The Annals of Statistics
2018, Vol. 46, No. 1, 149–179
https://doi.org/10.1214/17-AOS1545

ON SEMIDEFINITE RELAXATIONS FOR THE BLOCK MODEL[1]

BY ARASH A. AMINI AND ELIZAVETA LEVINA

University of California, Los Angeles and University of Michigan

The stochastic block model (SBM) is a popular tool for community detection in networks, but fitting it by maximum likelihood (MLE) involves a computationally infeasible optimization problem. We propose a new semidefinite programming (SDP) solution to the problem of fitting the SBM, derived as a relaxation of the MLE. We put ours and previously proposed SDPs in a unified framework, as relaxations of the MLE over various subclasses of the SBM, which also reveals a connection to the well-known problem of sparse PCA. Our main relaxation, which we call SDP-1, is tighter than other recently proposed SDP relaxations, and thus previously established theoretical guarantees carry over. However, we show that SDP-1 exactly recovers true communities over a wider class of SBMs than those covered by current results. In particular, the assumption of strong assortativity of the SBM, implicit in consistency conditions for previously proposed SDPs, can be relaxed to weak assortativity for our approach, thus significantly broadening the class of SBMs covered by the consistency results. We also show that strong assortativity is indeed a necessary condition for exact recovery for previously proposed SDP approaches and not an artifact of the proofs. Our analysis of SDPs is based on primal-dual witness constructions, which provides some insight into the nature of the solutions of various SDPs. In particular, we show how to combine features from SDP-1 and already available SDPs to achieve the most flexibility in terms of both assortativity and block-size constraints, as our relaxation has the tendency to produce communities of similar sizes. This tendency makes it the ideal tool for fitting network histograms, a method gaining popularity in the graphon estimation literature, as we illustrate on an example of a social networks of dolphins. We also provide empirical evidence that SDPs outperform spectral methods for fitting SBMs with a large number of blocks.

1. Introduction. Community detection, one of the fundamental problems in network analysis, has attracted a lot of attention in a number of fields, including computer science, statistics, physics and sociology. The stochastic block model (SBM) [30] is a well established and widely used model for community detection, attractive for its analytical tractability and connections to fundamental properties of random graphs [6, 15, 42], but fitting it to data is a challenge due to the need to

Received January 2016; revised November 2016.

[1]Supported in part by NSF Grants DMS-11-06772, DMS-11-59005 and DMS-15-21551.

MSC2010 subject classifications. 62G20, 90C22, 62H99.

Key words and phrases. Community detection, network, semidefinite programming, stochastic block model.

optimize over K^n assignments of n nodes to K communities. Many fitting methods have been proposed, including profile likelihood [15], MCMC [45, 51], variational approaches [4, 14, 18], belief propagation [24] and pseudo-likelihood [8], the latter two being more or less the current state of the art in speed and accuracy. However, all these methods rely on a good initial value and can be sensitive to starting points. In contrast, spectral clustering methods do not require an initial value, are fast and have also been popular in community detection [19, 35, 49, 50]. Spectral clustering works reasonably well in dense networks with balanced communities but fails on sparse networks [34]. Regularization can help [8, 19, 32], but even regularized spectral clustering does not achieve the accuracy of likelihood-based methods when they are given a good initial value [8].

Recently, semidefinite programming (SDP) approaches to fitting the SBM have appeared in the literature [17, 20, 21], which rely on a SDP relaxation of the computationally infeasible likelihood optimization problem. They are attractive because, on one hand, they solve a global optimization problem and require no initial value, and on the other hand, they are still maximizing the likelihood, and one can therefore hope for better performance than from generic methods that do not use the likelihood in any way. As global optimization methods, they are easier to analyze than iterative methods depending on a starting value. It also appears that SDP relaxations in themselves have a regularization effect, which makes their solutions more robust to noise and outliers (see Remark 3.1). One drawback of SDP methods is the higher computational cost of SDP solvers. However, by formulating the problem as a SDP, we can benefit from continuous advances in solving large scale SDPs, an active area of research in optimization.

In this paper, we propose a new SDP relaxation of the likelihood optimization problem, which is tighter than any of the previously proposed SDP relaxations [17, 20, 21]. We also put all these relaxations into a unified framework, by viewing them as versions of the MLE restricted to different parameter spaces, and show their connection to the well-studied problem of sparse PCA. Empirically, the tighter relaxation gives better results, and we derive a first-order SDP implementation via ADMM which keeps computing costs reasonable.

On the theoretical side, our focus for the most part will be on balanced models, that is, those with equal community sizes. We obtain sufficient conditions on the parameters of the block model for strong consistency (i.e., exact recovery of communities) of our relaxation, SDP-1. These conditions guarantee success over a wider class of SBMs than in previous literature. Current conditions for the success of SDP relaxations implicitly impose what we will call strong assortativity, whereas our SDP succeeds for any weakly assortative SBM (cf. Definition 4.1), when the expected degree grows as $\Omega(\log n)$. We also show that the requirement of strong assortativity is necessary for the success of previous SDP relaxations (SDP-2 and SDP-3 in Table 1), and it is not an artifact of proof techniques (Section 5). Our proof of the success of SDP-1 is based on a primal-dual witness construction

TABLE 1
SDP relaxations

	SDP-1	**SDP-2**	**SDP-3**	**EVT**
Maximize	$\langle A, X\rangle$	$\langle A, X\rangle$	$\langle A, X\rangle - \lambda\langle \boldsymbol{E}_n, X\rangle$	$\langle A, X\rangle$
Subject to	$X\mathbf{1}_n = (n/K)\mathbf{1}_n$	$\langle \boldsymbol{E}_n, X\rangle = n^2/K$		
	$\operatorname{diag}(X) = \mathbf{1}_n$	$\operatorname{tr}(X) = n$		$\operatorname{tr}(X) = n$
	$X \geq 0$	$0 \leq X \leq 1$		$\|\|\|X\|\|\|_2 \leq n/K$
		$X \succeq 0$		
Model	$\mathrm{PP}^{\mathrm{bal}}(p,q) \equiv \mathcal{X}_{\mathrm{orbit}}(X_0)$		$\mathrm{PP}(p,q) \equiv \mathcal{X}_{\mathrm{free}}$	

which has already been used successfully in the context of sparse recovery problems; see, for example, [10, 54]. In the context of SDP relaxations for the SBM, however, the only instance of this approach that we know of is the recent work of [1], for the case of the $K = 2$ SBM. Our approach can be viewed as a nontrivial extension of [1] to the case of general K, and a more complex SDP with the doubly nonnegative cone constraint and more equality constraints. As a by-product, we also recover the current results for SDP-2 for the class of strongly assortative SBMs.

SDP-1, in its basic form, tends to partition the network into blocks of similar sizes. This is sometimes an unwelcome feature in practice, and sometimes a desirable one, since very large and very small communities are generally difficult to interpret. If this feature is not desirable, SDP-1 can be modified to allow for different block sizes, as discussed in Section 6. The equal sized blocks are especially suitable for constructing network histograms, a method for graphon estimation proposed by [46]. Viewing the SBM as a nonparametric approximation to a general reasonably smooth mean function of the adjacency matrix (the graphon) is analogous to constructing a histogram to approximate a general smooth density function. A number of methods for graphon estimation have been proposed recently [5, 55, 58], and the network histogram as a graphon estimator has been proposed in [46]. A histogram is appealing because it is controlled by the number of bins (blocks) K, which is a single parameter that can be chosen to balance fitting the data with robustness to noise. In this case, it is particularly appropriate to fit blocks of equal or similar sizes, just like in the usual histogram. We show empirically in Section 8 that our SDP relaxation provides the best tool for histogram estimation, as well as generally cleaner solutions, compared to other less tight SDP relaxations and generic methods like spectral clustering.

The rest of the paper is organized as follows. In Section 2, we introduce the SBM and its submodels. We derive a general blueprint for MLE relaxations in Section 3, introduce our proposed SDP and compare with the ones existing in the literature, including a brief discussion of the connection with sparse PCA. Section 4 presents our consistency results for balanced block models, along with an overview of the proofs. A result showing the failure of SDP-2 in the absence of strong assortativity (which is not needed for our relaxation) appears in Section 5. Extension to the case of unbalanced communities is discussed in Section 6. Section 7 presents application of SDP-1 to graphon estimation via fitting network histograms. Section 8 compares several SDPs numerically, and we conclude with a discussion in Section 9. Technical details of the proofs and a brief discussion of a first-order method for implementing SDP-1 can be found in the supplementary material [9].

Notation. We use $\otimes$ to denote Kronecker product of matrices, and $\circ$ to denote Schur (element-wise) product of matrices. $\mathbb{S}^n$ denotes the set of symmetric $n \times n$ matrices, and $\langle A, X\rangle := \operatorname{tr}(AX)$ the corresponding inner product. $(\mathbb{S}^n_+, \succeq)$ is the cone of positive semidefinite (PSD) $n \times n$ matrices, and its natural partial order, namely, $A \succeq B$ iff $A - B \in \mathbb{S}^n_+$. $\boldsymbol{E}_{n,m}$ is the $n \times m$ matrix of all ones and $\boldsymbol{E}_n$ is the $n \times n$ matrix of all ones. A $n \times 1$ vector of all ones is denoted $\mathbf{1}_n$. $\|u\| = \|u\|_2$ is the ℓ_2 norm of vector u, and $|||A||| = |||A|||_2$ is the $\ell_2 \to \ell_2$ operator norm of matrix A. $\ker(A)$ and $\operatorname{range}(A)$ denote the kernel (null space) and the range (column space) of matrix A. $\operatorname{diag} : \mathbb{R}^{n\times n} \to \mathbb{R}^n$ acts on square matrices and extracts the diagonal. $\operatorname{diag}^* : \mathbb{R}^n \to \mathbb{R}^{n\times n}$ is the adjoint of diag, acting on vectors, producing the natural diagonal matrix. For a matrix X, let $\operatorname{supp}(X) := \{(i, j) : X_{ij} \neq 0\}$ be its support. More specialized notation is introduced in Section 4.2.

2. The stochastic block model.

We now formally introduce the SBM. The network data (nodes and edges connecting them) are represented by a simple undirected graph on n nodes via its $n \times n$ adjacency matrix A, a binary symmetric matrix with $A_{ij} = 1$ if there is an edge between nodes i and j, and 0 otherwise. Each node belongs to exactly one community, specified by its membership vector $z_i \in \{0, 1\}^K$, with exactly one nonzero entry, $z_{ik} = 1$, indicating that node i belongs to community k. The vectors z_i are not observed. The SBM is parametrized through the symmetric probability matrix $\Psi \in [0, 1]^{K\times K}$, where Ψ_{kr} is the probability of an edge forming between a pair of nodes from communities k and r. For simplicity, we assume n is a multiple of K.

Given z_i and Ψ, $\{A_{ij}, i < j\}$ are drawn independently as Bernoulli random variables with $\mathbb{E}[A_{ij}|z_i, z_j] = z_i^T \Psi z_j$. Let Z be the $n \times K$ matrix with rows $z_1^T, \dots, z_n^T$. Then we can write the model as

$$M_Z := \mathbb{E}[A|Z] = Z\Psi Z^T. \tag{2.1}$$

Note that A_{ii}'s are so far undefined. They can be defined based on convenience, but we will always assume that they are defined so that (2.1) holds over all elements (e.g., one possibility is to set $A_{ii} := [M_Z]_{ii}$). We do not treat $\{A_{ii}\}$ as part of the observed data.

M_Z is a block constant, rank K matrix and we can think of the operation $\Psi \mapsto Z\Psi Z^T$ as a block constant embedding of a $K \times K$ matrix into the space of $n \times n$

matrices. This provides us with a simple but useful property: for any matrix M and function f on $\mathbb{R}$, let $f \circ M$ be the pointwise application of f to the entries of M, $[f \circ M]_{ij} = f(M_{ij})$. Then we have

$$f \circ (Z\Psi Z^T) = Z(f \circ \Psi)Z^T. \tag{2.2}$$

Using (2.2), we can write the log-likelihood of the SBM in a compact form. First, note that

$$\begin{aligned}\ell(Z, \Psi) &= \sum_{i<j} A_{ij} \log[M_Z]_{ij} + (1 - A_{ij}) \log(1 - [M_Z]_{ij}) \\ &= \sum_{i<j} A_{ij}[f \circ M_Z]_{ij} + [g \circ M_Z]_{ij},\end{aligned}$$

where $f(x) := \log \frac{x}{1-x}$ and $g(x) := \log(1 - x)$ are functions on $[0, 1]$. Recall that for symmetric matrices A and B, we defined $\langle A, B\rangle := \operatorname{tr}(AB)$, and $\boldsymbol{E}_n$ is the $n \times n$ matrix of ones. Let $\langle A, B\rangle_0 := \langle A, B\rangle - \sum_i A_{ii} B_{ii}$, that is, the inner product defined through off-diagonal elements of A and B. Using (2.2),

$$\begin{aligned}2\ell(Z, \Psi) &= \langle A, f \circ M_Z\rangle_0 + \langle \boldsymbol{E}_n, g \circ M_Z\rangle_0 \\ &= \langle A, Z(f \circ \Psi)Z^T\rangle_0 + \langle \boldsymbol{E}_n, Z(g \circ \Psi)Z^T\rangle_0.\end{aligned} \tag{2.3}$$

In deriving the SDPs, our focus will be on the following two special cases of the SBM. In Section 4, we will show the SDPs work for larger classes of SBMs than those they are derived for.

(PP) The planted partition (PP) model, $\mathrm{PP}(p, q)$, defined by just two parameters p and q via

$$\Psi = q\boldsymbol{E}_K + (p - q)I_K, \tag{2.4}$$

where I_K is the $K \times K$ identity matrix, and following the PP literature we assume $p > q$. Note that (2.4) simply means that the diagonal elements are p and the off-diagonal elements are q.

($\mathrm{PP}^{\mathrm{bal}}$) The balanced planted partition model, $\mathrm{PP}^{\mathrm{bal}}(p, q)$, which is $\mathrm{PP}(p, q)$ with the additional assumption that the blocks have equal sizes.

