Rate-optimal community detection near the KS threshold via node-robust algorithms

Jingqiu Ding JINGQIU.DING@INF.ETHZ.CH
Yiding Hua YIDING.HUA@INF.ETHZ.CH
Kasper Lindberg KASPER.LINDBERG@INF.ETHZ.CH
David Steurer DAVID.STEURER@INF.ETHZ.CH
ETH Zürich

Aleksandr Storozhenko AS7649@PRINCETON.EDU
Princeton University

Editors: Steve Hanneke and Tor Lattimore

Abstract

We study community detection in the *symmetric k-stochastic block model*, where n nodes are evenly partitioned into k clusters with intra- and inter-cluster connection probabilities p and q, respectively. Our main result is a polynomial-time algorithm that achieves the optimal misclassification rate

$$\exp\Big(-\big(1 \pm o(1)\big)\tfrac{C}{k}\Big), \quad \text{where } C = (\sqrt{pn} - \sqrt{qn})^2,$$

whenever $C \geqslant K\, k^2 \log k$ for some universal constant K, matching the Kesten–Stigum (KS) threshold up to a $\log k$ factor. Notably, this rate holds even when an adversary corrupts an $\eta \leqslant \exp\big(-(1 \pm o(1))\tfrac{C}{k}\big)$ fraction of the nodes.

To the best of our knowledge, this optimal error rate was previously only attainable either via computationally inefficient procedures (Zhang and Zhou, 2015) or via polynomial-time algorithms that require strictly stronger assumptions such as $C \geqslant Kk^3$ (Gao et al., 2017). In the node-robust setting, the best known algorithm requires the substantially stronger condition $C \geqslant Kk^{102}$ (Liu and Moitra, 2022). Our results close this gap by providing the first polynomial-time algorithm that achieves the optimal error rate near the KS threshold in both settings.

Our work has two key technical contributions: (1) we robustify majority voting via the Sum-of-Squares framework, (2) we develop a novel graph bisectioning algorithm via robust majority voting, which allows us to significantly improve the misclassification rate to $1/\operatorname{poly}(k)$ for the initial estimation near the KS threshold.

Keywords: Stochastic Block Model, Community Detection, Sum-of-Squares, Robust Algorithms.

1. Introduction

The stochastic block model (SBM), introduced by Holland et al. (1983), is a fundamental statistical model in network analysis and has been extensively studied over the years (Abbe, 2018).

In this work, we focus on the symmetric balanced k-stochastic block model (k-SBM). Given $k, n \in \mathbb{Z}^+$, let $(\boldsymbol{G}, \boldsymbol{Z}) \sim \mathsf{SBM}_n(d, \varepsilon, k)$ denote the k-SBM with expected degree $d > 0$ and bias $\varepsilon \in [0, 1]$. The label matrix $\boldsymbol{Z} \in \{0, 1\}^{n \times k}$ is generated by partitioning the n vertices of the graph uniformly at random into k communities of size n/k each; the ith row $\boldsymbol{Z}(i)$ is the indicator vector of the community containing vertex i. Conditioned on the latent labels $\boldsymbol{Z}$, the edges between distinct vertices $i, j \in [n]$ are sampled independently with probability

$$\mathbb{P}\{(i,j) \in E(\boldsymbol{G})\} \coloneqq \begin{cases} p_1 = \left(1 + \left(1 - \frac{1}{k}\right)\varepsilon\right) \cdot \frac{d}{n}, & \text{if } \boldsymbol{Z}(i) = \boldsymbol{Z}(j), \\ p_2 = \left(1 - \frac{\varepsilon}{k}\right) \cdot \frac{d}{n}, & \text{otherwise.} \end{cases} \tag{1.1}$$

Given an observation of the graph $\boldsymbol{G} \sim \mathsf{SBM}_n(d, \varepsilon, k)$, the central statistical problem is to recover the latent community structure $\boldsymbol{Z}$. This naturally raises the question of how accurately the labels $\boldsymbol{Z}$ can be recovered as a function of ε and d; see the survey Abbe (2018) for a detailed overview.

The strongest possible goal is to recover all labels without error, a task known as *exact recovery*. This is not always possible. For instance, when $d = O(1)$, the graph $\boldsymbol{G}$ typically contains isolated vertices, whose labels cannot be inferred from the graph. Hajek et al. (2014) and Abbe et al. (2015); Abbe and Sandon (2015) showed that exact recovery is achievable by efficient algorithms above the *Chernoff–Hellinger threshold*, that is, when $C_{d,\varepsilon} \geqslant k \log n$, and is information-theoretically impossible otherwise.

When exact recovery is out of reach, a natural weaker task is to output a labeling that performs better than random guessing. This task, known as *weak recovery*, has been a central topic in the study of the SBM. The key threshold governing weak recovery is the *Kesten–Stigum (KS) threshold*, given by $\varepsilon^2 d > k^2$. For the symmetric balanced k-SBM, a sequence of works Decelle et al. (2011); Massoulié (2014); Bordenave et al. (2015); Abbe and Sandon (2016); Mossel et al. (2018) established that weak recovery can be achieved in polynomial time whenever $\varepsilon^2 d > k^2$; see also Mohanty et al. (2024) for a related above-KS algorithmic result for more general constant-community SBMs. Conversely, for fixed k, and more generally for growing $k \ll \sqrt{n}$, a rich line of work Hopkins and Steurer (2017); Hopkins (2018); Bandeira et al. (2021); Ding et al. (2025); Sohn and Wein (2025); Chin et al. (2025) suggests that weak recovery below the KS threshold, $\varepsilon^2 d < k^2$, is computationally hard.

These results delineate the regimes in which exact recovery is possible and those where only weak recovery can be hoped for. Between these two extremes, it is natural to ask: for fixed ε and d, what is the best achievable recovery rate, either information-theoretically or by polynomial-time algorithms? This question leads to the notion of the *optimal misclassification rate* for the SBM.

Optimal misclassification rate. For a family Θ of k-SBMs, we define the misclassification error as

$$\operatorname{error}_k(\hat{Z}, \boldsymbol{Z}) \coloneqq \min_{\pi \in S_k} \frac{1}{2n} \sum_{j=1}^{k} \left\| \hat{Z}(\cdot, j) - \boldsymbol{Z}(\cdot, \pi(j)) \right\|_1 .$$

The corresponding worst-case risk is then

$$\inf_{\hat{Z}} \sup_{\theta \in \Theta} \mathbb{E}_\theta \left[\operatorname{error}_k(\hat{Z}, \boldsymbol{Z})\right] ,$$

where the infimum is taken over all estimators $\hat{Z}$, the supremum over parameters $\theta \in \Theta$, and $\mathbb{E}_\theta$ denotes expectation with respect to the k-SBM with parameter θ. For the symmetric balanced model of Eq. (1.1), the family Θ is obtained by fixing the model parameters n, k, d, ε

and varying only the planted balanced partition. Equivalently, θ specifies a partition of $[n]$ into k communities of size n/k, and conditional on this partition the edges are sampled with probabilities p_1, p_2 as in Eq. (1.1). Thus, the optimality claim in this paper is uniform over balanced community assignments, not over a larger heterogeneous SBM class with varying community sizes or edge probabilities.

Zhang and Zhou (2015) showed that, for this balanced homogeneous family in the sparse regime $d = o(n)$, whenever $C_{d,\varepsilon}/(k \log k) \to \infty$, the optimal misclassification rate is

$$\exp\left(-(1+o(1))\frac{C_{d,\varepsilon}}{k}\right),$$

where $C_{d,\varepsilon}$[1] is defined as

$$C_{d,\varepsilon} := (\sqrt{np_1} - \sqrt{np_2})^2 = \left(\sqrt{\left(1+\left(1-\frac{1}{k}\right)\varepsilon\right)d} - \sqrt{\left(1-\frac{\varepsilon}{k}\right)d}\right)^2 . \tag{1.2}$$

In the same work, they constructed a statistically optimal but computationally inefficient estimator that attains this rate for the same k-SBM family.

Polynomial-time algorithms achieving the optimal error rate are known in two regimes. For $k = 2$, Mossel et al. (2014); Fei and Chen (2019) gave polynomial-time algorithms, based respectively on belief propagation and semidefinite programming, that achieve the optimal error rate when $C_{d,\varepsilon}$ is sufficiently large. For general k, Gao et al. (2017) gave a polynomial-time algorithm based on majority voting that achieves the optimal error rate under the stronger assumption $\varepsilon^2 d \geqslant \Omega(k^3)$. Thus, there remains a gap between the KS threshold $\varepsilon^2 d > k^2$ and the threshold $\varepsilon^2 d \geqslant \Omega(k^3)$ achieved by known efficient algorithms. This motivates the following question.

Question 1 *Can the optimal error rate* $\exp(-C_{d,\varepsilon}/k)$ *be achieved by efficient algorithms under the condition* $\varepsilon^2 d \geqslant Kk^2$ *for some universal constant* K*?*

Node robustness. Robustness to adversarial perturbations is a desirable property for community detection algorithms, as real-world networks often contain corrupted or noisy data. A particularly challenging corruption model introduced in Acharya et al. (2022) is *node corruption*, where an adversary is allowed to arbitrarily modify all edges incident to an η fraction of the vertices. For the symmetric k-SBM, Liu and Moitra (2022) recently gave the first polynomial-time algorithm that achieves the optimal error rate under $\eta \leqslant \exp(-C_{d,\varepsilon}/k)$ node corruption. However, their result requires $\varepsilon^2 d \geqslant \Omega(k^{102})$, which is very far from the KS threshold. This leads to the following natural question.

Question 2 *Given an* η*-node-corrupted* k*-SBM and assuming* $\varepsilon^2 d \geqslant Kk^2$*, can we achieve the optimal error rate* $\exp(-C_{d,\varepsilon}/k)$ *in polynomial time when* $\eta \leqslant \exp(-C_{d,\varepsilon}/k)$*?*

1. $C_{d,\varepsilon}$ differs from $\varepsilon^2 d$ by a constant factor.

1.1. Results

In this work, we give the first polynomial-time node-robust algorithm that achieves the optimal error rate when $\varepsilon^2 d \geqslant \Omega(k^2 \log k)$. This resolves the above questions affirmatively up to a multiplicative $\log k$ factor. Moreover, our algorithm applies to all $k \leqslant n^{0.001}$, substantially improving over prior robust results that handle only constant k.

Theorem 3 (Informal main theorem, see Theorem 24 for the formal version) *There exist universal constants $C_0 > C_1 > 0$ and $K > 0$ such that the following holds. Let $k \leqslant n^{0.001}$, $\eta \leqslant k^{-C_0}$, and $d = o(n)$. Suppose $\varepsilon^2 d \geqslant K k^2 \log k$. Then there exists a polynomial-time algorithm that, given an η-node-corrupted graph $\boldsymbol{G} \sim \mathsf{SBM}_n(d, \varepsilon, k)$, produces, with high probability, a labeling with error at most*

$$\exp\left(-\frac{C_{d,\varepsilon}}{k} + O\left(\sqrt{d}\varepsilon\right)\right) + k^{C_1}\eta\,.$$

For comparison, the best known node-robust algorithm by Liu and Moitra (2022) achieves error

$$\exp\left(-\frac{C_{d,\varepsilon}}{k} + O\left(k^{50}\sqrt{d}\,\varepsilon\right)\right) + O(k\eta)\,,$$

which exhibits a substantially worse dependence on k. Moreover, even to obtain weak recovery, their algorithm requires $\varepsilon^2 d \geqslant k^{102}$, which is far above the KS threshold.

Open problem. We leave as an open problem whether there exist polynomial-time algorithms that achieve the optimal error rate $\exp(-C_{d,\varepsilon}/k)$ under the weaker condition $\varepsilon^2 d \geqslant \Omega(k^2)$. This problem remains open even without adversarial corruption. Note that, since $C_{d,\varepsilon}/k = \Omega(k)$ when $\varepsilon^2 d \geqslant \Omega(k^2)$, the optimal error $\exp(-C_{d,\varepsilon}/k)$ is at most $\exp(-\Omega(k))$. Thus, a positive resolution would yield the optimal error rate $\exp(-(1+o(1))C_{d,\varepsilon}/k)$; at the KS scale $\varepsilon^2 d = \Theta(k^2)$, this corresponds to an error rate of $\exp(-\Theta(k))$.

1.2. Comparison to previous work

Liu and Moitra (2022) gave the first polynomial-time algorithm that attains the optimal error rate under node corruption. Their algorithm is based on an iterative refinement framework: it starts from a weak initialization and repeatedly boosts the accuracy via an SDP-based majority-vote procedure. For this boosting step to succeed, however, their analysis requires an initialization with error at most $1/\operatorname{poly}(k)$, which they obtain only under the stringent condition $\varepsilon^2 d \geqslant \Omega(k^{102})$.

We circumvent this bottleneck by designing an initialization algorithm that reaches the same accuracy under the much milder condition $\varepsilon^2 d \geqslant \Omega(k^2 \log k)$. Our key ingredient is a new robust graph bisectioning algorithm that reduces the initialization error from a constant to $1/\operatorname{poly}(k)$. We apply this procedure recursively, reducing the k-community detection problem to a sequence of two-community problems, and thereby obtain a final initialization with error $1/\operatorname{poly}(k)$.

Building on this initialization, we design a *one-shot* boosting algorithm based on the Sum-of-Squares (SoS) method. This avoids both the iterative refinement framework and the auxiliary test-function machinery of Liu and Moitra (2022). Our use of SoS for robust community detection is inspired by the recent work of Ding et al. (2023); Chen et al. (2024a,b, 2025). The main innovation in our SoS algorithm is a new family of constraints enforcing *pairwise majority-vote consistency*, which allows us to boost the accuracy to the optimal error rate in two rounds.

2. Techniques

2.1. Two communities

For notational simplicity, we first illustrate some of the ideas underlying our algorithm for the case of two communities. Let $(\boldsymbol{G}, \boldsymbol{x}) \sim \mathsf{SBM}_n(d, \varepsilon)$ denote the stochastic block model with degree parameter $d > 0$ and bias parameter $\varepsilon \in [0, 1]$. Here, we choose $\boldsymbol{x} \in \{\pm 1\}^n$ uniformly at random among all unbiased vectors so that $\sum_{i=1}^n \boldsymbol{x}(i) = 0$, where we take $n \geqslant 2$ to be even. We can view $\boldsymbol{x}$ as a balanced bipartition of the vertex set $[n]$. We refer to its parts $\boldsymbol{x}^{-1}(1)$ and $\boldsymbol{x}^{-1}(-1)$ as blocks or communities. Then, for every pair of distinct vertices $i, j \in [n]$, we decide independently at random whether to join them by an edge with probability,

$$\mathbb{P}\{ij \in E(\boldsymbol{G}) \mid \boldsymbol{x}\} = (1 + \varepsilon \cdot \boldsymbol{x}(i) \cdot \boldsymbol{x}(j)) \cdot \frac{d}{n} \,. \tag{2.1}$$

Basic recovery approach. In the stochastic block model without adversarial corruptions, the most basic approach for approximately recovering the underlying blocks uses spectral techniques (Boppana, 1987; Guédon and Vershynin, 2016; Coja-Oghlan, 2010; Massoulié, 2014; Chin et al., 2015). As a slight abuse of notation, for an n-vertex graph G with degree parameter d, let A_G denote its ordinary adjacency matrix, with zero diagonal. For the fixed bias parameter ε, we also write G for the centered adjacency matrix

$$G := A_G - \frac{d}{n}\Big(\mathbf{1}\mathbf{1}^\top - I_n\Big) + \frac{\varepsilon d}{n} I_n. \tag{2.2}$$

Equivalently, the off-diagonal entries of G are $1 - \frac{d}{n}$ on edges and $-\frac{d}{n}$ on nonedges, and $G(i,i) = \frac{\varepsilon d}{n}$. Then, for the stochastic block model $(\boldsymbol{G}, \boldsymbol{x}) \sim \mathsf{SBM}_n(d, \varepsilon)$,

$$\mathbb{E}[\boldsymbol{G} \mid \boldsymbol{x}] = \tfrac{\varepsilon d}{n} \cdot \boldsymbol{x}\boldsymbol{x}^\mathsf{T} \,. \tag{2.3}$$

At the same time, conditioned on $\boldsymbol{x}$, the matrix $\boldsymbol{E} := \frac{n}{\varepsilon d}\boldsymbol{G} - \boldsymbol{x}\boldsymbol{x}^\mathsf{T}$ has zero diagonal and, up to symmetry, independent mean-zero off-diagonal entries, and enjoys a spectral norm bound $\|\boldsymbol{E}\| \lesssim n/(\varepsilon\sqrt{d})$ for $d \gg \log n$. In this way, the best rank-one approximation of G allows us to approximately recover $\boldsymbol{x}$ up to sign as long as $\varepsilon \cdot \sqrt{d} \gg 1$ and $d \gg \log n$. After carefully truncating high-degree nodes, we can even drop the logarithmic-degree condition $d \gg \log n$. It turns out that for approximately recovering the underlying blocks, the only remaining condition, $d \gg 1/\varepsilon^2$, is also information-theoretically necessary.

Sum-of-squares lens on the basic approach. It will be instructive to formulate this basic approach in terms of the sum-of-squares method. While this sum-of-squares formulation is certainly overkill for approximate recovery in the vanilla version of the model, it will give us node-robustness essentially for free (whereas basic spectral techniques or even semidefinite programming techniques cannot handle node corruption). The basic spectral algorithms and majority voting algorithms are known to be brittle under node corruptions, while the basic semidefinite programming algorithms are only known to be robust against $O(nd\varepsilon)$ corrupted edges.

As a starting point for our sum-of-squares formulation, we note that a spectral norm bound $\|E\| \leqslant \lambda \cdot n$ certifies the following universally quantified inequality over the hypercube,

$$\forall u, v \in \{\pm 1\}^n, \quad \langle u, Ev \rangle \leqslant \lambda \cdot n^2 \,. \tag{2.4}$$

Moreover, this spectral norm bound constitutes a degree-2 sum-of-squares proof of the above inequality.

For our purposes, it will be useful to introduce sum-of-squares proofs as mathematical objects in their own right. For n-variate polynomials $p, q_1, \ldots, q_m \in \mathbb{R}[x]$, we say that a real-valued square matrix R is a *degree-ℓ sum-of-squares proof* of the inequality $p \geqslant 0$ subject to the constraints $q_1 \geqslant 0, \ldots, q_m \geqslant 0$ if

$$p = \operatorname{Tr} RR^{\mathsf{T}} \cdot \operatorname{Mom}_\ell(q_1, \ldots, q_m) \,, \tag{2.5}$$

where $\operatorname{Mom}_\ell(q_1, \ldots, q_m)$ is a block-diagonal matrix with entries in $\mathbb{R}[x]$ and (possibly empty) blocks M_S indexed by multisubsets $S \subseteq [m]$ such that for all pairs of monomials x^α, x^β with $\deg x^\alpha, \deg x^\beta \leqslant \frac{1}{2} \cdot (\ell - \sum_{i \in S} \deg q_i)$,

$$M_S(\alpha, \beta) := \left(\prod\nolimits_{i \in S} q_i\right) \cdot x^\alpha \cdot x^\beta \,. \tag{2.6}$$

We remark that sum-of-squares proofs are more commonly defined in terms of sum-of-squares polynomial multipliers for (products of) the constraint polynomials $q_1, \ldots, q_m$. However, it will be beneficial for our purposes to represent them as matrices (corresponding to the monomial-coefficients of the polynomial multipliers). Crucially, our definition makes it apparent that sum-of-squares proofs are solutions to a system of quadratic equations of Eq. (2.6).

Note that each block M_S has size $O(n^{\ell/2})$ by $O(n^{\ell/2})$ and the number of non-empty blocks is $O(m^\ell)$ (as we may assume the polynomials $q_1, \ldots, q_m$ to have degree at least 1). Therefore, R has size $O(nm)^\ell$ by $O(nm)^\ell$ and the number of quadratic equations in Eq. (2.5) is polynomial in n and m for every constant ℓ.

We denote that R is a *degree-ℓ sum-of-squares proof* of the inequality $p \geqslant 0$ subject to the constraints $q_1 \geqslant 0, \ldots, q_m \geqslant 0$ by

$$R\colon \{q_1(x) \geqslant 0, \ldots, q_m(x) \geqslant 0\} \,\Big|\frac{x}{\ell}\, p(x) \geqslant 0 \,. \tag{2.7}$$

If the variables x quantified by the proof or the degree bound ℓ are clear from the context, we may drop them in the above notation.

A fundamental property of sum-of-squares proofs is that they are automatizable. If a matrix R as above exists, we can find it in time polynomial in its bit complexity (Barak and Steurer, 2014, 2016; Raghavendra et al., 2018). A simple but very important consequence for estimation problems is that given as input a system of polynomial constraints $\mathcal{A}(x)$, we can compute in polynomial-time a vector $\hat{x}$ that satisfies all convex polynomial inequalities that can be derived from $\mathcal{A}(x)$ by a sum-of-squares proof with polynomial bit complexity.

In this way, referred to as the *proof-to-algorithm paradigm*, we obtain efficient estimation algorithms from the mere existence of small identifiability proofs.

In order to apply this terminology for approximate recovery of the stochastic block model, we consider the following polynomial constraints for the set of possible label assignments (for the case of two communities),

$$\mathcal{A}_{\text{label}}(x) := \left\{ x_1^2 = \ldots = x_n^2 = 1, \sum\nolimits_{i=1}^{n} x_i = 0 \right\}. \tag{2.8}$$

Next, we consider a system of quadratic equations for matrices E that are certifiably pseudo-random, in the sense that they are uncorrelated with all possible label assignments,

$$\mathcal{A}_{\text{rand}}(E, R; \lambda) := \left\{ R \colon \mathcal{A}_{\text{label}}(u), \mathcal{A}_{\text{label}}(v) \,\Big|\frac{u,v}{2}\, \langle u, Ev \rangle \leqslant \lambda \cdot n^2 \right\}. \tag{2.9}$$

Here, the auxiliary variables R represent the sum-of-squares certificate for the correlation bound. When we combine systems of polynomial constraints, we usually take these auxiliary variables to be unrelated. In this case, we do not keep track of these variables explicitly and suppress them in the above notation. We emphasize that the system Eq. (2.9) has only E and R as variables but not u or v. Instead u, v determine the quadratic equations we impose for the variables E and R. We also remark that Eq. (2.9) is a proxy for requiring that E satisfies the universally quantified inequality Eq. (2.4). In this sense, Eq. (2.9) allows us to encode 2^n constraints as a polynomial number of constraints.

With the above two polynomial systems at hand, we can mimic the basic recovery approach by the following combined system,[2]

$$\mathcal{A}_{\text{basic}}(G, x; \lambda) := \mathcal{A}_{\text{label}}(x) \cup \mathcal{A}_{\text{rand}}\left(\tfrac{n}{\varepsilon d} G - xx^{\mathsf{T}}; \lambda\right), \tag{2.10}$$

where we choose $\lambda \lesssim 1/(\varepsilon\sqrt{d})$. These constraints are useful for recovery because for the same graph G, essentially only one label assignment x can satisfy the constraints. Concretely, we have

$$\mathcal{A}_{\text{basic}}(G, x; \lambda), \mathcal{A}_{\text{basic}}(G, x'; \lambda) \,\Big|\frac{G,x,x'}{16}\, \tfrac{1}{n^2} \langle x, x' \rangle^2 \geqslant 1 - 2\lambda\,. \tag{2.11}$$

We emphasize that the above fact is deterministic and thus independent of any statistical model. Ignoring some details, its proof boils down to the following direct calculation,

$$n^2 - \langle x, x' \rangle^2 = \left\langle xx^{\mathsf{T}} - (x')(x')^{\mathsf{T}}, xx^{\mathsf{T}} \right\rangle = \left\langle E' - E, xx^{\mathsf{T}} \right\rangle \tag{2.12}$$

$$= \langle x, E'x \rangle + \langle -x, Ex \rangle\,, \tag{2.13}$$

where $E = \frac{n}{\varepsilon d} G - xx^{\mathsf{T}}$ and $E' = \frac{n}{\varepsilon d} G - (x')(x')^{\mathsf{T}}$. At the same time, the pseudorandomness constraints Eq. (2.9) allow us to bound both terms on the right in the same way by $\lambda \cdot n^2$,

$$\mathcal{A}_{\text{rand}}(E', R; \lambda), \mathcal{A}_{\text{label}}(x) \,\Big|\frac{E',R,x}{L}\, \langle x, E'x \rangle = \lambda n^2 - \operatorname{Tr} RR^{\mathsf{T}} \cdot \operatorname{Mom}_2(\mathcal{A}_{\text{label}}(x)) \tag{2.14}$$

$$\leqslant \lambda n^2\,. \tag{2.15}$$

2. In the constraint system $\mathcal{A}_{\text{basic}}$ we choose to treat the graph G as a variable. As a consequence, our sum-of-squares proof will be for statements universally quantified over G. While this property would not be necessary for recovery in the non-robust setting, it will turn out to be crucial in the robust setting.

In the second step, we use that $\mathcal{A}_{\text{rand}}(E', R)$ contains the constraint $\lambda \cdot n^2 - \langle u, E'v\rangle = \operatorname{Tr} RR^{\mathsf{T}} \cdot \operatorname{Mom}_2(\mathcal{A}_{\text{label}}(u), \mathcal{A}_{\text{label}}(v))$, where u, v are formal variables, and that we can substitute x for both u and v. In the third step, we use that $\operatorname{Tr} RR^{\mathsf{T}} \cdot \operatorname{Mom}_2(\mathcal{A}_{\text{label}}(x))$ is a sum-of-squares polynomial in variables x and R modulo the constraints we have on x[3]. Combining the previous two calculations, we derive a bound of 2λ on the distance $1 - \langle x, x'\rangle^2/n^2$ from the constraints $\mathcal{A}_{\text{basic}}(G, x), \mathcal{A}_{\text{basic}}(G, x')$.

The stochastic block model $(\boldsymbol{G}, \boldsymbol{x}) \sim \mathsf{SBM}_n(d, \varepsilon)$ satisfies the constraints $\mathcal{A}_{\text{basic}}(G, x; \lambda)$ for $\lambda \lesssim (\varepsilon\sqrt{d})^{-1}$ with high probability if $\varepsilon\sqrt{d} \gg 1$ (even for constant degree d (Guédon and Vershynin, 2016)). Thus, the now standard proof-to-algorithm paradigm for sum-of-squares (Ding et al., 2022, 2023) yields a polynomial-time algorithm to recover the underlying blocks up to an error of $O\big(\varepsilon\sqrt{d}\big)^{-1}$.

Handling node corruptions using sum-of-squares. We will now see that the sum-of-squares formulation Eq. (2.11) of the basic recovery approach directly implies that we can handle a constant fraction of node corruptions.

To this end, we formulate the node distance between graphs in terms of polynomial equations,

$$\mathcal{A}_{\text{close}}(G_1, G_2, z; \eta) := \left\{ \mathrm{D}(z)^2 = \mathrm{D}(z), \mathrm{D}(z) \cdot (G_1 - G_2) \cdot \mathrm{D}(z) = 0, \operatorname{Tr} \mathrm{D}(z) = (1 - \eta) \cdot n \right\}, \tag{2.16}$$

where $D(z)$ denotes the diagonal matrix with z on the diagonal. A pair of graphs G_1 and G_2 satisfy the above constraints for some z if and only if the two graphs differ in at most $\eta \cdot n$ nodes. Here, the binary vector z indicates the set of nodes where the two graphs agree. As before, we will omit the auxiliary variables z in the above notation whenever there are no relations to other variables.

To handle node corruptions, we use that $\mathcal{A}_{\text{basic}}$ is Lipschitz with respect to this node distance and restrictions,

$$\mathcal{A}_{\text{close}}(G, G', z; \eta), \mathcal{A}_{\text{basic}}(G, x; \lambda) \,\Big|\frac{G, G', x, z}{16}\, \mathcal{A}_{\text{basic}}\big(\mathrm{D}(z) \cdot G' \cdot \mathrm{D}(z), x; \lambda + 2\eta\big). \tag{2.17}$$

The above statement is not surprising: If we have a good label assignment x for G, then it continues to be one if we were to remove a small fraction of nodes (the complement of the set indicated by z). After removing this set of nodes, the two graphs G and G' agree. (Note that, in general, we cannot hope for G' to satisfy the pseudorandomness constraints $\mathcal{A}_{\text{rand}}$ with respect to x and that the restriction to the vertex set indicated by z is necessary.) Letting $E = \frac{n}{\varepsilon d} G - xx^{\mathsf{T}}$ and $E' = \frac{n}{\varepsilon d} \mathrm{D}(z) \cdot G' \cdot \mathrm{D}(z) - xx^{\mathsf{T}}$ and omitting some technical details, the above Lipschitz property follows from the fact,

$$\mathcal{A}_{\text{close}}(G, G', z; \eta), \mathcal{A}_{\text{label}}(u), \mathcal{A}_{\text{label}}(v) \,\Big|_{\overline{16}}\, \langle u, E'v\rangle \leqslant \langle u, \mathrm{D}(z) \cdot E \cdot \mathrm{D}(z) v\rangle + 2\eta n^2, \tag{2.18}$$

where $2\eta n^2$ accounts for the entries where E' and $\mathrm{D}(z) \cdot E \cdot \mathrm{D}(z)$ differ. Each of these entries is either 1 or -1 and is located in the union of ηn rows and ηn columns. Hence, the sum of the absolute values of these entries is at most $2\eta n^2$.

Given a corrupted graph $G_{\text{corrupted}}$ as input, we can recover the underlying blocks using the following polynomial constraints in variables G and x,

$$\mathcal{A}_{\text{robust}}(G, x; \lambda, \eta) := \mathcal{A}_{\text{close}}(G, G_{\text{corrupted}}; \eta) \cup \mathcal{A}_{\text{basic}}(G, x; \lambda + 4\eta). \tag{2.19}$$

3. See Theorem 55 for details

The previous Lipschitz property of $\mathcal{A}_{\text{basic}}$ together with the vanilla analysis of $\mathcal{A}_{\text{basic}}$ for recovery implies that any two solutions to the above system for the same input graph $G_{\text{corrupted}}$ are close (and this fact has low-degree sum-of-squares proofs),

$$\mathcal{A}_{\text{robust}}(G, x; \lambda, \eta), \mathcal{A}_{\text{robust}}(G', x'; \lambda, \eta) \vdash_{16} 1 - \tfrac{1}{n^2}\langle x, x' \rangle^2 \leqslant 2\lambda + 16\eta\,. \tag{2.20}$$

Indeed, since we impose both G and G' to be η-close to $G_{\text{corrupted}}$, the graphs G and G' have to be 2η-close and we can derive $\mathcal{A}_{\text{close}}(G, G', z; 2\eta)$ for an appropriately chosen z by a low-degree sum-of-squares proof. By the Lipschitz property of $\mathcal{A}_{\text{basic}}$, we get that x and x' satisfy $\mathcal{A}_{\text{basic}}$ for the same graph $H := \mathrm{D}(z) \cdot G \cdot \mathrm{D}(z)$. Concretely, since the robust constraints already give $\mathcal{A}_{\text{basic}}(G, x; \lambda + 4\eta)$ and $\mathcal{A}_{\text{basic}}(G', x'; \lambda + 4\eta)$, we can derive $\mathcal{A}_{\text{basic}}(H, x; \lambda+4\eta+2\cdot 2\eta)$ and $\mathcal{A}_{\text{basic}}(H, x'; \lambda+4\eta+2\cdot 2\eta)$ by Eq. (2.17). Applying Eq. (2.11) with parameter $\lambda+8\eta$, we conclude that x and x' are close by Eq. (2.11). As before, Eq. (2.20) means that the sum-of-squares meta algorithm recovers the underlying blocks with error $O(\varepsilon\sqrt{d})^{-1} + O(\eta)$ for the stochastic block model with degree parameter d and bias ε in the presence of an η fraction of adversarial node corruptions, recovering one of the results of Liu and Moitra (2022).

The algorithm obtained in this way is essentially the same as in Liu and Moitra (2022) with only superficial differences. However, we believe the analyses differ substantially. While the analysis in Liu and Moitra (2022) requires significant ingenuity tailored to the particular algorithm, our analysis in contrast is purely mechanical.

Boosting accuracy robustly. While the techniques discussed so far capture up to constant factors the information-theoretically necessary parameters to achieve small error, say error at most 0.01, they do not capture the optimal error rate.

In the non-robust setting, we can boost the accuracy of a solution with small error by several rounds of majority voting, where we assign each node the most common label among its neighbours. This strategy however fails dramatically in the presence of node corruptions. Concretely, when $\eta \gtrsim \varepsilon$, even without changing the degrees of nodes, an adversary can fool majority voting by corrupting only an $O(\eta)$ fraction of nodes (Liu and Moitra, 2022).

For simplicity, suppose that we start with the ground-truth label assignment $\boldsymbol{x}$ and an adversary can select random vertex subsets S and T with $|S| = O(\eta n)$ and $|T| = n/2$. Then, the adversary can use S to poison the votes for T by setting all edges from S to random nodes in T with the opposite label. This corruption flips the majority vote outcome for most of the nodes in T with high probability. Hence, majority voting applied to $\boldsymbol{x}$ results in an assignment that agrees with $\boldsymbol{x}$ in a random subset of roughly half of the nodes and thus has large error.

In this example, a tiny fraction of nodes is responsible for roughly half of the majority vote outcomes. However, in a random-like graph, the expander mixing lemma certifies that small fractions of nodes cannot change many majority vote outcomes. Concretely, in the previous example, the expander mixing lemma implies for the uncorrupted graph that the number of edges between S and T is $\frac{d}{n}|S| \cdot |T| \pm O(d \cdot |S| \cdot |T|)^{1/2} = \frac{1}{2}\eta dn \cdot (1 \pm O((\eta d)^{-1/2}))$. On the other hand, in the corrupted graph we have about ηdn edges between these sets, which is about two times larger than before the corruption as $\eta d \gtrsim \varepsilon d \geqslant \varepsilon^2 d \gg 1$ when $\eta \gtrsim \varepsilon$ and the blocks are recoverable.

