# Exact Community Recovery in Correlated Stochastic Block Models

**Julia Gaudio**                                          JULIA.GAUDIO@NORTHWESTERN.EDU
*Department of Industrial Engineering and Management Sciences, Northwestern University*

**Miklós Z. Rácz**                                                 MRACZ@PRINCETON.EDU
*Department of Operations Research and Financial Engineering, Princeton University*

**Anirudh Sridhar**                                             ANIRUDHS@PRINCETON.EDU
*Department of Electrical and Computer Engineering, Princeton University*

**Editors:** Po-Ling Loh and Maxim Raginsky

## Abstract

We consider the problem of learning latent community structure from multiple correlated networks. We study edge-correlated stochastic block models with two balanced communities, focusing on the regime where the average degree is logarithmic in the number of vertices. Our main result derives the precise information-theoretic threshold for exact community recovery using multiple correlated graphs. This threshold captures the interplay between the community recovery and graph matching tasks. In particular, we uncover and characterize a region of the parameter space where exact community recovery is possible using multiple correlated graphs, even though (1) this is information-theoretically impossible using a single graph and (2) exact graph matching is also information-theoretically impossible. In this regime, we develop a novel algorithm that carefully synthesizes algorithms from the community recovery and graph matching literatures.

**Keywords:** Stochastic block model, community recovery, graph matching, correlated random graphs, information-theoretic limits

## 1. Introduction

Recovering communities in networks is a fundamental learning task that has myriad applications in sociology, biology, and beyond. Increasingly, network data is supplemented with further data that is correlated with the underlying communities, such as latent feature vectors (e.g., the interests of individuals in a social network) or further correlated networks (e.g., personal and professional social networks overlap, yet contain complementary information). Synthesizing information from these different data sources presents an opportunity to obtain improved community recovery algorithms and guarantees, yet this comes with algorithmic and statistical challenges. In particular, integrating information from correlated networks is often hindered because the graphs are not aligned, due to node labels that are missing, erroneous, anonymized, or otherwise unknown. This highlights the importance of graph matching, which is an important learning task in its own right.

Recently, Rácz and Sridhar (2021) determined the information-theoretic limits for exact graph matching in edge-correlated stochastic block models, and as an application they showed how to exactly recover communities from two correlated graphs in a regime where it is impossible to do so using just a single graph. The main contribution of our work is to go *beyond* exact graph matching, and we determine the precise information-theoretic threshold for exact community recovery from two correlated block models. In particular, we uncover and characterize a region of the parameter space where exact community recovery is possible despite exact graph matching being impossible (and exact community recovery from a single graph also being impossible), positively resolving a

conjecture of Rácz and Sridhar (2021). To do so, we develop a novel algorithm that carefully synthesizes community recovery and graph matching algorithms. Overall, our work highlights the subtle interplay between community recovery and graph matching, two canonical and widely-studied learning problems.

## 2. Community recovery in correlated stochastic block models

**The stochastic block model (SBM).** The SBM is the canonical probabilistic generative model for networks with community structure. Introduced by Holland et al. (1983), it has received enormous attention over the past decades; in particular, it serves as a natural theoretical testbed for evaluating and comparing clustering algorithms on average-case networks (see, e.g., Dyer and Frieze (1989); Bui et al. (1984); Boppana (1987)). The SBM allows a precise understanding of when community information can be extracted from network data, due to the fact that it exhibits sharp information-theoretic phase transitions for various inference tasks. Such phase transitions were first conjectured by Decelle et al. (2011) and were subsequently proven rigorously by several authors (Mossel et al., 2014; Massoulié, 2014; Mossel et al., 2018; Abbe et al., 2016; Mossel et al., 2016; Abbe and Sandon, 2015; Bordenave et al., 2015; Abbe, 2017).

Here we focus on the SBM with two symmetric communities, arguably the simplest setting. For a positive integer $n$ and $p, q \in [0, 1]$, we construct the graph $G \sim \mathrm{SBM}(n, p, q)$ as follows. The graph $G$ has $n$ vertices, labeled by the elements of $[n] := \{1, \ldots, n\}$. Each vertex $i \in [n]$ has a community label $\sigma_*(i) \in \{+1, -1\}$; these are drawn i.i.d. uniformly at random across all $i \in [n]$. The vector of community labels is denoted by $\boldsymbol{\sigma_*} := \{\sigma_*(i)\}_{i=1}^n$, with the two communities given by the sets $V^+ := \{i \in [n] : \sigma_*(i) = +1\}$ and $V^- := \{i \in [n] : \sigma_*(i) = -1\}$. Given the community labels $\boldsymbol{\sigma_*}$, the edges of $G$ are drawn independently across vertex pairs as follows. For distinct $i, j \in [n]$, if $\sigma_*(i)\sigma_*(j) = 1$ (i.e., $i$ and $j$ are in the same community), then the edge $(i, j)$ is in $G$ with probability $p$; else, $(i, j)$ is in $G$ with probability $q$.

