Semidefinite Programs for Exact Recovery of a Hidden Community

Bruce Hajek   B-HAJEK@ILLINOIS.EDU
Department of ECE, University of Illinois at Urbana-Champaign, Urbana, IL

Yihong Wu   YIHONGWU@ILLINOIS.EDU
Department of ECE, University of Illinois at Urbana-Champaign, Urbana, IL

Jiaming Xu   JIAMINGXU@BERKELEY.EDU
Simons Institute for the Theory of Computing, University of California, Berkeley, Berkeley, CA

Abstract

We study a semidefinite programming (SDP) relaxation of the maximum likelihood estimation for exactly recovering a hidden community of cardinality K from an $n \times n$ symmetric data matrix A, where for distinct indices i, j, $A_{ij} \sim P$ if i, j are both in the community and $A_{ij} \sim Q$ otherwise, for two known probability distributions P and Q. We identify a sufficient condition and a necessary condition for the success of SDP for the general model. For both the Bernoulli case ($P = \text{Bern}(p)$ and $Q = \text{Bern}(q)$ with $p > q$) and the Gaussian case ($P = \mathcal{N}(\mu, 1)$ and $Q = \mathcal{N}(0, 1)$ with $\mu > 0$), which correspond to the problem of planted dense subgraph recovery and submatrix localization respectively, the general results lead to the following findings: (1) If $K = \omega(n/\log n)$, SDP attains the information-theoretic recovery limits with sharp constants; (2) If $K = \Theta(n/\log n)$, SDP is order-wise optimal, but strictly suboptimal by a constant factor; (3) If $K = o(n/\log n)$ and $K \to \infty$, SDP is order-wise suboptimal. The same critical scaling for K is found to hold, up to constant factors, for the performance of SDP on the stochastic block model of n vertices partitioned into multiple communities of equal size K. A key ingredient in the proof of the necessary condition is a construction of a primal feasible solution based on random perturbation of the true cluster matrix.

Keywords: Semidefinite programming relaxations; Planted dense subgraph recovery; Submatrix localization; Stochastic block model

1. Introduction

1.1. Motivation and problem setup

Consider the stochastic block model (SBM) Holland et al. (1983) with a single community, where out of n vertices a community consisting of K vertices are chosen uniformly at random; two vertices are connected by an edge with probability p if they both belong to the community and with probability q if either one of them is not in the community. The goal is to recover the community based on observation of the graph, which, when $p > q$, is also known as the *planted dense subgraph* recovery problem McSherry (2001); Chen and Xu (2014); Hajek et al. (2015a); Montanari (2015).

In the special case of $p = 1$ and $q = 1/2$, planted dense subgraph recovery reduces to the widely-studied planted clique problem, i.e., finding a hidden clique of size K in the Erdős-Rényi random graph $\mathcal{G}(n, 1/2)$. It is well-known that the maximum likelihood estimator (MLE), which is computationally intractable, finds any clique of size $K \geq 2(1 + \epsilon) \log_2 n$ for any constant $\epsilon > 0$; however, existing polynomial-time algorithms, including spectral methods Alon et al. (1998), message passing Deshpande and Montanari (2015a), and semi-definite programming (SDP) relaxation of

MLE Feige and Krauthgamer (2000), are only known to find a clique of size $K \geq \epsilon\sqrt{n}$. In fact, impossibility results for the more powerful s-round Lovász-Schrijver relaxations and, more recently, degree-$2r$ sums-of-squares (SOS) relaxation (with $s = 1$ and $r = 1$ corresponding to SDP) have been recently obtained in Feige and Krauthgamer (2003) and Deshpande and Montanari (2015b); Meka et al. (2015); Hopkins et al. (2015); Raghavendra and Schramm (2015), showing that relaxations of constant rounds or degrees lead to order-wise suboptimality even for detecting the clique. In other words, for the planted clique problem there is a significant gap between the state of the art of polynomial-time algorithms and what is information-theoretically possible.

In sharp contrast, for sparser graphs and larger community size, SDP relaxations have been shown to achieve the information-theoretic recovery limit up to sharp constants. For $p = a \log n / n, q = b \log n / n$ and $K = \rho n$ for fixed constants $a, b > 0$ and $0 < \rho < 1$, the recent work Hajek et al. (2016) identified a sharp threshold $\rho^* = \rho^*(a, b)$ such that if $\rho > \rho^*$, an SDP relaxation of MLE recovers the hidden community with high probability; if $\rho < \rho^*$, exact recovery is information theoretically impossible. This optimality result of SDP has been extended to multiple communities as long as their sizes scale linearly with the graph size n Bandeira (April, 2015); Hajek et al. (2015b); Agarwal et al. (2015); Perry and Wein (2015); Montanari and Sen (April, 2015).

The dichotomy between the optimality of SDP up to sharp constants in the relatively sparse regime and the order-wise suboptimality of SDP in the dense regime prompts us to investigate the following question:

> *When do SDP relaxations cease to be optimal for planted dense subgraph recovery*?

In this paper, we address this question under the more general hidden community model considered in Deshpande and Montanari (2015a).

Definition 1 (Hidden Community Model) *Let C^* be drawn uniformly at random from all subsets of $[n]$ of cardinality K. Given probability measures P and Q on a common measurable space $\mathcal{X}$, let A be an $n \times n$ symmetric matrix with zero diagonal where for all $1 \leq i < j \leq n$, A_{ij} are mutually independent, and $A_{ij} \sim P$ if $i, j \in C^*$ and $A_{ij} \sim Q$ otherwise.*

The distributions P and Q as well as the community size K vary with the matrix size n in general. In this paper we assume that these model parameters are known to the estimator, and focus on exact recovery of the hidden community based on the data matrix A, namely, constructing an estimator $\widehat{C} = \widehat{C}(A)$, such that as $n \to \infty$, $\mathbb{P}\{\widehat{C} \neq C^*\} \to 0$ uniformly in the choice of the true cluster C^*.

We are particularly interested in the following choices of P and Q:

- Bernoulli case: $P = \text{Bern}(p)$ and $Q = \text{Bern}(q)$ with $0 \leq q < p \leq 1$. In this case, the data matrix A corresponds to the adjacency matrix of a graph, and the problem reduces to planted dense subgraph recovery.

- Gaussian case: $P = \mathcal{N}(\mu, 1)$ and $Q = \mathcal{N}(0, 1)$ with $\mu > 0$. In this case, the submatrix of A with row and column indices in C^* has a positive mean μ except on the diagonal, while the rest of A has zero mean, and the problem corresponds to a symmetric version of the *submatrix localization* problem studied in Shabalin et al. (2009); Kolar et al. (2011); Butucea et al. (2015); Ma and Wu (2015); Chen and Xu (2014); Cai et al. (2015).

1.2. Main results

We show that for both planted dense subgraph recovery and submatrix localization, SDP relaxations of MLE achieve the information-theoretic optimal threshold if and only if the hidden community size satisfies $K = \omega(\frac{n}{\log n})$. More specifically,

- $K = \omega(\frac{n}{\log n})$, SDP attains the information-theoretic recovery limits with sharp constants. This extends the previous result in Hajek et al. (2016) obtained for $K = \Theta(n)$ and the Bernoulli case.

- $K = \Theta(\frac{n}{\log n})$, SDP is order-wise optimal, but strictly suboptimal by a constant factor;

- $K = o(\frac{n}{\log n})$ and $K \to \infty$, SDP is order-wise suboptimal.

To establish our main results, we derive a sufficient condition and a necessary condition under which the optimal solution to SDP is unique and coincides with the true cluster matrix. In particular, for planted dense subgraph recovery, whenever SDP does not achieve the information-theoretic threshold, our sufficient condition and necessary condition are within constant factors of each other; for submatrix localization, we characterize the minimal signal-to-noise ratio required by SDP within a factor of four when $K = \omega(\sqrt{n})$. The sufficiency proof is similar to those in Hajek et al. (2016) based on the dual certificate argument; we extend the construction and validation of dual certificates for the success of SDP to the general distributions P, Q. The necessity proof is via constructing a high-probability feasible solution to the SDP by means of random perturbation of the ground truth that leads to a higher objective value. One could instead adapt the existing constructions in the SOS literature for planted clique Deshpande and Montanari (2015b); Meka et al. (2015); Hopkins et al. (2015); Raghavendra and Schramm (2015) to our setting, but it falls short of establishing the impossibility of SDP to attain the optimal recovery threshold in the critical regime of $K = \Theta(n/\log n)$; see Remark 5 for details.

An alternative approach to establish impossibility results for SDP, thanks to strong duality that holds for the specific program, is to prove the non-existence of dual certificates, which turns out to yield the same condition given by the aforementioned explicit construction of primal solutions. The dual-based method has been previously used for proving necessary conditions for related nuclear-norm constrained optimization problems, see e.g., Kolar et al. (2011); Vinayak et al. (2014); Chen and Xu (2014); however, the constants in the derived conditions are often loose or unspecified. In comparison, we aim to obtain necessary conditions for SDP relaxations with explicit constants. Another difference is that the specific SDP considered here is more complicated involving the stringent positive semi-definite constraint and a set of equality and non-negativity constraints.

Using similar techniques, we obtain analogous results for SDP relaxation for SBM with logarithmically many communities. Specifically, consider the network of $n = rK$ vertices partitioned into r communities of cardinality K each, with edge probability p for pairs of vertices within communities and q for other pairs of vertices. Then SDP relaxation, in contrast to the MLE, is constantwise suboptimal if $r \geq C \log n$ for sufficiently large C, and orderwise suboptimal if $r = \omega(\log n)$. That is, it is constantwise suboptimal if $K \leq \frac{cn}{\log n}$ for sufficiently small c, and orderwise suboptimal if $K = o\left(\frac{n}{\log n}\right)$. This result complements the sharp optimality for SDP previously established in Hajek et al. (2015b) for $r = O(1)$ and extended to $r = o(\log n)$ in Agarwal et al. (2015).

In closing, we comment on the barrier which prevents SDP from being optimal. It is known that, see e.g., Chen and Xu (2014); Montanari and Sen (April, 2015), spectral methods which estimate the

communities based on the leading eigenvector of the data matrix A suffer from a spectral barrier: the spectrum of the "signal part" $\mathbb{E}[A]$ must escape that of the "noise part" $A - \mathbb{E}[A]$, i.e., the smallest nonzero singular value of $\mathbb{E}[A]$ needs to be much larger than the spectral norm $\|A - \mathbb{E}[A]\|$. Closely related to the spectral barrier, the SDP barrier originates from a key random quantity (see (6)), which is at most and, in fact, possibly much smaller than, the largest eigenvalue of $A - \mathbb{E}[A]$. Thus we expect the SDP barrier to be weaker than the spectral one. Indeed, for the submatrix localization problem, if the submatrix size is sufficiently small, i.e., $K = o(\sqrt{n/\log n})$, SDP recovers the community with high probability if $\mu = \Omega(\sqrt{\log n})$, while the spectral barrier requires a much stronger signal: $\mu = \Omega(\sqrt{n}/K)$; see Section 4.1 for details.

1.3. Notation

Let $\mathbf{I}$ and $\mathbf{J}$ denote the identity matrix and all-one matrix, respectively. For a matrix X we write $X \succeq 0$ if X is positive semidefinite, and $X \geq 0$ if X is non-negative entrywise. Let $\mathcal{S}^n$ denote the set of all $n \times n$ symmetric matrices. For $X \in \mathcal{S}^n$, let $\lambda_2(X)$ denote its second smallest eigenvalue. For an $m \times n$ matrix M, let $\|M\|$ and $\|M\|_{\mathrm{F}}$ denote its spectral and Frobenius norm, respectively. For any $S \subset [m], T \subset [n]$, let $M_{ST} \in \mathbb{R}^{S \times T}$ denote $(M_{ij})_{i \in S, j \in T}$ and for $m = n$ abbreviate $M_S = M_{SS}$. For a vector x, let $\|x\|$ denote its Euclidean norm. We use standard big O notations as well as their counterparts in probability, e.g., for any sequences $\{a_n\}$ and $\{b_n\}$, $a_n = \Theta(b_n)$ or $a_n \asymp b_n$ if there is an absolute constant $c > 0$ such that $1/c \leq a_n/b_n \leq c$. All logarithms are natural and we adopt the convention $0 \log 0 = 0$. Let $\mathrm{Bern}(p)$ denote the Bernoulli distribution with mean p and $\mathrm{Binom}(n, p)$ denote the binomial distribution with n trials and success probability p. Let $d(p\|q) = p \log \frac{p}{q} + (1 - p) \log \frac{1-p}{1-q}$ denote the Kullback-Leibler (KL) divergence between $\mathrm{Bern}(p)$ and $\mathrm{Bern}(q)$. We say a sequence of events $\mathcal{E}_n$ holds with high probability, if $\mathbb{P}\{\mathcal{E}_n\} \to 1$ as $n \to \infty$.

2. Semidefinite programming relaxations

Recall that $\xi^* \in \{0, 1\}^n$ denotes the indicator of the underlying cluster C^*, such that $\xi_i^* = 1$ if $i \in C^*$ and $\xi_i^* = 0$ otherwise. Let L denote an $n \times n$ symmetric matrix such that $L_{ij} = f(A_{ij})$ for $i \neq j$ and $L_{ii} = 0$, where $f : \mathcal{X} \to \mathbb{R}$ is any function possibly depending on the model parameters. Consider the following combinatorial optimization problem:

$$\widehat{\xi} = \arg\max_{\xi} \sum_{i,j} L_{ij} \xi_i \xi_j$$
$$\text{s.t.} \ \ \xi \in \{0, 1\}^n \tag{1}$$
$$\xi^\top \mathbf{1} = K,$$

which maximizes the sum of entries among all $K \times K$ principal submatrices of L.

If L is the log likelihood ratio (LLR) matrix with $f(A_{ij}) = \log \frac{dP}{dQ}(A_{ij})$ for $i \neq j$ and $L_{ii} = 0$, then $\widehat{\xi}$ is precisely the MLE of ξ^*. In general, evaluating the MLE requires knowledge of K and the distributions P, Q. Computing the MLE is NP hard in the worst case for general values of n and K since certifying the existence of a clique of a specified size in an undirected graph, which is known to be NP complete Karp (1972), can be reduced to computation of the MLE. This intractability of the MLE prompts us to consider its semidefinite programming relaxation as studied in Hajek et al.

4

(2016). Note that (1) can be equivalently[1] formulated as

$$
\max_{Z} \ \langle L, Z \rangle
$$
$$
\begin{aligned}
\text{s.t.} \ \ &\text{rank}(Z) = 1, \\
&Z_{ii} \leq 1, \quad \forall i \in [n] \\
&Z_{ij} \geq 0, \quad \forall i, j \in [n] \\
&\langle \mathbf{I}, Z \rangle = K \\
&\langle \mathbf{J}, Z \rangle = K^2.
\end{aligned} \tag{2}
$$

Replacing the rank-one constraint by the positive semidefinite constraint leads to the following convex relaxation of (2), which can be cast as a semidefinite program:[2]

$$
\widehat{Z}_{\text{SDP}} = \arg\max_{Z} \ \langle L, Z \rangle
$$
$$
\begin{aligned}
\text{s.t.} \ \ &Z \succeq 0 \\
&Z_{ii} \leq 1, \quad \forall i \in [n] \\
&Z \geq 0 \\
&\langle \mathbf{I}, Z \rangle = K \\
&\langle \mathbf{J}, Z \rangle = K^2.
\end{aligned} \tag{3}
$$

Let $\xi^* \in \{0, 1\}^n$ denote the indicator of the community such that $\text{supp}(\xi^*) = C^*$. Let $Z^* = \xi^*(\xi^*)^\top$ denote the cluster matrix corresponding to C^*. It is straightforward to retrieve the underlying cluster C^* from Z^*. Thus, if $\mathbb{P}\{\widehat{Z}_{\text{SDP}} = Z^*\} \to 1$ as $n \to \infty$, then exact recovery of C^* is attained. Note that by the symmetry of the SDP formulation and the distribution of L, the probability of success $\mathbb{P}\{\widehat{Z}_{\text{SDP}} = Z^*\}$ is the same conditioned on any realization of ξ^* and hence the worst-case probability of error coincides with the average-case one.

Recall that if L is the LLR matrix, then the solution $\widehat{\xi}$ to (1) is precisely the MLE of ξ^*. In the Gaussian case, $\log \frac{dP}{dQ}(A_{ij}) = \mu(A_{ij} - \mu/2)$ with $\mu > 0$ for $i \neq j$; in the Bernoulli case, $\log \frac{dP}{dQ}(A_{ij}) = \log \frac{p(1-q)}{q(1-p)} A_{ij} + \log \frac{1-p}{1-q}$ with $p > q$ for $i \neq j$. Thus, in both cases, (3) with $L = A$ corresponds to a semidefinite programming relaxation of the MLE, and the only model parameter needed for evaluating (3) is the cluster size K.

3. Analysis of SDP in the general model

In this section, we give a sufficient condition and a necessary condition, both deterministic, for the success of SDP (3) for exact recovery. Define

$$
e(i, C^*) = \sum_{j \in C^*} L_{ij}, \quad i \in [n] \tag{4}
$$

---

1. Here (1) and (2) are equivalent in the following sense: For any feasible ξ for (1), $Z = \xi\xi^\top$ is feasible for (2); Any feasible Z for (2) can be written as $Z = \xi\xi^\top$ such that either ξ or $-\xi$ is feasible for (1).
2. $\widehat{Z}_{\text{SDP}}$ as $\arg\max$ denotes the set of maximizers of the optimization problem (3). If Z^* is the unique maximizer, we write $\widehat{Z}_{\text{SDP}} = Z^*$.

and

$$\alpha = \mathbb{E}_P[L_{12}],$$
$$\beta = \mathbb{E}_Q[L_{12}].$$

We assume that $\alpha \geq \beta$, i.e., L has an elevated mean in the submatrix supported on $C^* \times C^*$ (excluding the diagonal). This assumption guarantees that Z^* is the optimal solution to (3) when L is replaced by its mean $\mathbb{E}[L]$, and is clearly satisfied when L is the LLR matrix, in which case $\alpha = D(P\|Q) \geq 0 \geq -D(Q\|P) = \beta$, or $L = A$ in the Gaussian and Bernoulli cases.

Theorem 2 (Sufficient condition for SDP: general case) *If*

$$\min_{i \in C^*} e(i, C^*) - \max \left\{ \max_{i \notin C^*} e(i, C^*), K\beta \right\} > \|L - \mathbb{E}[L]\| - \beta, \tag{5}$$

then $\widehat{Z}_{\mathrm{SDP}} = Z^*$.

The sufficient condition of Theorem 2 is derived via the dual certificate argument. That is, we give an explicit construction of dual variables which together with Z^* are shown to satisfy the KKT conditions under the condition (5).

