JMLR: Workshop and Conference Proceedings vol 49:1–34, 2016

Information-theoretic thresholds for community detection in sparse networks

Jess Banks BANKS.JESS.M@GMAIL.COM
Cristopher Moore MOORE@SANTAFE.EDU
Santa Fe Institute, Santa Fe, New Mexico

Joe Neeman JOENEEMAN@GMAIL.COM
Institute of Applied Mathematics, University of Bonn
Mathematics Department, University of Texas, Austin

Praneeth Netrapalli PRANEETH@MICROSOFT.COM
Microsoft Research, Cambridge MA

Abstract

We give upper and lower bounds on the information-theoretic threshold for community detection in the stochastic block model. Specifically, consider a symmetric stochastic block model with q groups, average degree d, and connection probabilities c_{in}/n and c_{out}/n for within-group and between-group edges respectively; let $\lambda = (c_{\text{in}} - c_{\text{out}})/(qd)$. We show that, when q is large, and $\lambda = O(1/q)$, the critical value of d at which community detection becomes possible—in physical terms, the condensation threshold—is

$$d_{\text{c}} = \Theta\left(\frac{\log q}{q\lambda^2}\right) ,$$

with tighter results in certain regimes. Above this threshold, we show that any partition of the nodes into q groups which is as 'good' as the planted one, in terms of the number of within- and between-group edges, is correlated with it. This gives an exponential-time algorithm that performs better than chance; specifically, community detection becomes possible below the Kesten-Stigum bound for $q \geq 5$ in the disassortative case $\lambda < 0$, and for $q \geq 11$ in the assortative case $\lambda > 0$ (similar upper bounds were obtained independently by Abbe and Sandon). Conversely, below this threshold, we show that no algorithm can label the vertices better than chance, or even distinguish the block model from an Erdős-Rényi random graph with high probability.

Our lower bound on d_{c} uses Robinson and Wormald's small subgraph conditioning method, and we also give (less explicit) results for non-symmetric stochastic block models. In the symmetric case, we obtain explicit results by using bounds on certain functions of doubly stochastic matrices due to Achlioptas and Naor; indeed, our lower bound on d_{c} is their second moment lower bound on the q-colorability threshold for random graphs with a certain effective degree.

1. Introduction

The Stochastic Block Model (SBM) is a random graph ensemble with planted community structure, where the probability of a connection between each pair of vertices is a function only of the groups or communities to which they belong. It was originally invented in sociology (Holland et al. (1983)); it was reinvented in physics and mathematics under the name "inhomogeneous random graph" (Söderberg (2002); Bollobás et al. (2007)), and in computer science as the planted partition problem (e.g. McSherry (2001)).

Given the current interest in network science, the block model and its variants have become popular parametric models for the detection of community structure. An interesting set of questions arise when we ask to what extent the communities, i.e., the labels describing the vertices' group memberships, can be recovered from the graph it generates. In the case where the average degree grows as $\log n$, if the structure is sufficiently strong then the underlying communities can be recovered (Bickel and Chen (2009)), and the threshold at which this becomes possible has recently been determined (Abbe et al. (2016); Abbe and Sandon (2015); Agarwal et al. (2015)). Above this threshold, efficient algorithms exist that recover the communities exactly, labeling every vertex correctly with high probability; below this threshold, exact recovery is information-theoretically impossible.

In the sparse case where the average degree is $O(1)$, finding the communities is more difficult, since we effectively have only a constant amount of information about each vertex. In this regime, our goal is to label the vertices better than chance, i.e., to find a partition with nonzero correlation or mutual information with the ground truth. This is sometimes called the *detection* problem to distinguish it from exact recovery. A set of phase transitions for this problem was conjectured in the statistical physics literature based on tools from spin glass theory (Decelle et al. (2011a,b)). Some of these conjectures have been made rigorous, while others remain as tantalizing open problems.

Besides the detection problem, it is natural to ask whether a graph generated by the stochastic block model can be distinguished from an Erdős-Rényi random graph with the same average degree. This is called the *distinguishability* problem, and it is believed to have the same threshold as the detection problem. Although distinguishing a graph from the stochastic block model from an Erdős-Rényi graph seems intuitively easier than actually detecting the communities, we do not know any rigorous proof of this statement.

1.1. The Kesten-Stigum bound, information-theoretic detection, and condensation

Although we will also deal with non-symmetric stochastic block models, in this discussion we focus on the symmetric case where the q groups are of equal expected size, and the probability of edges between vertices within and between groups are c_{in}/n and c_{out}/n respectively for constants $c_{\text{in}}, c_{\text{out}}$. The expected average degree of the resulting graph is then

$$d = \frac{c_{\text{in}} + (q-1)c_{\text{out}}}{q} \,. \tag{1}$$

It is convenient to parametrize the strength of the community structure as

$$\lambda = \frac{c_{\text{in}} - c_{\text{out}}}{qd} \,. \tag{2}$$

As we will see below, this is the second eigenvalue of a transition matrix describing how labels are "transmitted" between neighboring vertices. It lies in the range

$$-\frac{1}{q-1} \le \lambda \le 1 \,,$$

where $\lambda = -1/(q-1)$ corresponds to $c_{\text{in}} = 0$ (also known as the planted graph coloring problem) and $\lambda = 1$ corresponds to $c_{\text{out}} = 0$ where vertices only connect to others in the same group. We say that block models with $\lambda > 0$ and $\lambda < 0$ are *assortative* and *disassortative* respectively.

The conjecture of Decelle et al. (2011a,b) is that efficient algorithms exist if and only if we are above the threshold

$$d = \frac{1}{\lambda^2} \,. \tag{3}$$

This is known in information theory as the Kesten-Stigum threshold (Kesten and Stigum (1966a,b)), and in physics as the Almeida-Thouless line (de Almeida and Thouless (1978)).

Above the Kesten-Stigum threshold, Decelle et al. (2011a,b) claimed that community detection is computationally easy, and moreover that belief propagation—also known in statistical physics as the cavity method—is asymptotically optimal in that it maximizes the fraction of vertices labeled correctly (up to a permutation of the groups). For $q = 2$, this was proved in Mossel et al. (2014b); very recently Abbe and Sandon (2015) showed that a type of belief propagation performs better than chance for all q. In addition, Bordenave et al. (2015) showed that a spectral clustering algorithm based on the non-backtracking operator succeeds all the way down to the Kesten-Stigum threshold (proving a conjecture of Krzakala et al. (2013), who introduced the algorithm).

What happens below the Kesten-Stigum threshold is more complicated. Decelle et al. (2011a,b) conjectured that for sufficiently small q, community detection is information-theoretically impossible when $d < 1/\lambda^2$. Mossel et al. (2012) proved this in the case $q = 2$: first, they showed that the ensemble of graphs produced by the stochastic block model becomes *contiguous* with that produced by Erdős-Rényi graphs of the same average degree, making it impossible even to tell whether or not communities exist with high probability. Secondly, by relating community detection to the Kesten-Stigum reconstruction problem on trees (Evans et al. (2000)), they showed that for most pairs of vertices the probability, given the graph, that they are in the same group asymptotically approaches $1/2$. Thus it is impossible, even if we could magically compute the true posterior probability distribution, to label the vertices better than chance.

On the other hand, Decelle et al. (2011a,b) conjectured that for sufficiently large q, namely $q \geq 5$ in the assortative case $c_{\text{in}} > c_{\text{out}}$ and $q \geq 4$ in the disassortative case $c_{\text{in}} < c_{\text{out}}$, there is a "hard but detectable" regime where community detection is information-theoretically possible, but computationally hard. One indication of this is the extreme case where $c_{\text{in}} = 0$: this is equivalent to the planted graph coloring problem where we choose a uniformly random coloring of the vertices, and then choose $dn/2$ edges uniformly from all pairs of vertices with different colors. In this case, we have $\lambda = -1/(q-1)$ and (3) becomes $d > (q-1)^2$. However, while graphs generated by this case of the block model are q-colorable by definition, the q-colorability threshold for Erdős-Rényi graphs grows as $2q \ln q$ (Achlioptas and Naor (2005)), and falls below the Kesten-Stigum threshold for $q \geq 5$. In between these two thresholds, we can at least distinguish the two graph ensembles by asking whether a q-coloring exists; however, finding one might take exponential time.

More generally, planted ensembles where some combinatorial structure is built into the graph, and un-planted ensembles such as Erdős-Rényi graphs where these structures occur by chance, are believed to become distinguishable at a phase transition called *condensation* (Krzakala et al. (2007)). Below this point, the two ensembles are contiguous; above it, the posterior distribution of the partition or coloring conditioned on the graph—in physical terms, the Gibbs distribution—is dominated by a cluster of states surrounding the planted state. For instance, in random constraint satisfaction problems, the uniform distribution on solutions becomes dominated by those near the planted one; in our setting, the posterior distribution of partitions becomes dominated by those close to the ground truth (although, in the sparse case, with a Hamming distance that is still linear in n). Thus the condensation threshold is believed to be the threshold for information-theoretic community detection.

Below it, even optimal Bayesian inference will do no better than chance, while above it, typical partitions chosen from the posterior will be fairly accurate (though finding these typical partitions might take exponential time).

We note that some previous results show that community detection is possible below the Kesten-Stigum threshold when the sizes of the groups are unequal (Zhang et al. (2016)). In addition, even a vanishing amount of initial information can make community detection possible if the number of groups grows with the size of the network (Kanade et al. (2014)).

1.2. Our contribution

We give rigorous upper and lower bounds on the condensation threshold. Our bounds are most explicit in the case of symmetric stochastic block models, in which case we give upper and lower bounds for the condensation threshold as a function of q and λ. First, we use a first-moment argument to show that if

$$d > d_{\mathrm{c}}^{\mathrm{upper}} = \frac{2q \log q}{(1+(q-1)\lambda)\log(1+(q-1)\lambda) + (q-1)(1-\lambda)\log(1-\lambda)} , \tag{4}$$

then, with high probability, the only partitions that are as good as the planted one—that is, which have the expected number of edges within and between groups—have a nonzero correlation with the planted one. As a result, there is a simple exponential-time algorithm for labeling the vertices better than chance: simply test all partitions, and output the first good one.

We note that $d_{\mathrm{c}}^{\mathrm{upper}} < 1/\lambda^2$ for $q \geq 5$ when λ is sufficiently negative, including the case $\lambda = -1/(q-1)$ corresponding to graph coloring discussed above. Moreover, for $q \geq 11$, there also exist positive values of λ for $d_{\mathrm{c}}^{\mathrm{upper}} < 1/\lambda^2$. Thus for sufficiently large q, detectability is information-theoretically possible below the Kesten-Stigum threshold, in both the assortative and disassortative case. Similar (and somewhat tighter) results were obtained independently by Abbe and Sandon (2016).

We then show that community detection is information-theoretically impossible if

$$d < d_{\mathrm{c}}^{\mathrm{lower}} = \frac{2\log(q-1)}{q-1}\frac{1}{\lambda^2} . \tag{5}$$

Using the small subgraph conditioning method, we show that the block model and the Erdős-Rényi graph are contiguous whenever the second moment of the ratio between their probabilities—roughly speaking, the number of good partitions in an Erdős-Rényi graph—is appropriately bounded. We also show that this second moment bound implies non-detectability, in that the posterior distribution on any finite collection of vertices is asymptotically uniform. This reduces the proof of contiguity and non-detectability to a second moment argument; in the case of a symmetric stochastic block model, this consists of maximizing a certain function of doubly stochastic matrices.

Happily, this latter problem was largely solved by Achlioptas and Naor (2005), who used the second moment method to give nearly tight lower bounds on the q-colorability threshold. Our bound (5) corresponds to their lower bound on q-colorability for $G(n, d'/n)$ where $d' = d\lambda^2(q-1)^2$. Intuitively, d' is the degree of a random graph in which the correlations between vertices in the q-colorability problem are as strong as those in the stochastic block model with average degree d and eigenvalue λ.

Our bounds are tight in some regimes, and rather loose in others. Let μ denote $(c_{\text{in}} - c_{\text{out}})/d$. If μ is constant and q is large, we have

$$\lim_{q\to\infty} \frac{d_{\text{c}}^{\text{upper}}}{d_{\text{c}}^{\text{lower}}} = \frac{\mu^2}{(1+\mu)\log(1+\mu) - \mu} .$$

In the limit $\mu = -1$, corresponding to graph coloring, this ratio is 1, inheriting the tightness of previous upper and lower bounds on q-colorability. For other values of μ, our bounds match up to a multiplicative constant. In particular, when q is constant and $|\lambda|$ is small, they are about a factor of 2 apart:

$$\frac{2\log(q-1)}{q-1} \le d_{\text{c}}\lambda^2 \le \frac{4\log q}{q-1}(1 + O(q\lambda)) .$$

When $\lambda \ge 0$ is constant and q is large, we have

$$d_{\text{c}}^{\text{upper}} = \frac{2}{\lambda}(1 + O(1/\log q)) .$$

Thus, in the limit of large q, detectability is possible below the Kesten-Stigum threshold whenever $\lambda < 1/2$.

2. Definitions and results

A stochastic block model with $q \ge 2$ communities is parametrized by two quantities: the distribution $\pi \in \Delta_q$ of vertex classes and the symmetric matrix $M \in \mathbb{R}^{q\times q}$ of edge probabilities. Given these two parameters, a random graph from the block model $G(n, M/n, \pi)$ is generated as follows: for each vertex v, sample a label σ_v in $[q] = \{1, \ldots, q\}$ independently with distribution π. Then, for each pair (u, v), include the edge (u, v) in the graph independently with probability $n^{-1}M_{\sigma_u,\sigma_v}$. Since we will worq with a fixed M and π throughout, we denote $G(n, M/n, \pi)$ by $\mathbb{P}_n$. Note that according to the preceding description, we have the following explicit form for the density of $\mathbb{P}_n$:

$$\mathbb{P}_n(G, \sigma) = \prod_{v\in V(G)} \pi_{\sigma_v} \prod_{(u,v)\in E(G)} \frac{M_{\sigma_u,\sigma_v}}{n} \prod_{(u,v)\notin E(G)} \left(1 - \frac{M_{\sigma_u,\sigma_v}}{n}\right) .$$

We will assume throughout that every vertex in $G \sim \mathbb{P}_n$ has the same expected degree. (In terms of M and π, this means that $\sum_j M_{ij}\pi_j$ does not depend on i.) Without this assumption, reconstruction and distinguishability – at least in the way that we will define them – are trivial, since we gain non-trivial information on the class of a vertex just by considering its degree.

With the preceding assumption in mind, let $d = \sum_j M_{ij}\pi_j$ be the expected degree of an arbitrary vertex. In order to discuss distinguishability, we will compare $\mathbb{P}_n$ with the Erdős-Rényi distribution $\mathbb{Q}_n := G(n, d/n)$.