**Community recovery.** In this setting, a community recovery algorithm takes as input the graph $G$, without knowledge of the community labels $\boldsymbol{\sigma_*}$, and outputs a community labeling $\widehat{\boldsymbol{\sigma}}$. The success of an algorithm is measured by the *overlap* between the estimated labeling and the ground truth, defined as

$$\mathrm{ov}(\widehat{\boldsymbol{\sigma}}, \boldsymbol{\sigma_*}) := \frac{1}{n} \left| \sum_{i=1}^n \widehat{\sigma}(i)\sigma_*(i) \right|.$$

We take an absolute value in this formula since the labelings $\boldsymbol{\sigma_*}$ and $-\boldsymbol{\sigma_*}$ specify the same partition of communities, and it is only possible to recover $\boldsymbol{\sigma_*}$ up to its sign. Observe that $\mathrm{ov}(\widehat{\boldsymbol{\sigma}}, \boldsymbol{\sigma_*}) \in [0, 1]$, with a larger value corresponding to a better estimate. In particular, the algorithm succeeds in exactly recovering the communities (i.e., $\widehat{\boldsymbol{\sigma}} = \boldsymbol{\sigma_*}$ or $\widehat{\boldsymbol{\sigma}} = -\boldsymbol{\sigma_*}$) if and only if $\mathrm{ov}(\widehat{\boldsymbol{\sigma}}, \boldsymbol{\sigma_*}) = 1$.

In the logarithmic degree regime—that is, when $p = \alpha \log(n)/n$ and $q = \beta \log(n)/n$ for some fixed constants $\alpha, \beta \geq 0$—it is well-known that there is a sharp information-theoretic threshold for exactly recovering communities in the SBM (Abbe et al., 2016; Mossel et al., 2016; Abbe and Sandon, 2015; Abbe, 2017). This is governed by the quantity

$$\mathrm{D}_+(\alpha, \beta) := \left( \frac{\sqrt{\alpha} - \sqrt{\beta}}{\sqrt{2}} \right)^2 = \frac{\alpha + \beta}{2} - \sqrt{\alpha\beta}. \tag{1}$$

In the general setting, this quantity is known as the *Chernoff-Hellinger divergence* (Abbe and Sandon, 2015; Abbe, 2017); in the specific setting of two balanced communities, it simplifies to
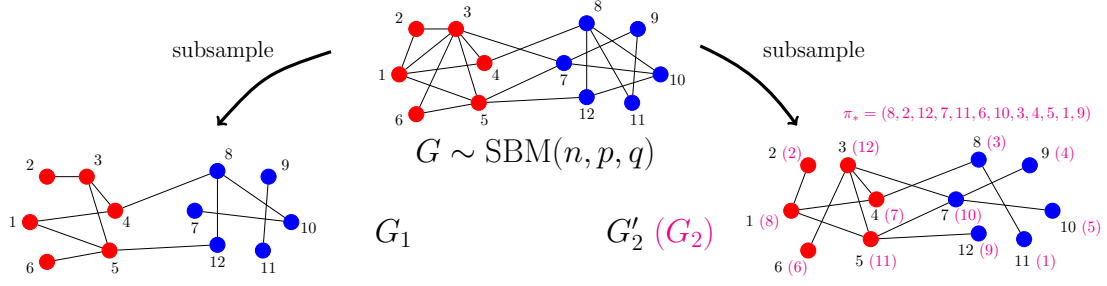
Figure 1: Schematic showing the construction of correlated SBMs (see text for details).

the Hellinger divergence of the vectors $(\alpha/2, \beta/2)$ and $(\beta/2, \alpha/2)$, giving (1). The information-theoretic threshold for exact community recovery is then given by

$$D_+(\alpha, \beta) = 1. \tag{2}$$

If $D_+(\alpha, \beta) > 1$, then exact community recovery is possible: there is a polynomial-time algorithm which outputs an estimator $\widehat{\sigma}$ satisfying $\lim_{n \to \infty} \mathbb{P}(\mathsf{ov}(\widehat{\sigma}, \sigma_*) = 1) = 1$. Moreover, if $D_+(\alpha, \beta) < 1$, then this is impossible: for *any* estimator $\widetilde{\sigma}$, we have that $\lim_{n \to \infty} \mathbb{P}(\mathsf{ov}(\widetilde{\sigma}, \sigma_*) = 1) = 0$.

**Correlated SBMs.** The goal of our work is to understand how the exact community recovery threshold given by (2) *changes* when the input data consists of multiple *correlated* SBMs. To this end, we study a natural model of correlated SBMs, which we describe next.

We construct $(G_1, G_2) \sim \mathrm{CSBM}(n, p, q, s)$ as follows, where the additional parameter $s \in [0, 1]$ controls the level of correlation between the two graphs. First, generate a parent graph $G \sim \mathrm{SBM}(n, p, q)$, and let $\sigma_*$ denote the community labels. Next, given $G$, we construct $G_1$ by independent subsampling: each edge of $G$ is included in $G_1$ with probability $s$, independently of everything else, and non-edges of $G$ remain non-edges in $G_1$. We obtain a second graph, $G_2'$, independently in the same way. The graphs $G_1$ and $G_2'$ inherit both the vertex labels and the community labels $\sigma_*$ from the parent graph $G$. Finally, we let $\pi_*$ be a uniformly random permutation of $[n]$, independently of everything else, and generate $G_2$ by relabeling the vertices of $G_2'$ according to $\pi_*$ (e.g., vertex $i$ in $G_2'$ is relabeled to $\pi_*(i)$ in $G_2$). This last step in the construction of $G_2$ reflects the fact that in applications, node labels are often obscured. To emphasize the effect of the vertex relabeling on the community labels, we define $\sigma_*^{\mathbf{1}} := \sigma_*$ and $\sigma_*^{\mathbf{2}} := \sigma_* \circ \pi_*^{-1}$, which are the community labels in $G_1$ and $G_2$, respectively. This construction is visualized in Figure 1.