The necessary condition relies on the following key quantity which is the value of an auxiliary SDP program. Let $m = n - K$ and $M = L_{(C^*)^c \times (C^*)^c}$ denote the submatrix of L outside the community. Then M is an $m \times m$ symmetric matrix with zero diagonal, where $\{M_{ij} : 1 \leq i < j \leq m\}$ are i.i.d. For $a \in \mathbb{R}$, consider the value (random variable) of the following SDP:

$$V_m(a) \triangleq \max_Z \langle M, Z \rangle \tag{6}$$
$$\text{s.t. } Z \succeq 0$$
$$Z \geq 0$$
$$\mathsf{Tr}(Z) = 1$$
$$\langle \mathbf{J}, Z \rangle = a.$$

There is no feasible solution to (6) unless $1 \leq a \leq m$, so by convention, let $V_m(a) = -\infty$ if $a < 1$ or $a > m$. Dropping the second and the last constraints in (6), yields $V_m(a) \leq \lambda_{\max}(M)$. Also, $V_m(1) = 0$, $V_m(m) = \langle M, \mathbf{J} \rangle / m$, and $a \mapsto V_m(a)$ is concave on $[1, m]$. Clearly, the distributions of M as well as $V_m(a)$ depend on the distribution Q but not P.

Fix K, n, C^*, the matrix L, and $a \in [1, K]$. Also, let $r = \frac{a}{K}$. For ease of notation suppose the indices are permuted so that $C^* = [K]$, index K minimizes $e(i, C^*)$ over all $i \in C^*$, and index $K+1$ maximizes $e(j, C^*)$ over all $j \notin C^*$. Let U be an $n \times n$ matrix corresponding to the solution of the SDP defining $V_m(a)$ with $M = L_{(C^*)^c \times (C^*)^c}$ in (6). That is, U is a symmetric $n \times n$ matrix with $U_{ij} = 0$ if $(i,j) \notin (C^*)^c \times (C^*)^c$, $V_m(a) = \langle L, U \rangle$, $U \succeq 0$, $U \geq 0$, $\mathsf{Tr}(U) = 1$, and $\langle \mathbf{J}, U \rangle = a = Kr$.

Next we give intuition about the construction of primal feasible solutions via random perturbation that lead to a necessary condition for SDP. Three positive semidefinite perturbations of Z^*, namely $Z^* + \delta_i$ for $1 \leq i \leq 3$, can be defined for $0 < \epsilon < 1/2$ by letting (dashed lines delineate the $K \times K$

submatrices and only nonzero entries are shown):

$$
\delta_1 = \begin{bmatrix} & & & -\epsilon & \vdots \\ & & & -\epsilon & \\ -\epsilon & \cdots & -\epsilon & -2\epsilon + \epsilon^2 & \\ \hline & & & & \\ & & & & \end{bmatrix} \tag{7}
$$

$$
\delta_2 = (1-r) \begin{bmatrix} & & & & \epsilon \\ & & & & \vdots \\ & & & & \epsilon \\ \hline \epsilon & \cdots & \epsilon & \epsilon^2 \\ & & & & \end{bmatrix} \tag{8}
$$

$$
\delta_3 = 2\epsilon U \tag{9}
$$

It turns out that $Z^* + \delta_1 + \delta_2 + \delta_3$ is close to being positive semidefinite for small ϵ. Also,

$$
\begin{array}{lll}
\langle \mathbf{I}, \delta_1 \rangle = -2\epsilon + o(\epsilon) & \langle \mathbf{J}, \delta_1 \rangle = -2K\epsilon + o(\epsilon) & \langle L, \delta_1 \rangle = -2\epsilon \min_{i \in C^*} e(i, C^*) \\
\langle \mathbf{I}, \delta_2 \rangle = o(\epsilon) & \langle \mathbf{J}, \delta_2 \rangle = 2(K-a)\epsilon + o(\epsilon) & \langle L, \delta_2 \rangle = 2(1-r)\epsilon \max_{j \notin C^*} e(j, C^*) \\
\langle \mathbf{I}, \delta_3 \rangle = 2\epsilon & \langle \mathbf{J}, \delta_3 \rangle = 2a\epsilon & \langle L, \delta_3 \rangle = 2\epsilon V_{n-K}(a)
\end{array}
$$

Therefore, up to $o(\epsilon)$ terms, $Z^* + \delta_1 + \delta_2 + \delta_3$ satisfies the two equality constraints of the SDP (3) and is near a feasible solution of the SDP (3), suggesting that a necessary condition for the optimality of Z^* is $\langle L, \delta_1 + \delta_2 + \delta_3 \rangle \leq 0$. Note that

$$
\langle L, \delta_1 + \delta_2 + \delta_3 \rangle = 2\epsilon \left((1-r) \max_{j \notin C^*} e(j, C^*) + V_{n-K}(a) - \min_{i \in C^*} e(i, C^*) \right) + o(\epsilon).
$$

Hence the term inside the parenthesis must be non-positive. This leads the following deterministic necessary condition for SDP. The proof, given in Section 8.2, is a minor variation of the heuristic argument just presented.

Theorem 3 (Necessary condition for SDP: general case) *If $Z^* \in \widehat{Z}_{\mathrm{SDP}}$, then*

$$
\min_{i \in C^*} e(i, C^*) - \max_{j \notin C^*} e(j, C^*) \geq \sup_{1 \leq a \leq K} \left\{ V_{n-K}(a) - \frac{a}{K} \max_{j \notin C^*} e(j, C^*) \right\}. \tag{10}
$$

Remark 4 *Note that* (10) *is equivalent to*

$$
\min_{i \in C^*} e(i, C^*) \geq \sup_{1 \leq a \leq K} \left\{ V_{n-K}(a) + \left(1 - \frac{a}{K}\right) \max_{i \notin C^*} e(i, C^*) \right\},
$$

Setting $a = K$ in (10) *yields the weaker necessary condition:* $\min_{i \in C^*} e(i, C^*) \geq V_{n-K}(K)$.

Remark 5 *The problem formulation as well as proof technique of Theorem 3 differ from existing results on the planted clique problem for the sum of squares (SoS) hierarchy Deshpande and Montanari (2015b); Meka et al. (2015); Hopkins et al. (2015); Raghavendra and Schramm (2015) in an essential way. Aside from the fact that those papers consider more powerful convex relaxations, they address the clique detection problem (which do have implications for clique estimation), which can be viewed as testing the null hypothesis H_0 : clique absent versus the alternative H_1 : clique present, using the value of the SOS program as the test statistic. The approach of these papers involves only the null hypothesis, showing that a feasible solution to SOS program can be constructed based on the $\mathcal{G}(n, 1/2)$ graph whose objective value is much larger than the size of the largest clique in $\mathcal{G}(n, 1/2)$, leading to a large integrality gap. This further induces a high false-positive error probability if the size of the planted clique K is small. In comparison, since we are dealing with recovery as opposed to detection using SDP, the impossibility result in Theorem 3 follows from the fact that, if the true cluster matrix Z^* is an optimal solution, then certain random perturbations of Z^* must not lead to a strictly larger objective value. More precisely, the perturbation argument involves three directions (7)-(9). Note that the matrix U in (9) is the maximizer of (6) and can be constructed using similar techniques in the SoS literature. However, this perturbation alone is not enough to separate the performance of SDP from MLE in the critical regime $K = \Theta(n/\log n)$, and it is necessary to exploit the other perturbations terms (7)-(8) that depend on the true cluster matrix.*

Remark 6 *Since Slater's condition and hence strong duality holds for the SDP (3), the fulfillment of the KKT conditions is necessary for Z^* to be a maximizer. We provide an alternative proof of Theorem 3 in Section 8.2, showing that (10) is necessary for the existence of dual variables to satisfy the KKT conditions together with Z^*.*

By comparing Theorem 2 and Theorem 3, we find that both the sufficient and necessary conditions are in terms of the separation between $\min_{i \in C^*} e(i, C^*)$ and $\max_{j \notin C^*} e(j, C^*)$. In comparison, for the optimal estimator, MLE, to succeed in exact recovery, it is necessary that $\min_{i \in C^*} e(i, C^*) \geq \max_{j \notin C^*} e(j, C^*)$; otherwise, one can form a candidate community C by swapping the node i in C^* achieving the minimum $e(i, C^*)$ with the node j not in C^* achieving the maximum $e(j, C^*)$, so that the new community C has a likelihood at least as large as that of C^*.

Capitalizing on Theorems 2 and 3, we will derive explicit sufficient and necessary results for the success of SDP in the Gaussian and Bernoulli cases. Interestingly, in both cases, if $K = \omega(n/\log n)$, the sufficient condition of SDP coincides in the leading terms with the information-theoretic necessary condition for $\min_{i \in C^*} e(i, C^*) \geq \max_{j \notin C^*} e(j, C^*)$, thus resulting in the optimality of SDP with the sharp constants.

4. Submatrix localization

In this section we consider the submatrix localization problem corresponding to the Gaussian case of Definition 1. The SDP relaxation of MLE is given by (3) with $L = A$.

Theorem 7 (Sufficient conditions for SDP: Gaussian case) *Assume that $K \geq 2$ and $n - K \asymp n$. Let $\epsilon > 0$ be an arbitrary constant. If either $K \to \infty$ and*

$$\mu(1 - \epsilon) \geq \frac{1}{\sqrt{K}} \left(\sqrt{2 \log K} + \sqrt{2 \log(n - K)} \right) + \frac{2\sqrt{n}}{K}, \tag{11}$$

or

$$\mu(1 - \epsilon) \geq 2\sqrt{\log K} + 2\sqrt{\log n}, \tag{12}$$

then $\mathbb{P}\{\widehat{Z}_{\mathrm{SDP}} = Z^*\} \to 1$ *as* $n \to \infty$.

Remark 8 *To deduce (11) from the general sufficient condition (5), we first show that*

$$\min_{i \in C^*} e(i, C^*) \geq (K - 1)\mu - \sqrt{2(K - 1)\log K} + o_P(\sqrt{K}) \tag{13}$$

$$\max_{j \notin C^*} e(j, C^*) \leq \sqrt{2K \log(n - K)} + o_P(\sqrt{K}). \tag{14}$$

Then (11) follows since $A - \mathbb{E}[A]$ *is an* $n \times n$ *Wigner matrix whose spectral norm is* $(2 + o_P(1))\sqrt{n}$. *Under the condition (12), with high probability,* $\min_{(i,j) \in C^* \times C^*: i \neq j} A_{ij} > \max_{(i,j) \notin C^* \times C^*} A_{ij}$, *and thus* $\widehat{Z}_{\mathrm{SDP}} = Z^*$. *In this case, the community can be also trivially recovered with probability tending to one using entrywise hard thresholding, and not surprisingly, but SDP as well.*

Next we present a converse result for the exact recovery performance of SDP in a strong sense:

Theorem 9 (Necessary condition for SDP in Gaussian case) *Assume that* $L = A$ *in the SDP (3),* $K \to \infty$, *and* $K = o(n)$. *Suppose that* $\liminf_{n \to \infty} \mathbb{P}\{Z^* \in \widehat{Z}_{\mathrm{SDP}}\} > 0$. *Then for any fixed* $\epsilon > 0$,

- *if* $K = \omega(\sqrt{n})$, *then*

$$\mu(1 + \epsilon) \geq \frac{1}{\sqrt{K}}\left(\sqrt{2\log K} + \sqrt{2\log(n - K)}\right) + \frac{\sqrt{n}}{2K}. \tag{15}$$

- *if* $K = \Theta(\sqrt{n})$, *then*

$$\mu(1 + \epsilon) \geq \sqrt{\log\left(1 + \frac{n}{4K^2}\right)}. \tag{16}$$

- *if* $K = o(\sqrt{n})$, *then*

$$\mu(1 + \epsilon) \geq \sqrt{\frac{1}{3}\log\frac{n}{K^2}}. \tag{17}$$

Remark 10 *To deduce Theorem 9 from the general necessary condition given in Theorem 3, we first show that the inequalities in (13) and (14) are in fact equalities. Then, we prove a high-probability lower bound to* $V_m(a)$ *and choose* $a = o(K)$ *in (10) when* $K = \omega(\sqrt{n})$, *and* $a = K$ *when* $K = O(\sqrt{n})$.

By comparing sufficient condition (11) and necessary condition (15), we can see that the sufficient condition and necessary condition are within a factor of 4 *in the case of* $K = \omega(\sqrt{n})$.

4.1. Comparison to the information-theoretic limits

It is instructive to compare the performance of the SDP to the information-theoretic fundamental limits. We focus on the most interesting regime of $K \to \infty$ and $n - K \asymp n$. It has been shown (cf. (Hajek et al., 2015e, Theorem 4)) that, for any $\epsilon > 0$, the MLE (which minimizes the probability of error) achieves exact recovery if

$$\mu(1 - \epsilon) \geq \frac{1}{\sqrt{K}} \max \left\{ \sqrt{2 \log K} + \sqrt{2 \log n}, 2\sqrt{\log(n/K)} \right\}; \tag{18}$$

conversely, if

$$\mu(1 + \epsilon) \leq \frac{1}{\sqrt{K}} \max \left\{ \sqrt{2 \log K} + \sqrt{2 \log n}, 2\sqrt{\log(n/K)} \right\}, \tag{19}$$

no estimator can exactly recover the community with high probability.

Comparing (18) – (19) with (11), (12), and (15)– (17), we arrive at the following conclusion on the performance of the SDP relaxation:

- $K = \omega(n/\log n)$: Since $\sqrt{n} = o(\sqrt{K \log n})$, in this regime SDP attains the information-theoretically optimal recovery threshold with sharp constant.

- $K = \Theta(n/\log n)$: SDP is order-wise optimal but strictly suboptimal in terms of constants. More precisely, consider the critical regime of

$$K = \frac{\rho n}{\log n}, \quad \mu = \frac{\mu_0 \log n}{\sqrt{n}} \tag{20}$$

  for fixed constants $\rho, \mu_0 > 0$. Then MLE succeeds (resp. fails) if $\rho\mu_0^2 >$ (resp. $<$) 8. If $\rho\mu_0 > 2\sqrt{2\rho}+2$, then SDP succeeds; conversely, if SDP succeeds, then $\rho\mu_0 \geq 2\sqrt{2\rho}+1/2$. Moreover, it is shown in Hajek et al. (2015c) that a message passing algorithm plus clean-up succeeds if $\rho\mu_0^2 > 8$ and $\rho\mu_0 > 1/\sqrt{e}$, while a linear message passing algorithm corresponds to a spectral method succeeds if $\rho\mu_0^2 > 8$ and $\rho\mu_0 > 1$. Therefore, SDP is strictly suboptimal comparing to MLE, message passing, and linear message passing for $\rho > 0$, $\rho > (1/\sqrt{e} - 1/2)^2/8$, and $\rho > 1/32$, respectively. See Fig. 1 for an illustration.

- $\omega(1) \leq K = o(n/\log n)$: Comparing to MLE, SDP is order-wise suboptimal. Moreover, when $K \leq n^{1/2-\delta}$ for any fixed constant $\delta > 0$, $\mu = \Omega(\sqrt{\log n})$ is necessary for SDP to achieve exact recovery, while the entrywise hard-thresholding or simply picking the largest entries attains exact recovery when $\mu(1 - \epsilon) \geq 2\sqrt{\log n}$. Thus in this regime, the more sophisticated SDP procedure is only possible to outperform the trivial thresholding algorithm by a constant factor. Similar phenomena has been observed in the bi-clustering problem Kolar et al. (2011), which is an asymmetric version of the submatrix localization problem, and the sparse PCA Krauthgamer et al. (2015).

- $K = \Theta(1)$: In this case the sufficient condition of SDP is within a constant factor of the information limit. For the extreme case of $K = 2$, SDP achieves the information limit with optimal constant, namely, $\mu(1 - \epsilon) \geq 2\sqrt{\log n}$; however, in this case exact recovery can be trivially achieved by entrywise hard-thresholding or simply picking the largest entries.
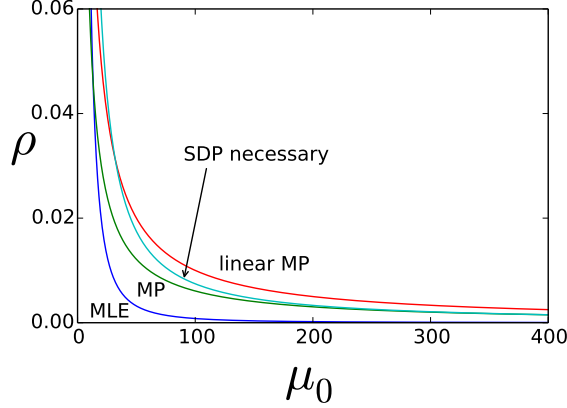
Figure 1: Phase diagram for the Gaussian model with $K = \rho n / \log n$ and $\mu = \mu_0 \log n / \sqrt{n}$. The curve MLE: $\rho \mu_0^2 = 1$ denotes the information-theoretic threshold for exact recovery. The threshold for optimized message passing (MP): $\rho^2 \mu^2 e = 1$, and linear message passing: $\rho^2 \mu^2 = 1$, parallel each other. The curve SDP necessary: $\rho \mu_0 \geq 2\sqrt{2\rho} + 1/2$ is a lower bound below which SDP does not provide exact recovery. The sufficient curve for SDP, above which SDP provides exact recovery, is not shown, and lies above the four curves shown.

5. Planted densest subgraph

In this section, we turn to the planted densest subgraph problem corresponding to the Bernoulli case of Definition 1, where $P = \mathrm{Bern}(p)$ and $Q = \mathrm{Bern}(q)$ with $0 \leq q < p \leq 1$. We prove both positive and negative results for the SDP relaxation of the MLE, i.e., (3) with $L = A$ being the adjacency matrix of the graph, to exactly recover the community C^*.

The following assumption on the community size and graph sparsity will be imposed:

Assumption 1 *As $n \to \infty$, $K \to \infty$, $n - K \asymp n$, q is bounded away from* 1*, and $nq = \Omega(\log n)$.*

Our SDP results are in terms of the following quantities:[3]

$$
\begin{aligned}
\tau_1 &= \text{solution of } Kd(\tau\|p) = \log K \text{ in } \tau \in (0, p) \\
\tau_2 &= \text{solution of } Kd(\tau\|q) = \log(n - K) \text{ in } \tau \in (q, 1)
\end{aligned}
\tag{21}
$$

Theorem 11 (Sufficient conditions for SDP: Bernoulli case) *Suppose that Assumption* 1 *holds. If*

$$
K(\tau_1 - \tau_2) \geq \kappa \left(\sqrt{nq(1-q)} + \sqrt{Kp(1-p)} \right),
\tag{22}
$$

where

$$
\kappa = \begin{cases}
O(1) & nq = \Omega(\log n) \\
4 + o(1) & nq = \omega(\log n) \\
2 + o(1) & nq = \omega(\log^4 n)
\end{cases},
\tag{23}
$$

then $\mathbb{P}\{\widehat{Z}_{\mathrm{SDP}} = Z^*\} \to 1$ as $n \to \infty$.

---

3. It can be shown that τ_1 and τ_2 are well-defined whenever exact recovery is information-theoretically possible; see Lemma 35.

