

Consistent adjacency-spectral partitioning for the stochastic block model when the model parameters are unknown

Donniell E. Fishkind, Daniel L. Sussman, Minh Tang, Joshua T. Vogelstein
and Carey E. Priebe

Department of Applied Mathematics and Statistics, Johns Hopkins University

August 23, 2012

Abstract

For random graphs distributed according to a stochastic block model, we consider the inferential task of partitioning vertices into blocks using spectral techniques. Spectral partitioning using the normalized Laplacian and the adjacency matrix have both been shown to be consistent as the number of vertices tend to infinity. Importantly, both procedures require that the number of blocks and the rank of the communication probability matrix are known, even as the rest of the parameters may be unknown. In this article, we prove that the (suitably modified) adjacency-spectral partitioning procedure, requiring only an upper bound on the rank of the communication probability matrix, is consistent. Indeed, this result demonstrates a robustness to model misspecification; an overestimate of the rank may impose a moderate performance penalty, but the procedure is still consistent. Furthermore, we extend this procedure to the setting where adjacencies may have multiple modalities and we allow for either directed or undirected graphs.

1 Background and overview

Our setting is the *stochastic block model* [12, 25]—a random graph model in which a set of n vertices is randomly partitioned into K *blocks* and then, conditioned on the partition, existence of edges between all pairs of vertices are independent Bernoulli trials with parameters determined by the block membership of the pair. (The model details are specified in Section 2.1.)

The realized partition of the vertices is not observed, nor are the Bernoulli trial parameters known. However, the realized vertex adjacencies (edges) are observed, and the main inferential task is to estimate the partition of the vertices, using the realized adjacencies as a guide. Such an estimate will be called consistent if and when, in considering a sequence of realizations for $n = 1, 2, 3, \dots$ with common model parameters, it happens almost surely that the fraction of misassigned vertices converges to zero as $n \rightarrow \infty$.

Rohe et al. [20] proved the consistency of a block estimator that is based on spectral partitioning applied to the normalized Laplacian, and Sussman et al. [24] extended this to prove the consistency of a block estimator that is based on spectral partitioning applied to the adjacency matrix. Importantly, both of these procedures assume that K and the rank of M are known (where $M \in [0, 1]^{K \times K}$ is the matrix consisting of the Bernoulli parameters for all pairs of blocks), even as the rest of the parameters may be unknown. In this article, we prove that the (suitably modified) adjacency-spectral partitioning procedure, requiring only an upper bound for $\text{rank} M$, gives consistent block estimation. We demonstrate a robustness to mis-specification of $\text{rank} M$; in particular, if a practitioner overestimates the rank of M in carrying out adjacency spectral partitioning to estimate the blocks, then the consistency of the procedure is not lost. Indeed, this is a model selection result, and we provide estimators for K and prove their consistency.

Our analysis and results are valid for both directed and undirected graphs. We also allow for more than one modality of adjacency. For instance, the stochastic block model can model a social network in which the vertices are people, and the blocks are different communities within the network such that probabilities of communication between individual people are community dependent, and there is available information about several different modes of communication between the people; e.g. who phoned whom on cell phones, who phoned whom on land lines, who sent email to whom, who sent snail mail to whom, with a separate adjacency matrix for each modality of communication. Indeed, if there are

different matrices M for each mode of communication, even if there is dependence in the communications between two people across different modalities, our analysis and results will hold—provided that every pair of blocks is “probabilistically discernable” within at least one mode of communication. (This will be made more precise in Section 2.1.)

Latent space models (e.g. Hoff et al. [11]) and, specifically, random dot product models (e.g. Young and Scheinerman [26]) give rise to the stochastic block model. Indeed, the techniques that we use in this article involve generating latent vectors for a random dot product model structure which we then use in our analysis. Nonetheless, our results can be used without awareness of such random-dot-product-graph underlying structure, and we do not concern ourselves here with estimating latent vectors for the blocks. (In any event, latent vectors are not uniquely determinable here).

Consistent block estimation in stochastic block models has received much attention. Fortunato [10] and Fjallstrom [9] provide reviews of partitioning techniques for graphs in general. Consistent partitioning of stochastic block models for two blocks was accomplished by Snijders and Nowicki [23] in 1997 and for equal-sized blocks by Condon and Karp [7] in 2001. For the more general case, Bickel and Chen [1] in 2009 demonstrated a stronger version of consistency via maximizing Newman-Girvan modularity [18] and other modularities. For a growing number of blocks, Choi et al. [3] in 2010 proved consistency of likelihood based methods. In 2012, Bickel et al. [2] provided a method to consistently estimate the stochastic block model parameters using subgraph counts and degree distributions. This work and the work of Bickel and Chen [1] both consider the case of very sparse graphs.

Rohe et al. [20] in 2011 used spectral partitioning on the normalized Laplacian to consistently estimate a growing number of blocks and they allow the minimum expected degree to be at least $\Theta(n/\sqrt{\log n})$. Sussman et al. [24] extended this to prove consistency of spectral partitioning directly on the adjacency matrix for directed and undirected graphs. Finally, Rohe et al. [21] proved consistency of bi-clustering on a directed version of the Laplacian for directed graphs. Unlike modularity and likelihood based methods, these spectral partitioning methods are computationally fast and easy to implement. Our work extends these spectral partitioning results to the situation when the number of blocks and the rank of the communication matrix is unknown. We present the situation for fixed parameters, and in Section 9 we discuss possible extensions.

The adjacency matrix has been previously used for block estimation in stochastic block models by McSherry [17], who proposed a randomized algorithm when the number of blocks

as well as the block sizes are known. Coja-Oghlan [6] further investigate the methods proposed in McSherry and extend the work to sparser graphs. This method relies on bounds in the operator norm which have also been investigated by Oliveira [19] and Chung et al. [5]. In 2012, Chaudhuri et al. [4] used an algorithm similar to the one in McSherry [17] to prove consistency for the degree corrected planted partition model, a slight restriction of the degree corrected stochastic block model proposed in [15]. Notably, Chaudhuri et al. [4] do not assume the number of blocks is known and provide an alternative method to estimate the number of blocks. This represents another important line of work for model selection in the stochastic block model.

The organization of the remainder of this article is as follows. In Section 2 we describe the stochastic block model, then we describe the inferential task and the adjacency-spectral partitioning procedure for the task—when very little is known about the parameters of the stochastic block model. In Section 3 ancillary results and bounds are proven, followed in Section 4 by a proof of the consistency of our adjacency-spectral partitioning. However, through Section 4, there is an extra assumption that the number of blocks K is known. In Section 5 we provide a consistent estimator for K , and in Section 6 we prove the consistency of an extended adjacency-spectral procedure that does not assume that K is known. Indeed, at that point, the only aspect of the model parameters which is still assumed to be known is just an upper bound for the rank of the communication probability matrix M .

Bickel et al. [2] mention the work of Rohe et al. [20] as an important step, and then opine that “unfortunately this does not deal with the problem [of] how to pick a block model which is a good approximation to the nonparametric model.” Taking these words to heart, our focus in this article is on showing a robustness in the consistency of spectral partitioning in the stochastic block model when using the adjacency matrix. Our focus is on removing the need to know a priori the parameters, and to still attain consistency in partitioning. This robustness opens the door to explore principled use of spectral techniques even for settings where the stochastic block model assumptions do not strictly hold, and we anticipate more future progress in consistency results for spectral partitioning in nonparametric models.

We conclude the article with additional discussion of consistent estimation of K (Section 7), illustrative simulations (Section 8), and a brief discussion (Section 9).

2 The model, the adjacency-spectral partitioning procedure, and its consistency

2.1 The stochastic block model

The random graph setting in which we work is the *stochastic block model*, which has parameters K, ρ, M where positive integer K is the number of blocks, the *block probability vector* $\rho \in (0, 1]^K$ satisfies $\sum_{k=1}^K \rho_k = 1$, and the *communication probability matrix* $M \in [0, 1]^{K \times K}$ satisfies the model identifiability requirement that, for all $p, q \in \{1, 2, \dots, K\}$ distinct, either it holds that $M_{p,\cdot} \neq M_{q,\cdot}$ (i.e. the p th and q th rows of M are not equal) or $M_{\cdot,p} \neq M_{\cdot,q}$ (i.e. the p th and q th columns of M are not equal). The model is defined (and the parameters have roles) as follows:

There are n vertices, labeled $1, 2, \dots, n$, and they are each randomly assigned to blocks labeled $1, 2, \dots, K$ by a random *block membership function* $\tau : \{1, 2, \dots, n\} \rightarrow \{1, 2, \dots, K\}$ such that for each vertex i and block k , independently of the other vertices, the probability that $\tau(i) = k$ is ρ_k .

Then there is a random adjacency matrix $A \in \{0, 1\}^{n \times n}$ where, for all pairs of vertices i, j that are distinct, $A_{i,j}$ is 1 or 0 according as there is an i, j edge or not. Conditioned on τ , the probability of there being an i, j edge is $M_{\tau(i), \tau(j)}$, independently of the other pairs of vertices. Our analysis and results will cover both the *undirected setting* in which edges are unordered pairs (in particular, A and M are symmetric) and also the *directed setting* in which edges are ordered pairs (in particular, A and M are not necessarily symmetric). In both settings the diagonals of A are all 0's (i.e. there are no "loops" in the graph).

We assume that the parameters of the stochastic block model are not known, except for one underlying assumption; namely, that a positive integer R is known that satisfies $\text{rank} M \leq R$. (Of course, R may be taken to be $\text{rank} M$ or K if either of these happen to be known.) However, for now through Section 4, we also assume that K is known; in Section 5 we will provide a consistent estimator for K if K is not known, and then in Section 6 we utilize this consistent estimator for K to extend all of the previous procedures and results to the scenario where K is also not known (and then the only remaining assumption is our one underlying assumption that a positive integer R is known such that $\text{rank} M \leq R$).

Although the realized adjacency matrix A is observed, the block membership function τ is not observed and, indeed, the inferential task here is to estimate τ . In Section 2.2, adjacency-

spectral partitioning is used to obtain a *block assignment function* $\hat{\tau} : \{1, 2, \dots, n\} \rightarrow \{1, 2, \dots, K\}$ that serves as an estimator for τ , up to permutation of the block labels $1, 2, \dots, K$ on the K blocks. Then Theorem 1 in Section 2.3 asserts that almost always the number of misassignments $\min_{\text{bijections } \pi: \{1, 2, \dots, K\} \rightarrow \{1, 2, \dots, K\}} |\{j = 1, 2, \dots, n : \tau(j) \neq \pi(\hat{\tau}(j))\}|$ is negligible.