More formally, we can limit the effect of small sets on majority vote outcomes if we have a label assignment x consistent with the majority votes for most of the nodes such that the matrix $E = \frac{n}{\varepsilon d} G - xx^{\mathsf{T}}$ satisfies the following pseudorandomness constraints, strengthening $\mathcal{A}_{\text{rand}}(E, R; \lambda)$,

$$\mathcal{A}_{\text{mix}}(E, R; \lambda) := \left\{ R\colon \mathop{\vert\!\frac{u,v}{2}} \langle u, Ev \rangle \leqslant \lambda n \cdot \tfrac{1}{2}(\|u\|^2 + \|v\|^2) \right\}. \tag{2.21}$$

Since the previous pseudorandomness constraints $\mathcal{A}_{\text{rand}}$ defined in Eq. (2.9) consider vectors u, v with $\|u\| = \|v\| = \sqrt{n}$, the above normalization for λ is the same as in Eq. (2.9). We remark that in our applications E is symmetric. Hence, due to the characterization of the above inequality in terms of matrix factorizations, an equivalent formulation of the above system would be

$$\left\{ R = (R_+, R_-)\colon R_+ {R_+}^{\mathsf{T}} = \lambda n \cdot I_n - E, \quad R_- {R_-}^{\mathsf{T}} = \lambda n \cdot I_n + E \right\}.$$

We use the following polynomial constraints to encode that x is consistent with the majority votes for most of the nodes,

$$\mathcal{A}_{\text{maj}}(G, x, R; \beta, \gamma) := \left\{ R\colon \mathcal{A}_{\text{set}}(z) \mathop{\vert\!\frac{z}{2}} \langle \mathrm{D}(z)x, Gx \rangle \geqslant (1-\gamma)\varepsilon d(|z| - \beta n) \right\}, \tag{2.22}$$

where $|z| = z_1 + \cdots + z_n$ and $\mathcal{A}_{\text{set}}(z) := \left\{ \mathrm{D}(z)^2 = \mathrm{D}(z) \right\}$. Note that for the stochastic block model $(\boldsymbol{G}, \boldsymbol{x}) \sim \mathsf{SBM}_n(d, \varepsilon)$, we expect that $(Gx)_i \approx \varepsilon d \cdot x_i$ for most nodes $i \in [n]$. Since the inequality in Eq. (2.22) is linear subject to simple constraints, sum-of-squares proofs capture all valid inequalities of this type up to a small error. If the inequality in Eq. (2.22) is valid, then we can have at most $\beta n/\gamma$ nodes i with $(Gx)_i \cdot x_i \leqslant (1-2\gamma)\varepsilon d$. At the same time, if the total deficit $\sum_i \left((1-\gamma)\varepsilon d - (Gx)_i \cdot x_i \right)_+$ is at most $(1-\gamma)\varepsilon d \cdot \beta n$, then the above inequality is valid.[4]

Liu and Moitra (2022) show that the stochastic block model satisfies the above constraints for $\lambda \lesssim (\varepsilon \sqrt{d})^{-1}$ and $\ln 1/\beta \approx_{o(1)} \gamma C_{d,\varepsilon}/2$ with high probability (after removing a negligible number of nodes). They also provide an iterative algorithm that boosts the accuracy of a solution to a rate that is asymptotically optimal.

Simplifying their approach significantly, we show that the sum-of-squares meta-algorithm can boost to such a rate in one shot. To this end, we combine the previous constraints as follows,

$$\mathcal{A}_{\text{boost}}(G, x) := \mathcal{A}_{\text{label}}(x) \cup \mathcal{A}_{\text{mix}}(\tfrac{n}{\varepsilon d} G - xx^{\mathsf{T}}; \lambda) \cup \mathcal{A}_{\text{maj}}(G, x; \beta_1, \gamma_1) \cup \mathcal{A}_{\text{maj}}(G, x; \beta_2, \gamma_2), \tag{2.23}$$

where we choose $\lambda \lesssim (\varepsilon \sqrt{d})^{-1}$, $\gamma_1 = 0.01$, $\beta_1 = \lambda$, $\gamma_2 = 1 - 10\lambda$, and $\beta_2 = \exp(-C_{d,\varepsilon}/2 + \chi \sqrt{C}_{d,\varepsilon})$ for some sufficiently large absolute constant χ. The following property of these constraints allows us to boost the accuracy of a solution,

$$\tfrac{1}{n}\|x - x'\|^2 \leqslant 0.01, \mathcal{A}_{\text{boost}}(G, x), \mathcal{A}_{\text{boost}}(G, x') \mathop{\vert\!\frac{G,x,x'}{}} \tfrac{1}{n}\|x - x'\|^2 \leqslant \exp\Big(-C_{d,\varepsilon}/2 + O(\sqrt{C}_{d,\varepsilon}) \Big). \tag{2.24}$$

For similar reasons as before, we can use Lipschitz and restriction properties of these constraints in order to obtain error bounds in the robust setting at the cost of an additional $O(\eta)$ error for an $O(\eta)$ fraction of node corruptions.

4. Note that for random-like graphs, the expander mixing lemma implies a bound on this total deficit once we have a bound on the number of nodes i with $(Gx)_i \cdot x_i \leqslant (1-\gamma)\varepsilon d$.

Toward establishing Eq. (2.24), we consider the vector z with $z_i = \frac{1}{4}\big(x_i - x'_i\big)^2$. Since x and x' have coordinates in $\{\pm 1\}$, the vector z has coordinates in $\{0,1\}$. Indeed, $\mathcal{A}_{\text{label}}(x), \mathcal{A}_{\text{label}}(x') \vdash \mathcal{A}_{\text{set}}(z)$. Furthermore, the vector z also satisfies $\mathrm{D}(z)x = -\mathrm{D}(z)x'$ and $x - x' = \mathrm{D}(z)\cdot(x-x')$ because z indicates the set of coordinates, where x and x' differ.

Now since the labelings x and x' agree in at least a 0.99 fraction of the nodes, the majority voting based on x or x' will agree in all but a β fraction of the nodes (when there is no node corruption). This gives us the intuition that the constraints from expander mixing lemmas (namely $\mathcal{A}_{\text{maj}}(G,x;\beta,\gamma)$ and $\mathcal{A}_{\text{maj}}(G,x';\beta,\gamma)$) ensure that x and x' are close to each other in distance β.

Concretely, using these relations between z, x, and x' together with majority-voting consistency constraints $\mathcal{A}_{\text{maj}}(G,x;\beta,\gamma)$ and $\mathcal{A}_{\text{maj}}(G,x';\beta,\gamma)$, we get

$$\begin{aligned}\mathcal{A}_{\text{maj}}(G,x;\beta,\gamma), \mathcal{A}_{\text{maj}}(G,x';\beta,\gamma) \,\Big|\!\frac{\;G,x,x'\;}{}\quad & (1-\gamma)\varepsilon d(|z| - \beta n) \\ & \leqslant \tfrac{1}{2}\Big(\langle \mathrm{D}(z)x, Gx\rangle + \langle \mathrm{D}(z)x', Gx'\rangle\Big) \\ & = \tfrac{1}{2}\langle \mathrm{D}(z)x, G(x-x')\rangle \\ & = \langle \mathrm{D}(z)x, G\,\mathrm{D}(z)x\rangle\,.\end{aligned}$$

We can upper bound the quantity on the right using the constraints $\mathcal{A}_{\text{mix}}(E;\lambda)$ for the matrix $E = \frac{n}{\varepsilon d}G - xx^{\mathsf{T}}$ as well as the label-assignment and set constraints for x and z,

$$\begin{aligned}\mathcal{A}_{\text{mix}}(E;\lambda), \mathcal{A}_{\text{set}}(z), \mathcal{A}_{\text{label}}(x) \,\Big|\!\frac{\;E,z,x\;}{}\quad & \langle \mathrm{D}(z)x, G\,\mathrm{D}(z)x\rangle \\ & = \tfrac{\varepsilon d}{n}\big(\langle x, \mathrm{D}(z)x\rangle^2 + \langle \mathrm{D}(z)x, E\,\mathrm{D}(z)x\rangle\big) \\ & \leqslant \tfrac{\varepsilon d}{n}\big(\|\mathrm{D}(z)x\|^4 + \lambda n\cdot\|\mathrm{D}(z)x\|^2\big) \\ & = \tfrac{\varepsilon d}{n}\big(|z|^2 + \lambda n\cdot|z|\big)\end{aligned}$$

Combining the lower and upper bounds on $\langle \mathrm{D}(z)x, G\,\mathrm{D}(z)x\rangle$, we get the following upper bound on $|z|$,

$$\mathcal{A}_{\text{maj}}(G,x;\beta,\gamma), \mathcal{A}_{\text{maj}}(G,x';\beta,\gamma), \mathcal{A}_{\text{mix}}(E;\lambda), \mathcal{A}_{\text{set}}(z), \mathcal{A}_{\text{label}}(x), \mathcal{A}_{\text{label}}(x') \tag{2.25}$$

$$\Big|\!\frac{\;G,x,x',E,z,x\;}{}\; |z| \leqslant \tfrac{1-\gamma}{1-\gamma-\lambda}\beta n + \tfrac{1}{(1-\gamma-\lambda)n}|z|^2\,. \tag{2.26}$$

While this bound alone is not enough to conclude an absolute bound on $|z|$, it does allow us to boost assumed bounds on $|z|$. Suppose $\gamma+\lambda \leqslant 1$ and $\lambda \leqslant \frac{1}{4}(1-\gamma)$. Then, $\frac{1}{1-\gamma-\lambda} \leqslant \frac{4}{3(1-\gamma)}$. Moreover, if we add the constraint $|z| \leqslant \frac{1-\gamma}{4}n$ to the previous constraints, we can derive $\frac{1}{(1-\gamma-\lambda)n}|z|^2 \leqslant \frac{1}{3}|z|$ and thus, by subtracting $\frac{1}{3}|z|$ from both sides of the inequality Eq. (2.25),

$$|z| \leqslant \tfrac{1-\gamma}{4}n, |z| \leqslant \tfrac{1-\gamma}{1-\gamma-\lambda}\beta n + \tfrac{1}{(1-\gamma-\lambda)n}|z|^2 \,\Big|\!\frac{\;z\;}{}\; |z| \leqslant 2\beta n\,. \tag{2.27}$$

Our target derivation Eq. (2.24) contains the inequality $4|z| = \|x-x'\|^2 \leqslant 0.01n$ as an initial assumption. Hence, we get to assume $|z| \leqslant \frac{1-\gamma_1}{4}n$ from the get-go as $\gamma_1 = 0.01$. By substituting γ_1 and β_1 for γ and β in Eq. (2.27), we derive $|z| \leqslant 2\beta_1 n$. Since $\gamma_2 = 1-2\lambda-8\beta_1$, we can compose this derivation with Eq. (2.27) with γ_2 and β_2 substituted for β and γ. In this way, we derive the desired bound $|z| \leqslant 2\beta_2 n$ for $\ln\beta_2 = -C_{d,\varepsilon}/2 + O(\sqrt{C}_{d,\varepsilon})$ on the distance between the label assignments x and x'.

Concluding the two-community case. Using the polynomial constraints and sum-of-squares proofs developed above, we obtain a polynomial-time algorithm for robust recovery in two-community stochastic block models with asymptotically optimal error rates. Concretely, supposing $d \geqslant \Omega(1/\varepsilon^2)$ and η sufficiently small, we use the constraints $\mathcal{A}_{\text{robust}}$ in Eq. (2.19) to obtain, with high probability, a label assignment $\tilde{x} \in \{\pm 1\}^n$ with error at most $0.005n$. Since the true bipartition is invariant under a global sign flip, we fix the sign of x so that $\|\tilde{x} - x\|^2 \leqslant 0.005n$. We then solve the sum-of-squares program for the following system of constraints:

$$\begin{aligned}\mathcal{A}_{\text{robust}-\text{boost}}(G, x; \eta) := \mathcal{A}_{\text{close}}(G, G_{\text{corrupted}}; \eta) \cup \left\{ \|x - \tilde{x}\|^2 \leqslant 0.005n \right\} \\ \cup\, \mathcal{A}_{\text{boost}}(G, x)\,.\end{aligned} \tag{2.28}$$

Here, $G_{\text{corrupted}}$ denotes the η-node-corrupted input graph, and $\tilde{x}$ is the rough label assignment obtained in the preprocessing phase, with the global sign chosen so that $\frac{1}{n}\|\tilde{x} - \boldsymbol{x}\|^2 \leqslant 0.005$. As discussed above, after pruning a tiny fraction of high-degree nodes, the stochastic block model $(\boldsymbol{G}, \boldsymbol{x}) \sim \mathsf{SBM}_n(d, \varepsilon)$ satisfies the above constraints with high probability, uniformly over all η-corruptions of $\boldsymbol{G}$, when we substitute $\boldsymbol{G}$ and $\boldsymbol{x}$ for G and x. By Eq. (2.24), any two solutions x, x' for the same graph G are $\exp(-C_{d,\varepsilon}/2 + O(\sqrt{C_{d,\varepsilon}}))$-close. The required hypothesis $\frac{1}{n}\|x - x'\|^2 \leqslant 0.01$ follows from the constraints $\frac{1}{n}\|x - \tilde{x}\|^2, \frac{1}{n}\|x' - \tilde{x}\|^2 \leqslant 0.005$, since $\frac{1}{4}\|a - b\|^2$ is Hamming distance on $\{\pm 1\}^n$ and Hamming distance satisfies the triangle inequality. Using restriction arguments analogous to Eq. (2.17), the same bound extends to η-close graphs G and G' at the cost of an additional $O(\eta)$ error.

We emphasize that this algorithm relies on the same structural properties of the stochastic block model as the algorithm by Liu and Moitra (2022). Concretely, both algorithms try to rule out a small fraction of nodes swaying a disproportionate number of majority vote outcomes. However, the final algorithms end up being quite different. While our algorithm and its analysis exploit the structural properties directly and don't require additional concepts besides the generic SOS framework, Liu and Moitra introduce several non-trivial technical tools such as the notion of resolvable matrices. Furthermore, they need to keep track of invariants satisfied throughout the iterations of their algorithm. Our algorithm avoids this complication by boosting to "full accuracy" in one shot.

2.2. More than two communities

Extending our techniques to more than two communities requires overcoming several new challenges. First, the initialization step must be adapted to handle $k > 2$ communities. Second, pairwise majority voting requires a symmetry-breaking initialization with small per-community error, whereas basic recovery only gives a global error guarantee. We address these issues below.

Basic recovery for more than two communities. The basic recovery step from the two-community case extends in a straightforward way by replacing $\mathbb{E}[\boldsymbol{G} \mid \boldsymbol{x}] = \frac{\varepsilon d}{n}\boldsymbol{x}\boldsymbol{x}^{\mathsf{T}}$ by its natural analogue for k communities $\mathbb{E}[\boldsymbol{G} \mid \boldsymbol{Z}] = \frac{\varepsilon d}{n}\sum_{a=1}^{k}(\boldsymbol{Z}_a)(\boldsymbol{Z}_a)^{\mathsf{T}} - \frac{\varepsilon d}{kn}\mathbf{1}\mathbf{1}^{\mathsf{T}}$, where $\boldsymbol{Z}_a \in \{0,1\}^n$ is the indicator vector for community $a \in [k]$.

Boosting basic recovery accuracy via robust bisectioning. The basic recovery step yields a labeling with misclassification error at most 0.001 whenever $\varepsilon^2 d \gg k^2$. Our goal is to refine this rough clustering to the optimal error rate $\exp\bigl(-(1-o(1))C_{d,\varepsilon}/k\bigr)$ using pairwise majority voting, as in the two-community case. However, pairwise majority voting requires that each community contain a clear majority of correctly labeled nodes, and an overall 0.001 error bound does not preclude the errors from concentrating on a small number of communities.

Our key idea for the initialization procedure is to reduce the k-community detection problem to a sequence of two-community problems, which can be solved using the two-community majority-vote procedure. More precisely, we design a *robust bisectioning algorithm* that, given a graph with k communities and a basic recovery with error at most 0.001, outputs a bisection that approximates the union of $k/2$ ground-truth communities versus its complement, with bisection misclassification error $\exp(-\varepsilon^2 d/k^2) + \operatorname{poly}(k)\cdot\eta$. The basic framework of the robust bisectioning algorithm is the following:

- Given a basic recovery with error at most 0.001, identify $k/2$ clusters such that at least a 0.99 fraction of the nodes in each cluster come from the same ground-truth community. This is done by a *robust verification algorithm* that, given a candidate cluster of size n/k, can certify whether at least a 0.99 fraction of its nodes belong to a single ground-truth community.
- Place the $k/2$ verified clusters on one side of the bisection and the remaining nodes on the other, yielding an initial bisection with constant error (e.g. 0.01). Then boost this bisection error to $\exp(-\varepsilon^2 d/k^2) + \operatorname{poly}(k)\eta$ using pairwise majority voting between the two sides.

By recursively applying the robust bisectioning algorithm, we obtain a rough k-clustering with misclassification error $1/\operatorname{poly}(k) + \operatorname{poly}(k)\cdot\eta$ when $\varepsilon^2 d \geqslant Kk^2\log k$ for a sufficiently large universal constant $K > 0$.

Achieving the optimal error rate. Finally, we boost the rough k-clustering with misclassification error $1/\operatorname{poly}(k)+\operatorname{poly}(k)\cdot\eta$ to the optimal error rate $\exp\Bigl(-(1-o(1))\frac{C_{d,\varepsilon}}{k}\Bigr)+\operatorname{poly}(k)\cdot\eta$. This is done using a sum-of-squares program similar to the two-community case, but with new robust mixing and pairwise majority-vote constraints tailored to the k-community setting. These constraints prevent a small corrupted set of nodes from swaying a disproportionate number of vote outcomes, and imply that any feasible solution of the resulting sum-of-squares program must agree with the ground truth on all but an $\exp(-(1-o(1))C_{d,\varepsilon}/k)$ fraction of nodes.[5] To obtain the final clustering, we round the sum-of-squares solution using the standard node-distance Lipschitzness argument, incurring an additional $\operatorname{poly}(k)\cdot\eta$ loss and obtaining the desired error rate.

Acknowledgments

This work is supported by funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 815464).

5. Technically, achieving the optimal error requires two rounds of majority voting with different parameters.

References

Emmanuel Abbe. Community detection and stochastic block models: Recent developments. *Journal of Machine Learning Research*, 18(177):1–86, 2018. URL http://jmlr.org/papers/v18/16-480.html.

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

Emmanuel Abbe and Colin Sandon. Achieving the ks threshold in the general stochastic block model with linearized acyclic belief propagation. In D. Lee, M. Sugiyama, U. Luxburg, I. Guyon, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 29. Curran Associates, Inc., 2016. URL https://proceedings.neurips.cc/paper_files/paper/2016/file/6c29793a140a811d0c45ce03c1c93a28-Paper.pdf.

Emmanuel Abbe, Afonso S Bandeira, and Georgina Hall. Exact recovery in the stochastic block model. *IEEE Transactions on information theory*, 62(1):471–487, 2015.

Jayadev Acharya, Ayush Jain, Gautam Kamath, Ananda Theertha Suresh, and Huanyu Zhang. Robust estimation for random graphs. In *Conference on Learning Theory*, pages 130–166. PMLR, 2022.

Afonso S Bandeira, Jess Banks, Dmitriy Kunisky, Christopher Moore, and Alex Wein. Spectral planting and the hardness of refuting cuts, colorability, and communities in random graphs. In *Conference on Learning Theory*, pages 410–473. PMLR, 2021.

Boaz Barak and David Steurer. Sum-of-squares proofs and the quest toward optimal algorithms. *arXiv preprint arXiv:1404.5236*, 2014.

Boaz Barak and David Steurer. Proofs, beliefs, and algorithms through the lens of sum-of-squares. *Course notes: http://www. sumofsquares. org/public/index. html*, 1:3, 2016.

Boaz Barak, Fernando GSL Brandao, Aram W Harrow, Jonathan Kelner, David Steurer, and Yuan Zhou. Hypercontractivity, sum-of-squares proofs, and their applications. In *Proceedings of the forty-fourth annual ACM symposium on Theory of computing*, pages 307–326, 2012.

Ravi B Boppana. Eigenvalues and graph bisection: An average-case analysis. In *28th Annual Symposium on Foundations of Computer Science (sfcs 1987)*, pages 280–285. IEEE, 1987.

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

Hongjie Chen, Jingqiu Ding, Tommaso d'Orsi, Yiding Hua, Chih-Hung Liu, and David Steurer. Private graphon estimation via sum-of-squares. In *Proceedings of the 56th Annual ACM Symposium on Theory of Computing*, pages 172–182, 2024a.

Hongjie Chen, Jingqiu Ding, Yiding Hua, and David Steurer. Private edge density estimation for random graphs: Optimal, efficient and robust. *Advances in Neural Information Processing Systems*, 37:90771–90817, 2024b.

Hongjie Chen, Jingqiu Ding, Yiding Hua, and Stefan Tiegel. Improved robust estimation for erd\h {o} sr\'enyi graphs: The sparse regime and optimal breakdown point. *arXiv preprint arXiv:2503.03923*, 2025.

Byron Chin, Elchanan Mossel, Youngtak Sohn, and Alexander S Wein. Stochastic block models with many communities and the kesten–stigum bound. *arXiv preprint arXiv:2503.03047*, 2025.

Peter Chin, Anup Rao, and Van Vu. Stochastic block model and community detection in sparse graphs: A spectral algorithm with optimal rate of recovery. In *Conference on Learning Theory*, pages 391–423. PMLR, 2015.

Amin Coja-Oghlan. Graph partitioning via adaptive spectral techniques. *Combinatorics, Probability and Computing*, 19(2):227–284, 2010.

Aurelien Decelle, Florent Krzakala, Cristopher Moore, and Lenka Zdeborová. Asymptotic analysis of the stochastic block model for modular networks and its algorithmic applications. *Physical Review E*, 84(6), Dec 2011. ISSN 1550-2376. doi: 10.1103/physreve.84.066106. URL http://dx.doi.org/10.1103/physreve.84.066106.

Jingqiu Ding, Tommaso d'Orsi, Rajai Nasser, and David Steurer. Robust recovery for stochastic block models. In *2021 IEEE 62nd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 387–394. IEEE, 2022.

Jingqiu Ding, Tommaso d'Orsi, Yiding Hua, and David Steurer. Reaching kesten-stigum threshold in the stochastic block model under node corruptions. In Gergely Neu and Lorenzo Rosasco, editors, *Proceedings of Thirty Sixth Conference on Learning Theory*, volume 195 of *Proceedings of Machine Learning Research*, pages 4044–4071. PMLR, 12–15 Jul 2023. URL https://proceedings.mlr.press/v195/hua23a.html.

Jingqiu Ding, Yiding Hua, Lucas Slot, and David Steurer. Low degree conjecture implies sharp computational thresholds in stochastic block model. *arXiv preprint arXiv:2502.15024*, 2025.

Yingjie Fei and Yudong Chen. Achieving the bayes error rate in stochastic block model by sdp, robustly. In Alina Beygelzimer and Daniel Hsu, editors, *Proceedings of the Thirty-Second Conference on Learning Theory*, volume 99 of *Proceedings of Machine Learning Research*, pages 1235–1269. PMLR, 25–28 Jun 2019. URL https://proceedings.mlr.press/v99/fei19a.html.

Uriel Feige and Eran Ofek. Spectral techniques applied to sparse random graphs. *Random Structures & Algorithms*, 27(2):251–275, September 2005. ISSN 1042-9832. doi: 10.1002/rsa.20089.

Chao Gao, Zongming Ma, Anderson Y. Zhang, and Harrison H. Zhou. Achieving optimal misclassification proportion in stochastic block models. *Journal of Machine Learning Research*, 18(60):1–45, 2017. URL http://jmlr.org/papers/v18/16-245.html.

Fabrizio Grandoni, Rafail Ostrovsky, Yuval Rabani, Leonard J. Schulman, and Rakesh Venkat. A refined approximation for euclidean k-means, 2021. URL https://arxiv.org/abs/2107.07358.

Olivier Guédon and Roman Vershynin. Community detection in sparse networks via grothendieck's inequality. *Probability Theory and Related Fields*, 165(3):1025–1049, 2016.

Bruce E. Hajek, Yihong Wu, and Jiaming Xu. Achieving exact cluster recovery threshold via semidefinite programming. *IEEE Transactions on Information Theory*, 62:2788–2797, 2014. URL https://api.semanticscholar.org/CorpusID:917315.

Paul W Holland, Kathryn Blackmond Laskey, and Samuel Leinhardt. Stochastic blockmodels: First steps. *Social networks*, 5(2):109–137, 1983.

Samuel B Hopkins. Mean estimation with sub-gaussian rates in polynomial time. *The Annals of Statistics*, 48(2):1193–1213, 2020.

Samuel B. Hopkins and Jerry Li. Mixture models, robustness, and sum of squares proofs. In *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2018, page 1021–1034, New York, NY, USA, 2018. Association for Computing Machinery. ISBN 9781450355599. doi: 10.1145/3188745.3188748. URL https://doi.org/10.1145/3188745.3188748.

Samuel B. Hopkins and David Steurer. Efficient bayesian estimation from few samples: Community detection and related problems. In *2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 379–390, 2017. doi: 10.1109/FOCS.2017.42.

Samuel.B Hopkins. *Statistical inference and the sum of squares method*. Cornell University, 2018.

Pravesh K. Kothari, Jacob Steinhardt, and David Steurer. Robust moment estimation and improved clustering via sum of squares. Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing, New York, NY, USA, 2018. Association for Computing Machinery.

Allen Liu and Ankur Moitra. Minimax rates for robust community detection. In *2022 IEEE 63rd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 823–831, 2022. doi: 10.1109/FOCS54457.2022.00083.

Laurent Massoulié. Community detection thresholds and the weak ramanujan property. In *Proceedings of the forty-sixth annual ACM symposium on Theory of computing*, pages 694–703, 2014.

Sidhanth Mohanty, Prasad Raghavendra, and David X Wu. Robust recovery for stochastic block models, simplified and generalized. In *Proceedings of the 56th Annual ACM Symposium on Theory of Computing*, pages 367–374, 2024.

Elchanan Mossel, Joe Neeman, and Allan Sly. Belief propagation, robust reconstruction and optimal recovery of block models. In Maria Florina Balcan, Vitaly Feldman, and Csaba Szepesvári, editors, *Proceedings of The 27th Conference on Learning Theory*, volume 35 of *Proceedings of Machine Learning Research*, pages 356–370, Barcelona, Spain, 13–15 Jun 2014. PMLR. URL https://proceedings.mlr.press/v35/mossel14.html.

Elchanan Mossel, Joe Neeman, and Allan Sly. A proof of the block model threshold conjecture. *Combinatorica*, 38(3):665–708, 2018.

Aaron Potechin and David Steurer. Exact tensor completion with sum-of-squares. In *Proceedings of the 2017 Conference on Learning Theory*, 2017.

Prasad Raghavendra and Benjamin Weitz. On the bit complexity of sum-of-squares proofs. *arXiv preprint arXiv:1702.05139*, 2017.

Prasad Raghavendra, Tselil Schramm, and David Steurer. High dimensional estimation via sum-of-squares proofs. In *Proceedings of the International Congress of Mathematicians: Rio de Janeiro 2018*, pages 3389–3423. World Scientific, 2018.

Youngtak Sohn and Alexander S Wein. Sharp phase transitions in estimation with low-degree polynomials. In *Proceedings of the 57th Annual ACM Symposium on Theory of Computing*, pages 891–902, 2025.

Anderson Y. Zhang and Harrison H. Zhou. Minimax rates of community detection in stochastic block models, 2015. URL https://arxiv.org/abs/1507.05313.

Appendix A. Organization

In Appendix B, we cover notations and basics of the Sum-of-Squares method. In Appendix C, we establish the statistical properties of the k-SBM model with a focus on the majority voting concentration result. In Appendix D, we outline our algorithmic framework and present our formal main results. The proofs of these results are subsequently detailed in Appendix E (initial rough clustering), Appendix F (identifying clusters with a 0.99-fraction of the vertices recovered), and Appendix G (robustly boosting the rough bisection accuracy to the optimal bisection error rate), respectively. These ingredients are then combined in Appendix H to obtain our robust bisectioning algorithm. In Appendix I, we show how to recursively apply the robust bisectioning algorithm to robustly find a k-clustering of the graph with error rate $1/\operatorname{poly}(k)$ for symmetry breaking. Finally, we conclude in Appendix J by boosting the error rate from $1/\operatorname{poly}(k)$ to the optimal error rate using robust pairwise majority voting. In Appendix K, we provide additional details on the Sum-of-Squares method. In Appendix L, we collect several spectral norm bounds used in the analysis. In Appendix M, we provide the proofs of several statistical properties of the k-SBM model stated in Appendix C.

Appendix B. Preliminary

B.1. Notation

Random variables. We use boldface to denote random variables, e.g., $\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z}$.

Asymptotics. We write $f \lesssim g$ to denote the inequality $f \leqslant C \cdot g$ for some absolute constant $C > 0$. We write $O(f)$ and $\Omega(f)$ to denote quantities f_- and f_+ satisfying $f_- \lesssim f$ and $f \lesssim f_+$, respectively.

Vector. For any vector, we use the notation $\|\cdot\|_1$ for the ℓ_1-norm, $\|\cdot\|_2$ for the ℓ_2-norm, and $\|\cdot\|_\infty$ for the ℓ_∞-norm. Furthermore, for any two vectors $x, y \in \mathbb{R}^n$, we use $x \odot y$ for the Hadamard (entrywise) product of the two vectors. We write $\mathbf{1}$ and $\mathbf{0}$ for the all-ones and all-zeros vectors (or matrices) of the appropriate dimension. For a subset $S \subseteq [n]$, we denote its indicator vector by $\mathbf{1}_S \in \{0,1\}^n$, where $\mathbf{1}_S(i) = 1$ if $i \in S$ and $\mathbf{1}_S(i) = 0$ otherwise.

Matrix. For a matrix $M \in \mathbb{R}^{n\times m}$, we denote its (i,j)-th entry by $M(i,j)$, its i-th row by $M(i,\cdot)$, and its j-th column by $M(\cdot,j)$. For two matrices of the same dimension, we also use $M \odot N$ for their Hadamard (entrywise) product. We use $\|M\|$ (and also $\|M\|_{\text{op}}$) for the spectral norm of M and $\|M\|_{\text{F}}$ for the Frobenius norm of M. We denote by $\|M\|_{\text{sum}}$ and $\|M\|_{\max}$ the sum and the maximum of the absolute values of the entries in M, respectively. For two matrices $M, N \in \mathbb{R}^{n\times m}$, we denote their inner product by $\langle M, N\rangle = \operatorname{Tr} MN^{\mathsf{T}} = \sum_{i,j} M(i,j)N(i,j)$. We write I_n for the $n \times n$ identity matrix, and $J = \mathbf{1}\mathbf{1}^\top$ for the all-ones matrix.

Graph. We write $G \in \{0,1\}^{n\times n}$ for the adjacency matrix of a (possibly corrupted) graph on vertex set $[n]$. When an average degree parameter d is specified, we define the centered adjacency matrix

$$\bar{G} \coloneqq G - \frac{d}{n} J.$$

When we need to distinguish the uncorrupted graph sampled from the stochastic block model from a corrupted observation, we write G° for the uncorrupted adjacency matrix and $\bar{G}^\circ \coloneqq G^\circ - \frac{d}{n} J$ for its centered version.

B.2. Sum-of-Squares hierarchy

In this paper, we employ the sum-of-squares hierarchy Barak and Steurer (2014, 2016); Raghavendra et al. (2018) for both algorithm design and analysis. As a broad category of semidefinite programming algorithms, sum-of-squares algorithms provide many optimal or state-of-the-art results in algorithmic statistics Hopkins and Li (2018); Kothari et al. (2018); Potechin and Steurer (2017); Hopkins (2020). We provide here a brief introduction to pseudo-distributions, sum-of-squares proofs, and sum-of-squares algorithms.

Pseudo-distribution. We can represent a finitely supported probability distribution over $\mathbb{R}^n$ by its probability mass function $\mu\colon \mathbb{R}^n \to \mathbb{R}$ such that $\mu \geqslant 0$ and $\sum_{x\in\text{supp}(\mu)} \mu(x) = 1$. We define pseudo-distributions as generalizations of such probability mass distributions, by relaxing the constraint $\mu \geqslant 0$ and only requiring that μ passes certain low-degree non-negativity tests.

Definition 4 (Pseudo-distribution) *A* level-ℓ pseudo-distribution μ *over* $\mathbb{R}^n$ *is a finitely supported function* $\mu : \mathbb{R}^n \to \mathbb{R}$ *such that* $\sum_{x \in \mathrm{supp}(\mu)} \mu(x) = 1$ *and* $\sum_{x \in \mathrm{supp}(\mu)} \mu(x) f(x)^2 \geqslant 0$ *for every polynomial* f *of degree at most* $\ell/2$*.*

We can define the formal expectation of a pseudo-distribution in the same way as the expectation of a finitely supported probability distribution.

Definition 5 (Pseudo-expectation) *Given a pseudo-distribution* μ *over* $\mathbb{R}^n$*, we define the* pseudo-expectation *of a function* $f : \mathbb{R}^n \to \mathbb{R}$ *by*

$$\tilde{\mathbb{E}}_\mu f := \sum_{x \in \mathrm{supp}(\mu)} \mu(x) f(x) \,. \tag{B.1}$$

In later sections, when the underlying pseudo-distribution μ is implicit, we write $\tilde{\mathbb{E}}$ for its pseudo-expectation operator and use the shorthand $\tilde{\mathbb{E}}[p]$ for $\tilde{\mathbb{E}}_\mu p$.

The following definition formalizes what it means for a pseudo-distribution to satisfy a system of polynomial constraints.

Definition 6 (Constrained pseudo-distributions) *Let* $\mu : \mathbb{R}^n \to \mathbb{R}$ *be a level-*ℓ *pseudo-distribution over* $\mathbb{R}^n$*. Let* $\mathcal{A} = \{f_1 \geqslant 0, \ldots, f_m \geqslant 0\}$ *be a system of polynomial constraints. We say that* μ satisfies $\mathcal{A}$ *at level* r*, denoted by* $\mu \models_r \mathcal{A}$*, if for every multiset* $S \subseteq [m]$ *and every sum-of-squares polynomial* h *such that* $\deg(h) + \sum_{i \in S} \max\{\deg(f_i), r\} \leqslant \ell$*,*

$$\tilde{\mathbb{E}}_\mu h \cdot \prod_{i \in S} f_i \geqslant 0 \,. \tag{B.2}$$

We say μ *satisfies* $\mathcal{A}$ *and write* $\mu \models \mathcal{A}$ *(without further specifying the degree) if* $\mu \models_0 \mathcal{A}$*.*

We remark that if μ is an actual finitely supported probability distribution, then we have $\mu \models \mathcal{A}$ if and only if μ is supported on solutions to $\mathcal{A}$.

Sum-of-squares algorithm. The *sum-of-squares algorithm* searches through the space of pseudo-distributions that satisfy a given system of polynomial constraints, by solving semidefinite programming.

Theorem 7 (Sum-of-squares algorithm) *There exists an* $(n+m)^{O(\ell)}$*-time algorithm that, given any explicitly bounded*[6] *and satisfiable system*[7] $\mathcal{A}$ *of* m *polynomial constraints in* n *variables, outputs a level-*ℓ *pseudo-distribution that satisfies* $\mathcal{A}$ *approximately.*

Sum-of-squares proof. We introduce sum-of-squares proofs as the dual objects of pseudo-distributions, which can be used to reason about properties of pseudo-distributions. We say a polynomial p is a sum-of-squares polynomial if there exist polynomials (q_i) such that $p = \sum_i q_i^2$.

6. A system of polynomial constraints is *explicitly bounded* if it contains a constraint of the form $\|x\|^2 \leqslant M$.
7. Here we assume that the bit complexity of the constraints in $\mathcal{A}$ is $(n+m)^{O(1)}$.

