Reaching Kesten-Stigum Threshold in the Stochastic Block Model under Node Corruptions

Jingqiu Ding JINGQIU.DING@INF.ETHZ.CH
ETH Zürich

Tommaso d'Orsi TOMMASO.DORSI@INF.ETHZ.CH
ETH Zürich

Yiding Hua YIDING.HUA@INF.ETHZ.CH
ETH Zürich

David Steurer DAVID.STEURER@INF.ETHZ.CH
ETH Zürich

Editors: Gergely Neu and Lorenzo Rosasco

Abstract

We study robust community detection in the context of node-corrupted stochastic block model, where an adversary can arbitrarily modify all the edges incident to a fraction of the n vertices. We present the first polynomial-time algorithm that achieves weak recovery at the Kesten-Stigum threshold even in the presence of a small constant fraction of corrupted nodes. Prior to this work, even state-of-the-art robust algorithms were known to break under such node corruption adversaries, when close to the Kesten-Stigum threshold.

We further extend our techniques to the $\mathbb{Z}_2$ synchronization problem, where our algorithm reaches the optimal recovery threshold in the presence of similar strong adversarial perturbations.

The key ingredient of our algorithm is a novel identifiability proof that leverages the push-out effect of the Grothendieck norm of principal submatrices.

Keywords: Community Detection, Robust Statistics, Sum-of-Squares Hierarchy

1. Introduction

Community detection is the problem of identifying hidden communities in random graphs that are generated based on some planted structures. The stochastic block models are a family of random graph models that, for a variety of reasons, play a central role in the study of community detection (see the excellent survey of Abbe (2017)). In this work, we focus on balanced two-community stochastic block model, which is defined as follows:

Definition 1 (Balanced two-community stochastic block model) *For parameters $\varepsilon \in (0,1), n > 0, d > 0$, for some label vector $x^* \in \{\pm 1\}^n$ such that $\mathbf{1}^\top x^* = 0$, the balanced two-community stochastic block model $\mathsf{SBM}_{d,\varepsilon}(x^*)$ describes the following distribution over graph with n vertices: every pair of distinct vertices $i, j \in [n]$ in the graph is connected by an edge independently with probability $(1 + \varepsilon x_i^* x_j^*)\frac{d}{n}$.*

In the above definition, ε is the bias parameter, d is the average degree and each vertex i gets a label x_i^*. Given a graph G sampled according to this model, the goal is to recover the (unknown) underlying vector of labels x^* as accurate as possible.

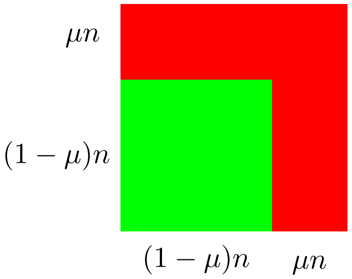

Figure 1: For the adjacency matrix generated from definition 3, the entries in the red area are adversarially corrupted, while the entries in the green area are sampled as in stochastic block model.

Definition 2 (Weak recovery) *For $G \sim \mathsf{SBM}_{d,\varepsilon}(x^*)$, an estimator $\hat{x}(G) \in \{\pm 1\}^n$ is said to achieve weak recovery if and only if $\mathbb{E}\langle \hat{x}(G), x^* \rangle^2 \geqslant \Omega(n^2)$.*

In the last decade, a celebrated line of works established that weak recovery can be achieved (and efficiently so) if and only if $\varepsilon^2 d > 1$ when $d = o(n)$ (Decelle et al. (2011); Massoulié (2014); Mossel et al. (2014, 2015); Banerjee (2018)). This threshold is called the *Kesten-Stigum threshold* (henceforth KS threshold). The algorithms introduced in these works, however, are brittle and can be fooled by very small adversarial perturbations.[1] Later works (Montanari and Sen (2016); Ding et al. (2022)) showed that, perhaps surprisingly, we can achieve weak recovery at the KS threshold, even in the presence of edge perturbations.

It is easy to see that weak recovery becomes information theoretically impossible when more than $\varepsilon \cdot d \cdot n$ edges are corrupted.[2] In light of this, Liu and Moitra (2022) explored an alternative model of corruption, where a constant fraction of vertices have all of their incident edges modified.

Definition 3 (Node-corrupted balanced two-community stochastic block model) *Given $\mu \in [0,1)$ and $G^0 \sim \mathsf{SBM}_{d,\varepsilon}(x^*)$, an adversary may choose up to μn vertices in G^0 and arbitrarily modify incident edges of them to produce the corrupted graph G.*

In this node corruption model, the adversary is allowed to change an arbitrary number of edges for each corrupted vertex and could introduce up to $O(\mu n^2)$ edges. While vertices of untypically large degree are algorithmically easy to identify and remove, less naive adversaries may remove all edges of an arbitrary subset of μn vertices, and replace them by roughly d spurious edges of their choosing. Such an adversary would introduce $O(\mu \cdot d \cdot n)$ edges, causing the algorithms of Montanari and Sen (2016); Ding et al. (2022) to fail when $\mu \geqslant \varepsilon$.

As a result, node corruption settings present new algorithmic challenges that previous approaches fail to overcome. Liu and Moitra (2022) developed a polynomial-time algorithm which achieves minimax error rates when $\varepsilon^2 d > C$ and the corruption ratio is $\mu = o(1)$ (where C is a large universal constant). [3]

[1] In the sense that, altering $n^{o(1)}$ edges is enough to break them.

[2] An adversary may remove each intra-cluster edge with probability $\varepsilon \cdot d/n$ and add each inter-cluster edge with probability $\varepsilon \cdot d/n$. With high probability such process alter at most $\varepsilon \cdot d \cdot n(1+o(1))$ edges. The graph now is indistinguishable from an Erdős-Rényi graph.

[3] The algorithm in Liu and Moitra (2022) is further robust against a different type of adversary corruptions with unbounded monotone changes from semi-random model defined in Moitra et al. (2016).

However, their algorithm cannot achieve weak recovery when the signal-to-noise ratio is close to the KS threshold, i.e. when $\varepsilon^2 d = 1 + \delta$ for small δ.

In this work, we develop algorithms that can exploit the special structure of node corruptions and *achieve node-robust weak recovery* as long as $\mu \leqslant \mu_\delta$, where μ_δ is a constant that only depends on δ but not on ε or d.

1.1. Results

Our main result is an efficient algorithm to achieve robust weak recovery under node corruptions.

Theorem 4 *Let $n > 1, d > 1, \varepsilon \in (0,1)$ and balanced label vector $x^* \in \{\pm 1\}^n$. Let $\delta := \varepsilon^2 d - 1$. Let G be a corrupted graph generated from definition 3 given $G^0 \sim \mathsf{SBM}_{d,\varepsilon}(x^*)$ and $\mu \in [0,1)$. When $\delta \geqslant \Omega(1)$ and $\mu \leqslant \Omega_\delta(1)$ [4], there exists a polynomial-time algorithm that outputs a labelling $\hat{x} \in \{\pm 1\}^n$ such that*

$$\mathbb{E}[\langle \hat{x}, x^* \rangle^2] \geqslant \Omega(n^2)$$

Our algorithm is the first one that succeeds in this setting, Liu and Moitra (2022) cannot work unless δ is sufficiently large, and Montanari and Sen (2016); Ding et al. (2022) cannot tolerate the corruption of $\Omega_\delta(1)$ vertices. Additionally, our algorithm is optimal in the sense that for $\varepsilon^2 d < 1$, it is information theoretically impossible to achieve weak recovery (see Mossel et al. (2015)).

Based on the techniques we use for the robust stochastic block model, we also give an algorithm for the closely related robust $\mathbb{Z}_2$ synchronization problem, which can be formulated as follows:

Definition 5 (Row/column-corrupted $\mathbb{Z}_2$ Synchronization model) *Given a hidden vector $x^* \in \{\pm 1\}^n$ and $\sigma > 0$, let A^0 be the uncorrupted $\mathbb{Z}_2$ synchronization matrix*

$$A^0 = \sigma x^* (x^*)^\top + W$$

where $W \in \mathbb{R}^{n \times n}$ is a symmetric random matrix whose upper triangular entries are i.i.d sampled from $N(0, n)$. An adversary may select μn elements of $[n]$ and arbitrarily modify the corresponding rows and columns of A^0 to produce a corrupted matrix A that we observe.

When $\sigma \leqslant 1$, even with no corruptions (i.e. $\mu = 0$), it is information theoretically impossible to achieve weak recovery (Perry et al. (2018)). When $\sigma \geqslant 1 + \Omega(1)$ and $\mu = 0$, a polynomial-time algorithm is known to output estimator $\hat{x} \in \{\pm 1\}^n$ such that $\langle \hat{x}, x \rangle^2 \geqslant \Omega(n^2)$ with high probability. This is due to the thresholding phenomenon called the BBP transition (Baik et al. (2005)). However, for reasons similar to those described in the SBM settings, when $\mu \geqslant \Omega(1)$, the analysis of known algorithms such as semidefinite programming (Montanari and Sen (2016)) or spectral algorithm (Perry et al. (2018)) breaks down. In this paper, we give an algorithm that can achieve the constant sharp threshold for robust $\mathbb{Z}_2$ synchronization:

Theorem 6 (Proved in section G) *Given a row/column corrupted matrix A generated from definition 5, when $\sigma \geqslant 1 + \Omega(1)$, there is a polynomial-time algorithm that outputs an estimator $\hat{x} \in \{\pm 1\}^n$ such that $\mathbb{E}\left[\langle \hat{x}, x^* \rangle^2\right] \geqslant \Omega(n^2)$ with high probability over x^* and A^0.*

[4] $\mu \leqslant \Omega_\delta(1)$ here means that μ is bounded by a constant depending on δ. The dependence on δ is necessary: if μ is a fixed constant, then the recovery impossible for a small enough constant δ (see section D for details).

2. Related previous work

Edge corruption robust algorithms A large body of work, concerning stochastic block model, has focused on the settings where an adversary may arbitrarily modify $\Omega(n)$ edges. When the average degree d diverges, Montanari and Sen (2016) showed that a simple semidefnite program is robust to such perturbations. Ding et al. (2022) presented a complementary result, providing weak recovery guarantees for the case when the average degree is bounded.

Node corruption robust algorithms The study of node corruptions in the stochastic block model was initiated in Cai and Li (2014). Subsequent work by Stephan and Massoulié (2019) provided a fast algorithm that achieves the KS threshold under the corruptions of $O(n^{0.001})$ nodes. However, this level of robustness is significantly weaker compared to our paper, which can tolerate constant fraction corrupted nodes.

Constant-fraction node corruptions in the stochastic block model were first studied in Liu and Moitra (2022), where the authors obtained a polynomial-time algorithm that achieves optimal recovery rates when $\varepsilon^2 d - 1$ is a sufficiently large constant. Although this provides weak recovery for large values of $\varepsilon^2 d$, it falls short of reaching the KS threshold, where $\varepsilon^2 d$ approaches 1.

Notably, the same node corruption model has also been studied in other random graph estimation problems, including the estimation of edge density in an Erdős–Rényi random graph, as studied by Acharya et al. (2022).

Other notions of robustness The algorithm of Liu and Moitra (2022) is further robust against unbounded monotone changes. *Monotone adversaries* have been shown to make it information theoretically impossible to reach the KS threshold (Moitra et al. (2016)). This implies that finding algorithms robust against monotone adversaries is infeasible in our settings.

Abbe et al. (2020) proposed a spectral powering algorithm that is able to practically achieve the KS threshold. This algorithm is robust against tangles and cliques, making it applicable to the geometric block model. However, it is not expected to provide comparable robustness guarantees under stronger adversaries.