A more complicated scenario is where there are multiple “modalities of communication” for the vertices. Specifically, instead of one probability communication matrix, there are several probability communication matrices $M^{(1)}, M^{(2)}, \dots, M^{(S)} \in [0, 1]^{K \times K}$ which are all parameters of the model, and there are corresponding random adjacency matrices $A^{(1)}, A^{(2)}, \dots, A^{(S)} \in \{0, 1\}^{n \times n}$ such that for each *modality* $s = 1, 2, \dots, S$ and for each pair of vertices i, j that are distinct, $A_{i,j}$ is 1 with probability $M_{\tau(i), \tau(j)}^{(s)}$ independently of the other pairs of vertices but possibly with dependence across the modalities. As above, for model identifiability purposes we assume that, for each $p, q \in \{1, 2, \dots, K\}$ distinct, there exists an $s \in \{1, 2, \dots, S\}$ such that $M_{p,\cdot}^{(s)} \neq M_{q,\cdot}^{(s)}$ or $M_{\cdot,p}^{(s)} \neq M_{\cdot,q}^{(s)}$. Also, it is assumed that we know positive integers $R^{(1)}, R^{(2)}, \dots, R^{(S)}$ which are upper bounds on $\text{rank} M^{(1)}, \text{rank} M^{(2)}, \dots, \text{rank} M^{(S)}$ respectively. We will also describe next in Section 2.2 how the adjacency-spectral partitioning procedure of that section can be modified for this more complicated scenario so that Theorem 1 will still hold for it.

2.2 The adjacency-spectral partitioning procedure

The adjacency-spectral partitioning procedure that we work with is given as follows:

First, take the realized adjacency matrix A , and compute a singular value decomposition $A = [U|U_r](\Sigma \oplus \Sigma_r)[V|V_r]^T$ where $U, V \in \mathbb{R}^{n \times R}$, $U_r, V_r \in \mathbb{R}^{n \times (n-R)}$, $\Sigma \in \mathbb{R}^{R \times R}$, and $\Sigma_r \in \mathbb{R}^{(n-R) \times (n-R)}$ are such that $[U|U_r]$ and $[V|V_r]$ are each real-orthogonal matrices, and $\Sigma \oplus \Sigma_r$ is a diagonal matrix with its diagonals non-increasingly ordered $\sigma_1 \geq \sigma_2 \geq \sigma_3 \dots \geq \sigma_n$. Let $\sqrt{\Sigma} \in \mathbb{R}^{R \times R}$ denote the diagonal matrix whose diagonals are the nonnegative square roots of the respective diagonals of Σ , and then compute $X := U\sqrt{\Sigma}$ and $Y := V\sqrt{\Sigma}$.

Then, cluster the rows of X or Y or $[X|Y]$ into at most K clusters using the minimum least squares criterion, as follows: If it is known that the rows of M are pairwise not equal, then compute $C \in \mathbb{R}^{n \times R}$ which minimizes $\|C - X\|_F$ over all matrices $C \in \mathbb{R}^{n \times R}$ such that there are at most K distinct-valued rows in C , otherwise, if it is known that the columns of M are pairwise not equal, then compute $C \in \mathbb{R}^{n \times R}$ which minimizes $\|C - Y\|_F$ over all matrices

$C \in \mathbb{R}^{n \times R}$ such that there are at most K distinct-valued rows in C , otherwise compute $\mathcal{C} \in \mathbb{R}^{n \times 2R}$ which minimizes $\|C - [X|Y]\|_F$ over all matrices $C \in \mathbb{R}^{n \times 2R}$ such that there are at most K distinct-valued rows in C . (Although our analysis will assume the use of this minimum least squares criterion, note that popular clustering algorithms such as K -means will also (empirically) produce good results for our inferential task of block assignment.)

The clusters obtained are estimates for the true blocks; i.e. define the block assignment function $\hat{\tau} : \{1, 2, \dots, n\} \rightarrow \{1, 2, \dots, K\}$ such that the inverse images $\{\hat{\tau}^{-1}(i) : i = 1, 2, \dots, K\}$ partition the rows of \mathcal{C} (by index) so that rows in each part are equal-valued. This concludes the procedure.

In the more complicated scenario of multiple modalities of communication, carry out the above procedure in the same way, *mutatis mutandis*: For each modality s , compute the singular value decomposition $A^{(s)} = [U^{(s)}|U_r^{(s)}](\Sigma^{(s)} \oplus \Sigma_r^{(s)})[V^{(s)}|V_r^{(s)}]^T$ for $U^{(s)}, V^{(s)} \in \mathbb{R}^{n \times R^{(s)}}$, $U_r^{(s)}, V_r^{(s)} \in \mathbb{R}^{n \times (n - R^{(s)})}$, $\Sigma \in \mathbb{R}^{R^{(s)} \times R^{(s)}}$, and $\Sigma_r \in \mathbb{R}^{(n - R^{(s)}) \times (n - R^{(s)})}$ such that $[U^{(s)}|U_r^{(s)}]$ and $[V^{(s)}|V_r^{(s)}]$ are each real-orthogonal matrices and $\Sigma^{(s)} \oplus \Sigma_r^{(s)}$ is a diagonal matrix with its diagonals non-increasingly ordered, then define $X^{(s)} := U^{(s)}\sqrt{\Sigma^{(s)}}$ and $Y^{(s)} := V^{(s)}\sqrt{\Sigma^{(s)}}$ and then, according as the rows of all $M^{(s)}$ are known to be distinct-valued, the columns of $M^{(s)}$ are known to be distinct-valued, or neither, compute \mathcal{C} which minimizes $\|C - [X^{(1)}|X^{(2)}|\dots|X^{(S)}]\|_F$ or $\|C - [Y^{(1)}|Y^{(2)}|\dots|Y^{(S)}]\|_F$ or $\|C - [X^{(1)}|X^{(2)}|\dots|X^{(S)}|Y^{(1)}|Y^{(2)}|\dots|Y^{(S)}]\|_F$ such that there are at most K distinct-valued rows in C , and then define $\hat{\tau}$ as the partition of the vertices into K blocks according to equal-valued corresponding rows in \mathcal{C} .

2.3 Consistency of the adjacency-spectral partitioning of Section 2.2

We consider a sequence of realizations of the stochastic block model given in Section 2.1 for successive values $n = 1, 2, 3, \dots$ with all stochastic block model parameters being fixed. In this article, an event will be said to hold *almost always* if almost surely the event occurs for all but a finite number of n . The following consistency result asserts that the number of misassignments in the adjacency-spectral procedure of Section 2.2 is negligible; it will be proven in Section 4.

Theorem 1. *With the adjacency-spectral partitioning procedure of Section 2.2, for any fixed $\epsilon > \frac{3}{4}$, the number of misassignments $\min_{\text{bijections } \pi: \{1, 2, \dots, K\} \rightarrow \{1, 2, \dots, K\}} |\{j = 1, 2, \dots, n : \tau(j) \neq \pi(\hat{\tau}(j))\}|$ is almost always less than n^ϵ .*

Theorem 1 holds for all of the scenarios we described in Section 2.1; whether the edges are directed or undirected, whether there is one modality of communication or multiple modalities. It also doesn't matter if for each successive n the partition function and adjacencies are re-realized for all vertices or if instead they are carried over from previous n 's realization with just one new vertex randomly assigned to a block and just this vertex's adjacencies to the previous vertices being newly realized. (Note that if the partition function and adjacencies are re-realized for all vertices for successive n then when we invoke the Strong Law of Large Numbers we will be using the version of the Law in [14].)

In Sussman et al. [24], it was shown that if $R = \text{rank}M$ then the number of misassignments of the adjacency spectral procedure in Section 2.2 is almost always less than a constant times $\log n$ (where the constant is a function of the model parameters). Indeed, both $\log n$ and n^ϵ , when divided by the number of vertices n , converge to zero, and in that sense we can now say that whether $\text{rank}M$ is known or if it is overestimated then either way the number of misassignments of spectral-adjacency partitioning is negligible. This is a useful robustness result.

3 Ancillary results

3.1 Latent vectors and constants from the model parameters

In this section we identify relevant constants α , β , and γ which depend on the specific values of the stochastic block model parameters; these constants will be used in our analysis. We also consider a particular decomposition of a model parameter (the communication probability matrix M) into *latent vectors* which we may then usefully associate with the respective blocks.

We first emphasize that knowing the values of these constants α , β , and γ which we are about to identify and knowing the values of the latent vectors which we are about to define are not at all needed to actually **perform** the adjacency-spectral clustering procedure of Section 2.2, nor is any such knowledge needed in order to invoke and **use** the consistency result Theorem 1. These constants and latent vectors will be used here in developing the analysis and then proving Theorem 1.

The stochastic block model parameters are K , ρ , M ; the constants α , β , γ are defined as follows: Recall that $\rho_k > 0$ for all k ; choose constant $\alpha > 0$ such that $\alpha < \rho_k$ for all k .

Next, choose matrices $\mu, \nu \in \mathbb{R}^{K \times \text{Rank} M}$ such that $M = \mu\nu^T$; indeed, such matrices μ and ν (exist and) can be easily computed using a singular value decomposition of M . It is trivial to see that if any two rows of M are not equal-valued then those two corresponding rows of μ must be not equal-valued, and if any two columns of M are not equal-valued then those two corresponding rows of ν are not equal-valued. Choose constant $\beta > 0$ be such that, for all pairs of nonequal-valued rows $\mu_{k,\cdot}, \mu_{k',\cdot}$ of μ it holds that $\|\mu_{k,\cdot} - \mu_{k',\cdot}\|_2 > \beta$, and for all pairs of nonequal-valued rows $\nu_{k,\cdot}, \nu_{k',\cdot}$ of ν it holds that $\|\nu_{k,\cdot} - \nu_{k',\cdot}\|_2 > \beta$. Lastly, since μ and ν are full column rank, choose constant $\gamma > 0$ such that the eigenvalues of $\mu^T \mu$ and $\nu^T \nu$ are all greater than γ .