For $\mathrm{PP}(p, q)$, the likelihood greatly simplifies, since $f \circ \Psi$ and $g \circ \Psi$ take only two values,

$$f \circ \Psi = f(q)\boldsymbol{E}_K + [f(p) - f(q)]I_K$$

and similarly for $g \circ \Psi$. Since $Z\boldsymbol{E}_K Z^T = \boldsymbol{E}_n$, (2.3) becomes

$$2\ell(Z, \Psi) = [f(p) - f(q)]\langle A, ZZ^T\rangle_0 + [g(p) - g(q)]\langle \boldsymbol{E}_n, ZZ^T\rangle_0 + \text{const},$$

where the constant term does not depend on Z. With the condition $p > q$, we have $f(p) > f(q)$ and $g(p) < g(q)$. Then we obtain

$$\begin{aligned}&\frac{2\ell(Z,\Psi)}{f(p)-f(q)}\\ &\quad=\langle A, ZZ^T\rangle - \lambda\langle \boldsymbol{E}_n, ZZ^T\rangle + \text{const}, \qquad \lambda := \frac{g(q)-g(p)}{f(p)-f(q)} > 0.\end{aligned} \tag{2.5}$$

Note that we have safely replaced $\langle\cdot,\cdot\rangle_0$ with $\langle\cdot,\cdot\rangle$, possibly changing the constant, since $[ZZ^T]_{ii} = 1, \forall i$ regardless of Z. A similar calculation appears in [17], albeit in a slightly different form.

3. Relaxing the maximum likelihood estimator (MLE).

Given the adjacency matrix A, the MLE for (Z, Ψ) is obtained by maximizing the likelihood of the SBM. It is known to have desirable consistency and in some sense optimality properties [15], but the exact computation of the MLE is in general NP-hard, due to the optimization over Z. However, it can be relaxed to computationally feasible convex problems.

We can obtain a class of MLEs by varying the domain over which the likelihood (2.5) is maximized. That is, we have the general estimator:

$$\widehat{Z} := \operatorname*{argmax}_{Z\in\mathcal{Z}} \langle A, ZZ^T\rangle - \lambda\langle \boldsymbol{E}_n, ZZ^T\rangle. \tag{3.1}$$

Each Z corresponds to a *clustering* matrix $X = ZZ^T \in \{0,1\}^{n\times n}$, where $X_{ij} = 1$ if i and j belong to the same community, and $X_{ij} = 0$ otherwise. Any subset $\mathcal{Z}$ in the Z-space induces a corresponding subset $\mathcal{X}$ in the X-space. We can consider estimators of X, and our blueprint for deriving different relaxations will be varying the space $\mathcal{X}$ in the optimization problem

$$\widehat{X} := \operatorname*{argmax}_{X\in\mathcal{X}} \langle A, X\rangle - \lambda\langle \boldsymbol{E}_n, X\rangle. \tag{3.2}$$

3.1. *Our relaxation*: *SDP*-1.

Our relaxation corresponds to the balanced model $\mathrm{PP}^{\mathrm{bal}}(p,q)$, in which each community is of size n/K. In this case, all admissible Z's can be obtained by permutation of any fixed admissible $Z_0 = I_K \otimes \mathbf{1}_{n/K}$, and we can take the feasible set $\mathcal{Z}$ in (3.1) to be

$$\mathcal{Z}_{\text{orbit}}(Z_0) := \{PZ_0Q : P, Q \text{ are permutation matrices}\},$$

where $\otimes$ is the Kronecker product and $\mathbf{1}_{n/K}$ is the vector of all ones of length n/K. This choice of Z_0 is for convenience and corresponds to assigning nodes consecutively to communities 1 through K. Recalling $X = ZZ^T$, the corresponding feasible set in the X-space is

$$\begin{aligned}&\mathcal{X}_{\text{orbit}}(X_0)\\ &\quad:= \{PX_0P^T : P \text{ is a permutation matrix}\}, \qquad X_0 = I_K \otimes \boldsymbol{E}_{n/K}.\end{aligned} \tag{3.3}$$

Note that X_0 is block-diagonal with all the diagonal blocks equal to $\boldsymbol{E}_{n/K}$.

In order to relax $\mathcal{X}_{\text{orbit}}(X_0)$, we first note that any X in this set is clearly positive semidefinite (PSD), denoted by $X \succeq 0$, since $X = (PZ_0)(PZ_0)^T$. In addition, $0 \le X_{ij} \le 1$ for all i, j, which we write as $0 \le X \le 1$, and $\text{diag}(X) = \mathbf{1}_n$. Note that the latter condition, $X \succeq 0$ and $X \ge 0$ imply $X \le 1$, since $1 - X_{ij}^2 = X_{ii}X_{jj} - X_{ij}^2 \ge 0$ implying $X_{ij} = |X_{ij}| \le 1$. Finally, it is easy to see that each row of X should sum to n/K, that is, $X\mathbf{1}_n = (n/K)\mathbf{1}_n$. This implies we can remove the term $\lambda\langle \boldsymbol{E}_n, X\rangle$ from the objective function in (3.2), since

$$\langle X, \boldsymbol{E}_n\rangle = \text{tr}(X\mathbf{1}_n\mathbf{1}_n^T) = \mathbf{1}_n^T X\mathbf{1}_n = \mathbf{1}_n^T(n/K)\mathbf{1}_n = n^2/K, \tag{3.4}$$

which is a constant. Thus, we arrive at our proposed relaxation, **SDP-1**:

$$\begin{aligned} &\text{argmax}_X \quad \langle A, X\rangle \\ &\text{subject to} \quad X\mathbf{1}_n = (n/K)\mathbf{1}_n, \qquad \text{diag}(X) = \mathbf{1}_n, \qquad X \succeq 0, X \ge 0. \end{aligned} \tag{3.5}$$

3.2. *Other relaxations*: *SDP*-2 *and SDP*-3. Two other interesting SDP relaxations have recently appeared in the literature. First, we will consider the relaxation of Chen and Xu [21]; see also [20]. They essentially work with the same $\text{PP}^{\text{bal}}(p, q)$, although their model is slightly more general (see Remark 3.1). The main relaxation proposed in [21] is via constraining the nuclear norm of X, a common heuristic for constraining the rank. Since X is PSD, we obtain $|||X|||_* = \text{tr}(X) = n$. In addition, they impose a single affine constraint, namely (3.4). Thus, their main focus is on the relaxation which replaces $\mathcal{X}_{\text{orbit}}(X_0)$ with $\{X : |||X|||_* \le n, \langle X, \boldsymbol{E}_n\rangle = n^2/K, 0 \le X \le 1\}$. However, they briefly mention a much tighter SDP relaxation which imposes positive semidefiniteness directly. This is what we have called **SDP-2**, shown in Table 1.

Note that $X \succeq 0$ and $\text{tr}(X) = n$ imply $|||X|||_* = n$, which is much tighter than $|||X|||_* \le n$. The main difference between SDP-2 and our relaxation is that we impose the constraint $\langle \boldsymbol{E}_n, X\rangle = n^2/K$ more restrictively, by breaking it into n separate affine constraints. We also break the $\text{tr}(X) = n$ into n pieces, but that does not seem to make much of a difference.

Next, we consider the relaxation of Cai and Li [17], though in a slightly different form. This relaxation works for the more general model $\text{PP}(p, q)$. In this case, we are looking at the feasible set

$$\mathcal{X}_{\text{free}} = \{X = ZZ^T : Z \text{ is an admissible membership matrix}\}. \tag{3.6}$$

For $X \in \mathcal{X}_{\text{free}}$, we still have $X \succeq 0$ and $X_{ij} \in \{0, 1\}$. Thus, one can simply relax to the problem denoted by **SDP-3** in Table 1.

Note that $\lambda\langle \boldsymbol{E}_n, X\rangle$ remains in the objective, since there are no constraints to make it constant. We cannot enforce an affine constraint involving $\langle \boldsymbol{E}_n, X\rangle$ directly for $\mathcal{X}_{\text{free}}$ without knowing the block sizes. In fact, let $\underline{\boldsymbol{n}} = (n_1, \dots, n_K)$ be the vector of block sizes, and let $\boldsymbol{E}_{\underline{\boldsymbol{n}}} := \text{diag}^*(\boldsymbol{E}_{n_1}, \dots, \boldsymbol{E}_{n_K})$ be the block-diagonal matrix with diagonal blocks of all ones with sizes given by $\underline{\boldsymbol{n}}$. It is easy to see that $\mathcal{X}_{\text{free}}$ is the union of orbits of all possible $\boldsymbol{E}_{\underline{\boldsymbol{n}}}$,

$$\mathcal{X}_{\text{free}} = \bigcup_{\underline{\boldsymbol{n}}:\|\underline{\boldsymbol{n}}\|_1=n} \mathcal{X}_{\text{orbit}}(\boldsymbol{E}_{\underline{\boldsymbol{n}}}) = \bigcup_{\|\underline{\boldsymbol{n}}\|_1=n} \{P\boldsymbol{E}_{\underline{\boldsymbol{n}}}P^T : P \text{ is a permutation matrix}\} \tag{3.7}$$

from which it follows that $\langle \boldsymbol{E}_n, X\rangle = \|\underline{\boldsymbol{n}}\|_2^2 = \sum_j n_j^2$, a function of the unknown $\{n_j\}$.

The optimal value for parameter λ, assuming the model is $\text{PP}(p,q)$, is given in (2.5) as a function of p and q. However, one can think of λ as a general regularization parameter controlling the sparseness of X, noticing $\langle \boldsymbol{E}_n, X\rangle = \|X\|_1$ since $X \geq 0$. It is well known that the ℓ_1 norm is a good surrogate for a cardinality constraint when enforcing sparseness, which leads us to a link to sparse PCA discussed in Section 3.3.

REMARK 3.1. Both [21] and [17] consider the effect of outliers on their SDPs. Cai and Li [17] derive the SDP for the model we described but they modify it by penalizing the trace, which is justified by their theory for a fairly general model of outliers. Chen and Xu [21] start with a generalized version of $\text{PP}^{\text{bal}}(p,q)$ which allows for a subset of nodes that belong to no community, and relax that model. Our relaxation SDP-1 can also work for this generalized model if we replace $X\mathbf{1}_n = (n/K)\mathbf{1}_n$ with the inequality version $X\mathbf{1}_n \leq (n/K)\mathbf{1}_n$. This has an advantage over Chen and Xu's approach, since one does not need to know the number of outliers a priori.

3.3. *Connection with nonnegative sparse PCA*. Representation (3.7) suggests another natural direction to restrict the parameter space. Note that $\|\underline{\boldsymbol{n}}\|_\infty = \max_j n_j \in [n/K, n]$, as a consequence of $\|\underline{\boldsymbol{n}}\|_1 = n$. The closer $\|\underline{\boldsymbol{n}}\|_\infty$ is to n/K, the more balanced the communities are. This suggests the following class:

$$\mathcal{X}_{\text{free}}^{\gamma} := \bigcup \{\mathcal{X}_{\text{orbit}}(\boldsymbol{E}_{\underline{\boldsymbol{n}}}) : \|\underline{\boldsymbol{n}}\|_1 = n, \|\underline{\boldsymbol{n}}\|_\infty \leq \gamma(n/K)\}, \tag{3.8}$$

where $\gamma \in [1, K]$ measures the deviation from completely balanced communities. For $X \in \mathcal{X}_{\text{free}}^{\gamma}$, note that $|||X|||_2 = |||\boldsymbol{E}_{\underline{\boldsymbol{n}}}|||_2 = \max_j |||\boldsymbol{E}_{n_j}|||_2 = \|\underline{\boldsymbol{n}}\|_\infty \leq \gamma(n/K)$. As before, we have $\text{tr}(X) = n$, $\|X\|_1 = \langle \boldsymbol{E}_n, X\rangle$, and $X \in \mathcal{N}_+^n := \{X : X \succeq 0, X \geq 0\}$, the doubly nonnegative cone. Letting $\widetilde{X} = (K/n)X$, we have

$$\begin{aligned} &\text{argmax}_{\widetilde{X}} \quad \langle A, \widetilde{X}\rangle - \lambda\|\widetilde{X}\|_1 \\ &\text{subject to} \quad |||\widetilde{X}|||_2 \leq \gamma, \qquad \text{tr}(\widetilde{X}) = K, \qquad \widetilde{X} \succeq 0, \widetilde{X} \geq 0. \end{aligned} \tag{3.9}$$

Apart from the nonnegative constraint $\widetilde{X} \geq 0$ (which can be removed to obtain a further relaxation), this is a generalization of the SDP relaxation for sparse PCA. Specifically, $\gamma = 1$ corresponds to the now well-known relaxation for recovering a sparse K-dimensional leading eigenspace of A. The corresponding solution $\widetilde{X}$

can be considered a generalized projection into this subspace (see, e.g., [22, 53]), and note that $\widetilde{X} \succeq 0$, $|||\widetilde{X}|||_2 \le 1$ is equivalent to $0 \preceq \widetilde{X} \preceq I$. We will not pursue this direction here, but it opens up possibilities for leveraging sparse PCA results in network models.

3.4. *Connection with adjacency-based spectral clustering*. The first step in spectral clustering based on the adjacency matrix is the truncation of A to its K largest eigenvalues, which we call eigenvalue truncation (EVT). The resulting matrix $\widetilde{X}$ is the solution of a SDP maximizing $\langle A, \widetilde{X}\rangle$ subject to $\widetilde{X} \succeq 0, \mathrm{tr}(\widetilde{X}) = K, |||\widetilde{X}|||_2 \le 1$. We can consider $X := (n/K)\widetilde{X}$ as an estimate of the cluster matrix by EVT. The resulting SDP appears in Table 1, and will be our surrogate to compare the other SDPs to this particular version of spectral clustering. We should note that the more common form of spectral clustering, based on truncation to K largest eignevalues in *absolute* value is equivalent to applying EVT to $|A| = \sqrt{A^2}$.