$\mathbb{Z}_2$ synchronization In the related $\mathbb{Z}_2$ synchronization model, similar guarantees have also been obtained. When the number of corrupted entries is limited to $\tilde{O}(n^{3/2})$, the weak recovery threshold $\sigma > 1$ (BBP threshold) can be achieved via basic semidefinite programming (Montanari and Sen (2016)). Recently, Liu and Moitra (2022) showed that, under the row/column-corruption model (see definition 5), achieving the minimax error rate is possible with $o(n)$ corrupted rows/columns, when σ is greater than a sufficiently large constant. However, when σ is close to 1 and $\Omega(n)$ rows/columns are corrupted, no weak recovery algorithms (including inefficient algorithms) are known.

3. Techniques

We outline here the main ideas behind theorem 4 and theorem 6. The algorithm is splitted into two components, each tailored to handle one degree regime. For the regime with average degree $d > d_\delta$, where d_δ is a constant that only depends on $\delta := d\varepsilon^2 - 1$, our starting point is the result of Montanari and Sen (2016). For the sparse regime with $d \leqslant d_\delta$, we will borrow from Ding et al. (2022).

Push-out effect of the basic SDP Consider the settings $d > d_\delta$ and, for simplicity of the exposition, assume $d = \omega(1)$ and $\mu = o(1)$. Montanari and Sen (2016) proved that the following SDP program –which we refer to as the basic SDP–

$$\mathrm{SDP}(M) = \max\left\{\langle M, X\rangle : X \succeq 0, X_{ii} = 1 \forall i \in [n]\right\}$$

achieves weak recovery at the KS threshold. Concretely, for an uncorrupted graph $G^0 \sim \mathsf{SBM}_{d,\varepsilon}(x^*)$ with centered adjacency matrix $\tilde{A}^0$, they showed for some constant $\Delta_\delta > 0$,

$$\mathrm{SDP}(\tilde{A}^0) \geqslant (2+\Delta_\delta)n\sqrt{d}\,, \tag{3.1}$$

$$\mathrm{SDP}(\tilde{A}^0 - \frac{\varepsilon d}{n}x^* x^{*\mathsf{T}}) \leqslant \left(2+\frac{\Delta_\delta}{2}\right)n\sqrt{d}\,. \tag{3.2}$$

with probability $1-\exp(-\Omega_\delta(n))$. Taken together, these inequalities highlight a significant shift in the SDP value resulting from the subtraction of a rank-1 matrix from $\tilde{A}^0$. This phenomenon is often referred to as the *push-out effect* and has often been exploited to design algorithms (Perry et al. (2018); Montanari and Sen (2016); Ding et al. (2022)). Additionally, the exponential concentration probability allows us to demonstrate that the *push-out effect* occurs for every principal submatrix of size $(1-o(1))n \times (1-o(1))n$

It is easy to see that the basic SDP is robust against *some* adversarial perturbations. A single edge alteration can change both eq. (3.1) and eq. (3.2) by at most 2. As $\mathrm{SDP}(\tilde{A}^0) - \mathrm{SDP}(\tilde{A}^0 - \frac{\varepsilon d}{n}x^* x^{*\mathsf{T}}) \geqslant \Omega(n\sqrt{d})$, as long as the number of edge corruptions is bounded from above by $O(n\sqrt{d})$, the algorithm can still approximately recover the communities.

While this algorithm is robust to some edge adversarial perturbations, it is highly non-trivial whether it still works in presence of $\Omega(1)$-fraction of corrupted nodes, even if we assume all corrupted nodes have degree $O(d)$. In the node corruption model, the number of modified edges can reach $\Omega(n \cdot d)$, which is far more than $n\sqrt{d}$ when $d = \omega(1)$. The gap between nd and $n\sqrt{d}$ indicates that we need a fundamentally different approach to find algorithms robust to $\Omega(n)$ corrupted nodes.

Push-out effect of submatrices A priori it is not clear whether it is possible to recover the signal in presence of node corruptions, or if such an adversary has the capability of hiding all the information. A good news is that, while the basic SDP is fragile to node corruptions, it *suggests* a plausible direction to design an (inefficient!) algorithm robust to node corruptions. The key observation is that there is always a principal submatrix of size $(1-\mu)n \times (1-\mu)n$ free from corruption. More specifically, let $\tilde{A}$ be the adjacency matrix of the corrupted graph, the structure of node corruptions implies that the uncorrupted vertices $S^* \subseteq [n]$ satisfies $\tilde{A}_{S^*} = \tilde{A}^0_{S^*}$, where $\tilde{A}_{S^*}$ and $\tilde{A}^0_{S^*}$ denote the submatrix of $\tilde{A}$ and $\tilde{A}^0$ restricted to the set $S^* \times S^*$. Moreover, it can be shown that, with high probability, the push-out effect *still holds* for this submatrix. That is:

$$\mathrm{SDP}(\tilde{A}_{S^*}) \geqslant (2+\Delta_\delta)(1-\mu)n\sqrt{d}\,, \tag{3.3}$$

$$\mathrm{SDP}\left(\tilde{A}_{S^*} - \frac{\varepsilon d}{n}x^*_{S^*}x^{*\top}_{S^*}\right) \leqslant \left(2+\frac{\Delta_\delta}{2}\right)(1-\mu)n\sqrt{d}\,. \tag{3.4}$$

In other words, if we *knew* the set of uncorrupted nodes, then we would still be able to approximately recover the communities.

Unfortunately, the set of uncorrupted nodes S^* is not immediately known. Moreover, even disregarding computational issues, it remains unclear how one could identify such a set. A rudimentary strategy to address this challenge would be to identify a subset $S \subseteq [n]$ such that the objective value of the $\mathrm{SDP}(\tilde{A}_S)$ is large and to use the optimizer X as an estimator. However, this approach presents a problem in that the selected set S may contain corrupted vertices, leading to a situation where the optimizer X may align with the corruption rather than accurately reflecting the true labels.[5]

A natural way to circumvent this issue, is to search over *pairs* (S, X) where $S \subseteq [n]$ and X is a positive semidefinite matrix that fulfills the *submatrix push-out constraints* that is described below.

Definition 7 *Given a corrupted graph G as described in definition 3 and its centered adjacency matrix $\tilde{A}$, consider a set $S \subseteq [n]$ such that $|S| = (1-\mu)n$ and a positive semidefinite matrix X where $X_{ii} = 1$ for all $i \in [n]$, we say that the triplet $(\tilde{A}, S, X)$ satisfies submatrix push-out constraints if and only if for every subset $S' \subseteq S$ such that $|S'| \geqslant (1-2\mu)n$, it holds that*

$$\langle \tilde{A}_{S'}, X_{S'} \rangle \geqslant (2+\Delta_\delta)(1-o(1))n\sqrt{d}\,.$$

The *submatrix push-out constraints* is useful because, if one can find S, X such that $(\tilde{A}, S, X)$ satisfies the *submatrix push-out constraints*, then for $S' = S \cap S^*$, which is the set of uncorrupted nodes in set S and has size at least $(1-2\mu)n$, it follows from definition 7 that

$$\langle \tilde{A}_{S'}, X_{S'} \rangle \geqslant (2+\Delta_\delta)(1-o(1))n\sqrt{d}\,.$$

Therefore, we know that X correlates well with uncorrputed nodes S' in set S. By the basic SDP push-out effect of submatrices, we obtain

$$\langle \tilde{A}_{S'} - X^*_{S'}, X_{S'} \rangle \leqslant \left(2 + \frac{\Delta_\delta}{2}\right)(1-o(1))n\sqrt{d}\,.$$

Thus, one can deduce that $\langle X_{S'}, X^*_{S'} \rangle \geqslant \Omega(n^2)$. Since the entries of X are within $[-1, 1]$, we can further obtain $\langle X, X^* \rangle \geqslant \Omega(n^2)$ as well. Subsequently, after applying the standard rounding procedure outlined in lemma 25, we obtain an estimator $\hat{x} \in \{\pm 1\}^n$ with a weak recovery guarantee $\langle x^*, \hat{x} \rangle^2 \geqslant \Omega(n^2)$.

Certificates for the submatrix push-out effect Even with the *submatrix push-out constraints*, two fundamental challenges remain. *First*, we need to prove the existence of a pair (S, X) that satisfies the submatrix push-out constraints. *Second*, we need to be able to find such a pair efficiently.

With regard to the first challenge, ideally we would like to prove that the set of uncorrupted nodes S^* and the optimizer X of $\mathrm{SDP}(\tilde{A}_{S^*})$ fulfill the submatrix push-out constraints. However, it is difficult prove this: even though we have established that $\langle \tilde{A}_{S^*}, X \rangle \geqslant (2+\Omega(1))(1-\mu)n\sqrt{d}$, it remains unclear whether $\langle \tilde{A}_{S^*} - \tilde{A}_{S'}, X \rangle$ is small for all $S' \subseteq S^*$ of size $(1-2\mu)n$.

To overcome this barrier, we make the following crucial observation:

Lemma 8 (Formal statement and proof in section H.3) *Given S of size $(1-\mu)n$, if $\left\| \tilde{A}_S \right\|_{\mathrm{op}} \leqslant O(\sqrt{d})$, then $\mathrm{SDP}(\tilde{A}_S - \tilde{A}_{S'}) \leqslant O(\mu n\sqrt{d})$ for all $S' \subseteq S$ of size at least $(1-2\mu)n$.*

[5] Note that even with no corruption, when δ is a small constant, the optimizer X is only weakly correlated with $x^* x^{*\mathsf{T}}$.

This result suggests us to consider the following program:

$$\begin{aligned} \max_{X,S} \quad & \langle \tilde{A}_S, X\rangle \\ \text{s.t.} \quad & X \succeq 0 \\ & X_{ii} = 1 \quad \forall i \in [n] \\ & \left\| \tilde{A}_S \right\|_{\text{op}} \leqslant O(\sqrt{d}) \end{aligned} \tag{3.5}$$

We begin by establishing the feasibility of the program. Although the spectral norm of $\tilde{A}_{S^*}$ can potentially reach $\operatorname{polylog}(n)$, we can leverage the results of Feige and Ofek (2005) and reduce it to $O(\sqrt{d})$ through the pruning of high-degree nodes. The feasibility of the program can then be confirmed by taking S as the set of uncorrputed nodes and have degree at most $O(d)$. Furthermore, by union bound and the push-out effect established in eq. (3.1) and eq. (3.2), we have $\mathrm{SDP}(\tilde{A}_S) \geqslant (2+\Delta_\delta)\cdot(1-2\mu)n\sqrt{d}$ with high probability. Therefore the objective value of this program is at least $(2+\Delta_\delta)\cdot(1-2\mu)n\sqrt{d}$.

The optimizer of program 3.5, denoted by the pair $(\hat{X},\hat{S})$, can then be shown to satisfy the submatrix push-out constraints as defined in definition 7. It follows from our previous argument that the objective value of this program is at least $(2+\Delta_\delta)\cdot(1-2\mu)n\sqrt{d}$, which implies $\langle \tilde{A}_{\hat{S}}, \hat{X}\rangle \geqslant (2+\Delta_\delta)\cdot(1-2\mu)n\sqrt{d}$. Moreover, the program constraints enforce the bound $\left\|\tilde{A}_{\hat{S}}\right\|_{\text{op}} \leqslant O(\sqrt{d})$. Together with lemma 8, these implies that $\mathrm{SDP}(\tilde{A}_{\hat{S}} - \tilde{A}_{S'}) \leqslant O(\mu n\sqrt{d})$ for all $S' \subseteq \hat{S}$ with size at most $(1-2\mu)n$. When $\mu = o(1)$, it follows that $\mathrm{SDP}(\tilde{A}_{S'}) \geqslant \mathrm{SDP}(\tilde{A}_{\hat{S}}) + \mathrm{SDP}(\tilde{A}_{S'} - \tilde{A}_{\hat{S}}) \geqslant (2+\Delta_\delta)\cdot(1-o(1))\cdot n\sqrt{d}$ for all $S' \subseteq S_{\max}$ with size at most $(1-2\mu)n$.