Remark 12 *To deduce sufficient condition (22) from the general result (5), we first show that with high probability,*

$$\min_{i \in C^*} e(i, C^*) \geq (K-1)\tau_1 \tag{24}$$

$$\max_{j \notin C^*} e(j, C^*) \leq K\tau_2. \tag{25}$$

Then, we prove that with high probability,

$$\|A - \mathbb{E}[A]\| \leq \kappa \left(\sqrt{nq(1-q)} + \sqrt{Kp(1-p)} \right).$$

Note that $\|A - \mathbb{E}[A]\|$ behaves roughly the same as $\|A' - \mathbb{E}[A']\|$, where A' is the adjacency matrix of $\mathcal{G}(n, q)$. In light of the concentration results for Wigner matrices and the fact that $\|A' - \mathbb{E}[A']\| = \omega_P(\sqrt{nq})$ whenever $nq = o(\log n)$ (cf. (Hajek et al., 2016, Appendix A)), it is reasonable to expect that $\|A' - \mathbb{E}[A']\| = \sqrt{nq}(2 + o_P(1))$ whenever the average degree satisfies $nq = \Omega(\log n)$; however, this still remains an open problem cf. Le and Vershynin (2015) and the best known upper bounds depends on the scaling of nq. This explains the piecewise expression of κ in (23).

Theorem 13 (Necessary conditions for SDP: Bernoulli case) *Suppose that Assumption 1 holds, and $K = o(n)$. If $\liminf_{n \to \infty} \mathbb{P}\{Z^* = \widehat{Z}_{\mathrm{SDP}}\} > 0$, then*

$$K \geq \frac{1}{\kappa}\sqrt{\frac{nq}{1-q}} + 1, \tag{26}$$

$$K(\tau_1 - \tau_2) \geq \frac{1}{\kappa}\sqrt{\frac{nq}{1-q}}(1-\tau_2) - 6\sqrt{\frac{Kp}{\log K}} - \frac{K(p-q)(2\log\log K + 1)}{\log K}, \tag{27}$$

where κ is defined in (23).

Remark 14 *We prove (26) by contradiction: assuming (26) is violated, we construct explicitly a high-probability feasible solution Z to (3) based on the optimal solution of SDP defining $V_{n-K}(K)$ given in (6), and show that $\langle A, Z \rangle = \langle A, Z^* \rangle$, contracting the unique optimality of Z^*. Notice that in the special case of $p = 1$ (planted clique), Z^* is always a maximizer of the SDP (3) therefore the failure of SDP amounts to multiple maximizers.*

To deduce the necessary condition (27) from Theorem 3, we first establish some inequalities similar to (24) and (25) but in the reverse direction. Then, we prove a high-probability lower bound to $V_m(a)$ and choose $a = \frac{1}{\kappa}\sqrt{\frac{nq}{1-q}} + 1$.

Remark 15 *Particularizing Theorem 11 and Theorem 13 to the planted clique problem ($p = 1$ and $q = 1/2$), we conclude that: for any fixed $\epsilon > 0$, if $K \geq 2(1+\epsilon)\sqrt{n}$, then SDP succeeds (namely, Z^* is the unique optimal solution to (3)) with high probability; conversely, if $K \leq (1-\epsilon)\sqrt{n}/2$, SDP fails with high probability. In comparison, a message passing algorithm plus clean-up is shown in Deshpande and Montanari (2015a) to succeed if $K > (1+\epsilon)\sqrt{n/e}$.*

Assume that $\log \frac{p(1-q)}{q(1-p)}$ is bounded. If $\frac{K(p-q)}{\sqrt{nq}} = O(1)$, then the sufficient condition of SDP given in Theorem 11 reduces to $\frac{K(\tau_1 - \tau_2)}{\sqrt{nq}} \geq \Omega(1)$, while the necessary condition of SDP given in

Theorem 13 reduces to $\frac{K(\tau_1 - \tau_2)}{\sqrt{nq}} \geq \Omega(1)$. Thus, the sufficient and necessary conditions are within constant factors of each other. If instead $\frac{K(p-q)}{\sqrt{nq}} = \omega(1)$, then SDP attains the information-theoretic recovery threshold with sharp constants, as shown in the next subsection.

5.1. Comparison to the information-theoretic limits

In this section, we compare the performance limits of SDP with the information-theoretic limits of exact recovery obtained in Hajek et al. (2015e) under the assumption that $\log \frac{p(1-q)}{q(1-p)}$ is bounded and K/n is bounded away from 1. Let

$$\tau^* \triangleq \frac{\log \frac{1-q}{1-p} + \frac{1}{K} \log \frac{n}{K}}{\log \frac{p(1-q)}{q(1-p)}}. \tag{28}$$

It is shown in (Hajek et al., 2015e, Theorem 3) that, the optimal estimator, MLE, achieves exact recovery if

$$\liminf_{n \to \infty} \frac{K d(\tau^* \| q)}{\log n} > 1, \text{ and } \liminf_{n \to \infty} \frac{K d(p \| q)}{\log(n/K)} > 2. \tag{29}$$

Conversely, if

$$\limsup_{n \to \infty} \frac{K d(\tau^* \| q)}{\log n} < 1, \text{ or } \limsup_{n \to \infty} \frac{K d(p \| q)}{\log(n/K)} < 2, \tag{30}$$

no estimator can exactly recover the community with high probability.

Next we compare the SDP conditions (Theorems 11 and 13) to the information limit (29)–(30). Without loss of generality, we can assume the MLE necessary conditions holds. Our results on the performance limits of SDP lead to the following observations:

- $K = \omega(n/\log n)$. In this case, (29) implies (22) and thus SDP attains the information-theoretic recovery threshold with sharp constants. To see this, note that Lemma 35 shows that $\tau_1 \geq (1-\epsilon)\tau^* + \epsilon p$ and $\tau_2 \leq (1-\epsilon)\tau^* + \epsilon q$ for some small constant $\epsilon > 0$. Moreover, Lemma 32 and Lemma 34 imply that

$$\frac{K d(\tau^* \| q)}{\log n} \asymp \frac{K(p-q)^2}{q \log n} = \left[\frac{n}{K \log n} \right] \frac{K^2 (p-q)^2}{nq}. \tag{31}$$

Therefore, if $K = \omega(n/\log n)$, (29) implies that $Kq = \Omega(\log n)$ and $K(p-q)/\sqrt{nq} \to \infty$, and consequently

$$K(\tau_1 - \tau_2) \geq \epsilon K(p-q) = \omega(\sqrt{nq}),$$

which in turn implies condition (22). This result recovers the previous result in the special case of $K = \rho n$, $p = a \log n/n$, and $q = b \log n/n$ with fixed constants ρ, a, b, where SDP has been shown to attain the information-theoretic recovery threshold with sharp constants Hajek et al. (2016).

- $K = o(n/\log n)$. In this case, condition (27) together with $q \leq \tau_2 \leq p$ and $\tau_1 \leq p$ implies that $K(p-q)/\sqrt{nq} = \Omega(1)$. In comparison, in view of (31), $K(p-q)/\sqrt{nq} = \omega(\sqrt{K \log n/n})$ is sufficient for the information-theoretic sufficient condition (29) to hold. Hence, in this regime, SDP is order-wise suboptimal.

The above observations imply that a gap between the performance limit of SDP and information-theoretic limit emerges at $K = \Theta(n/\log n)$. To elaborate on this, consider the following regime:

$$K = \frac{\rho n}{\log n}, \quad p = \frac{a \log^2 n}{n}, \quad q = \frac{b \log^2 n}{n}, \tag{32}$$

where $\rho > 0$ and $a > b > 0$ are fixed constants. Let $I(x, y) \triangleq x - y \log(ex/y)$ for $x, y > 0$. Let γ_1 satisfy $\gamma_1 < a$ and $\rho I(a, \gamma_1) = 1$ and γ_2 satisfy $\gamma_2 > b$ and $\rho I(b, \gamma_2) = 1$. The following corollary follows from the performance limit of MLE given by (29)-(30) and that of SDP given by (22)-(27).

Corollary 16 *Assume the scaling (32).*

- *If $\gamma_1 > \gamma_2$, then MLE attains exact recovery; conversely, if MLE attains exact recovery, then $\gamma_1 \geq \gamma_2$.*

- *If $\rho(\gamma_1 - \gamma_2) > 4\sqrt{b}$, then SDP attains exact recovery; conversely, if SDP attains exact recovery, then $\rho(\gamma_1 - \gamma_2) \geq \sqrt{b}/4$.*

The proof is deferred to Appendix D. The above corollary implies that in the regime of (32), SDP is order-wise optimal, but strictly suboptimal by a constant factor. In comparison, as shown in Hajek et al. (2015d), belief propagation plus clean-up succeeds if $\gamma_1 > \gamma_2$ and $\rho(a - b) > \sqrt{b/e}$, while a linear message-passing algorithm corresponding to spectral method succeeds if $\gamma_1 > \gamma_2$ and $\rho(a - b) > \sqrt{b}$.

6. Stochastic block model with $\Omega(\log n)$ communities

In this section, we consider the stochastic block model with $r \geq 2$ communities of size K in a network of $n = rK$ nodes. Derived in Hajek et al. (2015b); Agarwal et al. (2015); Perry and Wein (2015), the following SDP is a natural convex relaxation of MLE:[4]

$$\widehat{Y}_{\text{SDP}} = \underset{Y \in \mathbb{R}^{n \times n}}{\arg \max} \ \langle A, Y \rangle$$
$$\text{s.t.} \ \ Y \succeq 0,$$
$$Y_{ii} = 1, \quad i \in [n]$$
$$Y_{ij} \geq -\frac{1}{r-1}, \quad i, j \in [n]$$
$$\langle Y, \mathbf{J} \rangle = 0. \tag{33}$$

Define the $n \times n$ symmetric matrix Y^* corresponding to the true clusters by $Y^*_{ij} = 1$ if vertices i and j are in the same cluster, including the case $i = j$, and $Y^*_{ij} = -\frac{1}{r-1}$ otherwise.

Consider $p = \frac{\alpha \log n}{K}$ and $q = \frac{\beta \log n}{K}$ for fixed constants $\alpha > \beta > 0$. For constant number of communities, namely $r = O(1)$, the sharp optimality of SDP has been established in Hajek et al. (2015b): if $\sqrt{\alpha} - \sqrt{\beta} > 1$, $\widehat{Y}_{\text{SDP}} = Y^*$ with high probability; conversely, if $\sqrt{\alpha} - \sqrt{\beta} < 1$ and the clusters are uniformly chosen at random among all r-equal-sized partitions of $[n]$, then for any sequence of estimators $\widehat{Y}_n$, $\mathbb{P}\{\widehat{Y}_n = Y^*\} \to 0$ as $n \to \infty$. The optimality of SDP has been extended

---

4. There are slightly different but equivalent ways to impose the constraints. Under the condition $Y \succeq 0$, the constraint $\langle Y, \mathbf{J} \rangle = 0$ is equivalent to $Y\mathbf{1} = 0$, which is the formulation used in Hajek et al. (2015b).

to $r = o(\log n)$ communities in Agarwal et al. (2015). Determining whether SDP continues to be optimal for $r = \Omega(\log n)$, or equivalently, for communities of size $K = O(\frac{n}{\log n})$, is left as an open question in Agarwal et al. (2015). Next, we settle this question by proving that in contrast to the MLE, SDP is constantwise suboptimal when $r \geq C \log n$ for sufficiently large C, and orderwise suboptimal when $r \gg \log n$. What remains open is to assert the suboptimality of SDP for *all* $r = \Theta(\log n)$ similar to the single-community case.

Theorem 17 *Suppose $p = o(1)$, $q = \Theta(p)$, and $r \to \infty$. If $\liminf_{n \to \infty} \mathbb{P}\{\widehat{Y}_{\mathrm{SDP}} = Y^*\} > 0$, then*

$$K(p-q)^2 \geq \frac{rq^2(1+o(1))}{p\kappa^2}, \tag{34}$$

where κ is the constant defined in (23).

Proof Section 8.2.3. ∎

Remark 18 *Under the assumption of $q = \Theta(p)$, the information-theoretic condition has been established in Chen and Xu (2014): MLE succeeds with high probability if and only if*

$$K(p-q)^2 \asymp q \log n. \tag{35}$$

Comparing (35) to the necessary condition (34) for SDP, we immediately conclude that SDP is orderwise suboptimal if $r = \omega(\log n)$, or equivalently, $K = o(\frac{n}{\log n})$. Furthermore, if $r \geq C \log n$ for a sufficiently large constant C, SDP is suboptimal in terms of constants, which is consistent with the single-community result in Section 1.2.

7. Discussions

In this paper, we derive a sufficient condition and a necessary condition for the success of an SDP relaxation (3) for exact recovery under the general P/Q model. For both the Gaussian and Bernoulli cases, the general results imply that the SDP attains the information-theoretic recovery limits with sharp constants if and only if $K = \omega(n/\log n)$. Loosely speaking, there are two types of perturbation which can lead to a higher objective value and prevent the true cluster matrix Z^* being the unique maximizer of the SDP. One is the *local* perturbation of the ground truth corresponding to swapping a node in the community with one outside. In order for exact recovery to be informationally possible, the optimal estimator, MLE, must also remain insensitive to this local perturbation. The other is the *global* perturbation induced by the solution of the auxiliary SDP (6). This global perturbation is closely related to the spectral perturbation, i.e., $\|A - \mathbb{E}[A]\|$, which is responsible for the suboptimality of the spectral algorithms. It turns out that when $K = \omega(n/\log n)$, the local perturbation dominates the global one, leading to the attainability of the optimal threshold by SDP; however, when $K = O(n/\log n)$, the local perturbation is dominated by the global one, resulting in the suboptimality of SDP.

An interesting future direction is to establish upper and lower bounds of SOS relaxations for the problem of finding a hidden community in relatively sparse SBM.

8. Proofs

In this section, we prove our main theorems. In particular, Section 8.1 contains the proofs of SDP sufficient conditions given in Theorem 2, Theorem 7, and Theorem 11. The proofs of SDP necessary conditions given in Theorem 3, Theorem 9, and Theorem 13 are presented in Section 8.2.

8.1. Sufficient Conditions

In this subsection, we provide the proof of Theorem 2, as well as the proofs of its further consequence in the Gaussian and Bernoulli cases.

Before the main proofs, we need a dual certificate lemma, providing a set of deterministic conditions which is both sufficient and necessary for the success of SDP (3).

Lemma 19 Z^* *is an optimal solution to* (3) *if and only if the following KKT conditions hold: there exist* $D = \mathsf{diag}\{d_i\} \geq 0$, $B \in \mathcal{S}^n$ *with* $B \geq 0$, $\lambda, \eta \in \mathbb{R}$ *such that* $S \triangleq D - B - L + \eta\mathbf{I} + \lambda\mathbf{J}$ *satisfies* $S \succeq 0$, *and*

$$S\xi^* = 0, \tag{36}$$

$$d_i(Z_{ii}^* - 1) = 0, \quad \forall i, \tag{37}$$

$$B_{ij}Z_{ij}^* = 0, \quad \forall i, j. \tag{38}$$

If further

$$\lambda_2(S) > 0, \tag{39}$$

or

$$\min_{i \in C^*} d_i > 0, \ and \ \min_{(i,j)\notin C^* \times C^*: i \neq j} B_{ij} > 0, \tag{40}$$

then Z^* *is the unique optimal solution to* (3).

Proof Notice that $Z = \frac{K(n-K)}{n(n-1)}\mathbf{I} + \frac{K(K-1)}{n(n-1)}\mathbf{J}$ is strictly feasible to (3), i.e., the Slater's condition holds, which implies, via Slater's theorem for SDP, that strong duality holds (see, e.g., (Boyd and Vandenberghe, 2004, Section 5.9.1)). Thus the KKT conditions given in (36)–(38) are both sufficient and necessary for the optimality of Z^*.

To show the uniqueness of Z^* under condition (39) or condition (40), consider another optimal solution $\widetilde{Z}$. Then,

$$\langle S, \widetilde{Z} \rangle = \langle D - B - L + \eta\mathbf{I} + \lambda\mathbf{J}, \widetilde{Z} \rangle \overset{(a)}{=} \langle D - B - L, \widetilde{Z} \rangle + \eta K + \lambda K^2$$
$$\overset{(b)}{\leq} \langle D - L, Z^* \rangle + \eta K + \lambda K^2 = \langle S, Z^* \rangle = 0.$$

where (a) holds because $\langle \mathbf{I}, \widetilde{Z} \rangle = K$ and $\langle \mathbf{J}, \widetilde{Z} \rangle = K^2$; (b) holds because $\langle L, \widetilde{Z} \rangle = \langle L, Z^* \rangle$, $B, \widetilde{Z} \geq 0$, and $\langle D, \widetilde{Z} \rangle \leq \sum_{i \in C^*} d_i = \langle D, Z^* \rangle$ in view of $d_i \geq 0$ and $\widetilde{Z}_{ii} \leq 1$ for all $i \in [n]$. It follows that the inequality (b) holds with equality, and thus $\langle D, \widetilde{Z} - Z^* \rangle = 0$ and $\langle B, \widetilde{Z} \rangle = 0$.

Suppose (39) holds. Since $\widetilde{Z} \succeq 0$, $S \succeq 0$, and $\langle S, \widetilde{Z} \rangle = 0$, $\widetilde{Z}$ needs to be a multiple of $Z^* = \xi^*(\xi^*)^\top$. Then $\widetilde{Z} = Z^*$ since $\mathsf{Tr}(\widetilde{Z}) = \mathsf{Tr}(Z^*) = K$.

16

Suppose instead (40) holds. Since $\langle B, \widetilde{Z} \rangle = 0$ and $B, \widetilde{Z} \geq 0$, it follows that $\widetilde{Z}_{ij} = 0$ for all $i \neq j$ such that $(i,j) \notin C^* \times C^*$. Also, in view of $\langle D, \widetilde{Z} - Z^* \rangle = 0$ and $\widetilde{Z}_{ii} \leq 1$, we have that $\widetilde{Z}_{ii} = 1$ for all $i \in C^*$. Hence, $\widetilde{Z}_{ii} = 0$ for all $i \notin C^*$ due to $\langle \mathbf{I}, \widetilde{Z} \rangle = K$. Finally, it follows from $\langle \mathbf{J}, \widetilde{Z} \rangle = K^2$ that $\widetilde{Z}_{ij} = 1$ for all $(i,j) \in C^* \times C^*$. Hence, we conclude that $\widetilde{Z} = Z^*$. $\blacksquare$

Proof [Proof of Theorem 2] We construct (λ, η, S, D, B) which satisfy the conditions in Lemma 19. Observe that to satisfy (36), (37), and (38), we need that $D = \text{diag}\{d_i\}$ with

$$d_i = \begin{cases} \sum_{j \in C^*} L_{ij} - \eta - \lambda K & \text{if } i \in C^* \\ 0 & \text{otherwise} \end{cases}, \tag{41}$$

and $B_{ij} = 0$ for $i, j \in C^*$, and

$$\sum_{j \in C^*} B_{ij} = \lambda K - \sum_{j \in C^*} L_{ij}, \quad \forall i \notin C^*, \tag{42}$$

where, given λ, η can be chosen without loss of generality to be:

$$\eta = \min_{i \in C^*} e(i, C^*) - \lambda K.$$

There remains flexibility in the choice of λ and the completion of the specification of B. Recall that $\alpha = \mathbb{E}_P[L_{12}]$ and $\beta = \mathbb{E}_Q[L_{12}]$. We let

$$\lambda = \max\left\{ \max_{i \notin C^*} e(i, C^*)/K, \beta \right\}$$

$$B_{ij} = b_i \mathbf{1}_{\{i \notin C^*, j \in C^*\}} + b_j \mathbf{1}_{\{i \in C^*, j \notin C^*\}},$$

where $b_i \triangleq \lambda - \frac{1}{K} \sum_{j \in C^*} L_{ij}$ for $i \notin C^*$. By definition, we have $d_i(Z_{ii}^* - 1) = 0$ and $B_{ij}Z_{ij}^* = 0$ for all $i, j \in [n]$. Moreover, for all $i \in C^*$,

$$d_i \xi_i^* = d_i = \sum_j L_{ij}\xi_j^* - \eta - \lambda K = \sum_j L_{ij}\xi_j^* + \sum_j B_{ij}\xi_j^* - \eta - \lambda K,$$

where the last equality holds because $B_{ij} = 0$ if $(i,j) \in C^* \times C^*$; for all $i \notin C^*$,

$$\sum_j L_{ij}\xi_j^* + \sum_j B_{ij}\xi_j^* - \lambda K = \sum_{j \in C^*} L_{ij} + Kb_i - \lambda K = 0,$$

where the last equality follows from our choice of b_i. Hence, $D\xi^* = L\xi^* + B\xi^* - \eta\xi^* - \lambda K \mathbf{1}$ and consequently $S\xi^* = 0$. Also, by definition, $\min_{i \in C^*} d_i \geq 0$ and $\min_{i \notin C^*} b_i \geq 0$, and thus $D \geq 0$, $B \geq 0$.