First studied by Onaran et al. (2016), this model of correlated SBMs is the natural generalization of correlated Erdős-Rényi random graphs, which were introduced by Pedarsani and Grossglauser (2011) (see Section 5 for discussion of further related work). In particular, marginally $G_1$ and $G_2$ are both SBMs. Specifically, since the subsampling probability is $s$, we have that $G_1 \sim \mathrm{SBM}(n, ps, qs)$. Therefore, from (2) it follows that, in the logarithmic degree regime where $p = \alpha \log(n)/n$ and $q = \beta \log(n)/n$, the communities can be exactly recovered from $G_1$ *alone* if $D_+(\alpha s, \beta s) > 1$. Since $D_+(\alpha s, \beta s) = s D_+(\alpha, \beta)$, this condition is equivalent to $s D_+(\alpha, \beta) > 1$, which we can also write as $D_+(\alpha, \beta) > 1/s$.

The central question of our work is how to go *beyond* this single-graph threshold by incorporating the information in $G_2$. This question was initiated in recent work of Rácz and Sridhar (2021), whose starting observation was the following. If $\pi_*$ were *known*, then we can reconstruct $G_2'$ from $G_2$, and then "overlay" $G_1$ and $G_2'$ to obtain a new graph $H_*$ that combines the information in the

3

two graphs. In particular, $(i, j)$ is an edge in $H_*$ if and only if $(i, j)$ is an edge in the parent graph $G$ and it is included in either $G_1$ or $G_2'$ in the subsampling process. It thus follows that $H_*$ is also an SBM, specifically, $H_* \sim \mathrm{SBM}\left(n, \alpha(1 - (1 - s)^2) \log(n)/n, \beta(1 - (1 - s)^2) \log(n)/n\right)$. In particular, this argument implies that if $\pi_*$ were *known* and

$$\frac{1}{1 - (1 - s)^2} < \mathrm{D}_+(\alpha, \beta) < \frac{1}{s},$$

then it is information-theoretically impossible to exactly recover $\boldsymbol{\sigma}_*$ from $G_1$ alone, but one can recover $\boldsymbol{\sigma}_*$ exactly by combining information from $G_1$ *and* $G_2$.

**Graph matching.** Since $\pi_*$ is not known, the argument above raises the question of when $\pi_*$ can be exactly recovered from $(G_1, G_2)$, a task known as *graph matching*. The main result of Rácz and Sridhar (2021) answers this question (see also Section 5 for discussion of related work). Specifically, they show that the information-theoretic threshold for exactly recovering $\pi_*$ is given by $s^2 (\alpha + \beta)/2 = 1$. Note that this is precisely the *connectivity threshold* for the *intersection graph* of $G_1$ and $G_2'$ (the edges of this intersection graph are the edges present in the parent graph $G$ that survived both subsampling processes). Letting

$$\mathrm{T}_{\mathrm{c}}(\alpha, \beta) := \frac{\alpha + \beta}{2} \tag{3}$$

denote the connectivity threshold in $\mathrm{SBM}(n, \alpha \log(n)/n, \beta \log(n)/n)$, we can write the threshold for exactly recovering $\pi_*$ as $\mathrm{T}_{\mathrm{c}}(\alpha s^2, \beta s^2) = 1$, or equivalently, $s^2 \mathrm{T}_{\mathrm{c}}(\alpha, \beta) = 1$. Thus, if $s^2 \mathrm{T}_{\mathrm{c}}(\alpha, \beta) > 1$, then $\pi_*$ can be exactly recovered from $(G_1, G_2)$, while if $s^2 \mathrm{T}_{\mathrm{c}}(\alpha, \beta) < 1$, then this is impossible.

To summarize the two previous paragraphs, Rácz and Sridhar (2021) showed that if

$$s^2 \mathrm{T}_{\mathrm{c}}(\alpha, \beta) > 1 \qquad \text{and} \qquad \left(1 - (1 - s)^2\right) \mathrm{D}_+(\alpha, \beta) > 1, \tag{4}$$

then exact community recovery is possible, that is, it is possible to exactly recover $\boldsymbol{\sigma}_*$ using $(G_1, G_2)$.

**The interplay between community recovery and graph matching.** The work of Rácz and Sridhar (2021) leaves open the question of what happens when exact graph matching is impossible. In particular, is there a parameter regime where exact community recovery is possible from $(G_1, G_2)$, even though (1) this is information-theoretically impossible using a single graph and (2) exact graph matching is also impossible?