As a result, using similar analysis as the previous paragraph, due to the basic SDP push-out effect, the optimizer $\hat{X}$ will now have non-trivial correlation with the ground truth x^*, that is $\langle \hat{X}, X^*\rangle \geqslant \Omega(n^2)$.

The last step is to turn this exponential-time algorithm into an efficient one. Fortunately, the above argument can be captured by the Sum-of-Squares proof system, thereby enabling us to use the Sum-of-Squares relaxation of program 3.5 to obtain an estimator $\hat{X}$ such that $\langle \hat{X}, X^*\rangle \geqslant \Omega(n^2)$.

Node robust algorithms for sparse graphs In the degree regime $d \leqslant d_\delta$, a simpler approach works: *remove high-degree vertices iteratively.* Although all vertices in the graph could have degree $\omega(1)$ under corruption, our strategy limits the number of removed vertices to $O(\mu n)$ by iteratively removing the highest degree node and one of its random neighbors. In this way, in each round, the number of corrupted nodes in the remaining graph is reduced by $\Omega(1)$ in expectation, meaning that the algorithm will terminate in $O(\mu n)$ rounds in expectation. As a result, the remaining graph differs from the uncorrupted graph by $O(n)$ edges, which allows us to apply the edge robust algorithm from Ding et al. (2022).

Comparison with Liu and Moitra (2022) In Liu and Moitra (2022), a weak recovery algorithm is presented that is robust to $o(n)$ node corruptions when $\varepsilon^2 d$ is sufficiently large. The algorithm conceptually resembles ours, as it also aims to identify a subgraph with desired properties. However, it falls short of reaching the KS threshold. In particular, when there are no corruptions, their algorithm is reduced to a combination of degree pruning and existing spectral algorithms (Feige and Ofek (2005)), which is not known to provide weak recovery guarantees close to the KS threshold.

4. Preliminaries and Notations

In this section, we formally define notations and cover necessary preliminaries that will be used throughout the paper.

Matrix and vector notations We use $\mathbf{1}$ to denote the all 1's vector and J to denote the all 1's matrix, i.e. $J = \mathbf{1}\,\mathbf{1}^\top$. For a vector u, we use u_i to denote its i-th entry. For a matrix M, we use M_{ij} to denote the (i,j)-th entry of M, $M \succeq 0$ to denote that M is positive semidefinite, $\mathrm{Tr}(M)$ to denote the trace of M, $\|M\|_{\mathrm{op}}$ to denote the spectral/operator norm of M and $\|M\|_{\mathrm{F}}$ to denote the Frobenius norm of M. For two matrices X and Y of the same size, we use $\odot$ to denote the Hadamard product and we define their inner product by $\langle X, Y\rangle = \sum_{i,j=0}^n X_{ij}Y_{ij} = \mathrm{Tr}(X^\top Y)$. Additionally, given a set $S \subseteq [n]$, we use v_S to denote the subvector restricted to the set S and M_S to denote the submatrix of M where we only keep entries in the set $S \times S$, that is $M_S = M \odot (\mathbf{1}_S\,\mathbf{1}_S^\top)$.

Stochastic block model notations We use $\delta = \varepsilon^2 d - 1$ to denote the distance to the KS threshold, use A^0 to denote the adjacency matrix of the uncorrupted graph G^0, use A to denote the adjacency matrix of the corrupted graph G, use $X^* = x^*(x^*)^\top$ to denote the label matrix, use S^* to denote the uncorrupted set of vertices, use $\tilde{A}^0 = A^0 - \frac{d}{n}J$ to denote the centered uncorrpted adjacency matrix and use $\tilde{A} = A - \frac{d}{n}J$ to denote the centered corrupted adjacency matrix.

Basic SDP and Grothendieck norm We define basic SDP and Grothendieck norm as follows

Definition 9 (Basic SDP) *We define basic SDP as follows*

$$\mathrm{SDP}(M) = \max\left\{\langle M, X\rangle : X \succeq 0, X_{ii} = 1 \forall i \in [n]\right\} \tag{4.1}$$

An equivalent definition (can be easily verified using eigendecomposition of X) is

$$\mathrm{SDP}(M) = \max\left\{\sum_{i,j=1}^{n} M_{ij}\langle \sigma_i, \sigma_j\rangle : \sigma_i \sim S^{n-1}\right\} \tag{4.2}$$

where S^{n-1} is the n-dimensional unit sphere.

Definition 10 (Grothendieck norm) *Let matrix function $P_\Gamma : \mathbb{R}^{n\times n} \to \mathbb{R}^{2n\times 2n}$ be defined as*

$$P_\Gamma(M) = \begin{bmatrix} 0 & M \\ 0 & 0 \end{bmatrix}$$

We define Grothendieck norm $\|\cdot\|_{Gr} : \mathbb{R}^{n\times n} \to \mathbb{R}$ as

$$\|M\|_{Gr} = \max\left\{\langle P_\Gamma(M), X\rangle : X \succeq 0, X_{ii} = 1 \forall i \in [2n]\right\} \tag{4.3}$$

An equivalent definition (the equivalence can be easily verified using eigendecomposition of X) is

$$\|M\|_{Gr} = \max\left\{\sum_{i,j=1}^{n} M_{ij}\langle \sigma_i, \delta_j\rangle : \sigma_i \sim S^{n-1}, \delta_i \sim S^{n-1}\right\} \tag{4.4}$$

where S^{n-1} is the n-dimensional unit sphere.

From definition 9 and definition 10, it is easy to get the following inequalities between the basic SDP and Grothendieck norm.

Claim 11 (Proved in section H.1) *Given matrix M, we have* $\mathrm{SDP}(M) \leqslant \|M\|_{Gr}$.

Claim 12 (Proved in section H.2) *Let M be an $n \times n$ matrix whose diagonal entries are 0 and $S \subseteq [n]$ be a subset of indices, we have* $\mathrm{SDP}(M_S) \leqslant \mathrm{SDP}(M)$.

Grothendieck inequality The celebrated Grothendieck inequality relates Grothendieck norm and the $\infty \to 1$ norm.

Definition 13 ($\infty \to 1$ norm) *Let us define $\infty \to 1$ norm $\|\cdot\|_{\infty\to1} : \mathbb{R}^{n\times n} \to \mathbb{R}$ as*

$$\|M\|_{\infty\to1} = \max\left\{\langle x, My\rangle : x, y \in \{\pm1\}^n\right\}$$

Theorem 14 (Grothendieck inequality, see Alon and Naor (2004)) *Let M be a real matrix of size $n \times n$. We have*

$$\|M\|_{\infty\to1} \leqslant \|M\|_{Gr} \leqslant K_G \|M\|_{\infty\to1}$$

where K_G is a universal constant called the Grothendieck constant.

Sum-of-Squares algorithms In this paper, we employ the Sum-of-Squares hierarchy for both algorithm design and analysis. As a broad category of semidefinite programming algorithms, Sum-of-Squares algorithms provide optimal or state-of-the-art results in algorithmic statistics, as demonstrated by numerous studies, including Hopkins and Li (2018); Kothari et al. (2018); Potechin and Steurer (2017); Hopkins (2020) (see review Barak and Steurer (2014); Raghavendra et al. (2018)).

Definition 15 (Sum-of-Squares proof) *Given a set of polynomial inequalities $\mathcal{A} = \{p_i(x) \geqslant 0\}_{i\in[m]}$ in variables $x_1, x_2, \ldots, x_n$, a sum-of-squares proof of the inequality $q(x) \geqslant 0$ is*

$$q(x) = \sum_\alpha a_\alpha^2(x)\bar{p}_\alpha(x) + \sum_\beta b_\beta^2(x)$$

where $\{a_\alpha\}$, $\{b_\beta\}$ are real polynomials and $\bar{p}_\alpha$ is a product of a subset of the polynomials in $\mathcal{A}$. It is a Sum-of-Squares proof of degree-d if all the polynomials in the summation $\left\{a_\alpha^2(x)\bar{p}_\alpha(x), b_\beta^2(x)\right\}$ have degrees no greater than d, and we denote this proof as $\mathcal{A} \left|\frac{x}{d}\right. q(x) \geqslant 0$.

Definition 16 (Sum-of-Squares refutation) *Given a set of polynomial inequalities $\mathcal{A} = \{p_i(x) \geqslant 0\}_{i\in[m]}$ in variables $x_1, x_2, \ldots, x_n$, a sum-of-squares refutation of $\mathcal{A}$ is*

$$-1 = \sum_\alpha a_\alpha^2(x)\bar{p}_\alpha(x) + \sum_\beta b_\beta^2(x)$$

where $\{a_\alpha\}$, $\{b_\beta\}$ are real polynomials and $\bar{p}_\alpha$ is a product of a subset of the polynomials in $\mathcal{A}$. It is a Sum-of-Squares proof of degree-d if all the polynomials in the summation $\left\{a_\alpha^2(x)\bar{p}_\alpha(x), b_\beta^2(x)\right\}$ have degrees no greater than d.

Definition 17 (Pseudo-expectation) *Let $\mathcal{A} = \{p_i(x) \geqslant 0\}_{i \in [m]}$ be a set of polynomial inequalities. A degree-d pseudo-expectation $\tilde{\mathbb{E}}$ for $\mathcal{A}$ is a linear operator that maps polynomials to real numbers such that:*

- $\tilde{\mathbb{E}}[1] = 1$,
- $\tilde{\mathbb{E}}[q^2(x)] \geqslant 0$ *for every polynomial q with* $\deg(q) \leqslant \frac{d}{2}$,
- $\tilde{\mathbb{E}}[q^2(x) \cdot p_i(x)] \geqslant 0$ *for every polynomial $p_i \in \mathcal{A}$ and every polynomial q with* $\deg(q) \leqslant \frac{d - \deg(p_i)}{2}$.

The following theorem reveals the key relationship between SOS proofs and SOS algorithms.

Theorem 18 (Informal restatement of Parrilo (2000); Lasserre (2001); Barak and Steurer (2014)) *For a system of polynomial inequalities $\mathcal{A}$ of size m, there is an algorithm that either finds a degree-d SOS refutation of $\mathcal{A}$ or finds a degree-d pseudo-expectation for $\mathcal{A}$ in time* $\operatorname{poly}(mn^d)$.

5. Reaching the KS threshold for diverging degree

In this section, we give an SOS algorithm when average degree d is larger than some constant d_δ which depends only on $\delta := \varepsilon^2 d - 1$.

We begin by presenting our main technical theorem, which implies theorem 4.