Throughout this work, we will make use of the matrix T defined by

$$T_{ij} = \frac{1}{d}\pi_i M_{ij},$$

or in other words, $T = \frac{1}{d}\operatorname{diag}(\pi)M$. Note that T is a stochastic matrix, in the sense that it has non-negative elements and all its rows sum to 1. The Perron-Frobenius eigenvectors of T are π on the right, and $\mathbf{1}$ on the left (where $\mathbf{1}$ denotes the vector of ones), and the corresponding eigenvalue is 1. We let $\lambda_1, \ldots, \lambda_q$ be the eigenvalues of T, arranged in order of decreasing absolute value (so that $\lambda_1 = 1$ and $|\lambda_2| \leq 1$). The second of these turns out to be the most important for us; therefore, set $\lambda = \lambda_2$.

There is an important probabilistic interpretation of the matrix T relating to the local structure of $G \sim \mathbb{P}_n$; although we will not rely on this interpretation in the current work, it played an important role in Mossel et al. (2014a). Indeed, one can show that for any fixed radius R, the R-neighborhood of a vertex in $G \sim \mathbb{P}_n$ has almost the same distribution as a Galton-Watson tree with radius R and offspring distribution $\mathrm{Poisson}(d)$. Then, the class labels on the neighborhood can be generated by first choosing the label of the root according to π and then, conditioned on the root's label being i, choosing its children's labels independently to be j with probability T_{ij}. This procedure continues down the tree: any vertex with parent u has probability $T_{\sigma_u j}$ to receive the label j. Thus, T is the transition matrix of a certain Markov process that describes a procedure for approximately generating the class labels on a local neighborhood in G.

In part of this work, we will deal with the symmetric case, in which $\pi_i = \frac{1}{q}$ for all i and

$$M_{i,j} = \begin{cases} c_{\text{in}} & \text{if } i = j \\ c_{\text{out}} & \text{if } i \neq j\,. \end{cases} \tag{6}$$

In this case, the expected average degree is

$$d = \frac{c_{\text{in}} + (q-1)c_{\text{out}}}{q}\,,$$

the Markov transition matrix (which is symmetric, and hence doubly stochastic) is

$$T = \frac{1}{qd}\begin{pmatrix} c_{\text{in}} & & c_{\text{out}} \\ & \ddots & \\ c_{\text{out}} & & c_{\text{in}} \end{pmatrix} = \lambda \mathbb{I} + (1-\lambda)\frac{\mathbb{J}}{q}, \tag{7}$$

where $\mathbb{I}$ is the identity matrix, $\mathbb{J}$ is the matrix of all 1s, and where

$$\lambda = \frac{c_{\text{in}} - c_{\text{out}}}{qd}$$

is T's second eigenvalue. We can think of λ as the probability that information is transmitted from u to v: with probability λ we copy u's group label to v, and with probability $1-\lambda$ we choose v's group uniformly from $[q]$. The parameter λ interpolates between the case $\lambda = 1$ where all edges are within-group, to an Erdős-Rényi graph where $\lambda = 0$ and edges are placed uniformly at random, to $\lambda < 0$ where edges are more likely between groups than within them. This gives a useful reparametrization of the model in terms of c and λ, where

$$\begin{aligned} c_{\text{in}} &= d(1 + (q-1)\lambda) \\ c_{\text{out}} &= d(1-\lambda)\,. \end{aligned} \tag{8}$$

For labellings σ and τ in $[q]^n$, define their *overlap* by

$$\mathrm{overlap}(\sigma, \tau) = \frac{1}{n}\max_{\rho}\sum_{i=1}^{q}\left(|\sigma^{-1}(i) \cap \tau^{-1}(\rho(i))| - \frac{1}{n}|\sigma^{-1}(i)||\tau^{-1}(\rho(i))|\right),$$

where the supremum runs over all permutations ρ of $[q]$. In words, σ and τ have a positive overlap if there is some relabelling of $[q]$ so that they are positively correlated.

Definition 1 *We say that the block model $\mathbb{P}_n = G(n, M/n, \pi)$ is* detectable *if there is some $\delta > 0$ and an algorithm $\mathcal{A}$ mapping graphs to labellings such that if $(G, \sigma) \sim \mathbb{P}_n$ then*

$$\lim_{n\to\infty} \Pr(\mathrm{overlap}(\mathcal{A}(G), \sigma) > \delta) > 0.$$

Definition 2 *We say that $\mathbb{P}_n$ and $\mathbb{Q}_n$ are* asymptotically orthogonal *if there is a sequence A_n of events such that $\mathbb{P}_n(A_n) \to 0$ and $\mathbb{Q}_n(A_n) \to 1$.*

We say that $\mathbb{P}_n$ and $\mathbb{Q}_n$ are contiguous *if for every sequence A_n of events, $\mathbb{P}_n(A_n) \to 0$ if and only if $\mathbb{Q}_n(A_n) \to 0$.*

Our main result is the following:

Theorem 3 *Consider the symmetric stochastic block model $\mathbb{P}_n$ with q communities, average degree d, and second-eigenvalue λ. Define*

$$d_{\mathrm{c}}^{\mathrm{upper}} = \frac{2q \log q}{(1 + (q-1)\lambda) \log(1 + (q-1)\lambda) + (q-1)(1-\lambda)\log(1-\lambda)} \tag{9}$$

$$d_{\mathrm{c}}^{\mathrm{lower}} = \frac{2\log(q-1)}{q-1} \frac{1}{\lambda^2} \, . \tag{10}$$

If $d > d_{\mathrm{c}}^{\mathrm{upper}}$ then $\mathbb{P}_n$ and $\mathbb{Q}_n$ are asymptotically orthogonal, and $\mathbb{P}_n$ is detectable. If $d < d_{\mathrm{c}}^{\mathrm{lower}}$ then $\mathbb{P}_n$ and $\mathbb{Q}_n$ are contiguous, and $\mathbb{P}_n$ is not detectable.

The lower bound in Theorem 3 comes from a more general (but less explicit) bound that holds also for block models that are not symmetric. In order to state the more general result, we must first introduce some notation.

Definition 4 *Let Δ_m denote the probability simplex in $\mathbb{R}^m$:*

$$\Delta_m := \{p \in \mathbb{R}^m : p_i \geq 0, \sum_{i=1}^m p_i = 1\}.$$

Define $D : \Delta_m \times \Delta_m \to \mathbb{R}$ by

$$D(p, \tilde{p}) = \sum_{i=1}^m p_i \log(p_i / \tilde{p}_i).$$

Note that if we interpret $p, \tilde{p} \in \Delta_m$ as probability distributions on a m-point set, then $D(p, \tilde{p})$ is exactly the Kullback-Leibler divergence of p with respect to $\tilde{p}$.

Definition 5 *For $\pi \in \Delta_q$, define*

$$\Delta_{q^2}(\pi) := \{(p_{ij})_{i,j=1}^q \in \Delta_{q^2} : \sum_{i=1}^q p_{ij} = \pi_j \text{ and } \sum_{j=1}^q p_{ij} = \pi_i \text{ for all } i, j\}.$$

In other words, elements of $\Delta_{q^2}(\pi)$ are probability distributions on $[q]^2$ that have π as their marginal distributions.

Definition 6 *For $\pi \in \Delta_q$ and a $q \times q$ matrix A, let $p = \pi \otimes \pi$, where $\otimes$ denotes Kronecker product and define*

$$Q(\pi, A) = \sup_{\alpha \in \Delta_{q^2}(\pi)} \frac{(\alpha - p)^\mathsf{T}(A \otimes A)(\alpha - p)}{D(\alpha, p)}.$$

Although we do not know any simple algebraic expression for Q, one can easily compute numerical approximations. For non-symmetric stochastic block models, our main result is that Q gives a lower bound on the detectability threshold:

Theorem 7 *Let $\mathbb{P}_n = G(n, M/n, \pi)$ and $\mathbb{Q}_n = G(n, d/n)$, where $d = \sum_j M_{ij}\pi_j$. If*

$$Q(\pi, (M - d\mathbb{J})/\sqrt{2d}) < 1$$

then $\mathbb{P}_n$ and $\mathbb{Q}_n$ are contiguous and $\mathbb{P}_n$ is non-detectable.

For comparison with the Kesten-Stigum bound, note that $Q(\pi, (M - d\mathbb{J})/\sqrt{2d}) < 1$ implies that $\lambda^2 d < 1$. This comes from comparing the second derivatives at p in the numerator and denominator of Q: if $Q < 1$ then the Hessian of the numerator must be smaller (in the semidefinite order) than that of the denominator, and this turns out to be equivalent to $\lambda^2 d < 1$.

We remark that while $Q(\pi, (M - d\mathbb{J})/\sqrt{2d}) < 1$ is only a sufficient condition for the contiguity of $\mathbb{P}_n$ and $\mathbb{Q}_n$, it is actually a sharp condition for a certain second moment to exist:

Proposition 8 *Fix a sequence a_n with $a_n = o(n)$ and $a_n = \omega(\sqrt{n})$. Let Ω_n be the event that for all $i \in [q]$, $|\sigma^{-1}(i)| = n\pi_i \pm a_n$. With the notation of Theorem 7, take $\hat{\mathbb{P}}_n$ to be $\mathbb{P}_n$ conditioned on Ω_n. If $Q(\pi, (M - d\mathbb{J})/\sqrt{2d}) < 1$ then*

$$\lim_{n\to\infty} \mathbb{E}_{\mathbb{Q}_n} \left(\frac{\hat{\mathbb{P}}_n}{\mathbb{Q}_n}\right)^2 = (1 + o(1)) \prod_{i,j=2}^{q} \psi(d\lambda_i\lambda_j) < \infty, \tag{11}$$

where $\lambda_1, \cdots, \lambda_q$ are the eigenvalues of T (cf. (2)) such that $1 = \lambda_1 \geq |\lambda_2| \geq \cdots \geq |\lambda_q|$, and $\psi(x) = (1-x)^{-1/2}e^{-x/2-x^2/4}$. On the other hand, if $Q(\pi, (M - d\mathbb{J})/\sqrt{2d}) > 1$ then

$$\lim_{n\to\infty} \mathbb{E}_{\mathbb{Q}_n} \left(\frac{\hat{\mathbb{P}}_n}{\mathbb{Q}_n}\right)^2 = \infty.$$

2.1. Outline of the paper

We prove the upper bound of Theorem 3 in Section 3. In Section 4, we prove the lower bound of Theorem 3 assuming Theorem 7. In Section 5, we prove Proposition 8. Finally, in Section 6, we prove Theorem 7. Some auxiliary results are proved in Appendix A.

2.2. Outline of the proofs

The part of Theorem 3 regarding $d_\mathrm{c}^\mathrm{upper}$ follows from union bounds. First, note that under $\mathbb{P}_n$, groups in the planted partition have average in-degree of about c_in/k and average out-degree of about $(k-1)c_\mathrm{out}/k$. We call such partitions "good." In order to show orthogonality, we show that with high probability, graphs from $\mathbb{Q}_n$ have no good partitions. (That is, the events A_n witnessing orthogonality are $A_n = \{G \text{ has no good partitions}\}$.) We show this by computing the probability that a particular partition is good and comparing it to the number of all partitions. In order to show detectability, we show that with high probability under $\mathbb{P}_n$, every good partition is correlated with the planted partition: we bound the probability that a given partition is good, and sum the probabilities over all partitions that are uncorrelatd with the planted one.

The part of Theorem 3 regarding $d_{\mathrm{c}}^{\mathrm{lower}}$ follows from Theorem 7. We recognize that the optimization problem in the definition of Q may be written as an optimization over the set of doubly-stochastic matrices. Using tools due to Achlioptas and Naor (2005) (Theorem 11 and Lemma 12), we prove that $d < d_{\mathrm{c}}^{\mathrm{lower}}$ implies that $Q < 1$, and we conclude by applying Theorem 7.

Proposition 8 is the main technical step in the proof of contiguity in Theorem 7. With Proposition 8 in hand, we apply the small subgraph conditioning method (see Theorem 20) which is a type of conditional second moment method. In order to apply it, we only need to know the limiting distribution of small subgraphs under $\hat{\mathbb{P}}_n$ and $\mathbb{Q}_n$ (which are already known) and (11) from Proposition 8.

The proof of Proposition 8 itself is tedious but elementary: we expand the square and write the result as the exponential of a quadratic form in multinomial random variables. Shifted and renormalized, the multinomial variables have a Gaussian limit; the expectation of an exponentiated quadratic form of Gaussian variables can be computed exactly, and gives (11). In order to apply the central limit theorem in the above argument, one needs to check that the exponentiated quadratic form in multinomial variables is uniformly integrable. This naturally leads to the condition on Q: we need to compare an exponentiated quadratic form with the multinomial probability mass function, which is essentially an exponentiated entropy. In the end, we need the entropy to dominate the quadratic form (which is exactly what happens with $Q < 1$).

Finally, to prove non-detectability in Theorem 7 we compare the distribution $\mathbb{P}_n$ to the distribution (call it $\tilde{\mathbb{P}}_n$) obtained by conditioning on the labels of a constant number of vertices. If we can show that the resulting distributions are close in total variation, it implies that the labels of those vertices cannot be statistically inferred. Applying the Cauchy-Schwarz inequality to the total variation distance, it is enough to show that

$$\mathbb{E}_{\mathbb{Q}_n} \mathbb{1}_{\Omega_n} \left(\frac{\mathbb{P}_n}{\mathbb{Q}_n} - \frac{\tilde{\mathbb{P}}_n}{\mathbb{Q}_n} \right)^2$$

is small. This naturally leads to a computation very similar to the proof of Proposition 8. The only difference is that we are now conditioning on the labels of a constant number of vertices, but that has very little effect.

2.3. Conclusions and future work

We (and, independently, Abbe and Sandon (2016)) have shown that community detection is information-theoretically possible below the Kesten-Stigum threshold. However, we have not given any evidence that it is computationally hard. Of course, we cannot hope to prove this without knowing that P $\neq$ NP, but we could hope to prove that certain classes of algorithms take exponential time. In particular, we could show that Monte Carlo algorithms or belief propagation take exponential time to find a good partition, assuming their initial states or messages are uniformly random.

Physically, we believe this occurs because there is a free energy barrier between a “paramagnetic” phase of partitions which are essentially random, and a “ferromagnetic” or “retrieval” phase which is correlated with the planted partition (Decelle et al. (2011a,b); Zhang and Moore (2014)). Proving this seems within reach: rigorous results have been obtained in random constraint satisfaction problems (Achlioptas and Coja-Oghlan (2008); Coja-Oghlan and Efthymiou (2015)) showing that solutions become clustered with $O(n)$ Hamming distance and $O(n)$ energy barriers between them. In particular, Markov chain Monte Carlo algorithms for sampling the posterior distribution, such as Metropolis-Hastings or Glauber dynamics that update the label of one vertex at time according to its marginal distribution conditioned on the current labels of its neighbors, take exponential time to travel from one cluster to another. The goal in this case would be to show in a planted model that Monte Carlo takes exponential time to find the cluster corresponding to the planted solution.

Finally, both our upper and lower bounds can be improved. Our upper bound requires that w.h.p. all good partitions are correlated with the planted one. We could obtain better bounds by requiring that this is true w.h.p. of *most* good partitions, which would require a lower bound on the typical number of good partitions with large overlap. In the limit $\lambda \to 1$ of strong assortative structure, for instance, one can use the fact that vertices of degree 1 can be set to match their neighbors, or set freely to give the same typical overlap as the planted partition. Using these and other ideas, Abbe and Sandon (2016) showed that $d_{\mathrm{c}} \to 1$ as $\lambda \to 1$, while our bounds only give $d_{\mathrm{c}} \leq 2$. (For regimes where d_{c} is large, their bounds and ours are asymptotically equivalent.) Further improvements seem possible.