We answer this question affirmatively, developing an algorithm that carefully combines community recovery and graph matching steps. Moreover, we determine the precise information-theoretic threshold for when exact community recovery is possible. If $\left(1 - (1 - s)^2\right) \mathrm{D}_+(\alpha, \beta) > 1$, then this threshold is given by

$$\mathrm{T}_{\mathrm{c}}\left(\alpha s^2, \beta s^2\right) + \mathrm{D}_+\left(\alpha s(1 - s), \beta s(1 - s)\right) = 1. \tag{5}$$

The threshold in (5) cleanly showcases the interplay between community recovery and graph matching: the first term in (5) comes from graph matching, while the second term comes from community recovery. We now turn to describing our results formally.

## 3. Results

We determine the information-theoretic threshold for exact community recovery from two correlated stochastic block models $(G_1, G_2) \sim \mathrm{CSBM}(n, \alpha \log(n)/n, \beta \log(n)/n, s)$. This result has two parts and we start with the positive one.

**Theorem 1** *Fix constants $\alpha, \beta > 0$ and $s \in [0,1]$. Let $(G_1, G_2) \sim \mathrm{CSBM}\left(n, \frac{\alpha \log n}{n}, \frac{\beta \log n}{n}, s\right)$.*
*Suppose that*

$$\left(1 - (1-s)^2\right) \mathrm{D}_+(\alpha, \beta) > 1 \tag{6}$$

*and that*

$$s^2 \mathrm{T}_c(\alpha, \beta) + s(1-s) \mathrm{D}_+(\alpha, \beta) > 1. \tag{7}$$

*Then there is an estimator $\widehat{\boldsymbol{\sigma}} = \widehat{\boldsymbol{\sigma}}(G_1, G_2)$ such that $\lim\limits_{n \to \infty} \mathbb{P}\left(\mathrm{ov}\left(\widehat{\boldsymbol{\sigma}}, \boldsymbol{\sigma}_*\right) = 1\right) = 1$.*

In the prior work of Rácz and Sridhar (2021) it was shown that (6) is necessary for exact community recovery, and that the conditions in (4) suffice. As described in Section 2, Rácz and Sridhar (2021) focused on determining the exact graph matching threshold and then using community recovery algorithms as a black box. The main contribution of Theorem 1 is to go *beyond* exact graph matching, to showcase how exact community recovery is possible from $(G_1, G_2)$ even in regimes where (1) this is impossible from $G_1$ alone and (2) exact graph matching is impossible. This necessitates developing algorithms that combine information from $G_1$ and $G_2$ in more delicate ways, integrating ideas from community recovery and graph matching algorithms. Indeed, at a high level, the algorithm we develop to prove Theorem 1 has four main steps:

(1) Obtain a partial almost exact graph matching $\widehat{\mu}$ between $G_1$ and $G_2$;

(2) Obtain an almost exact community labeling of vertices in $G_1$;

(3) For vertices in $G_1$ that are part of the matching $\widehat{\mu}$: refine the almost exact labeling obtained in Step (2) via a majority vote in the (denser) graph consisting of edges that are either in $G_1$ or in $G_2$ (determined using $\widehat{\mu}$).

(4) For vertices in $G_1$ that are not part of the matching $\widehat{\mu}$: classify them according to a majority vote of the labels of their neighbors, where we use only the edges in $G_1$.

In order to make such an algorithm work, the devil is in the details, with careful choices in each step; we refer to Section 4 for a more detailed overview of the algorithm.

The threshold in (7) highlights the interplay between the community recovery and graph matching tasks. Indeed, the first term in (7) comes from graph matching: $s^2 \mathrm{T}_c(\alpha, \beta) = 1$ is the threshold for exact graph matching; moreover, when $s^2 \mathrm{T}_c(\alpha, \beta) < 1$, the best possible almost exact graph matching makes $n^{1 - s^2 \mathrm{T}_c(\alpha, \beta) + o(1)}$ errors, which is relevant for Step (1) of the algorithm. On the other hand, the second term in (7) comes from community recovery; in particular, this term arises from the majority vote in Step (4). Note that while we use all edges in $G_1$ for this step, the unmatched nodes are isolated in the intersection graph, and hence the relevant edges are not present in $G_2$, leading to the "effective" factor of $s(1-s)$. Since exact community recovery in $\mathrm{SBM}(n, \alpha s(1-s) \log(n)/n, \beta s(1-s) \log(n)/n)$ is governed by the quantity $\mathrm{D}_+(\alpha s(1-s), \beta s(1-s)) = s(1-s) \mathrm{D}_+(\alpha, \beta)$, this leads to the second term in (7).

As the following impossibility result shows, Theorem 1 is tight.