Definition 8 (Sum-of-squares proof) *A* sum-of-squares *proof that a system of polynomial constraints $\mathcal{A} = \{f_1 \geqslant 0, \ldots, f_m \geqslant 0\}$ implies $q \geqslant 0$ consists of sum-of-squares polynomials $(p_S)_{S\subseteq[m]}$ such that*[8]

$$q = \sum_{\text{multiset } S\subseteq[m]} p_S \cdot \prod_{i\in S} f_i \,.$$

If such a proof exists, we say that $\mathcal{A}$ (sos-)proves *$q \geqslant 0$ within degree ℓ, denoted by $\mathcal{A} \,\big|_{\overline{\ell}}\, q \geqslant 0$. In order to clarify the variables quantified by the proof, we often write $\mathcal{A}(x) \,\big|\frac{x}{\ell}\, q(x) \geqslant 0$. We say that the system $\mathcal{A}$ is* sos-refuted *within degree ℓ if $\mathcal{A} \,\big|_{\overline{\ell}}\, -1 \geqslant 0$. Otherwise, we say that the system is* sos-consistent *up to degree ℓ, which also means that there exists a level-ℓ pseudo-distribution satisfying the system.*

B.3. Stochastic block model

Definition 9 (Symmetric balanced k-community stochastic block model) *Assume that k is a power of 2 and n is a multiple of k. Let $(\boldsymbol{G}, \boldsymbol{Z}^\circ) \sim \mathsf{SBM}_n(d, \varepsilon, k)$ denote the symmetric balanced k-community stochastic block model with degree parameter $d > 0$ and bias parameter $\varepsilon \in [0, 1]$. The model is defined as follows:*

- ***Communities.*** *The communities $\boldsymbol{Z}^\circ$ are selected uniformly at random from even partitions of the nodes, i.e. uniformly at random partition the graph into k sets of $\frac{n}{k}$ vertices each.*

- ***Edges.*** *For every pair of distinct vertices $i, j \in [n]$, they are connected by an edge independently at random with probability*

$$\mathbb{P}\{(i,j) \in E(\boldsymbol{G})\} = \begin{cases} p_1 = \big(1 + \big(1 - \frac{1}{k}\big)\varepsilon\big)\dfrac{d}{n}, & \text{if } \boldsymbol{Z}^\circ(i) = \boldsymbol{Z}^\circ(j), \\ p_2 = \Big(1 - \dfrac{\varepsilon}{k}\Big)\dfrac{d}{n}, & \text{otherwise.} \end{cases}$$

Definition 10 (Community membership matrix) *For a given community membership matrix $Z \in \{0,1\}^{n\times k}$ we define the associated centered community membership matrix*

$$X \coloneqq ZZ^\top - \frac{1}{k}J.$$

In particular, the true centered block matrix is

$$X^\circ \coloneqq Z^\circ(Z^\circ)^\top - \frac{1}{k}J\,.$$

Let G° be the uncorrupted adjacency matrix and $\bar{G}^\circ \coloneqq G^\circ - \frac{d}{n}J$ be the centered adjacency matrix, it follows that

$$\mathbb{E}\big[\bar{G}^\circ \,\big|\, Z^\circ\big] = \frac{\varepsilon d}{n} X^\circ.$$

8. Here we follow the convention that $\prod_{i\in S} f_i = 1$ for $S = \emptyset$.

Definition 11 (Signal-to-noise ratio (SNR)) *The* signal-to-noise ratio (SNR) *for the symmetric balanced k-community stochastic block model (Theorem 9) is defined as*

$$nI, \quad \textit{where} \quad I = D_{1/2}(Ber(p_1) \,\|\, Ber(p_2))$$

is the Rényi divergence of order $1/2$ between two Bernoulli distributions with parameters p_1 and p_2, as specified in Theorem 9. In the sparse regime where $d = o(n)$, this expression simplifies to

$$nI = (\sqrt{p_1 n} - \sqrt{p_2 n})^2 \cdot (1 + o(1)).$$

Throughout this paper, we denote this simplified expression by $C_{d,\varepsilon}$.

Appendix C. Statistical properties of stochastic block models

In this section, we focus on the statistical properties of majority voting in SBM. At a high level, majority voting for a single vertex u asks whether its centered signed degree $(\bar{G}y)_u \cdot y_u$ is positive. Aggregating this test over a set of candidate vertices S amounts to lower bounding the linear form $\langle \bar{G}\, y \odot y,\, z\rangle$ for indicators $z \in \{0,1\}^n$ supported on the level-i bisection (or on a pair of communities in the k-clustering step). The voting bounds in Theorems 14 and 16 show that, with high probability, this inner product is uniformly large whenever $\|z\|_1$ is not too small and remains not too negative even for very small $\|z\|_1$, while the masked variants in Theorems 15 and 17 guarantee the same after restricting to a large trusted subset. We will use these properties to certify that a small set of possibly adversarial vertices cannot flip many majority votes at once and to enable robust boosting. The proofs of the results in this section are deferred to Appendix M.

Concentration bound of single-vertex majority voting. The following result provides a concentration bound for a mixture of binomial distributions. The argument follows a generic Chernoff bound approach similar to lemma 5.6 in Liu and Moitra (2022). The main difference is that we need to find the right bound (i.e. SNR) for induced bisections of k-SBM. This boils down to bounding the moments of the following summation of binomial distributions.

Theorem 12 *Fix parameters $\beta \in (0,1]$ and $\alpha \in (0,\beta]$. Consider the distribution*

$$\mathcal{D} = \mathrm{Binom}(\alpha n, a/n) + \mathrm{Binom}\big((\beta-\alpha)n,\, b/n\big) - \mathrm{Binom}(\beta n,\, b/n)\ ,$$

where the three binomials are independent. Let $\gamma \coloneqq \alpha/\beta$ and define

$$\widetilde{a} \coloneqq a^{\gamma} b^{1-\gamma}\ , \qquad \widetilde{b} \coloneqq b\ , \qquad \widetilde{C} \coloneqq \big(\sqrt{\widetilde{a}} - \sqrt{\widetilde{b}}\big)^2\ , \qquad R(p,q) \coloneqq \frac{p(1-q)}{q(1-p)}\ .$$

Then for every $\theta \in \mathbb{R}$,

$$\mathop{\mathbb{P}}_{\boldsymbol{X}\sim\mathcal{D}}[\,\boldsymbol{X} \leqslant \theta\,] \;\leqslant\; \exp\Big(-\beta\,\widetilde{C} \;+\; \frac{\theta}{2}\log R(\widetilde{a}/n,\, \widetilde{b}/n)\Big)\ .$$

As an immediate corollary, we get:

Theorem 13 *Fix a level $i \in \{1, 2, \ldots, \log_2 k\}$ and let $\beta_i \coloneqq 2^{-i}$ (with $\alpha_i = 1/k$). Let $\widetilde{a}, \widetilde{b}, \widetilde{C}$ and $R(\cdot,\cdot)$ be as in Theorem 62, evaluated at this (α, β). Then for every $\theta \in \mathbb{R}$,*

$$\mathbb{P}[\boldsymbol{X} \leqslant \theta] \leqslant \exp\Big(-\frac{(\log 2)^2}{4} \cdot \frac{d\,\varepsilon^2}{\beta_i\, k^2} + \frac{\theta}{2}\log R(\widetilde{a}/n, \widetilde{b}/n)\Big).$$

Majority voting for bisections. We first give a theorem for the error of majority voting in induced bisections of the graph.

Theorem 14 *Fix a level i and set $\beta_i = 2^{-i}$, $n_i = 2\beta_i n$, and $k_i = \beta_i k$. Let $\gamma \in [0, 0.99]$. Choose $\rho_\gamma := \exp(-\gamma\beta_i\tilde{C}_i/2)$, and set $t = 0.001(1-\gamma)\tilde{C}_i$. Then with probability at least $1 - \exp(-100k) - \frac{1}{n^3}$, for every $z \in \{0,1\}^n$, and for every valid bisections $y \in \{0, \pm 1\}^n$ at level i. we have*

$$\langle \bar{G}y \odot y, z\rangle \geqslant \frac{(1-\gamma)\varepsilon d}{8k}\left(\|z\|_1 - \frac{96\rho_\gamma k n_i}{1-\gamma}\right),$$

where $\bar{G}$ is the centered adjacency matrix.

For establishing the feasibility of the program constraints in Appendix G and Appendix J, we need to bound the error of majority voting when a small fraction of the vertices in the graph are removed, i.e. the following corollary.

Corollary 15 *Fix a level i and set $\beta_i = 2^{-i}$, $n_i = 2\beta_i n$, and $k_i = \beta_i k$. Let $\gamma \in [0, 0.99]$. Choose $\rho_\gamma := \exp(-\gamma\beta_i\tilde{C}_i/2)$, and set $t = 0.001(1-\gamma)\tilde{C}_i$. Then with probability at least $1 - \exp(-100k) - \frac{2}{n^3}$, for every $z \in \{0,1\}^n$, for every $s \in \{0,1\}^n$ such that $\|s\|_1 \geqslant (1 - \exp(-2C_{d,\varepsilon}))n$, and for every valid bisections $y \in \{0, \pm 1\}^n$ at level i, we have*

$$\langle \bar{G} \odot (ss^\top)y \odot y, z\rangle \geqslant \frac{(1-\gamma)\varepsilon d}{16k}\left(\|z\|_1 - \frac{640\rho_\gamma k n_i}{1-\gamma}\right),$$

where $\bar{G}$ is the centered adjacency matrix.

Majority voting for pairwise communities. We give a similar theorem for bounding the error of majority voting for each pair of communities.

Theorem 16 *Let $C_{d,\varepsilon} = (\sqrt{a} - \sqrt{b})^2$. Let $\gamma \in [0, 1 - \frac{1000\chi k}{\varepsilon\sqrt{d}}]$. Choose $\rho_\gamma := \exp(-\gamma C_{d,\varepsilon}/k)$. Then with probability at least $1 - \exp(-100k) - \frac{1}{n^3}$, for every $z \in \{0,1\}^n$, and for every pair of communities $y \in \{0, \pm 1\}^n$. We have*

$$\langle \bar{G}y \odot y, z\rangle \geqslant \frac{(1-\gamma)\varepsilon d}{8k}\left(\|z\|_1 - \frac{96\rho_\gamma k n}{1-\gamma}\right),$$

where $\bar{G}$ is the centered adjacency matrix.

Similarly, for establishing the feasibility of the program constraints in Appendix J, We need the following corollary, which bounds the error of majority voting when a small fraction of the vertices are removed from the graph.

Corollary 17 *Let $\gamma \in [0, 1 - \frac{1000\chi k}{\varepsilon\sqrt{d}}]$. Choose $\rho_\gamma := \exp(-\gamma C)$. Then with probability at least $1 - \exp(-100k) - \frac{2}{n^3}$, for every $z \in \{0,1\}^n$, for every $s \in \{0,1\}^n$ such that $\|s\|_1 \geqslant (1 - \exp(-2C_{d,\varepsilon}))n$, and for every pair of communities $y \in \{0, \pm 1\}^n$. We have*

$$\langle \bar{G} \odot (ss^\top)y \odot y, z\rangle \geqslant \frac{(1-\gamma)\varepsilon d}{16k}\left(\|z\|_1 - \frac{640\rho_\gamma k n}{1-\gamma}\right),$$

where $\bar{G}$ is the centered adjacency matrix.

Appendix D. Algorithmic framework and results

In this section, we give a sketch of our robust optimal recovery algorithm (see Algorithm 42) with its key building blocks. Algorithm 42 contains two main steps: *initialization* and *boosting based on pairwise majority voting*. The goal of initialization is to get a $1/\operatorname{poly}(k)$ approximation such that, in each community, we have error rate at most 0.001. Given the $1/\operatorname{poly}(k)$-initial estimation, we can apply pairwise majority voting to boost the accuracy to the optimal error rate.

D.1. Robust bisection algorithm

The key component of our initialization algorithm is the following robust bisectioning algorithm that finds bisections in the graph with bisection error $\exp\Big(-\big(1-o(1)\big)\,\frac{C_{d,\varepsilon}}{k^2}\Big) + \operatorname{poly}(k)\eta$.

Algorithm 18 (Robust bisection algorithm)
Input: *A graph G sampled from the k-stochastic block model $\mathsf{SBM}_n(d,\varepsilon,k)$ with ηn corrupted nodes.*

1. ***Graph splitting:*** *We let G_1 be the graph obtained by subsampling each edge in G independently with probability 0.99 and let $G_2 \coloneqq G \setminus G_1$.*
2. ***Rough initialization:*** *Run a rough k-clustering algorithm on graph G_1 to obtain rough initialization Z_{rough} with error rate $0.001 + 10^4\eta$.*
3. ***Identifying well recovered blocks:*** *Use graph G_2 and Z_{rough} to identify $k/2$ clusters in which 0.99-fraction of the nodes belongs to the same community. Construct rough bisection x_{rough} by putting the $k/2$ identified clusters on one side of the bisection.*
4. ***Bisection boosting:*** *Use majority voting on bisections with x_{rough} as the initialization to obtain a bisection $\hat{x}$ with error rate $\exp\Big(-\big(1-o(1)\big)\,\frac{C_{d,\varepsilon}}{k^2}\Big) + \operatorname{poly}(k)\eta$.*

Output: *$\hat{x}$.*

We will now briefly explain and state the results for the subroutines *rough initialization*, *identifying well recovered blocks* and *bisection boosting*, then state the initialization guarantee that is achieved by recursively applying the bisection algorithm.

Rough initialization. Previous work based on semidefinite programming and spectral algorithms can get a rough initialization with error $0.001 + 10^4\,k\,\eta$ (Lemma 7.3 of Liu and Moitra (2022)). In our work, we show that a rough initialization procedure based on Sum-of-Squares (SoS) can get an improved guarantee (see Appendix E for details).

Theorem 19 (Robust rough initialization) *Let $(G^\circ, Z^\circ) \sim \mathsf{SBM}_n(d,\varepsilon,k)$ be generated from the k-stochastic block model and G be generated by adversarially corrupting η-fraction of the nodes in G°. Assume $k \leqslant n^{0.001}$, $\eta \leqslant \frac{1}{\operatorname{poly}(k)}$, $d = o(n)$, and $\varepsilon^2 d \geqslant K k^2$ for some* *sufficiently large constant K. There exists a polynomial-time algorithm that, given observation of G, outputs an estimator $\hat{Z} \in \{0,1\}^{n\times k}$ such that, with probability $1-\exp(-\Omega(k)) - \frac{1}{\mathrm{poly}(n)}$,*

$$\mathrm{error}_k(\hat{Z}, Z^\circ) \leqslant 0.001 + 10^4\,\eta.$$

Identifying well recovered clusters. Once we have a rough initialization with misclassification rate at most 0.001, at least $k/2$ clusters must be "well recovered", in the sense that at least 0.99 of the nodes in these well recovered clusters belong to the same community. We show that there exists an algorithm that can robustly identify these well recovered clusters (see Appendix F for details).

Theorem 20 (Robust identification of well recovered clusters) *Let $(G^\circ, Z^\circ) \sim \mathsf{SBM}_n(d,\varepsilon,k)$ be generated from the k-stochastic block model and G be generated by adversarially corrupting η-fraction of the nodes in G°. Assume $k \leqslant n^{0.001}$, $\eta \leqslant \frac{1}{\mathrm{poly}(k)}$, $d = o(n)$, and $\varepsilon^2 d \geqslant Kk^2$ for some sufficiently large constant K. For any set $S \subset V$ of n/k vertices, there exists a polynomial-time algorithm that, with probability $1-\exp(-\Omega(k)) - \frac{1}{\mathrm{poly}(n)}$, outputs*

- *YES if at least 0.99 of the vertices in S belong to the same community;*
- *NO if no more than 0.98 of the vertices in S belong to any single community.*

Bisection boosting. Given $k/2$ well recovered clusters, we can construct a rough bisection x_{rough} by putting them on one side of the bisection. We show that, given this rough bisection with error rate 0.001, we can use majority voting to boost the bisection accuracy to obtain a bisection with optimal bisection error (see Appendix G for details).

Theorem 21 (Robust bisection boosting) *Let $(G^\circ, Z^\circ) \sim \mathsf{SBM}_n(d,\varepsilon,k)$ be generated from the k-stochastic block model and G be generated by adversarially corrupting η-fraction of the nodes in G°. Assume $k \leqslant n^{0.001}$, $\eta \leqslant \frac{1}{\mathrm{poly}(k)}$, $d = o(n)$, and $\varepsilon^2 d \geqslant Kk^2$ for some sufficiently large constant K. Let $S_1, S_2, \ldots, S_{k/2} \subset V$ be disjoint subsets of size n/k such that in every S_i at least 0.99-fraction of the nodes belong to the same community. Let $x^\circ \in \{\pm 1\}^n$ be the ground-truth community bisection with the underlying communities of $S_1, S_2, \ldots, S_{k/2}$ on the same side. There exists a polynomial-time algorithm that, given observation of G and $S_1, S_2, \ldots, S_{k/2}$, outputs $\hat{x} \in \{\pm 1\}^n$ such that, with probability $1-\exp(-\Omega(k)) - \frac{1}{\mathrm{poly}(n)}$,*

$$\frac{1}{n}\|\hat{x} - x^\circ\|^2 \leqslant \exp\left(-\big(1-o(1)\big)\frac{\tilde{C}}{8}\right) + \mathrm{poly}(k)\eta\,,$$

where $\tilde{C} = \left(\sqrt{a^{\frac{2}{k}} b^{1-\frac{2}{k}}} - \sqrt{b}\right)^2$ is the bisection SNR.

Robust bisection algorithm. As a corollary, by combining Theorem 19, Theorem 20 and Theorem 21, we have a robust bisection algorithm for the k-stochastic block model with optimal bisection error rate (see Appendix H for details).

Theorem 22 (Robust bisection algorithm) *Let* $(G^\circ, Z^\circ) \sim \mathsf{SBM}_n(d, \varepsilon, k)$ *be generated from the* k*-stochastic block model and* G *be generated by adversarially corrupting* η*-fraction of the nodes in* G°*. Assume* $k \leqslant n^{0.001}$, $\eta \leqslant \frac{1}{\mathrm{poly}(k)}$, $d = o(n)$, *and* $\varepsilon^2 d \geqslant Kk^2$ *for some sufficiently large constant* K*. There exists a polynomial-time algorithm that, given observation of* G*, outputs* $\hat{x} \in \{\pm 1\}^n$ *such that, with probability* $1 - \exp(-\Omega(k)) - \frac{1}{\mathrm{poly}(n)}$,

$$\frac{1}{n}\|\hat{x} - x^\circ\|^2 \leqslant \exp\left(-\big(1 - o(1)\big)\frac{\tilde{C}}{8}\right) + O(k\eta).$$

where $\tilde{C} = \left(\sqrt{a^{\frac{2}{k}} b^{1-\frac{2}{k}}} - \sqrt{b}\right)^2$ *is the bisection SNR and* $x^\circ \in \{\pm 1\}^n$ *is a true community bisection of* G°*.*

D.2. Robust initialization for symmetry breaking

By recursively applying Theorem 22, we can cluster the vertices into k communities with error rate $\exp\Big(-\big(1 - o(1)\big)\,\frac{C_{d,\varepsilon}}{k^2}\Big) + \mathrm{poly}(k)\eta$ (see Appendix I for details). When $C_{d,\varepsilon} \geqslant \Omega(k^2 \log k)$, we get an error rate of $1/\mathrm{poly}(k) + \mathrm{poly}(k)\eta$.

Theorem 23 (Robust initialization for symmetry breaking) *Let* $(G^\circ, Z^\circ) \sim \mathsf{SBM}_n(d, \varepsilon, k)$ *be generated from the* k*-stochastic block model and* G *be generated by adversarially corrupting* η*-fraction of the nodes in* G°*. Assume* $k \leqslant n^{0.001}$, $\eta \leqslant \frac{1}{\mathrm{poly}(k)}$, $d = o(n)$, *and* $\varepsilon^2 d \geqslant Kk^2$ *for some sufficiently large constant* K*. There exists a polynomial-time algorithm that, given observation of* G*, outputs an estimator* $\hat{Z} \in \{0,1\}^{n \times k}$ *such that, with probability* $1 - \exp(-\Omega(k)) - \frac{1}{\mathrm{poly}(n)}$,

$$\mathit{error}_k(\hat{Z}, Z^\circ) \leqslant \exp\left(-\big(1 - o(1)\big)\,\frac{C_{d,\varepsilon}}{k^2}\right) + \mathrm{poly}(k)\eta.$$

D.3. Boosting via pairwise majority voting

Given the initialization with error $1/\mathrm{poly}(k) + \mathrm{poly}(k)\eta$, we can apply pairwise majority voting to boost the accuracy of the estimator to the optimal error rate. The formal main theorem of our robust clustering algorithm for node-corrupted k-stochastic block model is as follows. The algorithm and proofs of Theorem 24 are presented in Appendix J.

Theorem 24 (Robust recovery in k-SBM with optimal rate) *There exist universal constants* $C_0 > C_1 > 0$ *and* $K > 0$ *such that the following holds. Let* $(G^\circ, Z^\circ) \sim \mathsf{SBM}_n(d, \varepsilon, k)$ *be generated from the* k*-stochastic block model and* G *be generated by adversarially corrupting* η*-fraction of the nodes in* G°*. Assume* $k \leqslant n^{0.001}$, $\eta \leqslant k^{-C_0}$, $d = o(n)$, *and* $\varepsilon^2 d \geqslant Kk^2 \log k$*. There exists a polynomial-time algorithm (see Algorithm 42) that, given observation of* G*, outputs an estimator* $\hat{Z} \in \{0,1\}^{n \times k}$ *such that, with probability* $1 - \exp(-\Omega(k)) - \frac{1}{\mathrm{poly}(n)}$,

$$\mathit{error}_k(\hat{Z}, Z^\circ) \leqslant \exp\Big(-\big(1 - o(1)\big)\frac{C_{d,\varepsilon}}{k}\Big) + k^{C_1}\eta\,.$$

Here $o(1)$ *denotes a quantity tending to zero in the regime* $\varepsilon^2 d/k^2 \to \infty$*.*

Appendix E. Robust rough initialization

In this section, we prove Theorem 19 via a natural SoS-based initialization procedure. Our analysis is of independent interest as it both simplifies and strictly improves upon the initialization guarantees of Liu and Moitra (2022). In particular, we distinguish two regimes based on the adversarial corruption fraction η:

- If $\eta \sim \exp\big(-C_{d,\varepsilon}/k\big)$ (the critical, minimax regime), our goal is to obtain an initial recovery error bounded by a small constant (e.g. 0.001).
- If η is significantly larger than $\exp\big(-C_{d,\varepsilon}/k\big)$, the recovery error is dominated by the corruption level, meaning that one can only hope to recover the community structure up to the inherent corruption threshold.

In the former case, both our method and that of Liu and Moitra (2022) achieve the desired 0.001 recovery error, which is sufficient to bootstrap our subsequent boosting step. In the latter regime, however, our SoS formulation offers a significant improvement by tolerating a constant corruption fraction η independently of k. By contrast, the guarantee in Liu and Moitra (2022) requires that $\eta \lesssim 1/k$ in order to maintain the misclassification error (on the order of $k\eta$) bounded, rendering their bound vacuous as k diverges.

Theorem *[Restatement of Theorem 19] Let $(G^\circ, Z^\circ) \sim \mathsf{SBM}_n(d,\varepsilon,k)$ be generated from the k-stochastic block model and G be generated by adversarially corrupting η-fraction of the nodes in G°. Assume $k \leqslant n^{0.001}$, $\eta \leqslant \frac{1}{\mathrm{poly}(k)}$, $d = o(n)$, and $\varepsilon^2 d \geqslant Kk^2$ for some sufficiently large constant K. There exists a polynomial-time algorithm that, given observation of G, outputs an estimator $\hat{Z} \in \{0,1\}^{n\times k}$ such that, with probability $1-\exp(-\Omega(k)) - \frac{1}{\mathrm{poly}(n)}$,*

$$\mathit{error}_k(\hat{Z}, Z^\circ) \leqslant 0.001 + 10^4\,\eta.$$

We consider a constant-degree sum-of-squares relaxation for the following program:

$$\mathcal{A}_{\mathrm{init}}(\xi, G, Z; G_{\mathrm{input}}, \eta) := \left\{ \begin{array}{l} \xi \odot \xi = \xi, \quad \langle \xi, \mathbf{1}\rangle \geqslant \Big(1-\eta-\exp\Big(-\dfrac{2\varepsilon^2 d}{k}\Big)\Big)n, \\ Z \odot Z = Z, \quad \displaystyle\sum_{a\in[k]} Z(i,a) = 1, \quad \sum_{i\in[n]} Z(i,a) = \frac{n}{k}, \\ \Big\| G - \dfrac{d}{n}\,11^\top - \dfrac{\varepsilon d}{n}\,(ZZ^\top - \dfrac{1}{k}\,J) \Big\|_{\mathrm{op}} \leqslant \chi\sqrt{d}, \\ G \odot \xi\xi^\top = G_{\mathrm{input}} \odot \xi\xi^\top. \end{array} \right\} \tag{E.1}$$

For simplicity of notation, we let

$$\mathcal{A}_{\mathrm{bool}}(\xi, Z) := \left\{ \xi \odot \xi = \xi, Z \odot Z = Z, \sum_{a\in[k]} Z(i,a) = 1, \sum_{i\in[n]} Z(i,a) = n/k \right\}$$

$$\mathcal{A}_{\mathrm{spec}}(G, Z) := \left\{ \Big\| G - \frac{d}{n}\,11^\top - \frac{\varepsilon d}{n}\,(ZZ^\top - \frac{1}{k}\,J) \Big\|_{\mathrm{op}} \leqslant \chi\sqrt{d} \right\}$$

$$\mathcal{A}_{\text{corr}}(\xi, G; G_{\text{input}}, \eta) := \left\{ G \odot \xi\xi^\top = G_{\text{input}} \odot \xi\xi^\top, \langle \xi, \mathbf{1} \rangle \geqslant \big(1 - \eta - \exp(-2\varepsilon^2 d/k)\big)n \right\}$$

Then we have

$$\mathcal{A}_{\text{init}}(\xi, G, Z; G_{\text{input}}, \eta) = \mathcal{A}_{\text{bool}}(\xi, Z) \cup \mathcal{A}_{\text{spec}}(G, Z) \cup \mathcal{A}_{\text{corr}}(\xi, G; G_{\text{input}}, \eta)\,.$$

We first show that the program is feasible with high probability.

Lemma 25 *Let G_{input} be an η-corrupted stochastic block model (SBM) graph on n nodes partitioned into k communities. Suppose that*

$$\varepsilon^2 d \geqslant K\, k^2,$$

for a sufficiently large constant $K > 0$. Then, with high probability, the feasibility conditions in Eq. (E.1) are satisfied by the choice $\xi = \xi_0 \odot \mathbf{1}_S$ (where S is as in Cor. 58), $Z = Z^\circ$, and a suitable G agreeing with G_{input} on the subgraph induced by the set of vertices indicated by ξ.

Proof By Corollary 58, there exists a subset $S \subseteq [n]$ with $|S| \geqslant \big(1 - \exp(-2C_{d,\varepsilon})\big)n$ such that

$$\left\| \Big(G^\circ - \frac{d}{n} J - \frac{\varepsilon d}{n} \left(Z^\circ Z^{\circ\top} - \tfrac{1}{k} J \right) \Big) \odot \mathbf{1}_S \mathbf{1}_S^\top \right\|_{\text{op}} \leqslant \chi\sqrt{d}\ ,$$

where χ is a universal constant. Let ξ be the indicator of the uncorrupted nodes intersected with S, i.e., $\xi := \xi_0 \odot \mathbf{1}_S$, and take $Z^\circ \in \{0,1\}^{n\times k}$ to be the ground-truth membership matrix. Define G to agree with G_{input} on the block indexed by $\text{supp}(\xi)$ and to equal the model mean $\frac{d}{n}J + \frac{\varepsilon d}{n}\big(Z^\circ (Z^\circ)^\top - \frac{1}{k}J\big)$ outside this block. Then the centered matrix $G - \frac{d}{n}J - \frac{\varepsilon d}{n}\big(Z^\circ (Z^\circ)^\top - \frac{1}{k}J\big)$ vanishes outside $\text{supp}(\xi)$ and, on $\text{supp}(\xi)$, its operator norm is bounded by $\chi\sqrt{d}$ by Corollary 58. Hence the spectral constraint in Eq. (E.1) holds. Moreover, $\langle \xi, \mathbf{1} \rangle \geqslant \big(1 - \eta - \exp(-2C_{d,\varepsilon})\big)n$, which satisfies $\mathcal{A}_{\text{corr}}$. Therefore, with these choices of ξ, $Z = Z^\circ$, and G, all feasibility conditions in Eq. (E.1) are satisfied with high probability. ∎

Now, we give the SoS proof for the recovery guarantees.

Lemma 26 *Let G_{input} be an η-corrupted SBM graph with n nodes and k communities, and assume that*

$$\varepsilon^2 d \geqslant K k^2,$$

for some sufficiently large constant K. Then, we have

$$\mathcal{A}_{init}(\xi^{(1)}, G^{(1)}, Z^{(1)}; G_{input}, \eta), \quad \mathcal{A}_{init}(\xi^{(2)}, G^{(2)}, Z^{(2)}; G_{input}, \eta)$$
$$\Big|\frac{\xi^{(1)}, \xi^{(2)}, G^{(1)}, G^{(2)}, Z^{(1)}, Z^{(2)}}{8}\ \left\| Z^{(1)}(Z^{(1)})^\top - Z^{(2)}(Z^{(2)})^\top \right\|_{\text{F}}^2 \leqslant (0.001 + 10\eta)n^2/k\,.$$

Proof Since the constraints $G^{(1)} \odot \xi^{(1)}(\xi^{(1)})^\top = G_{\text{input}} \odot \xi^{(1)}(\xi^{(1)})^\top$ and $G^{(2)} \odot \xi^{(2)}(\xi^{(2)})^\top = G_{\text{input}} \odot \xi^{(2)}(\xi^{(2)})^\top$ hold, let $\xi = \xi^{(1)} \odot \xi^{(2)}$. We then have a degree-8 sum-of-squares proof that

$$G^{(1)} \odot \xi\xi^\top = G_{\text{input}} \odot \xi\xi^\top \quad \text{and} \quad G^{(2)} \odot \xi\xi^\top = G_{\text{input}} \odot \xi\xi^\top\,.$$

Therefore, we have $G^{(1)} \odot \xi\xi^\top = G^{(2)} \odot \xi\xi^\top$. By the min–max characterization of the spectral norm,

$$\mathcal{A}_{\text{bool}}(\xi^{(1)}, Z^{(1)}) \cup \mathcal{A}_{\text{spec}}(G^{(1)}, Z^{(1)}) \,\Big|\!\frac{\;G^{(1)},Z^{(1)}\;}{8}\; \left\| \left(G^{(1)} - \frac{d}{n} J - \frac{\varepsilon d}{n} (Z^{(1)} (Z^{(1)})^\top - \frac{1}{k} J) \right) \odot \xi\xi^\top \right\|_{\text{op}} \leqslant \chi\sqrt{d}$$

$$\mathcal{A}_{\text{bool}}(\xi^{(2)}, Z^{(2)}) \cup \mathcal{A}_{\text{spec}}(G^{(2)}, Z^{(2)}) \,\Big|\!\frac{\;G^{(2)},Z^{(2)}\;}{8}\; \left\| \left(G^{(2)} - \frac{d}{n} J - \frac{\varepsilon d}{n} (Z^{(2)} (Z^{(2)})^\top - \frac{1}{k} J) \right) \odot \xi\xi^\top \right\|_{\text{op}} \leqslant \chi\sqrt{d}$$

By the triangle inequality, this yields a degree-8 sum-of-squares certificate that

$$\left\| \frac{\varepsilon d}{n} \left(Z^{(1)} (Z^{(1)})^\top - Z^{(2)} (Z^{(2)})^\top \right) \odot \xi\xi^\top \right\|_{\text{op}} \leqslant 2\chi\sqrt{d}\,.$$

Since $Z^{(1)}$ and $Z^{(2)}$ are community membership matrices, each Gram matrix $Z^{(t)}(Z^{(t)})^\top$ has rank at most k, so $\left(Z^{(1)}(Z^{(1)})^\top - Z^{(2)}(Z^{(2)})^\top\right) \odot \xi\xi^\top$ has rank at most $2k$[9]. Thus, we obtain a degree-8 sum-of-squares proof that

$$\left\| \left(Z^{(1)} (Z^{(1)})^\top - Z^{(2)} (Z^{(2)})^\top \right) \odot \xi\xi^\top \right\|_{\text{F}}^2 \leqslant \frac{8\chi^2 k n^2}{\varepsilon^2 d}\,.$$

Finally, since $\|\xi\|_1 \geqslant (1 - 2\eta - 2\exp(-2\varepsilon^2 d/k))n$, we have

$$\mathcal{A}_{\text{init}}(\xi^{(1)}, G^{(1)}, Z^{(1)}; G_{\text{input}}, \eta), \mathcal{A}_{\text{init}}(\xi^{(2)}, G^{(2)}, Z^{(2)}; G_{\text{input}}, \eta)$$
$$\Big|\!\frac{\;\xi^{(1)},\xi^{(2)},G^{(1)},G^{(2)},Z^{(1)},Z^{(2)}\;}{8}\; \left\| Z^{(1)} (Z^{(1)})^\top - Z^{(2)} (Z^{(2)})^\top \right\|_{\text{F}}^2 \leqslant \frac{8\chi^2 k n^2}{\varepsilon^2 d} + 10\eta n^2/k\,.$$

■

Finally, we conclude with the rounding step.

Proof [Proof of Theorem 19] Let us write $M^\circ \coloneqq Z^\circ (Z^\circ)^\top$ for the ground-truth matrix. By Theorem 25, the ground-truth community matrix Z° is feasible for Eq. (E.1). Hence, any pseudo-distribution satisfying these constraints obeys

$$\|ZZ^\top - M^\circ\|_F^2 \leqslant 10\eta\, n^2/k + \frac{8\chi^2 k\, n^2}{\varepsilon^2\, d}\,.$$

Standard rounding then yields a community matrix $M = \tilde{\mathbb{E}} ZZ^T \in [0,1]^{n\times n}$ such that

$$\|M - M^\circ\|_F^2 \leqslant \Big(\frac{10\,\eta}{k} + \frac{8\,\chi^2\,k}{\varepsilon^2\,d}\Big)\, n^2. \tag{E.2}$$

Having established the Frobenius-norm bound, the rest of the proof mirrors Lemma 7.3 of Liu and Moitra (2022); we present it below for completeness. Since the rows of M lie in $\mathbb{R}^n$, we can apply a Euclidean k-means algorithm with a $\gamma = 7$ approximation guarantee Grandoni et al. (2021). Evaluating the k-means objective at the ground-truth partition $(S_1, \ldots, S_k)$ shows that

$$k\text{-means}_{\text{opt}}(M) \;\leqslant\; \|M - M^\circ\|_F^2 \;\leqslant\; \left(\frac{10\,\eta}{k} + \frac{8\,\chi^2\,k}{\varepsilon^2\,d}\right) n^2,$$

9. Indeed, letting $D = \operatorname{diag}(\xi)$ and $\Delta = Z^{(1)}(Z^{(1)})^\top - Z^{(2)}(Z^{(2)})^\top$, we have $\Delta \odot \xi\xi^\top = D\Delta D$, hence $\operatorname{rank}(\Delta \odot \xi\xi^\top) \leqslant \operatorname{rank}(\Delta) \leqslant 2k$.

where the last inequality is Eq. (E.2).

