1) are given by

$$\begin{aligned} |\alpha\rangle &\mapsto |\alpha'\rangle = \boldsymbol{\Pi}\,|\alpha\rangle\ , \\ |p\rangle &\mapsto |p'\rangle = \boldsymbol{\Gamma}\,|1\rangle + (\boldsymbol{I} - 2\boldsymbol{\Gamma})\boldsymbol{\Omega}\,|p\rangle \\ &\qquad\equiv \boldsymbol{\Omega}\boldsymbol{\Gamma}'\,|1\rangle + \boldsymbol{\Omega}(\boldsymbol{I} - 2\boldsymbol{\Gamma}')\,|p\rangle\ , \end{aligned}$$

where $\boldsymbol{\Gamma}$ is a diagonal matrix with element γ_{rs} on the diagonal, where $\boldsymbol{I}$ is the identity matrix, and where $\boldsymbol{\Gamma}' = \boldsymbol{\Omega}^{-1}\Gamma$ is also a diagonal matrix.

Proof. The proof of the first part of Theorem 1 (form of the transformations) is given in the main text, see Sec. IV C 1.

To prove the second part of the theorem (constrained transformations), we look for the subset of all transformations of the form (C1) that also preserve $\langle\mathcal{L}\rangle$, i.e., transformations T in $S_{q^*} \times B_{q^*}$ that map $(\boldsymbol{\alpha},\boldsymbol{P})$ to $(\boldsymbol{\alpha}',\boldsymbol{P}')$ and that satisfy

$$\langle\alpha|u(\boldsymbol{\alpha},\boldsymbol{P})\rangle = \langle\alpha'|u(\boldsymbol{\alpha}',\boldsymbol{P}')\rangle\ .$$

It is easy to check that if $\boldsymbol{\Omega} = \boldsymbol{\Pi}$ and $\boldsymbol{\Gamma} = \gamma\boldsymbol{I}$ with $\gamma \in \{0,1\}$, then the average density and the normalized log-likelihood are both preserved. Therefore, if the transformations are of the proposed form, then λ is preserved.

To complete the proof we must show that $\langle\mathcal{L}\rangle$ is conserved *only if* $\boldsymbol{\Gamma} = \gamma\boldsymbol{I}$ and $\boldsymbol{\Omega} = \boldsymbol{\Pi}$. First, we note that by the properties of the scalar product and permutation matrices, we have the following obvious symmetry

$$\langle\alpha|u\rangle = \langle\boldsymbol{\Pi}\alpha|\boldsymbol{\Pi}u\rangle\ ,$$

which is valid for all permutation matrices $\boldsymbol{\Pi}$. We use this symmetry to "shift" all permutation matrices to the second part of the scalar product representation of $\langle\mathcal{L}\rangle$, i.e., we write

$$\langle\alpha|u\rangle \mapsto \langle\alpha'|u'\rangle = \langle\boldsymbol{\Pi}\alpha|u'\rangle = \langle\alpha|\boldsymbol{\Pi}^{-1}u'\rangle\ .$$

Now, from Eq. (C2), it is clear that we will have $\langle\mathcal{L}(\boldsymbol{\alpha},\boldsymbol{P})\rangle = \langle\mathcal{L}(\boldsymbol{\alpha}',\boldsymbol{P}')\rangle$ if and only if

$$\langle\alpha|u - \boldsymbol{\Pi}^{-1}u'\rangle = 0\ , \tag{C3}$$

where $|u'\rangle := |u(\boldsymbol{\alpha}',\boldsymbol{P}')\rangle$. Since $|u - \boldsymbol{\Pi}^{-1}u'\rangle$ is analytic in $\boldsymbol{\alpha}$, we can expand it by using Taylor series; this creates an infinite series of constraints that must all be satisfied. In particular, condition (C3) will be satisfied only if

$$|u - \boldsymbol{\Pi}^{-1}u'\rangle = |0\rangle\ .$$

This is true if and only if, for all (r,s), one has

$$\begin{aligned} p_{rs}\log\frac{p_{rs}}{\langle\alpha|p\rangle} + (1-p_{rs})\log\frac{1-p_{rs}}{1-\langle\alpha|p\rangle} \\ = \bar{p}_{rs}\log\frac{\bar{p}_{rs}}{\langle\alpha|\bar{p}\rangle} + (1-\bar{p}_{rs})\log\frac{1-\bar{p}_{rs}}{1-\langle\alpha|\bar{p}\rangle}\ , \end{aligned} \tag{C4}$$

where $|\bar{p}\rangle = \mathbf{\Pi}^{-1}\,|p'\rangle$. Here, $|\bar{p}\rangle$ is the transformed vector $|p'\rangle$, on which the inverse of permutation $\pi(r,s)$ is also applied.