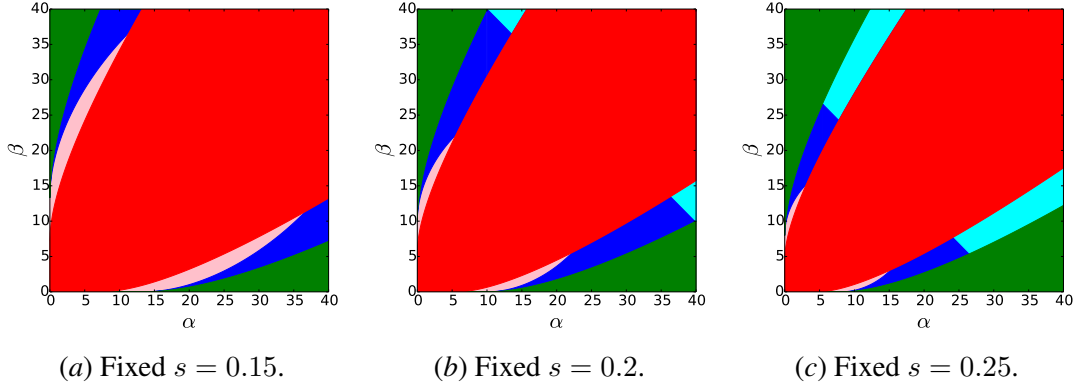(a) Fixed $s = 0.15$.  (b) Fixed $s = 0.2$.  (c) Fixed $s = 0.25$.

Figure 2: Phase diagram for exact community recovery for fixed $s$, with $\alpha \in [0, 40]$ and $\beta \in [0, 40]$ on the axes. *Green region:* exact community recovery is possible from $G_1$ alone; *Cyan region:* exact community recovery is impossible from $G_1$ alone, but exact graph matching is possible, and subsequently exact community recovery is possible from $(G_1, G_2)$; *Dark Blue region:* exact community recovery is impossible from $G_1$ alone, exact graph matching is also impossible, yet exact community recovery is nonetheless possible from $(G_1, G_2)$; *Pink region:* exact community recovery is impossible from $(G_1, G_2)$ (even though it would be possible if $\pi_*$ were known). *Red region:* exact community recovery is impossible from $(G_1, G_2)$ (even if $\pi_*$ is known). Characterizing the Dark Blue and Pink regions is the main result of this paper.

**Theorem 2** *Fix constants $\alpha, \beta > 0$ and $s \in [0, 1]$. Let $(G_1, G_2) \sim \mathrm{CSBM}\left(n, \frac{\alpha \log n}{n}, \frac{\beta \log n}{n}, s\right)$. Suppose that*

$$\left(1 - (1 - s)^2\right) \mathrm{D}_+(\alpha, \beta) < 1 \tag{8}$$

*or that*

$$s^2 \mathrm{T_c}(\alpha, \beta) + s(1 - s)\mathrm{D}_+(\alpha, \beta) < 1. \tag{9}$$

*Then for any estimator $\widetilde{\sigma} = \widetilde{\sigma}(G_1, G_2)$, we have that $\lim_{n \to \infty} \mathbb{P}(\mathrm{ov}(\widetilde{\sigma}, \sigma_*) = 1) = 0$.*

Impossibility of exact community recovery under the condition (8) was shown by Rácz and Sridhar (2021), so the contribution of Theorem 2 is to show impossibility under the condition (9).

As discussed above, the condition (9) highlights the interplay between the community recovery and graph matching tasks. In particular, Theorem 2 uncovers and characterizes a region of the parameter space where exact community recovery from $(G_1, G_2)$ is impossible, despite the fact that if $\pi_*$ were known, then exact community recovery would be possible from the correctly matched union graph $G_1 \vee_{\pi_*} G_2$.

Putting together Theorems 1 and 2, we obtain the information-theoretic threshold for exact community recovery in correlated SBMs, see (5). These results are illustrated in the phase diagrams of Figures 2 and 3.
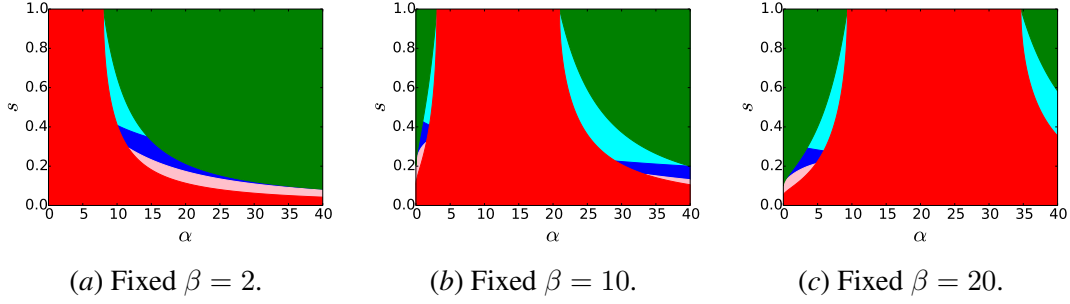
(a) Fixed $\beta = 2$.      (b) Fixed $\beta = 10$.      (c) Fixed $\beta = 20$.

Figure 3: Phase diagrams for exact community recovery for fixed $\beta$, with $\alpha \in [0, 40]$ and $s \in [0, 1]$ on the axes. (Colors as in Fig. 2.)

## 4. Overview of algorithms and proofs

We next expand upon the very high-level steps of the recovery algorithm presented in Section 3, detailing choices made in each step, and highlighting technical challenges that arise in the analysis. We also give an overview of the impossibility proof.