Theorem 19 *Let G be a graph as described in definition 3, suppose $\delta \geqslant \Omega(1)$, there exists constants $d_\delta \leqslant O(1)$ and $\mu_\delta \geqslant \Omega(1)$ which only depend on δ, such that when $d \geqslant d_\delta$ and $\mu \leqslant \mu_\delta$, there exists a polynomial-time algorithm (algorithm 20) that outputs $\hat{x} \in \{\pm 1\}^n$ satisfying*

$$\mathbb{E}\langle \hat{x}, x^* \rangle^2 \geqslant \Omega(n^2)\,.$$

Our algorithm is based on the deg-4 SOS relaxation of the following contraint set. Given a node-corrupted graph G generated according to definition 3 and its centered adjacency matrix $\tilde{A} = A - \frac{d}{n}\mathbf{1}\mathbf{1}^\top$, we consider the following system of polynomial equations in PSD matrix X of size $n \times n$ and $\{0,1\}$-vector w of size n:

$$\mathcal{A} := \left\{ \begin{array}{ll} w_i^2 = w_i & \forall i \in [n] \\ \sum_i w_i = (1 - \mu - \beta)n & \\ X \succeq 0 & \\ X_{ii} = 1 & \forall i \in [n] \\ \langle \tilde{A} \odot (ww^\top), X \rangle \geqslant (2 + \Delta)(1 - \mu - \beta)n\sqrt{d} & \\ \left\| \tilde{A} \odot (ww^\top) \right\|_{\text{op}} \leqslant C_s \sqrt{d} & \end{array} \right\} \tag{5.1}$$

Here Δ and C_s are constants depending on δ, and β is the small fraction of high degree nodes we need to prune to get bounded spectral norm according to corollary 34.

The outline of our algorithm is given below:

Algorithm 20 (Algorithm reaching KS threshold for diverging degree)
Input: *Graph G from node-corrupted SBM.*

1. *Run deg-4 SOS relaxation of program 5.1 and obtain pseudo-expectation $\tilde{\mathbb{E}}$.*
2. *Compute $\hat{X} := \tilde{\mathbb{E}}[X]$.*
3. *Apply the rounding procedure in lemma 25 on $\hat{X}$ to get estimator $\hat{x}$.*

The design and analysis of our SOS algorithm is based on the push-out effect of the basic SDP (Montanari and Sen (2016)) and spectral properties of the adjacency matrix (Feige and Ofek (2005); Chin et al. (2015); Liu and Moitra (2022)) (see section E and section F for more details). Essentially, we identify a subset of the vertices whose adjacency matrix has large enough basic SDP value and is spectrally bounded. Then, we use the spectral norm bound and the Grothendieck inequality to bound the basic SDP value of the submatrix formed by corrupted vertices in the selected subset.

5.1. Proof of correctness

Now, we present the main body of the proof and leave the rounding scheme to section B in the appendix.

Theorem 21 *Consider the constraint set in program 5.1, when $\delta \geqslant \Omega(1)$, there exists functions $d_\delta \leqslant O(1)$ and $\mu_\delta \geqslant \Omega(1)$ which only depend on δ, such that when $d \geqslant d_\delta$ and $\mu \leqslant \mu_\delta$, the following holds with probability at least $1 - o(1)$*

$$\mathcal{A} \vdash_{4}^{X,w} \langle X, X^* \rangle \geqslant \Omega(n^2)$$

We break down the proof of theorem 21 into lemma 22, lemma 23 and lemma 24. For simplicity, let us refer to the set of vertex i with $w_i = 1$ as set S, that is $S = \{i \in [n] | w_i = 1\}$.

In lemma 22, we prove the feasibility of program 5.1.

Lemma 22 (Proof deferred to section H.4) *Program 5.1 is feasible with probability $1 - o(1)$.*

Then, in lemma 23, we give a deg-4 SOS proof to show that $\langle X_{S'}, X^*_{S'} \rangle$ is large for some set S' with size at least $(1 - 2\mu - \beta)n$.

Lemma 23 *Consider set $S' = S \cap S^*$, which is the set of uncorrupted vertices in the set S found by the program. For X and w that satisfy the SOS program in eq. (5.1), we have*

$$\mathcal{A} \vdash_{4}^{X,w} \langle X_{S'}, X^*_{S'} \rangle \geqslant \frac{\Delta'(1-\beta)n^2}{\varepsilon\sqrt{d}} - O(\frac{\mu n^2}{\varepsilon\sqrt{d}})$$

where β is the small constant fraction of high degree nodes we need to prune to get bounded spectral norm according to corollary 34 and $\Delta' = \Delta'(\delta)$ for some value $\Delta'(\delta)$ that only depends on δ.

Proof We will apply the identity $\langle X_{S'}, X^*_{S'} \rangle = \langle X_{S'}, \frac{n}{\varepsilon d}\tilde{A}_{S'} \rangle - \langle X_{S'}, \frac{n}{\varepsilon d}\tilde{A}_{S'} - X^*_{S'} \rangle$ and bound the value of $\langle X_{S'}, \frac{n}{\varepsilon d}\tilde{A}_{S'} - X^*_{S'} \rangle$ and $\langle X_{S'}, \frac{n}{\varepsilon d}\tilde{A}_{S'} \rangle$ separately.

The value of $\langle X_{S'}, \frac{n}{\varepsilon d}\tilde{A}_{S'} - X^*_{S'}\rangle$ is easy to bound. From theorem 32 and union bound, we can get that, with probability $1 - o(1)$, we have

$$\langle X_{S'}, \tilde{A}_{S'} - \frac{\varepsilon d}{n} X^*_{S'}\rangle \leqslant \mathrm{SDP}(\tilde{A}_{S'} - \frac{\varepsilon d}{n} X^*_{S'}) \leqslant (2+\rho)(1-2\mu-\beta)n\sqrt{d}$$

Now, goal is to bound $\langle X_{S'}, \tilde{A}_{S'}\rangle$. We decompose it as follows

$$\langle X_{S'}, \tilde{A}_{S'}\rangle = \langle X_S, \tilde{A}_S\rangle - \langle X_S, \tilde{A}_S - \tilde{A}_{S'}\rangle \tag{5.2}$$

From the constraints of eq. (5.1), we have

$$\langle X_S, \tilde{A}_S\rangle \geqslant (2+\Delta)(1-\mu-\beta)n\sqrt{d} \tag{5.3}$$

To bound the value of $\langle X_S, \tilde{A}_S - \tilde{A}_{S'}\rangle$, we note that, by constraint $\|\tilde{A}_S\|_{\mathrm{op}} \leqslant C_s\sqrt{d}$, we can apply lemma 8 to get

$$\langle X_S, \tilde{A}_S - \tilde{A}_{S'}\rangle \leqslant \mathrm{SDP}(\tilde{A}_S - \tilde{A}_{S'}) \leqslant C'_s \mu n\sqrt{d} \tag{5.4}$$

for some constant C'_s.

Plug eq. (5.3) and eq. (5.4) into eq. (5.2), we get

$$\langle X_{S'}, \tilde{A}_{S'}\rangle = \langle X_S, \tilde{A}_S\rangle - \langle X_S, \tilde{A}_S - \tilde{A}_{S'}\rangle \geqslant (2+\Delta)(1-\mu-\beta)n\sqrt{d} - C'_s\mu n\sqrt{d}$$

Now, we can apply the identity $\langle X_{S'}, X^*_{S'}\rangle = \langle X_{S'}, \frac{n}{\varepsilon d}\tilde{A}_{S'}\rangle - \langle X_{S'}, \frac{n}{\varepsilon d}\tilde{A}_{S'} - X^*_{S'}\rangle$ and get

$$\begin{aligned}\langle X_{S'}, X^*_{S'}\rangle =& \langle X_{S'}, \frac{n}{\varepsilon d}\tilde{A}_{S'}\rangle - \langle X_{S'}, \frac{n}{\varepsilon d}\tilde{A}_{S'} - X^*_{S'}\rangle \\ \geqslant& \frac{n}{\varepsilon d}\Big((2+\Delta)(1-\mu-\beta)n\sqrt{d} - C'_s\mu n\sqrt{d}\Big) - \frac{n}{\varepsilon d}(2+\rho)(1-2\mu-\beta)n\sqrt{d} \\ \geqslant& \frac{\Delta'(1-\beta)n^2}{\varepsilon\sqrt{d}} - O(\frac{\mu n^2}{\varepsilon\sqrt{d}})\end{aligned}$$

■

Finally, since X is positive semidefinite and $X_{ii} = 1$ for all $i \in [n]$, we can conclude that there is a deg-4 SOS proof to show that correlation $\langle X, X^*\rangle$ is large.

Lemma 24 (Proof deferred to section H.5) *For X and w that satisfy the SOS program in eq. (5.1), we have*

$$\mathcal{A} \left|\frac{X,w}{4}\right. \langle X, X^*\rangle \geqslant \frac{\Delta'(1-\beta)n^2}{\varepsilon\sqrt{d}} - O(\frac{\mu n^2}{\varepsilon\sqrt{d}}) - 2\beta n^2$$

where β is the small constant fraction of high degree nodes we need to prune to get bounded spectral norm according to corollary 34 and $\Delta' = \Delta'(\delta)$ for some value $\Delta'(\delta)$ that only depends on δ.

Now, we have all the ingredients to prove theorem 21

Proof [Proof of theorem 21] From lemma 22, we know that the SOS program in eq. (5.1) is feasible with probability $1 - o(1)$. Combine this with lemma 24, we know that, with probability $1 - o(1)$, the SOS program in eq. (5.1) finds X and w such that they satisfy

$$\mathcal{A} \vdash^{X,w}_{4} \langle X, X^* \rangle \geqslant \frac{\Delta'(1-\beta)n^2}{\varepsilon\sqrt{d}} - O(\frac{\mu n^2}{\varepsilon\sqrt{d}}) - 2\beta n^2$$

for some β that is the small constant fraction of high degree nodes we need to prune to get bounded spectral norm according to corollary 34 and $\Delta' = \Delta'(\delta)$ for some value $\Delta'(\delta)$ that only depends on δ.

When $\mu \leqslant \mu_\delta$ for some value μ_δ that only depends on δ and $\beta = \beta(\delta)$ for some value $\beta(\delta)$ that only depends on δ, we have:

$$\mathcal{A} \vdash^{X,w}_{4} \langle X, X^* \rangle \geqslant \frac{\Delta'(1-\beta)n^2}{\varepsilon\sqrt{d}} - O(\frac{\mu n^2}{\varepsilon\sqrt{d}}) - 2\beta n^2 = \theta(\delta)n^2$$

for some $\theta(\delta)$ that only depends on δ. Thus, when $\delta \geqslant \Omega(1)$, we can get the weak recovery guarantee:

$$\mathcal{A} \vdash^{X,w}_{4} \langle X, X^* \rangle \geqslant \Omega(n^2)$$

■

In order to fully establish the validity of theorem 19, it remains to apply the standard rounding procedure from Hopkins and Steurer (2017) on the pseudo-expectation of matrix X (as depicted in algorithm 20). We will address this part in section B.

Acknowledgments

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 815464). Yiding Hua and Jingqiu Ding are grateful to Afonso S. Bandeira for very helpful discussions.

References

Emmanuel Abbe. Community detection and stochastic block models: recent developments. *The Journal of Machine Learning Research*, 18(1):6446–6531, 2017.

Emmanuel Abbe, Enric Boix-Adsera, Peter Ralli, and Colin Sandon. Graph powering and spectral robustness. *SIAM Journal on Mathematics of Data Science*, 2(1):132–157, 2020.

Jayadev Acharya, Ayush Jain, Gautam Kamath, Ananda Theertha Suresh, and Huanyu Zhang. Robust estimation for random graphs. In Po-Ling Loh and Maxim Raginsky, editors, *Proceedings of Thirty Fifth Conference on Learning Theory*, volume 178 of *Proceedings of Machine Learning Research*, pages 130–166. PMLR, 02–05 Jul 2022. URL https://proceedings.mlr.press/v178/acharya22a.html.