It remains to verify $S \succeq 0$ with $\lambda_2(S) > 0$, i.e.,

$$\inf_{x \perp \xi^*, \|x\|_2 = 1} x^\top S x > 0. \tag{43}$$

Since

$$\mathbb{E}[L] = (\alpha + \beta)Z^* + \beta \mathbf{J} - \alpha \begin{bmatrix} \mathbf{I}_{K \times K} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} - \beta \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{(n-K) \times (n-K)} \end{bmatrix},$$

17

it follows that for any $x \perp \xi^*$ and $\|x\|_2 = 1$,

$$
\begin{aligned}
x^\top S x &= x^\top D x - x^\top B x + (\lambda - \beta) x^\top \mathbf{J} x + \alpha \sum_{i \in C^*} x_i^2 + \beta \sum_{i \notin C^*} x_i^2 + \eta - x^\top (L - \mathbb{E}[L]) x \\
&\overset{(a)}{=} \sum_{i \in C^*} d_i x_i^2 + (\lambda - \beta) x^\top \mathbf{J} x + \alpha \sum_{i \in C^*} x_i^2 + \beta \sum_{i \notin C^*} x_i^2 + \eta - x^\top (L - \mathbb{E}[L]) x \\
&\overset{(b)}{\geq} \min_{i \in C^*} \sum_{i \in C^*} d_i x_i^2 + (\lambda - \beta) x^\top \mathbf{J} x + \beta + \eta - \|L - \mathbb{E}[L]\| \\
&\overset{(c)}{>} \left(\min_{i \in C^*} d_i \right) \sum_{i \in C^*} x_i^2 \geq 0.
\end{aligned}
\tag{44}
$$

where (a) holds because $B_{ij} = 0$ for all $(i, j) \in C^* \times C^*$ and

$$
x^\top B x = 2 \sum_{i \notin C^*} \sum_{j \in C^*} x_i x_j B_{ij} = 2 \sum_{i \notin C^*} x_i b_i \sum_{j \in C^*} x_j = 0;
$$

(b) follows due to the assumption that $\alpha \geq \beta$ and the fact that $x^\top (L - \mathbb{E}[L]) x \leq \|L - \mathbb{E}[L]\|$; (c) holds because by assumption $\eta > \|L - \mathbb{E}[L]\| - \beta$ and $\lambda \geq \beta$; the last inequality follows due to $\min_{i \in C^*} d_i \geq 0$. Therefore, the desired (43) holds in view of (44), completing the proof. ∎

8.1.1. GAUSSIAN CASE

We need the following standard result in extreme value theory (e.g., see (David and Nagaraja, 2003, Example 10.5.3) and use union bound).

Lemma 20 *Let $\{Z_i\}$ be a sequence of standard normal random variables. Then*

$$
\max_{i \in [m]} Z_i \leq \sqrt{2 \log m} + o_P(1), \quad m \to \infty,
$$

with equality if the random variables are independent.

Proof [Proof of Theorem 7] In the Gaussian case, $\mathbb{E}_P[A_{12}] = \mu$ and $\mathbb{E}_Q[A_{12}] = 0$. Hence, in view of Theorem 2, it suffices to show that with probability tending to one,

$$
\min_{i \in C^*} \sum_{j \in C^*} A_{ij} - \max \left\{ \max_{i \notin C^*} \sum_{j \in C^*} A_{ij}, 0 \right\} > \|A - \mathbb{E}[A]\|.
\tag{45}
$$

By Lemma 20,

$$
\max_{i \notin C^*} \sum_{j \in C^*} A_{ij} \leq \sqrt{2K \log(n - K)} + o_P(\sqrt{K}).
$$

Note that $\left\{ \sum_{j \in C^*} A_{ij} : i \in C^* \right\}$ are not mutually independent. By Lemma 20 applied to $-A_{ij}$,

$$
\min_{i \in C^*} \sum_{j \in C^*} A_{ij} \geq (K - 1)\mu - \sqrt{2(K - 1) \log K} + o_p(\sqrt{K}).
$$

By Lemma 25, for any sequence $t_n \to \infty$,

$$\|A - \mathbb{E}[A]\| \le 2\sqrt{n} + t_n$$

with probability converging to one. Hence, in view of the assumption (11), we have that (45) holds with high probability.

In the remainder, we prove (12) for any $K \ge 2$ implies that Z^* is the unique optimal solution of the SDP. We write $T = \{(i,j) \in C^* \times C^* : i \ne j\}$ and $T^c = \{(i,j) \in [n] \times [n] : i \ne j\} \backslash T$. Recall that for distinct i, j, $A_{ij} \sim \mathcal{N}(\mu, 1)$ if $i, j \in C^*$ and $\mathcal{N}(0, 1)$ otherwise. Using Lemma 20 and the assumption (12), we have

$$\min_{(i,j) \in T} A_{ij} > \max_{(i,j) \in T^c} A_{ij} \tag{46}$$

with probability converging to 1. Hence, without loss of generality, we can and do assume that (46) holds in the following. Let Z be any feasible solution of SDP (3). Since $Z_{ii} \le 1$ for all i and $Z \succeq 0$, it follows that $|Z_{ij}| \le 1$ for all i, j. Hence $0 \le Z \le \mathbf{J}$. Also, $\langle \mathbf{J} - \mathbf{I}, Z \rangle = K(K-1)$. So $\langle Z, A \rangle$ is a weighted sum of the terms $(A_{ij} : i \ne j)$, where the weights Z_{ij} are nonnegative, with values in [0,1], and total weight equal to $K(K-1)$. The sum is thus maximized if and only if all the weight is placed on the $K(K-1)$ largest terms, namely A_{ij} with $(i,j) \in T$, which are each strictly larger than the other terms. Thus, Z^* is the unique maximizer.

∎

8.1.2. BERNOULLI CASE

Proof [Proof of Theorem 11] In the Bernoulli case, $\mathbb{E}_P[A_{12}] = p$ and $\mathbb{E}_Q[A_{12}] = q$. Hence, in view of Theorem 2, it reduces to show that with probability tending to one,

$$\min_{i \in C^*} \sum_{j \in C^*} A_{ij} - \max\left\{\max_{i \notin C^*} \sum_{j \in C^*} A_{ij}, Kq\right\} > \|A - \mathbb{E}[A]\| - q. \tag{47}$$

We will use the following upper bounds for the binomial distribution tails (Zubkov and Serov, 2013, Theorem 1):

$$\mathbb{P}\{\mathrm{Binom}(m, p) \le m\tau - 1\} \le Q\left(\sqrt{2md(\tau\|p)}\right), \quad 2/m \le \tau \le p, \tag{48}$$

$$\mathbb{P}\{\mathrm{Binom}(m, q) \ge m\tau + 1\} \le Q\left(\sqrt{2md(\tau\|q)}\right), \quad q \le \tau \le 1 - 1/m, \tag{49}$$

where $Q(\cdot)$ denotes the standard normal tail probability. By the definition of τ_1 and τ_2, it follows that

$$\mathbb{P}\left\{\sum_{j \in C^*} A_{ij} \le (K-1)\tau_1 - 1\right\} \le Q(\sqrt{2(K-1)\log K/K}) = o(1/K), \quad \forall i \in C^*$$

$$\mathbb{P}\left\{\sum_{j \in C^*} A_{ij} \ge K\tau_2 + 1\right\} \le Q(\sqrt{2\log(n-K)}) = o(1/(n-K)), \quad \forall i \notin C^*.$$

By the union bound, with high probability,

$$\min_{i \in C^*} \sum_{j \in C^*} A_{ij} > (K-1)\tau_1 - 1$$

$$\max_{i \notin C^*} \sum_{j \in C^*} A_{ij} < K\tau_2 + 1.$$

We decompose $A = A_1 + A_2$, where A_1 is obtained from A by setting all entries not in $C^* \times C^*$ to be zero; similar, A_2 is obtained from A by setting all entries in $C^* \times C^*$ to be zero. Applying Lemma 30 yields that with high probability,

$$\|A - \mathbb{E}[A]\| \leq \|A_1 - \mathbb{E}[A_1]\| + \|A_2 - \mathbb{E}[A_2]\|$$
$$\leq \kappa \left(\sqrt{Kp(1-p)} + \sqrt{nq(1-q)} \right),$$

where κ is defined in (23). Hence, in view of the assumption (22), we have that (47) holds with high probability. ∎

8.2. Necessary conditions

Proof [Proof of Theorem 3] The proof is a slight variation of the heuristic derivation given before the statement of Theorem 3. Fix K, n, C^*, the matrix L, and a constant a with $1 \leq a \leq K$ and let $r = \frac{a}{K}$. Suppose the indices are ordered and the matrix U is defined as in the heuristic derivation.

Let Z be defined as a function of $\epsilon \geq 0$ as follows. We shall specify α and β depending on ϵ for sufficiently small ϵ in such a way that

$$\alpha \leq 1, \ \alpha = 1 + O(\epsilon^2), \ \beta \geq 1 - r, \ \beta = 1 - r + O(\epsilon). \tag{50}$$

Let ξ_ϵ be the column vector with $K+1$ nonzero entries, defined by $\xi_\epsilon = (1, \ldots, 1, 1-\epsilon, \beta\epsilon, 0, \ldots, 0)^\top$. Finally, let $Z = \alpha \xi_\epsilon \xi_\epsilon^T + 2\epsilon U$. In expanded form:

$$Z = \alpha \begin{bmatrix} 1 & \cdots & 1 & 1-\epsilon & \vdots & \beta\epsilon & \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \\ 1 & \cdots & 1 & 1-\epsilon & \vdots & \beta\epsilon & \\ 1-\epsilon & \cdots & 1-\epsilon & (1-\epsilon)^2 & \vdots & \beta\epsilon(1-\epsilon) & \\ \beta\epsilon & \cdots & \beta\epsilon & \beta\epsilon(1-\epsilon) & \vdots & \beta^2\epsilon^2 & \\ & & & & & & \\ & & & & & & \end{bmatrix} + 2\epsilon U$$

Up to $o(\epsilon)$ terms, Z is equal to the matrix $Z^* + \delta_1 + \delta_2 + \delta_3$ described in the heuristic derivation. Clearly for ϵ sufficiently small, $Z \geq 0$, $Z \succeq 0$, and $Z_{ii} \leq 1$. It is also straightforward to see that

$$\langle L, Z - Z^* \rangle = 2\epsilon \left((1-r) \max_{i \notin C^*} e(i, C^*) + V_{n-K}(a) - \min_{i \notin C^*} e(i, C^*) \right) + o(\epsilon),$$

20

so that once we establish the feasibility of Z, the proof will be complete. That is, it remains to show that α and β can be selected for sufficiently small ϵ so that (50), $\langle \mathbf{I}, Z \rangle = K$, and $\langle \mathbf{J}, Z \rangle = K^2$ hold true. The later two equations can be written as

$$\alpha \left\{ K - 2\epsilon + (1 + \beta^2)\epsilon^2 \right\} = K - 2\epsilon \tag{51}$$

$$\alpha \left\{ K - (1 - \beta)\epsilon \right\}^2 = K^2 - 2\epsilon K r . \tag{52}$$

Combining (51) and (52) to eliminate α and simplifying yields the following equation for β :

$$K^2(1 - \beta - r) + \epsilon K(\beta - 2(1 - \beta - r)) + \epsilon^2 \left((1 - \beta^2) - Kr(1 + \beta^2) \right) = 0$$

This equation has the form $F(\epsilon, \beta) = 0$ (K and r are fixed) with a solution at $(\epsilon, \beta) = (0, 1 - r)$. Also, $\frac{\partial F}{\partial \epsilon}(0, 1 - r) = K(1 - r)$ and $\frac{\partial F}{\partial \beta}(0, 1 - r) = -K^2 \neq 0$. Therefore, by the implicit function theorem, the equation determines β as a continuously differentiable function of ϵ for small enough epsilon, and

$$\beta = (1 - r)\left(1 + \frac{\epsilon}{K}\right) + O(\epsilon^2).$$

This expression for β together with (51) yields that for sufficiently small ϵ, $\alpha < 1$ and

$$\alpha = 1 - \epsilon^2 \left(\frac{1 + (1 - r)^2}{K}\right) + O(\epsilon^3).$$

∎

Proof [Alternative proof of Theorem 3] Here is an alternative proof of Theorem 3 via a dual-based approach. If $Z^* = \xi^*(\xi^*)^\top$ maximizes (3), then by Lemma 19 there exist dual variables (S, D, B, λ, η) with $S = D - B - L + \eta\mathbf{I} + \lambda\mathbf{J} \succeq 0$, $B \geq 0$, $D = \operatorname{diag}\{d_i\} \geq 0$, such that (36), (37) and (38) are satisfied. As a consequence, the choice of D is fixed, namely,

$$d_i = \begin{cases} \sum_{j \in C^*} L_{ij} - \eta - \lambda K & \text{if } i \in C^* \\ 0 & \text{otherwise} \end{cases} . \tag{53}$$

Therefore, the condition $\min_{i \in C^*} d_i \geq 0$ implies that

$$\min_{i \in C^*} e(i, C^*) \geq \lambda K + \eta. \tag{54}$$

Moreover, the dual variable B satisfies $B_{C^* C^*} = 0$ and the off-diagonal block $B_{(C^*)^c C^*}$ satisfies

$$\sum_{j \in C^*} B_{ij} = \lambda K - \sum_{j \in C^*} L_{ij}, \quad \forall i \notin C^*. \tag{55}$$

Denote all possible choices of B by the following convex set:

$$\mathcal{B} = \{B : B \in \mathcal{S}^n, B \geq 0, B_{C^* C^*} = 0, B_{(C^*)^c C^*} \text{ satisfies } (55)\}.$$

In particular, we have $\sum_{j \in C^*} B_{ij} \geq 0$ for all $i \notin C^*$, which implies that

$$\lambda K \geq \max_{i \notin C^*} e(i, C^*). \tag{56}$$

Finally, $S = D + \lambda \mathbf{J} - B - L + \eta \mathbf{I} \succeq 0$ and $S\xi^* = 0$ imply that there exists $B \in \mathcal{B}$ and η such that $\eta \geq \sup_{\|x\|=1} x^\top (B + L - D - \lambda \mathbf{J})x$ and (54) holds. Hence,

$$
\begin{aligned}
\eta &\geq \inf_{B \in \mathcal{B}} \sup_{\|x\|=1} x^\top (B + L - D - \lambda \mathbf{J})x \\
&= \inf_{B \in \mathcal{B}} \lambda_{\max} (B + L - D - \lambda \mathbf{J}) \\
&\geq \inf_{B \geq 0} \lambda_{\max} (B + L - D - \lambda \mathbf{J}) \\
&= \inf_{B \geq 0} \sup_{U \succeq 0, \langle U, \mathbf{I} \rangle = 1} \langle L - D - \lambda \mathbf{J} + B, U \rangle \\
&\stackrel{(a)}{=} \sup_{U \succeq 0, \langle U, \mathbf{I} \rangle = 1} \inf_{B \geq 0} \langle L - D - \lambda \mathbf{J} + B, U \rangle && (57) \\
&= \sup_{U \geq 0, U \succeq 0, \langle U, \mathbf{I} \rangle = 1} \langle L - D - \lambda \mathbf{J}, U \rangle, && (58)
\end{aligned}
$$

where (a) follows because $U = (1/n)\mathbf{I} + \mathbf{J}$ is strictly feasible for the supremum in (57) (i.e. it satisfies Slater's condition) so the strong duality holds.

Restricting U in (58) to satisfy $U_{ij} = 0$ except for those $i, j \notin C^*$, and $\langle U, \mathbf{J} \rangle = a \in [1, K]$, we get that $\eta \geq \sup_{1 \leq a \leq K} \{V_{n-K}(a) - a\lambda\}$. It follows from (54) that

$$
\begin{aligned}
\min_{i \in C^*} e(i, C^*) &\geq \sup_{1 \leq a \leq K} \{V_{n-K}(a) - a\lambda\} + \lambda K \\
&\geq \sup_{1 \leq a \leq K} \left\{ V_{n-K}(a) - \frac{a}{K} \max_{i \notin C^*} e(i, C^*) \right\} + \max_{i \notin C^*} e(i, C^*),
\end{aligned}
$$

where the last inequality follows from $a \leq K$ and (56). ∎

8.2.1. GAUSSIAN CASE

Consider the Gaussian case $P = \mathcal{N}(\mu, 1)$ and $Q = \mathcal{N}(0, 1)$. Before the proof of Theorem 9, we need to introduce a key lemma to lower bound the value of $V_m(a)$ given in (6). Recall that $m = n - K$. By the assumption, $L = A$ and hence M has the same distribution as an $m \times m$ symmetric random matrix W with zero-diagonal and $W_{ij} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1)$ for $1 \leq i < j \leq m$. The following lemma provides a high-probability lower bound on $V_m(a)$ defined in (6); its proof is deferred to Appendix E.

Lemma 21 *Assume that $a > 1$ and $a = o(m)$ as $m \to \infty$. Let $M = W$ be an $m \times m$ symmetric random matrix with zero-diagonal and independent standard normal entries in the definition of $V_m(a)$ in (6). Then with probability tending to one,*

$$
V_m(a) \geq \begin{cases} \frac{\sqrt{m}}{2} - r & a = \omega(\sqrt{m}) \\ a\sqrt{\log\left(1 + \frac{m}{4a^2}\right)} - o(a) & a = \Theta(\sqrt{m}) \\ (a-1)\sqrt{\frac{1}{3}\log\frac{m}{a^2}} - O(a \log\log\frac{m}{a^2}) & a = o(\sqrt{m}) \end{cases} \tag{59}
$$

where $r \triangleq \frac{m^{3/4}}{\sqrt{8(a-1)}} + \frac{2a}{\sqrt{m}} = o(\sqrt{m})$ if $a = \omega(\sqrt{m})$.