Now fix any alternative clustering $S'_1, \ldots, S'_k$ of the rows, and consider the same clustering evaluated on M°. For one cluster, say S'_1, let

$$s_b = |S'_1 \cap S_b| \quad \text{for } b \in [k].$$

As in Lemma 7.3 of Liu and Moitra (2022), the k-means objective contributed by S'_1 on the rows of M° is at least

$$\Big(s_1 + \cdots + s_k - \max\{s_1, \ldots, s_k\}\Big) \cdot \frac{n}{4k}. \tag{E.3}$$

Summing (E.3) over all clusters S'_a lower-bounds the k-means objective on M° for the partition $S'_1, \ldots, S'_k$. Let the computed clustering be $S'_1, \ldots, S'_k$. Define

$$\delta = \frac{1}{n} \min_{\substack{\pi:[k]\to[k] \\ \pi \text{ invertible}}} \sum_{a=1}^{k} |S'_a \setminus S_{\pi(a)}|, \qquad \delta' = \frac{1}{n} \min_{f:[k]\to[k]} \sum_{a=1}^{k} |S'_a \setminus S_{f(a)}|.$$

Here δ is exactly the misclassification error of $S'_1, \ldots, S'_k$ relative to the ground truth. As in Liu and Moitra (2022), one has $\delta \leqslant 2\delta'$. Moreover, using (E.3) and summing over clusters shows that the k-means objective of $S'_1, \ldots, S'_k$ on M° is at least $\frac{n^2}{4k}\,\delta'$. Combining this lower bound with the γ-approximation guarantee on M and Eq. (E.2) (exactly as in Lemma 7.3 of Liu and Moitra (2022)) yields

$$\delta \leqslant 10^4\, k \left(\frac{\eta}{k} + \frac{\chi^2\, k}{\varepsilon^2\, d} \right) ,$$

which is the desired bound. ■

Appendix F. Robustly identifying well-recovered clusters

In this section, we show that given a set of n/k vertices in the graph sampled from $\mathsf{SBM}_n(d, \varepsilon, k)$, we can decide between

- Case I: at least 0.99 of the vertices in the set belong to the same community;
- Case II: no more than 0.98 of the vertices in the set belong to any single community.

Restriction to the given set. Let $S \subseteq [n]$ denote the given set of $|S| = n/k$ vertices, and let $\mathbf{1} = \mathbf{1}_S \in \{0,1\}^n$ be its indicator. We define the matrices restricted to S as

$$G_s \coloneqq G \odot (\mathbf{1}_S\, \mathbf{1}_S^\top) \qquad \text{and} \qquad J_s \coloneqq \mathbf{1}_S\, \mathbf{1}_S^\top\,.$$

Throughout this section, every appearance of G and J inside operator norms or inner products is to be understood as G_s and J_s respectively. We also view all vectors (e.g., z, τ, ζ) as elements of $\mathbb{R}^n$ supported on S.

Theorem *[Restatement of Theorem 20] Let $(G^\circ, Z^\circ) \sim \mathsf{SBM}_n(d, \varepsilon, k)$ be generated from the k-stochastic block model and G be generated by adversarially corrupting η-fraction of the nodes in G°. Assume $k \leqslant n^{0.001}$, $\eta \leqslant \frac{1}{\mathrm{poly}(k)}$, $d = o(n)$, and $\varepsilon^2 d \geqslant K k^2$ for some sufficiently large constant K. For any set $S \subset V$ of n/k vertices, there exists a polynomial-time algorithm that, with probability $1 - \exp(-\Omega(k)) - \frac{1}{\mathrm{poly}(n)}$, outputs*

- *YES if at least 0.99 of the vertices in S belong to the same community;*
- *NO if no more than 0.98 of the vertices in S belong to any single community.*

We will show that the feasibility of the degree-2 SoS relaxation of the following program satisfies Theorem 20.

$$\mathcal{A}_{\text{verify}}(z; G, \eta) := \begin{cases} z \odot z = z,\ \langle z, \mathbf{1} \rangle \geqslant \left(\frac{0.99}{k} - \eta - \rho\right) n \\ \left\| \left(G_s - \frac{\alpha}{n} J_s\right) \odot z z^\top \right\|_{\mathrm{op}} \leqslant \chi \sqrt{\alpha / k} \\ \left\langle \left(G_s - \frac{d}{n} J_s\right) \odot z z^\top, \mathbf{1}\mathbf{1}^\top \right\rangle \geqslant \frac{0.97(k-1)\varepsilon d n}{k^3} \end{cases} , \tag{F.1}$$

where $\alpha = d + (1 - \frac{1}{k})\varepsilon d$, η is the corruption rate, χ is the universal constant defined in the spectral result in Theorem 57, and $\rho = \exp(-2C_{d,\varepsilon})$ is the fraction of high-degree vertices that need to be trimmed to obtain the spectral norm bound (see Theorem 58).

Lemma 27 *Given a set of n/k vertices in the graph sampled from $\mathsf{SBM}_n(d, \varepsilon, k)$ where at least 0.99 fraction of the vertices in the set belongs to the same community (Case I), when $\eta \leqslant C_\eta$ for some constant C_η that is small enough, Program Eq. (F.1) is feasible with probability at least $1 - n^{-\Omega(1)}$.*

Proof Let $\tau \in \{0, 1\}^n$ (supported on S) be the indicator vector for the largest set of uncorrupted and degree-bounded nodes (according to Theorem 57) that are in the same community. We will show that, with probability $1 - n^{-\Omega(1)}$, τ is a feasible solution to the program Eq. (F.1).

By assumption, $\|\tau\|_1 \geqslant (\frac{0.99}{k} - \eta - \rho) n$. Therefore, τ satisfies the integrality and norm constraints of program Eq. (F.1).

By Theorem 57, with probability at least $1 - \frac{k^2}{n^2}$, we have $\left\| \left(G_s - \frac{\alpha}{n} J_s\right) \odot \tau\tau^\top \right\|_{\mathrm{op}} \leqslant \chi \sqrt{\frac{\alpha}{k}}$. Therefore, τ satisfies the spectral constraint of program Eq. (F.1).

Now, consider $\left\langle G_s - \frac{d}{n} J_s, \tau\tau^\top \right\rangle$. Notice that $\left\langle G_s, \tau\tau^\top \right\rangle$ is equal to 2 times the number of edges in the induced subgraph $G_s(\tau)$ where each edge is sampled independently with probability $\frac{\alpha}{n}$. Therefore, by Chernoff bound, it follows that, with probability at least $1 - n^{-100}$,

$$\left\langle G_s, \tau\tau^\top \right\rangle \geqslant \frac{\alpha}{n} \|\tau\|_1^2 - O\left(\sqrt{\frac{\alpha \log(n)}{n}} \|\tau\|_1^2 \right).$$

Hence,

$$\left\langle G_s - \frac{d}{n} J_s, \tau\tau^\top \right\rangle \geqslant \frac{\alpha - d}{n} \|\tau\|_1^2 - O\left(\sqrt{\frac{\alpha \log(n)}{n}} \|\tau\|_1^2 \right).$$

Since $\alpha = d + (1-\frac{1}{k})\varepsilon d$, $\alpha \leqslant 2d$ and $\|\tau\|_1 \geqslant (\frac{0.99}{k} - \eta - \rho)n$, it follows that

$$\left\langle G_s - \frac{d}{n}J_s, \tau\tau^\top \right\rangle \geqslant \frac{0.98(k-1)\varepsilon dn}{k^3} - O\Big(\sqrt{d\log(n)}\Big) \geqslant \frac{0.975(k-1)\varepsilon dn}{k^3}\,.$$

Thus, τ satisfies the sum constraint of program Eq. (F.1), which finishes the proof. ■

Lemma 28 *Given a set of n/k vertices in the graph sampled from $\mathsf{SBM}_n(d,\varepsilon,k)$ where at most a 0.98 fraction of the vertices in the set belongs to the same community (Case II), when $\eta \leqslant C_\eta$ for some constant C_η that is small enough, Program Eq. (F.1) is not feasible with probability at least $1-n^{-\Omega(1)}$.*

Proof We will show that there is an SoS proof of

$$\mathcal{A}_{\text{verify}}(z;G,\eta)\Big|\frac{z}{2}\left\langle \left(G_s - \frac{d}{n}J_s\right)\odot zz^\top, \mathbf{1}\,\mathbf{1}^\top \right\rangle \leqslant \frac{0.963(k-1)\varepsilon dn}{k^3}\,,$$

which is a refutation to constraint $\left\langle\left(G_s - \frac{d}{n}J_s\right)\odot zz^\top, \mathbf{1}\,\mathbf{1}^\top\right\rangle \geqslant \frac{0.97(k-1)\varepsilon dn}{k^3}$ in $\mathcal{A}_{\text{verify}}(z;G,\eta)$.

Let us denote the set of uncorrupted and degree-bounded (according to Theorem 58) nodes by $\zeta \in \{0,1\}^n$ (supported on S). By Theorem 58, it follows that $\|\zeta\|_1 \geqslant (1-\eta-\rho)\frac{n}{k}$ and $\left\|\left((G_0)_s - \frac{d}{n}J_s\right)\odot\zeta\zeta^\top\right\|_{\text{op}} \leqslant \chi\sqrt{d/k}$, where $(G_0)_s \coloneqq G_0 \odot (\mathbf{1}\,\mathbf{1}^\top)$. Consider the largest set of vertices in ζ that are from the same community, denoted by $\tau \in \{0,1\}^n$ (supported on S). By our assumption, it follows that $\|\tau\|_1 \leqslant 0.98\frac{n}{k}$.

We will decompose $\left\langle\left(G_s - \frac{d}{n}J_s\right)\odot zz^\top, \mathbf{1}\,\mathbf{1}^\top\right\rangle$ into the following three terms and bound them separately

$$\begin{aligned}\left\langle \left(G_s - \frac{d}{n}J_s\right)\odot zz^\top, \mathbf{1}\,\mathbf{1}^\top \right\rangle =& \left\langle \left(G_s - \frac{d}{n}J_s\right)\odot zz^\top, \tau\tau^\top + (\zeta-\tau)(\zeta-\tau)^\top \right\rangle \\ &+ \left\langle \left(G_s - \frac{d}{n}J_s\right)\odot zz^\top, 2(\zeta-\tau)\tau^\top \right\rangle + \left\langle \left(G_s - \frac{d}{n}J_s\right)\odot zz^\top, \mathbf{1}\,\mathbf{1}^\top - \zeta\zeta^\top \right\rangle.\end{aligned} \tag{F.2}$$

Term 1. For the first term $\left\langle\left(G_s - \frac{d}{n}J_s\right)\odot zz^\top, \tau\tau^\top + (\zeta-\tau)(\zeta-\tau)^\top\right\rangle$, we can further decompose it into the following two terms

$$\begin{aligned}\left\langle \left(G_s - \frac{d}{n}J_s\right)\odot zz^\top, \tau\tau^\top + (\zeta-\tau)(\zeta-\tau)^\top \right\rangle =& \left\langle \left(G_s - \frac{d}{n}J_s\right)\odot zz^\top, \tau\tau^\top \right\rangle \\ &+ \left\langle \left(G_s - \frac{d}{n}J_s\right)\odot zz^\top, (\zeta-\tau)(\zeta-\tau)^\top \right\rangle.\end{aligned}$$

The two terms can be bounded using the same method. Consider the first term $\left\langle\left(G_s - \frac{d}{n}J_s\right)\odot zz^\top, \tau\tau^\top\right\rangle$, it follows that

$$\left\langle \left(G_s - \frac{d}{n}J_s\right)\odot zz^\top, \tau\tau^\top \right\rangle = \left\langle \left(G_s - \frac{\alpha}{n}J_s\right)\odot zz^\top, \tau\tau^\top \right\rangle + \left\langle \frac{\alpha - d}{n}J_s \odot zz^\top, \tau\tau^\top \right\rangle$$

$$= \left\langle \left(G_s - \frac{\alpha}{n} J_s\right) \odot zz^\top, \tau\tau^\top \right\rangle + \frac{\alpha - d}{n} \langle \tau, z\rangle^2 .$$

Since $\mathcal{A}_{\text{verify}}$ certifies $\left\|(G_s - \frac{\alpha}{n} J_s) \odot zz^\top\right\|_{\text{op}} \leqslant \chi\sqrt{\alpha/k}$, it follows that

$$\mathcal{A}_{\text{verify}}(z; G, \eta) \,\Big|\tfrac{z}{2}\, \left\langle \left(G_s - \frac{d}{n} J_s\right) \odot zz^\top, \tau\tau^\top \right\rangle \leqslant \chi\sqrt{\alpha/k}\|\tau\|^2 + \frac{\alpha - d}{n}\langle \tau, z\rangle^2 .$$

Using the same analysis by replacing τ by $\zeta - \tau$, we can also obtain

$$\mathcal{A}_{\text{verify}}(z; G, \eta) \,\Big|\tfrac{z}{2}\, \left\langle \left(G_s - \frac{d}{n} J_s\right) \odot zz^\top, (\zeta - \tau)(\zeta - \tau)^\top \right\rangle \leqslant \chi\sqrt{\alpha/k}\|\zeta - \tau\|^2 + \frac{\alpha - d}{n}\langle \zeta - \tau, z\rangle^2 .$$

Combining the two inequalities, we get

$$\begin{aligned} \mathcal{A}_{\text{verify}}(z; G, \eta) \,\Big|\tfrac{z}{2} &\left\langle \left(G_s - \frac{d}{n} J_s\right) \odot zz^\top, \tau\tau^\top + (\zeta - \tau)(\zeta - \tau)^\top \right\rangle \\ &\leqslant \frac{\alpha - d}{n}\left(\langle \tau, z\rangle^2 + \langle \zeta - \tau, z\rangle^2\right) + \chi\sqrt{\alpha/k}\left(\|\tau\|^2 + \|\zeta - \tau\|^2\right) . \end{aligned} \tag{F.3}$$

Since $\mathcal{A}_{\text{verify}}(z; G, \eta) \,\Big|\tfrac{z}{2}\, 0 \leqslant z \leqslant 1$, it follows that $\langle \tau, z\rangle \leqslant \|\tau\|_1$ and $\langle \zeta - \tau, z\rangle \leqslant \|\zeta - \tau\|_1$. Moreover, since $\|\tau\|_1 \leqslant 0.98\frac{n}{k}$ and $\|\zeta\|_1 \leqslant \frac{n}{k}$, the value of $\|\tau\|_1^2 + \|\zeta - \tau\|_1^2$ is upper bounded by $(0.98\frac{n}{k})^2 + (0.02\frac{n}{k})^2 \leqslant 0.961\frac{n^2}{k^2}$. Therefore, we can obtain

$$\mathcal{A}_{\text{verify}}(z; G, \eta) \,\Big|\tfrac{z}{2}\, \langle \tau, z\rangle^2 + \langle \zeta - \tau, z\rangle^2 \leqslant \|\tau\|_1^2 + \|\zeta - \tau\|_1^2 \leqslant 0.961\frac{n^2}{k^2} . \tag{F.4}$$

Since τ and ζ are Boolean vectors and $\tau \subseteq \zeta$, it follows that

$$\|\tau\|^2 + \|\zeta - \tau\|^2 = \|\zeta\|^2 \leqslant \frac{n}{k} . \tag{F.5}$$

Plugging Eq. (F.4) and Eq. (F.5) into Eq. (F.3), we can obtain

$$\begin{aligned} \mathcal{A}_{\text{verify}}(z; G, \eta) \,\Big|\tfrac{z}{2} &\left\langle \left(G_s - \frac{d}{n} J_s\right) \odot zz^\top, \tau\tau^\top + (\zeta - \tau)(\zeta - \tau)^\top \right\rangle \\ &\leqslant \frac{0.961(\alpha - d)n}{k^2} + \frac{\chi\sqrt{\alpha}n}{\sqrt{k^3}} . \end{aligned}$$

Plugging in $\alpha = d + (1 - \frac{1}{k})\varepsilon d$ and $\alpha \leqslant 2d$, we get

$$\mathcal{A}_{\text{verify}}(z; G, \eta) \,\Big|\tfrac{z}{2}\, \left\langle \left(G_s - \frac{d}{n} J_s\right) \odot zz^\top, \tau\tau^\top + (\zeta - \tau)(\zeta - \tau)^\top \right\rangle \leqslant \frac{0.961(k-1)\varepsilon dn}{k^3} + \frac{2\chi\sqrt{d}n}{\sqrt{k^3}} . \tag{F.6}$$

Term 2. For the second term $\left\langle (G_s - \frac{d}{n} J_s) \odot zz^\top, 2(\zeta - \tau)\tau^\top \right\rangle$, since we are summing over entries in the set of uncorrupted vertices ζ, it follows that

$$\left\langle \left(G_s - \frac{d}{n} J_s\right) \odot zz^\top, 2(\zeta - \tau)\tau^\top \right\rangle = 2\left\langle \left((G_0)_s - \frac{d}{n} J_s\right) \odot \zeta\zeta^\top, (\zeta - \tau)\tau^\top \odot zz^\top \right\rangle .$$

To simplify notation, let us denote $M_d = \left((G_0)_s - \frac{d}{n}J_s\right) \odot \zeta\zeta^\top$. It follows by SoS Cauchy-Schwarz (Theorem 52) that

$$\begin{aligned} 2\left\langle M_d, (\zeta-\tau)\tau^\top \odot zz^\top \right\rangle &= 2\langle M_d \cdot (\tau \odot z), (\zeta-\tau)\odot z\rangle \\ &\leqslant \frac{\|M_d \cdot (\tau\odot z)\|^2}{\chi\sqrt{d/k}} + \chi\sqrt{d/k}\|(\zeta-\tau)\odot z\|^2 . \end{aligned}$$

Since we chose ζ such that $\|M_d\|_{\mathrm{op}} \leqslant \chi\sqrt{d/k}$, it follows that

$$2\left\langle M_d, (\zeta-\tau)\tau^\top \odot zz^\top \right\rangle \leqslant \chi\sqrt{d/k}\left(\|(\zeta-\tau)\odot z\|^2 + \|\tau\odot z\|^2\right).$$

Since $\mathcal{A}_{\text{verify}}(z;G,\eta) \left|\frac{z}{2}\right. 0 \leqslant z \leqslant 1$, it follows that $\|(\zeta-\tau)\odot z\|^2 \leqslant \|\zeta-\tau\|^2$ and $\|\tau\odot z\|^2 \leqslant \|\tau\|^2$. Moreover, since ζ and τ are Boolean, we have $\|\tau\|^2 + \|\zeta-\tau\|^2 = \|\zeta\|^2 \leqslant \frac{n}{k}$. Hence, we can obtain

$$\mathcal{A}_{\text{verify}}(z;G,\eta) \left|\frac{z}{2}\right. \left\langle G_s - \frac{d}{n}J_s, \left(2(\zeta-\tau)\tau^\top\right)\odot zz^\top \right\rangle \leqslant \frac{\chi\sqrt{d}n}{\sqrt{k^3}} . \tag{F.7}$$

Term 3. For the third term $\left\langle \left(G_s - \frac{d}{n}J_s\right)\odot zz^\top, \mathbf{1}\mathbf{1}^\top - \zeta\zeta^\top\right\rangle$, we can further decompose it into

$$\left\langle \left(G_s - \frac{d}{n}J_s\right)\odot zz^\top, \mathbf{1}\mathbf{1}^\top - \zeta\zeta^\top\right\rangle = \left\langle \left(G_s - \frac{\alpha}{n}J_s\right)\odot zz^\top, \mathbf{1}\mathbf{1}^\top - \zeta\zeta^\top\right\rangle + \left\langle \left(\frac{\alpha-d}{n}J_s\right)\odot zz^\top, \mathbf{1}\mathbf{1}^\top - \zeta\zeta^\top\right\rangle . \tag{F.8}$$

Since $\mathcal{A}_{\text{verify}}$ certifies $\left\|\left(G_s - \frac{\alpha}{n}J_s\right)\odot zz^\top\right\|_{\mathrm{op}} \leqslant \chi\sqrt{\alpha/k}$ and $\|\zeta\|_1 \geqslant (1-\eta-\rho)\frac{n}{k}$, we can apply Frobenius Cauchy–Schwarz and get

$$\mathcal{A}_{\text{verify}}(z;G,\eta) \left|\frac{z}{2}\right. \left\langle \left(G_s - \frac{\alpha}{n}J_s\right)\odot zz^\top, \mathbf{1}\mathbf{1}^\top - \zeta\zeta^\top\right\rangle \leqslant \frac{2\chi\sqrt{\eta+\rho}\,\sqrt{\alpha}n}{\sqrt{k^3}} .$$

Since $\alpha \leqslant 2d$, $\rho \ll 1$ (due to Theorem 58) and the assumption that $\eta \ll 1$, it follows that

$$\mathcal{A}_{\text{verify}}(z;G,\eta) \left|\frac{z}{2}\right. \left\langle \left(G_s - \frac{\alpha}{n}J_s\right)\odot zz^\top, \mathbf{1}\mathbf{1}^\top - \zeta\zeta^\top\right\rangle \leqslant \frac{\chi\sqrt{d}n}{\sqrt{k^3}} . \tag{F.9}$$

Since $\mathcal{A}_{\text{verify}}(z;G,\eta) \left|\frac{z}{2}\right. 0 \leqslant z \leqslant 1$ and $\alpha = d + (1-\frac{1}{k})\varepsilon d$, it follows that

$$\mathcal{A}_{\text{verify}}(z;G,\eta) \left|\frac{z}{2}\right. \left\langle \left(\frac{\alpha-d}{n}J_s\right)\odot zz^\top, \mathbf{1}\mathbf{1}^\top - \zeta\zeta^\top\right\rangle \leqslant \frac{2(k-1)(\eta+\rho)\varepsilon dn}{k^3} .$$

Since $\rho \ll 1$ (due to Theorem 58) and the assumption that $\eta \ll 1$, it follows that

$$\mathcal{A}_{\text{verify}}(z;G,\eta) \left|\frac{z}{2}\right. \left\langle \left(\frac{\alpha-d}{n}J_s\right)\odot zz^\top, \mathbf{1}\mathbf{1}^\top - \zeta\zeta^\top\right\rangle \leqslant \frac{0.001(k-1)\varepsilon dn}{k^3} . \tag{F.10}$$

Plugging Eq. (F.9) and Eq. (F.10) into Eq. (F.8), it follows that

$$\mathcal{A}_{\text{verify}}(z;G,\eta) \left|\frac{z}{2}\right. \left\langle \left(G_s - \frac{d}{n}J_s\right)\odot zz^\top, \mathbf{1}\mathbf{1}^\top - \zeta\zeta^\top\right\rangle \leqslant \frac{0.001(k-1)\varepsilon dn}{k^3} + \frac{\chi\sqrt{d}n}{\sqrt{k^3}} . \tag{F.11}$$

Conclusion. Plugging Eq. (F.6), Eq. (F.7) and Eq. (F.11) into Eq. (F.2), we can obtain

$$\mathcal{A}_{\text{verify}}(z;G,\eta)\,\Big|\tfrac{z}{2}\,\left\langle \left(G_s - \frac{d}{n}J_s\right)\odot zz^\top, \mathbf{1}\mathbf{1}^\top \right\rangle \leqslant \frac{0.962(k-1)\varepsilon dn}{k^3} + \frac{4\chi\sqrt{d}n}{\sqrt{k^3}}\,.$$

Since $k^2 \ll \varepsilon^2 d$, we can obtain

$$\frac{4\chi\sqrt{d}n}{\sqrt{k^3}} \leqslant \frac{0.00001}{\sqrt{k}}\cdot\frac{\varepsilon dn}{k^2} \leqslant \frac{0.001(k-1)\varepsilon dn}{k^3}\,.$$

Thus, it follows that

$$\mathcal{A}_{\text{verify}}(z;G,\eta)\,\Big|\tfrac{z}{2}\,\left\langle \left(G_s - \frac{d}{n}J_s\right)\odot zz^\top, \mathbf{1}\mathbf{1}^\top \right\rangle \leqslant \frac{0.963(k-1)\varepsilon dn}{k^3}\,.$$

This is a refutation to constraint $\langle (G_s - \frac{d}{n}J_s)\odot zz^\top, \mathbf{1}\mathbf{1}^\top\rangle \geqslant \frac{0.97(k-1)\varepsilon dn}{k^3}$ in $\mathcal{A}_{\text{verify}}(z;G,\eta)$, which finishes the proof. ∎

Proof [Proof of Theorem 20] By Theorem 27 and Theorem 28, it follows that, given a set of n/k vertices in the graph sampled from $\mathsf{SBM}_n(d,\varepsilon,k)$, we can check feasibility of $\mathcal{A}_{\text{verify}}(z;G,\eta)$ (interpreted with G_s, J_s as above) in time $n^{O(1)}$ to decide whether the set of vertices is in Case I or Case II:

- Case I (at least 0.99 of the vertices in the set belong to the same community): feasibility follows by Theorem 27. Since the SoS SDP has small bit complexity and there exists a Boolean solution, by arguments analogous to those in Raghavendra and Weitz (2017), the ellipsoid method will be able to find a feasible solution in time $n^{O(1)}$.
- Case II (no more than 0.98 of the vertices in the set belong to any single community): By Theorem 28, the SoS program is infeasible; hence no feasible solution exists (and, in particular, none can be found in time $n^{O(1)}$).

The success probability is $1 - n^{-\Omega(1)}$. ∎

Appendix G. Robust boosting algorithm for bisection

In this section, we prove Theorem 21 for the robust bisection algorithm in the k-stochastic block model. The procedure essentially constitutes recovery in the 2-SBM model by aggregating communities into two global groups, each comprising $k/2$ local communities.

G.1. Algorithm and constraint systems

We first introduce the polynomial constraints that we use in the algorithm.

(i) Labeling constraints. The feasible label variable $Z \in \{0,1\}^{n\times k}$ satisfies

$$\mathcal{A}_{\text{label}}(Z) := \left\{ Z(i,a)^2 = Z(i,a),\ \sum_a Z(i,a) = 1,\ \sum_i Z(i,a) = \tfrac{n}{k} \right\}. \qquad \text{(G.1)}$$

(ii) Initialization. Let $Z_{\text{init}} \in \mathbb{R}^{n\times k}$ be an initializer 0.001-approximation error and recovers 0.99 fraction of the vertices in communities $a \in [k/2]$, we add constraints to enforce the SoS variable Z is close to Z_{init},

$$\mathcal{A}_{\text{init}}(Z; Z_{\text{init}}) := \begin{Bmatrix} \|Z(\cdot,a) - Z_{\text{init}}(\cdot,a)\|^2 \leqslant 0.001n/k \quad \forall a \in [k/2] \\ \left\| \sum_{a\in[k/2]} Z(\cdot,a) - \sum_{a\in[k/2]} Z_{\text{init}}(\cdot,a) \right\|^2 \leqslant 0.001n \end{Bmatrix}. \tag{G.2}$$

(iii) Corruption mask / node distance. We model node corruptions through a binary mask $\xi \in \{0,1\}^n$ and require the program matrix to agree with the observed graph outside the corrupted rows/columns:

$$\mathcal{A}_{\text{close}}(Y, \xi; \eta, \bar{G}) := \left\{ (Y - \bar{G}) \odot (\mathbf{1} - \xi)(\mathbf{1} - \xi)^\top = 0, \xi \odot \xi = \xi, \sum_i \xi_i \leqslant \delta_\eta n \right\}, \tag{G.3}$$

where $\delta_\eta = \exp(-2C_{d,\varepsilon}) + \eta$.

(iv) Community structure and spectral condition. We encode the decomposition of the centered adjacency into pairwise components and bound the spectral norm of the noise blocks:

$$\mathcal{A}_{\text{mix}}(Y, Z) := \left\{ X = ZZ^\top - \frac{1}{k}J, \left\| Y - \frac{\varepsilon d}{n} X \right\|_{\text{op}} \leqslant \left(\chi + \frac{1}{k} \right) \sqrt{d} \right\}. \tag{G.4}$$

(v) Majority-vote consistency. The final constraint system characterizes the constraints for the consistency of bisection majority voting, i.e. enforcing that the induced bisection of X (bisection of vertices in community $[\frac{k}{2}]$ and $(\frac{k}{2}, k]$) is consistent with the concentration result for majority voting:

$$\mathcal{A}_{\text{maj}}(Y, Z, R; \gamma, \tilde{C}) := \left\{ R\colon \mathcal{A}_{\text{set}}(z) \Big|\frac{u}{2}\, \langle Yx(Z), x(Z) \odot z \rangle \geqslant \alpha_\gamma \left(\sum_i z_i - \beta_{\gamma,\tilde{C}} n \right) \right\}, \tag{G.5}$$

where $\alpha_\gamma = \frac{(1-\gamma)\varepsilon d}{16k}$, $\beta_{\gamma,\tilde{C}} = \frac{640k}{1-\gamma} \exp(-\gamma \tilde{C}/2)$, $\mathcal{A}_{\text{set}}(z) := \{z \odot z = z\}$, and $x(Z) := 2 \sum_{a\in[k/2]} Z(\cdot,a) - \mathbf{1}$ [10].

For our algorithm, we define polynomial system $\mathcal{A}(Y, Z, \xi, R; \bar{G}, Z_{\text{init}}, \eta, \gamma, \tilde{C})$ based on the set of constraints defined above [11]

$$\begin{aligned} &\mathcal{A}(Y, Z, \xi, R; \bar{G}, Z_{\text{init}}, \eta, \gamma, \tilde{C}) \\ &:= \mathcal{A}_{\text{label}}(Z) \cup \mathcal{A}_{\text{init}}(Z; Z_{\text{init}}) \cup \mathcal{A}_{\text{close}}(Y, \xi; \eta, \bar{G}) \cup \mathcal{A}_{\text{mix}}(Y, Z, \xi) \cup \mathcal{A}_{\text{maj}}(Y, Z, R; \gamma, \tilde{C}) \,. \end{aligned} \tag{G.6}$$

10. We write x for simplicity when Z is clear from context

11. For simplicity, we use $\mathcal{A}(Y, Z, \xi)$ to denote $\mathcal{A}(Y, Z, \xi, R; \bar{G}, Z_{\text{init}}, \eta, \gamma, \tilde{C})$ when the input is clear from the context.

Algorithm. Now we describe our algorithm in Algorithm 29.

Algorithm 29 (Robust bisection boosting algorithm)
Input: *The graph G sampled from the k-stochastic block model $\mathsf{SBM}_n(d,\varepsilon,k)$ with ηn corrupted nodes, and initial clusters $S_1, S_2, \ldots, S_{k/2} \subseteq [n]$ with its corresponding label matrix Z_{init}.*

1. *Finding degree-$O(1)$ pseudo-distribution satisfying the polynomial constraints $\mathcal{A}(Y,Z,\xi,R;\bar{G},Z_{\text{init}},\eta,\gamma,\tilde{C})$ where $\bar{G} = G - \frac{d}{n}J$ is the centered adjacency matrix, $\gamma = \frac{1}{2}$ and $\tilde{C} = \left(\sqrt{a^{\frac{2}{k}}b^{1-\frac{2}{k}}} - \sqrt{b}\right)^2$.*
2. *Rounding the pseudo-distribution by sign to obtain the bisection $\hat{x} \in \{\pm 1\}^n$ from $x(Z) = 2\sum_{a\in[k/2]} Z(\cdot,a) - \mathbf{1}$.*

G.2. Feasibility and time complexity

We first show that the program is feasible with high probability and runs in polynomial time.

Lemma 30 *Let $\bar{G}^\circ$ be the centered adjacency matrix of the uncorrupted graph and S be the set of nodes with degree larger than $20d$. Let T be the set of corrupted nodes. Let Z° be the ground truth label matrix. Under the setting of Algorithm 29, program $\mathcal{A}(Y,Z,\xi)$ is satisfied by $(\bar{G}^\circ \odot (\mathbf{1}_{\bar{S}}\mathbf{1}_{\bar{S}}^\top), Z^\circ, \mathbf{1}_{S\cup T})$ with probability at least $1 - n^{-\Omega(1)} - \exp(-\Omega(k))$.*

Proof Note that $\mathcal{A}_{\text{label}}(Z^\circ)$ and $\mathcal{A}_{\text{init}}(Z;Z_{\text{init}})$ is satisfied by definition. Let $X^\circ = Z^\circ(Z^\circ)^\top - \frac{1}{k}J$. By Theorem 58, the size of set S is bounded by $\exp(-2C_{d,\varepsilon})\,n$ and the spectral norm of $\bar{G}^\circ - \frac{\varepsilon d}{n}X^\circ$ is bounded by $\chi\sqrt{d}$ with high probability. Notice that

$$\begin{aligned}
\left\|\bar{G}^\circ \odot (\mathbf{1}_{\bar{S}}\mathbf{1}_{\bar{S}}^\top) - \frac{\varepsilon d}{n}X^\circ\right\|_{\text{op}} &\leqslant \left\|\left(\bar{G}^\circ - \frac{\varepsilon d}{n}X^\circ\right)\odot(\mathbf{1}_{\bar{S}}\mathbf{1}_{\bar{S}}^\top)\right\|_{\text{op}} + \frac{\varepsilon d}{n}\left\|X^\circ \odot (J - \mathbf{1}_{\bar{S}}\mathbf{1}_{\bar{S}}^\top)\right\|_{\text{op}} \\
&\leqslant \chi\sqrt{d} + \frac{\varepsilon d}{n}\left\|X^\circ \odot (J - \mathbf{1}_{\bar{S}}\mathbf{1}_{\bar{S}}^\top)\right\|_{\text{F}} \\
&\leqslant \chi\sqrt{d} + \frac{2\varepsilon d}{n}\left\|X^\circ \odot (\mathbf{1}\mathbf{1}_{S}^\top)\right\|_{\text{F}} \\
&\leqslant \chi\sqrt{d} + \frac{4\varepsilon\sqrt{d}}{\sqrt{k}}\exp(-C_{d,\varepsilon})\sqrt{d} \\
&= \chi\sqrt{d} + 4C_{d,\varepsilon}\exp(-C_{d,\varepsilon})\sqrt{d} \\
&\leqslant \left(\chi + \frac{1}{k}\right)\sqrt{d}\,.
\end{aligned}$$

Therefore, $\mathcal{A}_{\text{mix}}(\bar{G}^\circ \odot (\mathbf{1}_{\bar{S}}\mathbf{1}_{\bar{S}}^\top), Z^\circ)$ is feasible with probability $1 - n^{-O(1)}$. Moreover, $\|\mathbf{1}_{S\cup T}\|_1 \leqslant |S| + |T| \leqslant \delta_\eta n$ with probability $1 - n^{-O(1)}$. Hence, $\mathcal{A}_{\text{close}}(\bar{G}^\circ \odot (\mathbf{1}_{\bar{S}}\mathbf{1}_{\bar{S}}^\top), \mathbf{1}_{S\cup T}; \eta, \bar{G})$ is feasible with probability $1 - n^{-O(1)}$.

By Theorem 15, the voting lower bound in $\mathcal{A}_{\text{maj}}(\bar{G}^\circ \odot (\mathbf{1}_{\bar{S}}\mathbf{1}_{\bar{S}}^\top), Z^\circ, R; \gamma, \tilde{C})$ is feasible for $\gamma = \frac{1}{2}$ and $\tilde{C} = \left(\sqrt{a^{\frac{2}{k}} b^{1-\frac{2}{k}}} - \sqrt{b}\right)^2$ with probability $1 - \exp(-\Omega(k)) - n^{-\Omega(1)}$. The existence of the SoS proof of Theorem 14 follows by Theorem 55 and Theorem 56.