The rows of μ and ν are respectively called *left latent vectors* and *right latent vectors*, and are associated with the vertices as follows. The matrices $\mathcal{X} \in \mathbb{R}^{n \times \text{rank} M}$ and $\mathcal{Y} \in \mathbb{R}^{n \times \text{rank} M}$ are defined such that for all $i = 1, 2, \dots, n$, $\mathcal{X}_{i,\cdot} := \mu_{\tau(i),\cdot}$ and $\mathcal{Y}_{i,\cdot} := \nu_{\tau(i),\cdot}$. The significance of the latent vectors is that for any pair of distinct vertices i and j the probability of an i, j edge is the inner product of the left latent vector associated with i (which is $\mathcal{X}_{i,\cdot}$) with the right latent vector associated with j (which is $\mathcal{Y}_{j,\cdot}$). Of course, these latent vectors are not observed; indeed, M is not known and τ is not observed.

Finally, let $\mathcal{X}\mathcal{Y}^T = \mathcal{U}\Lambda\mathcal{V}^T$ be a singular value decomposition, i.e. $\mathcal{U}, \mathcal{V} \in \mathbb{R}^{n \times \text{rank} M}$ each have orthonormal columns and $\Lambda \in \mathbb{R}^{\text{rank} M \times \text{rank} M}$ is a diagonal matrix with diagonals ordered in nonincreasing order $\varsigma_1 \geq \varsigma_2 \geq \varsigma_3 \geq \dots \geq \varsigma_{\text{rank} M}$. It is useful to observe that $\mathcal{X}(\mathcal{Y}^T \mathcal{V} \Lambda^{-1}) = \mathcal{U}$ and $(\Lambda^{-1} \mathcal{U}^T \mathcal{X}) \mathcal{Y}^T = \mathcal{V}^T$ imply that rows of \mathcal{X} which are equal-valued correspond to rows of \mathcal{U} that are equal-valued, and rows of \mathcal{Y} which are equal-valued correspond to rows of \mathcal{V} that are equal-valued.

In the more complicated scenario of more than one communication modality these definitions are made in the same way, mutatis mutandis: For all modalities s , choose $\mu^{(s)}, \nu^{(s)} \in \mathbb{R}^{K \times \text{Rank} M^{(s)}}$ such that $M^{(s)} = \mu^{(s)} \nu^{(s)T}$, then choose $\beta > 0$ such that for every modality s and all pairs of nonequal-valued rows $\mu_{k,\cdot}^{(s)}, \mu_{k',\cdot}^{(s)}$ of $\mu^{(s)}$ it holds that $\|\mu_{k,\cdot}^{(s)} - \mu_{k',\cdot}^{(s)}\|_2 > \beta$, and for all pairs of nonequal-valued rows $\nu_{k,\cdot}^{(s)}, \nu_{k',\cdot}^{(s)}$ of $\nu^{(s)}$ it holds that $\|\nu_{k,\cdot}^{(s)} - \nu_{k',\cdot}^{(s)}\|_2 > \beta$. Choose constant $\gamma > 0$ such that all eigenvalues of $\mu^{(s)T} \mu^{(s)}$ and $\nu^{(s)T} \nu^{(s)}$ for all modalities s are greater than γ . Then, for each modality s , define the rows of $\mathcal{X}^{(s)} \in \mathbb{R}^{n \times \text{Rank} M^{(s)}}$ and $\mathcal{Y}^{(s)} \in \mathbb{R}^{n \times \text{Rank} M^{(s)}}$ to be the rows from $\mu^{(s)}$ and $\nu^{(s)}$, respectively, corresponding to the blocks of the respective vertices, and then define $\mathcal{U}^{(s)}, \mathcal{V}^{(s)}$, and $\Lambda^{(s)}$ (with ordered diagonals $\varsigma_1^{(s)}, \varsigma_2^{(s)}, \dots, \varsigma_{\text{rank} M^{(s)}}^{(s)}$) to form singular value decompositions $\mathcal{X}^{(s)} \mathcal{Y}^{(s)T} = \mathcal{U}^{(s)} \Lambda^{(s)} \mathcal{V}^{(s)T}$.

3.2 Bounds

In this section we prove a number of bounds involving A , $\mathcal{X}\mathcal{Y}^T$, their singular values and matrices constructed from components of their singular value decompositions. These bounds will then be used in Section 4 to prove Theorem 1, which asserts the consistency of the adjacency-spectral partitioning procedure of Section 2.2.

The results in this section are stated and proved for both the directed setting and the undirected setting of Section 2.1. However, we directly treat only the setting with one modality of communication; if there are multiple modalities of communication then all of the statements and proofs in this section apply to each modality separately. Some of the results in this section can be found in similar or different form in [24]; we include all necessary results for completeness, and in order to incorporate many substantive changes needed for treatment of this article's focus.

Lemma 2. *It almost always holds that $\|AA^T - \mathcal{X}\mathcal{Y}^T(\mathcal{X}\mathcal{Y}^T)^T\|_F \leq \sqrt{3}n^{3/2}\sqrt{\log n}$ and it almost always holds that $\|A^T A - (\mathcal{X}\mathcal{Y}^T)^T \mathcal{X}\mathcal{Y}^T\|_F \leq \sqrt{3}n^{3/2}\sqrt{\log n}$.*

Proof: Let $\mathcal{X}_{i,\cdot}$ and $\mathcal{Y}_{i,\cdot}$ denote the i th rows of \mathcal{X} and \mathcal{Y} , respectively. For all $i \neq j$,

$$[AA^T]_{ij} - [\mathcal{X}\mathcal{Y}^T(\mathcal{X}\mathcal{Y}^T)^T]_{ij} = \sum_{l \neq i,j} (A_{il}A_{jl} - \mathcal{X}_{i,\cdot}\mathcal{Y}_{l,\cdot}^T\mathcal{X}_{j,\cdot}\mathcal{Y}_{l,\cdot}^T) - \mathcal{X}_{i,\cdot}\mathcal{Y}_{i,\cdot}^T\mathcal{X}_{j,\cdot}\mathcal{Y}_{i,\cdot}^T - \mathcal{X}_{i,\cdot}\mathcal{Y}_{j,\cdot}^T\mathcal{X}_{j,\cdot}\mathcal{Y}_{j,\cdot}^T \quad (1)$$

Hoeffding's inequality states that if Υ is the sum of m independent random variables that take values in the interval $[0, 1]$, and if $c > 0$ then $\mathbb{P}[(\Upsilon - \mathbb{E}[\Upsilon])^2 \geq c] \leq 2e^{-\frac{2c}{m}}$. Thus, for all i, j such that $i \neq j$, if we condition on \mathcal{X} and \mathcal{Y} , we have for $l \neq i, j$ that the $m := n - 2$ random variables $A_{il}A_{jl}$ have distribution $\text{Bernoulli}(\mathcal{X}_{i,\cdot}\mathcal{Y}_{l,\cdot}^T\mathcal{X}_{j,\cdot}\mathcal{Y}_{l,\cdot}^T)$ and are independent. Thus, taking $c = 2(n - 2) \log n$ in Equation (1), we obtain that

$$\mathbb{P}[(AA^T]_{ij} - [\mathcal{X}\mathcal{Y}^T(\mathcal{X}\mathcal{Y}^T)^T]_{ij})^2 \geq 2(n - 2) \log n + 4n - 4] \leq \frac{2}{n^4}. \quad (2)$$

Integrating Equation (2) over \mathcal{X} and \mathcal{Y} yields that Equation (2) is true unconditionally. By probability subadditivity, summing over i, j such that $i \neq j$ in Equation (2), we obtain that

$$\mathbb{P}\left[\sum_{i,j:i \neq j} ([AA^T]_{ij} - [\mathcal{X}\mathcal{Y}^T(\mathcal{X}\mathcal{Y}^T)^T]_{ij})^2 \geq 2n(n - 1)(n - 2) \log n + 4n(n - 1)^2\right] \leq \frac{2n(n - 1)}{n^4}. \quad (3)$$

By the Borel-Cantelli Lemma (which states that if a sequence of events have probabilities with bounded sum then almost always the events do not occur) we obtain from Equation

(3) that almost always

$$\sum_{i,j:i \neq j} ([AA^T]_{ij} - [\mathcal{X}\mathcal{Y}^T(\mathcal{X}\mathcal{Y}^T)^T]_{ij})^2 \leq \frac{5}{2}n^3 \log n$$

and thus almost always $\|AA^T - \mathcal{X}\mathcal{Y}^T(\mathcal{X}\mathcal{Y}^T)^T\|_F^2 \leq 3n^3 \log n$ because each of the diagonals of $AA^T - \mathcal{X}\mathcal{Y}^T(\mathcal{X}\mathcal{Y}^T)^T$ are bounded in absolute value by n . The very same argument holds mutatis mutandis for $\|A^T A - (\mathcal{X}\mathcal{Y}^T)^T \mathcal{X}\mathcal{Y}^T\|_F^2$. \square

The next lemma, Lemma 3, provides bounds on the singular values $\varsigma_1, \varsigma_2, \varsigma_3, \dots$ of matrix $\mathcal{X}\mathcal{Y}^T$ and then, in Corollary 4, we obtain bounds on the singular values $\sigma_1, \sigma_2, \sigma_3, \dots$ of matrix A . Recall that the rank of $\mathcal{X}\mathcal{Y}^T$ is (almost always) $\text{rank} M$, while A may in fact have rank n .

Lemma 3. *It almost always holds that $\alpha\gamma n \leq \varsigma_{\text{rank} M}$, and it always holds that $\varsigma_1 \leq n$.*

Proof: Because $\mathcal{X}\mathcal{Y}^T$ is in $[0, 1]^{n \times n}$, the nonnegative matrix $\mathcal{X}\mathcal{Y}^T(\mathcal{X}\mathcal{Y}^T)^T$ has all of its entries bounded by n , thus all of its row sums bounded by n^2 , and thus its spectral radius ς_1^2 is bounded by n^2 , ie we have $\varsigma_1 \leq n$ as desired.