Remark 22 *We also have the following simple observations on $V_m(a)$:*

- $V_m(1) = 0$.

- *Dropping the second and the last constraints in (6), we have $V_m(a) \leq \lambda_{\max}(W) = 2\sqrt{m}(1 + o_P(1))$.*

- *Since $\|W\|_{\ell_\infty} = \sqrt{2\log\binom{m}{2}} + o_P(1)$, it follows that $V_m(a) \leq (a-1)\|W\|_\infty = (a-1)\sqrt{2\log\binom{m}{2}} + o_P(a)$.*

We next prove Theorem 9 by combining Theorem 3 and Lemma 21.

Proof [Proof of Theorem 9] By assumption, $\liminf_{n\to\infty} \mathbb{P}\{\widehat{Z}_{\mathrm{SDP}} = Z^*\} > 0$. It follows from Theorem 3 that with a non-vanishing probability,

$$
\min_{i\in C^*} \sum_{j\in C^*} A_{ij} - \max_{i\notin C^*} \sum_{j\in C^*} A_{ij} \geq \sup_{0\leq a\leq K}\left\{ V_{n-K}(a) - \frac{a}{K}\max_{i\notin C^*}\sum_{j\in C^*} A_{ij}\right\}. \tag{60}
$$

In Appendix G we show that

$$
\min_{i\in C^*}\sum_{j\in C^*} A_{ij} \leq (K-1)\mu - \sqrt{2(K-1)\log K} + o_P(\sqrt{K}). \tag{61}
$$

In view of Lemma 20,

$$
\max_{i\notin C^*}\sum_{j\in C^*} A_{ij} \geq \sqrt{2K\log(n-K)} + o_P(\sqrt{K}). \tag{62}
$$

It follows from (60) that with a non-vanishing probability,

$$
(K-1)\mu - \sqrt{2(K-1)\log K} - \sqrt{2K\log(n-K)} + o(\sqrt{K})
$$
$$
\geq \sup_{0\leq a\leq K}\left\{ V_{n-K}(a) - a\sqrt{2\log(n-K)/K}\right\}. \tag{63}
$$

Case 1: $K = \omega(\sqrt{n})$. We show that the necessary condition (15) holds. In view of (63), to get a necessary condition as tight as possible, one should choose a so that $V_{n-K}(a)$ is large and a is small comparing to K. To this end, set $a = \sqrt{K}(n-K)^{1/4}$. Since $K = o(n)$ and $K = \omega(\sqrt{n})$ by assumption, we have $a = \omega(\sqrt{n-K})$ and $a = o(K)$. Applying Lemma 21, we conclude that

$$
V_{n-K}(a) \geq \frac{\sqrt{n-K}}{2} + o_p(\sqrt{n-K}). \tag{64}
$$

Combining (60), (61), (62), and (64), and using $\sqrt{n-K} \geq \sqrt{n} - K/(2\sqrt{n-K})$, we obtain the desired (15).

Case 2: $K = O(\sqrt{n})$. In view of the high-probability lower bounds to $V_{n-K}(a)$ for $a = O(\sqrt{n-K})$ given in (59), $V_{n-K}(a) - a\sqrt{2\log(n-K)/K}$ is maximized over $[1, K]$ at $a = K$. Hence, we set $a = K$, which satisfies $a = O(\sqrt{n-K})$. It follows from (63) that with a non-vanishing probability,

$$(K-1)\mu - \sqrt{2(K-1)\log K} + o(\sqrt{K}) \geq V_{n-K}(K).$$

The desired lower bound on μ follows from the high-probability lower bounds on $V_{n-K}(K)$ given in (59) for $a = O(\sqrt{n-K})$. ■

8.2.2. BERNOULLI CASE

Recall that $m = n - K$ and by assumption, $L = A$. In the Bernoulli case, M is an $m \times m$ symmetric random matrix with zero diagonal and independent entries such that $M_{ij} = M_{ji} \sim \text{Bern}(q)$ for all $i < j$. The following lemma provides a high-probability lower bound on $V_m(a)$ defined in (6); its proof is deferred to Appendix F.

Lemma 23 (Lower bound to $V_m(a)$ in Bernoulli case) *Assume that $a = o(m)$, q is bounded away from 1, $m^2 q \to \infty$. Recall that κ is defined in (23). With probability tending to one,*

- *If $a - 1 \geq \frac{1}{\kappa}\sqrt{mq/(1-q)}$, then*

$$V_m(a) \geq (a-1)q + \frac{\sqrt{mq(1-q)}}{\kappa}.$$

- *If $0 \leq a - 1 \leq \frac{1}{\kappa}\sqrt{mq/(1-q)}$, then $V_m(a) = a - 1$.*

Remark 24 *We have the following simple observations on $V_m(a)$:*

- *$V_m(1) = 0$ and $V_m(a) \leq (a-1)\|A\|_\infty = a - 1$.*

- *Dropping the second and the last constraints in (6), we have with high probability $V_m(a) \leq \lambda_{\max}(A) \leq \kappa\sqrt{mq(1-q)}$.*

We next prove Theorem 13 by combining Theorem 3 and Lemma 23.

Proof [Proof of Theorem 13] We first show that if Z^* is unique with some non-vanishing probability, then $K - 1 \geq \sqrt{nq/(1-q)}/\kappa$. We prove it by contradiction. Suppose that $K - 1 < \sqrt{(n-K)q/(1-q)}/\kappa$. Let $\widetilde{A}$ denote the $(n-K) \times (n-K)$ submatrix of A supported on $(C^*)^c \times (C^*)^c$. Take $a = K$ in Lemma 23; the last statement of the lemma implies that $V_{n-K}(K) = K - 1$ with high probability. Furthermore, the proof of the lemma shows that the $(n-K) \times (n-K)$ matrix $\widetilde{Z}$ defined by $\widetilde{Z}_{ii} = 1/(n-K)$ and $\widetilde{Z}_{ij} = (K-1)\widetilde{A}_{ij}/\langle \widetilde{A}, \mathbf{J} \rangle$ for $i \neq j$ satisfies $\langle \widetilde{Z}, \widetilde{A} \rangle = K - 1$ and, with high probability, $\widetilde{Z} \succeq 0$. Let Z be the $n \times n$ matrix such that $Z_{(C^*)^c(C^*)^c} = K\widetilde{Z}$ and $Z_{ij} = 0$ for all $(i,j) \notin (C^*)^c \times (C^*)^c$. Then one can easily verify that Z is feasible for (3) with high probability and $\langle Z, A \rangle = K(K-1)$. Since $\langle Z^*, A \rangle \leq K(K-1)$, it follows that with high probability Z^* is not the unique optimal solution to (3), arriving at a contradiction. The necessity of (26) is then proved.

Next, we prove the necessary condition (27). Since $\liminf_{n\to\infty} \mathbb{P}\{\widehat{Z}_{\text{SDP}} = Z^*\} > 0$ by assumption, Theorem 3 implies that with a non-vanishing probability,

$$\min_{i\in C^*} \sum_{j\in C^*} A_{ij} - \max_{i\notin C^*} \sum_{j\in C^*} A_{ij} \geq \sup_{1\leq a\leq K} \left\{ V_{n-K}(a) - \frac{a}{K} \max_{i\notin C^*} e(i, C^*) \right\}. \tag{65}$$

We use the following lower bounds for the binomial distribution tails (Zubkov and Serov, 2013, Theorem 1):

$$\mathbb{P}\left\{\text{Binom}(m,p) \leq m\tau\right\} \geq Q\left(\sqrt{2md(\tau\|p)}\right), \quad 1/m \leq \tau \leq p, \tag{66}$$

$$\mathbb{P}\left\{\text{Binom}(m,q) \geq m\tau\right\} \geq Q\left(\sqrt{2md(\tau\|q)}\right), \quad q \leq \tau \leq 1. \tag{67}$$

Let

$$\delta = \max\left\{\frac{2\log\log K}{\log K}, \frac{\log\log(n-K)}{\log(n-K)}\right\},$$

and define

$$\tau_1' = (1-\delta)\tau_1 + \delta p$$
$$\tau_2' = (1-\delta)\tau_2 + \delta q.$$

Let $K_o = \lceil \frac{K}{\log K}\rceil$ and $\sigma^2 = (K_o - 1)p$. Define events

$$\mathcal{E}_1 = \left\{\min_{i\in C^*} \sum_{j\in C^*} A_{ij} \leq (K - K_o)\tau_1' + (K_o - 1)p + 6\sigma\right\}$$

$$\mathcal{E}_2 = \left\{\max_{i\notin C^*} \sum_{j\in C^*} A_{ij} \geq K\tau_2'\right\}.$$

By the definition of τ_2' and the convexity of divergence, we have that $d(\tau_2'\|q) \leq (1-\delta)d(\tau_2\|q)$. Thus

$$\begin{aligned}
\mathbb{P}\left\{\text{Binom}(K,q) \geq K\tau_2'\right\} &\geq Q\left(\sqrt{2Kd(\tau_2'\|q)}\right) \\
&\geq Q\left(\sqrt{2K(1-\delta)d(\tau_2\|q)}\right) \\
&= Q\left(\sqrt{2(1-\delta)\log(n-K)}\right) \\
&= \Omega\left(\frac{(n-K)^{-(1-\delta)}}{\sqrt{\log(n-K)}}\right) \\
&= \Omega\left(\frac{\sqrt{\log(n-K)}}{n-K}\right),
\end{aligned}$$

where we used the bound $Q(x) \geq \frac{1}{\sqrt{2\pi}} \frac{t}{t^2+1} e^{-t^2/2}$ and the fact that $\delta \geq \frac{\log\log(n-K)}{\log(n-K)}$. Hence,

$$
\mathbb{P}\left\{\max_{i \notin C^*} \sum_{j \in C^*} A_{ij} \geq K\tau_2'\right\} \overset{(a)}{=} 1 - \prod_{i \notin C^*} \mathbb{P}\left\{\sum_{j \in C^*} A_{ij} < K\tau_2'\right\}
$$

$$
\overset{(b)}{=} 1 - \left(1 - \mathbb{P}\left\{\mathrm{Binom}(K, q) \geq K\tau_2'\right\}\right)^{n-K}
$$

$$
\overset{(c)}{\geq} 1 - \exp\left(-(n-K)\mathbb{P}\left\{\mathrm{Binom}(K, q) \geq K\tau_2'\right\}\right)
$$

$$
\geq 1 - \exp\left(-\Omega\left(\sqrt{\log(n-K)}\right)\right) \to 1,
$$

where (a) holds due to the independence of $\sum_{j \in C^*} A_{ij}$ for different $i \notin C^*$; (b) holds because for $i \notin C^*$, $\sum_{j \in C^*} A_{ij} \sim \mathrm{Binom}(K, q)$; (c) follows from the fact that $1 - x \leq e^{-x}$ for $x \geq 0$. Hence, we get that $\mathbb{P}\{\mathcal{E}_2\} \to 1$.

In Appendix H we show that $\mathbb{P}\{\mathcal{E}_1\} \to 1$, i.e.,

$$
\mathbb{P}\left\{\min_{i \in C^*} \sum_{j \in C^*} A_{ij} \leq (K - K_o)\tau_1' + (K_o - 1)p + 6\sigma\right\} \to 1. \tag{68}
$$

Let $\mathcal{E} = \mathcal{E}_1 \cap \mathcal{E}_2$. Then by union bound, $\mathbb{P}\{\mathcal{E}\} \to 1$. It follows from (65) that with a non-vanishing probability,

$$
(K - K_o)\tau_1' + (K_o - 1)p + 6\sigma - K\tau_2' \geq \sup_{1 \leq a \leq K}\left\{V_{n-K}(a) - a\tau_2'\right\} \tag{69}
$$

Applying Lemma 23, we have that with probability converging to 1,

$$
V_{n-K}(a) \geq \begin{cases} (a-1)q + \frac{1}{\kappa}\sqrt{(n-K)q(1-q)} & a - 1 \geq \frac{1}{\kappa}\sqrt{(n-K)q/(1-q)} \\ a - 1 & 0 \leq a - 1 \leq \frac{1}{\kappa}\sqrt{(n-K)q/(1-q)}. \end{cases} \tag{70}
$$

Recall that we have shown that $K - 1 \geq \frac{1}{\kappa}\sqrt{(n-K)q/(1-q)}$ in the first part of the proof. In view of $\tau_2' \geq q$ and (70), $V_{n-K}(a) - a\tau_2'$ is maximized at

$$
a = \frac{1}{\kappa}\sqrt{(n-K)q/(1-q)} + 1 \in [1, K],
$$

which gives $V_{n-K}(a) = a - 1$. Hence, it follows from (69) that

$$
(K - K_o)\tau_1' + (K_o - 1)p + 6\sigma - K\tau_2' \geq a - 1 - a\tau_2',
$$

which further implies that

$$
(K - K_o)(\tau_1' - \tau_2') \geq a - 1 - \tau_2'a + K_o\tau_2' - (K_o - 1)p - 6\sigma
$$

$$
= (a-1)(1 - \tau_2') + (K_o - 1)(\tau_2' - p) - 6\sigma
$$

$$
\geq \frac{1}{\kappa}\sqrt{\frac{(n-K)q}{1-q}}(1 - \tau_2') - 6\sqrt{\frac{Kp}{\log K}} - \frac{K(p-q)}{\log K}.
$$

Plugging in the definition of τ_1' and τ_2', we derive that

$$
\begin{aligned}
(K - K_o)(\tau_1 - \tau_2) &\geq \frac{1}{\kappa}\sqrt{\frac{(n-K)q}{1-q}}(1 - \tau_2') - 6\sqrt{\frac{Kp}{\log K}} - \frac{K(p-q)}{\log K} - \delta(p-q) \\
&\geq \frac{1}{\kappa}\sqrt{\frac{(n-K)q}{1-q}}(1 - \tau_2) - 6\sqrt{\frac{Kp}{\log K}} - \frac{K(p-q)(2\log\log K + 1)}{\log K},
\end{aligned}
$$

where the last inequality follows because $\tau_2' \leq \tau_2$ and $\delta \leq \frac{2\log\log K}{\log K}$. Hence, we arrive at the desired necessary condition (27). ∎

8.2.3. MULTIPLE-COMMUNITY STOCHASTIC BLOCK MODEL

Proof [Proof of Theorem 17] Since the MLE is optimal, in proving the theorem, we can assume without loss of generality that the necessary condition for consistency of the MLE, $K(p-q)^2 = \Omega(q\log n)$, holds (see Remark 18). Since $p = \Theta(q)$, it follows that we can assume without loss of generality that $K(p-q) = \Omega(\log n)$ and $Kq = \Omega(\log n)$.

Suppose (34) fails, namely, there exists $\epsilon > 0$ such that

$$
\frac{(p-q)\sqrt{np}}{rq} \leq \frac{1-\epsilon}{\kappa}. \tag{71}
$$

We construct a matrix Y which, with high probability, constitutes a feasible solution to the SDP program (33) with an objective value exceeding that of Y^*. The construction is a variant of that used in proving Lemma 23 in Appendix F. Let

$$
Y = sA + t(\mathbf{J} - \mathbf{I}) + w(\mathbf{d}\mathbf{1}^\top + \mathbf{1}\mathbf{d}^\top - 2D) + \mathbf{I}, \tag{72}
$$

where $\mathbf{d} = A\mathbf{1}$ is the vector of node degrees, $D = \text{diag}\{\mathbf{d}\}$, $s \geq 0$ and $t, w \in \mathbb{R}$ are to be specified. In other words, $Y_{ij} = sA_{ij} + t + w(d_i + d_j)$ for $i \neq j$ and $Y_{ii} = 1$. Let $z \triangleq \langle A, \mathbf{J} \rangle = \langle \mathbf{d}, \mathbf{1} \rangle$. Note that for any $Y \succeq 0$, the constraint $\langle Y, \mathbf{J} \rangle = 0$ is equivalent to $Y\mathbf{1} = 0$. Since

$$
Y\mathbf{1} = s\mathbf{d} + t(n-1)\mathbf{1} + w(n\mathbf{d} + z\mathbf{1} - 2\mathbf{d}) + \mathbf{1} = (s + w(n-2))\mathbf{d} + (t(n-1) + wz + 1)\mathbf{1},
$$

to satisfy $Y\mathbf{1} = 0$, we let

$$
s + w(n-2) = 0, \quad t(n-1) + wz + 1 = 0
$$

namely,

$$
w = -\frac{s}{n-2}, \quad t = \frac{sz}{(n-1)(n-2)} - \frac{1}{n-1}. \tag{73}
$$

Since $w \leq 0$, to satisfy the other constraints in (33), it suffices to ensure

$$
t + 2wd_{\max} \geq -\frac{1}{r-1} \tag{74}
$$

$$
Y \succeq 0, \tag{75}
$$

27

where $d_{\max} = \max_i d_i$ is the maximal degree.

Since $Y\mathbf{1} = 0$, (75) is equivalent to $PYP \succeq 0$, where $P = \mathbf{I} - \frac{1}{n}\mathbf{J}$ is the matrix for projection onto the subspace orthogonal to $\mathbf{1}$. Since

$$PYP = P(sA + (1-t)\mathbf{I} - 2wD)P,$$

in view of the facts that $\mathbb{E}[A] \succeq -p\mathbf{I}$, $D \succeq 0$, and $w \leq 0$, it suffices to verify that

$$s\|A - \mathbb{E}[A]\| \leq 1 - t - sp. \tag{76}$$

Next, we compute the objective value:

$$\langle A, Y \rangle = (s+t)\langle A, \mathbf{J} \rangle + 2w\|\mathbf{d}\|_2^2.$$

By the Chernoff bounds for binomial distributions,

$$\langle A, Y^* \rangle = \frac{n^2}{r}(p-q) + O_P\left(\sqrt{n^2 p/r}\right)$$

$$\langle A, \mathbf{J} \rangle = \frac{n^2}{r}\left(p + (r-1)q\right) + O_P(\sqrt{n^2 q}).$$

Then $\langle A, Y^* \rangle = nK(p-q)(1 + o_P(1))$ and $z = \langle A, \mathbf{J} \rangle = n^2 q(1 + o_P(1))$. By concentration,[5] $\|\mathbf{d}\|_2^2 = n^3 q^2(1 + o_P(1))$ and $d_{\max} = nq(1 + o_P(1))$. To ensure that $\langle A, Y \rangle > \langle A, Y^* \rangle$, we set $\langle A, Y \rangle = (1 + \epsilon)\langle A, Y^* \rangle$, or equivalently:

$$(s+t)z + 2w\|\mathbf{d}\|_2^2 = (1+\epsilon)\langle A, Y^* \rangle, \tag{77}$$

Solving (73) and (77) and by the assumption $p = o(1)$ and the fact $\frac{1}{n-1} = o(\frac{p-q}{r})$, we have:

$$s = (1 + \epsilon + o_P(1))\frac{p-q}{rq}, \quad t = (1 + \epsilon + o_P(1))\frac{p-q}{r}, \quad w = -(1 + \epsilon + o_P(1))\frac{p-q}{nrq}. \tag{78}$$

Hence $t + 2wd_{\max} = -(1 + \epsilon + o_P(1))\frac{p-q}{r} \geq -\frac{1}{r-1}$, i.e., (74), holds with high probability.