The SDP formulation of EVT can be considered a relaxation of the MLE in $\mathrm{PP}^{\mathrm{bal}}$, similar to the other SDPs we have considered. It is enough to note that $|||X|||_2 = |||X_0|||_2 = n/K$ for any $X \in \mathcal{X}_{\mathrm{orbit}}(X_0)$. Also note that SDP-1 is a strictly tighter relaxation than EVT. To see that, take any X which is feasible for SDP-1, and note that $X\mathbf{1}_n = (n/K)\mathbf{1}_n$ means that $\mathbf{1}_n$ is an eigenvector of X associated with eigenvalue n/K. The Perron–Frobenius theorem then implies that $|||X|||_2 \le n/K$; hence, X is feasible for EVT.

4. Strong consistency results. In this section, we provide consistency results for SDP-1 and a variant of SDP-2, which we will call SDP-2′. This version is obtained from SDP-2 by replacing $\mathrm{tr}(X) = n$ with $\mathrm{diag}(X) = \mathbf{1}_n$ and removing the now redundant condition $X \le 1$. This modification allows us to unify the treatment of these two SDPs. For example, optimality conditions in Section 4.2.2 are derived for a general blueprint (4.9), which includes both SDP-1 and SDP-2′ as special cases. The consistency results will go beyond the $\mathrm{PP}^{\mathrm{bal}}(p, q)$ model originally used in deriving them. Consider a general balanced block model, denoted as $\mathrm{BM}^{\mathrm{bal}}(\Psi) = \mathrm{BM}^{\mathrm{bal}}_m(\Psi)$ with block size m, and probability matrix $\Psi \in [0, 1]^{K\times K}$. Note the relationship $n = mK$ between the number of nodes n, the block size m and the number of blocks K. For notational consistency, we will denote diagonal and off-diagonal entries of Ψ differently:

$$p_k := \Psi_{kk}, \qquad q_{k\ell} := \Psi_{k\ell}, \qquad k \ne \ell. \tag{4.1}$$

The balanced planted partition model $\mathrm{PP}^{\mathrm{bal}}(p, q) = \mathrm{PP}^{\mathrm{bal}}_{m,K}(p, q)$ is a special case of $\mathrm{BM}^{\mathrm{bal}}_m(\Psi)$ where $p_k = p$ and $q_{k\ell} = q$ for all k, l.

We start with defining two notions of assortativity that will be key in our results. Let

$$q^*_k := \max_{r=k, s\ne k} q_{rs} = \max_{r\ne k, s=k} q_{rs}. \tag{4.2}$$

DEFINITION 4.1 (Strong and weak assortativity). Consider the balanced block model $\mathrm{BM}_m^{\mathrm{bal}}(\Psi)$ determined by (4.1):

- The model is strongly assortative (SA) if $\min_k p_k > \max_k q_k^*$.
- The model is weakly assortative (WA) if $p_k > q_k^*$ for all k.

An alternative way to state strong assortativity is $\min_k p_k > \max_{(k,\ell):k\neq\ell} q_{k,\ell}$. Strong assortativity implies weak assortativity. See (8.1) for an example where weak assortativity holds but not the strong one.

These definitions apply to general block models since they are defined only in terms of the edge probability matrix Ψ. We also define a partial order among balanced block models, which reflects the hardness of recovering the underlying cluster matrix X.

DEFINITION 4.2 (Strong assortativity (SA) ordering). The collection $\{\mathrm{BM}_m^{\mathrm{bal}}(\Psi) : \Psi \in [0,1]^{K\times K} \cap \mathbb{S}^K\}$ is partially ordered by

$$\mathrm{BM}_m^{\mathrm{bal}}(\widetilde{\Psi}) \geq \mathrm{BM}_m^{\mathrm{bal}}(\Psi) \quad \Longleftrightarrow \quad \widetilde{p}_k \geq p_k, \qquad \widetilde{q}_{k\ell} \leq q_{k\ell}, \qquad \forall k \neq \ell. \tag{4.3}$$

This ordering or the one induced on matrices in $[0,1]^{K\times K} \cap \mathbb{S}^K$ is referred to as SA-ordering.

Intuitively, for assortative models $\mathrm{BM}_m^{\mathrm{bal}}(\widetilde{\Psi}) \geq \mathrm{BM}_m^{\mathrm{bal}}(\Psi)$ implies that $\mathrm{BM}_m^{\mathrm{bal}}(\widetilde{\Psi})$ is easier than $\mathrm{BM}_m^{\mathrm{bal}}(\Psi)$ for cluster recovery. This will be made precise in Corollary 4.2 in Section 4.2.1. For example, consider a strongly assortative model $\mathrm{BM}^{\mathrm{bal}}(\Psi)$ where

$$p^- := \min_k p_k > \max_{(k,\ell):k\neq\ell} q_{k,\ell} =: q^+. \tag{4.4}$$

Then it is easy to see that $\mathrm{BM}^{\mathrm{bal}}(\Psi) \geq \mathrm{PP}^{\mathrm{bal}}(p^-, q^+)$, roughly meaning that fitting $\mathrm{BM}^{\mathrm{bal}}(\Psi)$ is not harder than fitting $\mathrm{PP}^{\mathrm{bal}}(p^-, q^+)$.

In order to study consistency, we always condition on the true cluster matrix, which is taken to be $X_0 := I_K \otimes \boldsymbol{E}_m$ without loss of generality. Let $\boldsymbol{S}_0 := \mathrm{supp}(X_0)$ be the index set of nonzero elements of X_0. We write $\mathrm{SDP}_{\mathrm{sol}}(A)$ for the solution set of the SDP for input A, where SDP is any of SDP-1, SDP-2 or SDP-3, which will be clear from the context. In this notation, $\mathrm{SDP}_{\mathrm{sol}}(A) = \{X_0\}$ means that X_0 is the unique solution of the SDP, in which case we say that the SDP is *strongly consistent* for cluster matrices.

REMARK 4.1. Our notion of consistency here is stronger than what is commonly called strong consistency in the literature [8, 15, 44, 59]. Strong consistency for an algorithm that outputs a set of community labels usually means exact recovery of labels up to a permutation of communities, with high probability. Viewing SDPs as algorithms that output the cluster matrix $\widehat{X}$, here by strong consistency we mean the exact recovery of X_0, which immediately implies exact label recovery. We note, however, that in some regimes one can recover labels exactly even when the output $\widehat{X}$ of an SDP is not exact. For example, one can run a community detection algorithm on $\widehat{X}$, say spectral clustering; see Algorithm 1 in Section 7. However, if the labels are inferred directly from $\widehat{X}$ (if $\widehat{X}$ corresponds to a graph with K disjoint connected components, then output the labels implied by the components; otherwise, output random labels), then our notion of strong consistency matches the standard one in the literature.

A key piece in our results will be the following matrix concentration inequality noted recently by many authors; see, for example, [21, 35, 52] and the references therein. Results of this type are often based on the refined discretization argument of [25].

PROPOSITION 4.1. *Let $A = (A_{ij}) \in \{0,1\}^{n\times n}$ be a symmetric binary matrix, with independent lower triangle and zero diagonal. There are universal positive constant (C, C', c, r) such that if*

$$\max_{ij} \operatorname{Var}(A_{ij}) \le \sigma^2 \qquad \text{for } n\sigma^2 \ge C' \log n$$

then with probability at least $1 - cn^{-r}$,

$$|||A - \mathbb{E}A||| \le C\sigma\sqrt{n}.$$

In what follows, (C, C', c, r) will always refer to the constants in this proposition. Our first result establishes consistency of SDP-2′ for the balanced planted partition models. We will work with two rescaled version of p, namely

$$\bar{p} := pm = p\frac{n}{K}, \qquad \widetilde{p} := \frac{\bar{p}}{\log n}, \tag{4.5}$$

and similarly for $\widetilde{q}$, $\bar{q}$ and q.

THEOREM 4.1 (Consistency of SDP-2′). *Let A be drawn from $\mathrm{PP}^{\mathrm{bal}}_{m,K}(p,q)$. For any $c_1, c_2 > 0$, let $C_1 := C' \vee \frac{4}{9}(c_1 + 1)$ and $C_2 := C + (\sqrt{4(c_1+1)} \vee 3\sqrt{4(c_2+1)})$. Assume $\widetilde{p} \ge C_1$. Then, if*

$$\widetilde{p} - \widetilde{q} > C_2(\sqrt{\widetilde{p}} + \sqrt{\widetilde{q}K}), \tag{4.6}$$

SDP-2′ is strongly consistent with probability at least $1 - c(Km^{-r} + n^{-r}) - n^{-c_1} - 2m^{-1}n^{-c_2}$.

As a consequence, we get consistency for a strongly assortative block model. More precisely, Theorem 4.1 combined with Corollary 4.2 in Section 4.2.1 gives the following.

COROLLARY 4.1 (Consistency of SDP-2′ for the strongly assortative case). *Let A be drawn from a strongly assortative* $\mathrm{BM}_m^{\mathrm{bal}}(\Psi)$. *Then the conclusion of Theorem 4.1 holds with* (p,q) *replaced with* (p^-, q^+) *as defined in* (4.4).

Note that Theorem 4.1 and its corollary automatically apply to SDP-1, because it is a tighter relaxation of the MLE than SDP-2′. However, SDP-1 succeeds for the much larger class of *weakly assortative* block models, as reflected in our main result, Theorem 4.2 below. Recall the notation q_k^* defined in (4.2) and write $q_{\max}^* = \max_k q_k^* = \max_{k\neq\ell} q_{k\ell}$. The scaled versions $\widetilde{q}_k^*$, $\bar{q}_k^*$ and $\widetilde{q}_{\max}^*, \bar{q}_{\max}^*$ are defined based on q_k^* and $q_{\max}^*$ as in (4.5).

THEOREM 4.2 (Consistency of SDP-1). *Let A be drawn from a weakly assortative* $\mathrm{BM}_m^{\mathrm{bal}}(\Psi)$. *For any* $c_1, c_2 > 0$, *let* $C_1 := C' \vee \frac{4}{9}(c_1+1)$ *and* $C_2 := (\sqrt{4(c_1+1)} + C) \vee (6\sqrt{2(c_2+1)})$. *Assume* $\min_k \widetilde{p}_k \geq C_1$. *Then, if*

$$\min_k \Big[(\widetilde{p}_k - \widetilde{q}_k^*) - C_2\Big(\sqrt{\widetilde{p}_k} + \sqrt{\widetilde{q}_k^*}\Big)\Big] > C\sqrt{\frac{\widetilde{q}_{\max}^* K}{\log n}}, \tag{4.7}$$

SDP-1 *is strongly consistent with probability at least* $1 - c(Km^{-r} + n^{-r}) - n^{-c_1} - 2m^{-1}n^{-c_2}$.

Note that for any weakly assortative $\mathrm{BM}_m^{\mathrm{bal}}(\Psi)$ with fixed K and constant entries of Ψ, condition (4.7) holds for large n and hence SDP-1 is strongly consistent. We show in Section 5 that SDP-2′ fails in general outside the class of strongly assortative block models.

REMARK 4.2. Our result for SDP-2 can be slightly strengthened by stating (4.6) as in (4.7) with $\widetilde{p}_k \equiv \widetilde{p}$ and $\widetilde{q}_{k,\ell} \equiv \widetilde{q}$. This gives a better threshold in the case where $\widetilde{q} \to \infty$ but $\widetilde{q}/\log n \to 0$.

REMARK 4.3. One can define strong and weak disassortativity by replacing p_k and $q_{k\ell}$ with $-p_k$ and $-q_{k\ell}$, respectively, in Definition 4.2. The results then hold if one applies the SDPs to $-A$ in the disassortative case.

REMARK 4.4. Another way to express conditions of Theorem 4.1 is in terms of the alternative parametrization (d, β) where $d := \bar{p} + (K-1)\bar{q}$ is the expected node degree, and $\beta := q/p = \bar{q}/\bar{p}$ is the out-in-ratio. A slight weakening of condition (4.6), using $(a+b)^2 \leq 2(a^2+b^2)$, gives

$$(\bar{p} - \bar{q})^2 \gtrsim (\bar{p} + \bar{q}K)\log n \quad \iff \quad d \gtrsim \left(\frac{1+K\beta}{1-\beta}\right)^2 \log n, \tag{4.8}$$

where we have used $d \asymp \bar{p} + K\bar{q}$. We also need $\bar{p} \gtrsim \log n$ which translates to $d \gtrsim (1+K\beta)\log n$, which is implied by (4.8). In particular, for fixed β, it is enough to have $d = \Omega(K^2 \log n)$ for SDP-2′ (and hence SDP-1) to be strongly consistent.

The proof of Theorem 4.2 appears in Section 4.3 with some of the more technical details deferred to the Appendices. The proof of Theorem 4.1 is similar and appears in the supplementary material [9] (Section 10.1).

4.1. *Comparison with other consistency results.* Rigorous results about the phase transition in the so-called reconstruction problem for the 2-block balanced PP model, that is, recovering a labeling positively correlated with the truth, in the sparse regime where $d = O(1)$, have appeared in [37, 42, 43] after originally conjectured by [23]. For $K = 2$, the problem of exact recovery in PP has recently been studied in [1, 44], where the exact recovery threshold is obtained when $d = \Omega(\log n)$, the minimal degree growth required for exact recovery. Mossel et al. [44] also discusses exact thresholds for weak consistency, that is, fraction of misclassified labels going to zero. Abbe et al. [1] also analyzed the MAXCUT SDP showing a consistency threshold within constant factor of the optimal. Since the earlier draft of our manuscript, more refined analyses of SDPs for balanced PP have appeared in [28, 29], as well as [2] which obtains the exact threshold for a general SBM, by a two-stage approach with no SDP involved. In [28], the argument in [1] is refined to show that MAXCUT SDP achieves the threshold of exact recovery with optimal constant, for the case $K = 2$. In [29], the analysis is extended to the general K, for an SDP which interestingly is equivalent to what we have called SDP-1, showing that it achieves optimal exact recovery threshold. This threshold is equivalent, up to constants, to that obtained in [21], and hence to (4.8) as will be discussed below. The analysis in [29] also provides the exact constant and an extension to the unbalanced case.