The second moment lower bound could be improved as it was for the k-colorability threshold in Coja-Oghlan and Vilenchik (2013). Indeed, the condensation threshold d_{c} for k-coloring was determined exactly in Bapst et al. (2014) for sufficiently large k. It is entirely possible that their techniques could work here. Note that constraint satisfaction problems correspond to zero-temperature models in physics, while the block model with $c_{\text{in}}, c_{\text{out}} \neq 0$ corresponds to a spin system at positive temperature; but some rigorous results have recently been obtained here as well by Bapst et al. (to appear).

3. Upper bound for symmetric SBMs: Proof of upper bound in Theorem 3

In this section, we prove the part of Theorem 3 relating to $d_{\mathrm{c}}^{\text{upper}}$. Recall that Theorem 3 assumes a symmetric block model; i.e., $\pi_i = 1/q$ for every i, and the connectivity matrix M is determined by only two parameters, c_{in} and c_{out}.

Our upper bound on the detectability threshold hinges on the following observation. We say a partition is *balanced* if it has n/q vertices in each group. With high probability, a graph generated by the SBM has at least one balanced partition, close to the the planted one, where the number of within-group and between-group edges m_{in} and m_{out} are close to their expectations. That is,

$$|m_{\text{in}} - \overline{m}_{\text{in}}| < n^{2/3} \quad \text{and} \quad |m_{\text{out}} - \overline{m}_{\text{out}}| < n^{2/3} \tag{12}$$

where

$$\begin{aligned} \overline{m}_{\text{in}} &= \frac{c_{\text{in}}}{2q} n = \frac{d(1+(q-1)\lambda)}{2q} n \\ \overline{m}_{\text{out}} &= \frac{(q-1)c_{\text{out}}}{2q} n = \frac{d(q-1)(1-\lambda)}{2q} n \,. \end{aligned} \tag{13}$$

This follows from standard concentration inequalities on the binomial distribution: the number of vertices in each group in σ is w.h.p. $n/q + o(n^{2/3}/\log n)$, in which case (12) holds w.h.p. Since the maximum degree is w.h.p. less than $\log n$, we can modify σ to make it balanced while changing m_{in} and m_{out} by $o(n^{2/3})$.

Call such a partition *good*. We will show that if $d > d_{\text{c}}^{\text{upper}}$ all good partitions are correlated with the planted one. As a result, there is an exponential algorithm that performs better than chance: simply use exhaustive search to find a good partition, and output it.

3.1. Distinguishability from $G(n, d/n)$

As a warm-up, we show that if $d > d_{\text{c}}^{\text{upper}}$ the probability that an Erdős-Rényi graph has a good partition is exponentially small, so the two distributions $\mathbb{P}$ and $\mathbb{Q}$ are asymptotically orthogonal.

Let G be a graph generated by $G(n, d/n)$. We condition on the high-probability event that it has m edges with $|m - \overline{m}| < n^{2/3}$ with

$$\overline{m} = \overline{m}_{\text{in}} + \overline{m}_{\text{out}} = dn/2\,,$$

in which case G is chosen from $G(n, m)$. Since G is sparse, we can think of its m edges as chosen uniformly with replacement from the n^2 possible ordered pairs. With probability $\Theta(1)$ the resulting graph is simple, with no self-loops or multiple edges, and hence uniform in $G(n, m)$. Thus any event that holds with high probability in the resulting model holds with high probability in $G(n, m)$ as well. Call this model $G'(n, m)$.

For a given balanced partition σ, the probability in $G'(n, m)$ that a given edge has its endpoints in the same group is $1/q$. Thus, up to subexponential terms resulting from summing over the $n^{2/3}$ possible values of the error terms, the probability that a given σ is good is

$$\Pr[\text{Bin}(\overline{m}, 1/q) = \overline{m}_{\text{in}}] = \binom{\overline{m}}{\overline{m}_{\text{in}}} (1/q)^{\overline{m}_{\text{in}}} (1 - 1/q)^{\overline{m}_{\text{out}}}\,.$$

The rate of this large-deviation event is given by the Kullback-Leibler divergence between binomial distributions with success probability $1/q$ and $\overline{m}_{\text{in}}/\overline{m}$,

$$\begin{aligned}\lim_{\overline{m}\to\infty} \frac{1}{\overline{m}} \log \Pr[\text{Bin}(\overline{m}, 1/q) = \overline{m}_{\text{in}}] &= -\frac{\overline{m}_{\text{in}}}{\overline{m}} \log \frac{\overline{m}_{\text{in}}/\overline{m}}{1/q} - \frac{\overline{m}_{\text{out}}}{\overline{m}} \log \frac{\overline{m}_{\text{out}}/\overline{m}}{1 - 1/q} \\ &= -\frac{c_{\text{in}}}{qd} \log \frac{c_{\text{in}}}{d} - \left(1 - \frac{c_{\text{in}}}{qd}\right) \log \frac{qd - c_{\text{in}}}{d(q-1)}\,,\end{aligned}$$

where we used $\overline{m}_{\text{in}}/\overline{m} = c_{\text{in}}/(qd)$ and $\overline{m}_{\text{out}}/\overline{m} = 1 - c_{\text{in}}/(qd)$. Writing this in terms of d and λ as in (8) and simplifying gives

$$\lim_{n\to\infty} \frac{1}{n} \log \Pr[\sigma \text{ is good}] = -\frac{d}{2q} \big[(1+(q-1)\lambda) \log(1+(q-1)\lambda) + (q-1)(1-\lambda) \log(1-\lambda)\big]\,. \quad (14)$$

Now, by the union bound, since there are at most q^n balanced partitions, the probability that any good partitions exist is exponentially small whenever the function in (14) is less than $-\log q$. This tells us that the block model is distinguishable from an Erdős-Rényi graph whenever

$$d > d_{\text{c}}^{\text{upper}} = \frac{2q \log q}{(1 + (q-1)\lambda) \log(1 + (q-1)\lambda) + (q-1)(1-\lambda) \log(1-\lambda)}\,,$$

As noted above, the limit $\lambda = -1/(q-1)$ corresponds to the planted graph coloring problem. In this case $d_{\text{c}}^{\text{upper}}$ is simply the first-moment upper bound on the q-colorability threshold,

$$d_{\text{c}}^{\text{upper}} = \frac{2 \log q}{-\log(1 - 1/q)} < 2q \log q\,.$$

3.2. All good partitions are accurate

Next we show that, if $d > d_c^{\text{upper}}$, with high probability any good partition is correlated with the planted one. Essentially, the previous calculation for $G(n,m)$ corresponds to counting good partitions τ which are uncorrelated with σ, i.e., which have $\mathrm{overlap}(\sigma,\tau) = 0$. We will show that in order for a good partition to exist, its overlap with σ is strictly greater than 0.

Given a balanced partition τ, let m_{in} and m_{out} denote the number of edges (u,v) with $\tau_u = \tau_v$ and $\tau_u \neq \tau_v$ respectively. As in the previous section, we say that τ is *good* if (12) holds, i.e., $|m_{\text{in}} - \overline{m}_{\text{in}}|, |m_{\text{out}} - \overline{m}_{\text{out}}| < n^{2/3}$ where $\overline{m}_{\text{in}}$ and $\overline{m}_{\text{out}}$ are given by (13). Note that the right-hand side of (15) is an increasing function of β, and that it coincides with d_c^{upper} when $\beta = 0$.

Theorem 9 *Let G be generated by the stochastic block model with parameters c_{in} and c_{out}, and let d and λ be defined as in (1) and (2). If $d > d_c^{\text{upper}}$ then, with high probability, any good partition has overlap at least $\beta > 0$ with the planted partition σ, where β is the smallest root of*

$$d = \frac{2q\big(h(\beta + \frac{1}{q}) + (1 - \frac{1}{q} - \beta)\log(q-1)\big)}{(1+(q-1)\lambda)\log\frac{1+(q-1)\lambda}{1+q\beta\lambda} + (q-1)(1-\lambda)\log\frac{(q-1)(1-\lambda)}{q-1-q\beta\lambda}} \tag{15}$$

where $h = -\big(\beta + \frac{1}{q}\big)\log\big(\beta+\frac{1}{q}\big) - \big(1 - \frac{1}{q} - \beta\big)\log\big(1-\frac{1}{q}-\beta\big)$ is the entropy function. Therefore, an exponential-time algorithm exists that w.h.p. achieves overlap at least β.

Proof We start by conditioning on the high-probability event that G has m edges, where $|m - \overline{m}| < n^{2/3}$ and $\overline{m} = dn/2$. Call the resulting model $G_{\text{SBM}}(n,m)$ (with the matrix of parameters M implicit). It consists of the distribution over all simple graphs with m edges, with probability proportional to $\mathbb{P}(G \mid \sigma)$.

In analogy with the model $G'(n,m)$ defined above, we consider another version of the block model where the m edges are chosen independently as follows. For each edge, we first choose an ordered pair of groups r, s with probability proportional to M_{rs}, i.e., with probability T_{rs}/q where $T = M/(qd)$ is the doubly stochastic matrix defined in (7). We then choose the endpoints u and v uniformly from $\sigma^{-1}(r)$ and $\sigma^{-1}(s)$ (with replacement if $r = s$). Call this model $G'_{\text{SBM}}(n,m)$. In the sparse case $d = O(1/n)$, the resulting graph is simple with probability $\Theta(1)$, in which event it is generated by $G_{\text{SBM}}(n,m)$. Thus any event that holds with high probability in $G'_{\text{SBM}}(n,m)$ holds with high probability in $G_{\text{SBM}}(n,m)$ as well.

Now fix a balanced partition τ. Let θ denote the probability that an edge (u,v) chosen in this way is within-group with respect to τ. Define the $q \times q$ matrix α by

$$\alpha_{st} = \frac{q}{n}|\sigma^{-1}(s) \cap \tau^{-1}(t)|;$$

in other words, α_{st} is the probability that $\tau_u = t$ if u is chosen uniformly from those with $\sigma_u = s$. Up to $O(1/n)$ terms, the events that $\tau_u = t$ and $\tau_v = t$ are independent. Thus in the limit $n \to \infty$,

$$\begin{aligned}\theta := \Pr[\tau_u = \tau_v] &= \sum_{r,s,t}\Pr[\sigma_u = r \wedge \sigma_v = s \wedge \tau_u = \tau_v = t] \\ &= \frac{1}{q}\sum_{r,s,t} T_{rs}\alpha_{rt}\alpha_{st} \\ &= \frac{1}{q}\operatorname{tr}\alpha^{\mathsf{T}} T\alpha\,,\end{aligned}$$

where $^\intercal$ denotes the matrix transpose. Since $T = \lambda\mathbb{I} + (1-\lambda)\frac{\mathbb{J}}{q}$ and $\mathbb{J}\alpha = \alpha\mathbb{J} = \mathbb{J}$, this gives

$$\theta = \frac{1 + (|\alpha|^2 - 1)\lambda}{q} \, ,$$

where $|\alpha|$ denotes the Frobenius norm,

$$|\alpha|^2 = \operatorname{tr} \alpha^\intercal \alpha = \sum_{r,s} \alpha_{rs}^2 \, .$$

When τ and σ are uncorrelated and $\alpha = \mathbb{J}/q$, we have $\theta = 1/q$ as in the previous section. When $\sigma = \tau$ and $\alpha = \mathbb{I}$, we have $\theta = c_{\text{in}}/(qd) = (1 + (q-1)\lambda)/q$.

For τ to be good, we need $|m_{\text{in}} - \overline{m}_{\text{in}}| < n^{2/3}$. Since $|m - \overline{m}| < n^{2/3}$ as well, up to subexponential terms the probability that τ is good is

$$\Pr[\text{Bin}(\overline{m}, \theta) = \overline{m}_{\text{in}}] = \binom{\overline{m}}{\overline{m}_{\text{in}}} \theta^{\overline{m}_{\text{in}}} (1-\theta)^{\overline{m}_{\text{out}}} \, .$$

The rate at which this occurs is again a Kullback-Leibler divergence, between binomial distributions with success probabilities θ and $\overline{m}_{\text{in}}/\overline{m} = c_{\text{in}}/(qd)$. Following our previous calculations gives

$$\begin{aligned}
&\lim_{n\to\infty} \frac{1}{n} \log \Pr[\text{Bin}(\overline{m}, \theta) = \overline{m}_{\text{in}}] && (16)\\
&= -\frac{d}{2}\left(\frac{c_{\text{in}}}{qd} \log \frac{c_{\text{in}}}{\theta qd} + \left(1 - \frac{c_{\text{in}}}{qd}\right) \log \frac{1 - c_{\text{in}}/qd}{1-\theta} \right)\\
&= -\frac{d}{2q}\left[(1 + (q-1)\lambda) \log \frac{1 + (q-1)\lambda}{\theta q} + (q-1)(1-\lambda) \log \frac{(q-1)(1-\lambda)}{q(1-\theta)} \right]\\
&= -\frac{d}{2q}\left[(1 + (q-1)\lambda) \log \frac{1 + (q-1)\lambda}{1 + (|\alpha|^2 - 1)\lambda} + (q-1)(1-\lambda) \log \frac{(q-1)(1-\lambda)}{q - 1 - (|\alpha|^2 - 1)\lambda} \right] \, .
\end{aligned}$$

We pause to prove a lemma which relates the Frobenius norm to the overlap. This bound is far from tight except in the extreme cases $\alpha = \mathbb{J}/q$ and $\alpha = \mathbb{I}$, but it lets us derive an explicit lower bound on the overlap of a good partition.

Lemma 10 $|\alpha|^2 \le 1 + q \operatorname{overlap}(\sigma, \tau)$.