The feasibility proof is completed by a union bound over all failure probabilities. ■

Lemma 31 *Algorithm 29 runs in polynomial time.*

Proof Since the sum-of-squares relaxation is at a constant degree, the resulting semidefinite program involves only polynomially many constraints and can be solved in polynomial time. ■

G.3. SoS guarantees for robust bisection boosting

Now, we show that $\mathcal{A}(Y, Z, \xi)$ boosts the bisection error rate from 0.001 to $1/\operatorname{poly}(k)$. To do this, we consider two feasible solutions $(Y^{(1)}, Z^{(1)}, \xi^{(1)})$ and $(Y^{(2)}, Z^{(2)}, \xi^{(2)})$ of $\mathcal{A}(Y, Z, \xi)$. We will show that the constraints in $\mathcal{A}(Y, Z, \xi)$ allow us to prove that any two good solutions will be close to each other in the bisection sense. Since Z° will also be a good solution, this implies that any feasible solution of $\mathcal{A}(Y, Z, \xi)$ will be close to Z° in the bisection sense.

Throughout this section, we define the following terms:

- Let $x^{(1)} = 2\sum_{a\in[k/2]} Z^{(1)}(\cdot, a) - \mathbf{1}$ and $x^{(2)} = 2\sum_{a\in[k/2]} Z^{(2)}(\cdot, a) - \mathbf{1}$ be the induced bisections of $Z^{(1)}$ and $Z^{(2)}$.
- Let $v = \frac{\mathbf{1} - x^{(1)} \odot x^{(2)}}{2}$ be the set of nodes where the bisections $x^{(1)}$ and $x^{(2)}$ differ.
- Let $s = \mathbf{1} - (\mathbf{1} - \xi^{(1)}) \odot (\mathbf{1} - \xi^{(2)})$ be the set of nodes where $Y^{(1)}$ and $Y^{(2)}$ differ.
- Let $w = v \odot (\mathbf{1} - s)$ be the set of vertices in the shared part of $Y^{(1)}$ and $Y^{(2)}$ such that the bisections $x^{(1)}$ and $x^{(2)}$ differ.
- Let $g = \mathbf{1} - (\mathbf{1} - v) \odot (\mathbf{1} - s)$ be union of v and s, notice that, equivalently $g = w + s$.

To prove the algorithmic guarantee of $\mathcal{A}(Y, Z, \xi)$, we need the following observations on properties of the variables defined above.

Lemma 32 *$x^{(1)}$, $x^{(2)}$ and v satisfies*

$$x^{(1)} \odot v = -x^{(2)} \odot v \quad \textit{and} \quad x^{(1)} \odot (\mathbf{1} - v) = x^{(2)} \odot (\mathbf{1} - v)\,,$$

and,

$$v \odot \frac{\mathbf{1} - x^{(1)}}{2} = v \odot \frac{\mathbf{1} + x^{(2)}}{2} \quad \textit{and} \quad v \odot \frac{\mathbf{1} - x^{(2)}}{2} = v \odot \frac{\mathbf{1} + x^{(1)}}{2}\,.$$

Proof By the definition of $v = \frac{\mathbf{1} - x^{(1)} \odot x^{(2)}}{2}$, it follows that

$$x^{(1)} \odot v = \frac{x^{(1)} - x^{(2)}}{2} = -x^{(2)} \odot v\,,$$

and,

$$x^{(1)} \odot (\mathbf{1} - v) = \frac{x^{(1)} + x^{(2)}}{2} = x^{(2)} \odot (\mathbf{1} - v)\,.$$

Now, we show $v \odot \frac{\mathbf{1} - x^{(1)}}{2} = v \odot \frac{\mathbf{1} + x^{(2)}}{2}$, and the other side follows by symmetry. By plugging in $v = \frac{\mathbf{1} - x^{(1)} \odot x^{(2)}}{2}$, the left hand side is

$$v \odot \frac{\mathbf{1} - x^{(1)}}{2} = \frac{\mathbf{1} - x^{(1)} - x^{(1)} \odot x^{(2)} + x^{(2)}}{4}\,,$$

and the right hand side is

$$v \odot \frac{\mathbf{1} + x^{(2)}}{2} = \frac{\mathbf{1} + x^{(2)} - x^{(1)} \odot x^{(2)} - x^{(1)}}{4}\,.$$

Therefore, we have $v \odot \frac{\mathbf{1} - x^{(1)}}{2} = v \odot \frac{\mathbf{1} + x^{(2)}}{2}$. ■

Lemma 33

$$\mathcal{A}\Big(Y^{(1)}, Z^{(1)}, \xi^{(1)}\Big), \mathcal{A}\Big(Y^{(2)}, Z^{(2)}, \xi^{(2)}\Big)$$

$$\left|\frac{\;Y^{(1)}, Z^{(1)}, \xi^{(1)}, Y^{(2)}, Z^{(2)}, \xi^{(2)}\;}{O(1)}\right. \Big\langle Y^{(1)}(x^{(1)} \odot (\mathbf{1} - g)), x^{(1)} \odot w \Big\rangle = -\Big\langle Y^{(2)}(x^{(2)} \odot (\mathbf{1} - g)), x^{(2)} \odot w \Big\rangle\,.$$

Proof Plugging in the definition of g and w, the left hand side can be written as

$$\begin{aligned}\Big\langle Y^{(1)}(x^{(1)} \odot (\mathbf{1} - g)), x^{(1)} \odot w \Big\rangle &= \Big\langle Y^{(1)}(x^{(1)} \odot (\mathbf{1} - v) \odot (\mathbf{1} - s)), x^{(1)} \odot v \odot (\mathbf{1} - s) \Big\rangle \\ &= \Big\langle \Big(Y^{(1)} \odot (\mathbf{1} - s)(\mathbf{1} - s)^\top\Big)(x^{(1)} \odot (\mathbf{1} - v)), x^{(1)} \odot v \Big\rangle\,.\end{aligned}$$

By $\mathcal{A}_{\text{close}}(Y, \xi; \eta, \bar{G})$, we have

$$Y^{(1)} \odot (\mathbf{1} - s)(\mathbf{1} - s)^\top = \bar{G} \odot (\mathbf{1} - s)(\mathbf{1} - s)^\top = Y^{(2)} \odot (\mathbf{1} - s)(\mathbf{1} - s)^\top\,,$$

and, by Theorem 32, we have

$$x^{(1)} \odot v = -x^{(2)} \odot v \quad \text{and} \quad x^{(1)} \odot (\mathbf{1} - v) = x^{(2)} \odot (\mathbf{1} - v)\,,$$

Thus,

$$\begin{aligned}\Big\langle Y^{(1)}(x^{(1)} \odot (\mathbf{1} - g)), x^{(1)} \odot w \Big\rangle &= -\Big\langle \Big(Y^{(2)} \odot (\mathbf{1} - s)(\mathbf{1} - s)^\top\Big)(x^{(2)} \odot (\mathbf{1} - v)), x^{(2)} \odot v \Big\rangle \\ &= -\Big\langle Y^{(2)}(x^{(2)} \odot (\mathbf{1} - g)), x^{(2)} \odot w \Big\rangle\,.\end{aligned}$$

■

Lemma 34 *The norm of v, s and g satisfies the following SoS inequalities*

$$\mathcal{A}\Big(Y^{(1)}, Z^{(1)}, \xi^{(1)}\Big), \mathcal{A}\Big(Y^{(2)}, Z^{(2)}, \xi^{(2)}\Big) \left|\frac{Z^{(1)},Z^{(2)}}{O(1)}\right. \|v\|^2 \leqslant 0.004n\,,$$

and,

$$\mathcal{A}\Big(Y^{(1)}, Z^{(1)}, \xi^{(1)}\Big), \mathcal{A}\Big(Y^{(2)}, Z^{(2)}, \xi^{(2)}\Big) \left|\frac{\xi^{(1)},\xi^{(2)}}{O(1)}\right. \|s\|^2 \leqslant 2\delta_\eta n\,,$$

and,

$$\mathcal{A}\Big(Y^{(1)}, Z^{(1)}, \xi^{(1)}\Big), \mathcal{A}\Big(Y^{(2)}, Z^{(2)}, \xi^{(2)}\Big) \left|\frac{Z^{(1)},Z^{(2)},\xi^{(1)},\xi^{(2)}}{O(1)}\right. \|w\|^2 \leqslant 0.004n\,,$$

and, for any $b \in [\frac{k}{2}]$ and $t \in \{1,2\}$,

$$\mathcal{A}\Big(Y^{(1)}, Z^{(1)}, \xi^{(1)}\Big), \mathcal{A}\Big(Y^{(2)}, Z^{(2)}, \xi^{(2)}\Big) \left|\frac{Z^{(1)},Z^{(2)},\xi^{(1)},\xi^{(2)}}{O(1)}\right. \Big\langle Z^{(t)}(\cdot,b), w\Big\rangle \leqslant \frac{0.004n}{k}\,.$$

Proof For the bound on v, by constraints in $\mathcal{A}_{\text{init}}(Z; Z_{\text{init}})$, it follows that

$$\begin{aligned}
\|v\|^2 &= \left\|\frac{\mathbf{1} - x^{(1)} \odot x^{(2)}}{2}\right\|^2 = \left\|\frac{x^{(1)} - x^{(2)}}{2}\right\|^2 \\
&= \left\|\sum_{a\in[k/2]} Z^{(1)}(\cdot,a) - \sum_{a\in[k/2]} Z^{(2)}(\cdot,a)\right\|^2 \\
&\leqslant 2\left\|\sum_{a\in[k/2]} Z^{(1)}(\cdot,a) - \sum_{a\in[k/2]} Z_{\text{init}}(\cdot,a)\right\|^2 + 2\left\|\sum_{a\in[k/2]} Z^{(2)}(\cdot,a) - \sum_{a\in[k/2]} Z_{\text{init}}(\cdot,a)\right\|^2 \\
&\leqslant 0.004n\,.
\end{aligned}$$

For the bound on s, by constraints in $\mathcal{A}_{\text{close}}(Y, \xi; \eta, \bar{G})$, it follows that

$$\begin{aligned}
\|s\|^2 = \sum_i s_i &= \sum_i 1 - (1 - \xi_i^{(1)}) \odot (1 - \xi_i^{(2)}) = \sum_i \xi_i^{(1)} + \sum_i \xi_i^{(2)}(1 - \xi_i^{(1)}) \\
&\leqslant \sum_i \xi_i^{(1)} + \sum_i \xi_i^{(2)} \leqslant 2\delta_\eta n\,.
\end{aligned}$$

Now, consider $w = v \odot (\mathbf{1} - s)$. Therefore,

$$\|w\|^2 = \|v \odot (\mathbf{1} - s)\|^2 \leqslant \|v\|^2 \leqslant 0.004n\,.$$

For the last inequality, for $b \in [\frac{k}{2}]$, it follows that

$$\Big\langle Z^{(1)}(\cdot,b), w\Big\rangle = \sum_i Z^{(1)}(i,b) v_i (1 - s_i) \leqslant \sum_i Z^{(1)}(i,b) v_i\,.$$

Since $v = x^{(1)} \odot \Big(\sum_{a\in[k/2]} Z^{(1)}(\cdot,a) - \sum_{a\in[k/2]} Z^{(2)}(\cdot,a)\Big)$ and $x^{(1)} \odot Z^{(1)}(\cdot,b) = Z^{(1)}(\cdot,b)$ for $b \in [\frac{k}{2}]$, we have

$$\sum_i Z^{(1)}(i,b) v_i = \sum_i Z^{(1)}(i,b) x_i^{(1)} \Big(\sum_{a\in[k/2]} Z^{(1)}(i,a) - \sum_{a\in[k/2]} Z^{(2)}(i,a)\Big)$$

$$\begin{aligned}
&= \sum_i Z^{(1)}(i,b) - \sum_{a\in[k/2]} Z^{(2)}(i,a)Z^{(1)}(i,b)\\
&\leqslant \sum_i Z^{(1)}(i,b) - Z^{(2)}(i,b)Z^{(1)}(i,b)\\
&= \sum_i Z^{(1)}(i,b)\big(1 - Z^{(2)}(i,b)\big)\\
&\leqslant \sum_i \big(Z^{(1)}(i,b) - Z^{(2)}(i,b)\big)^2\\
&= \Big\|Z^{(1)}(\cdot,b) - Z^{(2)}(\cdot,b)\Big\|^2\\
&\leqslant 2\Big\|Z^{(1)}(\cdot,b) - Z_{\text{init}}(\cdot,b)\Big\|^2 + 2\Big\|Z^{(2)}(\cdot,b) - Z_{\text{init}}(\cdot,b)\Big\|^2\\
&\leqslant \frac{0.004\,n}{k}\,.
\end{aligned}$$

Thus,

$$\Big\langle Z^{(1)}(\cdot,b), w\Big\rangle \leqslant \sum_i Z^{(1)}(i,b)v_i \leqslant \frac{0.004\,n}{k}\,.$$

■

Lemma 35 *For any $a \in [\frac{k}{2}]$ and $t \in \{1,2\}$,*

$$\mathcal{A}\Big(Y^{(1)}, Z^{(1)}, \xi^{(1)}\Big),\mathcal{A}\Big(Y^{(2)}, Z^{(2)}, \xi^{(2)}\Big) \;\; \Big|\frac{Y^{(1)},Z^{(1)},\xi^{(1)},Y^{(2)},Z^{(2)},\xi^{(2)}}{O(1)} \Big\langle X^{(t)}\Big(x^{(t)}\odot w\Big), x^{(t)}\odot w\odot Z^{(t)}(\cdot,a)\Big\rangle \leqslant \frac{0.008n}{k}\Big\langle w, Z^{(t)}(\cdot,a)\Big\rangle\,,$$

and,

$$\mathcal{A}\Big(Y^{(1)}, Z^{(1)}, \xi^{(1)}\Big),\mathcal{A}\Big(Y^{(2)}, Z^{(2)}, \xi^{(2)}\Big) \;\; \Big|\frac{Y^{(1)},Z^{(1)},\xi^{(1)},Y^{(2)},Z^{(2)},\xi^{(2)}}{O(1)} \Big\langle X^{(t)}\Big(x^{(t)}\odot w\Big), x^{(t)}\odot w\odot Z^{(t)}(\cdot,a)\Big\rangle \geqslant -\frac{0.004n}{k}\Big\langle w, Z^{(t)}(\cdot,a)\Big\rangle\,.$$

Proof Recall that $X^{(t)} = Z^{(t)}(Z^{(t)})^\top - \frac{1}{k}J$. We consider $Z^{(t)}(Z^{(t)})^\top$ and $\frac{1}{k}J$ separately. For the $Z^{(t)}(Z^{(t)})^\top$ part, notice that, since $x_i = 2\sum_{a\in[k/2]} Z(i,a) - 1$, we have $x_i^{(t)} \cdot Z^{(t)}(i,a) = Z^{(t)}(i,a)$ for $a \in [k/2]$. Therefore, it follows that

$$\begin{aligned}
\Big\langle Z^{(t)}(Z^{(t)})^\top\Big(x^{(t)}\odot w\Big), x^{(t)}\odot w\odot Z(\cdot,a)\Big\rangle &= \sum_{b\in[k]}\Big\langle Z^{(t)}(\cdot,b)Z^{(t)}(\cdot,b)^\top\Big(x^{(t)}\odot w\Big), x^{(t)}\odot w\odot Z^{(t)}(\cdot,a)\Big\rangle\\
&= \sum_{b\in[k]} Z^{(t)}(\cdot,b)^\top\Big(x^{(t)}\odot w\Big)Z^{(t)}(\cdot,b)^\top(x^{(t)}\odot w\odot Z^{(t)}(\cdot,a))\\
&= Z^{(t)}(\cdot,a)^\top(x^{(t)}\odot w)Z^{(t)}(\cdot,a)^\top(x^{(t)}\odot w\odot Z^{(t)}(\cdot,a))\\
&= \Big\langle Z^{(t)}(\cdot,a), w\Big\rangle Z^{(t)}(\cdot,a)^\top(w\odot Z^{(t)}(\cdot,a))
\end{aligned}$$

$$\leqslant \frac{0.004n}{k}\Big\langle w, Z^{(t)}(\cdot,a)\Big\rangle\,,$$

where the last inequality is by Theorem 34. Moreover, from the last equality, we can also obtain

$$\Big\langle Z^{(t)}(Z^{(t)})^\top\Big(x^{(t)}\odot w\Big), x^{(t)}\odot w\odot Z(\cdot,a)\Big\rangle \geqslant 0$$

For the $\frac{1}{k}J$ part, we have

$$\begin{aligned}\Big\langle \frac{1}{k}J\Big(x^{(t)}\odot w\Big), x^{(t)}\odot w\odot Z(\cdot,a)\Big\rangle &= \frac{1}{k}\Big\langle \mathbf{1}, x^{(t)}\odot w\Big\rangle\Big\langle \mathbf{1}, x^{(t)}\odot w\odot Z^{(t)}(\cdot,a)\Big\rangle\\ &\leqslant \frac{1}{k}\langle \mathbf{1}, w\rangle\Big\langle w, Z^{(t)}(\cdot,a)\Big\rangle\\ &\leqslant \frac{0.004n}{k}\Big\langle w, Z^{(t)}(\cdot,a)\Big\rangle\,.\end{aligned}$$

Equivalently, we can obtain

$$\Big\langle \frac{1}{k}J\Big(x^{(t)}\odot w\Big), x^{(t)}\odot w\odot Z(\cdot,a)\Big\rangle \geqslant -\frac{0.004n}{k}\Big\langle w, Z^{(t)}(\cdot,a)\Big\rangle\,.$$

Therefore, we have

$$\begin{aligned}\Big\langle X^{(t)}\Big(x^{(t)}\odot w\Big), x^{(t)}\odot w\odot Z^{(t)}(\cdot,a)\Big\rangle &= \Big\langle \Big(Z^{(t)}(Z^{(t)})^\top - \frac{1}{k}J\Big)\Big(x^{(t)}\odot w\Big), x^{(t)}\odot w\odot Z^{(t)}(\cdot,a)\Big\rangle\\ &\leqslant \frac{0.008n}{k}\Big\langle w, Z^{(t)}(\cdot,a)\Big\rangle\,,\end{aligned}$$

and,

$$\Big\langle X^{(t)}\Big(x^{(t)}\odot w\Big), x^{(t)}\odot w\odot Z^{(t)}(\cdot,a)\Big\rangle \geqslant -\frac{0.004n}{k}\Big\langle w, Z^{(t)}(\cdot,a)\Big\rangle\,.$$

■

Corollary 36 *For any $t\in\{1,2\}$,*

$$\mathcal{A}\Big(Y^{(1)},Z^{(1)},\xi^{(1)}\Big),\mathcal{A}\Big(Y^{(2)},Z^{(2)},\xi^{(2)}\Big) \;\Big|\!\frac{Y^{(1)},Z^{(1)},\xi^{(1)},Y^{(2)},Z^{(2)},\xi^{(2)}}{O(1)}\; \Big\langle X^{(t)}\Big(x^{(t)}\odot w\Big), x^{(t)}\odot w\odot \frac{x^{(t)}+\mathbf{1}}{2}\Big\rangle \leqslant \frac{0.008n}{k}\|w\|^2\,,$$

and,

$$\mathcal{A}\Big(Y^{(1)},Z^{(1)},\xi^{(1)}\Big),\mathcal{A}\Big(Y^{(2)},Z^{(2)},\xi^{(2)}\Big) \;\Big|\!\frac{Y^{(1)},Z^{(1)},\xi^{(1)},Y^{(2)},Z^{(2)},\xi^{(2)}}{O(1)}\; \Big\langle X^{(t)}\Big(x^{(t)}\odot w\Big), x^{(t)}\odot w\odot \frac{x^{(t)}+\mathbf{1}}{2}\Big\rangle \geqslant -\frac{0.004n}{k}\|w\|^2\,.$$

Proof Recall that $x^{(t)} = 2\sum_{a\in[k/2]} Z^{(t)}(\cdot,a) - \mathbf{1}$. Therefore,

$$\left\langle X^{(t)}\Big(x^{(t)}\odot w\Big), x^{(t)}\odot w\odot \frac{(x^{(t)}+\mathbf{1})}{2}\right\rangle = \left\langle X^{(t)}\Big(x^{(t)}\odot w\Big), x^{(t)}\odot w\odot \sum_{a\in[k/2]} Z^{(t)}(\cdot,a)\right\rangle$$
$$= \sum_{a\in[k/2]}\Big\langle X^{(t)}\Big(x^{(t)}\odot w\Big), x^{(t)}\odot w\odot Z^{(t)}(\cdot,a)\Big\rangle\,.$$

Applying Theorem 35, we can get

$$\begin{aligned}\left\langle X^{(t)}\Big(x^{(t)}\odot w\Big), x^{(t)}\odot w\odot \frac{(x^{(t)}+\mathbf{1})}{2}\right\rangle &\leqslant \sum_{a\in[k/2]}\frac{0.008n}{k}\Big\langle w, Z^{(t)}(\cdot,a)\Big\rangle\\ &= \frac{0.008n}{k}\left\langle w, \frac{x^{(t)}+\mathbf{1}}{2}\right\rangle\\ &\leqslant \frac{0.008n}{k}\|w\|^2\,,\end{aligned}$$

and,

$$\begin{aligned}\left\langle X^{(t)}\Big(x^{(t)}\odot w\Big), x^{(t)}\odot w\odot \frac{(x^{(t)}+\mathbf{1})}{2}\right\rangle &\geqslant -\sum_{a\in[k/2]}\frac{0.004n}{k}\Big\langle w, Z^{(t)}(\cdot,a)\Big\rangle\\ &= -\frac{0.004n}{k}\left\langle w, \frac{x^{(t)}+\mathbf{1}}{2}\right\rangle\\ &\geqslant -\frac{0.004n}{k}\|w\|^2\,.\end{aligned}$$

■

The following bound on $\big\langle Y^{(t)}\big(x^{(t)}\odot g\big), x^{(t)}\odot w\big\rangle$ will be the key thing that we need in the proof of the algorithmic guarantees.

Lemma 37 *For $t\in\{1,2\}$,*

$$\mathcal{A}\Big(Y^{(1)},Z^{(1)},\xi^{(1)}\Big),\mathcal{A}\Big(Y^{(2)},Z^{(2)},\xi^{(2)}\Big)\;\frac{\;\vert^{Y^{(1)},Z^{(1)},\xi^{(1)},Y^{(2)},Z^{(2)},\xi^{(2)}}\;}{O(1)}\;\Big\langle Y^{(t)}\Big(x^{(t)}\odot g\Big), x^{(t)}\odot w\Big\rangle \leqslant \frac{0.016\varepsilon d}{k}\|w\|^2 + 4\varepsilon d\delta_\eta n\,.$$

Proof Without loss of generality, we prove the statement for $t=1$. Let $Y_S^{(1)} := Y^{(1)}\odot(\mathbf{1}-s)(\mathbf{1}-s)^\top$, we use decomposition

$$\Big\langle Y^{(1)}\Big(x^{(1)}\odot g\Big), x^{(1)}\odot w\Big\rangle = \Big\langle \Big(Y^{(1)}-Y_S^{(1)}\Big)\Big(x^{(1)}\odot g\Big), x^{(1)}\odot w\Big\rangle + \Big\langle Y_S^{(1)}\Big(x^{(1)}\odot g\Big), x^{(1)}\odot w\Big\rangle\,. \tag{G.7}$$

For the term $\Big\langle\Big(Y^{(1)}-Y_S^{(1)}\Big)\big(x^{(1)}\odot g\big), x^{(1)}\odot w\Big\rangle$, since $Y^{(1)} - Y_S^{(1)} = Y^{(1)}\odot\big(J-(\mathbf{1}-s)(\mathbf{1}-s)^\top\big)$, $g = w+s$, and $w = v\odot(\mathbf{1}-s)$, we have $\left\langle\left(Y^{(1)}-Y_S^{(1)}\right)\left(x^{(1)}\odot w\right),x^{(1)}\odot w\right\rangle = 0$ and $\left(J-(\mathbf{1}-s)(\mathbf{1}-s)^\top\right)\odot\left(x^{(1)}\odot s\right)\left(x^{(1)}\odot w\right)^\top=\left(x^{(1)}\odot s\right)\left(x^{(1)}\odot w\right)^\top$. Therefore,

$$\begin{aligned}\left\langle\left(Y^{(1)}-Y_S^{(1)}\right)\left(x^{(1)}\odot g\right),x^{(1)}\odot w\right\rangle =&\left\langle\left(Y^{(1)}-Y_S^{(1)}\right)\left(x^{(1)}\odot s\right),x^{(1)}\odot w\right\rangle\\ =&\left\langle Y^{(1)}\odot\left(J-(\mathbf{1}-s)(\mathbf{1}-s)^\top\right),\left(x^{(1)}\odot s\right)\left(x^{(1)}\odot w\right)^\top\right\rangle\\ =&\left\langle Y^{(1)},\left(x^{(1)}\odot s\right)\left(x^{(1)}\odot w\right)^\top\right\rangle.\end{aligned}$$

let $Y^{(1)}=\frac{\varepsilon d}{n}X^{(1)}+E^{(1)}$ where $X^{(1)}=Z^{(1)}(Z^{(1)})^\top-\frac{1}{k}J$ and $E^{(1)}=Y^{(1)}-\frac{\varepsilon d}{n}X^{(1)}$, it follows that

$$\begin{aligned}\left\langle\left(Y^{(1)}-Y_S^{(1)}\right)\left(x^{(1)}\odot g\right),x^{(1)}\odot w\right\rangle =&\frac{\varepsilon d}{n}\left\langle Z^{(1)}(Z^{(1)})^\top,\left(x^{(1)}\odot s\right)\left(x^{(1)}\odot w\right)^\top\right\rangle\\ &-\frac{\varepsilon d}{kn}\left\langle J,\left(x^{(1)}\odot s\right)\left(x^{(1)}\odot w\right)^\top\right\rangle\\ &+\left\langle E^{(1)},\left(x^{(1)}\odot s\right)\left(x^{(1)}\odot w\right)^\top\right\rangle.\end{aligned}$$

For the first term, since $\mathbf{0}\leqslant Z^{(1)}(Z^{(1)})^\top\leqslant J$ and $x^{(1)}\leqslant\mathbf{1}$, it follows that

$$\frac{\varepsilon d}{n}\left\langle Z^{(1)}(Z^{(1)})^\top,\left(x^{(1)}\odot s\right)\left(x^{(1)}\odot w\right)^\top\right\rangle\leqslant\frac{\varepsilon d}{n}\left\langle J,sw^\top\right\rangle=\frac{\varepsilon d}{n}\left(\sum_i s_i\right)\left(\sum_i w_i\right)\leqslant 2\varepsilon d\delta_\eta\|w\|^2\,.$$

For the second term, since $x^{(1)}\geqslant-\mathbf{1}$, it follows that

$$-\frac{\varepsilon d}{kn}\left\langle J,\left(x^{(1)}\odot s\right)\left(x^{(1)}\odot w\right)^\top\right\rangle\leqslant\frac{\varepsilon d}{kn}\left\langle J,sw^\top\right\rangle=\frac{\varepsilon d}{kn}\left(\sum_i s_i\right)\left(\sum_i w_i\right)\leqslant\frac{2\varepsilon d\delta_\eta}{k}\|w\|^2\,.$$

For the third term, $\mathcal{A}_{\mathrm{mix}}(Y,Z)$ and SoS Cauchy-Schwarz inequality, it follows that

$$\begin{aligned}\left\langle E^{(1)},\left(x^{(1)}\odot s\right)\left(x^{(1)}\odot w\right)^\top\right\rangle\leqslant&\left(\chi+\frac{1}{k}\right)\sqrt{d}\left\|x^{(1)}\odot s\right\|^2+\frac{1}{(\chi+\frac{1}{k})\sqrt{d}}\left\|E^{(1)}\right\|_{\mathrm{op}}^2\left\|x^{(1)}\odot w\right\|^2\\ \leqslant&\left(\chi+\frac{1}{k}\right)\sqrt{d}\delta_\eta n+\left(\chi+\frac{1}{k}\right)\sqrt{d}\|w\|^2\end{aligned}$$

Since $\varepsilon^2 d\gg k^2$, we have $\left(\chi+\frac{1}{k}\right)\sqrt{d}\ll\frac{\varepsilon d}{k}$, we can obtain

$$\left\langle E^{(1)},\left(x^{(1)}\odot s\right)\left(x^{(1)}\odot w\right)^\top\right\rangle\leqslant\frac{0.0001\varepsilon d\delta_\eta}{k}n+\frac{0.001\varepsilon d}{k}\|w\|^2\,.$$

Thus, it follows that

$$\left\langle\left(Y^{(1)}-Y_S^{(1)}\right)\left(x^{(1)}\odot g\right),x^{(1)}\odot w\right\rangle\leqslant\frac{0.001\varepsilon d}{k}\|w\|^2+4\varepsilon d\delta_\eta n\,.\tag{G.8}$$

Now, we consider the term $\left\langle Y_S^{(1)}\left(x^{(1)} \odot g\right), x^{(1)} \odot w\right\rangle$. By $\mathcal{A}_{\text{close}}(Y, \xi; \eta, \bar{G})$, we have

$$Y^{(1)} \odot (\mathbf{1} - s)(\mathbf{1} - s)^\top = \bar{G} \odot (\mathbf{1} - s)(\mathbf{1} - s)^\top = Y^{(2)} \odot (\mathbf{1} - s)(\mathbf{1} - s)^\top .$$

By Theorem 32 and $w = v \odot (\mathbf{1} - s)$,

$$\frac{\mathbf{1} - x^{(1)}}{2} \odot w = \frac{\mathbf{1} + x^{(2)}}{2} \odot w \quad \text{and} \quad x^{(1)} \odot w = -x^{(2)} \odot w .$$

Therefore, we have

$$\begin{aligned}
&\left\langle Y_S^{(1)}\left(x^{(1)} \odot g\right), x^{(1)} \odot w\right\rangle \\
&= \left\langle Y_S^{(1)}\left(x^{(1)} \odot g\right), x^{(1)} \odot w \odot \frac{1 + x^{(1)}}{2}\right\rangle + \left\langle Y_S^{(1)}\left(x^{(1)} \odot g\right), x^{(1)} \odot w \odot \frac{1 - x^{(1)}}{2}\right\rangle \\
&= \left\langle Y_S^{(1)}\left(x^{(1)} \odot g\right), x^{(1)} \odot w \odot \frac{1 + x^{(1)}}{2}\right\rangle - \left\langle Y_S^{(2)}\left(x^{(2)} \odot g\right), x^{(2)} \odot w \odot \frac{1 + x^{(2)}}{2}\right\rangle \\
&= \left\langle Y^{(1)}\left(x^{(1)} \odot w\right), x^{(1)} \odot w \odot \frac{1 + x^{(1)}}{2}\right\rangle - \left\langle Y^{(2)}\left(x^{(2)} \odot w\right), x^{(2)} \odot w \odot \frac{1 + x^{(2)}}{2}\right\rangle .
\end{aligned}$$

Consider the first term, it follows that

$$\begin{aligned}
\left\langle Y^{(1)}\left(x^{(1)} \odot w\right), x^{(1)} \odot w \odot \frac{1 + x^{(1)}}{2}\right\rangle =& \frac{\varepsilon d}{n}\left\langle X^{(1)}\left(x^{(1)} \odot w\right), x^{(1)} \odot w \odot \frac{1 + x^{(1)}}{2}\right\rangle \\
&+ \left\langle E^{(1)}\left(x^{(1)} \odot w\right), x^{(1)} \odot w \odot \frac{1 + x^{(1)}}{2}\right\rangle .
\end{aligned}$$

By Theorem 36, we have

$$\frac{\varepsilon d}{n}\left\langle X^{(1)}\left(x^{(1)} \odot w\right), x^{(1)} \odot w \odot \frac{1 + x^{(1)}}{2}\right\rangle \leqslant \frac{0.008\varepsilon d}{k}\|w\|^2 .$$

By $\mathcal{A}_{\text{mix}}(Y, Z)$ and SoS Cauchy-Schwarz inequality, we have

$$\left\langle E^{(1)}\left(x^{(1)} \odot w\right), x^{(1)} \odot w \odot \frac{1 + x^{(1)}}{2}\right\rangle \leqslant \left(\chi + \frac{1}{k}\right)\sqrt{d}\|w\|^2 .$$

Since $\varepsilon^2 d \gg k^2$, we have $\left(\chi + \frac{1}{k}\right)\sqrt{d} \ll \frac{\varepsilon d}{k}$, and, therefore,

$$\left\langle Y^{(1)}\left(x^{(1)} \odot w\right), x^{(1)} \odot w \odot \frac{1 + x^{(1)}}{2}\right\rangle \leqslant \frac{0.01\varepsilon d}{k}\|w\|^2 .$$

Using similar analysis on the lower bound side, we can obtain

$$\left\langle Y^{(2)}\left(x^{(2)} \odot w\right), x^{(2)} \odot w \odot \frac{1 + x^{(2)}}{2}\right\rangle \geqslant -\frac{0.005\varepsilon d}{k}\|w\|^2 .$$

Thus,

$$\left\langle Y_S^{(1)}\left(x^{(1)} \odot g\right), x^{(1)} \odot w\right\rangle \leqslant \frac{0.015\varepsilon d}{k}\|w\|^2 . \tag{G.9}$$

Plugging Eq. (G.8) and Eq. (G.9) into Eq. (G.7), it follows that

$$\left\langle Y^{(1)}\left(x^{(1)} \odot g\right), x^{(1)} \odot w\right\rangle \leqslant \frac{0.016\varepsilon d}{k}\|w\|^2 + 4\varepsilon d\delta_\eta n .$$

■

Now combining Theorem 33 and Theorem 37, we can prove Theorem 38.

Theorem 38 *Consider the setting of Algorithm 29. We have*

$$\frac{\mathcal{A}\left(Y^{(1)}, Z^{(1)}, \xi^{(1)}\right), \mathcal{A}\left(Y^{(2)}, Z^{(2)}, \xi^{(2)}\right)}{\Big|\frac{Y^{(1)}, Z^{(1)}, \xi^{(1)}, Y^{(2)}, Z^{(2)}, \xi^{(2)}}{O(1)}} \left\|\frac{\mathbf{1} - x^{(1)} \odot x^{(2)}}{2}\right\|^2 \leqslant O\left(\frac{k}{0.96-\gamma}\left(\exp(-\frac{\gamma\tilde{C}}{2}) + \delta_\eta\right) n\right) .$$

Proof By Theorem 33, we can obtain

$$\left\langle Y^{(1)}(x^{(1)} \odot (\mathbf{1}-g)), x^{(1)} \odot w\right\rangle = -\left\langle Y^{(2)}(x^{(2)} \odot (\mathbf{1}-g)), x^{(2)} \odot w\right\rangle .$$

Therefore, we have

$$\left\langle Y^{(1)}(x^{(1)}), x^{(1)} \odot w\right\rangle + \left\langle Y^{(2)}(x^{(2)}), x^{(2)} \odot w\right\rangle = \left\langle Y^{(1)}(x^{(1)} \odot g), x^{(1)} \odot w\right\rangle + \left\langle Y^{(2)}(x^{(2)} \odot g), x^{(2)} \odot w\right\rangle .$$

By $\mathcal{A}_{\text{maj}}(Y, Z, R; \gamma, \tilde{C})$, it follows that

$$\left\langle Y^{(1)}(x^{(1)}), x^{(1)} \odot w\right\rangle + \left\langle Y^{(2)}(x^{(2)}), x^{(2)} \odot w\right\rangle \geqslant 2\alpha_\gamma\left(\|w\|^2 - \beta_{\gamma,\tilde{C}} n\right) .$$

By Theorem 37, we have for each $t \in \{1, 2\}$,

$$\left\langle Y^{(t)}\left(x^{(t)} \odot g\right), x^{(t)} \odot w\right\rangle \leqslant \frac{0.016\varepsilon d}{k}\|w\|^2 + 4\varepsilon d\delta_\eta n .$$

Therefore, we have

$$\alpha_\gamma\left(\|w\|^2 - \beta_{\gamma,\tilde{C}} n\right) \leqslant \frac{0.016\varepsilon d}{k}\|w\|^2 + 4\varepsilon d\delta_\eta n .$$

By rearranging terms and plugging in $\alpha_\gamma = \frac{(1-\gamma)\varepsilon d}{16k}$ and $\beta_{\gamma,\tilde{C}} = \frac{640k}{1-\gamma}\exp(-\gamma\tilde{C}/2)$, we can obtain

$$\|w\|^2 \leqslant O\left(\frac{k}{0.96-\gamma}\left(\exp(-\frac{\gamma\tilde{C}}{2}) + \delta_\eta\right) n\right) .$$

Since $w = v \odot (\mathbf{1} - s)$, it follows that

$$\|v\|^2 = \|w + v \odot s\|^2 \leqslant 2\|w\|^2 + 2\|v \odot s\|^2 \leqslant O\left(\frac{k}{0.96-\gamma}\left(\exp(-\frac{\gamma\tilde{C}}{2}) + \delta_\eta\right) n\right) .$$

■

G.4. Algorithmic guarantees for robust bisection boosting

Now, we are ready to prove Theorem 21.