**Almost exact graph matching via the $k$-core estimator.** For a permutation $\pi : [n] \to [n]$, let $G_1 \wedge_\pi G_2$ be the corresponding intersection graph of $G_1$ and $G_2$, where $(i, j)$ is an edge in $G_1 \wedge_\pi G_2$ if and only if $(i, j)$ is an edge in $G_1$ and $(\pi(i), \pi(j))$ is an edge in $G_2$. The graph matching algorithm we study—called the $k$-core estimator—iterates over all permutations of $[n]$ and finds a permutation $\widehat{\pi}$ that induces the largest $k$-core[1] in the corresponding intersection graph. The output of the algorithm is a potentially incomplete vertex correspondence $\widehat{\mu}$, which is the restriction of $\widehat{\pi}$ to the vertex set of the $k$-core in $G_1 \wedge_{\widehat{\pi}} G_2$. Figure 4 provides an illustration.

While it is (information-theoretically) impossible that $\widehat{\mu} = \pi_*$ with probability larger than $o(1)$ in the parameter regime that we consider, $\widehat{\mu}$ has two important properties that we highlight. First, $\widehat{\mu}$ matches almost all vertices; that is, if we denote by $F$ the set of unmatched vertices, then $|F| = o(n)$. Second, the vertices matched by $\widehat{\mu}$ are all correct with probability $1 - o(1)$. This property is extremely important for downstream community recovery tasks, as it allows us to clearly leverage information from $G_1$ and $G_2$ in classifying a given vertex in the correspondence $\widehat{\mu}$.

Intuitively, graph matching algorithms tend to fail in aligning vertices that are not well-connected in $G_1 \wedge_{\pi_*} G_2$ (in particular, singletons), since such vertices have little common information across $G_1$ and $G_2$. On the other hand, the $k$-core is, by definition, a well-connected subgraph, explaining why the $k$-core estimator succeeds for an appropriate choice of $k$ (we choose $k = 13$ in this paper). Previously, Cullina et al. (2020) proved the almost exact correctness of the $k$-core estimator for sparse correlated Erdős-Rényi graphs. Here, we extend their results to correlated SBMs, which requires a novel analysis, since the prior work of Cullina et al. (2020) utilized certain probability generating functions that can only be tractably computed for Erdős-Rényi graphs. Our approach circumvents this issue while also providing a tight analysis for the logarithmic degree regime.

An important and novel consequence of our analysis is that for correlated SBMs in the logarithmic degree regime, the $k$-core estimator achieves *optimal performance*. Specifically, we show that the $k$-core estimator fails to match at most $|F| \leq n^{1 - s^2 \mathrm{T_c}(\alpha, \beta) + o(1)}$ vertex pairs, which is orderwise

---

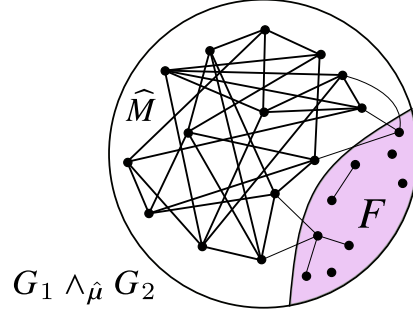1. The $k$-core of a graph is the largest induced subgraph for which all vertices have degree at least $k$.

Figure 4: Illustration of the $k$-core matching (here, $k = 4$). Pictured here is the corresponding $k$-core of the intersection graph $G_1 \wedge_{\widehat{\mu}} G_2$.

equal to the number of singletons of $G_1 \wedge_{\pi_*} G_2$, for which it is known that *any* graph matching algorithm will fail (Rácz and Sridhar, 2021; Cullina et al., 2016). This optimality of the $k$-core estimator is a fundamental reason why we utilize the $k$-core estimator to prove that the information-theoretic threshold can be achieved.

**Exact community recovery in the correctly matched region.** Since exact community recovery is impossible in $G_1$ or in $G_2$ alone, we will utilize the almost-exact vertex alignment $\widehat{\mu}$ to combine information from both graphs to recover communities. The algorithm we design does so by recovering communities through multiple subroutines, each of which is executed carefully in order to de-couple the complex dependencies between $G_1$, $G_2$, and $\widehat{\mu}$. Initially, we focus on recovering the community labels of vertices that are part of the (partial) matching $\widehat{\mu}$.

First, we run a community recovery algorithm on $G_1$ to generate an almost-exact community labeling (i.e., with $o(n)$ errors). Using $\widehat{\mu}$, we then identify the graph $G_1 \vee_{\widehat{\mu}} G_2$, which consists of edges $(i, j)$ such that $(i, j)$ is an edge in $G_1$ or $(\widehat{\mu}(i), \widehat{\mu}(j))$ is an edge in $G_2$. Using $G_1 \vee_{\widehat{\mu}} G_2$, we then refine the almost-exact community labeling by re-classifying vertices in $\widehat{\mu}$ according to a majority vote among the labels of neighbors in $G_1 \vee_{\widehat{\mu}} G_2$. See Figure 5 for an illustration.