Proof Since α is doubly stochastic, Birkhoff's theorem tells us it can be expressed as a convex combination of permutation matrices,

$$\alpha = \sum_\pi a_\pi \pi \quad \text{where} \quad \sum_\pi a_\pi = 1 \, .$$

Thus

$$|\alpha|^2 = \operatorname{tr} \alpha^\intercal \alpha = \operatorname{tr} \left(\sum_\pi a_\pi \pi^{-1} \right) \alpha = \sum_\pi a_\pi \operatorname{tr} \pi^{-1} \alpha \le \max_\pi \operatorname{tr} \pi^{-1} \alpha = 1 + q \operatorname{overlap}(\sigma, \tau),$$

where the last step follows from the fact that, for balanced partitions σ and τ, the overlap is a maximum, over all permutations π:

$$\begin{aligned}\mathrm{overlap}(\sigma,\tau) &= \frac{1}{n}\max_{\pi}\sum_{i=1}^{q}\left(|\sigma^{-1}(i)\cap\tau^{-1}(\pi(i))| - \frac{1}{n}|\sigma^{-1}(i)||\tau^{-1}(\pi(i))|\right) \\ &= \frac{1}{q}\max_{\pi}\operatorname{tr}\pi^{-1}\alpha - \frac{1}{q}\,,\end{aligned}$$

completing the proof. ■

The function in (16) is an increasing function of λ, since as λ increases the distributions $\mathrm{Bin}[\overline{m}, q]$ and $\mathrm{Bin}[\overline{m}, c_{\mathrm{in}}/(qd)]$ become closer in Kullback-Leibler distance. Thus if τ has overlap β, Lemma 10 implies

$$\begin{aligned}&\lim_{n\to\infty}\frac{1}{n}\Pr[\tau\text{ is good}] \\ &\quad\le -\frac{d}{2q}\left[(1+(q-1)\lambda)\log\frac{1+(q-1)\lambda}{1+q\beta\lambda} + (q-1)(1-\lambda)\log\frac{(q-1)(1-\lambda)}{q-1-q\beta\lambda}\right]\,. \end{aligned} \tag{17}$$

For fixed σ, the number of balanced partitions τ with overlap matrix α is the number of ways to partition each group $\sigma^{-1}(r)$ so that there are $\alpha_{rs}n/q$ vertices in $\sigma^{-1}(r)\cap\tau^{-1}(s)$:

$$\prod_{r=1}^{q}\binom{n/q}{\{\alpha_{rs}n/q \mid 1\le s\le q\}} = \prod_{r=1}^{q}\frac{(n/q)!}{\prod_s(\alpha_{r,s}n/q)!} \le \mathrm{e}^{nH(\alpha)}\,,$$

where $H(\alpha)$ is the average entropy of the rows of α_{rs}/q,

$$H(\alpha) = -\frac{1}{q}\sum_{r,s}\alpha_{rs}\log\alpha_{rs}\,. \tag{18}$$

By the union bound, the probability that there are any good partitions with overlap matrix α is exponentially small whenever the sum of $H(\alpha)$ and the right-hand side of (17) is negative. For a fixed overlap β, maximized by the permutation π, the entropy $H(\alpha)$ is maximized when

$$\alpha_{rs} = \begin{cases}\frac{1}{q}+\beta & \text{if } s=\pi(r)\\ \frac{1}{q}-\frac{\beta}{q-1} & \text{if } s\neq\pi(r)\,,\end{cases}$$

so we have

$$H(\alpha) \le h\left(\frac{1}{q}+\beta\right) + \left(1-\frac{1}{q}-\beta\right)\log(q-1)\,. \tag{19}$$

Combining the bounds (17) and (19), and requiring that their sum is at least zero, completes the proof. ■

3.3. Detection below the Kesten-Stigum bound

In §1.2 we commented on the asymptotic behavior of $d_{\mathrm{c}}^{\mathrm{upper}}$ in various regimes. In Table 1 we give, for various values of q, the point λ^* at which $d_{\mathrm{c}}^{\mathrm{upper}} = 1/\lambda^2$; then $d_{\mathrm{c}}^{\mathrm{upper}} < 1/\lambda^2$ for $\lambda < \lambda^*$. As stated above, in the limit $q\to\infty$ we have $d_{\mathrm{c}}^{\mathrm{upper}} = 2/\lambda$, so λ^* tends to $1/2$.

q	5	6	7	8	9	10	11	20	100	1000	10^4
λ^*	-0.239	-0.166	-0.112	-0.070	-0.036	-0.08	0.014	0.127	0.286	0.372	0.410

Table 1: For $\lambda < \lambda^*$ we have $d_{\rm c}^{\rm upper} < 1/\lambda^2$, so that community detection is information-theoretically possible below the Kesten-Stigum bound. For $q \geq 5$, this holds in the sufficiently disassortative case, including planted graph coloring where $\lambda = -1/(q-1)$. For $q \geq 11$, it occurs throughout the disassortative range $\lambda < 0$, and in some assortative cases.

4. Lower bound for symmetric SBMs: Proof of lower bound in Theorem 3

In this section we use the general bound of Theorem 7 to prove the part of Theorem 3 involving $d_{\rm c}^{\rm lower}$. In particular, we study the quantity Q—defined in Definition 6—in the case of symmetric stochastic block models. Note that Q is defined as the maximum of a certain function over the set of doubly stochastic matrices. This kind of maximization problem was studied extensively by Achlioptas and Naor (2005) on the way to proving their lower bound on the q-colorability threshold, allowing us to relate this problem to theirs.

First, note that $Q(\pi, (M - d\mathbb{J})/\sqrt{2d})$ simplifies considerably in the symmetric case, when $\pi_i = \frac{1}{q}$ for all i and M is determined by only two parameters. In this case, $\Delta_{q^2}(\pi)$ is (up to scaling) the set of doubly stochastic matrices, while

$$M - d\mathbb{J} = \lambda d \begin{pmatrix} q-1 & & -1 \\ & \ddots & \\ -1 & & q-1 \end{pmatrix}.$$

Going back to Definition 6, we see that $Q(\pi, (M - d\mathbb{J})/\sqrt{2d}) < 1$ if and only if $\Phi(\alpha) < 0$ for all doubly stochastic α, where

$$\Phi(\alpha) = H(\alpha) - \log q + \frac{d\lambda^2}{2}\left(|\alpha|^2 - 1\right), \tag{20}$$

$|\alpha|$ denoting the Frobenius norm and $H(\cdot)$ the average row entropy of α/q as in (18). By Theorem 7, if $\Phi(\alpha) < 0$ for all doubly stochastic α then (i) $\mathbb{P}_n$ and $\mathbb{Q}_n$ are contiguous, and (ii) $\mathbb{P}_n$ is non-detectable.

4.1. Maximizing Φ

Achlioptas and Naor (2005), in the process of proving a lower bound on the q-coloring threshold for Erdős-Rényi graphs, develop substantial machinery for optimizing Φ-like functions over the polytope of doubly stochastic matrices. Specifically, they relax the problem to maximizing over all row-stochastic matrices, and show that the maximizer is then a mixture of uniform rows and rows where all but one of the entries are identical. Although their bound is quite general, we quote here their results for the entropy. (Note that their definition of $H(\alpha)$ and ours differ by a factor of q.)

Theorem 11 *(Achlioptas and Naor, 2005, Theorem 9) Let α be doubly stochastic with $|\alpha|^2 = \rho$. Then*

$$H(\alpha) \le \max_{m \in \left[0, \frac{q(q-\rho)}{q-1}\right]} \left\{ \frac{m}{q} \log q + \left(1 - \frac{m}{q}\right) f\left(\frac{q\rho - m}{q(q-m)}\right) \right\}, \tag{21}$$

where

$$f(r) = g\left(\frac{1 + \sqrt{(q-1)(qr-1)}}{q}\right) + (q-1)\, g\left(\frac{1 - \frac{1+\sqrt{(q-1)(qr-1)}}{q}}{q-1}\right)$$

and $g(x) = -x \log x$.

With this result in hand and using $f(1/q) = q\, g(1/q) = \log q$, we know that for all α with $|\alpha|^2 = \rho$,

$$\Phi(\alpha) \le \max_{m \in \left[0, q(q-\rho)/(q-1)\right]} \left(1 - \frac{m}{q}\right)\left(f\left(\frac{q\rho - m}{q(q-m)}\right) - f(1/q)\right) + \frac{d\lambda^2}{2}(\rho - 1).$$

Achlioptas and Naor determined the value of $d\lambda^2/2$ for which the right-hand side is less than or equal to zero for all $m \in [0, q(q-\rho)/(q-1)]$ and all $\rho \in [1, q]$.

Lemma 12 *(Achlioptas and Naor, 2005, Proof of Theorem 7) When $\delta < (q-1)\log(q-1)$,*

$$\frac{\delta(\rho - 1)}{(q-1)^2} \le \left(1 - \frac{m}{q}\right)\left(f(1/q) - f\left(\frac{q\rho - m}{q(q-m)}\right)\right)$$

for all $m \in [0, q(q-\rho)/(q-1)]$ and all $\rho \in [1, q]$.

Our lower bound is an immediate corollary of this lemma. Substituting $\delta = d\lambda^2(q-1)^2/2$ and solving for d gives

$$d_{\mathrm{c}}^{\mathrm{lower}} = \frac{2\log(q-1)}{q-1}\frac{1}{\lambda^2}. \tag{22}$$

As we commented in §1.2, this corresponds to the lower bound on the q-colorability threshold of $G(n, d'/n)$ where $d' = 2\delta = d\lambda^2(q-1)^2$, scaling the eigenvalue on each edge to λ from its value $-1/(q-1)$ for q-coloring. This fits with the Kesten-Stigum threshold as well, since the amount of information (appropriately defined) transmitted along each edge is proportional to λ^2 (Janson and Mossel (2004)).

5. The second moment argument: Proof of Proposition 8

In this section, we will prove Proposition 8, thereby showing the link between the condition $Q(\pi, (M - d\mathbb{J})/\sqrt{2d}) < 1$ and the boundedness of certain second moments. Our first lemma expresses the second moment in question in terms of (centered and normalized) multinomial random variables. In order to state the lemma, we make the following notation. Given two assignments $\sigma, \tau \in [q]^n$, let $N_{ij} := N_{ij}(\sigma, \tau) := |\{v : \sigma_v = i, \tau_v = j\}|$, and $X_{ij} := X_{ij}(\sigma, \tau) := n^{-1/2}(N_{ij} - n\pi_i\pi_j)$. Recall that Ω_n is the event that the label frequencies are approximately their expected values, and let Y_n denote the restricted density $1_{\Omega_n}\frac{d\mathbb{P}_n}{d\mathbb{Q}_n}$. With a slight overloading of notation, we write $\sigma \in \Omega_n$ if for all $i \in [q]$, $|\{u : \sigma_u = i\}| = n\pi_i \pm a_n$. Set $A := M - d\mathbb{J}$.

Lemma 13 *We have:*

$$\mathbb{E}_{\mathbb{Q}_n} Y_n^2 = (1+O(n^{-1})) \sum_{\sigma,\tau\in\Omega_n} \mathbb{P}_n(\sigma)\mathbb{P}_n(\tau)\exp\left(\frac{1}{2d}\sum_{ijk\ell} X_{ij}X_{k\ell}A_{ik}A_{j\ell} + \nu_1 + \nu_2 + \xi_n\right),$$

where

$$\begin{aligned}\nu_1 &= -\frac{1}{2d}\sum_{ij} A_{ii}A_{jj}\pi_i\pi_j,\\ \nu_2 &= -\frac{1}{2d^2}\sum_{ijk\ell} A_{ik}^2 A_{j\ell}^2 \pi_i\pi_j\pi_k\pi_\ell, \text{ and}\\ \xi_n &= O(n^{-1/2})\sum_{ij}|X_{ij}| + O(n^{-1})\left(\sum_{ij}|X_{ij}|\right)^2.\end{aligned}$$

Proof For a graph G and assignment σ, define

$$W_{uv}(G,\sigma) = \begin{cases} \frac{M_{\sigma_u,\sigma_v}}{d} & \text{if } (u,v)\in E(G)\\ \frac{1-\frac{M_{\sigma_u,\sigma_v}}{n}}{1-\frac{d}{n}} & \text{if } (u,v)\notin E(G).\end{cases}$$

Then we may write out

$$\begin{aligned} Y_n &= \sum_{\sigma\in\Omega_n}\frac{\mathbb{P}_n(G,\sigma)}{\mathbb{Q}_n(G)}\\ &= \sum_{\sigma\in\Omega_n}\mathbb{P}_n(\sigma)\prod_{u,v} W_{uv}(G,\sigma).\end{aligned}$$

Squaring both sides and taking expectations,

$$\begin{aligned}\mathbb{E}_{\mathbb{Q}_n} Y_n^2 &= \mathbb{E}_{\mathbb{Q}_n}\sum_{\sigma,\tau\in\Omega_n}\mathbb{P}_n(\sigma)\mathbb{P}_n(\tau)\prod_{u,v}W_{uv}(G,\sigma)W_{uv}(G,\tau)\\ &= \sum_{\sigma,\tau\in\Omega_n}\mathbb{P}_n(\sigma)\mathbb{P}_n(\tau)\prod_{u,v}\mathbb{E}_{\mathbb{Q}_n}[W_{uv}(G,\sigma)W_{uv}(G,\tau)], \qquad (23)\end{aligned}$$

where the last equality holds because under $\mathbb{Q}_n$, and for any fixed σ, the variables $W_{uv}(G,\sigma)$ are independent as u and v vary.

Let us compute the inner expectation in (23). Recall that under $\mathbb{Q}_n$, $(u,v)\in E(G)$ with probability $\frac{d}{n}$. Writing (for brevity) s for $M_{\sigma_u\sigma_v}$ and t for $M_{\tau_u\tau_v}$, we have

$$\begin{aligned}\mathbb{E}_{\mathbb{Q}_n} W_{uv}(G,\sigma)W_{uv}(G,\tau) &= \frac{st}{d^2}\cdot\frac{d}{n} + \frac{(1-\frac{s}{n})(1-\frac{t}{n})}{(1-\frac{d}{n})^2}(1-\frac{d}{n})\\ &= \frac{st}{nd} + \left(1-\frac{s}{n}\right)\left(1-\frac{t}{n}\right)\left(1+\frac{d}{n}+\frac{d^2}{n^2}+O(n^{-3})\right)\\ &= 1 + \frac{(s-d)(t-d)}{nd} + \frac{(s-d)(t-d)}{n^2} + O(n^{-3})\end{aligned}$$

Setting $r = (s-d)(t-d)$, and using the fact that $1 + x = \exp(x - x^2/2 + O(x^3))$, we have

$$\mathbb{E}_{\mathbb{Q}_n} W_{uv}(G,\sigma) W_{uv}(G,\tau) = \exp\left(\frac{r}{dn} + \frac{r}{n^2} - \frac{r^2}{2d^2n^2} + O(n^{-3})\right).$$

Now, if $(\sigma_u, \tau_u, \sigma_v, \tau_v) = (i, j, k, \ell)$ then $(s-d)(t-d) = (M_{ik} - d)(M_{j\ell} - d) = A_{ik}A_{j\ell}$. Hence,

$$\mathbb{E}_{\mathbb{Q}_n} W_{uv}(G,\sigma) W_{uv}(G,\tau) = \exp\left(\frac{A_{ik}A_{j\ell}}{dn} + \frac{A_{ik}A_{j\ell}}{n^2} - \frac{(A_{ik}A_{j\ell})^2}{2d^2n^2} + O(n^{-3})\right). \tag{24}$$

Let $N_{ijk\ell} = |\{\{u,v\} : \sigma_u = i, \tau_u = j, \sigma_v = k, \tau_v = \ell\}|$. Plugging (24) into (23), we have

$$\mathbb{E}_{\mathbb{Q}_n} Y_n^2 = (1 + O(n^{-1})) \sum_{\sigma,\tau\in\Omega_n} \mathbb{P}_n(\sigma)\mathbb{P}_n(\tau) \exp\left(\sum_{ijk\ell=1}^{s} N_{ijk\ell}\left(\frac{A_{ik}A_{j\ell}}{dn} + \frac{A_{ik}A_{j\ell}}{n^2} - \frac{(A_{ik}A_{j\ell})^2}{2d^2n^2}\right)\right) \tag{25}$$

where the $(1 + O(n^{-1}))$ term arises because $\sum_{ijk\ell} N_{ijk\ell} \le n^2$. Applying Lemma 14 (below) now finishes the proof. ■

The last step in the proof of Lemma 13 requires us to replace $N_{ijk\ell}$ by its normalized version, X_{ij}, and then rearrange the sums in (25). We will do this step in slightly more generality, where we allow $N_{ijk\ell}$ to be defined on a subset of the vertices. For the purposes of this section it suffices to consider $S = [n]$, but the general form will be useful when we prove Theorem 7.