For the PP model with general K, [21] provides sufficient conditions for strong consistency of their nuclear norm relaxation of the MLE. These conditions automatically apply to SDP-2′ and SDP-1 since they are tighter relaxations. More precisely, their model, in the zero outliers case, coincides with $\mathrm{PP}^{\mathrm{bal}}(p, q)$ and their sufficient conditions translate to $(p - q)^2(n/K)^2 \gtrsim p(n/K)\log n + qn$. A slightly weaker version, obtained by replacing q with $q \log n$, reads $(\bar{p} - \bar{q})^2 \gtrsim (\bar{p} + \bar{q}K)\log n$ which is the one we have obtained in (4.8) as a consequence of Theorem 4.1. The stronger version also follows from our proof; see Remark 4.2. Interestingly, exactly the same condition (4.8), is established in [17] for SDP-3, when specialized to $\mathrm{PP}^{\mathrm{bal}}(p, q)$, the case with zero outliers. In other words, results of the form predicted by Theorem 4.1 already exist for SDP relaxations of the block model, albeit using different proof techniques. On the other hand, we are not aware of any results like Theorem 4.2, which guarantees success of SDP-1 for weakly assortative block models. A somewhat different condition amounting to $(\bar{p} - \bar{q})^2 \gtrsim mK^2\bar{p}$ and $np = K\bar{p} \geq \log n$ is implied by the results of [35] for spectral clustering based on the adjacency matrix, which we have called eigenvalue truncation (EVT). We note that the dependence on K is worse than in the SDP results, among other things. This is corroborated empirically in Section 8, which shows SDPs outperform EVT for larger values of K.

We should point out that there is a somewhat parallel line of work regarding relaxations for clustering problems. For example, a variant of SDP-1 [with $\operatorname{diag}(X) = \mathbf{1}_n$ replaced with $\operatorname{tr}(X) = n$] has been proposed as a relaxation of the K-means or normalized K-cut problems [47, 56]. However, theoretical analysis of SDPs in the clustering context have only recently began. See, for example, [11] for a recent analysis, using a probabilistic model of clusters. An earlier line of work reformulates the clustering problem as instances of the planted partition model and analyzes an SDP relaxation for cluster recovery [7, 38]. The planted K-disjoint clique model in [7] and the fully random model of [38] both can be considered as special case of the planted partition model. The analysis in [38] is in particular interesting for analyzing an SDP with triangle-inequality type constraints and providing approximation bounds relative to the optimal combinatorial solution.

Recently, a very interesting paper [27] analyzed the performance of SDP relaxations in the sparse regime where $d = O(1)$. They showed that as long as the feasible region is contained in the so-called Grothendieck set $\{X \succeq 0, \operatorname{diag}(X) \le 1\}$, the SDPs can achieve arbitrary accuracy, with high probability, assuming that $(\bar{p} - \bar{q})^2/(\bar{p} + \bar{q}K)$ is sufficiently large. These results are complementary to ours and show that all the SDPs in Table 1 are capable of approximate recovery in the sparse regime.

As a testament to the resurgence of interest in SDP relaxations for community detection, while this paper was under review, the following have appeared in the literature: On exact recovery for general K [3], [48] and for $K = 2$ [12], on the performance in the sparse regime [31, 41], on local optima of rank-restricted versions [40], on relation to Probably Certifiably Correct algorithms [13], on robustness [39] and on extensions to the case with node covariates [57]. In particular, [31, 40] provide encouraging results on scalability of SDPs for community detection. Their approach is based on solving rank-restricted versions of SDPs as nonconvex optimization problems, going back to the seminal work of [16]. The results in [31, 40] suggest that the rank can be taken to be much smaller for community detection problems than the previous known bound in [16].

4.2. *Some useful general results.* Here, we collect some general observations on solutions of SDPs which will be useful in proving Theorems 4.1 and 4.2. Let S_k be the indices of the kth community. We have $|S_k| = m$. Let $X_{S_k S_j}$ be the submatrix of X on indices $S_k \times S_j$, and $X_{S_k} := X_{S_k S_k}$. Let $\mathbf{1}_{S_k} \in \mathbb{R}^n$ be the indicator vector of S_k, equal to one on S_k and zero elsewhere. $\boldsymbol{E}_{\boldsymbol{S}_0} \in \{0, 1\}^{n \times n}$ denotes the indicator matrix of $\boldsymbol{S}_0 \subset [n]^2$. Let e_k^n, or simply e_k, be kth unit vector of $\mathbb{R}^n$. Let $\operatorname{span}\{\mathbf{1}_{S_k}\}$ and $\operatorname{span}\{\mathbf{1}_{S_k}\}^\perp$ denote the subspace spanned by $\{\mathbf{1}_{S_2}, \mathbf{1}_{S_2}, \dots, \mathbf{1}_{S_K}\}$ and its orthogonal complement. Let $d(S_k) \in \mathbb{R}^n$ be the vector of node degrees relative to the subgraph induced by S_k, $d(S_k) = A\mathbf{1}_{S_k} = A_{S_k}\mathbf{1}_m$. Note that $[d(S_k)]_{S_k} \in \mathbb{R}^m$ is the subvector of $d(S_k)$ on indices S_k.

4.2.1. *SDPs respect SA-ordering.* The following lemma formalizes an intuitive fact on how SDPs interact with the SA-ordering of Definition 4.2. The proof is given in the supplementary material [9] (Section 9).

LEMMA 4.1. *Let $\widetilde{A} \in \mathbb{S}^n$ be obtained from A by setting some elements off $\boldsymbol{S}_0$ to zero and some elements on $\boldsymbol{S}_0$ to one. Then, for either of SDP-1 or SDP-2′,*

$$\mathrm{SDP}_{\mathrm{sol}}(A) = \{X_0\} \quad \Longrightarrow \quad \mathrm{SDP}_{\mathrm{sol}}(\widetilde{A}) = \{X_0\}.$$

The lemma generalizes to any optimization problem that maximizes $X \mapsto \langle A, X \rangle$, and has its feasible region included in $\{X : 0 \le X \le 1\}$. An immediate consequence is the following probabilistic version for SBMs, stated conditionally on the true cluster matrix X_0.

COROLLARY 4.2. *Assume $\mathrm{BM}_m^{\mathrm{bal}}(\widetilde{\Psi}) \ge \mathrm{BM}_m^{\mathrm{bal}}(\Psi)$, and let $\widetilde{A} \sim \mathrm{BM}_m^{\mathrm{bal}}(\widetilde{\Psi})$ and $A \sim \mathrm{BM}_m^{\mathrm{bal}}(\Psi)$. Then, for either of SDP-1 or SDP-2′,*

$$\mathbb{P}\big(\mathrm{SDP}_{\mathrm{sol}}(\widetilde{A}) = \{X_0\}\big) \ge \mathbb{P}\big(\mathrm{SDP}_{\mathrm{sol}}(A) = \{X_0\}\big).$$

This corollary allows us to transfer consistency results for SDPs regarding a particular SBM to any SBM that dominates it. It also allows us to inflate off-diagonal entries of Ψ for a general $\mathrm{BM}^{\mathrm{bal}}(\Psi)$ without loss of generality. More precisely, we will assume in the course of the proof that off-diagonal entries of Ψ satisfy certain lower bounds to ensure concentration. These lower bounds can then be safely discarded at the end by Corollary 4.2.

4.2.2. *Optimality conditions.* Consider the following general SDP:

$$\begin{aligned} \max \quad & \langle A, X \rangle \\ \text{s.t.} \quad & \operatorname{diag}(X) = \mathbf{1}_n, \qquad \mathcal{L}_2(X) = b_2 \\ & X \succeq 0, \qquad X \ge 0, \end{aligned} \tag{4.9}$$

where $\mathcal{L}_2$ is a linear map from $\mathbb{S}^n$ to $\mathbb{R}^s$ for some integer s, and $b_2 \in \mathbb{R}^s$. This is a blueprint for both SDP-1 and SDP-2′. Let $\mathcal{L}_1(X) := \operatorname{diag}(X)$ and $b_1 = \mathbf{1}_n$. Then $\mathcal{L}(X) := (\mathcal{L}_1(X), \mathcal{L}_2(X)) = (b_1, b_2) =: b$ summarizes the linear constraints for the SDP. The dual problem is

$$\begin{aligned} \min \quad & \langle \mu, b_2 \rangle + \sum_i \nu_i \\ \text{s.t.} \quad & \mathcal{L}_2^*(\mu) + \operatorname{diag}^*(\nu) \succeq A + \Gamma, \qquad \Gamma \ge 0, \end{aligned}$$

where $\mu \in \mathbb{R}^s$, $\nu \in \mathbb{R}^n$ and $\Gamma \in \mathbb{S}^n$, and the minimization is over the triple (μ, ν, Γ) of dual variables. $\mathcal{L}_2^*$ is the adjoint of $\mathcal{L}_2$ and diag^* is the adjoint of diag. Letting

$$\Lambda := \Lambda(\mu, \nu, \Gamma) := \mathcal{L}_2^*(\mu) + \operatorname{diag}^*(\nu) - A - \Gamma, \tag{4.10}$$

the (KKT) optimality conditions are

Primal Feas.	$X \succeq 0,$	$X \geq 0,$	$\mathcal{L}(X) = b,$	
Dual Feas.	$\Lambda \succeq 0,$	$\Gamma \geq 0,$		
Comp. Slackness (a)	$\Gamma_{ij} X_{ij} = 0,$	$\forall i, j,$		(CSa)
Comp. Slackness (b)	$\langle \Lambda, X \rangle = 0.$			(CSb)

Another way to state (CSa) is to write $\Gamma \circ X = 0$ where $\circ$ denotes the Schur (element-wise) product of matrices.

The primal-dual witness approach that we will use in the proofs is based on finding a pair of primal and dual solutions that simultaneously satisfy the KKT conditions. The pair then witnesses strong duality between the primal and dual problems implying that it is an optimal pair.

4.2.3. *Sufficient conditions for exact recovery.* We would like to obtain sufficient conditions under which the true cluster matrix $X_0 = I_K \otimes \boldsymbol{E}_m$ is the unique solution of the primal SDP. Complementary slackness (a), or (CSa), implies that we need $\Gamma_{S_k} = 0$ for all k, while we are free to choose $\Gamma_{S_k S_j}$ for $j \neq k$, using the submatrix notation.

Since both X_0 and Λ are PSD, (CSb) is equivalent to $\Lambda X_0 = 0$, which is in turn equivalent to range$(X_0) \subset \ker(\Lambda)$. Note that X_0 has K nonzero eigenvalues, all equal to m, corresponding to eigenvectors $\{\mathbf{1}_{S_k}\}_{k=1}^K$, where $\mathbf{1}_{S_k} \in \mathbb{R}^n$ is the indicator vector of S_k. Hence, range$(X_0) = \text{span}\{\mathbf{1}_{S_k}\}$, and (CSb) for X_0 is equivalent to

$$\text{span}\{\mathbf{1}_{S_k}\} \subset \ker(\Lambda).$$

The following lemma, proved in the supplementary material [9] (Section 10), gives conditions for X_0 to be the unique optimal solution.

LEMMA 4.2. *Assume that Γ is dual feasible* (*i.e.*, $\Gamma \geq 0$), *and for some $\mu \in \mathbb{R}$ and $\nu \in \mathbb{R}^n$*:

(A1) $\ker(\Lambda(\mu, \nu, \Gamma)) = \text{span}\{\mathbf{1}_{S_k}\}$, *and* $\Lambda(\mu, \nu, \Gamma) \succeq 0$,
(A2) $\Gamma_{S_k} = 0, \forall k$,
(A3) *Each* $\Gamma_{S_k S_\ell}$, $k \neq \ell$ *has at least one nonzero element*.

Then X_0 is the unique primal optimal solution, *and* (μ, ν, Γ) *is dual optimal*.

Note that condition (A1) is satisfied if for some $\varepsilon > 0$,

$$\Lambda \mathbf{1}_{S_k} = 0, \qquad \forall k, \tag{4.11}$$

$$u^T \Lambda u \geq \varepsilon \|u\|_2^2, \qquad \forall u \in \text{span}\{\mathbf{1}_{S_k}\}^\perp. \tag{4.12}$$

4.3. *Proof of Theorem 4.2: Primal-dual witness for SDP-1.* Let $\boldsymbol{\Phi}_i = e_i \mathbf{1}_n^T + \mathbf{1}_n e_i^T \in \mathbb{S}^n$ where $e_i = e_i^n$ is the ith standard basis vector in $\mathbb{R}^n$. We note that $\langle X, \boldsymbol{\Phi}_i \rangle = \operatorname{tr}(X \boldsymbol{\Phi}_i) = 2(X\mathbf{1}_n)_i$. Thus, SDP-1 is an instance of (4.9), with $\mathcal{L}_2(X) = (\langle X, \boldsymbol{\Phi}_i \rangle)_{i=1}^n$ and $b_2 = 2m\mathbf{1}_n$. The corresponding adjoint operator is $\mathcal{L}_2^*(\mu) = \sum_{i=1}^n \mu_i \boldsymbol{\Phi}_i = \mu \mathbf{1}_n^T + \mathbf{1}_n \mu^T$. Thus [cf. (4.10)],

$$\Lambda = \Lambda(\mu, \nu, \Gamma) = \left(\mu \mathbf{1}_n^T + \mathbf{1}_n \mu^T\right) + \operatorname{diag}^*(\nu) - A - \Gamma. \tag{4.13}$$

The following summarizes our primal-dual construction in this case:

$$\begin{aligned} \nu_{S_k} &= \left[d(S_k)\right]_{S_k} - \phi_k m \mathbf{1}_m, \qquad \mu_{S_k} := \frac{1}{2}\phi_k \mathbf{1}_m, \\ \Gamma_{S_k} &:= 0, \end{aligned} \tag{4.14}$$

$$\begin{aligned} \Gamma_{S_k S_\ell} &:= \mu_{S_k} \mathbf{1}_m^T + \mathbf{1}_m \mu_{S_\ell}^T + P_{\mathbf{1}_m^\perp} A_{S_k S_\ell} P_{\mathbf{1}_m^\perp} - A_{S_k S_\ell}, \\ &= \frac{1}{2}(\phi_k + \phi_\ell)\boldsymbol{E}_m + P_{\mathbf{1}_m^\perp} A_{S_k S_\ell} P_{\mathbf{1}_m^\perp} - A_{S_k S_\ell}, \qquad k \neq \ell \end{aligned} \tag{4.15}$$

for some numbers $\{\phi_k\}_{k=1}^K$ to be determined later. Note that μ is chosen to be constant over blocks, but these constants can vary between blocks. We have the following analogue of Lemma 10.2. Recall that $\boldsymbol{E}_{\boldsymbol{S}_0^c}$ is the indicator matrix of $\boldsymbol{S}_0^c$ where $\boldsymbol{S}_0$ is the support of X_0.