Theorem *[Restatement of Theorem 21] Let $(G^\circ, Z^\circ) \sim \mathsf{SBM}_n(d, \varepsilon, k)$ be generated from the k-stochastic block model and G be generated by adversarially corrupting η-fraction of the nodes in G°. Assume $k \leqslant n^{0.001}$, $\eta \leqslant \frac{1}{\mathrm{poly}(k)}$, $d = o(n)$, and $\varepsilon^2 d \geqslant Kk^2$ for some sufficiently large constant K. Let $S_1, S_2, \ldots, S_{k/2} \subset V$ be disjoint subsets of size n/k such that in every S_i at least 0.99-fraction of the nodes belong to the same community. Let $x^\circ \in \{\pm 1\}^n$ be the ground-truth community bisection with the underlying communities of $S_1, S_2, \ldots, S_{k/2}$ on the same side. There exists a polynomial-time algorithm that, given observation of G and $S_1, S_2, \ldots, S_{k/2}$, outputs $\hat{x} \in \{\pm 1\}^n$ such that, with probability $1 - \exp(-\Omega(k)) - \frac{1}{\mathrm{poly}(n)}$,*

$$\frac{1}{n}\|\hat{x} - x^\circ\|^2 \leqslant \exp\left(-\big(1 - o(1)\big)\frac{\tilde{C}}{8}\right) + \mathrm{poly}(k)\eta\,,$$

where $\tilde{C} = \left(\sqrt{a^{\frac{2}{k}} b^{1-\frac{2}{k}}} - \sqrt{b}\right)^2$ is the bisection SNR.

Proof Since the program is feasible for the spectrally truncated graph $\bar{G}^\circ \odot (\mathbf{1}_{\bar{S}}\mathbf{1}_{\bar{S}}^\top)$ with ground-truth label Z° with the ground-truth bisection be $x^\circ \in \{\pm 1\}^n$, by Theorem 38, it follows that

$$\frac{1}{n}\left\|\tilde{\mathbb{E}}x - x^\circ\right\|^2 \leqslant O\left(\frac{k}{0.96 - \gamma}\left(\exp\Big(-\frac{\gamma\tilde{C}}{4}\Big) + \delta_\eta\right)\right),$$

where $\tilde{C} = \left(\sqrt{a^{\frac{2}{k}} b^{1-\frac{2}{k}}} - \sqrt{b}\right)^2$. Plug in $\gamma = \frac{1}{2}$ and $\delta_\eta = \exp(-2C_{d,\varepsilon}) + \eta$, it follows that

$$\frac{1}{n}\left\|\tilde{\mathbb{E}}x - x^\circ\right\|^2 \leqslant O\left(k\left(\exp\Big(-\frac{\tilde{C}}{8}\Big) + \eta\right)\right).$$

Let $\hat{x} = \mathrm{sign}(\tilde{\mathbb{E}}x)$. Since the entrywise error of sign rounding increases by at most a multiplicative factor of $O(1)$, it follows that

$$\frac{1}{n}\|\hat{x} - x^\circ\|^2 \leqslant O\left(k\left(\exp\Big(-\frac{\tilde{C}}{8}\Big) + \eta\right)\right) = \exp\left(-\big(1 - o(1)\big)\frac{\tilde{C}}{8}\right) + O(k\eta)\,.$$

■

Appendix H. Robust bisection algorithm

In this section, we combine results from Appendix E, Appendix F and Appendix G to design an algorithm for robust bisectioning and prove Theorem 22.

Algorithm 39 (Robust bisection algorithm)
Input: *A graph G sampled from the k-stochastic block model $\mathsf{SBM}_n(d,\varepsilon,k)$ with ηn corrupted nodes.*

1. ***Graph splitting:*** *We let G_1 be the graph obtained by subsampling each edge in G independently with probability 0.99 and let $G_2 \coloneqq G \setminus G_1$.*
2. ***Rough initialization:*** *Apply the algorithm from Theorem 19 on graph G_1 to obtain a rough k-clustering Z_{rough} with error rate $0.001 + 10^4\eta$.*
3. ***Identifying well recovered blocks:*** *Apply the algorithm from Theorem 20 on graph G_2 and Z_{rough} to identify $k/2$ clusters $S_1, S_2, \ldots, S_{k/2}$ in which 0.99-fraction of the nodes belongs to the same community.*
4. ***Bisection boosting:*** *Apply the algorithm from Theorem 21 on graph G with $S_1, S_2, \ldots, S_{k/2}$ and Z_{rough} as input, find a bisection $\hat{x}$ that separates the identified blocks from the remaining blocks with an error rate $\exp\left(-\left(1-o(1)\right)\frac{C_{d,\varepsilon}}{k^2}\right) + \mathrm{poly}(k)\eta$.*

Output: *$\hat{x}$.*

Theorem *[Restatement of Theorem 22] Let $(G^\circ, Z^\circ) \sim \mathsf{SBM}_n(d,\varepsilon,k)$ be generated from the k-stochastic block model and G be generated by adversarially corrupting η-fraction of the nodes in G°. Assume $k \leqslant n^{0.001}$, $\eta \leqslant \frac{1}{\mathrm{poly}(k)}$, $d = o(n)$, and $\varepsilon^2 d \geqslant Kk^2$ for some sufficiently large constant K. There exists a polynomial-time algorithm that, given observation of G, outputs $\hat{x} \in \{\pm 1\}^n$ such that, with probability $1 - \exp(-\Omega(k)) - \frac{1}{\mathrm{poly}(n)}$,*

$$\frac{1}{n}\|\hat{x} - x^\circ\|^2 \leqslant \exp\left(-\left(1-o(1)\right)\frac{\tilde{C}}{8}\right) + O(k\eta).$$

where $\tilde{C} = \left(\sqrt{a^{\frac{2}{k}} b^{1-\frac{2}{k}}} - \sqrt{b}\right)^2$ is the bisection SNR and $x^\circ \in \{\pm 1\}^n$ is a true community bisection of G°.

Proof For Theorem 21, it suffices to establish the guarantees required by the initialization phase. Specifically, we must obtain $k/2$ disjoint subsets $S_1, S_2, \ldots, S_{k/2}$, each of size exactly n/k, [12] such that, with high probability, at least a 0.99 fraction of the vertices within each subset belong to the same community.

By Theorem 19, we obtain with high probability disjoint subsets $V_1, V_2, \ldots, V_k \subseteq [n]$, such that for at least 0.99 fraction of these subsets, each contains at least 0.999 fraction of its vertices from one ground-truth community, with fewer than $0.001n/k$ vertices from other communities. Without loss of generality, let these well-aligned subsets be $V_1, V_2, \ldots, V_{0.99k}$.

To guarantee that each subset has exactly n/k vertices, we rebalance the subsets by redistributing vertices appropriately. Specifically, surplus vertices are transferred from

12. We assume that k is a power of 2 and n is divisible by k, ensuring that n/k is an integer and eliminating rounding issues in the partitioning process.

subsets exceeding size n/k to subsets with fewer than n/k vertices. Let the resulting balanced subsets be denoted by $S_1, S_2, \dots, S_k$.

The key observation here is that, initially, each of the subsets $V_1, \dots, V_{0.99k}$ contains between $0.999n/k$ and $1.001n/k$ vertices. Consequently, the number of vertices added to or removed from each of these subsets during rebalancing is at most $0.001n/k$. Thus, each subset retains at least $0.998n/k$ vertices from its dominant community. Therefore, for at least 0.99 fraction of the subsets, a 0.998 fraction of their vertices still belongs to a single ground-truth community.

Now, we have $k/2$ disjoint subsets $S_1, S_2, \dots, S_{k/2}$, each of size n/k, satisfying the condition that at least 0.998 fraction of vertices in each subset originate from the same community. By Theorem 20, the algorithm will accept these subsets as valid clusters and reject subsets with recovery rates below 0.99. Combining this with Theorem 21 concludes the proof of the theorem. ∎

Appendix I. Robust initialization for symmetry breaking

Building on the robust optimal bisection algorithm from Appendix H, we present a robust polynomial-time algorithm for clustering vertices into k communities. Our approach attains a misclassification error strictly below the critical $1/k$ symmetry-breaking threshold.

Algorithm 40 (Robust k-clustering via recursive bisections)
Input: *Graph G drawn from the k-stochastic block model $\mathsf{SBM}_{n_i}(d_i, \varepsilon, k_i)$ with corruption fraction at most η_i.*

1. *Apply the robust bisection algorithm (Algorithm 39) on G to obtain bisection $x \in \{\pm 1\}^{n_i}$.*
2. *Let $x^+ := \{i \in [n] : x_i = 1\}$ and $x^- := \{i \in [n] : x_i = -1\}$. Partition G into two induced subgraphs $G_1 = G[x^+]$ and $G_2 = G[x^-]$ such that each distributes as $\mathsf{SBM}_{n_{i+1}}(d_{i+1}, \varepsilon, k_{i+1})$.*
3. *Recursively apply Algorithm 40 to G_1 and G_2 with parameters $n_{i+1}, d_{i+1}, \varepsilon, k_{i+1}$.*
4. *After $\lfloor \log k \rfloor$ levels of recursion, merge the resulting clusters into a final community assignment vector $\hat{Z} \in \{0,1\}^{n \times k}$.*

Output: *$\hat{Z}$.*

Now, we restate and prove the algorithmic guarantees of Algorithm 40.

Theorem *[Restatement of Theorem 23] Let $(G^\circ, Z^\circ) \sim \mathsf{SBM}_n(d, \varepsilon, k)$ be generated from the k-stochastic block model and G be generated by adversarially corrupting η-fraction of the nodes in G°. Assume $k \leqslant n^{0.001}$, $\eta \leqslant \frac{1}{\mathrm{poly}(k)}$, $d = o(n)$, and $\varepsilon^2 d \geqslant K k^2$ for some sufficiently large constant K. There exists a polynomial-time algorithm that, given observation of G, outputs an estimator $\hat{Z} \in \{0,1\}^{n \times k}$ such that, with probability $1 - \exp(-\Omega(k)) - \frac{1}{\mathrm{poly}(n)}$,*

$$\mathit{error}_k(\hat{Z}, Z^\circ) \leqslant \exp\left(-(1 - o(1))\,\frac{C_{d,\varepsilon}}{k^2}\right) + \mathrm{poly}(k)\eta.$$

Proof We follow the level-wise parametrization of Appendix M.1: for level $i \in \{1, \dots, \lfloor \log_2 k \rfloor\}$ set $\beta_i := 2^{-i}$, $n_i := 2\beta_i n$, and $k_i := \beta_i k$. There are 2^i subproblems at level i, each on n_i vertices and k_i communities. Let $\widetilde{C}_i$ be the level-i bisection SNR and err_i the (per-vertex) bisection error of Algorithm 18 on a level-i subproblem.

The original instance has at most ηn corrupted vertices. Restricting to any level-i vertex set V_i of size n_i gives at most ηn corruptions inside V_i, so the effective corruption fraction satisfies

$$\eta_i \leqslant \frac{\eta n}{n_i} = \frac{\eta}{2\beta_i} \leqslant \frac{\eta}{\beta_i} \leqslant k\eta,$$

since $\beta_i \geqslant 1/k$ for $i \leqslant \log_2 k$. Applied at level i, Theorem 22 (with k_i communities and corruption η_i) gives, for some absolute $C_0 > 0$,

$$\mathrm{err}_i \leqslant \exp\Big(-\big(1 - o(1)\big)\frac{\widetilde{C}_i}{8}\Big) + C_0 k_i \eta_i.$$

By Theorem 63, $\beta_i \widetilde{C}_i$ is nondecreasing in i and $\beta_1 \widetilde{C}_1 = \Theta(C_{d,\varepsilon}/k^2)$, hence

$$\exp\Big(-\big(1 - o(1)\big)\frac{\widetilde{C}_i}{8}\Big) \leqslant \exp\Big(-\big(1 - o(1)\big)\frac{C_{d,\varepsilon}}{k^2}\Big) \ .$$

Moreover $k_i \eta_i \leqslant (\beta_i k)(\eta/\beta_i) = k\eta$, so

$$\mathrm{err}_i \leqslant \exp\Big(-\big(1 - o(1)\big)\frac{C_{d,\varepsilon}}{k^2}\Big) + C_0 k\eta \quad \text{for all } i. \tag{I.1}$$

At level i there are 2^i disjoint subproblems, each with error at most err_i, so the number of newly misclustered vertices at that level is at most $2^i \mathrm{err}_i n_i$. Since $n_i = 2\beta_i n = 2^{1-i} n$, we have $2^i n_i = 2n$, and the contribution of level i to the global error is at most

$$\frac{2^i \mathrm{err}_i n_i}{n} \leqslant 2\mathrm{err}_i.$$

Summing over $i = 1, \dots, \lfloor \log_2 k \rfloor$ and using (I.1),

$$\mathrm{error}_k(\hat{Z}, Z^\circ) \leqslant \sum_{i=1}^{\lfloor \log_2 k \rfloor} 2\mathrm{err}_i \leqslant O(\log k)\Big(\exp\Big(-\big(1 - o(1)\big)\frac{C_{d,\varepsilon}}{k^2}\Big) + C_0 k\eta\Big).$$

Absorbing the $O(\log k)$ and C_0 factors into $\mathrm{poly}(k)$, we obtain

$$\mathrm{error}_k(\hat{Z}, Z^\circ) \leqslant \exp\Big(-\big(1 - o(1)\big)\frac{C_{d,\varepsilon}}{k^2}\Big) + \mathrm{poly}(k)\,\eta,$$

as claimed. ∎

As a direct consequence of Theorem 23, we obtain the following corollary, capturing the symmetry-breaking improvement:

Corollary 41 *If $d\varepsilon^2 \geqslant Kk^2 \log k$ for some sufficiently large constant K, then Algorithm 40 achieves, with probability at least $1 - \exp(-\Omega(k)) - \frac{1}{\mathrm{poly}(n)}$, a labeling error at most $O\Big(\frac{1}{\mathrm{poly}(k)}\Big)$.*

Appendix J. Robust optimal recovery algorithm

In this section, we build on the guarantees from Appendix I and give a robust polynomial-time algorithm that partitions the vertices into k communities with the optimal error rate. The algorithm is a constant-degree sum-of-squares (SoS) relaxation of polynomial constraints that encode pairwise majority voting. These constraints closely mirror those used for robust bisectioning. The key difference is the symmetry-breaking guarantee from Theorem 41, which ensures that across all feasible solutions the community labels are aligned. This alignment allows us to analyze the majority voting guarantees for each pair of communities.

J.1. Algorithm and constraint systems

We first introduce the polynomial constraints that are used in our robust optimal recovery algorithm.

(i) Labeling constraints. The feasible label variable $Z \in \{0,1\}^{n\times k}$ satisfies

$$\mathcal{A}_{\text{label}}(Z) := \left\{ Z(i,a)^2 = Z(i,a), \quad \sum_a Z(i,a) = 1, \quad \sum_i Z(i,a) = \frac{n}{k} \right\}. \tag{J.1}$$

(ii) Initialization (symmetry breaking). Let $Z_{\text{init}} \in \mathbb{R}^{n\times k}$ be an initializer with μ-approximation error. We constrain the SoS variable Z to remain close to Z_{init}:

$$\mathcal{A}_{\text{sym-break}}(Z; Z_{\text{init}}, \mu) := \left\{ \|Z - Z_{\text{init}}\|_{\mathrm{F}}^2 \leqslant \mu \right\}. \tag{J.2}$$

(iii) Corruption mask / node distance. We model node corruptions through a binary mask $\xi \in \{0,1\}^n$ and require the program matrix to agree with the observed graph outside the corrupted rows/columns:

$$\mathcal{A}_{\text{close}}(Y, \xi; \bar{G}, \eta) := \left\{ (Y - \bar{G}) \odot (\mathbf{1} - \xi)(\mathbf{1} - \xi)^\top = 0,\ \xi \odot \xi = \xi,\ \sum_i \xi_i \leqslant \delta_\eta n \right\}, \tag{J.3}$$

where $\delta_\eta = \exp(-2C_{d,\varepsilon}) + \eta$.

(iv) Community structure and spectral condition. We encode the decomposition of the centered adjacency into pairwise components and bound the spectral norm of the noise blocks:

$$\mathcal{A}_{\text{mix}}(Y, Z) := \left\{ X = ZZ^\top - \frac{1}{k}J, \left\| Y - \frac{\varepsilon d}{n} X \right\|_{\text{op}} \leqslant \left(\chi + \frac{1}{k} \right) \sqrt{d} \right\}. \tag{J.4}$$

(v) Majority-vote consistency. For each pair $a, b \in [k]$ define $x_{ab} = Z(\cdot, a) - Z(\cdot, b)$. We require that the pairwise labels satisfy the majority voting constraints.

$$\mathcal{A}_{\text{maj}}(Y, Z, R; \gamma, \beta) := \left\{ \begin{aligned} & R\colon \mathcal{A}_{\text{set}}(z) \,\Big|\tfrac{u}{2}\, \langle Y x_{ab}, z \odot x_{ab} \rangle \geqslant \frac{(1-\gamma)\varepsilon d}{16k} \left(\|z \odot x_{ab}\|^2 - 640\beta n \right) \\ & x_{ab} = Z(\cdot, a) - Z(\cdot, b) \end{aligned} \right\}, \tag{J.5}$$

where $\mathcal{A}_{\text{set}}(z) := \{z \odot z = z\}$. We will use two parameter settings, corresponding to a two-round majority vote that achieves the optimal error rate [13]

13. Two rounds are needed to sharpen the error from a constant-factor exponent to the optimal exponent.

- $\mu_1 = \frac{n}{\text{poly}(k)}$, $\gamma_1 = 0.99$, $\beta_1 = 1000k \exp\left(-0.99 \frac{C_{d,\varepsilon}}{k}\right)$.
- $\mu_{\text{opt}} = \exp\left(-0.99(1-o(1))\frac{C_{d,\varepsilon}}{k}\right)$, $\gamma_{\text{opt}} = 1 - \frac{10\chi k}{\sqrt{C_{d,\varepsilon}}}$, $\beta_{\text{opt}} = \frac{\sqrt{C_{d,\varepsilon}}}{10\chi} \exp\left(-\left(1 - \frac{10\chi k}{\sqrt{C_{d,\varepsilon}}}\right)\frac{C_{d,\varepsilon}}{k}\right)$.

For our algorithm, we define polynomial system $\mathcal{A}_{\text{opt}}(Y, Z, \xi, R; \bar{G}, Z_{\text{init}}, \eta, \mu, \gamma, \beta)$ based on the set of constraints defined above [14]

$$\begin{aligned}&\mathcal{A}_{\text{opt}}(Y, Z, \xi, R; \bar{G}, Z_{\text{init}}, \eta, \mu, \gamma, \beta) \\ &:= \mathcal{A}_{\text{label}}(Z) \cup \mathcal{A}_{\text{sym-break}}(Z; Z_{\text{init}}, \mu) \cup \mathcal{A}_{\text{close}}(Y, \xi; \bar{G}, \eta) \cup \mathcal{A}_{\text{mix}}(Y, Z) \cup \mathcal{A}_{\text{maj}}(Y, Z, R; \gamma, \beta)\,.\end{aligned} \tag{J.6}$$

Algorithm. The complete robust optimal recovery algorithm for k-SBM is described in Algorithm 42.

Algorithm 42 (Robust optimal recovery algorithm)
Input: *The centered adjacency matrix $\bar{G}$ drawn from $\mathsf{SBM}_n(d, \varepsilon, k)$ with at most ηn corrupted nodes.*

1. *Run Algorithm 40 to obtain an initializer $\hat{Z}_{init}$.*
2. *Find degree-$O(1)$ pseudo-expectation $\tilde{\mathbb{E}}$ satisfying $\mathcal{A}_{\text{opt}}(Y, Z, \xi, R; \bar{G}, Z_{\text{init}}, \eta, \gamma, \beta)$ with $\hat{Z}_{init}$, $\gamma = 0.99$, $\mu = \frac{n}{\text{poly}(k)}$, $\beta = 1000k \exp\left(-0.99\frac{C_{d,\varepsilon}}{k}\right)$.*
3. *Round $\tilde{\mathbb{E}}$ by labelling vertex i with label $\text{argmax}_a \tilde{\mathbb{E}}[Z(i,a)]$ to $\hat{Z}$.*
4. *Find degree-$O(1)$ pseudo-expectation $\tilde{\mathbb{E}}_{opt}$ satisfying $\mathcal{A}_{\text{opt}}(Y, Z, \xi, R; \bar{G}, Z_{\text{init}}, \eta, \gamma, \beta)$ with $\hat{Z}$, $\mu = \exp\left(-0.99(1-o(1))\frac{C_{d,\varepsilon}}{k}\right)$, $\gamma = 1 - \frac{10\chi k}{\sqrt{C_{d,\varepsilon}}}$, $\beta = \frac{\sqrt{C_{d,\varepsilon}}}{10\chi}\exp\left(-\left(1-\frac{10\chi k}{\sqrt{C_{d,\varepsilon}}}\right)\frac{C_{d,\varepsilon}}{k}\right)$.*
5. *Round $\tilde{\mathbb{E}}_{opt}$ by labelling vertex i with label $\text{argmax}_a \tilde{\mathbb{E}}_{opt}[Z(i,a)]$ to Z_{opt}.*

Output: *Z_{opt}.*

J.2. Feasibility and time complexity

We first establish feasibility of $\mathcal{A}_{\text{opt}}(Y, Z, \xi)$ and time complexity of Algorithm 42.

Lemma 43 *Let $\bar{G}^\circ$ be the centered adjacency matrix of the uncorrupted graph and S be the set of nodes with degree larger than $20d$. Let T be the set of corrupted nodes. Let Z° be the ground truth label matrix. Under the conditions of Theorem 24, program $\mathcal{A}_{\text{opt}}(Y, Z, \xi)$ is satisfied by $(\bar{G}^\circ \odot (\mathbf{1}_{\bar{S}}\mathbf{1}_{\bar{S}}^\top), Z^\circ, \mathbf{1}_{S\cup T})$ with probability at least $1 - \exp(-100k) - 1/n^3$.*

14. For simplicity, we use $\mathcal{A}_{\text{opt}}(Y, Z, \xi)$ to denote $\mathcal{A}_{\text{opt}}(Y, Z, \xi, R; \bar{G}, Z_{\text{init}}, \eta, \gamma, \beta)$ when the input is clear from the context.

Proof The proof parallels Theorem 30. For $\mathcal{A}_{\text{mix}}$, feasibility follows from Theorem 17. The mixing constraints $\left\|Y - \frac{\varepsilon d}{n} X\right\|_{\text{op}} \leqslant \left(\chi + \frac{1}{k}\right)\sqrt{d}$ follow from applying Theorem 58 and similar analysis as Theorem 30. The feasibility of initialization constraint follows from Theorem 23. The feasibility of corruption constraint follows from Theorem 58. ■

Lemma 44 *Algorithm 42 runs in polynomial time.*

Proof The SoS relaxation has constant degree and therefore produces an SDP with polynomially many variables and constraints, which can be solved in polynomial time. ■

J.3. SoS guarantees for robust optimal boosting

We now prove $\mathcal{A}_{\text{opt}}(Y, Z, \xi)$ boosts the accuracy to the desired rate. To do this, we consider two feasible solutions $(Y^{(1)}, Z^{(1)}, \xi^{(1)})$ and $(Y^{(2)}, Z^{(2)}, \xi^{(2)})$ of $\mathcal{A}_{\text{opt}}(Y, Z, \xi)$. We will show that the constraints in $\mathcal{A}_{\text{opt}}(Y, Z, \xi)$ allow us to prove that any two good solutions will be within the optimal error scale of each other. Since Z° is also a good solution, this implies that any feasible solution of $\mathcal{A}_{\text{opt}}(Y, Z, \xi)$ achieves the optimal error rate.

Throughout this section, we define the following terms:

- Let $l_{ab}^{(t)} = x_{ab}^{(t)} \odot x_{ab}^{(t)}$ be the support of $x_{ab}^{(t)}$ for $t \in \{1, 2\}$. Notice that, equivalently, $l_{ab}^{(t)} = Z^{(t)}(\cdot, a) + Z^{(t)}(\cdot, b)$.
- Let $l_{ab} = l_{ab}^{(1)} \odot l_{ab}^{(2)}$ be the intersection of $l_{ab}^{(1)}$ and $l_{ab}^{(2)}$.
- Let $v_{ab} = \frac{l_{ab}^{(1)} \odot l_{ab}^{(2)} - x_{ab}^{(1)} \odot x_{ab}^{(2)}}{2}$ be the coordinates where two feasible pairwise labels $x_{ab}^{(1)}$ and $x_{ab}^{(2)}$ differ.
- Let $s = \mathbf{1} - (\mathbf{1} - \xi^{(1)}) \odot (\mathbf{1} - \xi^{(2)})$ be the set of nodes where $Y^{(1)}$ and $Y^{(2)}$ differ.
- Let $w_{ab} = v_{ab} \odot (\mathbf{1} - s)$ be the set of vertices in the shared part of $Y^{(1)}$ and $Y^{(2)}$ such that the bisections $x^{(1)}$ and $x^{(2)}$ differ.
- Let $g_{ab} = \mathbf{1} - (\mathbf{1} - v_{ab}) \odot (\mathbf{1} - s)$ be union of v_{ab} and s, notice that, equivalently $g_{ab} = w_{ab} + s$.

To prove the algorithmic guarantee of $\mathcal{A}_{\text{opt}}(Y, Z, \xi)$, we need the following observations on properties of the variables defined above.

Lemma 45 *For any $a, b \in [k]$, $x_{ab}^{(1)}$, $x_{ab}^{(2)}$ and v_{ab} satisfies*

$$x_{ab}^{(1)} \odot v_{ab} = -x_{ab}^{(2)} \odot v_{ab} \quad \textit{and} \quad x_{ab}^{(1)} \odot (l_{ab} - v_{ab}) = x_{ab}^{(2)} \odot (l_{ab} - v_{ab}),$$

and,

$$v_{ab} \odot \frac{l_{ab} - x_{ab}^{(1)}}{2} = v_{ab} \odot \frac{l_{ab} + x_{ab}^{(2)}}{2} \quad \textit{and} \quad v_{ab} \odot \frac{l_{ab} - x_{ab}^{(2)}}{2} = v_{ab} \odot \frac{l_{ab} + x_{ab}^{(1)}}{2}.$$

Proof The proof follows by similar analysis as Theorem 32. ■

Lemma 46 *The norm of v_{ab}, s and g_{ab} satisfies the following SoS inequalities*

$$\mathcal{A}_{\text{opt}}\left(Y^{(1)}, Z^{(1)}, \xi^{(1)}\right), \mathcal{A}_{\text{opt}}\left(Y^{(2)}, Z^{(2)}, \xi^{(2)}\right) \Big|\frac{\xi^{(1)},\xi^{(2)}}{O(1)} \|s\|^2 \leqslant 2\delta_\eta n\,,$$

and,

$$\mathcal{A}_{\text{opt}}\left(Y^{(1)}, Z^{(1)}, \xi^{(1)}\right), \mathcal{A}_{\text{opt}}\left(Y^{(2)}, Z^{(2)}, \xi^{(2)}\right) \Big|\frac{Z^{(1)},Z^{(2)}}{O(1)} \|v_{ab}\|^2 \leqslant \mu\,,$$

and,

$$\mathcal{A}_{\text{opt}}\left(Y^{(1)}, Z^{(1)}, \xi^{(1)}\right), \mathcal{A}_{\text{opt}}\left(Y^{(2)}, Z^{(2)}, \xi^{(2)}\right) \Big|\frac{Z^{(1)},Z^{(2)},\xi^{(1)},\xi^{(2)}}{O(1)} \|w_{ab}\|^2 \leqslant \mu\,,$$

and,

$$\mathcal{A}_{\text{opt}}\left(Y^{(1)}, Z^{(1)}, \xi^{(1)}\right), \mathcal{A}_{\text{opt}}\left(Y^{(2)}, Z^{(2)}, \xi^{(2)}\right) \Big|\frac{Z^{(1)},Z^{(2)},\xi^{(1)},\xi^{(2)}}{O(1)} \|g_{ab}\|^2 \leqslant \mu + 4\delta_\eta n\,.$$

Proof For the bound on s, by constraints in $\mathcal{A}_{\text{close}}(Y, \xi; \eta, \bar{G})$, it follows that

$$\begin{aligned}\|s\|^2 &= \sum_i s_i = \sum_i 1 - (1 - \xi_i^{(1)}) \odot (1 - \xi_i^{(2)}) = \sum_i \xi_i^{(1)} + \sum_i \xi_i^{(2)}(1 - \xi_i^{(1)}) \\ &\leqslant \sum_i \xi_i^{(1)} + \sum_i \xi_i^{(2)} \leqslant 2\delta_\eta n\,.\end{aligned}$$

For the bound on v_{ab}, by constraints in $\mathcal{A}_{\text{init}}(Z; Z_{\text{init}})$, it follows that

$$\begin{aligned}\|v_{ab}\|^2 =& \left\|\frac{l_{ab}^{(1)} \odot l_{ab}^{(2)} - x_{ab}^{(1)} \odot x_{ab}^{(2)}}{2}\right\|^2 = \left\|\frac{x_{ab}^{(1)} \odot l_{ab}^{(2)} - x_{ab}^{(2)} \odot l_{ab}^{(1)}}{2}\right\|^2 \\ \leqslant& \left\|\frac{x_{ab}^{(1)} - x_{ab}^{(2)}}{2}\right\|^2 \leqslant \frac{1}{2}\left\|Z^{(1)}(\cdot, a) - Z^{(2)}(\cdot, a)\right\|^2 + \frac{1}{2}\left\|Z^{(1)}(\cdot, b) - Z^{(2)}(\cdot, b)\right\|^2 \\ \leqslant& \left\|Z^{(1)}(\cdot, a) - Z_{\text{init}}(\cdot, a)\right\|^2 + \left\|Z^{(2)}(\cdot, a) - Z_{\text{init}}(\cdot, a)\right\|^2 \\ &+ \left\|Z^{(1)}(\cdot, b) - Z_{\text{init}}(\cdot, b)\right\|^2 + \left\|Z^{(2)}(\cdot, b) - Z_{\text{init}}(\cdot, b)\right\|^2 \\ \leqslant& \|Z - Z_{\text{init}}\|_{\text{F}}^2 \leqslant \mu\,.\end{aligned}$$

The bound on $\|w_{ab}\|^2$ follows by

$$\|w_{ab}\|^2 = \|v_{ab} \odot (\mathbf{1} - s)\|^2 \leqslant \|v_{ab}\|^2 \leqslant \mu\,.$$

For the bound on $\|g_{ab}\|^2$, since $g_{ab} = w_{ab} + s$,

$$\|g_{ab}\|^2 = \|w_{ab} + s\|^2 \leqslant 2\|w_{ab}\|^2 + 2\|s\|^2 \leqslant \mu + 4\delta_\eta n\,.$$

■

Fact 47 *Fix $a, b \in [k]$, there is a constant-degree SoS proof of*

$$\left\langle Y^{(1)}\left(x_{ab}^{(1)} \odot (\mathbf{1} - g_{ab})\right), x_{ab}^{(1)} \odot w_{ab}\right\rangle = -\left\langle Y^{(2)}\left(x_{ab}^{(2)} \odot (\mathbf{1} - g_{ab})\right), x_{ab}^{(2)} \odot w_{ab}\right\rangle .$$

Proof Plugging in the definition of g_{ab} and w_{ab}, the left hand side can be written as

$$\begin{aligned}\left\langle Y^{(1)}(x_{ab}^{(1)} \odot (\mathbf{1} - g_{ab})), x_{ab}^{(1)} \odot w_{ab}\right\rangle &= \left\langle Y^{(1)}(x_{ab}^{(1)} \odot (\mathbf{1} - v_{ab}) \odot (\mathbf{1} - s)), x_{ab}^{(1)} \odot v_{ab} \odot (\mathbf{1} - s)\right\rangle \\ &= \left\langle \left(Y^{(1)} \odot (\mathbf{1} - s)(\mathbf{1} - s)^\top\right)\left(x_{ab}^{(1)} \odot (\mathbf{1} - v_{ab})\right), x_{ab}^{(1)} \odot v_{ab}\right\rangle .\end{aligned}$$

By $\mathcal{A}_{\text{close}}(Y, \xi; \eta, \bar{G})$, we have

$$Y^{(1)} \odot (\mathbf{1} - s)(\mathbf{1} - s)^\top = \bar{G} \odot (\mathbf{1} - s)(\mathbf{1} - s)^\top = Y^{(2)} \odot (\mathbf{1} - s)(\mathbf{1} - s)^\top ,$$

and, by Theorem 45, we have

$$x_{ab}^{(1)} \odot v_{ab} = -x_{ab}^{(2)} \odot v_{ab} \quad \text{and} \quad x_{ab}^{(1)} \odot (l_{ab} - v_{ab}) = x_{ab}^{(2)} \odot (l_{ab} - v_{ab}) ,$$

Thus,

$$\begin{aligned}\left\langle Y^{(1)}(x_{ab}^{(1)} \odot (\mathbf{1} - g_{ab})), x_{ab}^{(1)} \odot w_{ab}\right\rangle &= -\left\langle \left(Y^{(2)} \odot (\mathbf{1} - s)(\mathbf{1} - s)^\top\right)(x_{ab}^{(2)} \odot (\mathbf{1} - v_{ab})), x_{ab}^{(2)} \odot v_{ab}\right\rangle \\ &= -\left\langle Y^{(2)}(x_{ab}^{(2)} \odot (\mathbf{1} - g_{ab})), x_{ab}^{(2)} \odot w_{ab}\right\rangle .\end{aligned}$$

■

We next upper bound the contribution of the signal on the set g.