Lemma 14 *Let $S \subseteq [n]$ such that $|S| = n - o(n)$. Further, let*

$$\begin{aligned}
N_{ijk\ell} &:= N_{ijk\ell}(\sigma,\tau) := |\{\{u,v\} : u, v \in S, \sigma_u = i, \tau_u = j, \sigma_v = k, \tau_v = \ell\}|\,, \\
N_{ij} &:= N_{ij}(\sigma,\tau) := |\{u : u \in S, \sigma_u = i, \tau_u = j\}| \text{ and,} \\
X_{ij} &:= X_{ij}(\sigma,\tau) := n^{-1/2}\left(N_{ij} - n\pi_i\pi_j\right) \\
t_{ijk\ell} &:= \frac{A_{ik}A_{j\ell}}{dn} + \frac{A_{ik}A_{j\ell}}{n^2} - \frac{(A_{ik}A_{j\ell})^2}{2d^2n^2}
\end{aligned}$$

Then, we have:

$$\sum_{ijk\ell} N_{ijk\ell} t_{ijk\ell} = \frac{1}{2d} \sum_{ijk\ell} X_{ij} X_{k\ell} A_{ik} A_{j\ell} + \nu_1 + \nu_2 + \xi_n,$$

where

$$\begin{aligned}
\nu_1 &= -\frac{1}{2d} \sum_{ij} A_{ii} A_{jj} \pi_i \pi_j, \\
\nu_2 &= -\frac{1}{4d^2} \sum_{ijk\ell} A_{ik}^2 A_{j\ell}^2 \pi_i \pi_j \pi_k \pi_\ell, \text{ and} \\
\xi_n &= O(n^{-1/2}) \sum_{ij} |X_{ij}| + O(n^{-1}) \left(\sum_{ij} |X_{ij}|\right)^2 + O(n^{-1}).
\end{aligned}$$

Proof We see that $N_{ijk\ell} = \frac{1}{2} N_{ij} N_{k\ell}$ unless $i = k$ and $j = \ell$, in which case $N_{ijk\ell} = \binom{N_{ij}}{2} = \frac{1}{2} N_{ij} N_{k\ell} - \frac{1}{2} N_{ij}$. So, we have

$$\sum_{ijk\ell} N_{ijk\ell} t_{ijk\ell} = \frac{1}{2} \sum_{ijk\ell} N_{ij} N_{k\ell} t_{ijk\ell} - \frac{1}{2} \sum_{ij} N_{ij} t_{ijij} \tag{26}$$

Recall that $\sum_i \pi_i M_{ik} = d$ for any fixed k and $\sum_k \pi_k M_{ik} = d$ for any fixed i. It follows that $\sum_i \pi_i A_{ij} = \sum_j \pi_j A_{ij} = 0$. Hence,

$$\sum_i \pi_i t_{ijk\ell} = -\sum_i \pi_i \frac{(A_{ik} A_{j\ell})^2}{2d^2 n^2}.$$

Writing $N_{ij} = \sqrt{n} X_{ij} + n \pi_i \pi_j$, we have

$$\begin{aligned}\sum_{ijk\ell} N_{ij} N_{k\ell} t_{ijk\ell} &= n \sum_{ijk\ell} X_{ij} X_{k\ell} t_{ijk\ell} - \sum_{ijk\ell} \frac{(A_{ik} A_{j\ell})^2}{2d^2 n^2} \left(n^{3/2} X_{ij} \pi_k \pi_\ell + n^{3/2} X_{k\ell} \pi_i \pi_j + n^2 \pi_i \pi_j \pi_k \pi_\ell \right) \\ &= n \sum_{ijk\ell} X_{ij} X_{k\ell} t_{ijk\ell} - \sum_{ijk\ell} \frac{(A_{ik} A_{j\ell})^2}{2d^2} \pi_i \pi_j \pi_k \pi_\ell + O(n^{-1/2}) \sum_{ij} |X_{ij}|,\end{aligned}$$

Next, note that $t_{ijk\ell} = \frac{1}{dn} A_{ik} A_{j\ell} + O(n^{-2})$, and so

$$\begin{aligned}\sum_{ijk\ell} N_{ij} N_{k\ell} t_{ijk\ell} = \frac{1}{d} \sum_{ijk\ell} X_{ij} X_{k\ell} A_{ik} A_{j\ell} - \frac{1}{2d^2} \sum_{ijk\ell} (A_{ik} A_{j\ell})^2 \pi_i \pi_j \pi_k \pi_\ell \\ + O(n^{-1/2}) \sum_{ij} |X_{ij}| + O(n^{-1}) \left(\sum_{ij} |X_{ij}| \right)^2;\end{aligned}$$

we recognize the second term as $2\nu_2$, and the last two terms as being part of ξ_n. This takes care of first term in (26); for the second term,

$$\sum_{ij} N_{ij} t_{ijij} = \sqrt{n} \sum_{ij} X_{ij} t_{ijij} + n \sum_{ij} \pi_i \pi_j t_{ijij} = O(n^{-1/2}) \sum_{ij} |X_{ij}| + \frac{1}{d} \sum_{ij} A_{ii} A_{jj} \pi_i \pi_j + O(n^{-1});$$

here, the second term is $2\nu_1$ and the others are part of ξ_n. ■

The following lemma gives a simpler form for ν_1 and ν_2 appearing above. In particular, this will allow us to relate ν_1 and ν_2 to the eigenvalues of T. We define $B := \frac{1}{d} \operatorname{diag}(\pi) A = T - \pi \otimes \mathbf{1}^\intercal$, where $^\top$ denotes the transpose and $\otimes$ denotes the Kronecker product.

Lemma 15 *Let ν_1 and ν_2 be as in Lemma 14. Then, we have:*

$$\begin{aligned}\nu_1 &= -\frac{d}{2} \operatorname{tr}(B)^2 \\ \nu_2 &= -\frac{d^2}{4} \operatorname{tr}(B^2)^2.\end{aligned}$$

Proof Note that $A_{ii}\pi_i = dB_{ii}$. Hence,

$$\nu_1 = -\frac{1}{2d}\sum_{ij} A_{ii}A_{jj}\pi_i\pi_j = -\frac{d}{2}\sum_{ij} B_{ii}B_{jj} = -\frac{d}{2}\operatorname{tr}(B)^2.$$

Similarly, since $A_{ik}\pi_i = B_{ik}$ and $A_{ik}\pi_k = A_{ki}\pi_k = B_{ki}$,

$$\nu_2 = -\frac{d^2}{4}\sum_{ijk\ell} B_{ik}B_{ki}B_{j\ell}B_{\ell j} = -\frac{d^2}{4}\operatorname{tr}\left((B^{\otimes 2})^2\right) = -\frac{d^2}{4}\operatorname{tr}(B^2)^2.$$

■

The following lemma shows that ξ_n in Lemma 14 is very small in an appropriate sense.

Lemma 16 *Let ξ_n be as in Lemma 14. If $a_n = o(n^{1/2})$ then $\mathbb{E}\exp(a_n\xi_n) \to 1$.*

Proof By the central limit theorem, each X_{ij} has a limit in distribution as $n \to \infty$; hence $a_n\xi_n \to 0$ in probability. It is therefore enough to show that the sequence $\exp(a_n\xi_n)$ is uniformly integrable, but this follows from Hoeffding's inequality: since X_{ij} is a centered, renormalized sum of independent indicator variables, Hoeffding's inequality implies that

$$\Pr(|X_{ij}| \ge t) \le 2e^{-t^2/2}.$$

Let $X = X_{11}$; the definition of ξ_n ensures that there is a constant C such that ξ_n is stochastically dominated by $C(Y + Y^2 + n^{-1})$, where $Y = n^{-1/2}q^2|X|$. Hence,

$$\Pr(\xi_n \ge C(t + t^2 + n^{-1})) \le \Pr(Y \ge t) \le 2e^{-\frac{nt^2}{2q^2}}.$$

Since q is a constant, this may be rearranged to state that

$$\Pr(\xi_n \ge t) \le 2e^{-cn\min\{t,t^2\}} \tag{27}$$

for some constant c and all $t \ge 0$. Finally, for any $M \ge 0$

$$\begin{aligned}\mathbb{E}[e^{a_n\xi_n}1_{\{e^{a_n\xi_n}\ge M\}}] &= \Pr(e^{a_n\xi_n} \ge M) + \int_M^\infty \Pr(e^{a_n\xi_n} \ge t)\,dt \\ &= \Pr\left(\xi_n \ge \frac{\log M}{a_n}\right) + \int_M^\infty \Pr\left(\xi_n \ge \frac{\log t}{a_n}\right)dt\end{aligned}$$

If $a_n = o(n^{1/2})$ then (27) implies that both terms above converge to zero (uniformly in n) as $M \to \infty$. ■

We now state the following three results before we prove the main result of this section. The following proposition characterizes when the exponential of a quadratic form of a sequence of multinomial random variables is uniformly integrable. Its proof can be found in Section A.

Proposition 17 *Define X_{ij} as in Lemma 14. Then*

$$\exp\left(\frac{1}{2d}\sum X_{ij}X_{k\ell}A_{ik}A_{j\ell}\right)$$

is uniformly integrable if $Q(\pi, A/\sqrt{2d}) < 1$, and fails to be uniformly integrable if $Q(\pi, A/\sqrt{2d}) > 1$.

Using Hölder's inequality, it is fairly straightforward to introduce the ξ_n term:

Lemma 18 *Define X_{ij} as in Lemma 14. Then*

$$\exp\left(\frac{1}{2d}\sum X_{ij}X_{k\ell}A_{ik}A_{j\ell} + \xi_n\right)$$

is uniformly integrable if $Q(\pi, A/\sqrt{2d}) < 1$, and fails to be uniformly integrable if $Q(\pi, A/\sqrt{2d}) > 1$.

Proof Supposing that $Q(\pi, A/\sqrt{2d}) < 1$, we find some $\epsilon > 0$ such that $Q(\pi, \sqrt{1+\epsilon}A/\sqrt{2d}) < 1$. Set $a_n = n^{1/3}$ and $b_n = \frac{a_n}{a_n - 1}$ to be the Hölder conjugate of a_n. Setting

$$W := \operatorname{vec}(X) \in \mathbb{R}^{q^2}, \tag{28}$$

Hölder's inequality and Lemma 16 give

$$\begin{aligned}
&\mathbb{E}_{\sigma,\tau} \exp\left((1+\frac{\epsilon}{2})\left(\frac{1}{2d}\sum_{ijk\ell} X_{ij}X_{k\ell}A_{ik}A_{j\ell} + \xi_n\right)\right) \\
&\le \left(\mathbb{E}_{\sigma,\tau}\exp\left(\frac{(1+\frac{\epsilon}{2})b_n}{2d} W^T(A^{\otimes 2})W\right)\right)^{1/b_n} \left(\mathbb{E}\exp((1+\frac{\epsilon}{2})a_n\xi_n)\right)^{1/a_n} \\
&\le \left(\mathbb{E}_{\sigma,\tau}\exp\left(\frac{(1+\frac{\epsilon}{2})b_n}{2d} W^T(A^{\otimes 2})W\right)\right)^{1/b_n}.
\end{aligned}$$

To check uniform integrability, we apply Proposition 17. For sufficiently large n, we have $b_n \le \frac{1+\epsilon}{1+\frac{\epsilon}{2}}$ and

$$\exp\left(\frac{(1+\frac{\epsilon}{2})b_n}{2d} W^T A^{\otimes 2} W\right) \le \max\left\{1, \exp\left(\frac{(1+\epsilon)}{2d} W^T A^{\otimes 2} W\right)\right\}.$$

We see from the fact that $Q(\pi, \sqrt{1+\epsilon}A/\sqrt{2d}) < 1$ and Proposition 17 that the right hand side above has a finite expectation.

To summarize, we have shown that if $Z = \exp(\frac{1}{2d}\sum X_{ij}X_{k\ell}A_{ik}A_{j\ell} + \xi_n)$ then $\mathbb{E}Z^{(1+\epsilon/2)} < \infty$ for some $\epsilon > 0$. It follows that Z is uniformly integrable, as claimed.

To show that $Q(\pi, A/\sqrt{2d}) > 1$ implies non-uniform integrability, requires an almost identical argument, but using the reverse Hölder inequality instead of the usual Hölder inequality. We omit the details. ■

The following lemma calculates the expected value of the exponential of a quadratic form of a Gaussian random vector.

Lemma 19 *Take $Z \sim \mathcal{N}(0, \Sigma)$, where $\Sigma = \operatorname{diag}(\pi)^{\otimes 2} - \left(\pi \otimes \pi\right)^{\otimes 2}$, where $a \otimes b$ denotes the Kronecker product of a and b, and $a^{\otimes 2}$ denotes the outer product of a with itself. Recall that λ_i denote the eigenvalues of T, with $1 = \lambda_1 \geq |\lambda_2| \geq \cdots \geq |\lambda_q|$. If $d\lambda_2^2 < 1$ then*

$$\mathbb{E} \exp\left(\frac{1}{2d} Z^T A^{\otimes 2} Z\right) = \prod_{i,j=2}^{q} \frac{1}{\sqrt{1 - d\lambda_i \lambda_j}}.$$

Otherwise, $\mathbb{E} \exp\left(\frac{1}{2d} Z^T A^{\otimes 2} Z\right) = \infty$.

Proof A standard computation (see, e.g. Mathai and Provost (1992)) shows that if $\mu_1, \ldots, \mu_s$ denote the eigenvalues of $\Sigma\tilde{A}$ then $\mathbb{E}\exp(Z^T \tilde{A} Z/2) = \prod_i \frac{1}{\sqrt{1-\mu_i}}$. Now,

$$\Sigma A^{\otimes 2} = \left(\operatorname{diag}(\pi)^{\otimes 2} - \left(\pi \otimes \pi\right)^{\otimes 2}\right) A^{\otimes 2} = (\operatorname{diag}(\pi) A)^{\otimes 2} - (\pi\pi^\intercal A)^{\otimes 2}.$$

Recall, however, that $A\pi = 0$. Hence, we are interested in the eigenvalues of $(\operatorname{diag}(\pi)A)^{\otimes 2} = (dB)^{\otimes 2}$. Since the top eigenvalue of T is 1 (with 1 as its right-eigenvector and π as its left-eigenvector), we see that if $\lambda_1, \cdots, \lambda_q$ are the eigenvalues of T with $\lambda_1 = 1$, then

$$\{d\lambda_i \lambda_j : i, j = 2, \ldots, q\}$$

are the eigenvalues of $\frac{1}{d}\Sigma(A \otimes A)$. ■

Proof [Proof of Proposition 8] First of all, note that

$$\frac{d\hat{\mathbb{P}}_n(G, \sigma)}{d\mathbb{Q}_n} = \frac{Y_n}{\mathbb{P}_n(\Omega_n)} = (1 + o(1))Y_n.$$

Hence, it suffices to compute the limit of $\mathbb{E}_{\mathbb{Q}_n} Y_n^2$.