Next, for all $k = 1, 2, \dots, K$, let random variable n_k denote the number of vertices in block k . The nonzero eigenvalues of $(\mathcal{X}\mathcal{Y}^T)(\mathcal{X}\mathcal{Y}^T)^T = \mathcal{X}\mathcal{Y}^T \mathcal{Y} \mathcal{X}^T$ are the same as the nonzero eigenvalues of $\mathcal{Y}^T \mathcal{Y} \mathcal{X}^T \mathcal{X}$. By the definition of α and the Law of Large Numbers, almost always $n_k > \alpha n$ for each k , thus we express $\mathcal{X}^T \mathcal{X} = \sum_{k=1}^K n_k \mu_{k,\cdot}^T \mu_{k,\cdot} = \alpha n \mu^T \mu + \sum_{k=1}^K (n_k - \alpha n) \mu_{k,\cdot}^T \mu_{k,\cdot}$ as the sum of two positive semidefinite matrices and obtain that the minimum eigenvalue of $\mathcal{X}^T \mathcal{X}$ is at least $\alpha\gamma n$. Similarly the minimum eigenvalue of $\mathcal{Y}^T \mathcal{Y}$ is at least $\alpha\gamma n$. The minimum eigenvalue of a product of positive semidefinite matrices is at least the product of their minimum eigenvalues [27], thus the minimum eigenvalue of $\mathcal{Y}^T \mathcal{Y} \mathcal{X}^T \mathcal{X}$ (which is equal to $\varsigma_{\text{rank} M}^2$) is at least $\alpha\gamma n \cdot \alpha\gamma n$, as desired. \square

Corollary 4. *It almost always holds that $\alpha\gamma n \leq \sigma_{\text{rank} M}$, it always holds that $\sigma_1 \leq n$, and it almost always holds that $\sigma_{\text{rank} M+1} \leq 3^{1/4} n^{3/4} \log^{1/4} n$.*

Proof: By Lemma 2 and Weyl's Lemma (e.g., see [13]), we obtain that for all m it almost always holds that $|\sigma_m^2 - \varsigma_m^2| \leq \|AA^T - \mathcal{X}\mathcal{Y}^T(\mathcal{X}\mathcal{Y}^T)^T\|_F \leq \sqrt{3}n^{3/2}\sqrt{\log n}$. For all $m > \text{rank} M$, the m th singular value of $\mathcal{X}\mathcal{Y}^T$ is zero, thus almost always $\sigma_{\text{rank} M+1} \leq 3^{1/4} n^{3/4} \log^{1/4} n$. Lemma 3 can in fact be strengthened to show that there is an $\delta > 0$ such that almost

always $(\alpha\gamma + \delta)n \leq \varsigma_{\text{rank}M}$, hence $(\alpha\gamma + \delta)^2 n^2 \leq \varsigma_{\text{rank}M}^2$, thus we have almost always that $(\alpha\gamma)^2 n^2 \leq \sigma_{\text{rank}M}^2$, as desired. Showing that $\sigma_1 \leq n$ is done the same way that $\varsigma_1 \leq n$ was shown in Lemma 3. \square

It is worth noting that a consequence of Corollary 4 is that, for any chosen real number ω such that $\frac{3}{4} < \omega < 1$, the random variable which counts the number of $\sigma_1, \sigma_2, \dots, \sigma_n$ which are greater than n^ω is a consistent estimator for $\text{rank}M$ (is almost always equal to $\text{rank}M$). Our goal in this article is to show a robustness result, that “overestimating” $\text{rank}M$ with R in the adjacency-spectral partitioning procedure does not ruin the consistency of the procedure.

Recall from Section 2.2 the singular value decomposition $A = [U|U_r](\Sigma \oplus \Sigma_r)[V|V_r]^T$. At this point it will be useful to further partition $U = [U_\ell|U_c]$, $V = [V_\ell|V_c]$, and $\Sigma = \Sigma_\ell \oplus \Sigma_c$ where $U_\ell, V_\ell \in \mathbb{R}^{n \times \text{rank}M}$, $U_c, V_c \in \mathbb{R}^{n \times (R - \text{rank}M)}$, $\Sigma_\ell \in \mathbb{R}^{\text{rank}M \times \text{rank}M}$, and $\Sigma_c \in \mathbb{R}^{(R - \text{rank}M) \times (R - \text{rank}M)}$. (The subscripts ℓ, c, r are mnemonics for “left”, “center”, and “right”, respectively.) Also define the matrices $X_\ell := U_\ell \sqrt{\Sigma_\ell}$, $Y_\ell := V_\ell \sqrt{\Sigma_\ell}$, $X_c := U_c \sqrt{\Sigma_c}$, $Y_c := V_c \sqrt{\Sigma_c}$, $X_r := U_r \sqrt{\Sigma_r}$, and $Y_r := V_r \sqrt{\Sigma_r}$. Referring back to the definition of X and Y in Section 2.2, note that $X = [X_\ell|X_c]$ and $Y = [Y_\ell|Y_c]$.

From the definition of β in Section 3.1 it follows that for any i and j such that $\mathcal{X}_{i,\cdot} \neq \mathcal{X}_{j,\cdot}$ (or $\mathcal{Y}_{i,\cdot} \neq \mathcal{Y}_{j,\cdot}$) it holds that $\|\mathcal{X}_{i,\cdot} - \mathcal{X}_{j,\cdot}\| \geq \beta$ (respectively, $\|\mathcal{Y}_{i,\cdot} - \mathcal{Y}_{j,\cdot}\| \geq \beta$). The next result shows how this separation extends to the rows of the singular vectors of $\mathcal{X}\mathcal{Y}^T$.

Lemma 5. *Almost always the following are true:*

For all i, j such that $\|\mathcal{X}_{i,\cdot} - \mathcal{X}_{j,\cdot}\|_2 \geq \beta$, it holds that $\|\mathcal{U}_{i,\cdot} - \mathcal{U}_{j,\cdot}\|_2 \geq \beta \sqrt{\frac{\alpha\gamma}{n}}$.

For all i, j such that $\|\mathcal{Y}_{i,\cdot} - \mathcal{Y}_{j,\cdot}\|_2 \geq \beta$, it holds that $\|\mathcal{V}_{i,\cdot} - \mathcal{V}_{j,\cdot}\|_2 \geq \beta \sqrt{\frac{\alpha\gamma}{n}}$.

For all i, j such that $\|\mathcal{X}_{i,\cdot} - \mathcal{X}_{j,\cdot}\|_2 \geq \beta$, it holds that $\|\mathcal{U}_{i,\cdot} Q \sqrt{\Sigma_\ell} - \mathcal{U}_{j,\cdot} Q \sqrt{\Sigma_\ell}\|_2 \geq \alpha\beta\gamma$ for any orthogonal matrix $Q \in \mathbb{R}^{\text{rank}M \times \text{rank}M}$.

For all i, j such that $\|\mathcal{Y}_{i,\cdot} - \mathcal{Y}_{j,\cdot}\|_2 \geq \beta$, it holds that $\|\mathcal{V}_{i,\cdot} Q \sqrt{\Sigma_\ell} - \mathcal{V}_{j,\cdot} Q \sqrt{\Sigma_\ell}\|_2 \geq \alpha\beta\gamma$ for any orthogonal matrix $Q \in \mathbb{R}^{\text{rank}M \times \text{rank}M}$.

Proof: Recall the singular value decomposition $\mathcal{X}\mathcal{Y}^T = \mathcal{U}\Lambda\mathcal{V}^T$ from Section 3.1 (where $\mathcal{U}, \mathcal{V} \in \mathbb{R}^{n \times \text{rank}M}$ each have orthonormal columns and $\Lambda \in \mathbb{R}^{\text{rank}M \times \text{rank}M}$ is diagonal). Let $\mathcal{Y}^T\mathcal{Y} = W\Delta^2 W^T$ be a spectral decomposition; that is, $W \in \mathbb{R}^{\text{rank}M \times \text{rank}M}$ is orthogonal and $\Delta \in \mathbb{R}^{\text{rank}M \times \text{rank}M}$ is a diagonal matrix with positive diagonal entries. Note that

$$(\mathcal{X}W\Delta)(\mathcal{X}W\Delta)^T = \mathcal{X}W\Delta^2 W^T \mathcal{X}^T = \mathcal{X}\mathcal{Y}^T\mathcal{Y}\mathcal{X}^T = \mathcal{U}\Lambda\mathcal{V}^T\mathcal{V}\Lambda\mathcal{U}^T = (\mathcal{U}\Lambda)(\mathcal{U}\Lambda)^T. \quad (4)$$

For any i, j distinct, let $e \in \mathbb{R}^n$ denote the vector with all zeros except for the value 1 in the i th coordinate and the value -1 in the j th coordinate. By Equation (4), we thus have that

$$\|(\mathcal{X}W\Delta)_{i,\cdot} - (\mathcal{X}W\Delta)_{j,\cdot}\|_2^2 = e^T(\mathcal{X}W\Delta)(\mathcal{X}W\Delta)^T e = e^T(\mathcal{U}\Lambda)(\mathcal{U}\Lambda)^T e = \|(\mathcal{U}\Lambda)_{i,\cdot} - (\mathcal{U}\Lambda)_{j,\cdot}\|_2^2. \quad (5)$$

From Lemma 3 and its proof, we have that the diagonals of Δ are almost always at least $\sqrt{\alpha\gamma n}$ and that the diagonals of Λ are at most n . Using this and Equation (5), we get that if i, j are such that $\|X_{i,\cdot} - X_{j,\cdot}\| \geq \beta$ then it holds that

$$\begin{aligned} \beta &\leq \|\mathcal{X}_{i,\cdot} - \mathcal{X}_{j,\cdot}\|_2 = \|(\mathcal{X}W)_{i,\cdot} - (\mathcal{X}W)_{j,\cdot}\|_2 \leq \frac{1}{\sqrt{\alpha\gamma n}} \|(\mathcal{X}W\Delta)_{i,\cdot} - (\mathcal{X}W\Delta)_{j,\cdot}\|_2 \\ &= \frac{1}{\sqrt{\alpha\gamma n}} \|(\mathcal{U}\Lambda)_{i,\cdot} - (\mathcal{U}\Lambda)_{j,\cdot}\|_2 \leq \frac{1}{\sqrt{\alpha\gamma n}} n \|\mathcal{U}_{i,\cdot} - \mathcal{U}_{j,\cdot}\|_2. \end{aligned}$$

Thus $\|\mathcal{U}_{i,\cdot} - \mathcal{U}_{j,\cdot}\|_2 \geq \beta\sqrt{\frac{\alpha\gamma}{n}}$, as desired. Now, if $Q \in \mathbb{R}^{\text{rank } M \times \text{rank } M}$ is any orthogonal matrix then, by Corollary 4,

$$\|\mathcal{U}_{i,\cdot} - \mathcal{U}_{j,\cdot}\|_2 = \|\mathcal{U}_{i,\cdot}Q - \mathcal{U}_{j,\cdot}Q\|_2 \leq \frac{1}{\sqrt{\alpha\gamma n}} \|\mathcal{U}_{i,\cdot}Q\sqrt{\Sigma_\ell} - \mathcal{U}_{j,\cdot}Q\sqrt{\Sigma_\ell}\|_2$$

which, together with $\|\mathcal{U}_{i,\cdot} - \mathcal{U}_{j,\cdot}\|_2 \geq \beta\sqrt{\frac{\alpha\gamma}{n}}$, implies $\|\mathcal{U}_{i,\cdot}Q\sqrt{\Sigma_\ell} - \mathcal{U}_{j,\cdot}Q\sqrt{\Sigma_\ell}\|_2 \geq \alpha\beta\gamma$, as desired. The same argument applies mutatis mutandis for $\|Y_{i,\cdot} - Y_{j,\cdot}\| \geq \beta$. \square

In the following, the sum of vector subspaces will refer to the subspace consisting of all sums of vectors from the summand subspaces; equivalently, it will be the smallest subspace containing all of the summand subspaces. The following theorem is due to Davis and Kahan [8] in the form presented in [20].