LEMMA 4.3. *Let* (μ, ν, Γ) *be as defined in* (4.14)–(4.15). *Then* Γ *verifies* (A2) *and* (4.11) *holds. In addition*:

(a) Γ *is dual feasible, that is,* $\Gamma \geq 0$, *if for all* $i \in S_k$, $j \in S_\ell$, $\ell \neq k$,

$$\frac{1}{2}(\phi_k + \phi_\ell) m \geq d_i(S_\ell) + d_j(S_k) - d_{av}(S_k, S_\ell), \tag{4.16}$$

and satisfies (A3) *if at least one inequality is strict for each pair* $k \neq \ell$.

(b) Γ *verifies* (4.12) *if for* $\rho_k := \min_{i \in S_k} d_i(S_k)/m$,

$$\min_k \left[(\rho_k - \phi_k)m - |||\Delta_k|||\right] > |||\boldsymbol{E}_{\boldsymbol{S}_0^c} \circ \Delta|||. \tag{4.17}$$

This lemma amounts to a set of deterministic conditions for the success of SDP-1. To complete the proof of Theorem 4.2, we develop a probabilistic analogue by choosing $\phi_k \approx \bar{q}_k^*$ and using the key inequality $\bar{q}_{k\ell} \leq \frac{1}{2}(\bar{q}_k^* + \bar{q}_\ell^*)$. See Appendix B for details.

5. Failure of SDP-2′ in the absence of strong assortativity. We now show that strong assortativity is a necessary condition for exact recovery in SDP-2′. For this purpose, it is enough to focus on the noiseless case, that is, when the input to the SDP is the mean matrix of the block model. If SDP-2′ fails on exact recovery of the true population mean, there is no hope of recovering its noisy version, that is, the adjacency matrix. The following result is deterministic and nonasymptotic. In particular, it holds without any constraints on the expected degrees (besides those imposed by assortativity assumptions). We will state it in a slightly more general form than is needed here, including the case of general block sizes. Keeping consistency with earlier notation, we let $\boldsymbol{E}_{S_k S_\ell} \in \{0,1\}^{n\times n}$ be the indicator matrix of the set $S_k \times S_\ell$, and $\boldsymbol{E}_{S_k} := \boldsymbol{E}_{S_k S_k}$. Similarly, I_{S_k} is the $n \times n$ identity matrix with elements outside $S_k \times S_k$ set to zero, that is, $\boldsymbol{E}_{S_k S_\ell}$ is not a submatrix of $\boldsymbol{E}_n$, but a masked version of it.

PROPOSITION 5.1. *Let $\mathbb{E}[A]$ be the mean matrix of a weakly assortative block model. Assume that the blocks are indexed by $S_k \subset [n]$ where $|S_k| = n_k$, for $k = 1,\dots,K$. For some $I \subset [n] = \{1,\dots,n\}$, to be determined, consider a solution of the form*

$$X = \sum_{k\in I}\sum_{\ell\in I} \alpha_{k\ell}\boldsymbol{E}_{S_k S_\ell} + \sum_{k\notin I}[\beta_k \boldsymbol{E}_{S_k} + (1-\beta_k) I_{S_k}], \qquad \alpha_{k\ell} = \alpha_{\ell k}, \tag{5.1}$$

with $\alpha_{kk} = 1$, $k \in I$ and $\beta_k \in [0,1)$ for $k \notin I$. Then the following hold:

(a) *Assume that* $\operatorname{argmax}_{k\neq\ell} q_{k\ell} = \{(k_0,\ell_0)\}$ *and let $I := \{k : p_k \ge q_{k_0\ell_0}\}$. Furthermore, let $m := \min_k n_k$, $\xi_k := n_k/m$ and*

$$\alpha^*_{k_0\ell_0} := \frac{1}{2\xi_{k_0}\xi_{\ell_0}}\left[\left(1-\frac{1}{m}\right)\sum_{k\notin I}\xi_k - \sum_{k\in I}\xi_k(\xi_k - 1)\right]. \tag{5.2}$$

If $\alpha^*_{k_0\ell_0} \in [0,1]$, then SDP-2′, applied with $A = \mathbb{E}[A]$ and $m = \min_k n_k$ has (5.1) as solution, with $\alpha_{k\ell} = \alpha^*_{k_0\ell_0} 1\{\{k,\ell\} = \{k_0,\ell_0\}\}$ and $\beta_k = 0$ for all $k \notin I$.

(b) *Assume that the given block model is balanced and let $I^c := [K] \setminus I = \{k : p_k < q_{k_0\ell_0}\}$ where I and (k_0,ℓ_0) are defined in part* (a). *If $|I^c| \le 2$, then the conclusion of part* (a) *holds with $\alpha^*_{k_0\ell_0} = \frac{1}{2}(1-1/m)|I^c|$.*

(c) *Assume that the given block model is balanced and weakly but not strongly assortative. Let $SDP_{\mathrm{sol}}(\cdot)$ be the solution set of SDP-2′. Then $SDP_{\mathrm{sol}}(\mathbb{E}[A]) \neq \{X_0\}$.*

We note that part (c) establishes the failure of SDP-2′ once strong assortativity is violated. We prove the proposition in Section 14 of the supplementary material [9]. The conclusions of both parts (a) and (b) hold even in the strongly assortative case. However, in that case, the set I^c will be empty and the conditions in part (a) cannot be met, whereas part (b) gives the expected result of $X = X_0$. The interesting case occurs when strong assortativity is violated, which gives a nonempty set I^c. Since $\beta_k = 0$ for $k \in I^c$, this shows that SDP-2′ fails to recover those blocks. The condition $|I^c| \le 2$, in the balanced case, might seem restrictive, but it is enough for our purpose of establishing part (c). In general, that is, with no assumption on $|I^c|$, SDP-2′ still misses the blocks violating strong associativity, though the nonzero-block portion of X, namely, $(\alpha_{k\ell})_{k,\ell\in I}$ takes a more complicated form. In parts (a)

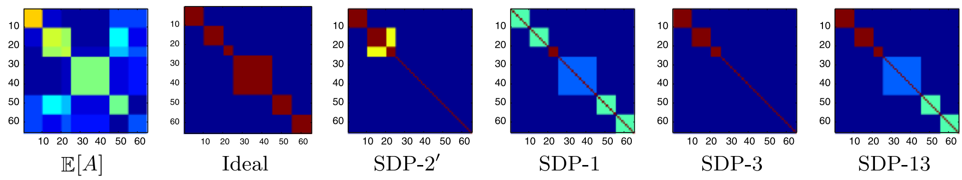


FIG. 1. *Illustration of Propositions* 6.1 *and* 5.1. *This block model is weakly but not strongly assortative, and has unequal block sizes* $\underline{\boldsymbol{n}} = (10, 10, 5, 20, 10, 10)$. *The leftmost column is the population mean, and the rest of the columns are the results of various SDPs, with* $m = \min_k n_k$, *and equal regularization parameters in the case of SDP*-3 *and SDP*-13 ($\lambda = \mu$). *The ideal cluster matrix is also shown for comparison. See Section* 17 *in the supplementary material* [9] *for more details.*

and (b), at most one nondiagonal element of $(\alpha_{k\ell})$ is nonzero whereas in general several such elements will be nonzero. These ideas are illustrated in Figure 1, with more detailed discussion in Section 6.1 of the supplementary material [9], in particular, with an application of part (a), in the unbalanced case with $|I^c| > 2$.

6. Extensions to the unbalanced case. Let us discuss how our results can be extended to the unbalanced case. Recall that, in general, $\underline{\boldsymbol{n}} = (n_1, \dots, n_K)$ denotes the vector of block sizes. One could argue that as long as $(\min_k n_k)/n \geq C$ for some constant $C > 0$, that is, $\{n_k/n\}_k$ is bounded away from zero, the problem of block model recovery is not inherently more difficult than that of the balanced case. To simplify our discussion, we focus on the noiseless case from which the results can be extended to the aforementioned bounded block-size regime. We will show that in a weakly assortative block model, SDP-1 applied with $m = \min_k n_k$ recovers all the blocks, albeit some imperfectly. We also consider the following mixture of SDP-1 and SDP-3, which we will call **SDP-13**:

$$
\begin{aligned}
&\max_X \quad \langle A, X\rangle - \mu\langle \boldsymbol{E}_n, X\rangle, \\
&\text{s.t.} \quad \operatorname{diag}(X) = \mathbf{1}_n, \qquad X\mathbf{1}_n \geq m\mathbf{1}_n, \\
&\qquad\quad X \succeq 0, \qquad X \geq 0,
\end{aligned}
\tag{6.1}
$$

and show that when applied with $m \leq \min_k n_k$ and appropriate choice of μ, it too recovers the blocks with improvements over SDP-1. Without loss of generality, let us sort the blocks so that $p_1 \geq p_2 \geq \cdots \geq p_K$.

PROPOSITION 6.1. *Let* $\mathbb{E}[A]$ *be the mean matrix of a weakly assortative block model, with blocks indexed by* $S_k \subset [n]$ *where* $|S_k| = n_k$, *for* $k = 1, \dots, K$. *Consider a solution of the form*

$$
X = \sum_{k=1}^{K} \alpha_k \boldsymbol{E}_{S_k} + (1 - \alpha_k) I_{S_k}, \qquad \alpha_k \in (0, 1]. \tag{6.2}
$$

The following hold:

(a) *SDP*-1 *applied with* $A = \mathbb{E}[A]$ *and* $m \le \min_k n_k$ *has* (6.2) *as a solution with* $\alpha_k = (m-1)/(n_k-1)$ *for all* k.

(b) *Consider* $I := \{k : n_k > m\}$ *and* $I_1(k) := \{r \in I : r \le k\}$. *Let* $J_k := \bigcap_{r=1}^{k} [q_r^*, p_r]$. *Define* $k_0 := \max\{k : J_k \ne \varnothing\}$. *Then SDP*-13, *applied with* $A = \mathbb{E}[A]$, $m \le \min_k n_k$ *and*

$$\mu \in J_{k_0} \cap [p_{k_0+1}, 1], \qquad (p_{K+1} := 0),$$

has (6.2) *as a solution with*

$$\alpha_k = \begin{cases} 1 & k \in I_1(k_0), \\ (m-1)/(n_k-1) & \textit{otherwise}. \end{cases}$$

The key difference between the solution presented in Proposition 6.1 and that of Proposition 5.1 is that in the former, all α_k are guaranteed to be nonzero, whereas in the latter, $\alpha_k \equiv \beta_k$ corresponding to blocks violating weak assortativity are zero. Let us call the blocks in (6.2) for which $\alpha_k \in (0, 1)$, as imperfectly-recovered, while those with $\alpha_k = 1$ as *perfectly recovered*. The result of Proposition 6.1 can be summarized as follows: Both SDP-1 and SDP-13, with properly set parameters, recover all the blocks at least imperfectly, while SDP-13 has the potential to recover more blocks perfectly. Note that this is in contrast to the *partial recovery* of SDP-2′ in Proposition 5.1 where some of the blocks are completely missing from the solution, namely, those with $\alpha_k \equiv \beta_k = 0$. (Both perfect and imperfect recovery of the blocks of X lead to exact label recovery; the difference matters when one adds noise, since one expects the imperfectly recovered blocks to need more SNR to be recovered in the noisy setting.) In particular, we always have $k_0 \ge 1$ in part (b), implying that SDP-13 recovers at least one more block perfectly relative to SDP-1. In the special case of a strongly assortative block model, we have $\varnothing \ne [\max_k q_k^*, \min_k p_k] \subset \bigcap_{k=1}^{K} [q_r^*, p_r]$, hence $k_0 = K$ and SDP-13 recovers all the blocks perfectly. It is also interesting to note that both SDP-1 and SDP-13 recover the smallest blocks (i.e., those in $\{k : n_k = m\}$) perfectly, when we set $m = \min_k n_k$ (which is the optimal choice if the minimum is known). These observations are illustrated in Figure 1. The proof of Proposition 6.1 appears in the supplementary material [9], along more details on Figure 1.

7. Application to network histograms. A balanced block model is ideally suited for computing network histograms as defined by [46], which have been proposed as nonparametric estimators of graphons. They have been shown to do well empirically and recent results of [33], Section 2.4, suggest rate optimality of the balanced models for reasonably sparse graphs; see also [26]. A graphon is a bivariate symmetric function $f : [0, 1]^2 \to [0, 1]$. The corresponding network model can be written as $\mathbb{E}[A|\xi] = f(\xi_i, \xi_j)$ where $\xi = (\xi_1, \dots, \xi_n) \in [0, 1]^n$ are (unobserved) latent node positions. Without loss of generality, (ξ_i) can be assumed to be i.i.d. uniform on $[0, 1]$. The goal is to recover (a version of) f given A. In general, f is identifiable up to a measure-preserving transformation σ of $[0, 1]$ onto itself, since $f^\sigma = f(\sigma(\cdot), \sigma(\cdot))$ produces the same network model as f.