From Lemma 13, we see that we need to calculate the limit of the quantity

$$\mathbb{E}_{\sigma,\tau \in \Omega_n} \exp\left(\frac{1}{2d} \sum_{ijk\ell} X_{ij} X_{k\ell} A_{ik} A_{j\ell} + \xi_n\right).$$

Lemma 18 establishes that the above sequence is uniformly integrable.

Now, note that $(N_{ij})_{i,j=1}^q$ is distributed as a multinomial random vector with n trials and probabilities $\pi_i \pi_j$. In particular, $\frac{1}{n}\mathbb{E}N_{ij} = \pi_i \pi_j$, $\frac{1}{n}\operatorname{Var}(N_{ij}) = \pi_i \pi_j - (\pi_i \pi_j)^2$, and $\frac{1}{n}\operatorname{Cov}(N_{ij} N_{k\ell}) = -\pi_i \pi_j \pi_k \pi_\ell$ if $\{i, j\} \neq \{k, \ell\}$. Since $X_{ij} = n^{-\frac{1}{2}}(N_{ij} - n\pi_i \pi_j)$, central limit theorem implies that $W := \operatorname{vec}(X) \in \mathbb{R}^{k^2}$ converges in distribution to a Gaussian random vector, Z with mean 0 and covariance matrix $\operatorname{diag}(\pi)^{\otimes 2} - \pi^{\otimes 4}$. Using Lemma 19 now gives us

$$\mathbb{E}_{\mathbb{Q}_n} Y_n^2 \to \exp(\nu_1 + \nu_2) \prod_{i,j=2}^{q} \frac{1}{\sqrt{1 - d\lambda_i \lambda_j}}. \tag{29}$$

Going back to Lemma 15, we have

$$\nu_1 = -\frac{d}{2}\operatorname{tr}(B)^2 = -\frac{1}{2}\sum_{i,j=2}^{q} d\lambda_i \lambda_j$$

and

$$\nu_2 = -\frac{d^2}{4}\operatorname{tr}(B^2)^2 = -\frac{1}{4}\sum_{i,j=2}^{q}(d\lambda_i\lambda_j)^2.$$

Hence, the right hand side of (29) is equal to

$$\prod_{i,j}\psi(d\lambda_i\lambda_j),$$

as claimed. ■

6. Proof of Theorem 7

6.1. Non-distinguishability

In this section, we use Proposition 8 to the contiguity claim in Theorem 7. Our main tool is the conditional second moment method, which was originally developed by Robinson and Wormald (1992) in their study of Hamiltonian cycles in d-regular graphs. Janson (1995) was the first to apply this method for proving contiguity. We use a formulation from (Wormald, 1999, Theorem 4.1):

Theorem 20 *Consider two sequences $\mathbb{P}_n, \mathbb{Q}_n$ of probability distributions on a sequence Ω_n of probability spaces. Suppose that there exist random variables $\{X_{m,n} : m \geq 3\}$, where $X_{m,n}$ is defined on Ω_n, such that for every m,*

$$X_{m,n} \xrightarrow{d} \text{Pois}(\mu_m) \text{ under } \mathbb{Q}_n \text{ as } n \to \infty; \text{ and} \tag{30}$$

$$X_{m,n} \xrightarrow{d} \text{Pois}(\mu_m(1+\delta_m)) \text{ under } \mathbb{P}_n \text{ as } n \to \infty. \tag{31}$$

Suppose also that for any m^*, the collection $X_{3,n}, \dots, X_{m^*,n}$ are asymptotically independent as $n \to \infty$ under both $\mathbb{P}_n$ and $\mathbb{Q}_n$, in the sense that every joint moment of $X_{3,n}, \dots, X_{m^*,n}$ converges to the same joint moment of the appropriate independent Poisson variables. If

$$\mathbb{E}_{\mathbb{Q}_n}\left(\frac{\mathbb{P}_n}{\mathbb{Q}_n}\right)^2 \leq (1+o(1))\exp\left(\sum_{m\geq 3}\mu_m\delta_m^2\right) < \infty \tag{32}$$

then $\mathbb{P}_n$ and $\mathbb{Q}_n$ are contiguous.

We will apply Theorem 20 with $\mathbb{P}_n$ replaced by $\hat{\mathbb{P}}_n = (\mathbb{P}_n \mid \Omega_n)$; i.e., the block model conditioned on having almost the expected label frequencies. We will take $X_{m,n}$ to be the number of m-cycles in the graph G (which is drawn either from $\hat{\mathbb{P}}_n$ or from $\mathbb{Q}_n$). In order to apply Theorem 20, we need to know that the number of m-cycles has a limiting Poisson distribution (and we need to know the parameters). For $\mathbb{Q}_n$, this is classical; for $\mathbb{P}_n$ it was proved by Bollobás et al. (2007) (and it follows for $\hat{\mathbb{P}}_n$ since $\hat{\mathbb{P}}_n$ is obtained from $\mathbb{P}_n$ by conditioning on an event that holds with probability converging to 1).

Proposition 21 *Let X_m be the number of m-cycles in G. Then*

$$X_m \xrightarrow{d} \mathrm{Pois}\left(\frac{1}{2m}d^m\right) \textit{ under } \mathbb{Q}_n\textit{, and}$$
$$X_m \xrightarrow{d} \mathrm{Pois}\left(\frac{1}{2m}d^m\,\mathrm{tr}(T^m)\right) \textit{ under } \mathbb{P}_n.$$

Moreover, for any fixed m^* the variables $\{X_3, \dots, X_{m^*}\}$ are asymptotically independent under both $\mathbb{P}_n$ and $\mathbb{Q}_n$, in the sense of Theorem 20.

Hence, we may apply Theorem 20 with $\mu_m = \frac{1}{2m}d^m$ and $\delta_m = \mathrm{tr}(T^m) - 1$. Recalling that $1 = \lambda_1 \geq \cdots \geq \lambda_q$ are the eigenvalues of T, we have $\delta_m = \sum_{i\geq 2} \lambda_i^m$. Hence,

$$\begin{aligned}\sum_{m=3}^{\infty} \mu_m \delta_m^2 &= \frac{1}{2}\sum_{m=3}^{\infty}\frac{d^m}{m}\sum_{i,j=2}^{q}\lambda_i^m\lambda_j^m \\ &= \frac{1}{2}\sum_{i,j=2}^{q}\sum_{m=3}^{\infty}\frac{(d\lambda_i\lambda_j)^m}{m} \\ &= \sum_{i,j=2}^{q}\log\psi(d\lambda_i\lambda_j),\end{aligned}$$

where $\psi(x) = (1-x)^{-1/2}e^{-x/2-x^2/4}$. In particular, condition (32) follows immediately from Proposition 8, which in turn proves that $\hat{\mathbb{P}}_n$ and $\mathbb{Q}_n$ are contiguous. Since $\mathbb{P}_n(\Omega_n) \to 1$, $\mathbb{P}_n$ and $\hat{\mathbb{P}}_n$ are contiguous also. This proves the first statement of Theorem 7: if $Q(\pi, (M - d\mathbb{J})/\sqrt{2d}) < 1$ then $\mathbb{P}_n$ and $\mathbb{Q}_n$ are contiguous.

6.2. Non-detectability

Finally, in this section we prove that if $Q(\pi, A/\sqrt{2d}) < 1$ (where $A = M - d\mathbb{J}$) then $\mathbb{P}_n$ is non-detectable; this will complete the proof of Theorem 7. The following proposition is the main technical result we need. It shows that if $Q(\pi, A/\sqrt{2d}) < 1$ then for any two fixed configurations on a finite set of nodes, the total variation distance between the distribution on graphs conditioned on these two configurations respectively goes to zero.

Proposition 22 *Suppose $Q\left(\pi, A/\sqrt{2d}\right) < 1$. Then, for any fixed $r > 0$, and for any two configurations $(a_1, a_2, \cdots, a_r)$ and $(b_1, b_2, \cdots, b_r)$, we have:*

$$TV\left(\mathbb{P}_n\left(G|\sigma_u = a_u \textit{ for } u \in [r]\right), \mathbb{P}_n\left(G|\sigma_u = b_u \textit{ for } u \in [r]\right)\right) = o(1),$$

where $TV(\mathbb{P}_1, \mathbb{P}_2)$ denotes the total variation distance between the two distributions $\mathbb{P}_1$ and $\mathbb{P}_2$.

Proof We will first prove the statement of the proposition with $\mathbb{P}_n$ replaced by $\hat{\mathbb{P}}_n = (\mathbb{P}_n \mid \Omega_n)$; i.e., the block model conditioned on having almost the expected label frequencies.

We start by using the definition of total variation distance:

$$
\begin{aligned}
&TV\left(\hat{\mathbb{P}}_n\left(G|\sigma_u = a_u \text{ for } u \in [r]\right), \hat{\mathbb{P}}_n\left(G|\sigma_u = b_u \text{ for } u \in [r]\right)\right)\\
&= \sum_G \left|\hat{\mathbb{P}}_n\left(G|\sigma_u = a_u \text{ for } u \in [r]\right) - \hat{\mathbb{P}}_n\left(G|\sigma_u = b_u \text{ for } u \in [r]\right)\right|\\
&= \sum_G \left|\hat{\mathbb{P}}_n\left(G|\sigma_u = a_u \text{ for } u \in [r]\right) - \hat{\mathbb{P}}_n\left(G|\sigma_u = b_u \text{ for } u \in [r]\right)\right| \frac{\sqrt{\mathbb{Q}_n(G)}}{\sqrt{\mathbb{Q}_n(G)}}\\
&\overset{(a)}{\leq} \left(\sum_G \mathbb{Q}_n(G)\right)^{1/2} \left(\sum_G \frac{\left(\hat{\mathbb{P}}_n\left(G|\sigma_u = a_u \text{ for } u \in [r]\right) - \hat{\mathbb{P}}_n\left(G|\sigma_u = b_u \text{ for } u \in [r]\right)\right)^2}{\mathbb{Q}_n(G)}\right)^{1/2}\\
&= \left(\sum_G \frac{\left(\sum_{\widetilde{\sigma}} \hat{\mathbb{P}}_n(\widetilde{\sigma})\left(\hat{\mathbb{P}}_n(G|a,\widetilde{\sigma}) - \hat{\mathbb{P}}_n(G|b,\widetilde{\sigma})\right)\right)^2}{\mathbb{Q}_n(G)}\right)^{1/2},
\end{aligned}
$$

where (a) follows from Cauchy-Schwartz inequality and $\widetilde{\sigma}$ denotes an assignment on $[n] \setminus [r]$. We can expand the numerator as follows:

$$
\begin{aligned}
&\left(\sum_{\widetilde{\sigma}} \hat{\mathbb{P}}_n(\widetilde{\sigma})\left(\hat{\mathbb{P}}_n(G|a,\widetilde{\sigma}) - \hat{\mathbb{P}}_n(G|b,\widetilde{\sigma})\right)\right)^2\\
&= \sum_{\widetilde{\sigma},\widetilde{\tau}} \hat{\mathbb{P}}_n(\widetilde{\sigma})\hat{\mathbb{P}}_n(\widetilde{\tau})\Big(\hat{\mathbb{P}}_n(G|a,\widetilde{\sigma})\hat{\mathbb{P}}_n(G|a,\widetilde{\tau}) + \hat{\mathbb{P}}_n(G|b,\widetilde{\sigma})\hat{\mathbb{P}}_n(G|b,\widetilde{\tau})\\
&\qquad\qquad - \hat{\mathbb{P}}_n(G|a,\widetilde{\sigma})\hat{\mathbb{P}}_n(G|b,\widetilde{\tau}) - \hat{\mathbb{P}}_n(G|b,\widetilde{\sigma})\hat{\mathbb{P}}_n(G|a,\widetilde{\tau})\Big).
\end{aligned}
$$

We will now show that the value of

$$
\sum_{\widetilde{\sigma},\widetilde{\tau}} \hat{\mathbb{P}}_n(\widetilde{\sigma})\hat{\mathbb{P}}_n(\widetilde{\tau}) \sum_G \frac{\hat{\mathbb{P}}_n(G|a,\widetilde{\sigma})\hat{\mathbb{P}}_n(G|b,\widetilde{\tau})}{\mathbb{Q}_n(G)},
$$

is independent of a and b up to an $o(1)$ error term. This will prove our claim. Define

$$
W_{uv}(G,\sigma) := \begin{cases} \frac{M_{\sigma_u,\sigma_v}}{d} & \text{if } (u,v) \in E(G),\\ \frac{1-\frac{M_{\sigma_u,\sigma_v}}{n}}{1-\frac{d}{n}} & \text{if } (u,v) \notin E(G), \end{cases}
$$

and let $s_{ijk\ell} = (M_{ik} - d)(M_{j\ell} - d)/n = A_{ik}A_{j\ell}/n$, and $t_{ijk\ell} = \frac{s_{ijk\ell}}{d} + \frac{s_{ijk\ell}}{n} - \frac{s_{ijk\ell}^2}{2d^2}$. We have:

$$\begin{aligned}
&\sum_{\widetilde{\sigma},\widetilde{\tau}} \hat{\mathbb{P}}_n(\widetilde{\sigma})\,\hat{\mathbb{P}}_n(\widetilde{\tau}) \sum_G \frac{\hat{\mathbb{P}}_n(G|a,\widetilde{\sigma})\,\hat{\mathbb{P}}_n(G|b,\widetilde{\tau})}{\mathbb{Q}_n(G)} \\
&= \sum_{\widetilde{\sigma},\widetilde{\tau}} \hat{\mathbb{P}}_n(\widetilde{\sigma})\,\hat{\mathbb{P}}_n(\widetilde{\tau}) \prod_{u,v\in[n]} \mathbb{E}_{\mathbb{Q}_n}[W_{uv}(G,a,\widetilde{\sigma})W_{uv}(G,b,\widetilde{\tau})] \\
&= \hat{\mathbb{E}}_{\widetilde{\sigma},\widetilde{\tau}} \prod_{u,v\in[n]\setminus[r]} (1 + t_{\widetilde{\sigma}_u\widetilde{\tau}_u\widetilde{\sigma}_v\widetilde{\tau}_v} + \epsilon_n) \prod_{\substack{u\in[r]\\ v\in[n]\setminus[r]}} (1 + t_{a_ub_u\widetilde{\sigma}_v\widetilde{\tau}_v} + \epsilon_n) \prod_{u,v\in[r]} (1 + t_{a_ub_ua_vb_v} + \epsilon_n) \\
&= \hat{\mathbb{E}}_{\widetilde{\sigma},\widetilde{\tau}} \prod_{i,j,k,\ell\in[q]} (1 + t_{ijk\ell} + \epsilon_n)^{\widetilde{N}_{ijk\ell}} \prod_{\substack{u\in[r]\\ i,j\in[q]}} (1 + t_{a_ub_uij} + \epsilon_n)^{\widetilde{N}_{ij}} \prod_{u,v\in[r]} (1 + t_{a_ub_ua_vb_v} + \epsilon_n)\,,
\end{aligned} \tag{33}$$

where $\widetilde{N}_{ijk\ell} = |\{\{u,v\} : \widetilde{\sigma}_u = i, \widetilde{\tau}_u = j, \widetilde{\sigma}_v = k, \widetilde{\tau}_v = \ell\}|$, $\widetilde{N}_{ij} = |\{v : \widetilde{\sigma}_v = i, \widetilde{\tau}_v = j\}|$, and $\epsilon_n = O(n^{-3})$. We first note that the last term in (33) is essentially constant:

$$\prod_{u,v\in[r]} (1 + t_{a_ub_ua_vb_v} + \epsilon_n) = \prod_{u,v\in[r]} \left(1 + O\left(\frac{1}{n}\right)\right) = \left(1 + O\left(\frac{1}{n}\right)\right)^{r^2} = 1 + O\left(\frac{1}{n}\right).$$