Let us now suppose that $\boldsymbol{\alpha}$ tends to the point $\tilde{\boldsymbol{\alpha}}$, which is such that $\tilde{\alpha}_{rs} = 0$ for all (r,s) except for $(r,s) = (a,b)$ (i.e., $\tilde{\alpha}_{ab} = 1$). In this limit, Eq. (C4) is trivially satisfied when $(r,s) = (a,b)$ but not otherwise. Let us suppose $(r,s) \neq (a,b)$ and expand the equation around $p_{ab} = \bar{p}_{ab} = \frac{1}{2}$. From this second series expansion one concludes that the equality is satisfied if either $\bar{p}_{ab} = p_{ab}$ or $\bar{p}_{ab} = 1 - p_{ab}$. In both cases, the indices must match, which implies that $(a,b) = \pi^{-1}\circ\omega(a,b)$. By repeating the same argument for all (a,b), we conclude that $\omega = \pi$. Thus, the map $T\,:\,(\boldsymbol{\alpha},\boldsymbol{P}) \mapsto (\boldsymbol{\alpha}',\boldsymbol{P}')$ is a symmetry only if $\mathbf{\Pi} = \mathbf{\Omega}$.

This leaves the proof that $\mathbf{\Gamma} = \gamma\boldsymbol{I}$. Let us, by contradiction, assume that γ_{rs} differs from one set of indices to the other and define the sets A and B by

$$A = \{(r,s) : \gamma_{rs} = 0\} \quad \text{and} \quad B = \{(r,s) : \gamma_{rs} = 1\}\ .$$

Then one can write

$$\rho = \langle\alpha|p\rangle = \langle p\rangle_A + \langle p\rangle_B\ , \tag{C5}$$

where $\langle p\rangle_X := \sum_{(r,s)\in X}\alpha_{rs}p_{rs}$. Returning to Eq. (C4) for $(r,s)\in A$ and using the newfound fact that $\mathbf{\Pi} = \mathbf{\Omega}$ which implies $\bar{p}_{rs} = \gamma_{rs} + (1-2\gamma_{rs})p_{rs}$ (no more permutations), we find

$$\begin{aligned} p_{rs}\log\frac{p_{rs}}{\rho} + (1-p_{rs})\log\frac{1-p_{rs}}{1-\rho} \\ = p_{rs}\log\frac{p_{rs}}{\langle p'\rangle_A + \langle p'\rangle_B} + (1-p_{rs})\log\frac{1-p_{rs}}{1-\langle p'\rangle_A - \langle p'\rangle_B}\ . \end{aligned}$$

This can only be true if $\rho = \langle p'\rangle_A + \langle p'\rangle_B$, i.e., if $A = \emptyset$ or $B = \emptyset$. Therefore, $\gamma_{rs} = \gamma\ \forall(r,s)$, with $\gamma\in\{0,1\}$. □

2. Convexity of $\langle\mathcal{L}\rangle$

Theorem 2. *$\langle\mathcal{L}(\boldsymbol{\alpha},\boldsymbol{P})\rangle$ is convex with respect to $\boldsymbol{P}$.*

This property of $\langle\mathcal{L}\rangle$ is—perhaps surprisingly—not a consequence of the convexity of the KL divergence. Instead, it follows from the log-sum inequality.

Proof. We prove that $\langle\mathcal{L}(\boldsymbol{\alpha},\boldsymbol{P})\rangle$ is convex with respect to $\boldsymbol{P}$ by showing that it satisfies the convexity condition

$$\begin{aligned} \langle\mathcal{L}(\boldsymbol{\alpha},(1-t)\boldsymbol{P}+t\boldsymbol{Q})\rangle \\ \leq (1-t)\langle\mathcal{L}(\boldsymbol{\alpha},\boldsymbol{P})\rangle + t\langle\mathcal{L}(\boldsymbol{\alpha},\boldsymbol{Q})\rangle\ , \end{aligned} \tag{C6}$$

explicitly for all $t\in[0,1]$. Again, for the sake of clarity, we will use the notation developed in the previous section, and, in particular, write the density as $\rho = \langle\alpha|p\rangle$. We write each terms on the LHS of Eq. (C6) as

$$\alpha_{rs}\left\{[(1-t)p_{rs}+tq_{rs}]\log\frac{(1-t)p_{rs}+tq_{rs}}{(1-t)\,\langle\alpha|p\rangle + t\,\langle\alpha|q\rangle} + [(1-t)(1-p_{rs})+t(1-q_{rs})]\log\frac{(1-t)(1-p_{rs})+t(1-q_{rs})}{(1-t)(1-\langle\alpha|p\rangle)+t(1-\langle\alpha|q\rangle)}\right\}$$