Let $\{I_1, I_2, \dots, I_K\}$ be a partition of $[0, 1]$ into equal-sized blocks, that is, $|I_k| = 1/K$ for $k \in [K]$. We associate to each node a label z_i, by letting $z_i := k$ if $\xi_i \in I_k$. With some abuse of notation, we identify z_i with an element $(z_{ik})_k$ of $\{0, 1\}^K$ as before, and let $Z = (z_{ik})_{ik}$. Then $M_Z := \mathbb{E}[A|Z]$ follows a block model as in (2.1) with $[\Psi]_{kk} = |I_k|^{-1} \int_{I_k} f(\xi, \xi)\, d\xi$ and $[\Psi]_{k\ell} = (|I_k||I_\ell|)^{-1} \int_{I_k} \int_{I_\ell} f(\xi, \xi')\, d\xi\, d\xi'$, for $k \neq \ell$. Asymptotically, as $n \to \infty$, this block model is very close to being balanced. It provides an approximation of f, via the mapping that sends Ψ to a block constant graphon $\widetilde{f}$, defined as $\widetilde{f}(\xi, \xi') = [\Psi]_{k\ell}$ if $\xi \in I_k, \xi' \in I_\ell$. One can show that under regularity assumptions (e.g., smoothness) on f, as $K \to \infty$, $\widetilde{f}$ approximates f, for example, in the quotient norm: $\inf_\sigma \|f - \widetilde{f}^\sigma\|_{L^2}$. Alternatively, one can consider the mean matrix $M_f := (f(\xi_i, \xi_j))_{ij} \in [0, 1]^{n\times n}$ of the graphon model as an empirical version of f. In which case, the mean matrix M_Z of the aforementioned block model serves as an approximation to M_f, for example, in the quotient norm: $\inf_P |||M_f - P M_Z P^T|||_F$, where P runs through permutation matrices. This is the approach we take here and, with some abuse of terminology, call M_f the "graphon."

Graphon estimation via block model approximation requires estimating the mean matrix M_Z, which is fairly straightforward once we have a good estimate of the cluster matrix X. Algorithm 1 details the procedure based on eigenvalue truncation and K-means (i.e., spectral clustering), leading to estimate $\widehat{M_{\widehat{Z}}}$ of M_f. We call $\widehat{M_{\widehat{Z}}}$ a *network histogram* or a *graphon estimator*, and note that it can be computed from any estimate of $\widehat{X}$. However, in practice, SDP-1 has advantages over other ways of estimating $\widehat{X}$ in this context. The likelihood-based estimators have no way of enforcing equal number of nodes in each block, whereas our empirical results in Section 8 show that SDP-1 has a high tendency to form equal-sized blocks, more so than SDP-2, making it an ideal choice for histograms. SDP-3 is not well suited for this task since it does not enforce either a particular number of blocks or a particular block size. It is more flexible due to the tuning parameter λ, but that flexibility is a disadvantage when the goal is to construct a histogram.

Algorithm 1 Graphon estimation by fitting $\mathrm{PP}^{\mathrm{bal}}(p, q)$

Require: Estimated cluster matrix $\widehat{X}$, and number of blocks K.
Ensure: Graphon estimator $\widehat{M_{\widehat{Z}}}$.

1: Compute the eigendecomposition $\widehat{X} = \widehat{U}\widehat{\Lambda}\widehat{U}^T$ and set $\widehat{U}^K = \widehat{U}(:, 1{:}\, K)$.
2: Apply `K-means` to rows of $\widehat{U}^K$ to get a label vector $e \in [K]^n$. Set $\widehat{Z}(i, e(i)) = 1$, otherwise 0.
3: Set $\widehat{\Psi}_{rk} = \frac{1}{n^2} \sum_{e_i = r, e_j = k} A_{ij}$ for $r \neq k$ and $\frac{1}{n(n-1)} \sum_{e_i = e_j = r} A_{ij}$ otherwise.
4: Change $\widehat{\Psi}$ to $Q\widehat{\Psi}Q^T$ so that its diagonal is decreasing. and update $\widehat{Z}$ to $\widehat{Z}Q^T$.
5: Change $\widehat{Z}$ to $P\widehat{Z}$ so that corresponding labels are in increasing order.
6: Set $\widehat{M_{\widehat{Z}}} = \widehat{Z}\widehat{\Psi}\widehat{Z}^T$.

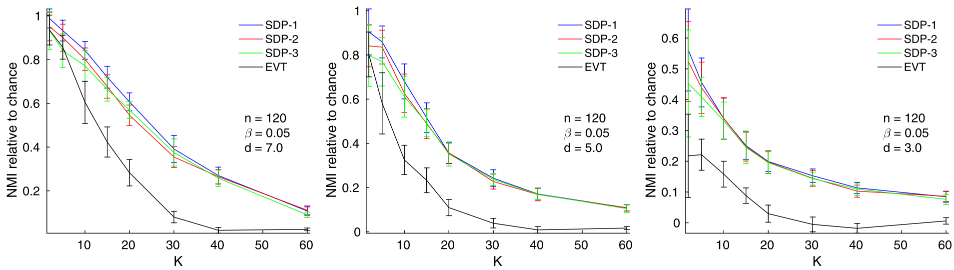


FIG. 2. *Bias-corrected NMI versus K in a balanced planted partition model, for various values of average degree d, with $n = 120$ and $\beta = 0.05$.*

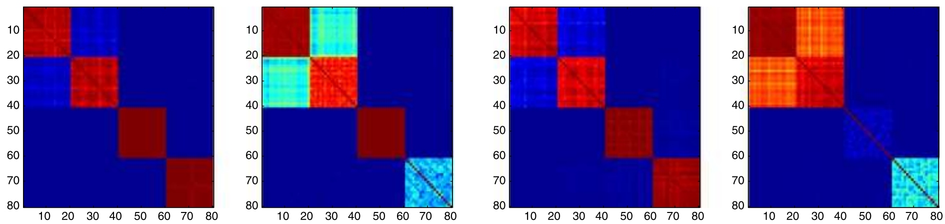
8. Numerical results. In this section, we present some experimental results comparing SDP-1 with SDP-2, SDP-3 and EVT, which amounts to spectral clustering on the adjacency matrix A. We chose EVT rather than a version of spectral clustering based on the graph Laplacian because SDPs also all operate on A itself. For SDP-3, in simulations we set the tuning parameter λ to the optimal value given in (2.5); a data-driven choice is given in [17].

We first consider the balanced symmetric model $\mathrm{PP}^{\mathrm{bal}}(p, q)$, reparametrized in terms of the average expected degree $d = p(\frac{n}{K} - 1) + q\frac{n}{K}(K - 1)$ and the out-in-ratio $\beta = q/p < 1$. Estimation becomes harder when d decreases (fewer edges) and when β increases (communities are not well separated). Figure 2 shows the agreement of estimated labels with the truth, as measured by the normalized mutual information (NMI), versus the number of communities K, averaged over 25 Monte Carlo replications. NMI takes values between 0 and 1, with higher values representing a better match. The labels are estimated from $\widehat{X}$ by Algorithm 1. As expected, the SDPs rank according to the tightness of relaxation, with SDP-1 dominating the other two, and all SDPs outperforing EVT. In Figure 2, the NMI is bias-adjusted, so that random guessing maps to NMI = 0. Without the adjustment, the NMI of random guessing increases as K approaches n, leading to a "dip" in the plots. See Figure 6 in the supplementary material [9] and the discussion that follows for more details.

Next, we consider a more general balanced block model $\mathrm{BM}^{\mathrm{bal}}(\Psi)$, with $K = 4$ to investigate the predictions of the theorems of Section 4. We consider the probability matrix

$$\Psi = \begin{pmatrix} 0.7 & 0.4 & 0.05 & 0.2 \\ 0.4 & 0.6 & 0.05 & 0.2 \\ 0.05 & 0.05 & p_3 & 0.05 \\ 0.2 & 0.2 & 0.05 & 0.4 \end{pmatrix} \tag{8.1}$$

and we vary p_3 from 0.7 down to 0.05. This model never satisfies the strong assortativity assumption over the range of p_3, because of the last row. However, it is at the boundary of strong assortativity if $p_3 > 0.4$, since $\Psi_{44} = \max_{k\neq\ell} \Psi_{k\ell}$ and



(a) SDP-1, $p_3 = 0.7$ (b) SDP-2, $p_3 = 0.7$ (c) SDP-1, $p_3 = 0.05$ (d) SDP-2, $p_3 = 0.05$

FIG. 3. *Mean estimated cluster matrices, $\widehat{X}$, for SDP-1 and SDP-2, for the weakly but not strongly assortative model (8.1) with $p_3 = 0.7$ and $p_3 = 0.05$. SDP-2 fails to recover one block at $p_3 = 0.7$ and two blocks at $p_3 = 0.05$.*

$\Psi_{jj} > \max_{k \neq \ell} \Psi_{k\ell}$ for $j \neq 4$. Its deviation from strong assortativity increases once p_3 falls below 0.4, and again once it crosses below 0.2. However, except for the boundary value of $p_3 = 0.05$, the model always remains weakly assortative. Figure 3 shows the results of Monte Carlo simulations with 25 replications, for SDP-1 and SDP-2. Mean cluster matrices $\widehat{X}$ obtained for the two SDPs are shown at the boundary points $p_3 = 0.05, 0.7$. SDP-2 has difficulty recovering the fourth block in both cases, and completely fails to recover the third block when $p_3 = 0.05$. The performance of SDP-1, however, remains more or less the same, surprisingly even at $p_3 = 0.05$. This can be clearly seen in Figure 4, which shows the relative errors $\|\widehat{X} - X_0\|_F / \|X_0\|_F$ for cluster matrices and the NMI for the labels reconstructed by Algorithm 1. Note how SDP-2 degrades as p_3 decreases to 0.05, with a sharp drop around 0.2, while SDP-1 behaves more or less the same. Note that for larger values of p_3, while SDP-2 does not reconstruct X_0 exactly as seen from the relative error plot, the resulting labels are nearly always exactly the truth as seen from the NMI plot. This may be due to the EVT truncation on $\widehat{X}$ implicit in Algorithm 1.

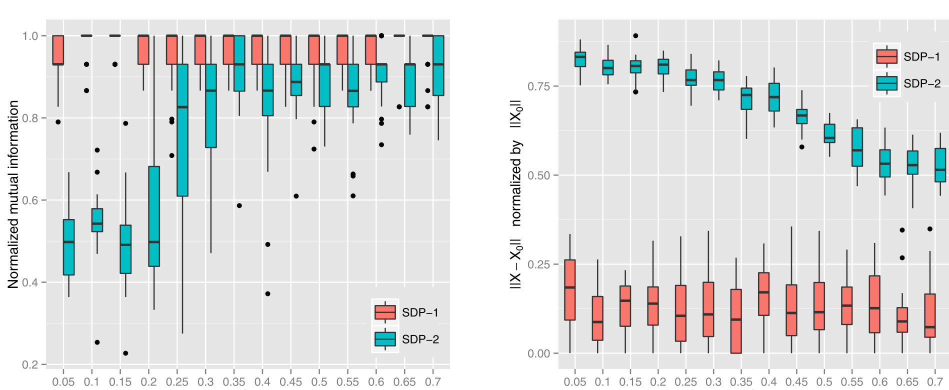
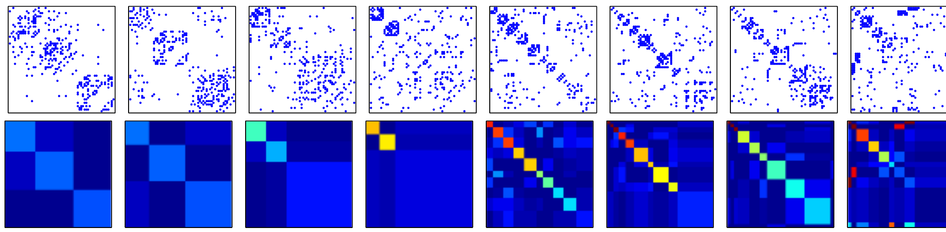


FIG. 4. *NMI and relative error of $\widehat{X}$ versus p_3 for the model with probability matrix (8.1).*



(a) SDP-1 (b) SDP-2 (c) SDP-3 (d) EVT (e) SDP-1 (f) SDP-2 (g) SDP-3 (h) EVT

FIG. 5. *Results for the dolphins network for* $K = 3$ (a)–(d) *and* $K = 10$ (e)–(h). *Row* 1: *adjacency matrix sorted according to the permutation of Algorithm* 1. *Row* 2: *Graphon estimator* $\widehat{M}_{\widehat{Z}}$ *of Algorithm* 1.

Finally, we apply various SDPs to graphon estimation for the dolphins network [36], with $n = 62$ nodes. Figure 5 shows the results for the three SDPs and the EVT with $K = 3$ and $K = 10$. For SDP-3, we used the median connectivity to set λ as suggested in [17]. The adjacency matrices in the first row and the graphon estimators in the second row are both permuted according to the ordering from Algorithm 1. The SDPs again provide a much cleaner picture of the communities in the data than the EVT. The blocks found by SDP-1 are similar in size and well separated from each other compared to the other two SDPs. We also applied the algorithms with $K = 2$ to compare to the the partition suggested by Figure 1(b) in [36], which can be considered the ground truth for a two-community structure. SDP-1, SDP-2, SDP-3 and EVT misclassify 7, 1, 4 and 11 nodes, respectively, out of 62. Since this partition into two blocks has unbalanced blocks (20 and 42), we expect SDP-1 to not match it as well. However, if we replace equality constraints with the inequality ones as discussed in Remark 3.1, SDP-1 misclassifies only 2 nodes. It is worth noting that the ground truth in this case is only one possible way to describe the network, taken from one scientific paper focused on the dolphins split, and there may well be more communities than two in the data. The nine strong (and one weak) clusters found by SDP-1 may be of interest for further understanding of this network.

9. Discussion. In this paper, we have put several SDP relaxations of the MLE into a unified framework (Table 1) by treating them as relaxations over different parameter spaces. SDP-1, the tighter relaxation we proposed, was shown to empirically dominate previous relaxations, and while all the SDPs we considered are strongly consistent on the strongly assortative class of block models, we showed that SDP-1 is strongly consistent on the much larger class of weakly assortative models, while SDP-2 fails outside the strongly assortative class. We proposed a mixture of SDP-1 and SDP-3 which combines the flexibilities of both, namely, consistency in weakly assortative and unbalanced models. It remains an open question whether a SDP relaxation can work for mixed networks with both assortative and dissortative communities. There are some indications that one can tackle mixed networks by applying SDPs to $|A| = \sqrt{A^2}$, the positive square-root of A.