For the second term in (33), since $\tilde{N}_{ij} < n$, we have

$$\begin{aligned}
\prod_{i,j\in[q]} (1 + t_{a_ub_uij} + \epsilon_n)^{\widetilde{N}_{ij}} &= \prod_{i,j\in[q]} \left(1 + \frac{s_{a_ub_uij}}{d} + O\left(\frac{1}{n^2}\right)\right)^{\widetilde{N}_{ij}} \\
&= (1 + o(1)) \prod_{i,j\in[q]} \exp\left(\frac{s_{a_ub_uij}}{d} \cdot \widetilde{N}_{ij}\right)
\end{aligned} \tag{34}$$

On the other hand,

$$\begin{aligned}
\prod_{i,j\in[q]} \exp\left(\frac{ns_{a_ub_uij}}{d} \cdot \pi_i\pi_j\right) &= \prod_{i,j\in[q]} \exp\left(\frac{\pi_i\pi_j A_{a_ui}A_{b_uj}}{d}\right) \\
&= \exp\left(\frac{\left(\sum_{i\in[q]} \pi_i A_{a_ui}\right)\left(\sum_{j\in[q]} \pi_j A_{b_uj}\right)}{d}\right) = 1,
\end{aligned}$$

and so we may write the second term of (33) as

$$\begin{aligned}
\prod_{i,j\in[q]} (1 + t_{a_ub_uij} + \epsilon_n)^{\widetilde{N}_{ij}} &= (1 + o(1)) \prod_{i,j\in[q]} \exp\left(\frac{ns_{a_ub_uij}}{d}\left(\frac{\widetilde{N}_{ij}}{n} - \pi_i\pi_j\right)\right) \\
&= (1 + o(1)) \prod_{i,j\in[q]} \exp\left(\frac{ns_{a_ub_uij}}{d} \cdot \frac{\widetilde{X}_{ij}}{\sqrt{n}}\right),
\end{aligned}$$

where $\widetilde{X}_{ij} := n^{-1/2}\left(\widetilde{N}_{ij} - n\pi_i\pi_j\right)$.

Going back to (33) and plugging in our estimates on the second and third terms,

$$\begin{aligned}
&\sum_{\widetilde{\sigma},\widetilde{\tau}} \hat{\mathbb{P}}_n\left(\widetilde{\sigma}\right)\hat{\mathbb{P}}_n\left(\widetilde{\tau}\right)\sum_G \frac{\hat{\mathbb{P}}_n\left(G|a,\widetilde{\sigma}\right)\hat{\mathbb{P}}_n\left(G|b,\widetilde{\tau}\right)}{\mathbb{Q}_n\left(G\right)} \\
&= (1+o(1))\hat{\mathbb{E}}_{\widetilde{\sigma},\widetilde{\tau}} \prod_{i,j,k,\ell\in[q]} \left(1+t_{ijk\ell}+\epsilon_n\right)^{\widetilde{N}_{ijk\ell}} \prod_{\substack{u\in[r]\\ i,j\in[q]}} \exp\left(\frac{ns_{a_ub_uij}}{d}\cdot\frac{\widetilde{X}_{ij}}{\sqrt{n}}\right) \\
&= (1+o(1))\hat{\mathbb{E}}_{\widetilde{\sigma},\widetilde{\tau}} \exp\left(\sum_{ijk\ell}\widetilde{N}_{ijk\ell}t_{ijk\ell}\right) \prod_{\substack{u\in[r]\\ i,j\in[q]}} \exp\left(\frac{ns_{a_ub_uij}}{d}\cdot\frac{\widetilde{X}_{ij}}{\sqrt{n}}\right) \\
&= (1+o(1))\hat{\mathbb{E}}_{\widetilde{\sigma},\widetilde{\tau}} \exp\left(\frac{1}{2d}\sum_{ijk\ell}\widetilde{X}_{ij}\widetilde{X}_{k\ell}A_{ik}A_{j\ell}+\nu_1+\nu_2+\widetilde{\xi}_n\right) \prod_{\substack{u\in[r]\\ i,j\in[q]}} \exp\left(\frac{ns_{a_ub_uij}}{d}\cdot\frac{\widetilde{X}_{ij}}{\sqrt{n}}\right),
\end{aligned}$$

where the last equality follows from Lemma 14. Note that $\exp\left(\frac{1}{2d}\sum_{ijk\ell}\widetilde{X}_{ij}\widetilde{X}_{k\ell}A_{ik}A_{j\ell}+\widetilde{\xi}_n\right)$ is independent of a and b and from Lemma 18, we also know that it is uniformly integrable. On the other hand, since $\left|\widetilde{X}_{ij}\right| \leq \sqrt{n}$, we see that $\exp\left(\sum_{u\in[r],i,j\in[q]}\frac{ns_{a_ub_uij}}{d}\cdot\frac{\widetilde{X}_{ij}}{\sqrt{n}}\right)$ is uniformly bounded; hence, the entire displayed expression above is uniformly integrable. Since $\widetilde{X}_{ij} \to \mathcal{N}\left(0,\pi_i\pi_j-(\pi_i\pi_j)^2\right)$, the displayed equation above converges to a finite quantity that is independent of a and b. This proves the statement of the proposition with $\mathbb{P}_n$ replaced by $\hat{\mathbb{P}}_n = (\mathbb{P}_n \mid \Omega_n)$. Noting that

$$TV\left(\mathbb{P}_n\left(G|\sigma_u = a_u \text{ for } u\in[r]\right), \hat{\mathbb{P}}_n\left(G|\sigma_u = a_u \text{ for } u\in[r]\right)\right) = o(1),\ \forall\ a$$

gives us the desired result. ■

As an easy consequence of Proposition 22, the posterior distribution of a single label is essentially unchanged if we know a bounded number of other labels:

Lemma 23 *Suppose $Q\left(\pi, A/\sqrt{2d}\right) < 1$. Then, for any set S such that $|S|$ is a constant, $u \notin S$, we have:*

$$\mathbb{E}\left(TV\left(\mathbb{P}_n\left(\sigma_u|G,\sigma_S\right),\pi\right)|\sigma_S\right) = o(1).$$

Proof

$$\begin{aligned}
\mathbb{E}\left(TV\left(\mathbb{P}_n\left(\sigma_u|G,\sigma_S\right),\pi\right)|\sigma_S\right) &= \sum_{\sigma_u}\mathbb{P}_n\left(\sigma_u\right)\sum_G\left|\frac{\mathbb{P}_n\left(G|\sigma_u,\sigma_S\right)}{\mathbb{P}_n\left(G|\sigma_S\right)}-1\right|\mathbb{P}_n\left(G|\sigma_S\right) \\
&= \sum_i \pi(i)TV\left(\mathbb{P}_n\left(G|\sigma_u=i,\sigma_S\right),\mathbb{P}_n\left(G|\sigma_S\right)\right) = o(1),
\end{aligned}$$

where the last step follows from Proposition 22. ■

Finally, we will show the non-detectability part of Theorem 7. By Markov's inequality, it is enough to show that $\lim_{n\to\infty}\mathbb{E}\left(\mathrm{overlap}(\mathcal{A}(G),\sigma)\right)=0$. We first bound $\mathbb{E}\left(\mathrm{overlap}(\mathcal{A}(G),\sigma)\right)$ as follows:

$$\begin{aligned}\mathbb{E}\left(\mathrm{overlap}(\sigma,\mathcal{A}(G))\right) &= \frac{1}{n}\mathbb{E}\left(\max_{\rho}\sum_{i=1}^{q}\left(N_{i\rho(i)}(\sigma,\mathcal{A}(G))-\frac{1}{n}N_i(\sigma)N_{\rho(i)}(\mathcal{A}(G))\right)\right)\\ &\le \frac{1}{n}\sum_{\rho}\mathbb{E}\left(\left|\sum_{i=1}^{q}\left(N_{i\rho(i)}(\sigma,\mathcal{A}(G))-\frac{1}{n}N_i(\sigma)N_{\rho(i)}(\mathcal{A}(G))\right)\right|\right). \qquad (35)\end{aligned}$$

We will now show that each of the terms in the above summation goes to zero. Without loss of generality, let ρ be the identity map. Fix $i\in[q]$ and consider the term $\mathbb{E}\left|\left(N_{ii}-\frac{1}{n}N_i(\sigma)N_i(\mathcal{A}(G))\right)\right|$ (for brevity, we suppress $\sigma,\mathcal{A}(G)$ in $N_{ii}(\sigma,\mathcal{A}(G))$). Using Jensen's inequality, it is sufficient to bound

$$\mathbb{E}\left(N_{ii}-\frac{1}{n}N_i(\sigma)N_i(\mathcal{A}(G))\right)^2=\mathbb{E}\left(N_{ii}^2-\frac{2}{n}N_{ii}N_i(\sigma)N_i(\mathcal{A}(G))+\frac{1}{n^2}N_i^2(\sigma)N_i^2(\mathcal{A}(G))\right). \qquad (36)$$

We will now calculate each of the above three terms.

$$\begin{aligned}\mathbb{E}N_{ii}^2=\mathbb{E}\left(\sum_u \mathbb{1}_{\{\sigma_u=i\}}\mathbb{1}_{\{\mathcal{A}(G)_u=i\}}\right)^2 &= \sum_{u,v}\mathbb{E}\mathbb{1}_{\{\sigma_u=i\}}\mathbb{1}_{\{\mathcal{A}(G)_u=i\}}\mathbb{1}_{\{\sigma_v=i\}}\mathbb{1}_{\{\mathcal{A}(G)_v=i\}}\\ &= \sum_{u,v}\mathbb{E}\mathbb{1}_{\{\sigma_u=i\}}\mathbb{1}_{\{\mathcal{A}(G)_u=i\}}\mathbb{1}_{\{\sigma_v=i\}}\mathbb{1}_{\{\mathcal{A}(G)_v=i\}}\\ &= \sum_{u,v}\mathbb{E}\left(\mathbb{E}\left(\mathbb{1}_{\{\sigma_u=i\}}\mathbb{1}_{\{\mathcal{A}(G)_u=i\}}\mathbb{1}_{\{\sigma_v=i\}}\mathbb{1}_{\{\mathcal{A}(G)_v=i\}}\middle|G\right)\right)\\ &= \sum_{u,v}\mathbb{E}\left(\mathbb{E}\left(\mathbb{1}_{\{\sigma_u=i\}}\mathbb{1}_{\{\sigma_v=i\}}\middle|G\right)\mathbb{1}_{\{\mathcal{A}(G)_u=i\}}\mathbb{1}_{\{\mathcal{A}(G)_v=i\}}\right)\\ &= \left(\pi(i)^2\mathbb{E}\left(\mathbb{1}_{\{\mathcal{A}(G)_u=i\}}\mathbb{1}_{\{\mathcal{A}(G)_v=i\}}\right)+o(1)\right)n^2, \qquad (37)\end{aligned}$$

where the last step follows from Lemma 23. Coming to the second term, we have:

$$\begin{aligned}\mathbb{E}N_{ii}N_i(\sigma)N_i(\mathcal{A}(G)) &= \mathbb{E}\left(\sum_u \mathbb{1}_{\{\sigma_u=i\}}\mathbb{1}_{\{\mathcal{A}(G)_u=i\}}\right)\left(\sum_u \mathbb{1}_{\{\sigma_u=i\}}\right)\left(\sum_u \mathbb{1}_{\{\mathcal{A}(G)_u=i\}}\right)\\ &= \sum_{u,v,w}\mathbb{E}\left(\mathbb{E}\left(\mathbb{1}_{\{\sigma_u=i\}}\mathbb{1}_{\{\mathcal{A}(G)_u=i\}}\mathbb{1}_{\{\sigma_v=i\}}\mathbb{1}_{\{\mathcal{A}(G)_w=i\}}\middle|G\right)\right)\\ &= \sum_{u,v,w}\mathbb{E}\left(\mathbb{E}\left(\mathbb{1}_{\{\sigma_u=i\}}\mathbb{1}_{\{\sigma_v=i\}}\middle|G\right)\mathbb{1}_{\{\mathcal{A}(G)_u=i\}}\mathbb{1}_{\{\mathcal{A}(G)_w=i\}}\right)\\ &= \left(\pi(i)^2\mathbb{E}\left(\mathbb{1}_{\{\mathcal{A}(G)_u=i\}}\mathbb{1}_{\{\mathcal{A}(G)_v=i\}}\right)+o(1)\right)n^3, \qquad (38)\end{aligned}$$

where the last step again follows from Lemma 23. A similar argument shows that

$$\mathbb{E}N_i^2(\sigma)N_i^2(\mathcal{A}(G))=\left(\pi(i)^2\mathbb{E}\left(\mathbb{1}_{\{\mathcal{A}(G)_u=i\}}\mathbb{1}_{\{\mathcal{A}(G)_v=i\}}\right)+o(1)\right)n^4. \qquad (39)$$

Plugging (37), (38) and (39) in (36) shows that

$$\mathbb{E}\left(N_{ii} - \frac{1}{n}N_i(\sigma)N_i(\mathcal{A}(G))\right)^2 = o(n^2).$$

This finishes the proof.

References

E. Abbe and C. Sandon. Detection in the stochastic block model with multiple clusters: proof of the achievability conjectures, acyclic BP, and the information-computation gap. *ArXiv e-prints*, 2015.

E. Abbe and C. Sandon. Crossing the KS threshold in the stochastic block model with information theory. *Proc. International Symposium on Information Theory (ISIT)*, 2016.

E. Abbe, A.S. Bandeira, and G. Hall. Exact recovery in the stochastic block model. *IEEE Transactions on Information Theory*, 62(1):471–487, 2016.

Emmanuel Abbe and Colin Sandon. Community detection in general stochastic block models: Fundamental limits and efficient algorithms for recovery. In *IEEE 56th Annual Symposium on Foundations of Computer Science, FOCS*, pages 670–688, 2015. doi: 10.1109/FOCS.2015.47. URL http://dx.doi.org/10.1109/FOCS.2015.47.