Noga Alon and Assaf Naor. Approximating the cut-norm via grothendieck's inequality. In *Proceedings of the thirty-sixth annual ACM symposium on Theory of computing*, pages 72–80, 2004.

Jinho Baik, Gérard Ben Arous, and Sandrine Péché. Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices. 2005.

Debapratim Banerjee. Contiguity and non-reconstruction results for planted partition models: the dense case. 2018.

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

T. Tony Cai and Xiaodong Li. Robust and computationally feasible community detection in the presence of arbitrary outlier nodes. *ArXiv*, abs/1404.6000, 2014.

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.

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):066106, 2011.

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.

Uriel Feige and Eran Ofek. Spectral techniques applied to sparse random graphs. *Random Structures & Algorithms*, 27(2):251–275, 2005.

Jun He and Xin Yao. A study of drift analysis for estimating computation time of evolutionary algorithms. *Natural Computing*, 3(1):21–35, 2004.

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*, pages 1021–1034, 2018.

Samuel B. Hopkins and David Steurer. Efficient bayesian estimation from few samples: Community detection and related problems. In *58th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2017, Berkeley, CA, USA, October 15-17, 2017*. IEEE Computer Society, 2017.

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.

Jean B Lasserre. Global optimization with polynomials and the problem of moments. *SIAM Journal on optimization*, 11(3):796–817, 2001.

Allen Liu and Ankur Moitra. Minimax rates for robust community detection. *arXiv preprint arXiv:2207.11903*, 2022.

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.

Ankur Moitra, William Perry, and Alexander S Wein. How robust are reconstruction thresholds for community detection? In *Proceedings of the forty-eighth annual ACM symposium on Theory of Computing*, pages 828–841, 2016.

Andrea Montanari and Subhabrata Sen. Semidefinite programs on sparse random graphs and their application to community detection. In *Proceedings of the forty-eighth annual ACM symposium on Theory of Computing*, pages 814–827, 2016.

Elchanan Mossel, Joe Neeman, and Allan Sly. Belief propagation, robust reconstruction and optimal recovery of block models. In *Conference on Learning Theory*, pages 356–370. PMLR, 2014.

Elchanan Mossel, Joe Neeman, and Allan Sly. Reconstruction and estimation in the planted partition model. *Probability Theory and Related Fields*, 162(3):431–461, 2015.

Pablo A Parrilo. *Structured semidefinite programs and semialgebraic geometry methods in robustness and optimization*. PhD thesis, California Institute of Technology, 2000.

Amelia Perry, Alexander S Wein, Afonso S Bandeira, and Ankur Moitra. Optimality and sub-optimality of pca i: Spiked random matrix models. *The Annals of Statistics*, 46(5):2416–2451, 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, 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.

Ludovic Stephan and Laurent Massoulié. Robustness of spectral methods for community detection. In *Conference on Learning Theory*, pages 2831–2860. PMLR, 2019.

Appendix A. Appendix organization

In section B, we complete the proof of theorem 19 by adding a rounding scheme. In section C, we give an algorithm that reduces node corruption to edge corruption in the sparse degree regime via degree pruning. In section G, we use similar techniques to obtain a polynomial-time algorithm for the row/column corrupted $\mathbb{Z}_2$ synchronization problem. In section D, we show that it is impossible to achieve weak recovery when the corruption fraction μ is larger than δ, therefore, μ has to be a value that depends on δ. In section E and section F, we present a small summary of previous results on the pushout effect of basic SDP and spectral bounds of degree-pruned adjacency matrix. In section H, we present proofs that are deferred to the appendix due to page limit constraint.

Appendix B. Rounding

Now, we complete the proof of theorem 19 by giving a rounding procedure adapted from Lemma 3.5 of Hopkins and Steurer (2017) [6].

Lemma 25 (Rounding procedure adapted from Lemma 3.5 of Hopkins and Steurer (2017)) *Let $\theta = \frac{1}{\|X\|_F n}\langle X, X^*\rangle$. Let Y be a matrix of minimum Frobenious norm such that $Y \succeq 0$, $\operatorname{diag} Y = 1$ and $\frac{1}{\|X\|_F n}\langle Y, X\rangle \geqslant \theta$. With probability $1 - o(1)$, the vector $\hat{x}$ obtained by taking coordiate-wise sign of a Gaussian vector with mean 0 and covariance Y satisfies*

$$\mathbb{E}[\langle \hat{x}, x^*\rangle^2] \geqslant \Omega(\theta)^2 n^2$$

Proof Apply Lemma 3.5 of Hopkins and Steurer (2017) by taking $P = X$, $y = x^*$ and $\delta' = \theta$, we can get

$$\mathbb{E}[\langle \hat{x}, x^*\rangle^2] \geqslant \Omega(\theta)^2 n^2$$

Notice that, because each entry of X is within ±1, we have $\|X\| \leqslant n$. Since $\langle X, X^*\rangle \geqslant \Omega(n^2)$ by theorem 21, we have $\theta = \Omega(1)$. Thus, $\hat{x}$ weakly recovers x^*. ■

Now we finish the proof of theorem 19.

Proof [Proof of theorem 19] By combining theorem 21 and theorem 18, we can compute the pseudo-expectation $\tilde{\mathbb{E}}$ for the SOS relaxtion of eq. (5.1) in polynomial time. Let $\hat{X} := \tilde{\mathbb{E}}[X]$ in eq. (5.1). By linearity of pseudo-expectation, we have $\hat{X} \succeq 0$, $\hat{X}_{ii} = 1$ and $\langle \hat{X}, X^*\rangle \geqslant \Omega(n^2)$ with probability $1 - o(1)$. Now applying rounding procedure in lemma 25, we can then obtain $\hat{x} \in \{\pm 1\}^n$ such that $\mathbb{E}\langle \hat{x}, x^*\rangle^2 \geqslant \Omega(n^2)$. ■

Appendix C. Reaching KS threshold for constant degree

In this section, we give an algorithm that reduces node corruption to edge corruption when $d < d_\delta$. This allows us to deal with graphs with small average degree.

C.1. Edge-robust algorithm

Before introducing our algorithm for the constant degree region, we restate the main theorem of Ding et al. (2022) here. Their main theorem shows that there exists a polynomial-time algorithm that is robust against $O(\rho n)$ edge perturbations, where ρ is a constant that depends on δ.

Theorem 26 (Informal restatement of Corollary 5.4 of Ding et al. (2022)) *Given a graph $G \sim \mathsf{SBM}_{d,\varepsilon}(x^*)$, suppose G' is an arbitrary graph that differs from G in at most $O(\rho n)$ edges for*

$$\rho \leqslant \Big(\frac{1}{\delta}\log\frac{1}{\varepsilon}\Big)^{-O(1/\delta)}$$

Then, there exists a polynomial-time algorithm that, given G' and δ, computes an n-dimensional unit vector $\hat{x}$ such that

$$\mathbb{E}[\langle \hat{x}, x^*\rangle^2] \geqslant \delta^{O(1)} n$$

[6] Other rounding procedures, such as random Gaussian rounding, also work here.

C.2. Degree-pruning based algorithm

The algorithm is based on degree pruning. The tricky part is that node corruption can arbitrarily increase the degree of uncorrupted vertices to μn. Therefore, simply pruning high-degree vertices can be quite difficult to analyse.

Our solution is to iteratively remove the highest degree node as well as one of its neighbours that is selected uniformly at random until all vertices have small enough degree. The goal is to make sure that, in each round, we remove $\Omega(1)$ corrupted vertices in expectation.

Notice that, in each round, if the highest degree node is corrupted, then it is good. If the highest degree node is uncorrupted, then we can show, with high probability, the majority of its neighbours are corrupted vertices and we are likely to remove a corrupted vertex if we select one of its neighbours uniformly at random. A key observation is that, this approach allows us to easily bound the total number of removed vertices using a simple and standard Markov Chain drift analysis.

After the degree pruning procedure, we will invoke the edge-robust algorithm from Ding et al. (2022) that is restated in theorem 26.

Algorithm 27 (Algorithm reaching KS threshold for constant degree)
Input: *A node-corrupted stochastic block model G.*

```
1. Set $G' \leftarrow G$
2. While there exist vertices with degree larger than $C_{\deg}(\mu)d$ in $G'$:
     – remove the highest-degree vertex $v$ from $G'$,
     – remove from $G'$ a neighbour $u$ of $v$ that is selected uniformly at random.
3. Run edge-robust algorithm from theorem 26 on the remaining graph $G'$.
4. Apply the rounding procedure in lemma 25 to get estimator $\hat{x}$.
```

In the following theorem, we will show that algorithm 27 outputs an estimator $\hat{x}$ that achieves weak recovery.

Theorem 28 *When $d < d_\delta$ and $\delta = \Omega(1)$, for some $C_{deg}(\mu)$ that only depends on μ, algorithm 27 outputs a vector $\hat{x} \in \{\pm 1\}^n$ such that*

$$\mathbb{E}[\langle \hat{x}, x^* \rangle^2] \geqslant \Omega(n^2)$$

Moreover, algorithm 27 runs in polynomial time.

C.3. Proof of correctness

To prove theorem 28, we will use the following two lemmas: lemma 29 and lemma 30. First, we prove lemma 29 which says that, with probability 0.99, the pruning step of algorithm 27 terminates in $O(\mu n)$ rounds. Then, in lemma 30, we prove that, with probability 0.99, algorithm 27 produces a graph G' that differs from G^0 by at most $O(\rho n)$ edges, such that we can apply theorem 26 on G' to get an estimator $\hat{x}$ that achieves weak recovery.

Lemma 29 *With probability at least 0.99, for some $C_{deg}(\mu)$ that only depends on μ, step 2 of algorithm 27 terminates in $O(\mu n)$ rounds.*

Proof Let S denote the set of uncorrupted vertices and let $G[S]$ denote the induced subgraph of the uncorrupted vertices. For vertices with degree more than $C_{deg}(\mu)d$ in G, we separate them into three cases:

1. corrupted vertices,
2. uncorrupted vertices with degree larger than or equal to $\frac{1}{2}C_{deg}(\mu)d$ in $G[S]$,
3. uncorrupted vertices with degree smaller than $\frac{1}{2}C_{deg}(\mu)d$ in $G[S]$.

We will prove that, with probability at least 0.99, all three cases can be eliminated in $O(\mu n)$ rounds. Therefore, with probability 0.99, step 2 of algorithm 27 terminates in $O(\mu n)$ rounds.

Case 1: Since there are at most μn corrupted vertices, it takes at most μn rounds to deal with corrupted vertices with degree more than $C_{deg}(\mu)d$ in G.

Case 2: For uncorrupted vertices with degree larger than or equal to $\frac{1}{2}C_{deg}(\mu)d$ in $G[S]$ and degree more than $C_{deg}(\mu)d$ in G, we bound it by the total number of vertices with degree larger than or equal to $\frac{1}{2}C_{deg}(\mu)d$ in G^0. By Chernoff Bound, we have that, for each vertex v, the probability that v has degree more than $\frac{1}{2}C_{deg}(\mu)d$ in G^0 is roughly bounded by

$$\mathbb{P}[\deg_{G^0}(v) \geqslant \frac{1}{2}C_{deg}(\mu)d] \leqslant O(\exp(-\frac{1}{2}C_{deg}(\mu)d))$$

Let T be the set of vertices with degree larger than or equal to $\frac{1}{2}C_{deg}(\mu)d$ in G^0. By Markov's inequality, we get that the probability that T contains more than μn vertices is roughly bounded by

$$\mathbb{P}[|T| \geqslant \mu n] \leqslant O(\frac{\exp(-\frac{1}{2}C_{deg}(\mu)d)}{\mu})$$

By setting $C_{deg}(\mu)$ to be large enough with respect to μ, we get that, with probability 0.999, there are at most μn vertices with degree larger than or equal to $\frac{1}{2}C_{deg}(\mu)d$ in G^0. Therefore, it takes at most μn rounds to remove the vertices that are uncorrupted and has degree more than $\frac{1}{2}C_{deg}(\mu)d$ in $G[S]$.