We also note that SDP-3 is harder to compare directly to SDP-1 or SDP-2 because it depends on a tuning parameter λ. However, Lagrange duality implies that for every A and K there exists a λ that makes SDP-3 equivalent to SDP-2. In general, SDP-3 is more flexible than SDP-2 because of the continuous parameter λ, but this also makes it unsuitable for certain tasks such as histogram estimation. Empirically, the SDPs outperformed adjacency-based spectral clustering (EVT), especially for a large number of communities K. This is reflected in current theoretical guarantees, where the conditions for the SDPs have better dependence on K that those available for the EVT. In addition, SDP formulation of EVT shows it to be a looser relaxation than, say, SDP-1 for balanced planted partition model. The three SDPs also seem to be inherently more robust to noise than the EVT, perhaps due to the implicit regularization effect of the doubly nonnegative cone.

APPENDIX A: PROOF OF LEMMA 4.3

As in the proof of Theorem 4.1, and in accordance with condition (A2), we set $\Gamma_{S_k} := 0$. Then

$$\begin{aligned} \Lambda_{S_k} &= \mu_{S_k}\mathbf{1}_m^T + \mathbf{1}_m\mu_{S_k}^T + \operatorname{diag}^*(\nu_{S_k}) - A_{S_k}, \\ \Lambda_{S_k^c S_k} &= \mu_{S_k^c}\mathbf{1}_m^T + \mathbf{1}_{n-m}\mu_{S_k}^T - (A+\Gamma)_{S_k^c S_k}. \end{aligned} \tag{A.1}$$

Recalling that $[d(S_k)]_{S_k} = A_{S_k}\mathbf{1}_m$, we can rewrite (4.11) as

$$\Lambda_{S_k}\mathbf{1}_m = 0 \quad\Longleftrightarrow\quad \mu_{S_k}m + \mathbf{1}_m\mu_{S_k}^T\mathbf{1}_m + \nu_{S_k} - [d(S_k)]_{S_k} = 0, \tag{A.2}$$

$$\begin{aligned} &\Lambda_{S_\ell S_k}\mathbf{1}_m = 0 \quad\Longleftrightarrow \\ &\quad [\mu_{S_\ell}\mathbf{1}_m^T + \mathbf{1}_m\mu_{S_k}^T - (A+\Gamma)_{S_\ell S_k}]\mathbf{1}_m = 0, \qquad k \neq \ell. \end{aligned} \tag{A.3}$$

As in the case of SDP-2′ (cf. the supplementary material [9]), (A.3) is equivalent to

$$\mu_{S_\ell}\mathbf{1}_m^T + \mathbf{1}_m\mu_{S_k}^T - (A+\Gamma)_{S_\ell S_k} = -B_{S_\ell S_k}$$

for some $B_{S_\ell S_k}$ acting on $\operatorname{span}\{\mathbf{1}_m\}^\perp$. As before, we set $B_{S_\ell S_k} := P_{\mathbf{1}_m^\perp} A_{S_\ell S_k} P_{\mathbf{1}_m^\perp}$, and note that

$$\Delta := A - \mathbb{E}A, \qquad [\mathbb{E}A]_{S_k} = p_k\boldsymbol{E}_m, \qquad [\mathbb{E}A]_{S_k S_\ell} = q_{k\ell}\boldsymbol{E}_m, \qquad k \neq \ell, \tag{A.4}$$

so that $B_{S_\ell S_k} = P_{\mathbf{1}_m^\perp}\Delta_{S_\ell S_k}P_{\mathbf{1}_m^\perp}$. Now, take $u \in \operatorname{span}\{\mathbf{1}_{S_k}\}^\perp$. Then $u = \sum_k u_{S_k} = \sum_k e_k \otimes u_k$, for some $\{u_k\} \subset \operatorname{span}\{\mathbf{1}_m\}^\perp$. We will work with expansion of $u^T\Lambda u$ obtained in the supplementary material ([9], equation (11.5)). Using $A_{S_k} = p_k\boldsymbol{E}_m + \Delta_{S_k}$ and (A.1), we have

$$\begin{aligned} u_k^T\Lambda_{S_k}u_k &= u_k^T[\mu_{S_k}\mathbf{1}_m^T + \mathbf{1}_m\mu_{S_k}^T - p_k\boldsymbol{E}_m + \operatorname{diag}^*(\nu_{S_k}) - \Delta_{S_k}]u_k \\ &= u_k^T(\operatorname{diag}^*(\nu_{S_k}) - \Delta_{S_k})u_k \end{aligned}$$

using $\mathbf{1}_m^T u_k = 0$. Let us now choose μ to be constant over blocks, that is, $\mu_{S_k} := \frac{1}{2}\phi_k \mathbf{1}_m, \forall k$ for some numbers $\{\phi_k\}$ to be determined later. Note that (A.2) reads $\phi_k m \mathbf{1}_m + \nu_{S_k} - [d(S_k)]_{S_k} = 0$, or equivalently

$$\text{(A.5)} \qquad \operatorname{diag}^*(\nu_{S_k}) = \operatorname{diag}^*\big([d(S_k)]_{S_k}\big) - \phi_k m I_m.$$

On the other hand, for $k \neq \ell$, we have $u_k^T \Lambda_{S_k S_\ell} u_\ell = -u_k^T B_{S_k S_\ell} u_\ell = -u_k^T \Delta_{S_k S_\ell} u_\ell$ since $\{u_k\} \subset \operatorname{span}\{\mathbf{1}_m\}^\perp$. We arrive at

$$\text{(A.6)} \quad u^T \Lambda u = \sum_k u_k^T \big(\operatorname{diag}^*\big([d(S_k)]_{S_k}\big) - \bar{\mu}_k m I_m - \Delta_{S_k}\big) u_k - \sum_{k \neq \ell} u_k^T \Delta_{S_k S_\ell} u_\ell.$$

PROOF OF (a) AND (b). To verify dual feasibility, recall that $P_{\mathbf{1}_m^\perp} e_j = e_j - \frac{1}{m}\mathbf{1}_m$. Then

$$\begin{aligned}[\Gamma_{S_k S_\ell}]_{ij} = e_i^T \Gamma_{S_k S_\ell} e_j &= \frac{1}{2}(\phi_k + \phi_\ell) + \Big(e_i - \frac{1}{m}\mathbf{1}_m\Big)^T A_{S_k S_\ell} \Big(e_j - \frac{1}{m}\mathbf{1}_m\Big) - A_{ij} \\ &= \frac{1}{2}(\phi_k + \phi_\ell) - \frac{1}{m}\big[d_i(S_\ell) + d_j(S_k) - d_{\text{av}}(S_k, S_\ell)\big] \geq 0.\end{aligned}$$

To verify (4.12), we recall representation (A.6). By assumption $\operatorname{diag}^*([d(S_k)]_{S_k}) \succeq \rho_k m I_m$ for all k. From (A.5) and (A.6), it follows that, for $u \in \operatorname{span}\{\mathbf{1}_{S_k}\}^\perp$,

$$\begin{aligned} u^T \Lambda u &\geq \sum_k u_k^T (\rho_k m I_m - \phi_k m I_m - \Delta_{S_k}) u_k - \sum_{k \neq \ell} u_k^T \Delta_{S_k S_\ell} u_\ell \\ &\geq \sum_k \big[(\rho_k - \phi_k) m - |||\Delta_{S_k}|||\big] \|u_k\|^2 - u^T (\boldsymbol{E}_{\boldsymbol{S}_0^c} \circ \Delta) u \\ &\geq \min_k \big[(\rho_k - \phi_k) m - |||\Delta_{S_k}|||\big] \|u\|^2 - |||\boldsymbol{E}_{\boldsymbol{S}_0^c} \circ \Delta||| \|u\|^2. \end{aligned}$$

□

APPENDIX B: PROBABILISTIC BOUNDS FOR BM^{bal}

We will complete the construction of (μ, ν, Γ) in (4.14)–(4.15) for $\text{BM}_m^{\text{bal}}(\Psi)$, by specifying $\{\phi_k\}$ and completing the the proof of Theorem 4.2. The following is the analogue of Lemma 12.1 in the supplementary material [9] for $\text{BM}_m^{\text{bal}}(\Psi)$. The proof is similar and is omitted.

LEMMA B.1. *Let* $\gamma_k := \sqrt{(4c_1 \log n)/\bar{p}_k}$ *and* $\zeta_{k\ell} := \sqrt{(4c_2 \log n)/\bar{q}_{k\ell}}$. *Assume* $\gamma_k, \zeta_{k\ell} \in [0, 3]$. *Then*

$$d_i(S_k) \geq \bar{p}_k(1 - \gamma_{k\ell}), \qquad i \in S_k, \forall k \qquad \textit{w.p. at least } 1 - n^{-(c_1 - 1)}, \textit{ and}$$

$$\big|d_i(S_\ell) - \bar{q}_{k\ell}\big| \leq \zeta_{k\ell}\bar{q}_{k\ell}, \qquad i \in S_k, \forall (k \neq \ell), \qquad \textit{w.p. at least } 1 - 2m^{-1}n^{-(c_2 - 2)}.$$

We also have the following corollary of Proposition 4.1 for $\mathrm{BM}_m^{\mathrm{bal}}(\Psi)$. Recall the chain of definitions and equivalences: $\bar{q}^*_{\max} := \max_k \bar{q}^*_k = \max_{k\neq\ell} \bar{q}_{k\ell} = m(\max_{k\neq\ell} q_{k\ell}) =: mq_{\max}$.

COROLLARY B.1. *Let $A \in \{0,1\}^{n\times n}$ be distributed as $\mathrm{BM}_m^{\mathrm{bal}}(\Psi)$ and $\Delta := A - \mathbb{E}A$. Assume that $p_k \geq (C' \log m)/m$ for all k and $q_{\max} \geq (C' \log n)/n$. Then*:

- $\max_k |||\Delta_{S_k}||| \leq C\sqrt{\bar{p}_k}$, *w.p. at least* $1 - cKm^{-r}$.
- $|||\boldsymbol{E}_{S_0^c} \circ \Delta||| \leq C\sqrt{\bar{q}^*_{\max} K}$ *w.p. at least* $1 - cn^{-r}$.

PROOF. The assertion about diagonal blocks follows as in Corollary 12.1 of the supplementary material [9]. For the second assertion, we note that $\boldsymbol{E}_{S_0^c} \circ \Delta$ is an $n \times n$ matrix whose entries have variance $\leq \max_{k,\ell}(q_{k,\ell}) = \max_{k,\ell}(\bar{q}_{k,\ell}/m) = \bar{q}^*_{\max}/m$, hence w.p. at least $1 - cn^{-r}$, $|||\boldsymbol{E}_{S_0^c} \circ \Delta||| \leq C\sqrt{(\bar{q}^*_{\max}/m)n}$. □

According to Lemma B.1, for sufficiently small γ_k and $\zeta_{k,\ell}$, we have w.h.p. that $d_{\mathrm{av}}(S_k, S_\ell)$ also lies in $[\bar{q}_{k\ell}(1-\zeta_{k\ell}), \bar{q}_{k\ell}(1+\zeta_{k\ell})]$, for $k \neq \ell$, so that

$$\begin{aligned} d_i(S_\ell) + d_j(S_k) - d_{\mathrm{av}}(S_k, S_\ell) &\leq \bar{q}_{k\ell}(1 + 3\zeta_{k\ell}) \\ &\leq \bar{q}_{k\ell} + 3\sqrt{4c_2\bar{q}_{k\ell}\log n}, \qquad (i,j) \in S_k \times S_\ell. \end{aligned}$$

Note that right-hand side is increasing in $\bar{q}_{k\ell}$. We also note the following key inequality $\bar{q}_{k\ell} \leq \frac{1}{2}(\bar{q}^*_k + \bar{q}^*_\ell)$ obtained by summing the following two inequalities:

$$\frac{1}{2}\bar{q}_{k\ell} \leq \frac{1}{2}\max_{r=k, s\neq k} \bar{q}_{rs}, \qquad \frac{1}{2}\bar{q}_{k\ell} \leq \frac{1}{2}\max_{r\neq\ell, s=\ell} \bar{q}_{rs}$$

which hold for $k \neq \ell$. Hence,

$$\begin{aligned} \bar{q}_{k\ell} + 3\sqrt{4c_2\bar{q}_{k\ell}\log n} &\leq \frac{1}{2}(\bar{q}^*_k + \bar{q}^*_\ell) + 3\sqrt{2c_2(\bar{q}^*_k + \bar{q}^*_\ell)\log n} \\ &\leq \frac{1}{2}\Big(\bar{q}^*_k + 6\sqrt{2c_2\bar{q}^*_k\log n}\Big) + \frac{1}{2}\Big(\bar{q}^*_\ell + 6\sqrt{2c_2\bar{q}^*_\ell\log n}\Big), \end{aligned}$$

where we have used $\sqrt{x+y} \leq \sqrt{x} + \sqrt{y}$ for $x, y \geq 0$. Thus, taking

$$\phi_k := \frac{1}{m}\bar{\phi}_k, \qquad \bar{\phi}_k := \bar{q}^*_k + 6\sqrt{2c_2\bar{q}^*_k\log n}$$

satisfies (4.16). We also have $m\rho_k := \min_{i\in S_k} d_i(S_k) \geq \bar{p}_k - \sqrt{4c_1\bar{p}_k\log n}$, and from Corollary B.1, $|||\Delta_{S_k}||| \leq C\sqrt{\bar{p}_k}$ for all k. It follows that

$$\begin{aligned} (\rho_k - \phi_k)m - |||\Delta_k||| &\geq \bar{p}_k - \sqrt{4c_1\bar{p}_k\log n} - \Big(\bar{q}^*_k + 6\sqrt{2c_2\bar{q}^*_k\log n}\Big) - C\sqrt{\bar{p}_k} \\ &\geq (\bar{p}_k - \bar{q}^*_k) - (C + \sqrt{4c_1})\sqrt{\bar{p}_k\log n} - 6\sqrt{2c_2\bar{q}^*_k\log n}. \end{aligned}$$

By Corollary B.1, $\|\|\boldsymbol{E}_{S_0^c} \circ \Delta\|\| \le C\sqrt{\bar{q}^*_{\max} K}$. Thus, to satisfy (4.17), it is enough to have

$$\min_k \Big[(\bar{p}_k - \bar{q}^*_k) - (C + \sqrt{4c_1})\sqrt{\bar{p}_k \log n} - 6\sqrt{2c_2 \bar{q}^*_k \log n} \Big] > C\sqrt{\bar{q}^*_{\max} K},$$

which is implied by

$$\min_k \Big[(\bar{p}_k - \bar{q}^*_k) - C_2\big(\sqrt{\bar{p}_k \log n} + \sqrt{\bar{q}^*_k \log n}\big) \Big] > C\sqrt{\bar{q}^*_{\max} K}. \tag{B.1}$$

Auxiliary conditions we needed on $\bar{p}_k$ and $\bar{q}_{k\ell}$ were $\bar{p}_k \ge (4c_1/9)\log n$ and $\bar{q}_{k\ell} \ge (4c_2/9)\log n$ from Lemma B.1 and $\bar{p}_k \ge C' \log m$ and $nq_{\max} > C' \log n$. As before, we can drop the lower bounds on $\{q_{k\ell}\}_{k \ne \ell}$ due to Corollary 4.2. The lower bounds on $\bar{p}_k$ are implied by $\bar{p}_k \ge (C' \vee (4c_1/9)) \log n$. This completes the proof. To get to the form in which the theorem is stated, replace c_1 with $c_1 + 1$ and c_2 with $c_2 + 2$, and divide (B.1) by $\log n$.