Theorem 6. (Davis and Kahan) *Let $H, H' \in \mathbb{R}^{n \times n}$ be symmetric, suppose $\mathcal{S} \subset \mathbb{R}$ is an interval, and suppose for some positive integer d that $\mathcal{W}, \mathcal{W}' \in \mathbb{R}^{n \times d}$ are such that the columns of \mathcal{W} form an orthonormal basis for the sum of the eigenspaces of H associated with the eigenvalues of H in \mathcal{S} , and the columns of \mathcal{W}' form an orthonormal basis for the sum of the eigenspaces of H' associated with the eigenvalues of H' in \mathcal{S} . Let δ be the minimum distance between any eigenvalue of H in \mathcal{S} and any eigenvalue of H not in \mathcal{S} . Then there exists an orthogonal matrix $Q \in \mathbb{R}^{d \times d}$ such that $\|\mathcal{W}Q - \mathcal{W}'\|_F \leq \frac{\sqrt{2}}{\delta} \|H - H'\|_F$.*

Corollary 7. *There almost always exist real orthogonal matrices $Q_U, Q_V \in \mathbb{R}^{\text{rank } M \times \text{rank } M}$ which satisfy $\|\mathcal{U}Q_U - U_\ell\|_F \leq \frac{\sqrt{6}}{\alpha^2\gamma^2} \cdot \sqrt{\frac{\log n}{n}}$ and $\|\mathcal{V}Q_V - V_\ell\|_F \leq \frac{\sqrt{6}}{\alpha^2\gamma^2} \cdot \sqrt{\frac{\log n}{n}}$. Furthermore, it holds that $\|\tilde{\mathcal{X}}_\ell - X_\ell\|_F \leq \frac{\sqrt{6}}{\alpha^2\gamma^2} \cdot \sqrt{\log n}$ and $\|\tilde{\mathcal{Y}}_\ell - Y_\ell\|_F \leq \frac{\sqrt{6}}{\alpha^2\gamma^2} \cdot \sqrt{\log n}$, where we define $\tilde{\mathcal{X}}_\ell := \mathcal{U}Q_U\sqrt{\Sigma_\ell}$ and $\tilde{\mathcal{Y}}_\ell := \mathcal{V}Q_V\sqrt{\Sigma_\ell}$.*

Proof: Take \mathcal{S} in Theorem 6 to be the interval $(\frac{1}{2}\alpha^2\gamma^2n^2, \infty)$. By Lemma 3 and Corollary 4, we have almost always that precisely the greatest $\text{rank}M$ eigenvalues of each of $H := \mathcal{X}\mathcal{Y}^T(\mathcal{X}\mathcal{Y}^T)^T$ and $H' := AA^T$ are in \mathcal{S} . By Lemma 3, almost always $\delta \geq \alpha^2\gamma^2n^2$ (for the δ in Theorem 6) so, by Lemma 2, almost always $\frac{\sqrt{2}}{\delta}\|H - H'\|_F \leq \frac{\sqrt{2}}{\alpha^2\gamma^2n^2}\sqrt{3n^{3/2}\sqrt{\log n}}$. With this, the first statements of Corollary 7 follow from the Davis and Kahan Theorem (Theorem 6). The last statements of Corollary 7 follow from postmultiplying $\mathcal{U}Q_{\mathcal{U}} - U_{\ell}$ with $\sqrt{\Sigma_{\ell}}$ and then using Corollary 4 and the definition of X_{ℓ} . \square

Now, choose $\mathcal{U}_c \in \mathbb{R}^{n \times (R - \text{rank}M)}$ and $\mathcal{U}_r \in \mathbb{R}^{n \times (n - R)}$ such that $[\mathcal{U}|\mathcal{U}_c|\mathcal{U}_r] \in \mathbb{R}^{n \times n}$ is an orthogonal matrix. In particular, note that the columns of \mathcal{U}_c together with the columns of \mathcal{U}_r form an orthonormal basis for the eigenspace associated with eigenvalue 0 in the matrix $H := \mathcal{X}\mathcal{Y}^T(\mathcal{X}\mathcal{Y}^T)^T$.

Corollary 8. *There almost always exists a real orthogonal matrix $Q \in \mathbb{R}^{(n - \text{rank}M) \times (n - \text{rank}M)}$ such that $\| [\mathcal{U}_c|\mathcal{U}_r] Q - [U_c|U_r] \|_F \leq \frac{\sqrt{6}}{\alpha^2\gamma^2} \cdot \sqrt{\frac{\log n}{n}}$. Define $\tilde{\mathcal{X}}_c \in \mathbb{R}^{n \times (R - \text{rank}M)}$ and $\tilde{\mathcal{X}}_r \in \mathbb{R}^{n \times (n - R)}$ such that $[\tilde{\mathcal{X}}_c|\tilde{\mathcal{X}}_r] := [\mathcal{U}_c|\mathcal{U}_r]Q\sqrt{\Sigma_c \oplus \Sigma_r}$. Then $\| [\tilde{\mathcal{X}}_c|\tilde{\mathcal{X}}_r] - [X_c|X_r] \|_F \leq \frac{3^{1/8}6^{1/2}}{\alpha^2\gamma^2} \cdot n^{-1/8} \log^{5/8} n$.*

Proof: The first statement of Corollary 8 is proven in the exact manner that we proved Corollary 7, except that \mathcal{S} is instead taken to be the *complement* of $(\frac{1}{2}\alpha^2\gamma^2n^2, \infty)$. The second statement of Corollary 8 follows by postmultiplying $[\mathcal{U}_c|\mathcal{U}_r]Q - [U_c|U_r]$ with $\sqrt{\Sigma_c \oplus \Sigma_r}$ and then using Corollary 4 and the definitions of X_c and X_r . \square

Note 9. *Almost always it holds that $\|\tilde{\mathcal{X}}_c\|_F \leq \sqrt{R - \text{rank}M} \cdot 3^{1/8}n^{3/8} \log^{1/8} n$.*

Proof: It is clear (with the matrix Q from Corollary 8) that $[\mathcal{U}_c|\mathcal{U}_r] Q$ has orthonormal columns, hence the Froebenius norm of the first $R - \text{rank}M$ columns is exactly $\sqrt{R - \text{rank}M}$. The result follows from postmultiplying these columns by $\sqrt{\Sigma_c}$ and using Corollary 4. \square

4 Proof of Theorem 1, consistency of the adjacency-spectral procedure of Section 2.2

In this section we prove Theorem 1. Assuming that the number of blocks K is known and that an upper bound R is known for $\text{rank}M$, Theorem 1 states that, for the adjacency-spectral

procedure described in Section 2.2, and for any fixed real number $\epsilon > \frac{3}{4}$, the number of misassignments $\min_{\text{bijections } \pi: \{1,2,\dots,K\} \rightarrow \{1,2,\dots,K\}} |\{j = 1, 2, \dots, n : \tau(j) \neq \pi(\hat{\tau}(j))\}|$ is almost always less than n^ϵ . We focus first on the scenario where there is a single modality of communication, and we also suppose for now that it is known that the rows of M are pairwise nonequal.

First, an observation: Recall from Section 3.1 that, for each vertex, the block that the vertex is a member of via the block **membership** function τ is characterized by which of the K distinct-valued rows of \mathcal{U} the vertex is associated with in \mathcal{U} . In Corollary 7, we defined $\tilde{\mathcal{X}}_\ell := \mathcal{U}Q_\mathcal{U}\sqrt{\Sigma_\ell}$. Because $\tilde{\mathcal{X}}_\ell$ is \mathcal{U} times an invertible matrix (since $\sqrt{\Sigma_\ell}$ is almost always invertible by Corollary 4), the block that the vertex is truly a member of is thus characterized by which of the K distinct-valued rows of $\tilde{\mathcal{X}}_\ell$ the vertex is associated with in $\tilde{\mathcal{X}}_\ell$. Also recall that the block which the vertex is assigned to by the block **assignment** function $\hat{\tau}$ is characterized by which of the at-most- K distinct-valued rows of \mathcal{C} the vertex is associated with in \mathcal{C} —where $\mathcal{C} \in \mathbb{R}^{n \times R}$ was defined as the matrix which minimized $\|\mathcal{C} - X\|_F$ over all matrices $\mathcal{C} \in \mathbb{R}^{n \times R}$ such that there are at most K distinct-valued rows in \mathcal{C} .

Denote by $0^{n \times (R - \text{rank} M)}$ the matrix of zeros in $\mathbb{R}^{n \times (R - \text{rank} M)}$. We next show the following:

$$\text{For any fixed } \xi > \frac{3}{8}, \text{ almost always it holds that } \|\mathcal{C} - [\tilde{\mathcal{X}}_\ell | 0^{n \times (R - \text{rank} M)}]\|_F \leq n^\xi. \quad (6)$$

Indeed, by the definition of \mathcal{C} , the fact that $[\tilde{\mathcal{X}}_\ell | 0^{n \times (R - \text{rank} M)}]$ has K distinct-valued rows, and the triangle inequality, we have that

$$\|\mathcal{C} - X\|_F \leq \|[\tilde{\mathcal{X}}_\ell | 0^{n \times (R - \text{rank} M)}] - X\|_F \leq \|[\tilde{\mathcal{X}}_\ell | \tilde{\mathcal{X}}_c] - X\|_F + \|\tilde{\mathcal{X}}_c\|_F. \quad (7)$$

Then, by two uses of the triangle inequality and then Equation (7), we have

$$\begin{aligned} \|\mathcal{C} - [\tilde{\mathcal{X}}_\ell | 0^{n \times (R - \text{rank} M)}]\|_F &\leq \|\mathcal{C} - [\tilde{\mathcal{X}}_\ell | \tilde{\mathcal{X}}_c]\|_F + \|\tilde{\mathcal{X}}_c\|_F \\ &\leq \|\mathcal{C} - X\|_F + \|X - [\tilde{\mathcal{X}}_\ell | \tilde{\mathcal{X}}_c]\|_F + \|\tilde{\mathcal{X}}_c\|_F \\ &\leq 2 \cdot \|[\tilde{\mathcal{X}}_\ell | \tilde{\mathcal{X}}_c] - X\|_F + 2 \cdot \|\tilde{\mathcal{X}}_c\|_F \end{aligned}$$

which, by Corollary 7, Corollary 8, and Note 9, is almost always bounded by

$$2 \left[\left(\frac{\sqrt{6}}{\alpha^2 \gamma^2} \cdot \sqrt{\log n} \right)^2 + \left(\frac{3^{1/8} 6^{1/2}}{\alpha^2 \gamma^2} \cdot n^{-1/8} \log^{5/8} n \right)^2 \right]^{1/2} + 2R^{1/2} 3^{1/8} n^{3/8} \log^{1/8} n,$$

which is almost always bounded by n^ξ for any fixed $\xi > \frac{3}{8}$. Thus Line (6) is shown.