To make this algorithm work, there are several technical roadblocks which require novel ideas to overcome. First, since the almost-exact community labels inferred from $G_1$ are subsequently analyzed in the context of $G_1 \vee_{\widehat{\mu}} G_2$, it is critical that the incorrectly-classified vertices are not well-connected. If, for instance, a vertex had only a small number of correctly-classified neighbors in $G_1 \vee_{\widehat{\mu}} G_2$, the final majority vote step would not be guaranteed to succeed. We therefore employ an algorithm previously developed by Mossel et al. (2016), for which we can control the geometry of the misclassified vertices and show that the incorrectly-classified vertices are indeed only weakly connected. We remark that, as a consequence of our analysis, we show that the algorithm of Mossel et al. (2016) is *optimal*, in the sense that it outputs a labeling which makes the smallest possible number of errors. We repeatedly leverage this property to show that our full algorithm works down to the information-theoretic threshold.

The other key technical hurdle concerns the structure of $G_1 \vee_{\widehat{\mu}} G_2$. Notice that, in the regime $\left(1 - (1 - s)^2\right) D_+(\alpha, \beta) > 1$, with high probability it is guaranteed that all vertices in $G_1 \vee_{\pi_*} G_2$ have the property that the majority of each community among their neighbors is the same as the community label of the vertex itself. However, it is unclear whether this is the case for $G_1 \vee_{\widehat{\mu}} G_2$,
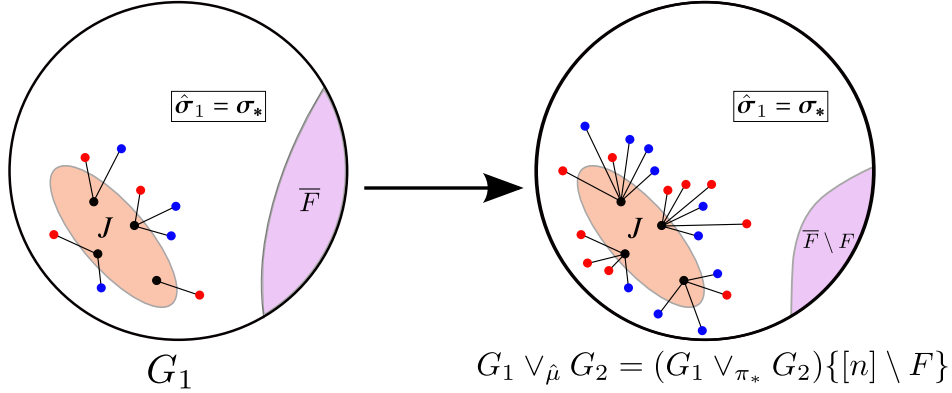
Figure 5: Illustration of how the recovery algorithm classifies vertices in $[n] \setminus \overline{F}$ (see Algorithm 2). In each graph, the orange region denotes the set of incorrectly classified vertices resulting from applying Algorithm 3 (almost exact community recovery) to $G_1$, the purple region denotes $\overline{F}$ or $\overline{F} \setminus F$ (which are classified later), and the white region denotes the set of correctly-classified vertices. When the two graphs are overlaid to form $G_1 \vee_{\widehat{\mu}} G_2$, all vertices in $J$ have a neighborhood majority corresponding to the correct community labeling.

since the removal of the vertices in $F$ that are not part of the (incomplete) vertex correspondence $\widehat{\mu}$ may skew the neighborhood majority of nodes that have many neighbors in $F$. To remedy this issue, we employ a method of Łuczak (1991) to find a set $\overline{F} \supset F$ that is guaranteed to be only *weakly connected* to vertices outside of $\overline{F}$ in $G_1 \vee_{\widehat{\mu}} G_2$ (and not much greater in size than $F$). As a consequence, if $v \in [n] \setminus \overline{F}$ is incorrectly classified in $G_1$, the majority of its correctly labeled neighbors in $G_1 \vee_{\widehat{\mu}} G_2$ will not be significantly affected by $\overline{F}$, resulting in the correct labeling.

**Classifying the vertices outside of the correctly matched region.** It now remains to classify vertices in $\overline{F}$. To classify $\overline{F} \setminus F$, it is useful to consider the graph $G_1 \setminus_{\widehat{\mu}} G_2$, which consists of edges $(i,j)$ that are in $G_1$ such that $(\widehat{\mu}(i), \widehat{\mu}(j))$ is not in $G_2$. In our construction of $\overline{F}$, we are careful to only use the structure of $G_2$, not $G_1$, so that the neighbors of $\overline{F}$ in $G_1 \setminus_{\widehat{\mu}} G_2$ do not depend strongly on the structure of $\overline{F}$. We then classify vertices in $\overline{F} \setminus F$ according to their majority among correctly-classified neighbors in $G_1 \setminus_{\widehat{\mu}} G_2$, which are given by the previous step of the algorithm. Due to the approximate independence between $\overline{F}$ and its neighbors in $G_1 \setminus_{\widehat{\mu}} G_2$, the probability that the majority vote fails can be computed in a straightforward manner and is shown to be $n^{-s(1-s)\mathrm{D}_+(\alpha,\beta)+o(1)}$. The factor of $s(1-s)$ in the exponent reflects the fact that $G_1 \setminus_{\widehat{\mu}} G_2$ is essentially constructed from edges in the parent graph $G$ that are sampled by $G_1$ but not by $G_2$. We show that $|\overline{F}| = n^{1-s^2\mathrm{T}_c(\alpha,\beta)+o(1)}$ with probability $1 - o(1)$, which, along with the probability that majority fails, implies that we can informally bound the probability that the algorithm fails as