Acknowledgments. We would like to thank Karl Rohe and Dave Choi for interesting discussions regarding the possibility of extending SDPs to mixed models.

SUPPLEMENTARY MATERIAL

Supplement to "On semidefinite relaxations for the block model" (DOI: 10.1214/17-AOS1545SUPP; .pdf). This supplement contains proofs of some of the results.

REFERENCES

[1] ABBE, E., BANDEIRA, A. S. and HALL, G. (2016). Exact recovery in the stochastic block model. *IEEE Trans. Inform. Theory* **62** 471–487. MR3447993

[2] ABBE, E. and SANDON, C. (2015). Community detection in general stochastic block models: Fundamental limits and efficient algorithms for recovery. In 2015 *IEEE* 56*th Annual Symposium on Foundations of Computer Science—FOCS* 2015 670–688. IEEE Computer Soc., Los Alamitos, CA. MR3473334

[3] AGARWAL, N., BANDEIRA, A. S., KOILIARIS, K. and KOLLA, A. (2017). Multisection in the stochastic block model using semidefinite programming. In *Applied and Numerical Harmonic Analysis* (H. Boche, G. Caire, R. Calderbank, et al., eds.) 125–162. Birkhäuser, Cham.

[4] AIROLDI, E. M., BLEI, D. M., FIENBERG, S. E. and XING, E. P. (2008). Mixed membership stochastic blockmodels. *J. Mach. Learn. Res.* **9** 1981–2014.

[5] AIROLDI, E. M., COSTA, T. B. and CHAN, S. H. (2013). Stochastic blockmodel approximation of a graphon: Theory and consistent estimation. In *Advances in NIPS* 26, 692–700.

[6] ALDOUS, D. J. (1981). Representations for partially exchangeable arrays of random variables. *J. Multivariate Anal.* **11** 581–598. MR0637937

[7] AMES, B. P. W. and VAVASIS, S. A. (2014). Convex optimization for the planted k-disjoint-clique problem. *Math. Program.* **143** 299–337. MR3152071

[8] AMINI, A. A., CHEN, A., BICKEL, P. J. and LEVINA, E. (2013). Pseudo-likelihood methods for community detection in large sparse networks. *Ann. Statist.* **41** 2097–2122. MR3127859

[9] AMINI, A. A. and LEVINA, E. (2018). Supplement to "On semidefinite relaxations for the block model." DOI:10.1214/17-AOS1545SUPP.

[10] AMINI, A. A. and WAINWRIGHT, M. J. (2009). High-dimensional analysis of semidefinite relaxations for sparse principal components. *Ann. Statist.* **37** 2877–2921. MR2541450

[11] AWASTHI, P., BANDEIRA, A. S., CHARIKAR, M., KRISHNASWAMY, R., VILLAR, S. and WARD, R. (2015). Relax, no need to round: Integrality of clustering formulations. In *ITCS'15—Proceedings of the 6th Innovations in Theoretical Computer Science* 191–200. ACM, New York. MR3419009

[12] BANDEIRA, A. S. (2015). Random Laplacian matrices and convex relaxations. Available at arXiv:1504.03987.

[13] BANDEIRA, A. S. (2016). A note on probably certifiably correct algorithms. *C. R. Math. Acad. Sci. Paris* **354** 329–333. MR3463033

[14] BICKEL, P., CHOI, D., CHANG, X. and ZHANG, H. (2013). Asymptotic normality of maximum likelihood and its variational approximation for stochastic blockmodels. *Ann. Statist.* **41** 1922–1943. MR3127853

[15] BICKEL, P. J. and CHEN, A. (2009). A nonparametric view of network models and Newman–Girvan and other modularities. *Proc. Natl. Acad. Sci. USA* **106** 21068–21073.

[16] BURER, S. and MONTEIRO, R. D. C. (2003). A nonlinear programming algorithm for solving semidefinite programs via low-rank factorization. *Math. Program.* **95** 329–357. MR1976484

[17] CAI, T. T. and LI, X. (2015). Robust and computationally feasible community detection in the presence of arbitrary outlier nodes. *Ann. Statist.* **43** 1027–1059. MR3346696

[18] CELISSE, A., DAUDIN, J.-J. and PIERRE, L. (2012). Consistency of maximum-likelihood and variational estimators in the stochastic block model. *Electron. J. Stat.* **6** 1847–1899. MR2988467

[19] CHAUDHURI, K., CHUNG, F. and TSIATAS, A. (2012). Spectral clustering of graphs with general degrees in the extended planted partition model. In *JMLR Workshop and Conference Proceedings* 23, 35.1–35.23.

[20] CHEN, Y., SANGHAVI, S. and XU, H. (2014). Improved graph clustering. *IEEE Trans. Inform. Theory* **60** 6440–6455. MR3265033

[21] CHEN, Y. and XU, J. (2016). Statistical-computational tradeoffs in planted problems and submatrix localization with a growing number of clusters and submatrices. *J. Mach. Learn. Res.* **17** 882–938.

[22] D'ASPREMONT, A., EL GHAOUI, L., JORDAN, M. I. and LANCKRIET, G. R. G. (2007). A direct formulation for sparse PCA using semidefinite programming. *SIAM Rev.* **49** 434–448. MR2353806

[23] DECELLE, A., KRZAKALA, F., MOORE, C. and ZDEBOROVÁ, L. (2011). Asymptotic analysis of the stochastic block model for modular networks and its algorithmic applications. *Phys. Rev. E* **84** 066106.

[24] DECELLE, A., KRZAKALA, F., MOORE, C. and ZDEBOROVÁ, L. (2012). Asymptotic analysis of the stochastic block model for modular networks and its algorithmic applications. *Phys. Rev. E* **84** 066106.

[25] FEIGE, U. and OFEK, E. (2005). Spectral techniques applied to sparse random graphs. *Random Structures Algorithms* **27** 251–275. MR2155709

[26] GAO, C., LU, Y. and ZHOU, H. H. (2015). Rate-optimal graphon estimation. *Ann. Statist.* **43** 2624–2652. MR3405606

[27] GUÉDON, O. and VERSHYNIN, R. (2016). Community detection in sparse networks via Grothendieck's inequality. *Probab. Theory Related Fields* **165** 1025–1049. MR3520025

[28] HAJEK, B., WU, Y. and XU, J. (2016). Achieving exact cluster recovery threshold via semidefinite programming. *IEEE Trans. Inform. Theory* **62** 2788–2797. MR3493879

[29] HAJEK, B., WU, Y. and XU, J. (2016). Achieving exact cluster recovery threshold via semidefinite programming: Extensions. *IEEE Trans. Inform. Theory* **62** 5918–5937. MR3552431

[30] HOLLAND, P. W., LASKEY, K. B. and LEINHARDT, S. (1983). Stochastic blockmodels: First steps. *Soc. Netw.* **5** 109–137. MR0718088

[31] JAVANMARD, A., MONTANARI, A. and RICCI-TERSENGHI, F. (2016). Phase transitions in semidefinite relaxations. *Proc. Natl. Acad. Sci. USA* **113** E2218–E2223. MR3494080

[32] JOSEPH, A. and YU, B. (2016). Impact of regularization on spectral clustering. *Ann. Statist.* **44** 1765–1791.

[33] KLOPP, O., TSYBAKOV, A. B., VERZELEN, N. et al. (2017). Oracle inequalities for network models and sparse graphon estimation. *Ann. Statist.* **45** 316–354. MR3611494

[34] LE, C. M., LEVINA, E. and VERSHYNIN, R. (2015). Sparse random graphs: Regularization and concentration of the Laplacian. Preprint. Available at arXiv:1502.03049.

[35] LEI, J. and RINALDO, A. (2015). Consistency of spectral clustering in stochastic block models. *Ann. Statist.* **43** 215–237. MR3285605

[36] LUSSEAU, D. and NEWMAN, M. E. J. (2004). Identifying the role that animals play in their social networks. *Proc. R. Soc. Lond., B Biol. Sci.* **271** S477–S481.

[37] MASSOULIÉ, L. (2014). Community detection thresholds and the weak Ramanujan property. In *STOC'*14*—Proceedings of the* 2014 *ACM Symposium on Theory of Computing* 694–703. ACM, New York. MR3238997

[38] MATHIEU, C. and SCHUDY, W. (2010). Correlation clustering with noisy input. In *Proceedings of the Twenty-First Annual ACM-SIAM Symposium on Discrete Algorithms* 712–728. SIAM, Philadelphia, PA. MR2768627

[39] MOITRA, A., PERRY, W. and WEIN, A. S. (2016). How robust are reconstruction thresholds for community detection? In *STOC'*16*—Proceedings of the* 48*th Annual ACM SIGACT Symposium on Theory of Computing* 828–841. ACM, New York. MR3536617

[40] MONTANARI, A. (2016). A Grothendieck-type inequality for local maxima. Preprint. Available at arXiv:1603.04064.

[41] MONTANARI, A. and SEN, S. (2016). Semidefinite programs on sparse random graphs and their application to community detection. In *Proceedings of the Forty-Eighth Annual ACM Symposium on Theory of Computing* 814–827. ACM, New York.

[42] MOSSEL, E., NEEMAN, J. and SLY, A. (2012). Stochastic block models and reconstruction. Available at arXiv:1202.1499.

[43] MOSSEL, E., NEEMAN, J. and SLY, A. (2015). Consistency thresholds for the planted bisection model. In *Proceedings of the Forty-Seventh Annual ACM Symposium on Theory of Computing*. *STOC* '15 69–75. ACM, New York.

[44] MOSSEL, E., NEEMAN, J. and SLY, A. (2017). A proof of the block model threshold conjecture. *Combinatorica*. DOI:10.1007/s00493-016-3238-8.

[45] NOWICKI, K. and SNIJDERS, T. A. B. (2001). Estimation and prediction for stochastic blockstructures. *J. Amer. Statist. Assoc.* **96** 1077–1087. MR1947255

[46] OLHEDE, S. C. and WOLFE, P. J. (2014). Network histograms and universality of blockmodel approximation. *Proc. Natl. Acad. Sci. USA* **111** 14722–14727.

[47] PENG, J. and WEI, Y. (2007). Approximating K-means-type clustering via semidefinite programming. *SIAM J. Optim.* **18** 186–205. MR2299680

[48] PERRY, A. and WEIN, A. S. (2017). A semidefinite program for unbalanced multisection in the stochastic block model. In *Sampling Theory and Applications* (*SampTA*), 2017 *International Conference on* 64–67. IEEE, New York.

[49] QIN, T. and ROHE, K. (2013). Regularized spectral clustering under the degree-corrected stochastic blockmodel. In *Advances in Neural Information Processing Systems* 3120–3128.

[50] ROHE, K., CHATTERJEE, S. and YU, B. (2011). Spectral clustering and the high-dimensional stochastic blockmodel. *Ann. Statist.* **39** 1878–1915. MR2893856

[51] SNIJDERS, T. A. B. and NOWICKI, K. (1997). Estimation and prediction for stochastic blockmodels for graphs with latent block structure. *J. Classification* **14** 75–100. MR1449742
[52] TOMOZEI, D.-C. and MASSOULIÉ, L. (2014). Distributed user profiling via spectral methods. *Stoch. Syst.* **4** 1–43. MR3353213
[53] VU, V. Q., CHO, J., LEI, J. and ROHE, K. (2013). Fantope projection and selection: A near-optimal convex relaxation of sparse PCA. In *Advances in Neural Information Processing Systems* 2670–2678.
[54] WAINWRIGHT, M. J. (2009). Sharp thresholds for high-dimensional and noisy sparsity recovery using ℓ_1-constrained quadratic programming (Lasso). *IEEE Trans. Inform. Theory* **55** 2183–2202. MR2729873
[55] WOLFE, P. J. and OLHEDE, S. C. (2013). Nonparametric graphon estimation. Preprint. Available at arXiv:1309.5936.
[56] XING, E. P. and JORDAN, M. I. (2003). On semidefinite relaxations for normalized k-cut and connections to spectral clustering. Technical report, Univ. California, Berkeley.
[57] YAN, B. and SARKAR, P. (2016). Convex relaxation for community detection with covariates. Preprint. Available at arXiv:1607.02675.
[58] ZHANG, Y., LEVINA, E. and ZHU, J. (2015). Estimating network edge probabilities by neighborhood smoothing. Available at arXiv:1509.08588.
[59] ZHAO, Y., LEVINA, E. and ZHU, J. (2012). Consistency of community detection in networks under degree-corrected stochastic block models. *Ann. Statist.* **40** 2266–2292. MR3059083

DEPARTMENT OF STATISTICS
UNIVERSITY OF CALIFORNIA, LOS ANGELES
LOS ANGELES, CALIFORNIA 90095
USA
E-MAIL: aaamini@ucla.edu

DEPARTMENT OF STATISTICS
UNIVERSITY OF MICHIGAN
ANN ARBOR, MICHIGAN 48109-1107
USA
E-MAIL: elevina@umich.edu