It remains to verify (76). Since $np = \Omega(\log n)$, applying Lemma 30 yields $\|A - \mathbb{E}[A]\| \leq \kappa\sqrt{np}$ with high probability. In view of the assumption (71), (76) holds with high probability, which completes the proof. $\blacksquare$

Acknowledgments

---

5. We use the following implication of the Chernoff bound: If X is the sum of independent Bernoulli random variables with mean μ, then for $\delta \geq 2e - 1$, $\mathbb{P}\{X \geq (1+\delta)\mu\} \leq 2^{-\delta\mu}$, and the assumptions $Kq = \Omega(\log n)$ and $r \to \infty$.

References

N. Agarwal, A. S. Bandeira, K. Koiliaris, and A. Kolla. Multisection in the stochastic block model using semidefinite programming. arXiv 1507.02323, July 2015.

N. Alon, M. Krivelevich, and B. Sudakov. Finding a large hidden clique in a random graph. *Random Structures and Algorithms*, 13(3-4):457–466, 1998.

Z.D. Bai and Y.Q. Yin. Necessary and sufficient conditions for almost sure convergence of the largest eigenvalue of a Wigner matrix. *The Annals of Probability*, 16(4):1729–1741, 1988.

A. S. Bandeira and R. van Handel. Sharp nonasymptotic bounds on the norm of random matrices with independent entries. arXiv 1408.6185, 2014.

A.S. Bandeira. Random Laplacian matrices and convex relaxations. arXiv 1504.03987, April, 2015.

S. Boucheron, G. Lugosi, and P. Massart. *Concentration inequalities: A nonasymptotic theory of independence*. Oxford University Press, 2013.

S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, New York, NY, USA, 2004.

C. Butucea, Y.I. Ingster, and I. Suslina. Sharp variable selection of a sparse submatrix in a high-dimensional noisy matrix. *ESAIM: Probability and Statistics*, 19:115–134, June 2015.

T. T. Cai, T. Liang, and A. Rakhlin. Computational and statistical boundaries for submatrix localization in a large noisy matrix. arXiv:1502.01988, Feb. 2015.

Y. Chen and J. Xu. Statistical-computational tradeoffs in planted problems and submatrix localization with a growing number of clusters and submatrices. In *Proceedings of ICML 2014 (Also arXiv:1402.1267)*, Feb 2014.

K.-L. Chung and P. Erdös. On the application of the Borel-Cantelli lemma. *Transactions of the American Mathematical Society*, pages 179–186, 1952.

H.A. David and H.N. Nagaraja. *Order Statistics*. Wiley-Interscience, Hoboken, New Jersey, USA, 3 edition, 2003.

K.R. Davidson and S. Szarek. Local operator theory, random matrices and Banach spaces. In W.B. Johnson and J. Lindenstrauss, editors, *Handbook on the Geometry of Banach Spaces*, volume 1, pages 317–366. Elsevier Science, 2001.

Y. Deshpande and A. Montanari. Finding hidden cliques of size $\sqrt{N/e}$ in nearly linear time. *Foundations of Computational Mathematics*, 15(4):1069–1128, August 2015a.

Y. Deshpande and A. Montanari. Improved sum-of-squares lower bounds for hidden clique and hidden submatrix problems. In *Proceedings of COLT 2015*, pages 523–562, June 2015b.

U. Feige and R. Krauthgamer. Finding and certifying a large hidden clique in a semirandom graph. *Random Structures & Algorithms*, 16(2):195–208, 2000.

U. Feige and R. Krauthgamer. The probable value of the Lovász–Schrijver relaxations for maximum independent set. *SIAM Journal on Computing*, 32(2):345–370, 2003.

B. Hajek, Y. Wu, and J. Xu. Computational lower bounds for community detection on random graphs. In *Proceedings of COLT 2015*, June 2015a.

B. Hajek, Y. Wu, and J. Xu. Achieving exact cluster recovery threshold via semidefinite programming: Extensions. arXiv 1502.07738, Feb. 2015b.

B. Hajek, Y. Wu, and J. Xu. Submatrix localization via message passing. arXiv 1510.09219, October 2015c.

B. Hajek, Y. Wu, and J. Xu. Recovering a hidden community beyond the spectral limit in $O(|E|\log^* |V|)$ time. arXiv 1510.02786, October 2015d.

B. Hajek, Y. Wu, and J. Xu. Information limits for recovering a hidden community. arXiv 1509.07859, September 2015e.

B. Hajek, Y. Wu, and J. Xu. Achieving exact cluster recovery threshold via semidefinite programming. *IEEE Transactions on Information Theory*, 62(5):2788–2797, May 2016. (arXiv 1412.6156 Nov. 2014).

P. W. Holland, K. B. Laskey, and S. Leinhardt. Stochastic blockmodels: First steps. *Social Networks*, 5(2):109–137, 1983.

S. B. Hopkins, P. K. Kothari, and A. Potechin. SoS and planted clique: Tight analysis of MPW moments at all degrees and an optimal lower bound at degree four. arXiv 1507.05230, July 2015.

R.M. Karp. Reducibility among combinatorial problems. In R.E. Miller and J.W. Thacher, editors, *Proceedings of a Symposium on the Complexity of Computer Computations*, pages 85–103. Plenum Press, March 1972.

M. Kolar, S. Balakrishnan, A. Rinaldo, and A. Singh. Minimax localization of structural information in large noisy matrices. In *Advances in Neural Information Processing Systems*, 2011.

R. Krauthgamer, B. Nadler, and D. Vilenchik. Do semidefinite relaxations solve sparse PCA up to the information limit? *The Annals of Statistics*, 43(3):1300–1322, June 2015.

R. Latała. Some estimates of norms of random matrices. *Proceedings of the American Mathematical Society*, 133(5):1273–1282, 2005.

C. M. Le and R. Vershynin. Concentration and regularization of random graphs. arXiv:1506.00669, June 2015.

Z. Ma and Y. Wu. Computational barriers in minimax submatrix detection. *The Annals of Statistics*, 43(3):1089–1116, 2015.

F. McSherry. Spectral partitioning of random graphs. In *42nd IEEE Symposium on Foundations of Computer Science*, pages 529 – 537, Oct. 2001.

R. Meka, A. Potechin, and A. Wigderson. Sum-of-squares lower bounds for planted clique. In *Proceedings of the Forty-Seventh Annual ACM on Symposium on Theory of Computing*, STOC '15, pages 87–96, New York, NY, USA, 2015. ACM.

A. Montanari. Finding one community in a sparse random graph. *Journal of Statistical Physics*, 161 (2):273–299, 2015. arXiv 1502.05680.

A. Montanari and S. Sen. Semidefinite programs on sparse random graphs. *arXiv:1504.05910*, April, 2015.

W. Perry and A.S. Wein. A semidefinite program for unbalanced multisection in the stochastic block model. arXiv 1507.05605, July 2015.

P. Raghavendra and T. Schramm. Tight lower bounds for planted clique in the degree-4 SOS program. arXiv:1507.05136, July 2015.

A. A. Shabalin, V. J. Weigman, C. M. Perou, and A. B Nobel. Finding large average submatrices in high dimensional data. *The Annals of Applied Statistics*, 3(3):985–1012, 2009.

T. Tao. *Topics in random matrix theory*. American Mathematical Society, Providence, RI, USA, 2012.

R. K. Vinayak, S. Oymak, and B. Hassibi. Sharp performance bounds for graph clustering via convex optimization. In *38th International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2014.

V. H. Vu. Spectral norm of random matrices. *Combinatorica*, 27(6):721–736, 2007. ISSN 0209-9683. doi: 10.1007/s00493-007-2190-z. URL http://dx.doi.org/10.1007/s00493-007-2190-z.

A. M. Zubkov and A. A. Serov. A complete proof of universal inequalities for the distribution function of the binomial law. *Theory of Probability & Its Applications*, 57(3):539–544, 2013.

Appendix A. Bounds on spectral norms of random matrices

For the convenience of the reader, this section collects known bounds on the spectral norms of random matrices that are used in this paper.

Lemma 25 (Gordon-Davidson-Szarek) *If Y is an $n \times n$ random matrix such that the random variables $(Y_{ij} : 1 \le i \le j \le n)$ are independent, Gaussian, with mean zero and $\mathsf{var}(Y_{ij}) \le 1$ then $\mathbb{P}\{\|Y\| \ge 2\sqrt{n} + t\} \le 2\exp(-t^2/4)$ for any $t > 0$.*

Lemma 25 is a slight generalization of (Davidson and Szarek, 2001, Theorem 2.11), which applies to the case of $\mathsf{var}(Y_{ij}) = 1$ and is based on Gordon's inequality on the expected norm $\mathbb{E}[\|Y\|]$, proved, in turn, by the Slepian-Gordon comparison lemma. Examining the proof shows that the assumption can be weakened to $\mathsf{var}(Y_{ij}) \le 1$.

Lemma 26 ((Latała, 2005, Theorem 2)) *There is a universal constant C such that whenever A is a random matrix (not necessarily square) of independent and zero-mean entries:*

$$\mathbb{E}\left[\|A\|\right] \leq C \left(\max_i \sqrt{\sum_j \mathbb{E}[a_{ij}^2]} + \max_j \sqrt{\sum_i \mathbb{E}[a_{ij}^2]} + \sqrt[4]{\sum_{ij} \mathbb{E}[a_{ij}^4]} \right)$$

Lemma 27 (Corollary of (Bai and Yin, 1988, Theorem A)) *Let $W = (X_{ij}, i, j \geq 1)$ be a symmetric infinite matrix such that entries above the diagonal are mean zero iid, entries on the diagonal are iid, and the diagonal of W is independent of the off-diagonal. Let $W_n = (X_{ij} : i, j \in \{1, \ldots, n\})$ for $n \geq 1$. Let $\sigma^2 = \mathsf{var}(X_{12})$. If $\mathbb{E}\left[X_{11}^2\right] < \infty$ and $\mathbb{E}\left[X_{12}^4\right] < \infty$, then $\|W_n\|/\sqrt{n} \to 2\sigma$ a.s. as $n \to \infty$.*

The following two lemmas are used in the proof of Lemma 30 below.

Lemma 28 ((Bandeira and van Handel, 2014, Corollary 3.6)) *Let X be an $n \times n$ symmetric random matrix with $X_{ij} = \xi_{ij} b_{ij}$, where $\{\xi_{ij} : i \geq j\}$ are independent symmetric random variables with unit variance, and $\{b_{ij} : i \geq j\}$ are given scalers. Then for any $\alpha \geq 3$,*

$$\mathbb{E}\left[\|X\|\right] \leq \mathrm{e}^{2/\alpha} \left\{ 2\sigma + 14\alpha \max_{ij} \|\xi_{ij} b_{ij}\|_{2\lceil \alpha \log n \rceil} \sqrt{\log n}, \right\}$$

where $\sigma^2 \triangleq \max_i \sum_j b_{ij}^2$.

Lemma 29 ((Vu, 2007, Theorem 1.4)) *There are universal constants C and C' such that the following holds. Let A be a symmetric random matrix such that $\{A_{ij} : 1 \leq i \leq j \leq n\}$ are independent, zero-mean, variance at most σ^2, and bounded in absolute value by K. If K and σ depend on n such that $\sigma \geq C' n^{-1/2} K \log^2 n$, then*

$$\|A\| \leq 2\sigma\sqrt{n} + C(K\sigma)^{1/2} n^{1/4} \log n, \tag{79}$$

with probability converging to one as $n \to \infty$.

For example, for the case the matrix entries are $\mathrm{Bern}(p)$, the second term in (79) becomes asymptotically negligible compared to the first if $\sqrt{pn} = \omega((np)^{1/4} \log n)$, or equivalently, $np = \omega(\log^4 n)$.

Lemma 30 *Let M denote a symmetric $n \times n$ random matrix with zero diagonals and independent entries such that $M_{ij} = M_{ji} \sim \mathrm{Bern}(p_{ij})$ for all $i < j$ with $p_{ij} \in [0, 1]$. Assume $p_{ij}(1 - p_{ij}) \leq r$ for all $i < j$ and $nr = \Omega(\log n)$. Then, with high probability,*

$$\|M - \mathbb{E}\left[M\right]\| \leq \kappa \sqrt{nr},$$

where

$$\kappa = \begin{cases} O(1) & nr = \Omega(\log n) \\ 4 + o(1) & nr = \omega(\log n) \\ 2 + o(1) & nr = \omega(\log^4 n) \end{cases} . \tag{80}$$

Proof It follows from the symmetrization argument and Lemma 28 (for this application of the lemma, $b_{ij} \leq \sqrt{r}$, $|\xi_{ij} b_{ij}| \leq 1$, and $\sigma^2 \leq nr$) that for any $\alpha \geq 3$,

$$\mathbb{E}\left[\|M - \mathbb{E}[M]\|\right] \leq 2\mathbb{E}\left[\|(M - \mathbb{E}[M]) \circ E\|\right] \leq 2e^{2/\alpha}(2\sqrt{nr} + 14\alpha\sqrt{\log n})$$

where E is an $n \times n$ zero-diagonal, symmetric random matrix whose entries are Rademacher and independent from M. Since $nr = \Omega(\log n)$, we have that $\mathbb{E}\left[\|M - \mathbb{E}[M]\|\right] = O(\sqrt{nr})$. If $nr = \omega(\log n)$, then by letting $\alpha = (nr/\log n)^{1/4}$, we have that $\mathbb{E}\left[\|M - \mathbb{E}[M]\|\right] \leq (4 + o(1))\sqrt{nr}$. Talagrand's inequality for Lipschitz convex functions (see Tao (2012) or (Boucheron et al., 2013, Theorem 7.12)) implies that for any constant $c' > 0$, there exists a constant $c_0 > 0$ such that with probability at least $1 - n^{-c'}$, $\|M - \mathbb{E}[M]\| \leq \mathbb{E}\left[\|M - \mathbb{E}[M]\|\right] + c_0\sqrt{\log n}$. Hence, we have proved the lemma for the case of $nr = \Omega(\log n)$ and the case of $nr = \omega(\log n)$. Finally, if $nr = \omega(\log^4 n)$, then the lemma is a direct consequence of Vu's result, Lemma 29, with $K = 1$ and $\sigma^2 = r$. ■

Appendix B. A concentration inequality for a random matrix of log normal entries

Let $g(x) = e^{\tau x - \tau^2/2}$ for some $\tau > 0$. Recall that W is an $m \times m$ symmetric, zero-diagonal random matrix with i.i.d. standard Gaussian entries up to symmetry. Let $g(W)$ denote an $m \times m$ symmetric, zero-diagonal random matrix whose (i, j)-th entry is $g(W_{ij})$ for $i \neq j$. We need the following matrix concentration inequality for $g(W)$.

Lemma 31 *There exists a universal constant $C > 0$ such that*

$$\mathbb{E}\left[\|g(W) - \mathbb{E}[g(W)]\|\right] \leq C\sqrt{m(e^{3\tau^2} - 1)}. \tag{81}$$

In addition, if $\tau \to 0$ as $m \to \infty$, then the following refined bound holds:

$$\mathbb{P}\left\{\|g(W) - \mathbb{E}[g(W)]\| > 2\sqrt{m}\tau + \Delta\right\} \leq O(\sqrt{\tau}) + 2e^{-m\tau/4}, \tag{82}$$

where $\Delta = 2\sqrt{m}\tau^{3/2} = o(\sqrt{m}\tau)$.

Proof We first prove (81). Let U be the upper-triangular part of $g(W) - \mathbb{E}[g(W)]$. Then $\mathbb{E}\left[\|g(W) - \mathbb{E}[g(W)]\|\right] \leq 2\mathbb{E}[\|U\|]$. Note that U consists of independent zero-mean entries, applying Latała's theorem (Lemma 26), we have for some universal constant $c' > 0$,

$$\mathbb{E}[\|U\|] \leq c'\left(\max_i\sqrt{\sum_j \mathbb{E}[U_{ij}^2]} + \max_j\sqrt{\sum_i \mathbb{E}[U_{ij}^2]} + \sqrt[4]{\sum_{i,j} \mathbb{E}[U_{ij}^4]}\right) \tag{83}$$

$$\leq c'\sqrt{m}\left(2\sqrt{\mathbb{E}[U_{12}^2]} + \mathbb{E}[U_{12}^4]^{1/4}\right). \tag{84}$$

Note that

$$\mathbb{E}[U_{12}^2] = \mathbb{E}[(e^{\tau W_{12} - \tau^2/2} - 1)^2] = e^{\tau^2} - 1$$

Similarly,

$$\mathbb{E}[U_{12}^4] = e^{6\tau^2} - 4e^{3\tau^2} + 6e^{\tau^2} - 3.$$

33

Combining the last three displays gives that $\mathbb{E}\left[\|U\|\right] \le C_0\sqrt{m}\mathrm{e}^{3\tau^2/2}$ holds for any $\tau > 0$ and some universal constant $C_0 > 0$. To complete the proof of (81), it remains to show that $\mathbb{E}\left[\|U\|\right] \le c\sqrt{m(e^{\tau^2}-1)}$ for all $\tau \in [0,1]$. Indeed,

$$\mathbb{E}[U_{12}^4] = \mathrm{e}^{6\tau^2} - 4e^{3\tau^2} + 6e^{\tau^2} - 3$$
$$= g(e^{\tau^2}-1) \le 200(e^{\tau^2}-1)^2 \le 800\tau^4,$$

where $g(s) \triangleq (s+1)^6 - 4(s+1)^3 + 6(s+1) - 3 = s^2(3 + 16s + 15s^2 + 6s^3 + s^4) \le 200s^2$ for all $s \in [0,2]$. Applying (84) again yields the desired result.