Now, it easily follows from Line (6) that

$$\begin{aligned} & \text{For any fixed } \epsilon > \frac{3}{4}, \text{ the number of rows of } \mathcal{C} - [\tilde{\mathcal{X}}_\ell | 0^{n \times (R - \text{rank} M)}] \\ & \text{with Euclidean norm at least } \frac{\alpha\beta\gamma}{3} \text{ is almost always less than } n^\epsilon; \end{aligned} \quad (8)$$

indeed, if this was not true, then $\|\mathcal{C} - [\tilde{\mathcal{X}}_\ell | 0^{n \times (R - \text{rank} M)}]\|_F \geq \sqrt{n^\epsilon \left(\frac{\alpha\beta\gamma}{3}\right)^2}$ would contradict Line 6.

Lastly, form balls B_1, B_2, \dots, B_K of radius $\frac{\alpha\beta\gamma}{3}$ about the K distinct-valued rows of $[\tilde{\mathcal{X}}_\ell | 0^{n \times (R - \text{rank} M)}]$; by Lemma 5, these balls are almost always disjoint. The number of vertices which the block membership function τ assigns to each block is almost always at least αn , thus (by Line (8) and the Pigeonhole Principle) almost always each ball B_1, B_2, \dots, B_K contains exactly one of the K distinct-valued rows of \mathcal{C} . And, for any fixed $\epsilon > \frac{3}{4}$, the number of misassignments from $\hat{\tau}$ is thus almost always less than n^ϵ . Theorem 1 is now proven in the scenario where there is a single modality of communication and it is known that the rows of M are pairwise nonequal.

In the general case where there are multiple modalities of communication and/or the rows of M are not known to be pairwise nonequal, then the above proof holds *mutatis mutandis* (affecting relevant bounds by at most a constant factor); in place of X use Y or $[X|Y]$ or $[X^{(1)}|X^{(2)}|\dots|X^{(S)}]$ or $[Y^{(1)}|Y^{(2)}|\dots|Y^{(S)}]$ or $[X^{(1)}|X^{(2)}|\dots|X^{(S)}|Y^{(1)}|Y^{(2)}|\dots|Y^{(S)}]$ and in place of $[\tilde{\mathcal{X}}_\ell|\tilde{\mathcal{X}}_c]$ use $[\tilde{\mathcal{Y}}_\ell|\tilde{\mathcal{Y}}_c]$ or $[\tilde{\mathcal{X}}_\ell|\tilde{\mathcal{X}}_c|\tilde{\mathcal{Y}}_\ell|\tilde{\mathcal{Y}}_c]$ or $[\tilde{\mathcal{X}}_\ell^{(1)}|\tilde{\mathcal{X}}_c^{(1)}|\tilde{\mathcal{X}}_\ell^{(2)}|\tilde{\mathcal{X}}_c^{(2)}|\dots|\tilde{\mathcal{X}}_\ell^{(S)}|\tilde{\mathcal{X}}_c^{(S)}]$, or $[\tilde{\mathcal{Y}}_\ell^{(1)}|\tilde{\mathcal{Y}}_c^{(1)}|\tilde{\mathcal{Y}}_\ell^{(2)}|\tilde{\mathcal{Y}}_c^{(2)}|\dots|\tilde{\mathcal{Y}}_\ell^{(S)}|\tilde{\mathcal{Y}}_c^{(S)}]$, or $[\tilde{\mathcal{X}}_\ell^{(1)}|\tilde{\mathcal{X}}_c^{(1)}|\tilde{\mathcal{X}}_\ell^{(2)}|\tilde{\mathcal{X}}_c^{(2)}|\dots|\tilde{\mathcal{X}}_\ell^{(S)}|\tilde{\mathcal{X}}_c^{(S)}|\tilde{\mathcal{Y}}_\ell^{(1)}|\tilde{\mathcal{Y}}_c^{(1)}|\tilde{\mathcal{Y}}_\ell^{(2)}|\tilde{\mathcal{Y}}_c^{(2)}|\dots|\tilde{\mathcal{Y}}_\ell^{(S)}|\tilde{\mathcal{Y}}_c^{(S)}]$, as appropriate, and similar kinds of adjustments.

5 Consistent estimation for the number of blocks K

In this section we provide a consistent estimator \hat{K} for the number of blocks K , if indeed K is not known. (The only assumption used is our basic underlying assumption that an upper bound R is known for $\text{rank} M$.)

To simplify the notation, in this section we assume that there is only one modality of communication and we also assume that it is known that the rows of M are distinct-valued. These simplifying assumptions do not affect the results we obtain, and the analysis can be easily generalized to the general case in the same manner as was done at the end of Section 4.

In the adjacency-spectral partitioning procedure from Section 2.2, recall that one of the steps was to compute $\mathcal{C} \in \mathbb{R}^{n \times R}$ which minimized $\|\mathcal{C} - X\|_F$ over all matrices $C \in \mathbb{R}^{n \times R}$ such that there are at most K distinct-valued rows in C . Then the block assignment function $\hat{\tau}$ was defined as partitioning the vertices into K blocks according to equal-valued corresponding rows in \mathcal{C} . Let us now generalize the procedure of Section 2.2. Suppose that, for any fixed positive integer K' , we instead compute $\mathcal{C} \in \mathbb{R}^{n \times R}$ which minimizes $\|\mathcal{C} - X\|_F$ over all matrices $C \in \mathbb{R}^{n \times R}$ such that there are at most K' distinct-valued rows in C . Then the block assignment function $\hat{\tau}$ is defined as partitioning the vertices into K' parts (some possibly empty) according to equal-valued corresponding rows in \mathcal{C} . We shall call this adjusted procedure “the adjacency-spectral partitioning procedure from Section 2.2 with K' parts.”

Theorem 10. *Let real number ξ such that $\frac{3}{8} < \xi < \frac{1}{2}$ be chosen and fixed. For the adjacency-spectral procedure from Section 2.2 with K' parts, if $K' = K$ then almost always $\|\mathcal{C} - X\|_F \leq n^\xi$, and if $K' < K$ then almost always $\|\mathcal{C} - X\|_F > n^\xi$.*

Proof: Using Equation (7), Corollary 7, Corollary 8, and Note 9 in the manner used to prove Line (6), we obtain that almost always $\|[\tilde{\mathcal{X}}_\ell | 0^{n \times (R - \text{rank} M)}] - X\|_F \leq n^\xi$, and that if $K' = K$ then almost always $\|\mathcal{C} - X\|_F \leq n^\xi$.

However, if $K' < K$ then, as we did in Section 4, consider balls B_1, B_2, \dots, B_K of radius $\frac{\alpha\beta\gamma}{3}$ about the K distinct-valued rows of $[\tilde{\mathcal{X}}_\ell | 0^{n \times (R - \text{rank} M)}]$. By Lemma 5, these balls are almost always disjoint and, in fact, their centers are almost always at least $\alpha\beta\gamma$ distance one from the other. By the pigeonhole principle, there is at least one ball that contains none of the K' distinct-valued rows of \mathcal{C} . Together with the fact that each block almost always has more than αn vertices, we obtain almost always that $\|\mathcal{C} - [\tilde{\mathcal{X}}_\ell | 0^{n \times (R - \text{rank} M)}]\|_F \geq \sqrt{\alpha n \left(\frac{\alpha\beta\gamma}{3}\right)^2}$. Thus, almost always $\|\mathcal{C} - X\| \geq \|\mathcal{C} - [\tilde{\mathcal{X}}_\ell | 0^{n \times (R - \text{rank} M)}]\|_F - \|[\tilde{\mathcal{X}}_\ell | 0^{n \times (R - \text{rank} M)}] - X\|_F > n^\xi$.

□

Let real number ξ such that $\frac{3}{8} < \xi < \frac{1}{2}$ be chosen and fixed. Define the random variable \hat{K} to be the least positive integer K' such that for the adjacency-spectral procedure from Section 2.2 with K' parts it happens that $\|\mathcal{C} - X\|_F \leq n^\xi$. By Theorem 10, we have the following consistency result for \hat{K} .

Theorem 11. *Almost always $\hat{K} = K$.*

6 The extended adjacency-spectral partitioning procedure

The adjacency-spectral partitioning procedure of Section 2.2 assumed that an integer R was known such that $R \geq \text{rank}M$, but it also assumed that the number of blocks K was known. We next extend the adjacency-spectral partitioning procedure of Section 2.2 (we call it “the extended adjacency-spectral partitioning procedure”) so that it only has the assumption that an integer R is known such that $R \geq \text{rank}M$, and it is not assumed that K is known. The procedure is as follows:

Let real number ξ such that $\frac{3}{8} < \xi < \frac{1}{2}$ be chosen and fixed. Successively for $K' = 1, 2, 3, \dots$, do the spectral partitioning procedure of Section 2.2 with K' parts until it happens that $\|\mathcal{C} - X\|_F \leq n^\xi$, then return the $\hat{\tau}$ from the last successive iteration (i.e. the iteration where $K' = \hat{K}$).

Theorem 12. *With the extended adjacency-spectral partitioning procedure, for any fixed $\epsilon > \frac{3}{4}$, the number of misassignments $\min_{\text{bijections } \pi: \{1,2,\dots,K\} \rightarrow \{1,2,\dots,K\}} |\{j = 1, 2, \dots, n : \tau(j) \neq \pi(\hat{\tau}(j))\}|$ is almost always less than n^ϵ .*

Proof: Indeed, almost always the last value of K' (which is \hat{K}) is equal to K by Theorem 11, and then almost always the number of misassignments is less than n^ϵ by Theorem 1. \square

7 Another consistent estimator for K

In Section 5 we provided the consistent estimator \hat{K} for the number of blocks K . It was based on Theorem 10, which contrasted—for the adjacency-spectral procedure from Section 2.2 with K' parts—what would happen when $K' = K$ versus when $K' < K$. In this section we are interested in contrasting—for the adjacency-spectral procedure from Section 2.2 with K' parts—what would happen when $K' = K$ versus when $K' > K$. This yields another consistent estimator for K .