Lemma 48 *In the setting of Theorem 24, for any $\delta \in [0, 0.001]$ and $t \in \{1, 2\}$,*

$$\mathcal{A}_{\text{opt}}\left(Y^{(1)}, Z^{(1)}, \xi^{(1)}\right), \mathcal{A}_{\text{opt}}\left(Y^{(2)}, Z^{(2)}, \xi^{(2)}\right) \Big|\frac{Y^{(1)}, Z^{(1)}, \xi^{(1)}, Y^{(2)}, Z^{(2)}, \xi^{(2)}}{6} \left\langle Y^{(t)}\left(x_{ab}^{(t)} \odot g_{ab}\right), x_{ab}^{(t)} \odot w_{ab}\right\rangle \leqslant \left(\frac{\varepsilon d}{\mathrm{poly}(k)} + 4\varepsilon d \delta_\eta\right)\|w_{ab}\|^2 + \frac{\varepsilon d \delta_\eta}{k} n .$$

Proof For notational simplicity, we ignore the superscript (t) throughout the proof. We decompose the left hand side in the following way

$$\langle Y(x_{ab} \odot g), x_{ab} \odot w_{ab}\rangle = \frac{\varepsilon d}{n}\langle X(x_{ab} \odot g_{ab}), x_{ab} \odot w_{ab}\rangle + \left\langle (Y - \frac{\varepsilon d}{n}X)(x_{ab} \odot g_{ab}), x_{ab} \odot w_{ab}\right\rangle . \tag{J.7}$$

For the first term, since $X = ZZ^\top - \frac{1}{k}J$, we have $-J \leqslant X \leqslant J$ such that

$$\frac{\varepsilon d}{n}\langle X(x_{ab} \odot g_{ab}), x_{ab} \odot w_{ab}\rangle \leqslant \frac{\varepsilon d}{n}\left(\sum_i (g_{ab})_i\right)\left(\sum_i (w_{ab})_i\right) = \frac{\varepsilon d}{n}\|g_{ab}\|^2\|w_{ab}\|^2 .$$

Since $\|g_{ab}\|^2 \leqslant \mu + 4\delta_\eta n$, it follows that

$$\frac{\varepsilon d}{n}\langle X(x_{ab} \odot g_{ab}), x_{ab} \odot w_{ab}\rangle \leqslant \left(\frac{\varepsilon d \mu}{n} + 4\varepsilon d \delta_\eta\right)\|w_{ab}\|^2 . \tag{J.8}$$

For the second term, by $\mathcal{A}_{\text{mix}}(Y,Z)$ and SoS Cauchy-Schwarz, we also have

$$\begin{aligned}\left\langle (Y-\frac{\varepsilon d}{n}X)(x_{ab}\odot g_{ab}), x_{ab}\odot w_{ab}\right\rangle &\leqslant \frac{1}{2}\left(\chi+\frac{1}{k}\right)\sqrt{d}\left(\|x_{ab}\odot g_{ab}\|^2+\|x_{ab}\odot w_{ab}\|^2\right)\\ &\leqslant \frac{1}{2}\left(\chi+\frac{1}{k}\right)\sqrt{d}\left(3\|w_{ab}\|^2+2\|s\|^2\right)\\ &\leqslant \frac{1}{2}\left(\chi+\frac{1}{k}\right)\sqrt{d}\left(3\|w_{ab}\|^2+2\delta_\eta n\right).\end{aligned}$$

Since $\sqrt{d}\ll\frac{\varepsilon d}{k}$, it follows that

$$\left\langle \left(Y-\frac{\varepsilon d}{n}X\right)(x_{ab}\odot g_{ab}), x_{ab}\odot w_{ab}\right\rangle \leqslant \frac{\varepsilon d}{k}\|w_{ab}\|^2+\frac{\varepsilon d\delta_\eta}{k}n\,. \tag{J.9}$$

Plugging Eq. (J.8) and Eq. (J.9) into Eq. (J.7), it follows that

$$\langle Y(x_{ab}\odot g), x_{ab}\odot w_{ab}\rangle \leqslant \left(\frac{\varepsilon d\mu}{n}+4\varepsilon d\delta_\eta\right)\|w_{ab}\|^2+\frac{\varepsilon d\delta_\eta}{k}n\,.$$

■

Now, we are ready to prove the algorithmic guarantee of $\mathcal{A}_{\text{opt}}$.

Theorem 49 *In the setting of Theorem 24,*

$$\mathcal{A}_{\text{opt}}\left(Y^{(1)},Z^{(1)},\xi^{(1)}\right),\mathcal{A}_{\text{opt}}\left(Y^{(2)},Z^{(2)},\xi^{(2)}\right)$$
$$\vdash_{6}^{Y^{(1)},Z^{(1)},\xi^{(1)},Y^{(2)},Z^{(2)},\xi^{(2)}}\sum_{a,b}\|v_{ab}\|^2\leqslant O\left(\frac{(1-\gamma)k^2\beta n+k^2\eta n}{1-\gamma-\mu k/n-\eta k}+k^2\mu n\right).$$

Proof For any pairs of communities $a,b\in[k]$, by Theorem 47,

$$\left\langle Y^{(1)}\,(x^{(1)}_{ab}\odot(\mathbf{1}-g_{ab})),x^{(1)}_{ab}\odot w_{ab}\right\rangle \;=\; -\left\langle Y^{(2)}\,(x^{(2)}_{ab}\odot(\mathbf{1}-g_{ab})),x^{(2)}_{ab}\odot w_{ab}\right\rangle.$$

Therefore,

$$\left\langle Y^{(1)}x^{(1)}_{ab},x^{(1)}_{ab}\odot w_{ab}\right\rangle+\left\langle Y^{(2)}x^{(2)}_{ab},x^{(2)}_{ab}\odot w_{ab}\right\rangle=\left\langle Y^{(1)}(x^{(1)}_{ab}\odot g),x^{(1)}_{ab}\odot w_{ab}\right\rangle+\left\langle Y^{(2)}(x^{(2)}_{ab}\odot g),x^{(2)}_{ab}\odot w_{ab}\right\rangle. \tag{J.10}$$

By $\mathcal{A}_{\text{maj}}(Y,Z,R;\gamma,\beta)$, it follows that

$$\left\langle Y^{(t)}x^{(t)}_{ab},x^{(t)}_{ab}\odot w_{ab}\right\rangle \geqslant \frac{(1-\gamma)\varepsilon d}{16k}\left(\|\odot w_{ab}\|^2-640\beta n\right), \tag{J.11}$$

for $t\in 1,2$. By Theorem 48, we have

$$\left\langle Y^{(t)}\left(x^{(t)}_{ab}\odot g_{ab}\right),x^{(t)}_{ab}\odot w_{ab}\right\rangle \leqslant \left(\frac{\varepsilon d\mu}{n}+4\varepsilon d\delta_\eta\right)\|w_{ab}\|^2+\frac{\varepsilon d\delta_\eta}{k}n\,. \tag{J.12}$$

for $t\in\{1,2\}$.

Therefore, plugging Eq. (J.11) and Eq. (J.12) into Eq. (J.10) and rearranging terms, we can obtain

$$\|w_{ab}\|^2 \leqslant \frac{(1-\gamma)\beta n + \eta n}{1-\gamma-\mu k/n - \eta k}\,.$$

Since $\|v_{ab}\|^2 \leqslant 2\|w_{ab}\|^2 + 2\|s\|^2$, it follows that

$$\sum_{a,b} \|v_{ab}\|^2 \leqslant O\left(\frac{(1-\gamma)k^2\beta n + k^2\eta n}{1-\gamma-\mu k/n-\eta k} + k^2\mu n\right).$$

■

J.4. Algorithmic guarantees for robust optimal recovery

In this section, we prove our main theorem and establish the algorithmic guarantees of our robust optimal recovery algorithm for k-SBM.

Theorem *[Restatement of Theorem 24] There exist universal constants $C_0 > C_1 > 0$ and $K > 0$ such that the following holds. Let $(G^\circ, Z^\circ) \sim \mathsf{SBM}_n(d,\varepsilon,k)$ be generated from the k-stochastic block model and G be generated by adversarially corrupting η-fraction of the nodes in G°. Assume $k \leqslant n^{0.001}$, $\eta \leqslant k^{-C_0}$, $d = o(n)$, and $\varepsilon^2 d \geqslant K k^2 \log k$. There exists a polynomial-time algorithm (see Algorithm 42) that, given observation of G, outputs an estimator $\hat{Z} \in \{0,1\}^{n\times k}$ such that, with probability $1-\exp(-\Omega(k)) - \frac{1}{\mathrm{poly}(n)}$,*

$$\mathit{error}_k(\hat{Z}, Z^\circ) \leqslant \exp\left(-(1-o(1))\frac{C_{d,\varepsilon}}{k}\right) + k^{C_1}\eta\,.$$

Here $o(1)$ denotes a quantity tending to zero in the regime $\varepsilon^2 d/k^2 \to \infty$.

Proof Let x°_{ab} be the ground-truth pairwise labels. Notice that Algorithm 40 provides a polynomial-time computable initializer $\hat{Z}_{\text{input}}$ with error at most $n/\operatorname{poly}(k)$ that we can use in step 2 of Algorithm 42. Plugging in $Z_{\text{input}} = \hat{Z}_{\text{input}}$, $\mu = \frac{n}{\mathrm{poly}(k)}$, $\gamma = 0.99$ and $\beta = 1000k\exp\left(-0.99\frac{C_{d,\varepsilon}}{k}\right)$ in Theorem 49 and rearranging terms, we can obtain $\tilde{\mathbb{E}}$ such that

$$\sum_{a,b} \|\tilde{\mathbb{E}} x^\circ_{ab} - x_{ab}\|_2^2 \leqslant \exp\left(-(1-o(1))\frac{0.99C_{d,\varepsilon}}{k}\right)n + \mathrm{poly}(k)\eta n\,.$$

Since the rounding introduces a constant multiplicative error, we can use the rounded solution $\hat{Z}$ in step 3. Repeating Theorem 49 with $Z_{\text{input}} = \hat{Z}$, $\mu = \exp\left(-0.99(1-o(1))\frac{C_{d,\varepsilon}}{k}\right)$, $\gamma = 1 - \frac{10\chi k}{\sqrt{C_{d,\varepsilon}}}$, $\beta = \frac{\sqrt{C_{d,\varepsilon}}}{10\chi}\exp\left(-\left(1-\frac{10\chi k}{\sqrt{C_{d,\varepsilon}}}\right)\frac{C_{d,\varepsilon}}{k}\right)$, we can obtain $\tilde{\mathbb{E}}_{\text{opt}}$ such that

$$\sum_{a,b} \|\tilde{\mathbb{E}}_{\text{opt}} x^\circ_{ab} - x_{ab}\|_2^2 \leqslant \exp\left(-(1-o(1))\frac{C_{d,\varepsilon}}{k}\right)n + \mathrm{poly}(k)\eta n\,.$$

Combine this with the rounding guarantees, we obtain the desired bound. ■

Appendix K. Sum-of-Squares proofs

In this section, we introduce some Sum-of-Squares (SoS) results that are used in our proofs.

K.1. Basic Sum-of-Squares inequalities

We start with a Cauchy-Schwarz inequality for pseudo-distributions.

Fact 50 (Cauchy-Schwarz for pseudo-distributions Barak et al. (2012)) *Let f, g be vector polynomials of degree at most d in indeterminate $x \in \mathbb{R}^n$. Then, for any level $2d$ pseudo-distribution D,*

$$\tilde{\mathbb{E}}_D[\langle f, g\rangle] \leqslant \sqrt{\tilde{\mathbb{E}}_D[\|f\|^2]} \cdot \sqrt{\tilde{\mathbb{E}}_D[\|g\|^2]}\,.$$

We will also repeatedly use the following two SoS versions of Cauchy-Schwarz inequality:

Fact 51 (Sum-of-Squares Cauchy-Schwarz I) *Let $x, y \in \mathbb{R}^d$ be indeterminates. Then,*

$$\Big|\frac{x,y}{4}\ \langle x, y\rangle^2 \leqslant \left(\sum_i x_i^2\right)\left(\sum_i y_i^2\right).$$

Fact 52 (Sum-of-Squares Cauchy-Schwarz II) *Let $x, y \in \mathbb{R}^d$ be indeterminates. Then, for any $C > 0$,*

$$\Big|\frac{x,y}{4}\ \langle x, y\rangle \leqslant \frac{C}{2}\|x\|^2 + \frac{1}{2C}\|y\|^2\,,$$

and,

$$\Big|\frac{x,y}{4}\ \langle x, y\rangle \geqslant -\frac{C}{2}\|x\|^2 - \frac{1}{2C}\|y\|^2\,,$$

We will use the following fact that shows how spectral certificates are captured within the SoS proof system.

Fact 53 (Spectral Certificates) *For any $m \times m$ matrix A,*

$$\Big|\frac{u}{2}\ \langle u, Au\rangle \leqslant \|A\|\|u\|_2^2\,.$$

Combining these facts, we have the following lemma as a corollary:

Lemma 54 *For any constant $L \geqslant 0$, for variables $E \in \mathbb{R}^{n\times n}, v \in \mathbb{R}^n, w \in \mathbb{R}^n$, we have constant degree SoS proof that*

$$L\mathrm{Id} - EE^\top = CC^\top \Big|\frac{E,C,v,w}{4}\ \langle Ev, w\rangle^2 \leqslant L\|v\|^2\|w\|^2\,.$$

Proof We have

$$\begin{aligned}\Big|\frac{E,C,v,w}{4}\ \langle Ev, w\rangle^2 &\leqslant \|v\|^2\|E^\top w\|^2 \\ &= \langle w, EE^\top w\rangle \cdot \|v\|^2\,.\end{aligned}$$

Since $L\mathrm{Id} - EE^\top = CC^\top$, we have $\langle w, EE^\top w\rangle \leqslant L\|w\|^2$. As a result,

$$L\mathrm{Id} - EE^\top = CC^\top \Big|\frac{E,C,v,w}{4}\ \langle Ev, w\rangle^2 \leqslant L\|v\|^2\|w\|^2\,.$$

■

K.2. Sum-of-Squares certificate

Theorem 55 *Let $x, y \in \mathbb{R}^d$ and $p_1(x), p_2(x), \ldots, p_m(x)$ be a set of polynomials $\mathbb{R}^d \to \mathbb{R}$. If there exists y° such that $p_1(x) \geqslant 0, p_2(x) \geqslant 0, \ldots, p_m(x) \geqslant 0 \,\Big|\!\frac{x}{\ell}\, q(x; y^\circ) \geqslant 0$, then the degree-$(\ell+2)$ SoS relaxation of the following program is feasible*

$$\mathcal{A} := \Big\{p_1(x) \geqslant 0, p_2(x) \geqslant 0, \ldots, p_m(x) \geqslant 0, q(x,y) = \mathrm{Tr}\Big(RR^\top \mathrm{Mom}_d(p_1, p_2, \ldots, p_m)\Big)\Big\}\,,$$

where $\mathrm{Mom}_d(p_1, \ldots, p_m)$ is a block-diagonal matrix with entries in $\mathbb{R}[x]$ and (possibly empty) blocks M_S indexed by subsets $S \subseteq [m]$ such that for all pairs of monomials x^α, x^β with $\deg x^\alpha, \deg x^\beta \leqslant \frac{1}{2} \cdot (\ell - \sum_{i \in S} \deg p_i)$,

$$M_S(\alpha, \beta) := \Big(\prod\nolimits_{i \in S} p_i\Big) \cdot x^\alpha \cdot x^\beta\,.$$

We also have the following SoS proof

$$\mathcal{A} \,\Big|\!\frac{x,y,R}{\ell+2}\, q(x,y) \geqslant 0\,.$$

Proof Let $R^\circ \in \mathbb{R}^{d^\ell \times d^\ell}$ be the matrix corresponding to a degree-ℓ SoS proof of $p_1(x) \geqslant 0, p_2(x) \geqslant 0, \ldots, p_m(x) \geqslant 0 \,\Big|\!\frac{x}{\ell}\, q(x; y^\circ) \geqslant 0$, i.e.

$$q(x; y^\circ) = \mathrm{Tr}\Big(R^\circ (R^\circ)^\top \mathrm{Mom}_d(p_1, p_2, \ldots, p_m)\Big)\,,$$

Now, the degree-$(\ell+2)$ SoS relaxation of $\mathcal{A}$ is feasible with $R = R^\circ$ and $y = y^\circ$. The SoS proof follows by

$$\mathcal{A} \,\Big|\!\frac{x,y,R}{\ell+2}\, q(x,y) = \mathrm{Tr}\Big(RR^\top \mathrm{Mom}_d(p_1, p_2, \ldots, p_m)\Big) \geqslant 0\,.$$

■

K.3. Sum-of-Squares subset sum

Next we give an SoS proof for a formulation related to the minimum subset sum.

Lemma 56 *For any vector $v \in \mathbb{R}^n$ such that for any subset $S \subseteq [n]$ we have $\sum_{i \in S} v_i \geqslant \gamma(|S| - \beta)$, there is a SoS proof that*

$$\{z \odot z = z\} \,\Big|\!\frac{z}{2}\, \langle z, v\rangle \geqslant \gamma(\|z\|_1 - \beta)\,.$$

Proof We note that lemma is equivalent of showing

$$\{z \odot z = z\} \,\Big|\!\frac{z}{2}\, \langle z, v - \gamma \mathbf{1}\rangle \geqslant -\gamma \cdot \beta\,.$$

If $v_i \leqslant \gamma$ for all i, then

$$\{z \odot z = z\} \,\Big|\!\frac{z}{2}\, \langle z, v - \gamma \mathbf{1}\rangle \geqslant \langle \mathbf{1}, v - \gamma \mathbf{1}\rangle \geqslant -\gamma \cdot \beta\,.$$

Else, w.l.o.g assume that v is in non-decreasing order and v_k is the last element that is at most γ, i.e. $v_1 \leqslant v_2 \leqslant \cdots \leqslant v_k \leqslant \gamma < v_{k+1} \leqslant v_n$. Now, we have

$$\begin{aligned}\{z \odot z = z\} \Big|\frac{z}{2}\ \langle z, v - \gamma \mathbf{1} \rangle &= \sum_{i=1}^{k} (v_i - \gamma) z_i + \sum_{i=k+1}^{n} (v_i - \gamma) z_i \\ &\geqslant \sum_{i=1}^{k} (v_i - \gamma) \\ &\geqslant -\gamma \cdot \beta \,.\end{aligned}$$

where the second last inequality is because $\{z \odot z = z\} \Big|\frac{z}{2}\ \mathbf{0} \leqslant z \leqslant \mathbf{1}$. ■

Appendix L. Spectral bounds

A well-known challenge in analyzing sparse SBMs is that high-degree nodes can inflate spectral bounds by an additional $\log n$ factor. To establish identifiability in Appendix G, we leverage a classical result from random matrix theory, which demonstrates that pruning nodes with degrees above a certain threshold suffices to achieve the desired spectral bounds.

Theorem 57 (Originally proved in Feige and Ofek (2005), restatement of theorem 6.7 in Liu an *Suppose M is a random symmetric matrix with zero on the diagonal whose entries above the diagonal are independent with the following distribution*

$$M_{ij} = \begin{cases} 1 - p_{ij} & w.p.\ p_{ij} \\ -p_{ij} & w.p.\ 1 - p_{ij} \end{cases}$$

Let σ be a quantity such that $p_{ij} \leqslant \sigma^2$ and M_1 be the matrix obtained from M by zeroing out all the rows and columns having more than $20\sigma^2 n$ positive entries. Then with probability $1 - 1/n^2$, we have $\|M_1\| \leqslant \chi\sigma\sqrt{n}$ for some universal constant χ.

As a consequence of the above, we admit the following spectral bound for the adjacency matrix of a pruned SBM.

Corollary 58 (Corollary 6.8 in Liu and Moitra (2022)) *Let G° be a graph sampled from the k-stochastic block model $\mathsf{SBM}_n(d, \varepsilon, k)$ and Z° be the true community membership matrix. Then, with probability at least $1 - \frac{2}{n^2}$, there exists a subset $S \subseteq [n]$ of size at least $\big(1 - \exp\big(-2C_{d,\varepsilon}\big)\big)n$ such that*

$$\left\| \left(G^\circ - \frac{d}{n} J - \frac{\varepsilon d}{n} X^\circ \right) \odot \mathbf{1}_S \mathbf{1}_S^\top \right\|_{\mathrm{op}} \leqslant \chi\sqrt{d},$$

where χ is a universal constant and $X^\circ = Z^\circ (Z^\circ)^\top - \frac{1}{k} J$ is the signal part of G°.

Appendix M. Concentration inequalities for stochastic block models

M.1. Level-wise SBM statistics.

Let us consider a level $i \in \{1, 2, \dots, \lfloor \log_2 k \rfloor\}$ in the recursive bisection as described in Appendix I. We define the parameters

$$\beta_i \coloneqq 2^{-i}, \qquad n_i \coloneqq 2\beta_i n, \qquad k_i \coloneqq \beta_i k,$$

so that one side of the level-i bisection has $n_i/2$ vertices and contains k_i communities. We let $\alpha_i \coloneqq 1/k$ denote the mass of a single ground-truth community and define the mixing parameter

$$\gamma_i \coloneqq \alpha_i/\beta_i \in [0, 1],$$

as the fraction of the side occupied by the community of a fixed vertex. Using γ_i, we introduce the *geometrically weighted* connection probabilities at level i, so that

$$\widetilde{a}_i \coloneqq a^{\gamma_i} b^{1-\gamma_i} \ , \qquad \widetilde{b}_i \coloneqq b \ .$$

This geometric weighting reflects that a random neighbor on the same side belongs to the relevant community with mass γ_i and to an irrelevant community with mass $1 - \gamma_i$. A simple uniform bound on the mixed probabilities will be used repeatedly through the section.

Lemma 59 *Fix $k \geqslant 2$ and $0 < \varepsilon < 1$. Then, for every level i,*

$$\sqrt{\widetilde{a}_i + \widetilde{b}_i} \leqslant \sqrt{a+b} = \sqrt{d\big(2 + \varepsilon(1 - \frac{2}{k})\big)} \leqslant \sqrt{3d}. \tag{M.1}$$

Proof Since $a \geqslant b$ and $\gamma_i \in [0, 1]$, we have $b^{1-\gamma_i} \leqslant a^{1-\gamma_i}$ and hence $\widetilde{a}_i = a^{\gamma_i} b^{1-\gamma_i} \leqslant a^{\gamma_i} a^{1-\gamma_i} = a$; also $\widetilde{b}_i = b$. Therefore $\widetilde{a}_i + \widetilde{b}_i \leqslant a + b$, which implies the first inequality in (M.1). A direct computation gives

$$a + b = \Big(1 + \big(1 - \frac{1}{k}\big)\varepsilon + 1 - \frac{\varepsilon}{k}\Big)d = \big(2 + \varepsilon(1 - \frac{2}{k})\big)d \leqslant (2 + \varepsilon)d \leqslant 3d.$$

■

To encode the same-side vs. opposite-side separation in a form amenable to multiplicative bounds used throughout the section we consider the per-edge log-odds ratio. For $p, q \in (0, 1)$ set

$$R(p, q) \coloneqq \frac{p(1-q)}{q(1-p)}, \qquad \widetilde{R}_i \coloneqq R\Big(\widetilde{a}_i/n, \widetilde{b}_i/n\Big).$$

The following lemma provides a useful lower bound on $\log \widetilde{R}_i$.

Lemma 60 *Fix $k \geqslant 2$ and $\varepsilon \in (0, 1)$, and let $d = o(n)$. Then, for every level i,*

$$\frac{\gamma_i \, \varepsilon}{2} \leqslant \log \widetilde{R}_i \leqslant \gamma_i \, \varepsilon \log 3 + o(1), \tag{M.2}$$

where $o(1) = o_n(1)$ is uniform in i.

Proof By definition,

$$\log \widetilde{R}_i = \log \frac{\widetilde{a}_i}{\widetilde{b}_i} + \log \frac{1-\widetilde{b}_i/n}{1-\widetilde{a}_i/n}.$$

Lower bound. Since $\widetilde{a}_i/\widetilde{b}_i = (a/b)^{\gamma_i}$ and $\widetilde{a}_i \geqslant \widetilde{b}_i$,

$$\log \widetilde{R}_i \geqslant \gamma_i \log \frac{a}{b}.$$

With $a = (1+(1-\frac{1}{k})\varepsilon)d$, $b = (1-\frac{\varepsilon}{k})d$,

$$\log \frac{a}{b} = \log\bigl(1+(1-\tfrac{1}{k})\varepsilon\bigr) - \log\bigl(1-\tfrac{\varepsilon}{k}\bigr) \geqslant (1-\tfrac{1}{k})\varepsilon - \tfrac{1}{2}(1-\tfrac{1}{k})^2\varepsilon^2 + \tfrac{\varepsilon}{k} \geqslant \varepsilon - \tfrac{\varepsilon^2}{2} \geqslant \tfrac{\varepsilon}{2},$$

hence $\log \widetilde{R}_i \geqslant \gamma_i \varepsilon/2$.

Upper bound. Using $1-\widetilde{b}_i/n \leqslant 1$ and $-\log(1-x) \leqslant x/(1-x)$ for $x \in (0,1)$,

$$\log \frac{1-\widetilde{b}_i/n}{1-\widetilde{a}_i/n} \leqslant -\log\Bigl(1-\frac{\widetilde{a}_i}{n}\Bigr) \leqslant \frac{\widetilde{a}_i}{n-\widetilde{a}_i} \leqslant \frac{a}{n-a},$$

so

$$\log \widetilde{R}_i \leqslant \gamma_i \log \frac{a}{b} + \frac{a}{n-a}.$$

Moreover, for $k \geqslant 2$,

$$\log \frac{a}{b} = \log \frac{1+(1-\frac{1}{k})\varepsilon}{1-\frac{\varepsilon}{k}} \leqslant \log \frac{1+\varepsilon/2}{1-\varepsilon/2} \leqslant \varepsilon \log 3 \ .$$

We, thus, get

$$\log \widetilde{R}_i \leqslant \gamma_i \, \varepsilon \log 3 + \frac{a}{n-a}.$$

Since $a = \Theta(d)$ and $d = o(n)$, we have $a/(n-a) = o(1)$ uniformly in i, yielding the desired upper bound. ■

M.2. Concentration inequalities

We now prove Theorem 63, a bound on the success probability of majority voting for the held-out vertex. Let us start with a short convexity lemma.

Lemma 61 *Let $a, b, n > 0$ and $\gamma \in (0,1]$. Assume $a \leqslant n$ and $b \leqslant n$. Define $f : [0,1] \to \mathbb{R}$ by*

$$f(x) = x - \bigl(x + \frac{n}{a} - 1\bigr)^{\gamma}\bigl(x + \frac{n}{b} - 1\bigr)^{1-\gamma} + \bigl(\frac{n}{a}\bigr)^{\gamma}\bigl(\frac{n}{b}\bigr)^{1-\gamma} - 1.$$

Then $f(x) \geqslant 0$ for all $x \in [0,1]$.

Proof [Proof of Lemma 61] If $\gamma = 1$ then $f(x) \equiv 0$, thus we restrict attention to $0 < \gamma < 1$. We certify nonnegativity by showing that f is convex on $[0,1]$ and minimized at $x = 1$.

Let $u(x) = x + \frac{n}{a} - 1$ and $v(x) = x + \frac{n}{b} - 1$. A direct differentiation gives

$$f'(x) = 1 - \gamma\, u(x)^{\gamma-1}\, v(x)^{1-\gamma} - (1-\gamma)\, u(x)^{\gamma}\, v(x)^{-\gamma}. \tag{M.3}$$

Evaluating (M.3) at $x = 1$ (so $u(1) = \frac{n}{a}$, $v(1) = \frac{n}{b}$) yields

$$f'(1) = 1 - \gamma \left(\frac{a}{b}\right)^{1-\gamma} - (1-\gamma)\left(\frac{b}{a}\right)^{\gamma} \leqslant 0,$$

where we used the weighted AM-GM inequality

$$\gamma a + (1-\gamma)\, b \geqslant a^{\gamma}\, b^{1-\gamma} \quad \iff \quad \gamma \left(\frac{a}{b}\right)^{1-\gamma} + (1-\gamma)\left(\frac{b}{a}\right)^{\gamma} \geqslant 1.$$

Differentiating (M.3) once more and factoring,

$$\begin{aligned} f''(x) &= \gamma(1-\gamma)\, u(x)^{\gamma-2}\, v(x)^{1-\gamma} - 2\,\gamma(1-\gamma)\, u(x)^{\gamma-1}\, v(x)^{-\gamma} + \gamma(1-\gamma)\, u(x)^{\gamma}\, v(x)^{-\gamma-1} \\ &= \gamma(1-\gamma)\, u(x)^{\gamma-2}\, v(x)^{-\gamma-1} \big(v(x) - u(x)\big)^2 \geqslant 0. \end{aligned}$$

Hence f is convex on $[0,1]$. In particular, f' is nondecreasing on the interval. Since $f'(1) \leqslant 0$, we have $f'(x) \leqslant 0$ for all $x \in [0,1]$, i.e., f is decreasing on $[0,1]$. Finally,

$$f(1) = 1 - \left(\frac{n}{a}\right)^{\gamma}\left(\frac{n}{b}\right)^{1-\gamma} + \left(\frac{n}{a}\right)^{\gamma}\left(\frac{n}{b}\right)^{1-\gamma} - 1 = 0,$$

so $f(x) \geqslant f(1) = 0$ for every $x \in [0,1]$, as claimed. ■

Theorem 62 *Assume $a \leqslant n$ and $b \leqslant n$. Fix parameters $\beta \in (0,1]$ and $\alpha \in (0,\beta]$. Consider the distribution*

$$\mathcal{D} = \mathrm{Binom}(\alpha n, a/n) + \mathrm{Binom}\big((\beta-\alpha)n,\, b/n\big) - \mathrm{Binom}(\beta n,\, b/n)\ ,$$

where the three binomials are independent. Let $\gamma := \alpha/\beta$ and define

$$\widetilde{a} := a^{\gamma} b^{1-\gamma}\ , \qquad \widetilde{b} := b\ , \qquad \widetilde{C} := \big(\sqrt{\widetilde{a}} - \sqrt{\widetilde{b}}\big)^2\ , \qquad R(p,q) := \frac{p(1-q)}{q(1-p)}\ .$$

Then for every $\theta \in \mathbb{R}$,

$$\mathop{\mathbb{P}}_{\boldsymbol{X} \sim \mathcal{D}}[\, \boldsymbol{X} \leqslant \theta \,] \leqslant \exp\Big(-\beta\, \widetilde{C} + \frac{\theta}{2} \log R(\widetilde{a}/n,\, \widetilde{b}/n) \Big)\ .$$

Proof [Proof of Theorem 62] Let $t > 0$ to be fixed later. By Markov's inequality,

$$\mathbb{P}[\boldsymbol{X} \leqslant \theta] = \mathbb{P}\big(e^{-t\boldsymbol{X}} \geqslant e^{-t\theta}\big) \leqslant \mathbb{E}[e^{-t\boldsymbol{X}}]\, e^{t\theta}\ .$$

Since the three binomials are independent and $\mathbb{E}[e^{s\,\mathrm{Binom}(m,p)}] = (pe^s + 1 - p)^m$, we have

$$\begin{aligned}\mathbb{E}[e^{-t\boldsymbol{X}}] &= \big(a/n\,e^{-t} + 1 - a/n\big)^{\alpha n}\big(b/n\,e^{-t} + 1 - b/n\big)^{(\beta-\alpha)n} \cdot \big(b/n\,e^{t} + 1 - b/n\big)^{\beta n} \\ &= \Big(\underbrace{\big(a/n\,e^{-t} + 1 - a/n\big)^{\gamma}\big(b/n\,e^{-t} + 1 - b/n\big)^{1-\gamma}}_{\text{apply Lemma 61}} \cdot \big(b/n\,e^{t} + 1 - b/n\big)\Big)^{\beta n} ,\end{aligned}$$

where $\gamma = \alpha/\beta \in [0,1]$. Applying Lemma 61 with $x = e^{-t} \in (0,1]$ gives

$$\big(a/n\,e^{-t} + 1 - a/n\big)^{\gamma}\big(b/n\,e^{-t} + 1 - b/n\big)^{1-\gamma} \leqslant (\widetilde{a}/n)\,e^{-t} + 1 - \widetilde{a}/n,$$

with $\widetilde{a} = a^{\gamma}b^{1-\gamma}$ and $\widetilde{b} = b$. Hence

$$\mathbb{E}[e^{-t\boldsymbol{X}}] \leqslant \Big(\big(\frac{\widetilde{a}}{n}e^{-t} + 1 - \frac{\widetilde{a}}{n}\big)\big(\frac{\widetilde{b}}{n}e^{t} + 1 - \frac{\widetilde{b}}{n}\big)\Big)^{\beta n} .$$

We choose

$$e^t = \sqrt{\frac{(\widetilde{a}/n)(1-\widetilde{b}/n)}{(\widetilde{b}/n)(1-\widetilde{a}/n)}} = \sqrt{R(\widetilde{a}/n,\,\widetilde{b}/n)} .$$

It then follows that

$$\begin{aligned}\mathbb{P}[\boldsymbol{X} \leqslant \theta] &\leqslant \Big(\big(\frac{\widetilde{a}}{n}e^{-t} + 1 - \frac{\widetilde{a}}{n}\big)\big(\frac{\widetilde{b}}{n}e^{t} + 1 - \frac{\widetilde{b}}{n}\big)\Big)^{\beta n} e^{t\theta} \\ &= \Big(\sqrt{\frac{\widetilde{a}\,\widetilde{b}}{n^2}} + \sqrt{\big(1-\frac{\widetilde{a}}{n}\big)\big(1-\frac{\widetilde{b}}{n}\big)}\Big)^{2\beta n} e^{t\theta} \\ &\leqslant \Big(1 - \frac{\big(\sqrt{\widetilde{a}} - \sqrt{\widetilde{b}}\big)^2}{2n}\Big)^{2\beta n} \exp\Big(\frac{\theta}{2}\log R(\widetilde{a}/n, \widetilde{b}/n)\Big) \\ &\leqslant \exp\Big(-\beta\,\widetilde{C} + \frac{\theta}{2}\log R(\widetilde{a}/n, \widetilde{b}/n)\Big) .\end{aligned}$$

■

Theorem 63 *Fix a level $i \in \{1, 2, \ldots, \lfloor \log_2 k \rfloor\}$ and let $\beta_i := 2^{-i}$ (with $\alpha_i = 1/k$). Let $\widetilde{a}, \widetilde{b}, \widetilde{C}$ and $R(\cdot,\cdot)$ be as in Theorem 62, evaluated at this (α, β). Then for every $\theta \in \mathbb{R}$,*

$$\mathbb{P}[\boldsymbol{X} \leqslant \theta] \leqslant \exp\bigg(-\frac{(\log 2)^2}{4} \cdot \frac{d\,\varepsilon^2}{\beta_i\,k^2} + \frac{\theta}{2}\log R(\widetilde{a}/n, \widetilde{b}/n)\bigg).$$

Proof [Proof of Theorem 63] By Theorem 62 (applied with $(\beta, \alpha) = (\beta_i, \alpha_i)$),

$$\mathbb{P}[\boldsymbol{X} \leqslant \theta] \leqslant \exp\Big(-\beta_i\,\widetilde{C}_i + \frac{\theta}{2}\log R(\widetilde{a}/n, \widetilde{b}/n)\Big).$$

It remains to lower bound $\beta_i \widetilde{C}_i$. Starting from the definition,

$$\widetilde{C}_i = \big(\sqrt{\widetilde{a}} - \sqrt{\widetilde{b}}\big)^2$$

$$= b\Big(\Big(\frac{a}{b}\Big)^{\gamma_i/2} - 1\Big)^2,$$

where $\gamma_i = \alpha_i/\beta_i$. Writing $r := a/b > 1$, we have

$$r^u - 1 = e^{u \log r} - 1 \geqslant u \log r \quad \text{for all } u > 0,$$

so with $u = \gamma_i/2$,

$$\Big(\frac{a}{b}\Big)^{\gamma_i/2} - 1 \;\geqslant\; \frac{\gamma_i}{2}\,\log\frac{a}{b}.$$

Hence

$$\begin{aligned}\widetilde{C}_i &\geqslant b\Big(\frac{\gamma_i}{2}\,\log\frac{a}{b}\Big)^2\\ &= b\,\frac{\gamma_i^2}{4}\,\log^2\frac{a}{b}.\end{aligned}$$

Multiplying by β_i and using $\gamma_i = \frac{\alpha_i}{\beta_i}$ and $\alpha_i = \frac{1}{k}$, we get

$$\begin{aligned}\beta_i\,\widetilde{C}_i &\geqslant \beta_i\,\frac{b\,\gamma_i^2}{4}\,\log^2\frac{a}{b}\\ &= \frac{b}{4}\cdot\frac{\alpha_i^2}{\beta_i}\,\log^2\frac{a}{b}\\ &= \frac{b}{4}\cdot\frac{1}{\beta_i\,k^2}\,\log^2\frac{a}{b}.\end{aligned}$$

Recall $a = (1 + (1 - \frac{1}{k})\varepsilon)d$ and $b = (1 - \frac{\varepsilon}{k})d$, so

$$\frac{a}{b} = \frac{1 + (1 - \frac{1}{k})\varepsilon}{1 - \frac{\varepsilon}{k}}, \qquad \frac{b}{d} = 1 - \frac{\varepsilon}{k},$$

and therefore

$$\beta_i\,\widetilde{C}_i \geqslant \frac{d}{4\,\beta_i\,k^2}\Big(1 - \frac{\varepsilon}{k}\Big)\,\log^2\frac{1 + (1 - \frac{1}{k})\varepsilon}{1 - \frac{\varepsilon}{k}}.$$

Define

$$H(\varepsilon, k) := \frac{1 - \frac{\varepsilon}{k}}{\varepsilon^2}\,\log^2\frac{1 + (1 - \frac{1}{k})\varepsilon}{1 - \frac{\varepsilon}{k}}, \qquad \varepsilon \in (0, 1],\; k \geqslant 2.$$

A direct computation shows that $H(\varepsilon, k)$ is decreasing in k, so

$$\inf_{k\geqslant 2} H(\varepsilon, k) = \lim_{k\to\infty} H(\varepsilon, k) = \frac{\log^2(1 + \varepsilon)}{\varepsilon^2}.$$

Moreover, $\log(1 + \varepsilon)/\varepsilon$ is decreasing on $(0, 1]$, hence

$$\frac{\log^2(1 + \varepsilon)}{\varepsilon^2} \geqslant \log^2 2 \quad \text{for all } \varepsilon \in (0, 1].$$

Thus $H(\varepsilon, k) \geqslant \log^2 2$ for all $\varepsilon \in (0,1]$, $k \geqslant 2$, and we obtain

$$\beta_i\, \widetilde{C}_i \geqslant \frac{(\log 2)^2}{4}\, \frac{d\, \varepsilon^2}{\beta_i\, k^2}.$$

Substituting this into the Theorem 62 inequality yields

$$\mathbb{P}[\boldsymbol{X} \leqslant \theta] \leqslant \exp\Big(-\frac{(\log 2)^2}{4}\, \frac{d\, \varepsilon^2}{\beta_i\, k^2} + \frac{\theta}{2} \log R(\widetilde{a}/n, \widetilde{b}/n)\Big),$$

as claimed. ■

M.3. Voting lower bound

In this section, we will leverage the results of Section M.2 to prove the global voting concentration theorems Theorem 14 and Theorem 16.