Next we establish the finer estimate (82) for $\tau \to 0$. The main idea is to linearize the function g. To this end, let $h(x) = g(x) - 1 - \tau x$. Since $\mathbb{E}\left[g(W)\right] = \mathbf{J} - \mathbf{I}$, it follows that

$$g(W) - \mathbb{E}\left[g(W)\right] = \tau W + h(W).$$

Lemma 25 yields that $\mathbb{P}\left\{\|W\| \ge 2\sqrt{m} + t\right\} \le 2\exp(-t^2/4)$ for any $t > 0$. Hence,

$$\mathbb{P}\left\{\|\tau W\| \ge 2\sqrt{m}\tau + \sqrt{m}\tau^{3/2}\right\} \le 2\mathrm{e}^{-m\tau/4}. \tag{85}$$

To bound $\|h(W)\|$, let B be the upper-triangular part of $h(W)$, namely, $B_{ij} = h(W_{ij})$ if $i < j$ and 0 elsewhere. Then $\|h(W)\| \le 2\|B\|$. Since B consists of independent zero-mean entries, Lemma 26 yields

$$\mathbb{E}\left[\|B\|\right] \le c\left(\max_i\sqrt{\sum_j \mathbb{E}[B_{ij}^2]} + \max_j\sqrt{\sum_i \mathbb{E}[B_{ij}^2]} + \sqrt[4]{\sum_{i,j}\mathbb{E}[B_{ij}^4]}\right)$$
$$\le c\sqrt{m}\left(2\sqrt{\mathbb{E}[B_{12}^2]} + \mathbb{E}[B_{12}^4]^{1/4}\right)$$

for some universal constant c. Note that

$$\mathbb{E}[B_{12}^2] = \mathbb{E}[(e^{\tau W_{12} - \tau^2/2} - 1 - \tau W_{12})^2] = e^{\tau^2} - 1 - \tau^2 = O(\tau^4)$$

as $\tau \to 0$. Similarly,

$$\mathbb{E}[B_{12}^4] = e^{6\tau^2} - 4e^{3\tau^2}\left(3\tau^2 + 1\right) + 6e^{\tau^2}\left(4\tau^4 + 5\tau^2 + 1\right) - 3 - 4\tau^6 - 21\tau^4 - 18\tau^2 = O(\tau^8).$$

Consequently, $\mathbb{E}\left[\|B\|\right] = O(\sqrt{m}\tau^2)$. Therefore

$$\mathbb{P}\left\{\|h(W)\| \ge \sqrt{m}\tau^{3/2}\right\} \le \mathbb{P}\left\{\|B\| \ge \sqrt{m}\tau^{3/2}/2\right\} = O(\sqrt{\tau}).$$

Combining the last displayed equation with (85) and applying the union bound complete the proof for the case $\tau \to 0$. ∎

Appendix C. Useful facts on binary divergences

Lemma 32 *For any $0 < q \leq p < 1$,*

$$\frac{(p-q)^2}{2p(1-q)} \leq d(p\|q) \leq \frac{(p-q)^2}{q(1-q)}. \tag{86}$$

$$\frac{(p-q)^2}{2p(1-q)} \leq d(q\|p) \leq \frac{(p-q)^2}{p(1-p)}. \tag{87}$$

Proof The upper bound follows by applying the inequality $\log x \leq x - 1$ for $x > 0$ and the lower bound is proved using $\frac{\partial^2 d(p\|q)}{\partial p^2} = \frac{1}{p(1-p)}$ and Taylor's expansion. ∎

Lemma 33 *Assume that $0 < q \leq p < 1$ and $u, v \in [q, p]$. Then for any $0 < \eta < 1$,*

$$d((1-\eta)u + \eta v \| v) \geq \left(1 - \frac{2\eta p(1-q)}{q(1-p)}\right) d(u\|v), \tag{88}$$

$$d((1-\eta)u + \eta v \| u) \geq \frac{\eta^2 q(1-p)}{2p(1-q)} \max\{d(u\|v), d(v\|u)\}. \tag{89}$$

Proof By the mean value theorem,

$$d((1-\eta)u + \eta v \| v) = d(u\|v) - \eta(u-v)d'(x\|v),$$

for some $x \in (\min\{u,v\}, \max\{u,v\})$. Notice that $d'(x\|v) = \log \frac{x(1-v)}{(1-x)v}$ and thus

$$|d'(x\|q)| \leq \log \frac{\max\{u,v\}(1 - \min\{u,v\})}{\min\{u,v\}(1 - \max\{u,v\})} \leq \frac{|u-v|}{q(1-p)},$$

where the last equality holds due to $\log(1+x) \leq x$ and $x \in (q, p)$. It follows that

$$d((1-\eta)u + \eta v \| v) \geq d(u\|v) - \frac{\eta(u-v)^2}{q(1-p)} \geq \left(1 - \frac{2\eta p(1-q)}{q(1-p)}\right) d(u\|v),$$

where the last inequality holds due to the lower bounds in (86) and (87). Thus the first claim follows. For the second claim,

$$d((1-\eta)u + \eta v \| u) \geq \frac{\eta^2(u-v)^2}{2p(1-q)} \geq \frac{\eta^2 q(1-p)}{2p(1-q)} \max\{d(u\|v), d(v\|u)\},$$

where the first inequality holds due to the lower bounds in (86) and (87); the last inequality holds due to the upper bounds in (86) and (87). ∎

Lemma 34 *Assume that $\log \frac{p(1-q)}{q(1-p)}$ is bounded from above. Suppose for some $\epsilon > 0$ that $Kd(p\|q) > (1+\epsilon)\log \frac{n}{K}$ for all sufficiently large n. Recall that τ^* is defined in (28). Then $p - \tau^* = \Theta(p-q)$ and $\tau^* - q = \Theta(p-q)$.*

Proof By the definition of τ^*,

$$p - \tau^* = \frac{d(p\|q) - \frac{1}{K}\log\frac{n}{K}}{\log\frac{p(1-q)}{q(1-p)}},$$

$$\tau^* - q = \frac{d(q\|p) + \frac{1}{K}\log\frac{n}{K}}{\log\frac{p(1-q)}{q(1-p)}}.$$

Notice that $d(p\|q) + d(q\|p) = (p-q)\log\frac{p(1-q)}{q(1-p)}$. Hence,

$$\frac{p - \tau^*}{p - q} = \frac{d(p\|q) - \frac{1}{K}\log\frac{n}{K}}{d(p\|q) + d(q\|p)},$$

$$\frac{\tau^* - q}{p - q} = \frac{d(q\|p) + \frac{1}{K}\log\frac{n}{K}}{d(p\|q) + d(q\|p)}.$$

By the boundedness assumption of $\log\frac{p(1-q)}{q(1-p)}$ and Lemma 32, $d(p\|q) \asymp d(q\|p)$. Since $Kd(p\|q) > (1+\epsilon)\log\frac{n}{K}$ for all sufficiently large n, it follows that $p - \tau^*$ and $\tau^* - q$ are both $\Theta(p-q)$. ∎

Lemma 35 *Assume that $\log\frac{p(1-q)}{q(1-p)}$ is bounded. Suppose that $Kd(p\|q) > (1+\epsilon)\log\frac{n}{K}$ for all sufficiently large n.*

- *If $\liminf_{n\to\infty}\frac{Kd(\tau^*\|q)}{\log n} \geq 1$, then τ_1 and τ_2 in (21) are well-defined and take values in the interval $[q, p]$.*

- *If $\liminf_{n\to\infty}\frac{Kd(\tau^*\|q)}{\log n} > 1$, then there exists a fixed constant $\eta > 0$ such that $\tau_1 \geq (1-\eta)\tau^* + \eta p$ and $\tau_2 \leq (1-\eta)\tau^* + \eta q$.*

Proof It follows from Lemma 34 that $p - \tau^* = \Omega(p-q)$ and $\tau^* - q = \Omega(p-q)$. In particular, there exists a fixed constant $\delta > 0$ such that $(1-\delta)q + \delta p \leq \tau^* \leq (1-\delta)p + \delta q$. By the monotonicity and convexity of divergence, $d(\tau^*\|q) \leq (1-\delta)d(p\|q)$ and $d(\tau^*\|p) \leq (1-\delta)d(q\|p)$. Hence, if $\liminf_{n\to\infty}\frac{Kd(\tau^*\|q)}{\log n} \geq 1$, then $Kd(p\|q) \geq (1+\delta')\log n$ and $Kd(q\|p) \geq (1+\delta')\log K$ for some fixed constant $\delta' > 0$. Thus, in view of the continuity of binary divergence functions, τ_1 and τ_2 are well-defined, and moreover $\tau_1 \geq q$ and $\tau_2 \leq p$.

Note that $(1-\eta)\tau^* + \eta p \in [q, p]$. In view of Lemma 33,

$$d\left((1-\eta)\tau^* + \eta q\|q\right) \geq \left(1 - \frac{2\eta p(1-q)}{q(1-p)}\right)d(\tau^*\|q)$$

If $\liminf_{n\to\infty}\frac{Kd(\tau^*\|q)}{\log n} > 1$, then there exists a fixed constant $\epsilon' > 0$ such that for sufficiently large n, $Kd(\tau^*q) \geq (1+\epsilon')\log n$. It follows from the last displayed equation that by choosing η sufficiently small, $d\left((1-\eta)\tau^* + \eta q\|q\right) \geq (1+\delta')\log n$ for some fixed constant $\delta' > 0$. Thus by definition, $\tau_2 \leq (1-\eta)\tau^* + \eta q$. Similarly, one can verify that $\tau_1 \geq (1-\eta)\tau^* + \eta p$. ∎

Appendix D. Proof of Corollary 16

We first show that if $\gamma_1 > \gamma_2$, then

$$\liminf_{n\to\infty} \frac{Kd(\tau^*\|q)}{\log n} > 1, \tag{90}$$

which implies that MLE achieves exact recovery in view of (29).

Recall that $I(x,y) = x - y\log(\mathrm{e}x/y)$ for $x, y > 0$. Define $\tau_0 = \frac{a-b}{\log(a/b)}$. Then $I(b, \tau_0) = I(a, \tau_0)$. Note that $I(b, \gamma_2) = I(a, \gamma_1) = 1/\rho$. Since $I(b, x)$ is strictly increasing over $[b, \infty)$ and $I(a, x)$ is strictly decreasing over $(0, a]$, it follows that $\gamma_2 < \tau_0 < \gamma_1$. Thus $I(b, \tau_0) > 1/\rho$. In the regime (32), we have $\tau^* = \frac{\log^2 n}{n}(\tau_0 + o(1))$. Taylor's expansion yields that

$$d(\tau\|q) = q - \tau\log\frac{\mathrm{e}q}{\tau} + O((\tau - q)^2) = I(q, \tau) + O((\tau - q)^2).$$

Therefore,

$$d(\tau^*\|q) = \frac{\log^2 n}{n}\left(I(b, \tau_0) + o(1)\right),$$

which implies the desired (90).

Secondly, suppose that MLE achieves exact recovery. We aim to show that $\gamma_1 \geq \gamma_2$. Suppose not. Then $\gamma_1 < \gamma_2$. By the similar argument as above, it follows that $\gamma_1 < \tau_0 < \gamma_2$. Thus $I(b, \tau_0) < 1/\rho$. As a consequence,

$$\frac{Kd(\tau^*\|q)}{\log n} \leq 1 - \epsilon,$$

for some positive constant $\epsilon > 0$, which contradicts the fact that $\liminf_{n\to\infty} \frac{Kd(\tau^*\|q)}{\log n} \geq 1$, the necessary condition (30) for MLE to achieve exact recovery.

Finally, we prove the claims for SDP. By definition, $\tau_1 = \log^2 n(\gamma_1 + o(1))/n$ and $\tau_2 = \log^2 n(\gamma_2 + o(1))/n$. Therefore, if $\rho(\gamma_1 - \gamma_2) > 4\sqrt{b}$, then the sufficient condition for SDP (22) holds; if the necessary condition for SDP (27) holds, then $\rho(\gamma_1 - \gamma_2) \geq \sqrt{b}/4$.

Appendix E. Proof of Lemma 21

Proof To prove the desired lower bound of $V_m(a)$, we construct an explicit feasible solution Z to (6). For a given $\tau \in \mathbb{R}$, let

$$g(x) = e^{\tau x - \tau^2/2}, \quad \alpha = \frac{2}{m(m-1)}\sum_{i<j} g(W_{ij}) > 0.$$

Define an $m \times m$ matrix Z by $Z_{ii} = \frac{1}{m}$ and $Z_{ij} = \frac{a-1}{\alpha m(m-1)}g(W_{ij})$ for $i \neq j$. By definition, $Z \geq 0, \mathsf{Tr}(Z) = 1$, and $\langle Z, \mathbf{J}\rangle = a$.

We pause to give some intuition on the construction of Z. Note that g is in fact the likelihood ratio between two shifted Gaussians: $g(x) = \frac{\mathrm{d}\mathcal{N}(\tau,1)}{\mathrm{d}\mathcal{N}(0,1)}(x)$ and thus $\mathbb{E}\left[g(W_{ij})\right] = 1$ and $\mathbb{E}\left[W_{ij}g(W_{ij})\right] = \tau$. Therefore we expect that α is concentrated near 1, and similarly

$$\langle Z, W\rangle \approx \frac{2(a-1)}{\alpha m(m-1)}\sum_{i<j}\mathbb{E}\left[W_{ij}g(W_{ij})\right] = \frac{(a-1)\tau}{\alpha} \approx (a-1)\tau.$$

Thus to get a lower bound to $V_m(a)$ as tight as possible, we would like to maximize τ so that $Z \succeq 0$ with high probability.

Recall that as defined in Appendix B, $g(W)$ is an $m \times m$ zero-diagonal matrix with the (i,j)-th entry given by $g(W_{ij})$. Then $\mathbb{E}\left[g(W)\right] = \mathbf{J} - \mathbf{I}$ and

$$
\begin{aligned}
Z &= \frac{1}{m}\mathbf{I} + \frac{a-1}{\alpha m(m-1)}g(W) \\
&= \frac{1}{m}\mathbf{I} + \frac{a-1}{\alpha m(m-1)}(\mathbf{J} - \mathbf{I}) + \frac{a-1}{\alpha m(m-1)}\left(g(W) - \mathbb{E}\left[g(W)\right]\right).
\end{aligned}
$$

Hence, since $\mathbf{J} \succeq 0$ and $a \geq 1$, to show $Z \succeq 0$, it suffices to verify that

$$
\|g(W) - \mathbb{E}\left[g(W)\right]\| \leq \frac{\alpha(m-1)}{a-1} - 1. \tag{91}
$$

Therefore, we aim to choose τ as large as possible to satisfy (91). Recall that Lemma 31 provides different upper bounds on $\|g(W) - \mathbb{E}\left[g(W)\right]\|$ depending on the value of τ. Accordingly, the best choice of τ depends on the particular regimes of a.

Case 1: $\omega(\sqrt{m}) \leq a \leq o(m)$. Let

$$
\tau = \frac{\sqrt{m}}{2(a-1)} - \frac{m^{3/4}}{2\sqrt{2}(a-1)^{3/2}} - \frac{1}{\sqrt{m}}. \tag{92}
$$

The last two terms in (92) are lower order terms comparing to the first term; thus $\tau = (1+o(1))\frac{\sqrt{m}}{2(a-1)}$. It follows that for sufficiently large m, $\tau > 0$, $\tau = o(1)$, and $\tau = \omega(1/\sqrt{m})$.

We next show that Z is feasible for (6) with high probability; it suffices to verify (91). For $i < j$, $\mathbb{E}\left[g(W_{ij})\right] = 1$ and $\mathsf{var}\left(g(W_{ij})\right) = e^{\tau^2} - 1 = O(\tau^2)$. It follows from Chebyshev's inequality that

$$
\mathbb{P}\left\{\left|\sum_{i<j}\left(g(W_{ij}) - \mathbb{E}\left[g(W_{ij})\right]\right)\right| \geq \frac{1}{2}\sqrt{m}(m-1)\tau\right\} \leq \frac{2(e^{\tau^2}-1)}{(m-1)\tau^2} \to 0. \tag{93}
$$

Thus, with probability tending to one,

$$
|\alpha - 1| \leq \tau/\sqrt{m}. \tag{94}
$$

As a consequence, we have that

$$
\frac{\alpha(m-1)}{a-1} - 1 \geq \frac{m-1}{a-1} - \frac{\sqrt{m}\tau}{a-1} - 1 \geq \frac{m-1}{a-1} - \frac{m}{2(a-1)^2} - 1. \tag{95}
$$

Since $\tau \to 0$ and $\tau m \to \infty$, applying Lemma 31 yields that with probability tending to one,

$$
\|g(W) - \mathbb{E}\left[g(W)\right]\| \leq 2\sqrt{m}\tau + 2\sqrt{m}\tau^{3/2}. \tag{96}
$$

Plugging in the definition of τ given in (92), (96) implies that with probability converging to one,

$$
\|g(W) - \mathbb{E}\left[g(W)\right]\| \leq \frac{m}{a-1} - 2. \tag{97}
$$

Since by assumption $a = \omega(\sqrt{m})$ and $a = o(m)$, combining (95) and (97) yields that with probability tending to one, (91) and hence $Z \succeq 0$ hold.

Finally, we compute the value of the objective function $\langle Z, W \rangle$. For $i < j$, $\mathbb{E}\left[W_{ij}g(W_{ij})\right] = \tau$ and $\mathbb{E}\left[W_{ij}^2 g^2(W_{ij})\right] = e^{\tau^2}(1 + 4\tau^2) = O(1)$. Thus, it follows from Chebyshev's inequality (and $a = \omega(m)$) that:

$$\mathbb{P}\left\{\left|\sum_{i<j}(W_{ij}g(W_{ij}) - \tau)\right| \geq \frac{m(m-1)}{2}\left[\frac{1}{\sqrt{m}} - \frac{1}{a-1}\right]\right\} = O\left(\frac{1}{m}\right). \tag{98}$$

In view of (92) and (94), with probability tending to one,

$$\begin{aligned}
\langle Z, W \rangle &= \frac{2(a-1)}{\alpha m(m-1)}\sum_{i<j}W_{ij}g(W_{ij}) \\
&\geq \frac{(a-1)\tau}{\alpha} - \frac{a}{\alpha\sqrt{m}} + \frac{1}{\alpha} \\
&\geq \left\{(a-1)\tau - \frac{a}{\sqrt{m}} + 1\right\}\left(1 - \frac{\tau}{\sqrt{m}}\right) \\
&\geq \left\{\frac{\sqrt{m}}{2} - \frac{m^{3/4}}{\sqrt{8(a-1)}} - \frac{2a}{\sqrt{m}} + 1\right\}\left(1 - \frac{1}{2(a-1)}\right) \\
&\geq \frac{\sqrt{m}}{2} - \frac{m^{3/4}}{\sqrt{8(1-a)}} - \frac{2a}{\sqrt{m}},
\end{aligned}$$

proving the first part of (59).

Case 2: $a = o(\sqrt{m})$. The desired lower bound given in the third part of (59) is trivially true for $a = 1$ so we suppose $a \geq 2$. The proof is almost identical to the first case except that we set

$$\tau = \sqrt{\frac{1}{3}\left(\log\frac{m}{a^2} - \log\log\frac{m}{a^2}\right)}. \tag{99}$$

First, we verify that (91) holds with high probability. By the choice of τ, $e^{\tau^2} = o(m^{1/3})$. Thus, (93) and hence (94) continue to hold. It follows from (94) that $\alpha = 1 + O_P\left(\frac{\log(m/a^2)}{\sqrt{m}}\right)$. Applying Lemma 31 and Markov's inequality, with probability at least $1 - (\log\frac{m}{a^2})^{-1/4}$,

$$\|g(W) - \mathbb{E}\left[g(W)\right]\| \leq C\left(\log\frac{m}{a^2}\right)^{1/4}\sqrt{m(e^{3\tau^2} - 1)}.$$

Plugging in the definition of τ given in (99), it further implies that with high probability,

$$\|g(W) - \mathbb{E}\left[g(W)\right]\| \leq C\frac{m}{a}\left(\log\frac{m}{a^2}\right)^{-1/4}.$$

Therefore (91) holds with high probability.