Case 3: For uncorrupted vertices that have degree smaller than $\frac{1}{2}C_{deg}(\mu)d$ in $G[S]$ but have degree more than $C_{deg}(\mu)d$ in G, the key observation is that more than half of their neighbours are corrupted vertices. Therefore, each time such a node is removed as the highest degree node, with probability more than $1/2$, the algorithm will remove a corrupted node as its random neighbour.

Now, let us only consider the rounds where the highest degree node is in case 3 and let t denote the total number of such rounds. Let X_i denote the number of corrupted vertices removed after round i and $X_0 = 0$. We know that $X_{i+1} = X_i + 1$ with probability more than $1/2$ and $X_{i+1} = X_i$ otherwise. We also know that the process has to terminate when $X_t = \mu n$. Therefore, by the standard Markov Chain drift analysis (see Lemma 1 in He and Yao (2004)), we have:

$$\mathbb{E}[t] \leqslant 2\mu n$$

and, by Markov inequality, the probability that vertices in case 3 are not eliminated after $1000\mu n$ rounds where the highest degree node is in case 3 is bounded by

$$\mathbb{P}[t \geqslant 1000\mu n] \leqslant \frac{2\mu n}{1000\mu n} = 0.002$$

Therefore, with probability at least 0.998, vertices in case 3 are eliminated in $1000n$ rounds.

Conclusion Taking union bound over the failure probabilities, we get that, with probability at least 0.99, step 2 algorithm 27 terminates in $1002\mu n = O(\mu n)$ rounds. ■

Notice that lemma 29 gives us an upper bound on the total number of removed vertices during the pruning step and algorithm 27 guarantees that G' will have bounded degree after pruning. These two observations allow us to have the following lemma, which says that the difference between G' and G^0 is at most $O(\rho n)$ edges, where $O(\rho n)$ is the number of edges that can be tolerated by the edge-robust algorithm in theorem 26.

Lemma 30 *Let G^0 be the uncorrupted graph, with probability 0.99, the remaining graph G' in step 3 of algorithm 27 differs from G^0 by $O(\rho n)$ edges, where $\rho \leqslant \left(\frac{1}{\delta}\log\frac{1}{\varepsilon}\right)^{-O(1/\delta)}$ as defined in theorem 26.*

Proof Graph G' differs from G^0 by two types of edges:

1. corrupted edges in G',
2. uncorrupted edges that are removed from pruning.

We will bound the two cases separately.

Case 1 algorithm 27 guarantees that the degree of each vertex in G' is bounded by $C_{\text{deg}}(\mu)d$. Since there are at most μn corrupted vertices in G', the maximum number of corrupted edges in G' is $C_{\text{deg}}(\mu)\mu dn$. Since $d < d_\delta$ for some d_δ, we can set $C_{\text{deg}}(\mu)$ to be small enough such that $C_{\text{deg}}(\mu)\mu dn \leqslant O(\rho n)$.

Case 2 For case 2, we consider two types of vertices that are removed in algorithm 27:

– vertices with degree smaller than or equal to $C_{\text{deg}}(\mu)d$ in G^0,

– vertices with degree larger than $C_{\text{deg}}(\mu)d$ in G^0.

For the first type of vertices, we observe that, with probability 0.99, algorithm 27 terminates in $O(\mu n)$ rounds by lemma 29. Therefore, with probability 0.99, there can be at most $O(\mu n)$ vertices in this case. Hence, the number of uncorrupted edges that are removed from pruning the first type of vertices can be bounded by $O(C_{\text{deg}}(\mu)\mu dn)$. Similar to case 1, we can set $C_{\text{deg}}(\mu)$ properly such that $O(C_{\text{deg}}(\mu)\mu dn) \leqslant O(\rho n)$.

For the second type of vertices, we know that, for each vertex v, the probability that v has degree larger than or equal to t is bounded by

$$\mathbb{P}[\deg_{G^0}(v) \geqslant t] \leqslant O(\exp(-t))$$

Let X_t denote the number of vertices with degree t in G^0. We have

$$\mathbb{E}[X_t] \leqslant \mathbb{P}[\deg_{G^0}(v) \geqslant t] \cdot n \leqslant O(\exp(-t)n)$$

Therefore, for some properly selected $C_{\text{deg}}(\mu)$, the expected total number of edges from vertices with degree larger than $C_{\text{deg}}(\mu)d$ in G^0 can be bounded by

$$\mathbb{E}\left[\sum_{t=C_{\text{deg}}(\mu)d}^{\infty} X_t t\right] = \sum_{t=C_{\text{deg}}(\mu)d}^{\infty} \mathbb{E}[X_t]t \leqslant O\left(\sum_{t=C_{\text{deg}}(\mu)d}^{\infty} \exp(-t)tn\right)$$

By setting $C_{\deg}(\mu)$ to be a large enough value, we have

$$\mathbb{E}\left[\sum_{t=C_{\deg}(\mu)d}^{\infty} X_t t\right] \leqslant A\rho n$$

for some universal constant A. By Markov inequality, with probability 0.99, the number of uncorrupted edges that are removed from pruning second type of vertices can be bounded by $O(\rho n)$.

Conclusion Take union bound over failure probabilities, we get that, with probability 0.98, G' differs from G^0 by $O(\rho n)$ edges. ■

Now, we prove theorem 28 using lemma 29 and lemma 30.

Proof [Proof of theorem 28] First, we prove recovery guarantees of algorithm 27. From lemma 30, we know that, with probability 0.98, G' differs from G^0 by $O(\rho n)$ edges. Combine the guarantees of theorem 26 and the rounding procedure in lemma 25, step 3 will output an estimator $\hat{x} \in \{\pm 1\}^n$ such that

$$\mathbb{E}[\langle \hat{x}, x^* \rangle^2] \geqslant 0.98 \cdot \Omega(n^2) = \Omega(n^2)$$

Now, we prove the time complexity of algorithm 27. For step 2 of the algorithm, each round takes at most $O(n^2)$ time and there can be at most n rounds. Therefore, step 2 takes at most $O(n^3)$ time. For step 3, the edge-robust algorithm from theorem 26 takes polynomial time. For step 4, the rounding procedure from lemma 25 takes polynomial time. Therefore, algorithm 27 runs in polynomial time. ■

Appendix D. Lower bound on the corruption fraction

As stated in theorem 4, our algorithm is robust against $\Omega_\delta(1)$ fraction of corrupted nodes. One might wonder whether we can remove the dependency on $\delta = \varepsilon^2 d - 1$, and find an algorithm robust against $\Omega(1)$ fraction of corrupted nodes (e.g 0.001 fraction of corrupted nodes). The following claim shows that this is impossible.

Claim 31 *Let $n > 1, d > 1, \varepsilon \in (0,1)$ and label vector $x^* \in \{\pm 1\}^n$. Let $\delta := \varepsilon^2 d - 1$ and suppose $\delta \geqslant \Omega(1)$. For $G^0 \sim \mathsf{SBM}_{d,\varepsilon}(x^*)$ and $\mu \geqslant \delta$, if an adversary removes μn vertices uniformly at random from G^0 to obtain the graph G that we observe, then it is information theoretically impossible to achieve weak recovery given G, i.e. for any estimator $\hat{x}(G) \in \{\pm 1\}^n$, we have*

$$\mathbb{E}\langle \hat{x}(G), x^* \rangle^2 \leqslant o(n^2)\,.$$

Proof Let us denote the set of remaining vertices as R. Note that the remaining graph follows distribution $\mathsf{SBM}_{d',\varepsilon'}(x^*_R)$, where $d' = (1-\mu)\cdot d$ and $\varepsilon' = \varepsilon$. When $\mu \geqslant \delta$, we have $\varepsilon'^2 d' \leqslant (1-\delta)^2 \cdot (1+\delta) \leqslant 1$. According to Mossel et al. (2015), it is information theoretically impossible to achieve weak recovery when $\varepsilon'^2 d' \leqslant 1$. Thus, it is information theoretically impossible to recover x^*. ■

Appendix E. Push-out effect of basic SDP

In this section, we present theorem 32 that captures the push-out effect of the basic SDP value of uncorrupted stochastic block model. This theorem is based on Theorem 5 and Theorem 8 of Montanari and Sen (2016) and is stated in a way that is easier for us to use in our analysis. It is intensively used in section 5, where we prove weak recovery guarantees of our SOS algorithm.

Theorem 32 (Restatement of Theorem 5 and Theorem 8 of Montanari and Sen (2016)) *Given a graph $G \sim \mathsf{SBM}_{d,\varepsilon}(x^*)$, there exists $C = C(\delta)$ and d_δ that only depend on $\delta = \varepsilon^2 d - 1$ such that when $d \geqslant d_\delta$, with probability at least $1 - Ce^{-n/C}$, we have*

$$\mathrm{SDP}(\tilde{A}) \geqslant (2+\Delta)n\sqrt{d}$$

and,

$$\mathrm{SDP}(\tilde{A} - \frac{\varepsilon d}{n}X^*) \leqslant (2+\rho)n\sqrt{d}$$

where $\rho = \frac{C\log d}{d^{1/10}}$ and $\Delta = \Delta(\delta)$ for some value $\Delta(\delta)$ that only depends on δ.

Appendix F. Spectral bound of degree-pruned submatrix

In this section, we use the following result to show that we can prune out a small fraction of the high degree vertices to get a spectrally bounded submatrix of the centered adjacency matrix.

Theorem 33 (Restatement of Feige and Ofek (2005); Chin et al. (2015); Liu and Moitra (2022)) *Suppose M is a random symmetric matrix with zero on the diagonal whose entries above the diagonal are independent with the following distribution*

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

Let α be a quantity such that $p_{ij} \leqslant \frac{\alpha}{n}$ and M_1 be the matrix obtained from M by zeroing out all the rows and columns having more than 20α positive entries. Then with probability $1 - \frac{1}{n^2}$, we have

$$\|M_1\|_{\mathrm{op}} \leqslant \chi\sqrt{\alpha}$$

for some constant χ.

From theorem 33, we can get the following spectral bound for degree-pruned adjacency matrix.

Corollary 34 *In the setting of definition 1, with probability at least $1 - o(1)$, there exists a subset $T \subseteq [n]$ of size at least $(1-\beta)n$ such that*

$$\left\|\tilde{A}_T\right\|_{\mathrm{op}} \leqslant C_s\sqrt{d}$$

where C_s is some constant and $\beta = \beta(\delta)$ is a value that only depends on δ.