For the adjacency-spectral procedure from Section 2.2 with K' parts, the at-most K' distinct-valued rows of \mathcal{C} will be called the *centroids*, the *centroid separation* will refer to the minimum Euclidean distance between all pairs of distinct centroids, and the *minimum part size* will refer to the least cardinality of the K' parts as partitioned by $\hat{\tau}$; in particular, if one

of the parts is empty then the minimum part size is zero, whereas the centroid separation would still be positive.

Theorem 13. *For the adjacency-spectral procedure from Section 2.2 with K' parts, if $K' = K$ then almost always the minimum part size is greater than αn and the centroid separation is at least $\frac{\alpha\beta\gamma}{3}$. Let $\zeta > 0$ and $\vartheta > 0$ be any fixed real numbers. If $K' > K$ then almost always it will **not** hold that the minimum part size is greater than ϑn and the centroid separation is at least ζ .*

Proof: As we did in Section 4, consider balls B_1, B_2, \dots, B_K of radius $\frac{\alpha\beta\gamma}{3}$ about the K distinct-valued rows of $[\tilde{\mathcal{X}}_\ell | 0^{n \times (R - \text{rank} M)}]$. By Lemma 5, these balls are almost always disjoint and, in fact, their centers are almost always at least $\alpha\beta\gamma$ distance one from the other. If $K = K'$ then recall from Section 4 that almost always each ball contains exactly one centroid. By the $\alpha\beta\gamma$ separation between the balls' centers, we thus have almost always that the centroid separation is at least $\frac{\alpha\beta\gamma}{3}$. Also, by Theorem 1 there is almost always a strictly sublinear number of misassignments, hence almost always the minimum part size is greater than αn .

Now to the case of $K' > K$. Suppose by way of contradiction that the minimum part size is greater than ϑn and the centroid separation is at least ζ . Since there are strictly more centroids than balls B_1, B_2, \dots, B_K , and because of the ζ separation between the centroids, by the pigeonhole principle there is at least one centroid with distance greater than $\frac{\zeta}{3}$ from each row of $[\tilde{\mathcal{X}}_\ell | 0^{n \times (R - \text{rank} M)}]$ (these rows are the centers of the balls). Since this centroid appears as a row of \mathcal{C} more than ϑn times, this would imply that $\|\mathcal{C} - [\tilde{\mathcal{X}}_\ell | 0^{n \times (R - \text{rank} M)}]\|_F \geq \sqrt{\vartheta n \left(\frac{\zeta}{3}\right)^2}$. However we have by the triangle inequality, the definition of \mathcal{C} , and the first few line of the proof of Theorem 10 that almost always $\|\mathcal{C} - [\tilde{\mathcal{X}}_\ell | 0^{n \times (R - \text{rank} M)}]\|_F \leq \|\mathcal{C} - X\|_F + \|X - [\tilde{\mathcal{X}}_\ell | 0^{n \times (R - \text{rank} M)}]\|_F \leq \|\mathcal{C}_K - X\|_F + \|X - [\tilde{\mathcal{X}}_\ell | 0^{n \times (R - \text{rank} M)}]\|_F \leq 2n^\xi < \sqrt{\vartheta n \left(\frac{\zeta}{3}\right)^2}$ (where ξ such that $\frac{3}{8} < \xi < \frac{1}{2}$ is fixed and \mathcal{C}_K denotes what \mathcal{C} would have been if we instead did the adjacency-spectral procedure from Section 2.2 with K parts instead of K' parts), which gives us the desired contradiction. \square

With Theorem 13 we obtain another consistent estimator for K . However, we would need to assume that positive real numbers ζ and ϑ are known that satisfy $\vartheta \leq \alpha$ and $\zeta \leq \frac{\alpha\beta\gamma}{3}$. Assuming that such ζ and ϑ are indeed known, we can define the random variable \check{K} to be the greatest positive integer K' among the values $1, 2, 3, \dots, \lfloor \frac{1}{\vartheta} \rfloor$ (note that $\frac{1}{\vartheta}$ is an upper

bound on K) such that for the adjacency-spectral procedure from Section 2.2 with K' parts the minimum part size is greater than ϑn and the centroid separation is at least ζ . By Theorem 13 we immediately obtain the following consistency result for \check{K} .

Theorem 14. *Almost always $\check{K} = K$.*

In order to define \check{K} , lower bounds on $\frac{\alpha\beta\gamma}{3}$ and α need to be known in addition to an upper bound on $\text{rank}M$ that needs to be known. This contrasts with \hat{K} , for which we only need to assume that an upper bound on $\text{rank}M$ is known. (Because \hat{K} requires fewer assumptions, the extended adjacency-spectral partitioning procedure in Section 6 utilizes \hat{K} and not \check{K} .) Nonetheless, it is useful to be aware of how the adjacency-spectral procedure from Section 2.2 with K' parts changes in behavior when K' becomes greater than K —besides how it changes in behavior when K' becomes less than K . And when lower bounds on $\frac{\alpha\beta\gamma}{3}$ and α are also known then, in practice for a single value of n , we can check for $\hat{K} = \check{K}$ in order to have more confidence that their common value is indeed K .

8 A simulated example and discussion

As an illustration, consider the stochastic block model with parameters

$$K = 3, \quad \rho = \begin{bmatrix} .3 \\ .3 \\ .4 \end{bmatrix} \quad M = \begin{bmatrix} .205 & .045 & .150 \\ .045 & .205 & .150 \\ .150 & .150 & .180 \end{bmatrix}, \quad (9)$$

(in particular, there is only one modality of communication) and suppose edges are undirected. Here $\text{rank}M = 2$,

For each of the values $R = 1, 2, 3, 10, 25$ and for each number of vertices $n = 100, 200, 300, \dots, 1400$, we generated 2500 Monte Carlo replications of this stochastic block model and to each of these 2500 realizations we applied the adjacency-spectral partitioning procedure of Section 2.2 using R as the upper bound on $\text{rank}M$ (which, in the case of $R = 1$, is purposely incorrect for illustration purposes) assuming that we know $K = 3$. Note that rather than finding the actual minimum of $\|\mathcal{C} - X\|_F$, we use the K -means algorithm which approximates this minimum. The five curves in Figure 1 correspond to $R = 1, 2, 3, 10, 25$ respectively, and they plot the mean fraction of misassignments (the number of misassigned vertices divided by the total number of vertices n , such fractions averaged over the 2500 Monte Carlo replicates) along the y -axis, against the value of n along the x -axis.

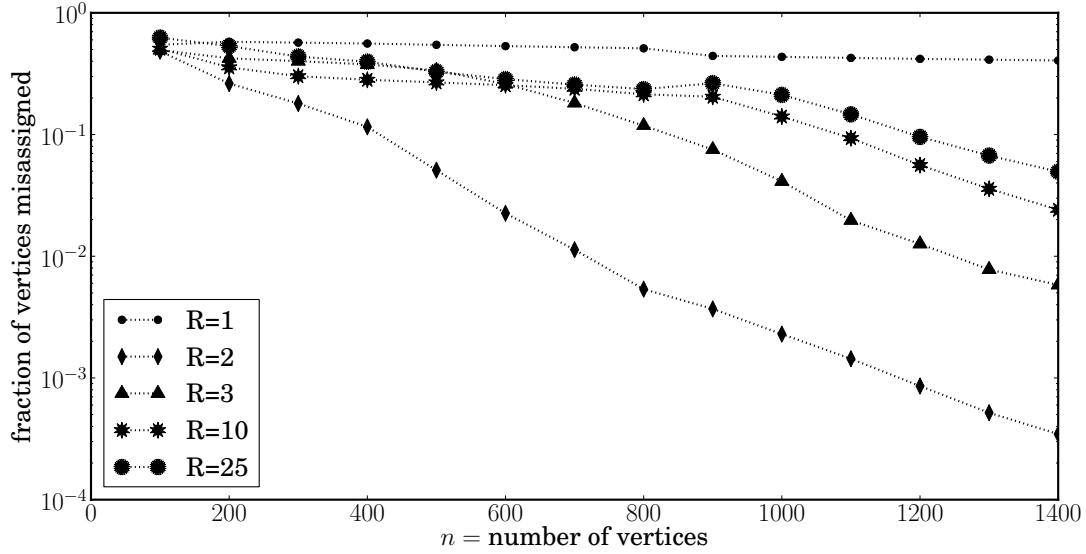


Figure 1: The mean misassignment fraction plotted against n , for each of $R = 1, 2, 3, 10, 25$.

Note that when $R = 2$ the performance of the adjacency-spectral partitioning is excellent (in fact, the number of misassignments becomes effectively zero as n gets to 1600). Indeed, even when $R = 10$ and $R = 25$ (which is substantially greater than $\text{rank}M = 2$) the adjacency-spectral partitioning performs very well. However, when $R = 1$, which is not an upper bound on $\text{rank}M$ (violating our one assumption in this article), the misassignment rate of adjacency-spectral partitioning is almost as bad as chance.