Dimitris Achlioptas and Amin Coja-Oghlan. Algorithmic barriers from phase transitions. In *49th Annual IEEE Symposium on Foundations of Computer Science, FOCS*, pages 793–802, 2008.

Dimitris Achlioptas and Assaf Naor. The two possible values of the chromatic number of a random graph. *Annals of Mathematics*, 162:1335–1351, 2005.

Naman Agarwal, Afonso S. Bandeira, Konstantinos Koiliaris, and Alexandra Kolla. Multisection in the stochastic block model using semidefinite programming. *CoRR*, abs/1507.02323, 2015. URL http://arxiv.org/abs/1507.02323.

Victor Bapst, Amin Coja-Oghlan, Samuel Hetterich, Felicia Raßmann, and Dan Vilenchik. The condensation phase transition in random graph coloring. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques, APPROX/RANDOM*, pages 449–464, 2014.

Victor Bapst, Amin Coja-Oghlan, and Felicia Raßmann. A positive temperature phase transition in random hypergraph 2-coloring. *Annals of Applied Probability*, abs/1410.2190, to appear. URL http://arxiv.org/abs/1410.2190.

Peter J. Bickel and Aiyou Chen. A nonparametric view of network models and Newman–Girvan and other modularities. *Proc. Natl. Acad. Sci. USA*, 106:21068–21073, 2009.

Béla Bollobás, Svante Janson, and Oliver Riordan. The phase transition in inhomogeneous random graphs. *Random Structures and Algorithms*, 31:3–122, 2007.

Béla Bollobás, Svante Janson, and Oliver Riordan. The phase transition in inhomogeneous random graphs. *Random Structures & Algorithms*, 31(1):3–122, 2007.

Charles Bordenave, Marc Lelarge, and Laurent Massoulié. Non-backtracking spectrum of random graphs: Community detection and non-regular ramanujan graphs. In *IEEE 56th Annual Symposium on Foundations of Computer Science, FOCS*, pages 1347–1357, 2015.

Amin Coja-Oghlan and Charilaos Efthymiou. On independent sets in random graphs. *Random Struct. Algorithms*, 47(3):436–486, 2015.

Amin Coja-Oghlan and Dan Vilenchik. Chasing the k-colorability threshold. In *54th Annual IEEE Symposium on Foundations of Computer Science, FOCS*, pages 380–389, 2013.

J. R. L. de Almeida and D. J. Thouless. Stability of the Sherrington-Kirkpatrick solution of a spin-glass model. *J. Phys. A*, 11:983–990, 1978.

Aurelien Decelle, Florent Krzakala, Cristopher Moore, and Lenka Zdeborová. Inference and phase transitions in the detection of modules in sparse networks. *Physical Review Letters*, 107:065701, 2011a.

Aurelien Decelle, Florent Krzakala, Cristopher Moore, and Lenka Zdeborová. Asymptotic analysis of the stochastic block model for modular networks and its algorithmic applications. *Physical Review E*, 84:066106, 2011b.

W. Evans, C. Kenyon, Y. Peres, and L.J. Schulman. Broadcasting on trees and the Ising model. *The Annals of Applied Probability*, 10(2):410–433, 2000.

P. W. Holland, K. B. Laskey, and S. Leinhardt. Stochastic blockmodels: Some first steps. *Social Networks*, 5:109–137, 1983.

Svante Janson. Random regular graphs: asymptotic distributions and contiguity. *Combinatorics, Probability and Computing*, 4(04):369–405, 1995.

Svante Janson and Elchanan Mossel. Robust reconstruction on trees is determined by the second eigenvalue. *Annals of Probability*, pages 2630–2649, 2004.

Varun Kanade, Elchanan Mossel, and Tselil Schramm. Global and local information in clustering labeled block models. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques, APPROX/RANDOM*, pages 779–792, 2014.

Harry Kesten and Bernt P. Stigum. A limit theorem for multidimensional Galton-Watson processes. *The Annals of Mathematical Statistics*, 37(5):1211–1223, 1966a.

Harry Kesten and Bernt P. Stigum. Additional limit theorems for indecomposable multidimensional Galton-Watson processes. *The Annals of Mathematical Statistics*, 37(6):1463–1481, 1966b.

F. Krzakala, A. Montanari, F. Ricci-Tersenghi, G. Semerjian, and L. Zdeborová. Gibbs states and the set of solutions of random constraint satisfaction problems. *Proc. Natl. Acad. Sci. USA*, 104: 10318, 2007.

Florent Krzakala, Cristopher Moore, Elchanan Mossel, Joe Neeman, Allan Sly, Lenka Zdeborová, and Pan Zhang. Spectral redemption in clustering sparse networks. *Proc. Natl. Acad. Sci. USA*, 110:20935–20940, 2013.

A.M. Mathai and Serge B. Provost. *Quadratic Forms in Random Variables*. Statistics Series. Taylor & Francis, 1992. ISBN 9780824786915. URL http://books.google.com/books?id=tFOqQgAACAAJ.

Frank McSherry. Spectral partitioning of random graphs. In *42nd Annual Symposium on Foundations of Computer Science, FOCS*, pages 529–537, 2001. doi: 10.1109/SFCS.2001.959929. URL http://dx.doi.org/10.1109/SFCS.2001.959929.

E. Mossel, J. Neeman, and A. Sly. Stochastic Block Models and Reconstruction. *ArXiv e-prints*, 2012.

E. Mossel, J. Neeman, and A. Sly. Stochastic block models and reconstruction. *Probability Theory and Related Fields*, 2014a. (to appear).

Elchanan Mossel, Joe Neeman, and Allan Sly. Belief propagation, robust reconstruction and optimal recovery of block models. In *Proceedings of The 27th Conference on Learning Theory, COLT 2014, Barcelona, Spain, June 13-15, 2014*, pages 356–370, 2014b. URL http://jmlr.org/proceedings/papers/v35/mossel14.html.

R.W. Robinson and N.C. Wormald. Almost all cubic graphs are Hamiltonian. *Random Structures and Algorithms*, 3(2):117–125, 1992.

Bo Söderberg. General formalism for inhomogeneous random graphs. *Physical Review E*, 66: 066121, 2002.

N.C. Wormald. Models of random regular graphs. *London Mathematical Society Lecture Note Series*, pages 239–298, 1999.

Pan Zhang and Cristopher Moore. Scalable detection of statistically significant communities and hierarchies, using message passing for modularity. *Proceedings of the National Academy of Sciences*, 111(51):18144–18149, 2014. doi: 10.1073/pnas.1409770111. URL http://www.pnas.org/content/111/51/18144.abstract.

Pan Zhang, Cristopher Moore, and M. E. J. Newman. Community detection in networks with unequal groups. *Phys. Rev. E*, 93:012303, 2016.

Appendix A. UI and multinomials

Here, we restate and prove Proposition 17. Recall that Δ_q denotes the set $\{(\alpha_1, \dots, \alpha_q) : \alpha_i \geq 0 \text{ and } \sum_i \alpha_i = 1\}$, and that $\Delta_{q^2}(\pi)$ denotes the set of $(\alpha_{11} \dots, \alpha_{qq})$ such that

$$\begin{aligned} \alpha_{ij} &\geq 0 \text{ for all } i, j, \\ \sum_{i=1}^q \alpha_{ij} &= \pi_j \text{ for all } j, \text{ and} \\ \sum_{j=1}^q \alpha_{ij} &= \pi_i \text{ for all } i. \end{aligned}$$

In what follows, we fix an $q^2 \times q^2$ matrix A and some $\pi \in \Delta_q$. We define $p \in \Delta_{q^2}(\pi)$ by $p_{ij} = \pi_i \pi_j$ (or alternatively, $p = \pi^{\otimes 2}$), and we take $N \sim \text{Multinom}(n, p)$ and $X = (N - np)/\sqrt{n}$. Finally, fix a sequence a_n such that $\sqrt{n} \ll a_n \ll n$ and define Ω_n to be the event that

$$\max_j \left| \sum_i N_{ij} - n\pi_j \right| \leq a_n \tag{40}$$

$$\max_i \left| \sum_j N_{ij} - n\pi_i \right| \leq a_n. \tag{41}$$

Note that the condition $\sqrt{n} \ll a_n$ ensures that the probability of Ω_n converges to 1.

Proposition 24 *Define*

$$\lambda = \sup_{\alpha \in \Delta_{q^2}(\pi)} \frac{(\alpha - p)^T A (\alpha - p)}{D(\alpha, p)}.$$

If $\lambda < 1$ *then*

$$\mathbb{E}[\mathbb{1}_{\Omega_n} \exp(X^T A X)] \to \mathbb{E} \exp(Z^T A Z) < \infty,$$

as $n \to \infty$*, where* $Z \sim \mathcal{N}(0, \text{diag}(p) - pp^T)$*. On the other hand, if* $\lambda > 1$ *then*

$$\mathbb{E}[\mathbb{1}_{\Omega_n} \exp(X^T A X)] \to \infty$$

as $n \to \infty$*.*

Lemma 25 *For any* $\epsilon > 0$*, any* $q = 2, 3, \dots$*, and any* $p \in \Delta_q$*, there is a constant* $C < \infty$ *such that for any* n*,*

$$n^{-q/2} \sum_{r_1 + \dots + r_q = n} \exp\left(-n\epsilon \left| \frac{r}{n} - p \right|^2 \right) \leq C.$$

Proof We have

$$\begin{aligned} n^{-q/2} \sum_{r_1 + \dots + r_q = n} \exp\left(-n\epsilon \left| \frac{r}{n} - p \right|^2 \right) &\leq n^{-q/2} \sum_{r_1, \dots, r_q = 1}^{n} \exp\left(-n\epsilon \left| \frac{r}{n} - p \right|^2 \right) \\ &= \prod_{i=1}^q \left[n^{-1/2} \sum_{r=1}^n \exp\left(-n\epsilon \left(\frac{r}{n} - p_i \right)^2 \right) \right]. \end{aligned}$$

The problem has now reduced to the case $q = 1$; i.e., we need to show that

$$n^{-1/2} \sum_{r=1}^{n} \exp(-n\epsilon(r/n - p)^2) < C(p, \epsilon).$$

We do this by dividing the sum above into $\ell = \lceil \sqrt{n} \rceil$ different sums. Note that if $\frac{r}{n} \geq p$ then

$$\left(\frac{r + \ell}{n} - p \right)^2 = \left(\frac{r}{n} - p \right)^2 + \frac{\ell^2}{n^2} + \frac{2\ell}{n} \left(\frac{r}{n} - p \right) \geq \left(\frac{r}{n} - p \right)^2 + \frac{1}{n}. \tag{42}$$

Hence, $r \geq np$ implies

$$\exp\left(-n\epsilon \left(\frac{r + \ell}{n} - p \right)^2 \right) \leq e^{-\epsilon} \exp\left(-n\epsilon \left(\frac{r}{n} - p \right)^2 \right).$$

Stratifying the original sum into strides of length ℓ,

$$n^{-1/2} \sum_{r=\lceil pn \rceil}^{n} \exp(-n\epsilon(r/n - p)^2) \leq n^{-1/2} \sum_{r=\lceil pn \rceil}^{\lceil pn \rceil + \ell - 1} \sum_{m=0}^{\infty} \exp(-n\epsilon((r + m\ell)/n - p)^2).$$

Now, (42) implies that the inner sum may be bounded by a geometric series with initial value less than 1, and ratio $e^{-\epsilon}$. Hence,

$$n^{-1/2} \sum_{r=\lceil pn \rceil}^{n} \exp(-n\epsilon(r/n - p)^2) \leq n^{-1/2} \ell \frac{1}{1 - e^{-\epsilon}},$$

which is bounded. A similar argument for the case $r \leq pn$ completes the proof. ■

Proof [Proof of Proposition 24] First, recall that for any $\alpha = (\alpha_{11}, \dots, \alpha_{qq}) \in \Delta_{q^2}$, we have $\Pr(N = \alpha n) \asymp \exp(-nD(\alpha, p))$; this just follows from Stirling's approximation. Next, note that $D(\alpha, p)$ is zero only for $\alpha = p$, and that $D(\alpha, p)$ is strongly concave in α. Therefore, $\lambda < 1$ implies that there is some $\epsilon > 0$ such that

$$D(\alpha, p) \geq (1 + \epsilon)(\alpha - p)^T A(\alpha - p) + \epsilon |\alpha - p|^2$$

for all $\alpha \in \Delta_{q^2}(p)$. Hence, any $\alpha \in \Delta_{q^2}(p)$ satisfies

$$\Pr(N = \alpha n) \exp(n(1 + \epsilon)(\alpha - p)^T A(\alpha - p)) \leq C \exp(-n\epsilon|\alpha - p|^2). \tag{43}$$

Recalling the definition of Ω_n, we write (with a slight abuse of notation) $\alpha \in \Omega_n$ if $|\max_i \sum_j \alpha_{ij} - p_i| \leq n^{-1} a_n$ and similarly with i and j reversed. Note that for every $\alpha \in \Omega_n$, there is some $\tilde{\alpha} \in \Delta_{q^2}(\pi)$ with $|\alpha - \tilde{\alpha}|^2 = o(n^{-1})$; in particular, (43) also holds for all $\alpha \in \Omega_n$ (with a change in the constant C). Then

$$\begin{aligned}
\mathbb{E}[\mathbb{1}_{\Omega_n} \exp((1 + \epsilon)X^T AX)] &= \sum_{\alpha \in \Omega_n} \Pr(N = n\alpha) \exp\left(n(1 + \epsilon)(\alpha - p)^T A(\alpha - p)\right) \\
&\leq \sum_{\alpha \in \Omega_n} \exp\left(-n\epsilon|\alpha - p|^2\right) \\
&\leq C < \infty,
\end{aligned}$$

for some constant C independent of n, where the last line follows from Lemma 25. In particular, $\exp(X^T AX)$ has $1+\epsilon$ uniformly bounded moments, and so it is uniformly integrable as $n \to \infty$. Since $X \xrightarrow{d} \mathcal{N}(0, \mathrm{diag}(p) - pp^T)$, it follows that $\mathbb{E}\exp(X^T AX) \to \mathbb{E}\exp(X^T AX)$.

In the other direction, if $\lambda > 1$ then there is some $\alpha \in \Delta_{q^2}(p)$, $\alpha \neq p$ and some $\epsilon > 0$ such that $D(\alpha, p) \leq (\alpha - p)^T A(\alpha - p) - 2\epsilon$. By the continuity of $D(\alpha, p)$ and $(\alpha - p)^T A(\alpha - p)$, we see that for sufficiently large n, there exists $r \in n\Delta_{q^2}(p)$ such that

$$D(r/n, p) \leq (r/n - p)^T A(r/n - p) - \epsilon.$$

For any n, let $r^* = r^*(n)$ be such an r. Then

$$\begin{aligned}
\mathbb{E}\exp(X^T AX) &\geq \Pr(N = r^*(n)) \exp\left(n(r^*/n - p)^T A(r^*/n - p)\right) \\
&\asymp \exp\left(n\left((r^*/n - p)^T A(r^*/n - p) - D(r^*/n, p)\right)\right) \\
&\geq \exp(n\epsilon) \to \infty.
\end{aligned}$$

■