Proof For simplicity, let us set a to be $a = (1+\varepsilon)d$ and set $t = \beta n$. We apply theorem 33 by setting $\alpha > a$ to be a large enough constant. The probability that there exists more than βn vertices with degree at least 20α is at most

$$\binom{n}{t}\binom{tn}{10\alpha t}\left(\frac{a}{n}\right)^{10\alpha t} \leqslant \left(\frac{en}{t}\right)^t\left(\frac{en}{10\alpha}\right)^{10\alpha t}\left(\frac{a}{n}\right)^{10\alpha t} = \left(\frac{en}{t}\right)^t\left(\frac{ea}{10\alpha}\right)^{10\alpha t}$$

Since $\alpha > a$, we have

$$\left(\frac{en}{t}\right)^t\left(\frac{ea}{10\alpha}\right)^{10\alpha t} \leqslant \left(\frac{en}{t}\right)^t\left(\frac{e}{10}\right)^{10\alpha t} \leqslant e^{-10\alpha t + t(\log(n/t)+1)}$$

Plug in $t = \beta n$, we get the failure probability is $e^{-10\alpha\beta n + \beta n(\log(1/\beta)+1)}$. As long as $-10\alpha + \log(1/\beta) + 1 < 0$ for some α and β, the failure probability is $o(1)$. Take union bound with failure probability of theorem 33, we get that, with probability $1 - o(1)$, we have

$$\left\|\left(\tilde{A} - \frac{\varepsilon d}{n}X^*\right)_T\right\|_{\text{op}} \leqslant \chi\sqrt{\alpha}$$

Since $\alpha > a > d$, we have

$$\left\|\left(\tilde{A} - \frac{\varepsilon d}{n}X^*\right)_T\right\|_{\text{op}} \leqslant C_s'\sqrt{d}$$

for some constant C_s'. Notice that $\left\|(\frac{\varepsilon d}{n}X^*)_T\right\|_{\text{op}} \leqslant \varepsilon d$. Apply triangle inequality, we get

$$\left\|\tilde{A}_T\right\|_{\text{op}} \leqslant C_s'\sqrt{d} + \varepsilon d$$

When $\varepsilon\sqrt{d} = O(1)$, we get $\varepsilon d = O(\sqrt{d})$. Hence, with probability $1 - o(1)$, we have

$$\left\|\tilde{A}_T\right\|_{\text{op}} \leqslant C_s\sqrt{d}$$

for some constant C_s. ■

Appendix G. Robust $\mathbb{Z}_2$ Synchronization

In this section, we give an algorithm to solve the row/column-corrupted $\mathbb{Z}_2$ synchronization problem using techniques from section 5. The idea is similar to the robust SOS algorithm for node-corrupted stochastic block model: we find a subset of the rows/columns such that the submatrix formed by the subset has large enough basic SDP value and bounded spectral norm. Then, we use the spectral norm bound to upper bound the basic SDP value of the submatrix formed by corrupted rows/columns in the selected subset.

G.1. Phase transition for $\mathbb{Z}_2$ synchronization

Before introducing our algorithm, we give a small summary of the basic SDP value phase transition for $\mathbb{Z}_2$ synchronization. It is based on Theorem 5 of Montanari and Sen (2016), where they gave a very clean result for the phase transition of deformed GOE matrices. The phase transition can naturally be extended to the $\mathbb{Z}_2$ synchronization model using a simple argument based on rotational symmetry. The following theorem informally restates Theorem 5 of Montanari and Sen (2016) and provides the result we need for our robust $\mathbb{Z}_2$ synchronization algorithm.

Theorem 35 ($\mathbb{Z}_2$ synchronization phase transition Montanari and Sen (2016)) *Given an uncorrupted $\mathbb{Z}_2$ synchronization matrix A^0 that is generated acoodirng to definition 5,*

- *if $\sigma \in [0,1]$, then for any $\xi > 0$, we have $\mathrm{SDP}(A^0) \in [(2-\xi)n^2, (2+\xi)n^2]$ with probability $1 - o(1)$,*
- *if $\sigma > 1$, then there exists $\Delta(\sigma) > 0$ such that $\mathrm{SDP}(A^0) \geqslant (2 + \Delta(\sigma))n^2$ with probability $1 - o(1)$.*

G.2. SOS Algorithm

In this corruption model, the $\mathbb{Z}_2$ synchronization is easier than the stochastic block model in the sense that, with high probability, A^0 is already bounded in spectral norm. Therefore, we can omit the pruning step that we did for robust stochastic block model. Let A be the row/column corrupted Z_2 synchronization matrix generated according to definition 5, consider the following system of polynomial equations in PSD matrix X of size $n \times n$ and vector w of size n:

$$\mathcal{A} := \left\{ \begin{array}{ll} w_i^2 = w_i & \forall i \in [n] \\ \sum_i w_i = (1-\mu)n & \\ X \succeq 0 & \\ X_{ii} = 1 & \forall i \in [n] \\ \langle A \odot (ww^\top), X \rangle \geqslant (2 + \Delta(\sigma))(1-\mu)^2 n^2 & \\ \left\| A \odot (ww^\top) \right\|_{\mathrm{op}} \leqslant (\sigma + \sigma^{-1})n & \end{array} \right\} \tag{G.1}$$

where $\Delta(\sigma) > 0$ is value that only depends on σ.

In eq. (G.1), the vector w is a $\{0,1\}$ vector that finds a subset of the rows/columns whose submatrix behaves like uncorrupted $\mathbb{Z}_2$ synchronization matrix. Essentially, we require the submatrix $A \odot (ww^\top)$ to have two properties: *(a)* it has large enough basic SDP value and *(b)* it has small enough spectral norm. The matrix X is a PSD matrix that is a solution to the basic SDP, such that the inner product between X and $A \odot (ww^\top)$ is large enough. This certifies property (a) and also allows us to obtain our estimator X. Property (b) is easy to show because there is an SOS certificate for spectral norm.

We will prove the following theorem for the SOS relaxation of eq. (G.1), which implies theorem 6. It says that, with high probability, there is a deg-4 SOS proof which shows that any X which satisfies eq. (G.1) has non-trivial correlation with the true labels X^*.

Theorem 36 (SOS proof for robust $\mathbb{Z}_2$ synchronization) *When $\sigma > 1$ and $\mu \leqslant \mu^*(\sigma)$ for some value $\mu^*(\sigma)$ that only depends on σ, with probability at least $1 - o(1)$, we have:*

$$\mathcal{A} \,\Big|\frac{X,w}{4}\; \langle X, X^* \rangle \geqslant \Omega(n^2)$$

G.3. Proof of correctness

To prove theorem 36, we need to show two things:

1. $\mathcal{A} \Big|\frac{X,w}{4} \langle X, X^* \rangle \geqslant \Omega(n^2)$ with high probability,
2. the constraint set eq. (G.1) is feasible with high probability.

Now, we prove the first property in the following lemma.

Lemma 37 *For X and w satisfying eq. (G.1), we have, with probability $1 - o(1)$,*

$$\mathcal{A} \Big|\frac{X,w}{4} \langle X, X^* \rangle \geqslant \Omega(n^2)$$

Proof Let $s \in \{0,1\}^n$ be the indicator variable for the set of uncorrupted indices. We will use the following identity to prove the lemma

$$\langle X, X^* \rangle = \langle X, X^* \odot (ww^\top) \odot (ss^\top) \rangle + \langle X, X^* \odot (J - (ww^\top) \odot (ss^\top)) \rangle \tag{G.2}$$

Notice that we have $\mathcal{A} \Big|\frac{X,w}{4} X_{ij}^2 \leqslant 1$ for each $(i,j) \in [n] \times [n]$. Therefore, we can get the following bound for the second term of eq. (G.2)

$$\mathcal{A} \Big|\frac{X,w}{4} \langle X, X^* \odot (J - (ww^\top) \odot (ss^\top)) \rangle \geqslant -4\mu n^2 \tag{G.3}$$

Now, the goal is to show that $\langle X, X^* \odot (ww^\top) \odot (ss^\top) \rangle$ is large enough. To do this, we will use the following identity

$$\langle X, X^* \odot (ww^\top) \odot (ss^\top) \rangle = \frac{1}{\sigma}(\langle X, A \odot (ww^\top) \odot (ss^\top) \rangle - \langle X, (A - \sigma X^*) \odot (ww^\top) \odot (ss^\top) \rangle) \tag{G.4}$$

The easy part is to bound $\langle X, (A - \sigma X^*) \odot (ww^\top) \odot (ss^\top) \rangle$. We know that $(A - \sigma X^*) \odot (ww^\top) \odot (ss^\top) = (A^0 - \sigma X^*) \odot (ww^\top) \odot (ss^\top)$ since it is restricted to the set of uncorrupted rows/columns. Moreover, from theorem 35, we know that, with high probability, $\text{SDP}(A^0 - \sigma X^*) \leqslant (2 + \xi)n^2$ for any constant $\xi > 0$. Therefore, by monotonicity of the basic SDP from claim 12, we have:

$$\begin{aligned}
\mathcal{A} \Big|\frac{X,w}{4} & \langle X, (A - \sigma X^*) \odot (ww^\top) \odot (ss^\top) \rangle \\
&= \langle X, (A^0 - \sigma X^*) \odot (ww^\top) \odot (ss^\top) \rangle \\
&\leqslant \text{SDP}((A^0 - \sigma X^*) \odot (ww^\top) \odot (ss^\top)) \\
&\leqslant \text{SDP}(A^0 - \sigma X^*) \\
&\leqslant (2 + \xi)n^2
\end{aligned}$$

The hard part is to show that $\langle X, A \odot (ww^\top) \odot (ss^\top) \rangle$ is large enough. We show this via the following identity:

$$\langle X, A \odot (ww^\top) \odot (ss^\top) \rangle = \langle X, A \odot (ww^\top) \rangle - \langle X, A \odot (ww^\top) - A \odot (ww^\top) \odot (ss^\top) \rangle$$

For the first term $\langle X, A \odot (ww^\top) \rangle$, we can simply apply the program constraint and get:

$$\mathcal{A} \Big|\frac{X,w}{4} \langle X, A \odot (ww^\top) \rangle \geqslant (2 + \Delta(\sigma))(1 - \mu)^2 n^2$$

For the second term $\langle X, A\odot(ww^\top) - A\odot(ww^\top)\odot(ss^\top)\rangle$, we bound it by the Grothendieck norm of $A\odot(ww^\top) - A\odot(ww^\top)\odot(ss^\top)$. Since we have $\mathcal{A}\left|\frac{X,w}{4}\right. \left\|A\odot(ww^\top)\right\|_{\text{op}} \leqslant (\sigma+\sigma^{-1})n$ from the program constraints, we can apply lemma 8 to get

$$\mathcal{A}\left|\frac{X,w}{4}\right. \langle X, A\odot(ww^\top) - A\odot(ww^\top)\odot(ss^\top)\rangle \leqslant O((\sigma+\sigma^{-1})\mu n^2)$$

Thus, we have:

$$\mathcal{A}\left|\frac{X,w}{4}\right. \langle X, A\odot(ww^\top)\odot(ss^\top)\rangle \geqslant (2+\Delta(\sigma))(1-\mu)^2n^2 - O((\sigma+\sigma^{-1})\mu n^2)$$

Plug the two parts into eq. (G.4), we get:

$$\begin{aligned}\mathcal{A}\left|\frac{X,w}{4}\right. \langle X, X^*\odot(ww^\top)\odot(ss^\top)\rangle &\geqslant \frac{1}{\sigma}\left((2+\Delta(\sigma))(1-\mu)^2n^2 - O((\sigma+\sigma^{-1})\mu n^2) - (2+\xi)n^2\right)\\ &= \theta(\sigma)n^2\end{aligned} \tag{G.5}$$

for some $\theta(\sigma)$ that only depends on σ.