Then we compute the value of the objective function $\langle Z, W \rangle$. Entirely analogously to (98), we have

$$\mathbb{P}\left\{ \left| \sum_{i<j} (W_{ij} g(W_{ij}) - \tau) \right| \geq \frac{\tau m(m-1)}{2m^{1/3}} \right\} \leq \frac{2m^{2/3} e^{\tau^2}(1 + 4\tau^2)}{m(m-1)\tau^2} \to 0.$$

Therefore with probability tending to one,

$$\langle Z, W \rangle = \frac{2(a-1)}{\alpha m(m-1)} \sum_{i<j} W_{ij} g(W_{ij}) \geq (a-1)\tau \left(1 - O\left(\frac{\log(m/a^2)}{\sqrt{m}} \right) \right) \left(1 - m^{-1/3} \right).$$

By the choice of τ given in (99), we have that

$$\tau \geq \sqrt{\frac{1}{3} \log \frac{m}{a^2}} \left(1 - O\left(\frac{\log\log(m/a^2)}{\log(m/a^2)} \right) \right).$$

Combining the last two displayed equations yield that with high probability,

$$\langle Z, W \rangle \geq (a-1)\sqrt{\frac{1}{3} \log \frac{m}{a^2}} - O\left(\frac{a \log\log(m/a^2)}{\sqrt{\log(m/a^2)}} \right),$$

proving the desired lower bound to $V_m(a)$ given in the third part of (59).

Case 3: $a = \Theta(\sqrt{m})$. Let τ be a constant to be chosen later. The proof is similar to the previous two cases; the key difference is that the distributions of entries of $g(W)$ are independent of m, and thus we can the invoke Lemma 27, a corollary of the Bai-Yin theorem, instead of Lemma 31, to obtain

$$\|g(W) - \mathbb{E}[g(W)]\| = 2\sqrt{m(e^{\tau^2} - 1)}(1 + o_P(1)).$$

In view of (91), as long as τ is chosen to be a constant so that

$$0 < \tau < \liminf_{m \to \infty} \sqrt{\log\left(1 + \frac{m}{4a^2} \right)},$$

we have $Z \succeq 0$ with high probability.

Finally, we compute the value of the objective function $\langle Z, W \rangle$. Entirely analogously to (98), we have

$$\mathbb{P}\left\{ \left| \sum_{i<j} (W_{ij} g(W_{ij}) - \tau) \right| \geq \frac{m(m-1)}{2a} \right\} = O(1/m) \to 0.$$

It follows that with probability converging to 1,

$$\langle Z, W \rangle \geq \frac{(a-1)\tau - 1}{\alpha} \geq ((a-1)\tau - 1)\left(1 - O(m^{-1/2}) \right) = a\tau + O(1),$$

which yields the desired lower bound to $V_m(a)$. ∎

Appendix F. Proof of Lemma 23

The proof follows the same fashion as that in the Gaussian case. In particular, to prove the desired lower bound to $V_m(a)$, we construct an explicit feasible solution Z to (6); however, the particular construction is different. Recall that in the Bernoulli case, M is assumed to be an $m \times m$ symmetric random matrix with zero diagonal and independent entries such that $M_{ij} = M_{ij} \sim \mathrm{Bern}(q)$ for all $i < j$.

Let

$$R = \frac{\langle M, \mathbf{J} \rangle}{m(m-1)}$$

and assume that $R \in (0, 1)$ for the time being. For a given $\gamma \in (0, 1]$, define

$$\alpha = \frac{\gamma - R}{R(1-R)} \frac{(a-1)}{m(m-1)}, \quad \beta = \frac{1-\gamma}{1-R} \frac{(a-1)}{m(m-1)} \geq 0.$$

Define an $m \times m$ matrix Z by $Z_{ii} = 1/m$ and $Z_{ij} = \alpha M_{ij} + \beta$ for $i \neq j$. By definition, $\langle Z, \mathbf{I} \rangle = 1$, $\alpha + \beta = \frac{\gamma(a-1)}{Rm(m-1)} \geq 0$, and thus $Z \geq 0$. Moreover,

$$\langle Z, \mathbf{J} \rangle = \alpha \langle M, \mathbf{J} \rangle + \beta \langle \mathbf{J} - \mathbf{I}, \mathbf{J} \rangle + \frac{1}{m} \langle \mathbf{I}, \mathbf{J} \rangle = a$$

and

$$\langle Z, M \rangle = (\alpha + \beta) \langle M, \mathbf{J} \rangle = (a-1)\gamma. \tag{100}$$

Thus to get a lower bound to $V_m(a)$ as tight as possible, we would like to choose γ as large as possible to satisfy $Z \succeq 0$ with high probability.

Note that

$$Z = \alpha M + \beta(\mathbf{J} - \mathbf{I}) + (1/m)\mathbf{I} = (\beta + \alpha q)(\mathbf{J} - \mathbf{I}) + (1/m)\mathbf{I} + \alpha(M - \mathbb{E}[M]).$$

Thus, to show $Z \succeq 0$, it suffices to verify that

$$\frac{1}{m} - \beta - \alpha q - \|M - \mathbb{E}[M]\| \geq 0.$$

By Lemma 30, with high probability, $\|M - \mathbb{E}[M]\| \leq \kappa \sqrt{mq(1-q)}$, where κ is a universal positive constant defined in (23). Hence, to show $Z \succeq 0$, it further suffices to verify that

$$\frac{1}{m} - \beta - \alpha q - \kappa \alpha \sqrt{mq(1-q)} \geq 0. \tag{101}$$

As a result, we would like to choose $\gamma \in (0, 1]$ as large as possible to satisfy (101). We pause to give some intuitions on the choice of γ. By concentration inequalities, $R \approx q$ with high probability. Since $a = o(m)$, $\beta = o(1/m)$. Furthermore, $q \ll \sqrt{mq(1-q)}$. Hence, to satisfy (101), roughly it suffices that

$$\alpha \leq \frac{1}{\kappa \sqrt{mq(1-q)}m},$$

which further implies that

$$\gamma \leq q + \frac{\sqrt{mq(1-q)}}{\kappa(a-1)}.$$

This suggests that we should take γ to be the minimum of $q + \frac{\sqrt{mq(1-q)}}{\kappa(a-1)}$ and 1.

Before specifying the precise choice of γ, we first show that R is close to q with high probability. Let $c_m = \log(m\sqrt{q})$ which converges to infinity under the assumption that $m^2 q \to \infty$. Thus, by the Chernoff bound for the binomial distribution, with probability converging to 1, $|R - q| \le c_m\sqrt{q}/m$. Without loss of generality, we can and do assume that $|R - q| \le c_m\sqrt{q}/m$ in the remainder of the proof. Since q is bounded away from 1 and $m^2 q \to \infty$, R is also bounded away from 1 and $R > 0$. This verifies that α, β and hence Z are well-defined.

Let

$$\gamma = \begin{cases} q + (1-\epsilon)\frac{\sqrt{mq(1-q)}}{\kappa(a-1)} & a - 1 \ge \frac{1-\epsilon}{\kappa}\sqrt{\frac{mq}{1-q}} \\ 1 & 0 \le a - 1 \le \frac{1-\epsilon}{\kappa}\sqrt{\frac{mq}{1-q}}, \end{cases} \tag{102}$$

where $\epsilon = 2/\log\left(m\min\{\sqrt{q}, 1/a\}\right)$. Equivalently,

$$\gamma = \min\left\{q + (1-\epsilon)\frac{\sqrt{mq(1-q)}}{\kappa(a-1)}, 1\right\}. \tag{103}$$

The assumptions, $m^2 q \to \infty$ and $a = o(m)$, imply that $\epsilon = o(1)$ and hence $\gamma \in [q, 1]$.

Next, we compute the value of $\langle Z, M \rangle$. In view of (100), it suffices to evaluate $(a-1)\gamma$. By the choice of γ,

$$(a-1)\gamma = \begin{cases} (a-1)q + (1-\epsilon)\frac{\sqrt{mq(1-q)}}{\kappa} & a - 1 \ge \frac{1-\epsilon}{\kappa}\sqrt{\frac{mq}{1-q}} \\ a - 1 & 0 \le a - 1 \le \frac{1-\epsilon}{\kappa}\sqrt{\frac{mq}{1-q}} \end{cases}. \tag{104}$$

Since $\epsilon = o(1)$, absorbing the factor $1 - \epsilon$ in the last displayed equation into the definition of κ given in (23) yields the desired lower bound to $V_m(a)$.

To finish the proof, we are left to verify (101). Since $\beta + \alpha R = \frac{a-1}{m(m-1)}$, it follows that

$$\frac{1}{m} - \beta - \alpha q = \frac{1}{m} - \beta - \alpha R - \alpha(q - R) = \frac{m-a}{m(m-1)} - O\left(\frac{(a-1)\gamma c_m}{m^3\sqrt{q}}\right), \tag{105}$$

where we used the fact that $|R - q| \le c_m\sqrt{q}/m$ and $\alpha \le a\gamma/m^2 R$ in the last equality.

Let $\alpha_0 = \frac{\gamma - q}{q(1-q)}\frac{(a-1)}{m(m-1)}$. Next, we bound $|\alpha - \alpha_0|$ from the above. In view of $|R - q| \le c_m\sqrt{q}/m$ and $\gamma \ge q$,

$$\left|\frac{\gamma - R}{R(1-R)} - \frac{\gamma - q}{q(1-q)}\right| \le \left|\frac{\gamma - R}{R(1-R)} - \frac{\gamma - q}{R(1-R)}\right| + \left|\frac{\gamma - q}{R(1-R)} - \frac{\gamma - q}{q(1-q)}\right|$$

$$\le \frac{|R - q|}{R(1-R)} + (\gamma - q)\frac{|R - q||R + q - 1|}{R(1-R)q(1-q)}$$

$$= O\left(\frac{c_m}{m\sqrt{q}}\right) + O\left(\frac{c_m\gamma}{mq^{3/2}}\right) = O\left(\frac{c_m\gamma}{mq^{3/2}}\right)$$

Consequently,

$$\alpha - \alpha_0 = O\left(\frac{(a-1)\gamma c_m}{m^3 q^{3/2}}\right). \tag{106}$$

Combining (105) and (106) yields that

$$\frac{1}{m} - \beta - \alpha q - \kappa\alpha\sqrt{mq(1-q)}$$

$$= \frac{1}{m} - \beta - \alpha q - \kappa\alpha_0\sqrt{mq(1-q)} - (\alpha - \alpha_0)\kappa\sqrt{mq(1-q)}$$

$$= \frac{1}{m(m-1)}\left(m - a - \frac{(a-1)(\gamma-q)}{q(1-q)}\kappa\sqrt{mq(1-q)} - O\left(\frac{(a-1)\gamma c_m}{\sqrt{mq}}\right)\right). \qquad (107)$$

Thus, to verify (101), it reduces to show the right hand side of the last displayed equation is negative. In view of (103),

$$\frac{(a-1)(\gamma-q)}{q(1-q)}\kappa\sqrt{mq(1-q)} \le (1-\epsilon)m.$$

and

$$\frac{(a-1)c_m\gamma}{\sqrt{mq}} \le \frac{(a-1)c_m}{\sqrt{m}} + \frac{c_m}{\kappa\sqrt{q}} = o\left(\frac{m}{\log(m\sqrt{q})}\right),$$

where the last equality because $c_m = \log(m\sqrt{q})$ and the assumption that $a = o(m)$. Combining the last two displayed equations and plugging in the definition of ϵ yield that

$$m - a - \frac{(a-1)(\gamma-q)}{q(1-q)}\kappa\sqrt{mq(1-q)} - O\left(\frac{(a-1)\gamma c_m}{\sqrt{mq}}\right)$$

$$\ge \frac{2m}{\log\left(m\min\{\sqrt{q}, 1/a\}\right)} - a - o\left(\frac{m}{\log(m\sqrt{q})}\right) \ge 0.$$

Hence, it follows from (107) that (101) holds. Consequently, $Z \succeq 0$ holds with high probability. This completes the proof of the lemma.

Appendix G. Proof of (61)

Note that for each $i \in C$, $X_i \triangleq \sum_{j\in C} W_{ij}$ is distributed according to $\mathcal{N}(0, K-1)$ but not independently. Below we use the Chung-Erdös inequality Chung and Erdös (1952):

$$\mathbb{P}\left\{\bigcup_{i=1}^K A_i\right\} \ge \frac{\left(\sum_{i=1}^K \mathbb{P}\{A_i\}\right)^2}{\sum_{i=1}^K \mathbb{P}\{A_i\} + \sum_{i\neq j}\mathbb{P}\{A_iA_j\}}. \qquad (108)$$

For any $i \neq j$, $\mathbb{P}\{X_i \le -s, X_j \le -s\} = \mathbb{E}[Q^2((s+Z)/\sqrt{K-2})] \triangleq \mathbb{E}[g^2(Z)]$ where $Z \sim \mathcal{N}(0,1)$, and $\mathbb{P}\{X_i \le -s\} = \mathbb{E}[g(Z)] = Q(s/\sqrt{K-1})$. Therefore

$$\mathbb{P}\{X_i \le -s, X_j \le -s\} - \mathbb{P}\{X_i \le -s\}\mathbb{P}\{X_j \le -s\}$$

$$= \mathsf{var}(g(Z)) = \mathbb{E}\left[\left(Q\left(\frac{s+Z}{\sqrt{K-2}}\right) - Q\left(\frac{s}{\sqrt{K-1}}\right)\right)^2\right]$$

$$\le Q(s/4) + \varphi\left(\frac{3s/4}{\sqrt{K-2}}\right)^2 \mathbb{E}\left[\left(\frac{s+Z}{\sqrt{K-2}} - \frac{s}{\sqrt{K-1}}\right)^2\right]$$

$$\le \exp(-s^2/8) + \exp\left(-\frac{9s^2/16}{K-2}\right)\left[\frac{1}{K-2} + \left(\frac{1}{\sqrt{K-1}} - \frac{1}{\sqrt{K-2}}\right)^2 s^2\right].$$

Let $s = \sqrt{K-1}(\sqrt{2\log K} - \log\log K/\sqrt{2\log K})$. Then $\mathbb{P}\{X_i \leq -s\} = \Theta(\sqrt{\log K}/K)$ and

$$\mathbb{P}\{X_i \leq -s, X_j \leq -s\} - \mathbb{P}\{X_i \leq -s\}\mathbb{P}\{X_j \leq -s\} = O(K^{-17/8}).$$

Applying (108), we conclude that $\mathbb{P}\{\min_{i \in C} X_i \leq -s\} \geq 1 - O(1/\sqrt{\log K})$.

Appendix H. Proof of (68)

We show $\mathbb{P}\{\mathcal{E}_1\} \to 1$. In this section, by a slight abuse of notation, let $e(i, S) = \sum_{j \in S} A_{ij}$. The proof is complicated by the fact the random variables $e(i, C^*)$ for $i \in C^*$ are not independent. The trick is to fix C^* and a small set $T \subset C^*$ with $|T| = K_o$. Then for $i \in T$, $e(i, C^*) = e(i, C^* \backslash T) + e(i, T)$, and we can make use of the fact that the random variables $(e(i, C^* \backslash T) : i \in T)$ are independent, $(e(i, C^* \backslash T) : i \in T)$ is independent of $(e(i, T) : i \in T)$, and, with high probability, at least half of the random variables in $e(i, T)$ are not unusually large. (The same trick is used for proving Theorem 6 in Hajek et al. (2015e).)

Suppose for convenience of notation that C^* consists of the first K indices, and T consists of the first K_o indices: $C^* = [K]$ and $T = [K_o]$. Let $T' = \{i \in T : e(i, T) \leq (K_o - 1)p + 6\sigma\}$, Since[6]

$$\min_{i \in C^*} e(i, C^*) \leq \min_{i \in T'} e(i, C^*) \leq \min_{i \in T'} e(i, C^* \backslash T) + (K_o - 1)p + 6\sigma,$$

it follows that

$$\mathbb{P}\{E_1\} \geq \mathbb{P}\left\{\min_{j \in T'} e(j, C^* \backslash T) \leq (K - K_o)\tau_1'\right\}.$$

We show next that $\mathbb{P}\left\{|T'| \geq \frac{K_o}{2}\right\} \to 1$ as $n \to \infty$. For $i \in T$, $e(i, T) = X_i + Y_i$ where $X_i = e(i, \{1, \ldots, i-1\})$ and $Y_i = e(i, \{i+1, \ldots, K_o\})$. The X's are mutually independent, and the Y's are also mutually independent, and X_i has the $\mathrm{Binom}(i-1, p)$ distribution and Y_i has the $\mathrm{Binom}(K_o - i, p)$ distribution. Then $\mathbb{E}[X_i] = (i-1)p$ and $\mathrm{var}(X_i) \leq \sigma^2$. Thus, by the Chebyshev inequality, $\mathbb{P}\{X_i \geq (i-1)p + 3\sigma\} \leq \frac{1}{9}$ for all $i \in T$. Therefore, $|\{i : X_i \leq (i-1)p + 3\sigma\}|$ is stochastically at least as large as a $\mathrm{Binom}\left(K_o, \frac{8}{9}\right)$ random variable, so that, $\mathbb{P}\left\{|\{i : X_i \leq (i-1)p + 3\sigma\}| \geq \frac{3K_o}{4}\right\} \to 1$ as $K_o \to \infty$ (which happens as $n \to \infty$). Similarly, $\mathbb{P}\left\{|\{i : Y_i \leq (K_o - i)p + 3\sigma\}| \geq \frac{3K_o}{4}\right\} \to 1$. If at least 3/4 of the X's are small and at least 3/4 of the Y's are small, it follows that at least 1/2 of the $e(i, T)$'s for $i \in T$ are small. Therefore, $\mathbb{P}\left\{|T'| \geq \frac{K_o}{2}\right\} \to 1$ as claimed.

The set T' is independent of $(e(i, C^* \backslash T) : i \in T)$ and those variables each have the $\mathrm{Binom}(K - K_o, p)$ distribution. Using the tail lower bound (66), we have

$$\mathbb{P}\{E_1\} \geq 1 - \mathbb{E}\left[\prod_{j \in T'} \mathbb{P}\{e(j, C^* \backslash T) \geq K\tau^* - K_o p - 6\sigma\} \,\Big|\, |T'| \geq \frac{K_o}{2}\right] - \mathbb{P}\left\{|T'| < \frac{K_o}{2}\right\}$$

$$\geq 1 - \exp\left(-Q\left(\sqrt{2(K - K_o)d(\tau_1' \| p)}\right)K_o/2\right) - o(1).$$

---

6. In case $T' = \emptyset$ we use the usual convention that the minimum of an empty set of numbers is $+\infty$.

By definition of τ_1' and the convexity of divergence, $d(\tau_1'\|p) \leq (1-\delta)d(\tau_1\|p)$, it follows that

$$Q\left(\sqrt{2(K-K_o)d(\tau_1'\|p)}\right)K_o/2 \geq Q\left(\sqrt{2(K-K_o)(1-\delta)d(\tau_1\|p)}\right)K_o/2$$

$$\geq Q\left(\sqrt{2(1-\delta)\log K}\right)K_0/2 \geq \frac{\sqrt{\log K}}{2},$$

and $\mathbb{P}\{E_1\} \to 1$.