Next we will consider the estimator for K proposed in Section 5. Recall that this estimator is defined as $\hat{K} = \arg \min_{K'} \{ \|\mathcal{C}_{K'} - X\|_F \leq n^\xi \} = \arg \min_{K'} \{ \log_n(\|\mathcal{C}_{K'} - X\|_F) \leq \xi \}$ where $\mathcal{C}_{K'}$ is the $n \times R$ matrix of centroids associated with each vertex, the adjacency spectral clustering procedure in Section 2.2 is done with K' parts, and $\xi \in (3/8, 1/2)$ is fixed. We now consider stochastic block model parameters with stronger differences between blocks to illustrate the effectiveness of the estimator. In particular we let

$$K = 3, \quad \rho = \begin{bmatrix} .3 \\ .3 \\ .4 \end{bmatrix} \quad M = \begin{bmatrix} .5 & .1 & .1 \\ .1 & .5 & .1 \\ .1 & .1 & .5 \end{bmatrix} \quad (10)$$

so that $\text{rank}M = 3$. For each $n = 100, 200, 400, 800, 1600, 3200, 6400$ we generated 50 Monte Carlo replications of this stochastic block model. To each of these 50 realizations we per-

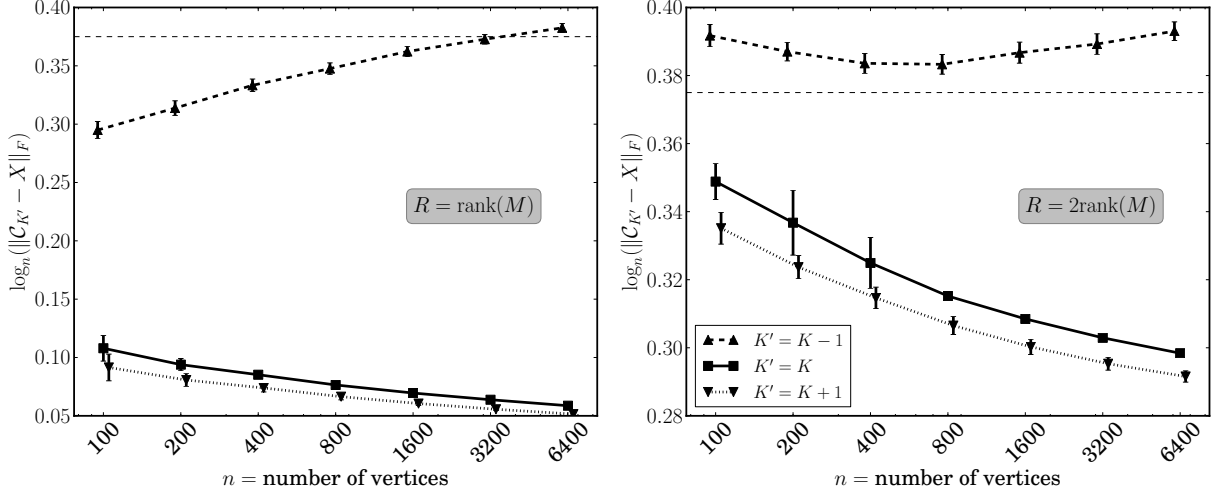


Figure 2: Test statistic for estimating K using the parameters in line (10) for $R = 3, 6$ and $K' = 2, 3, 4$. The unmarked dash line shows $\xi = 3/8$.

formed the adjacency spectral clustering procedure using $R = 3$ (Figure 2, left panel) and $R = 6$ (Figure 2, right panel) as our upper bound but this time assuming K is not known. We used $K' = 2, 3, 4$ and computed the statistic $\log_n(\|\mathcal{C}_{K'} - X\|_F)$. Figure 2 shows the mean and standard deviation of this test statistic over the 50 Monte Carlo replicates for each R , K' and n .

The results demonstrate that for $n = 6400$, \hat{K} is a good estimate when $R = 3 = \text{rank}M$ when we choose ξ close to $3/8$. On the other hand for smaller values of n , our estimator will select too few blocks regardless of the choice of $\xi \in (3/8, 1/2)$. Interestingly, choosing ξ close to $3/8$, \hat{K} always equals the true number of blocks when we let $R = 6 = 2\text{rank}M$, suggesting that this estimator has interesting behavior as a function of R . Note that for larger values of ξ , \hat{K} will tend to be smaller, and for smaller values of ξ , \hat{K} will tend to be larger.

9 Discussion

Our simulation experiment for estimating K demonstrates that good performance is possible for moderate n under certain parameter selections. This buttresses the theoretical and practical interest, as this estimator may serve as a stepping stone for the development of other more effective estimators. Indeed, bounds shown in [19] suggest that it may be possible

to allow ξ to be as small as $1/4$ using different proof methods. These methods in terms of the operator norm are an important area for further investigation when considering spectral techniques for inference on random graphs.

Note additionally that for our first simulation, we used k -means rather than minimizing $\|\mathcal{C} - X\|_F$ since the latter is computationally unfeasible. This, together with fast methods to compute the singular value decomposition, indicates that this method can be used even on quite large graphs. For even larger graphs, there are also techniques to approximate the singular value decomposition that should be considered in future work.

Further extensions of this work can be made in various directions. Rohe et al. [20] and others allow for the number of blocks to grow. We believe that this method could be extended to this scenario, though careful analysis is necessary to show that the estimator for the number of blocks is still consistent.

Another avenue is the problem of missing data, in the form of missing edges; results for this setting follow immediately provided that the edges are missing uniformly at random. This is because the observed graph will still be a stochastic block model with the same block structure. Other forms of missing data are deserving of further study. Sparse graphs are also of interest and this work can likely be extended to the case of moderately sparse graphs, for example with minimum degree $\Theta(n/\sqrt{\log n})$, without significant additional machinery. Another form of missing data is that since we consider graphs with no self-loops, the diagonal of the adjacency matrix are all zeros. Marchette et al. [16] and Scheinerman and Tucker [22] both suggest methods to impute the diagonals, and this has been shown to improve inference in practice.

This is related to one final point to mention: Is it better to do spectral partitioning on the adjacency matrix (as we do here in this article) or on the Laplacian (to be used in place of the adjacency matrix in our procedure of this article)? There doesn't currently seem to be a clear answer; for some choices of stochastic block model parameters it seems empirically that the adjacency matrix gives fewer misassignments than the Laplacian, and for other choices of parameters the Laplacian seems to be better. A determination of exact criterion (on the stochastic block model parameters) for which the adjacency matrix is better than the Laplacian and vice versa deserves attention in future work. But the analysis that we used here to reduce the required knowledge of the model parameters and to show robustness in the procedure will hopefully serve as an impetus to achieve formal results for spectral partitioning in the nonparametric setting for which the block model assumptions don't hold.

Acknowledgements: This work (all authors) is partially supported by National Security Science and Engineering Faculty Fellowship (NSSEFF) grant number N00244-069-1-0031, Air Force Office of Scientific Research (AFOSR), and Johns Hopkins University Human Language Technology Center of Excellence (JHU HLT COE). We also thank the editors and the anonymous referees for their valuable comments and critiques that greatly improved this work.

References

- [1] P.J. Bickel and A. Chen, A nonparametric view of network models and Newman-Girvan and other modularities, *Proceedings of the National Academy of Sciences of the United States of America* **106** (2009).
- [2] P.J. Bickel, A. Chen, and E. Levina, The method of moments and degree distributions for network models, *The Annals of Statistics* **39** (2011), pages 2280–2301.
- [3] D.S. Choi, P.J. Wolfe, and E.M. Airoldi, Stochastic blockmodels with growing number of classes (2010), preprint.
- [4] K. Chaudhuri, F. Chung, A. Tsiatas, Spectral Clustering of Graphs with General Degrees in the Extended Planted Partition Model, *Journal of Machine Learning Research: Workshop and Conference Proceedings*, (2012) pages 1–23
- [5] F. Chung, L. Lu, V. Vu, The spectra of random graphs with given expected degrees, *Internet Mathematics* **1** (3) (2004) 257–275.
- [6] A. Coja-Oghlan, Graph partitioning via adaptive spectral techniques, *Combinatorics, Probability and Computing* **19**(02) (2010) pages 227–284
- [7] A. Condon and R.M. Karp, Algorithms for graph partitioning on the planted partition model, *Random Structures and Algorithms* **18** (2001), pages 116–140.
- [8] C. Davis and W.M. Kahan, The rotation of eigenvectors by a perturbation III, *SIAM J. Numer. Anal.* **7** (1970), pages 1–46.

- [9] P. Fjallstrom, Algorithms for Graph Partitioning: A Survey, *Computer and Information Science*, **3**(10) (1998)
- [10] S. Fortunato, Community Detection in graphs, *Physics Reports* **486** (2010) pages 74-174
- [11] P. Hoff, A. Raftery, and M. Handcock, Latent space approaches to social network analysis. *Journal of the American Statistical Association* **97** (2002), pages 1090–1098.
- [12] P.W. Holland, K. Laskey, and S. Lienhardt, Stochastic blockmodels: First steps, *Social Networks* **5** (1983), pages 109–137.
- [13] R.A. Horn, C.R. Johnson, *Matrix Analysis*, Cambridge University Press, (1985).
- [14] T.C. Hu, F. Moricz, and R.L. Taylor, Strong laws of large numbers for arrays of rowwise independent random variables, *Acta Math. Hung* **54** (1989), pages 153–162.
- [15] B. Karrer, M. E. J. Newman, Stochastic blockmodels and community structure in networks, *Physical Review E*, **83** 1 (2011),
- [16] D. J. Marchette, C. E. Priebe, and G. Coppersmith, Vertex nomination via attributed random dot product graphs, In *Proceedings of the 57th ISI World Statistics Congress* (2011)
- [17] F. McSherry, Spectral partitioning of random graphs, *42nd IEEE Symposium on Foundations of Computer Science* (2001), pages 529–537.
- [18] M. Newman and M. Girvan, Finding and evaluating community structure in networks, *Physical Review* **69** (2004), pages 1–15.
- [19] R. I. Oliveira, Concentration of the adjacency matrix and of the laplacian in random graphs with independent edges, Arxiv preprint arXiv:0911.0600 (2010).
- [20] K. Rohe, S. Chatterjee, and B. Yu, Spectral clustering and the high-dimensional stochastic blockmodel, *The Annals of Statistics* **39** (2011), pages 1878–1915.
- [21] K. Rohe and B. Yu, Co-clustering for Directed Graphs; the Stochastic Co-Blockmodel and a Spectral Algorithm, Arxiv preprint arXiv:1204.2296 (2012).
- [22] E. Scheinerman and K. Tucker, Modeling graphs using dot product representations. *Computational Statistics*, **25** (2010).

- [23] T. Snijders and K. Nowicki, Estimation and prediction for stochastic block models for graphs with latent block structure, *Journal of Classification* **14** (1997), pages 75–100.
- [24] D. Sussman, M. Tang, D.E. Fishkind, C.E. Priebe, A consistent adjacency spectral embedding for stochastic blockmodel graphs, submitted for publication. Available at <http://arxiv.org/abs/1108.2228>
- [25] Y.J. Wang and G.Y. Wong, Stochastic blockmodels for directed graphs, *Journal of the American Statistical Association* **82** (1987).
- [26] S. Young and E. Scheinerman, Random dot product models for social networks, *Proceedings of the 5th International Conference on Algorithms and Models for the Web-graph* (2007), pages 138–149.
- [27] F. Zhang and Q. Zhang, Eigenvalue inequalities for matrix product, *IEEE Transaction on Automatic Control* **51** (2006), pages 1506–1509.