Theorem 64 (Claim 6.2 in Liu and Moitra (2022)) *Let n, n_1, n_2 be parameters with $n_1, n_2 \leqslant 10^{-6} n$. Let $M \in \mathbb{R}^{n \times n}$ be a random matrix whose entries are independent, have mean 0, variance at most σ^2 for some $\sigma \geqslant 20/\sqrt{n}$, and are almost surely bounded in $[-1, 1]$. Then, with probability at least*

$$1 - \exp\Big(-8\,(n_1 + n_2)\, \log\big(\frac{n}{n_1} + \frac{n}{n_2}\big)\Big),$$

the magnitude of the sum of the entries over any $n_1 \times n_2$ combinatorial subrectangle of M is at most $(n_1 + n_2)\, \sigma \sqrt{n}$.

As a corollary, we get the following theorem which applies for rectangles which are long and thin.

Theorem 65 *Let n, n_1, n_2 be parameters with $n_1 \leqslant 10^{-6} n$ and $n_2 \geqslant 10^{-6} n$. Let $M \in \mathbb{R}^{n \times n}$ be a random matrix whose entries are independent, have mean 0, variance at most σ^2 for some $\sigma \geqslant 20/\sqrt{n}$, and are almost surely bounded in $[-1, 1]$. Then, with probability at least*

$$1 - \exp\Big(-8\,(n_1 + 10^{-6} n)\, \log\big(\tfrac{n}{n_1} + 10^6\big)\Big),$$

the magnitude of the sum of the entries over any $n_1 \times n_2$ combinatorial subrectangle of M is at most $2\,(n_1 + n_2)\, \sigma \sqrt{n}$.

Proof Let us fix a set $S \subseteq [n]$ of rows with $|S| = n_1$ and define

$$v_j := \sum_{i \in S} M_{ij}, \qquad j \in [n].$$

Set $\delta := 10^{-6}$. By Theorem 64 with parameters $(n_1, \delta n)$, with the stated probability we have for all $U \subseteq [n]$ with $|U| = \delta n$ that

$$\Big|\sum_{j \in U} v_j\Big| = \Big|\sum_{(i,j) \in S \times U} M_{ij}\Big| \leqslant (n_1 + \delta n)\, \sigma \sqrt{n}.$$

We now use a simple averaging inequality: for any integers $K \geqslant L$ and any vector w,

$$\sup_{|T|=K} \Big|\sum_{j\in T} w_j\Big| \leqslant \frac{K}{L} \sup_{|U|=L} \Big|\sum_{j\in U} w_j\Big|.$$

Indeed, for any fixed T with $|T| = K$, each $j \in T$ lies in exactly $\binom{K-1}{L-1}$ subsets $U \subseteq T$ of size L, so

$$\sum_{j\in T} w_j = \binom{K-1}{L-1}^{-1} \sum_{\substack{U\subseteq T\\ |U|=L}} \sum_{j\in U} w_j,$$

and taking absolute values yields the desired result since $\binom{K}{L}/\binom{K-1}{L-1} = K/L$.

Applying this with $w = v$, $K = n_2$, and $L = \delta n$ yields, uniformly for all T with $|T| = n_2$,

$$\begin{aligned}\Big|\sum_{(i,j)\in S\times T} M_{ij}\Big| &= \Big|\sum_{j\in T} v_j\Big| \\ &\leqslant \frac{n_2}{\delta n} \sup_{|U|=\delta n} \Big|\sum_{j\in U} v_j\Big| \\ &\leqslant \frac{n_2}{\delta n} (n_1 + \delta n)\, \sigma\sqrt{n} \\ &\leqslant 2n_2\, \sigma\sqrt{n} \\ &\leqslant 2\,(n_1+n_2)\,\sigma\,\sqrt{n}\,.\end{aligned}$$

Since Theorem 64 is uniform over all choices of S and U of the specified sizes, the bound holds simultaneously for all $n_1 \times n_2$ combinatorial subrectangles, as claimed. ■

Theorem 66 *Fix a level $i \in \{1, 2, \ldots, \lfloor \log_2 k\rfloor\}$ and set $\beta_i := 2^{-i}$, $n_i := 2\beta_i n$, and $\gamma_i := \alpha_i/\beta_i = 2^i/k$ with $\alpha_i = 1/k$. Let $y \in \{0, \pm1\}^n$ be a valid level-i bisection labeling. Fix parameters $t \geqslant 1$ and $0 < \rho < 10^{-6}$, and set*

$$\widetilde{Q}_i = \beta_i \widetilde{C}_i - \log(1/\rho) - 3t.$$

Assume $a \geqslant b$, $a, b \leqslant n$, and $a \geqslant 400$. Then, with probability at least $1 - e^{-t\rho n_i} - 1/n_i^3$, for any z supported on $\{j : y_j \neq 0\}$ with $\|z\|_1 = \rho\, n_i$, we have

$$\langle \bar{G}\, y \odot y,\, z\rangle \geqslant 2\rho n_i \left(\frac{\widetilde{Q}_i}{\log \widetilde{R}_i} - \sqrt{a+b}\right).$$

Proof Imagine sampling the matrix G by independently drawing the entries G_{ij} for $i < j$ and then filling in the remainder symmetrically, with $G_{ji} = G_{ij}$ and $G_{ii} = 0$. If i and j are in the same community then $G_{ij} \sim \text{Bernoulli}(a/n)$; otherwise $G_{ij} \sim \text{Bernoulli}(b/n)$. Fix a level-$i$ bisection with sides S_i^+ and S_i^- of size $n_i/2$ each, and write $\alpha = 1/k$. For a vertex $u \in S_i^+$ consider the signed difference

$$\boldsymbol{X_u} = \sum_{v\in S_i^+\setminus\{u\}} G_{uv} - \sum_{v\in S_i^-} G_{uv}\ .$$

We observe that the neighbors of u split into three disjoint sets: its own community $C(u)\setminus\{u\}$ of size $\alpha n-1$ where edges have bias a/n; the remaining $(\beta_i-\alpha)n$ vertices on the same side where edges have bias b/n; and the $\beta_i n$ vertices on the opposite side where edges have bias b/n. Independence across disjoint unordered pairs implies that these three contributions are independent binomials, hence

$$\boldsymbol{X_u} \stackrel{d}{=} \mathrm{Binom}(\alpha n-1, a/n) + \mathrm{Binom}\big((\beta_i-\alpha)n, b/n\big) - \mathrm{Binom}(\beta_i n, b/n) \ .$$

We note that since $\langle Gy\odot y, z\rangle$ is linear in z and only coordinates with $y_j\neq 0$ contribute, it suffices to take $z=\mathbf{1}_S$ for some $S\subseteq\{j : y_j\neq 0\}$ with $|S|=\rho\, n_i$ (if $\rho n_i\notin\mathbb{N}$, take $\lfloor \rho n_i\rfloor$ and $\lceil \rho n_i\rceil$; any other $z\in[0,1]^n$ with the same ℓ_1-norm is a convex combination).

Let us first construct an auxiliary matrix G^{ind} in which, for each fixed row u, the variables $\{G^{\mathrm{ind}}_{uv}\}_v$ are independent Bernoullis with the same means as above (and different rows are also independent); define $\boldsymbol{X^{\mathrm{ind}}_u}$ analogously. Then

$$\langle G^{\mathrm{ind}} y\odot y,\, \mathbf{1}_S\rangle = \sum_{u\in S}\boldsymbol{X^{\mathrm{ind}}_u}.$$

We now apply Theorem 62 to the sum $\sum_{u\in S}\boldsymbol{X^{\mathrm{ind}}_u}$ by independence across rows with the parameter

$$\widetilde{\theta}_i = 2\rho n_i\frac{\beta_i\widetilde{C}_i-\log(1/\rho)-3t}{\log\widetilde{R}_i} = 2\rho n_i\frac{\widetilde{Q}_i}{\log\widetilde{R}_i}.$$

This yields

$$\mathbb{P}\Big[\sum_{u\in S}\boldsymbol{X^{\mathrm{ind}}_u}\leqslant\widetilde{\theta}_i\Big]\leqslant\exp\Big(-|S|\,\beta_i\widetilde{C}_i+\tfrac{\widetilde{\theta}_i}{2}\log\widetilde{R}_i\Big)=\exp\Big(-|S|(\log(1/\rho)+3t)\Big)\ .$$

A union bound over all $S\subseteq\{y\neq 0\}$ with $|S|=\rho n_i$ gives

$$\mathbb{P}\Big[\exists\, S:\sum_{u\in S}\boldsymbol{X^{\mathrm{ind}}_u}\leqslant\widetilde{\theta}_i\Big]\leqslant\binom{n_i}{\rho n_i}e^{-|S|(\log(1/\rho)+3t)}\leqslant e^{-t\rho n_i},$$

using $\binom{n_i}{\rho n_i}\leqslant\exp(\rho n_i(\log(1/\rho)+1))$ and $t\geqslant 1$. Therefore, with probability at least $1-e^{-t\rho n_i}$, simultaneously for all such S,

$$\sum_{u\in S}\boldsymbol{X^{\mathrm{ind}}_u}\geqslant\widetilde{\theta}_i = 2|S|\,\frac{\widetilde{Q}_i}{\log\widetilde{R}_i}. \tag{M.4}$$

It remains to pass from the independent sampling used above to the actual symmetric SBM with zero diagonal. We, first, note that zeroing the diagonal affects the sum by at most $|S|=\rho n_i$. We, further, note that only the block $S\times S$ differs between the two sampling procedures. Let $\Delta(S)$ be the difference between the contributions of $S\times S$ to $\sum_{u\in S}\boldsymbol{X^{\mathrm{ind}}_u}$ and to $\sum_{u\in S}\boldsymbol{X_u}$. As in Liu and Moitra (2022), $\Delta(S)$ is a sum/difference of independent, mean-0, bounded variables, each supported in $[-1,1]$ and nonzero with probability at most a/n (so $\sigma^2\leqslant(a+b)/n$). Apply Theorem 64 with $n_1=n_2=|S|$ and

$\sigma^2 = (a+b)/n$ (hence $\sigma \geqslant 20/\sqrt{n}$ because $a \geqslant 400$). Since $i \geqslant 1$, we have $n_i \leqslant n$ and hence $|S| = \rho n_i \leqslant \rho n \leqslant 10^{-6} n$, so the conditions hold. We obtain

$$\mathbb{P}\left[|\Delta(S)| > 2|S|\sqrt{a+b}\right] \leqslant \exp\Big(-16|S|\log\big(\tfrac{n}{|S|}\big)\Big) .$$

A union bound over all $\binom{n_i}{\rho n_i}$ choices of S shows that, with additional failure probability at most $1/n_i^3$ (for all large n), we have $|\Delta(S)| \leqslant 2|S|\sqrt{a+b}$ simultaneously for all such S.

Finally, intersecting the two high-probability events and noting that the level-i bisection is balanced so we get

$$\langle \bar{G}y \odot y,\, \mathbf{1}_S\rangle = \sum_{u\in S} y_u \sum_{v} y_v\,(G_{uv} - d_i) = \sum_{u\in S} \boldsymbol{X_u} \geqslant \widetilde{\theta}_i - 2|S|\sqrt{a+b}.$$

Substituting $|S| = \rho n_i$ and the definition of $\widetilde{\theta}_i$ yields the desired conclusion. ∎

Theorem 67 *Fix a level $i \in \{1,2,\ldots,\lfloor \log_2 k\rfloor\}$ and set $\beta_i \coloneqq 2^{-i}$, $n_i \coloneqq 2\beta_i n$, and $\gamma_i \coloneqq \alpha_i/\beta_i = 2^i/k$ with $\alpha_i = 1/k$. Assume $0 < \varepsilon < 1$ and $k \geqslant 2$, and take other parameters as in Theorem 62. Define*

$$\rho_\gamma \coloneqq \min\Big(\frac{1}{e}, \exp\big(-\frac{\varepsilon^2 d}{8k}\big)\Big) .$$

If, in addition, $\varepsilon^2 d \geqslant 9k^2$, then for any $0 < \rho \leqslant \min\{\rho_\gamma, 10^{-6}\}$ and $t = \varepsilon^2 d/k$, with probability at least $1 - e^{-t\rho n_i} - \frac{1}{n_i^3}$ we have

$$\langle Gy \odot y, z\rangle \geqslant -12\rho_\gamma n_i \varepsilon d \qquad \text{for every } z \text{ with } \|z\|_1 = \rho n_i,$$

where $y \in \{0, \pm 1\}^n$ is any valid bisection labeling at level i.

Proof Let $t = \varepsilon^2 d/k$ and fix any $0 < \rho \leqslant \min\{\rho_\gamma, 10^{-6}\}$. By Theorem 66, with probability at least $1 - e^{-t\rho n_i} - \frac{1}{n_i^3}$, we have for all z with $\|z\|_1 = \rho n_i$,

$$\langle Gy \odot y, z\rangle \geqslant 2\rho n_i \left(\frac{\beta_i \widetilde{C}_i - \log(1/\rho) - 3t}{\log \widetilde{R}_i} - \sqrt{a+b}\right) . \tag{M.5}$$

We now bound the three potentially negative contributions on the right-hand side of (M.5). For the $-\log(1/\rho)$ term, Theorem 60 yields

$$-\frac{2\rho n_i}{\log \widetilde{R}_i} \log\frac{1}{\rho} \geqslant -\frac{4\rho n_i}{\gamma_i \varepsilon} \log\frac{1}{\rho} .$$

By the monotonicity of $x\log(1/x)$ on $(0, e^{-1}]$ and $\rho \leqslant \rho_\gamma \leqslant e^{-1}$, we have

$$\rho \log\frac{1}{\rho} \leqslant \rho_\gamma \log\frac{1}{\rho_\gamma},$$

and therefore

$$-\frac{2\rho n_i}{\log \widetilde{R}_i}\log\frac{1}{\rho} \geqslant -\frac{4\rho_\gamma n_i}{\gamma_i\varepsilon}\log\frac{1}{\rho_\gamma} \geqslant -\frac{4\rho_\gamma n_i}{\gamma_i \varepsilon}\cdot\frac{\varepsilon^2 d}{k} = -4\rho_\gamma n_i \varepsilon d \cdot \frac{1}{\gamma_i k} \geqslant -4\rho_\gamma n_i \varepsilon d .$$

For the $-3t$ term, using $t = \varepsilon^2 d/k$ and Theorem 60 we obtain

$$-\frac{2\rho n_i \cdot 3t}{\log \widetilde{R}_i} \geqslant -\frac{12\rho n_i\, t}{\gamma_i\, \varepsilon} = -\frac{12\rho n_i}{\gamma_i\, \varepsilon} \cdot \frac{\varepsilon^2 d}{k} = -12\rho n_i \varepsilon d \cdot \beta_i \geqslant -6\, \rho_\gamma n_i\, \varepsilon d\ .$$

Finally, by Theorem 59, we have that

$$-2\rho n_i \sqrt{a+b} \geqslant -2\rho_\gamma n_i \sqrt{3d} \geqslant -2\rho_\gamma n_i \varepsilon d\ .$$

Combining these three bounds in (M.5) yields

$$\langle Gy \odot y, z\rangle \geqslant -\Big(4+6+2\Big)\, \rho_\gamma n_i \varepsilon d = -12\, \rho_\gamma n_i \varepsilon d \geqslant -12 \rho_\gamma n_i \varepsilon d,$$

which is the desired inequality. The probability statement is exactly that of Theorem 66 with $t = \varepsilon^2 d/k$, completing the proof. ■

Lemma 68 *Fix a level i and set $\beta_i = 2^{-i}$ and $n_i = 2\beta_i n$. Let $y \in \{0, \pm 1\}^n$ be a valid bisection labeling at level i. Choose $0 < \rho_\gamma \leqslant 10^{-6}$ and set $t = \varepsilon^2 d/k$. Then, with probability at least $1 - 2e^{-t} - n_i^{-2}$, the inequality*

$$\langle Gy \odot y, z\rangle \geqslant -12\rho_\gamma n_i \varepsilon d$$

holds simultaneously for every vector $z \in \{0, \pm 1\}^n$ with $\|z\|_1 \leqslant \rho_\gamma n_i$.

Proof We discretize over the admissible ℓ_1 sizes: since $z \in \{0, \pm 1\}^n$, the quantity $\|z\|_1$ is an integer, hence $\rho = \|z\|_1/n_i$ ranges over the set $\{j/n_i : 1 \leqslant j \leqslant \lfloor \rho_\gamma n_i \rfloor\}$. For a fixed value ρ we invoke the level-i voting lower bound (Theorem 67) with this ρ. It states that, with probability at least $1 - e^{-t\rho n_i} - n_i^{-3}$, the estimate

$$\langle Gy \odot y, z\rangle \geqslant -12\rho_\gamma n_i \varepsilon d$$

holds simultaneously for all z with $\|z\|_1 = \rho n_i$. Taking a union bound over the at most $\lfloor \rho_\gamma n_i \rfloor$ distinct values, the failure probability is at most

$$\sum_{j=1}^{\lfloor \rho_\gamma n_i \rfloor} \left(e^{-tj} + n_i^{-3}\right) \leqslant \frac{e^{-t}}{1-e^{-t}} + \frac{\lfloor \rho_\gamma n_i \rfloor}{n_i^3} \leqslant \frac{e^{-t}}{1-e^{-t}} + \frac{1}{n_i^2} \leqslant 2e^{-t} + \frac{1}{n_i^2}.$$

On the complementary event, the desired bound holds for every value in the set, and hence for every $z \in \{0, \pm 1\}^n$ with $\|z\|_1 \leqslant \rho_\gamma n_i$, as claimed. ■

Next we give a lemma for lower bounding the inner product when the size of the set is large.

Lemma 69 *Fix a level i and set $\beta_i = 2^{-i}$, $n_i = 2\beta_i n$, and $k_i = \beta_i k$. Let $y \in \{0, \pm 1\}^n$ be a valid bisection labeling at level i. Let $\gamma \in \big[0,\, 1 - \frac{1000\,\chi\, k_i}{\varepsilon\sqrt{d_i}}\big]$. Define*

$$\rho_\gamma \coloneqq \exp\Big(-\frac{\gamma\, \beta_i\, \widetilde{C}_i}{2}\Big), \qquad t \coloneqq 0.001\,(1-\gamma)\,\beta_i\,\widetilde{C}_i.$$

Then for every $\rho \geqslant \rho_\gamma$, with probability at least $1-\exp(-t\rho n_i)-\frac{1}{n_i^3}$, the inequality

$$\langle \bar{G}y \odot y,\, z\rangle \geqslant \frac{\varepsilon d}{5k}\,(1-\gamma)\,\rho_\gamma\, n_i$$

holds for all $z \in \{0,1\}^n$ with $\|z\|_1 = \rho n_i$, where $\bar{G}$ is the centered adjacency matrix.

Proof Let us fix $\rho \in [\rho_\gamma, 1]$ and $z \in \{0,1\}^n$ with $\|z\|_1 = \rho n_i$. By Theorem 66 applied to $\bar{G}$, with probability at least $1 - e^{-t\rho n_i} - n_i^{-3}$ we have

$$\langle \bar{G}y \odot y,\, z\rangle \geqslant 2\rho n_i\left(\frac{\widetilde{Q}_i}{\log \widetilde{R}_i} - \sqrt{a+b}\right), \qquad \widetilde{Q}_i \coloneqq \beta_i \widetilde{C}_i - \log\frac{1}{\rho} - 3t. \tag{M.6}$$

Since $\rho \geqslant \rho_\gamma = \exp(-\gamma\,\beta_i \widetilde{C}_i/2)$, we have

$$-\log(1/\rho) \geqslant -\gamma\,\beta_i\widetilde{C}_i/2,$$

and hence

$$\begin{aligned}\widetilde{Q}_i &\geqslant \Big(1-\frac{\gamma}{2}-0.003(1-\gamma)\Big)\,\beta_i\widetilde{C}_i\\ &\geqslant 0.997\big(1-\gamma\big)\,\beta_i\widetilde{C}_i. \end{aligned}\tag{M.7}$$

By Lemma 60, $\log \widetilde{R}_i \leqslant \gamma_i\,\varepsilon\log 3 + o(1)$. Therefore

$$\begin{aligned}\frac{\widetilde{Q}_i}{\log \widetilde{R}_i} &\geqslant 0.997\frac{(1-\gamma)\,\beta_i\,\widetilde{C}_i}{\gamma_i\,\varepsilon\log 3 + o(1)}\\ &\geqslant 0.997\cdot\frac{(\log 2)^2}{4\log 3}\,(1-\gamma)\,\frac{\varepsilon d}{k}, \end{aligned}\tag{M.8}$$

using

$$\beta_i\,\widetilde{C}_i \geqslant \frac{(\log 2)^2}{4}\,\frac{d\,\varepsilon^2}{\beta_i\,k^2}$$

by the monotonicity shown in Theorem 63.
For the last term, Lemma 59 gives $\sqrt{a+b} \leqslant \sqrt{3d}$. We thus have

$$\sqrt{a+b} \leqslant \sqrt{3d} \leqslant 0.001\,(1-\gamma)\,\frac{\varepsilon d}{k}. \tag{M.9}$$

Combining (M.6), (M.8), and (M.9), we obtain

$$\begin{aligned}\langle \bar{G}y \odot y,\, z\rangle &\geqslant 2\rho n_i\left(0.109\,(1-\gamma)\,\frac{\varepsilon d}{k} - 0.001\,(1-\gamma)\,\frac{\varepsilon d}{k}\right)\\ &\geqslant \frac{\varepsilon d}{5k}\,(1-\gamma)\,\rho n_i\\ &\geqslant \frac{\varepsilon d}{5k}\,(1-\gamma)\,\rho_\gamma n_i\ \ , \end{aligned}$$

as claimed. ∎

Corollary 70 *Fix a level i and set $\beta_i = 2^{-i}$ and $n_i = 2\beta_i n$. Let $y \in \{0, \pm 1\}^n$ be a valid bisection labeling at level i. Let $\gamma \in \left[0,\, 1 - \frac{1000\,\chi\, k_i}{\varepsilon\sqrt{d_i}}\right]$, and define*

$$\rho_\gamma := \exp\left(-\tfrac{\gamma\,\beta_i\,\widetilde{C}_i}{2}\right), \qquad t := 0.001\,(1-\gamma)\,\beta_i\,\widetilde{C}_i.$$

Then, with probability at least $1 - 2e^{-t\,\rho_\gamma n_i} - n_i^{-2}$, the inequality

$$\langle \bar{G}y \odot y, z\rangle \geqslant \frac{\varepsilon d}{5k}\,(1-\gamma)\,\rho_\gamma\, n_i$$

holds simultaneously for every vector $z \in \{0,1\}^n$ with $\|z\|_1 \geqslant \rho_\gamma n_i$, where $\bar{G}$ is the centered adjacency matrix.

Proof We discretize over the admissible ℓ_1 sizes. Since $z \in \{0,1\}^n$, the quantity $\|z\|_1$ is an integer, so $\rho = \|z\|_1/n_i$ ranges over $\{j/n_i : m_0 \leqslant j \leqslant n_i\}$, where $m_0 := \lceil \rho_\gamma n_i \rceil$. Fix $j \in \{m_0, \ldots, n\}$ and set $\rho = j/n_i \geqslant \rho_\gamma$. By Lemma 69 (applied with our choice of ρ), with probability at least $1 - e^{-t\rho n_i} - n_i^{-3} = 1 - e^{-tj} - n_i^{-3}$, we have

$$\langle \bar{G}y \odot y, z\rangle \geqslant \frac{\varepsilon d}{5k}\,(1-\gamma)\,\rho_\gamma\, n_i \quad \text{for all } z \in \{0,1\}^n \text{ with } \|z\|_1 = j.$$

Taking a union bound over all $j \in \{m_0, \ldots, n_i\}$, the total failure probability is at most

$$\sum_{j=m_0}^{n_i} \left(e^{-tj} + n_i^{-3}\right) \leqslant \frac{e^{-tm_0}}{1 - e^{-t}} + \frac{n_i - m_0 + 1}{n_i^3} \leqslant 2e^{-t\rho_\gamma n_i} + \frac{1}{n_i^2}.$$

On the complementary event, the claimed lower bound holds simultaneously for all sizes $j \geqslant m_0$, and hence for every $z \in \{0,1\}^n$ with $\|z\|_1 \geqslant \rho_\gamma n_i$, as desired. ∎

Now we conclude the proof of the majority voting lower bound of bisections.

Theorem *[Restatement of Theorem 14] Fix a level i and set $\beta_i = 2^{-i}$, $n_i = 2\beta_i n$, and $k_i = \beta_i k$. Let $\gamma \in [0, 0.99]$. Choose $\rho_\gamma := \exp(-\gamma\beta_i\tilde{C}_i/2)$, and set $t = 0.001(1-\gamma)\tilde{C}_i$. Then with probability at least $1 - \exp(-100k) - \frac{1}{n^3}$, for every $z \in \{0,1\}^n$, and for every valid bisections $y \in \{0, \pm 1\}^n$ at level i. we have*

$$\langle \bar{G}y \odot y, z\rangle \geqslant \frac{(1-\gamma)\varepsilon d}{8k}\left(\|z\|_1 - \frac{96\rho_\gamma k n_i}{1-\gamma}\right),$$

where $\bar{G}$ is the centered adjacency matrix.

Proof We first prove the claim for fixed bisection y. When the size of the set is at least $\rho_\gamma n$, the lower bound easily follows from Theorem 70. When the size of the set is at most $\rho_\gamma n$, the lower bound follows from Theorem 68 with probability $1 - \exp(-\varepsilon^2 d/k)$, as $\|z\|_1 \geqslant 0$. From the proof of Theorem 63, we know that $\beta_i\tilde{C}_i \geqslant \Omega(\varepsilon^2 d_i/k_i^2) \geqslant \Omega(2^i \log(k))$. It follows that $\tilde{C}_i\rho_\gamma n \geqslant 100k\log(k)$. Since there are at most 2^{2k} possible choices for community bisections y, we can take union bound and the claim follows. ∎

Finally we give the bound for larger values of γ which will be helpful for obtaining the optimal recovery rate in the final k-clustering.

Theorem *[Restatement of Theorem 16] Let $C_{d,\varepsilon} = (\sqrt{a} - \sqrt{b})^2$. Let $\gamma \in [0, 1 - \frac{1000\chi k}{\varepsilon\sqrt{d}}]$. Choose $\rho_\gamma := \exp(-\gamma C_{d,\varepsilon}/k)$. Then with probability at least $1 - \exp(-100k) - \frac{1}{n^3}$, for every $z \in \{0,1\}^n$, and for every pair of communities $y \in \{0, \pm 1\}^n$. We have*

$$\langle \bar{G}y \odot y, z \rangle \geqslant \frac{(1-\gamma)\varepsilon d}{8k}\left(\|z\|_1 - \frac{96\rho_\gamma kn}{1-\gamma}\right),$$

where $\bar{G}$ is the centered adjacency matrix.

Proof The proof is very similar to Theorem 14. However, now for the probability bound, we only need to use $\rho_\gamma n \geqslant 1$ and $C \geqslant 100k$, and we only need to take union bound over k^2 pairs of communities. ■

Corollary *[Restatement of Theorem 15] Fix a level i and set $\beta_i = 2^{-i}$, $n_i = 2\beta_i n$, and $k_i = \beta_i k$. Let $\gamma \in [0, 0.99]$. Choose $\rho_\gamma := \exp(-\gamma\beta_i\tilde{C}_i/2)$, and set $t = 0.001(1-\gamma)\tilde{C}_i$. Then with probability at least $1 - \exp(-100k) - \frac{2}{n^3}$, for every $z \in \{0,1\}^n$, for every $s \in \{0,1\}^n$ such that $\|s\|_1 \geqslant (1 - \exp(-2C_{d,\varepsilon}))n$, and for every valid bisections $y \in \{0, \pm 1\}^n$ at level i, we have*

$$\langle \bar{G} \odot (ss^\top)y \odot y, z \rangle \geqslant \frac{(1-\gamma)\varepsilon d}{16k}\left(\|z\|_1 - \frac{640\rho_\gamma kn_i}{1-\gamma}\right),$$

where $\bar{G}$ is the centered adjacency matrix.

Proof Notice that

$$\begin{aligned}\langle \bar{G} \odot (ss^\top)y \odot y, z \rangle &= \langle \bar{G} \odot (yy^\top), (z \odot s)s^\top \rangle \\ &= \langle \bar{G} \odot (yy^\top), (z \odot s)\mathbf{1}^\top \rangle - \langle \bar{G} \odot (yy^\top), (z \odot s)(\mathbf{1} - s)^\top \rangle\,.\end{aligned}$$

By Theorem 14, it follows that, with probability $1 - \exp(-100k) - \frac{1}{n^3}$,

$$\begin{aligned}\langle \bar{G} \odot (yy^\top), (z \odot s)\mathbf{1}^\top \rangle &\geqslant \frac{(1-\gamma)\varepsilon d}{8k}\left(\|z \odot s\|_1 - \frac{96\rho_\gamma kn}{1-\gamma}\right) \\ &\geqslant \frac{(1-\gamma)\varepsilon d}{8k}\left(\|z\|_1 - \|\mathbf{1} - s\|_1 - \frac{96\rho_\gamma kn}{1-\gamma}\right) \\ &\geqslant \frac{(1-\gamma)\varepsilon d}{8k}\left(\|z\|_1 - \exp(-2C_{d,\varepsilon})n - \frac{96\rho_\gamma kn}{1-\gamma}\right) \\ &\geqslant \frac{(1-\gamma)\varepsilon d}{8k}\left(\|z\|_1 - \frac{200\rho_\gamma kn}{1-\gamma}\right).\end{aligned}$$

By Theorem 65, it follows that, with probability $1 - \frac{1}{n^3}$

$$\langle \bar{G} \odot (yy^\top), (z \odot s)(\mathbf{1} - s)^\top \rangle \leqslant \sqrt{2d}(\|z\|_1 + \exp(-2C_{d,\varepsilon})n) \leqslant \frac{(1-\gamma)\varepsilon d}{16k}\left(\|z\|_1 + \frac{100\rho_\gamma kn}{1-\gamma}\right).$$

Therefore, with probability $1 - \exp(-100k) - \frac{2}{n^3}$,

$$\langle \bar{G} \odot (ss^\top)y \odot y, z \rangle \geqslant \frac{(1-\gamma)\varepsilon d}{8k}\left(\|z\|_1 - \frac{200\rho_\gamma kn}{1-\gamma}\right) - \frac{(1-\gamma)\varepsilon d}{16k}\left(\|z\|_1 + \frac{100\rho_\gamma kn}{1-\gamma}\right)$$

$$\geqslant \frac{(1-\gamma)\varepsilon d}{16k}\left(\|z\|_1 - \frac{640\rho_\gamma kn}{1-\gamma}\right).$$

■

Corollary *[Restatement of Theorem 17]* *Let* $\gamma \in [0, 1 - \frac{1000\chi k}{\varepsilon\sqrt{d}}]$. *Choose* $\rho_\gamma := \exp(-\gamma C)$. *Then with probability at least* $1 - \exp(-100k) - \frac{2}{n^3}$, *for every* $z \in \{0,1\}^n$, *for every* $s \in \{0,1\}^n$ *such that* $\|s\|_1 \geqslant (1 - \exp(-2C_{d,\varepsilon}))n$, *and for every pair of communities* $y \in \{0, \pm 1\}^n$. *We have*

$$\langle \bar{G} \odot (ss^\top) y \odot y, z\rangle \geqslant \frac{(1-\gamma)\varepsilon d}{16k}\left(\|z\|_1 - \frac{640\rho_\gamma kn}{1-\gamma}\right),$$

where $\bar{G}$ *is the centered adjacency matrix.*

Proof The proof is very similar to Theorem 15. The only difference is that the voting lower bound is via Theorem 16. ■