It is easy to see that the log-sum inequality

$$(a+\bar{a})\log\frac{a+\bar{a}}{b+\bar{b}} \leq a\log\frac{a}{b} + \bar{a}\log\frac{\bar{a}}{\bar{b}}$$

can be applied to both parts of Eq. (C 2) to separate terms by their coefficients $(1-t)$ and t. Repeating the same operation on all terms yields inequality (C6). □

[1] M. A. Porter, J.-P. Onnela, and P. J. Mucha, Notices of the AMS **56**, 1082 (2009).

[2] S. Fortunato, Phys. Rep. **486**, 75 (2010).

[3] M. E. J. Newman, Nat. Phys. **8**, 25 (2012).

[4] C. Seshadhri, T. G. Kolda, and A. Pinar, Phys. Rev. E. **85**, 056109 (2012).

[5] T. P. Peixoto, Phys. Rev. X **4**, 011047 (2014).

[6] J.-G. Young, L. Hébert-Dufresne, A. Allard, and L. J. Dubé, Phys. Rev. E **94**, 022317 (2016).

[7] L. Hébert-Dufresne, A. Allard, V. Marceau, P.-A. Noël, and L. J. Dubé, Phys. Rev. Lett. **107**, 158702 (2011).

[8] A. Nematzadeh, E. Ferrara, A. Flammini, and Y.-Y. Ahn, Phys. Rev. Lett. **113**, 088701 (2014).

[9] M. Rosvall and C. T. Bergstrom, PNAS **105**, 1118 (2008).

[10] L. Hébert-Dufresne, A. Allard, P.-A. Noël, J.-G. Young, and E. Libby, arXiv:1607.04632 (2016).