Now, plug eq. (G.3) and eq. (G.5) into eq. (G.2), we get

$$\begin{aligned}\mathcal{A}\left|\frac{X,w}{4}\right. \langle X, X^*\rangle &= \langle X, X^*\odot(ww^\top)\odot(ss^\top)\rangle + \langle X, X^*\odot(J-(ww^\top)\odot(ss^\top))\rangle\\ &\geqslant \theta(\sigma)n^2 - 4\mu n^2\end{aligned}$$

When $\mu \leqslant \mu^*(\sigma)$ for some value $\mu^*(\sigma)$ that only depends on σ, we have

$$\mathcal{A}\left|\frac{X,w}{4}\right. \langle X, X^*\rangle \geqslant \theta'(\sigma)n^2$$

where $\theta'(\sigma)$ is a value that only depends on σ. Thus, when $\sigma > 1$, we have, with probability $1-o(1)$,

$$\mathcal{A}\left|\frac{X,w}{4}\right. \langle X, X^*\rangle \geqslant \Omega(n^2)$$

■

Now, we are ready to prove theorem 36. In the proof, we first prove feasibility of the contraint set in eq. (G.1), then use lemma 37 to complete the proof.

Proof [Proof of theorem 36] The feasibility analysis is similar to the feasibility analysis in lemma 22. From theorem 35 and union bound, we get that the inequality $\langle A\odot(ww^\top), X\rangle \geqslant (2+\Delta(\sigma))(1-\mu)^2n^2$ is feasible with probability $1-o(1)$. The inequality $\left\|A\odot(ww^\top)\right\|_{\text{op}} \leqslant (\sigma+\sigma^{-1})n$ is feasible with probability $1-o(1)$ due to the famous BBP phase transition and monotonicity of spectral norm. Take union bound over failure probabilities of the two inequalities, we can conclude the feasibility analysis.

From lemma 37, we get that, with probability $1-o(1)$, we have $\mathcal{A}\left|\frac{X,w}{4}\right. \langle X, X^*\rangle \geqslant \Omega(n^2)$. Therefore, we can take union bound and conclude that, with probability $1-o(1)$, the program finds an X such that

$$\mathcal{A}\left|\frac{X,w}{4}\right. \langle X, X^*\rangle \geqslant \Omega(n^2)$$

■

Now we finish the proof of theorem 6.

Proof [Proof of theorem 6] By combining theorem 36 and theorem 18, we can compute the pseudo-expectation $\tilde{\mathbb{E}}$ for the SOS relaxtion of eq. (G.1) in polynomial time. Let $\hat{X} := \tilde{\mathbb{E}}[X]$ in eq. (5.1). By linearity of pseudo-expectation, we have $\hat{X} \succeq 0$, $\hat{X}_{ii} = 1$. Furthermore, we have $\langle \hat{X}, X^* \rangle \geqslant \Omega(n^2)$ with probability $1 - o(1)$. Now, applying rounding procedure in lemma 25, we can then obtain $\hat{x} \in \{\pm 1\}^n$ such that $\mathbb{E}\langle \hat{x}, x^* \rangle^2 \geqslant \Omega(n^2)$. ■

Appendix H. Deferred proofs

H.1. Proof of claim 11

Claim 38 (Restatement of claim 11) *Given matrix M, we have* $\mathrm{SDP}(M) \leqslant \|M\|_{Gr}$.

Proof If we look at the second definition of the basic SDP in eq. (4.2) and the second definition of Grothendieck norm in eq. (4.4), it is easy to check that the optimizer of eq. (4.2) is a solution to eq. (4.4) if we take $\delta_i = \sigma_i$. Hence, we have

$$\mathrm{SDP}(M) \leqslant \|M\|_{Gr}$$

■

H.2. Proof of claim 12

Claim 39 (Restatement of claim 12)

Let M be an $n \times n$ matrix whose diagonal entries are 0 and $S \subseteq [n]$ be a subset of indices, we have

$$\mathrm{SDP}(M_S) \leqslant \mathrm{SDP}(M)$$

Proof Let X be the optimizer of $\mathrm{SDP}(M_S)$ and $Z = X_S + \mathrm{Id}_{[n]\setminus s}$. We have

$$\begin{aligned}\mathrm{SDP}(M_S) =& \langle X, M_S \rangle \\ =& \langle X_S, M \rangle\end{aligned}$$

Since M has zero on diagonals, we have $\langle X_S, M \rangle = \langle Z, M \rangle$. Notice that $Z \succeq 0$ and $Z_{ii} = 1$ for all $i \in [n]$. Therefore, Z is a solution to the basic SDP, which implies that

$$\langle Z, M \rangle \leqslant \mathrm{SDP}(M)$$

Thus, we have

$$\begin{aligned}\mathrm{SDP}(M_S) =& \langle X_S, M \rangle \\ =& \langle Z, M \rangle \\ \leqslant& \,\mathrm{SDP}(M)\end{aligned}$$

■

H.3. Proof of lemma 8

Lemma 40 (Formal statement of lemma 8) *Let $\tilde{A} \in \mathbb{R}^{n\times n}$ and $S \subset [n]$ be a set of size $(1-\mu)n$. Suppose $\left\|\tilde{A}_S\right\|_{\text{op}} \leqslant C_s\sqrt{d}$ for some constant C_s, then for all $S' \subseteq S$ with size at least $(1-2\mu)n$, there is a deg-4 SOS proof that*

$$\text{SDP}(\tilde{A}_S - \tilde{A}_{S'}) \leqslant O(\mu n\sqrt{d})\,.$$

Proof Consider an arbitrary matrix $X \in \mathbb{R}^{n\times n}$ such that $X_{ii} = 1$ for $i \in [n]$ and $X \succeq 0$. It follows that

$$\begin{aligned}
\langle \tilde{A}_S - \tilde{A}_{S'}, X\rangle &= \langle \tilde{A}_S - \tilde{A}_{S'}, X_S - X_{S'}\rangle \\
&\leqslant \|\tilde{A}_S - \tilde{A}_{S'}\|_{\text{op}} \operatorname{Tr}(X_S - X_{S'}) \\
&\leqslant (\|\tilde{A}_S\|_{\text{op}} + \|\tilde{A}_{S'}\|_{\text{op}}) \cdot \mu n \\
&\leqslant 2\|\tilde{A}_S\|_{\text{op}} \cdot \mu n \\
&\leqslant 2C_s\sqrt{d} \cdot \mu n \\
&= O(\mu n\sqrt{d})
\end{aligned}$$

Notice that, every step of the proof can be made to be deg-4 SOS. Hence, the proof is deg-4 SOS. ■

H.4. Proof of lemma 22

Lemma 41 (Restatement of lemma 22) *The SOS program in eq. (5.1) is feasible with probability $1-o(1)$.*

Proof From corollary 34, we know that, with probability $1-o(1)$, there exists a submatrix $\tilde{A}_T$ of size $(1-\beta)n$ whose spectral norm is bounded by $C_s\sqrt{d}$. By monotonicity of spectral norm, the spectral norm of all submatrices of size $(1-\mu-\beta)n$ of $\tilde{A}_T$ are bounded by $C_s\sqrt{d}$. Therefore, if we consider the set $S = T \cap S^*$, it satisfies the spectral constraint with probability $1-o(1)$.

Now, we need to show that, with probability $1-o(1)$, the matrix $\tilde{A}_S$ with $S = T\cap S^*$ has large enough basic SDP value. Apply theorem 32, we get that with probability at least $1 - Ce^{-(1-\mu-\beta)n/C}$, a stochastic block model of size $(1-\mu-\beta)n$ has basic SDP value larger than or equal to $(2+\Delta)(1-\mu-\beta)n\sqrt{d}$. Consider all submatrices of size $(1-\mu-\beta)n$ of $\tilde{A}$ and take union bound, the failure probability is

$$\begin{aligned}
\binom{n}{(1-\mu-\beta)n} Ce^{-(1-\mu-\beta)n/C} \leqslant& C\left(\frac{en}{(\mu+\beta)n}\right)^{(\mu+\beta)n} e^{-(1-\mu-\beta)n/C} \\
=& Ce^{(\mu+\beta)n(\log(1/(\mu+\beta))+1)-(1-\mu-\beta)n/C}
\end{aligned}$$

When $\mu \leqslant \mu_\delta$ for some value μ_δ that only depends on δ and $\beta \ll 1/C$, the failure probability is $o(1)$. Therefore, with probability $1-o(1)$, for the uncorrputed stochastic block model, the basic SDP value of all its submatrices of size $(1-\mu-\beta)n$ is larger than or equal to $(2+\Delta)(1-\mu-\beta)n\sqrt{d}$, which include the submatrix defined by the set $S = T \cap S^*$.

Hence, with probability $1 - o(1)$, there exists a subset $S = T \cap S^*$ of size $(1 - \mu - \beta)n$ such that $\tilde{A}_S$ has basic SDP value larger than or equal to $(2 + \Delta)(1 - \mu - \beta)n\sqrt{d}$ and has spectral norm less than or equal to $C_s\sqrt{d}$.

For the value of X, we can simply take the optimizer of the basic SDP for $\tilde{A}_S$. This concludes the feasibility analysis of the program. ■

H.5. Proof of lemma 24

Lemma 42 (Restatement of lemma 24) *For X and w that satisfy the SOS program in eq. (5.1), we have*

$$\mathcal{A}\Big|\frac{X,w}{4}\ \langle X, X^*\rangle \geqslant \frac{\Delta'(1-\beta)n^2}{\varepsilon\sqrt{d}} - O(\frac{\mu n^2}{\varepsilon\sqrt{d}}) - 2\beta n^2$$

where β is the small constant fraction of high degree nodes we need to prune to get bounded spectral norm according to corollary 34 and $\Delta' = \Delta'(\delta)$ for some value $\Delta'(\delta)$ that only depends on δ.

Proof We decompose $\langle X, X^*\rangle$ into $\langle X, X^*\rangle = \langle X_{S'}, X^*_{S'}\rangle + \langle X - X_{S'}, X^*\rangle$. For $\langle X_{S'}, X^*_{S'}\rangle$, we can apply lemma 23 and get

$$\mathcal{A}\Big|\frac{X,w}{4}\ \langle X_{S'}, X^*_{S'}\rangle \geqslant \frac{\Delta'(1-\beta)n^2}{\varepsilon\sqrt{d}} - O(\frac{\mu n^2}{\varepsilon\sqrt{d}})$$

Now, we consider $\langle X - X_{S'}, X^*\rangle$. Notice that, since X is positive semidefinite whose diagonals are 1's, all its entries are within $[-1, 1]$. This is because all principle submatrices of a positive semidefinite matrix are positive semidefinite. If we consider the principle submatrix formed by X_{ii}, X_{ij}, X_{ji} and X_{jj}, its determinant is non-negative. Hence, $X_{ij}^2 \leqslant X_{ii}X_{jj} = 1$. Since there can be at most $(2\mu + \beta)n$ vertices that are not in S', $\langle X - X_{S'}, X^*\rangle$ is a summation of at most $2(2\mu + \beta)n^2$ entries whose absolute values are less than or equal to 1. Therefore, we have

$$|\langle X - X_{S'}, X^*\rangle| \leqslant 2(2\mu + \beta)n^2$$

Combine the bounds on $\langle X_{S'}, X^*_{S'}\rangle$ and $\langle X - X_{S'}, X^*\rangle$, we have

$$\begin{aligned}\mathcal{A}\Big|\frac{X,w}{4}\ \langle X, X^*\rangle =& \langle X_{S'}, X^*_{S'}\rangle + \langle X - X_{S'}, X^*\rangle \\ \geqslant& \frac{\Delta'(1-\beta)n^2}{\varepsilon\sqrt{d}} - O(\frac{\mu n^2}{\varepsilon\sqrt{d}}) - 2(2\mu + \beta)n^2 \\ \geqslant& \frac{\Delta'(1-\beta)n^2}{\varepsilon\sqrt{d}} - O(\frac{\mu n^2}{\varepsilon\sqrt{d}}) - 2\beta n^2\end{aligned}$$

which finishes the proof. ■