[11] S. P. Borgatti and M. G. Everett, Soc. Networks **21**, 375 (2000).
[12] J. Yang and J. Leskovec, Proc. IEEE **102**, 1892 (2014).
[13] T. P. Peixoto, Phys. Rev. E **85**, 056122 (2012).
[14] P. W. Holland, K. B. Laskey, and S. Leinhardt, Soc. Networks **5**, 109 (1983).
[15] P. W. Holland and S. Leinhardt, JASA **76**, 33 (1981).
[16] H. C. White, S. A. Boorman, and R. L. Breiger, Am. J. Sociol. , 730 (1976).
[17] T. A. Snijders and K. Nowicki, Journal of Classification **14**, 75 (1997).
[18] T. P. Peixoto, Phys. Rev. Lett. **110**, 148701 (2013).
[19] M. E. J. Newman and G. Reinert, Phys. Rev. Lett. **117**, 078301 (2016).
[20] T. Kawamoto and Y. Kabashima, arXiv:1606.07668 (2016).
[21] T. P. Peixoto, Phys. Rev. X **5**, 011033 (2015).
[22] S. Fortunato and M. Barthelemy, PNAS **104**, 36 (2007).
[23] E. Abbe, To appear in Special Issue of the Journal of Machine Learning Research (2017).
[24] By *correlated*, it is meant that the two partitions are more similar than two randomly constructed partitions. Our choice of measure will be made explicit at a later stage.
[25] P. J. Bickel and A. Chen, PNAS **106**, 21068 (2009).
[26] E. Abbe, A. S. Bandeira, and G. Hall, IEEE Transactions on Information Theory **62**, 471 (2016).
[27] J. Reichardt and M. Leone, Phys. Rev. Lett. **101**, 078701 (2008).
[28] A. Decelle, F. Krzakala, C. Moore, and L. Zdeborová, Phys. Rev. Lett. **107**, 065701 (2011).
[29] F. Krzakala, C. Moore, E. Mossel, J. Neeman, A. Sly, L. Zdeborová, and P. Zhang, PNAS **110**, 20935 (2013).
[30] L. Massoulié, in *Proceedings of the 46th Annual ACM Symposium on Theory of Computing* (ACM, 2014) pp. 694–703.
[31] A. Decelle, F. Krzakala, C. Moore, and L. Zdeborová, Phys. Rev. E **84**, 066106 (2011).
[32] E. Mossel, J. Neeman, and A. Sly, arXiv:1311.4115 (2013).
[33] J. Banks, C. Moore, J. Neeman, and P. Netrapalli, in *29th Annual Conference on Learning Theory* (2016) pp. 383–416.
[34] T. Kawamoto and Y. Kabashima, Phys. Rev. E **95**, 012304 (2017).
[35] There is no obstacle to a generalization to the directed case (with or without self-loops).
[36] M. E. J. Newman, *Networks: An Introduction* (Oxford University Press, 2010).
[37] S. van der Pas and A. van der Vaart, arXiv:1608.04242 (2016).
[38] M. E. J. Newman, Phys. Rev. E **94**, 052315 (2016).
[39] M. E. J. Newman, Phys. Rev. E **88**, 042822 (2013).
[40] A. Condon and R. M. Karp, Rand. Struct. Alg. **18**, 116 (2001).
[41] E. Abbe and C. Sandon, in *2015 IEEE 56th Annual Symposium on Foundations of Computer Science* (IEEE, 2015) pp. 670–688.
[42] X. Zhang, R. R. Nadakuditi, and M. E. J. Newman, Phys. Rev. E **89**, 042816 (2014).
[43] R. R. Nadakuditi and M. E. J. Newman, Phys. Rev. Lett. **108**, 188701 (2012).
[44] G. Ver Steeg, C. Moore, A. Galstyan, and A. Allahverdyan, Europhys. Lett. **106**, 48004 (2014).
[45] E. Mossel, J. Neeman, and A. Sly, Probab. Theory Related Fields **162**, 431 (2015).
[46] P. Zhang, C. Moore, and M. E. J. Newman, Phys. Rev. E **93**, 012303 (2016).
[47] T. P. Peixoto, "The graph-tool python library," (2014).
[48] T. M. Cover and J. A. Thomas, *Elements of Information Theory* (John Wiley & Sons, 2012).
[49] $D(\mathbb{P}||\mathbb{Q})$ also goes to 0 at $\rho = 0$, and a more careful scaling analysis is necessary to conclude on the detectability of sparse instances.
[50] H. S. M. Coxeter, *Regular polytopes* (Courier Corporation, 1973).
[51] S. Boyd and L. Vandenberghe, *Convex optimization* (Cambridge university press, 2004).
[52] Since $\mathcal{L} > 0$ is not sufficient for detectability, some instances could still be undetectable.
[53] M. Jerrum and G. B. Sorkin, Discrete Appl. Math **82**, 155 (1998).
[54] Another explanation is that there are effectively $q^* = 2$ pairs of blocks in the eyes of our formalism: A single inner pair and a single outer pair, with, respectively, a fraction β and $1 - \beta$ of all possible edges.
[55] We give a reference implementation of the algorithm in C++ at `www.github.com/jg-you/sbm_canonical_mcmc`.
[56] P. Zhang, arXiv:1501.03844 (2015).
[57] L. Danon, A. Diaz-Guilera, J. Duch, and A. Arenas, J. Stat. Mech. Theor. Exp. **2005**, P09008 (2005).
[58] T. O. Kvålseth, IEEE Trans. Syst., Man, Cybern. **3**, 517 (1987).
[59] We do not have a procedure to determine the value of λ within the information-theoretical framework itself. However, Random Matrix Theory and recent developments in Information–theory offers some insights as to why one should have $\lambda \propto 1/n$, see Appendix B and Ref. [62] for details.
[60] J.-G. Young, *De la détection de la structure communautaire des réseaux complexes*, Master's thesis, Université Laval (2014).
[61] T. P. Peixoto, Phys. Rev. Lett. **111**, 098701 (2013).
[62] T. Lesieur, F. Krzakala, and L. Zdeborová, in *Communication, Control, and Computing (Allerton), 2015 53rd Annual Allerton Conference on* (IEEE, 2015) pp. 680–687.
[63] P. Diao, D. Guillot, A. Khare, and B. Rajaratnam, arXiv:1608.03860 (2016).
[64] M. Molloy and B. A. Reed, Rand. Struct. Alg. **6**, 161 (1995).
[65] M. E. J. Newman, S. H. Strogatz, and D. J. Watts, Phys. Rev. E **64**, 026118 (2001).
[66] B. Karrer and M. E. J. Newman, Phys. Rev. E **83**, 016107 (2011).
[67] A. Clauset, C. Moore, and M. E. Newman, Nature **453**, 98 (2008).