$$\mathbb{P}(\text{exists a vertex in } \overline{F} \setminus F \text{ such that the majority vote fails}) \leq |\overline{F} \setminus F| \cdot \mathbb{P}(\text{majority vote fails})$$
$$\leq n^{1-s^2\mathrm{T}_c(\alpha,\beta)+o(1)} \cdot n^{-s(1-s)\mathrm{D}_+(\alpha,\beta)+o(1)} = n^{1-(s^2\mathrm{T}_c(\alpha,\beta)+s(1-s)\mathrm{D}_+(\alpha,\beta))+o(1)}.$$

9

In particular, the final expression is $o(1)$ if (7) holds. While the display above is quite informal, it captures the reason why our algorithm works in the regime (7).

Finally, we classify vertices in $F$, which we recall is the set of vertices outside of the $k$-core of $G_1 \wedge_{\pi_*} G_2$, with high probability. We classify these vertices according to a majority vote with respect to their correctly-labeled neighbors in $G_1$ outside of $F$. By definition, vertices in $F$ can have at most $k-1$ edges outside of $F$ in $G_1 \wedge_{\pi_*} G_2$, implying that the bulk of the neighbors will be in the graph $G_1 \setminus_{\pi_*} G_2$. Since $|F| \le n^{1-s^2 \mathrm{T_c}(\alpha,\beta)+o(1)}$ with high probability, we may repeat similar arguments as for the classification of $\overline{F} \setminus F$ above, to conclude that the majority vote will succeed in classifying $F$ provided that (7) holds.

**Impossibility results.** Since impossibility under the regime in (8) was proved by Rácz and Sridhar (2021), we focus on the regime in (9). To this end, notice that when $s^2 \mathrm{T_c}(\alpha,\beta) < 1$, there are $n^{1-s^2 \mathrm{T_c}(\alpha,\beta)+o(1)}$ singletons in $G_1 \wedge_{\pi_*} G_2$; these can be thought of as the vertices with non-overlapping information across $G_1$ and $G_2$. This property makes such vertices impossible to match correctly. As a result, the maximum a posteriori (MAP) estimator for community labels of the singletons of $G_1 \wedge_{\pi_*} G_2$ almost completely disregards information from $G_2$ and classifies the singletons according to their neighborhood majority in $G_1$ alone. To make this rigorous, we give the MAP estimator additional information in the form of $\boldsymbol{\sigma_*^2}$ (the community labels in $G_2$) and the correct matching $\pi_*(i)$ for all nodes $i$ that are not singletons in $G_1 \wedge_{\pi_*} G_2$, and show that even with this additional information, the MAP estimator fails. Specifically, we use the second moment method to show that under the condition (9), with high probability, at least one of the majority votes will lead to the wrong classification. Since the MAP estimator fails in this regime, so too does any other estimator.

## 5. Related work

Since our work focuses on the interplay between community recovery and graph matching, it naturally connects with and builds upon the extensive literatures on these two topics. We highlight here the most relevant related work.

**Community recovery in SBMs.** There is a vast literature on learning latent community structure in networks, and this question is by now understood well in SBMs (Holland et al., 1983; Dyer and Frieze, 1989; Bui et al., 1984; Boppana, 1987; Decelle et al., 2011; Mossel et al., 2014; Massoulié, 2014; Mossel et al., 2018; Abbe et al., 2016; Mossel et al., 2016; Abbe and Sandon, 2015; Bordenave et al., 2015); we refer the reader to Abbé's survey (Abbe, 2017) for an overview. We highlight in particular the works of Abbe et al. (2016) and Mossel et al. (2016), which characterized the threshold for exact community recovery in the balanced two-community SBM. We build and expand upon their algorithms and analyses, in particular dealing with the uncertainties and dependencies arising from the partial, inexact matching between the correlated graphs.

**Beyond SBMs.** Roughly speaking, there are two main strands of literature that go beyond SBMs, incorporating various types of additional information to aid in recovering communities: contextual SBMs and multi-layer networks. In contextual SBMs, the idea is to leverage *node-level* information (e.g., latent high-dimensional vectors) that is correlated with the community labels (Bothorel et al., 2015; Kanade et al., 2016; Mossel and Xu, 2016; Zhang et al., 2016; Binkiewicz et al., 2017; Deshpande et al., 2018; Abbe et al., 2020a; Lu and Sen, 2020; Saad and Nosratinia, 2020; Yan and Sarkar, 2021; Ma and Nandy, 2021). In particular, the information-theoretic limits have recently been characterized for both community detection and exact community recovery (Deshpande

et al., 2018; Abbe et al., 2020a; Lu and Sen, 2020), and in both cases these limits shift due to the high-dimensional node covariates.

Multi-layer SBMs were introduced by Holland, Laskey, and Leinhardt, in the same work that introduced SBMs (Holland et al., 1983). Here, given the underlying community structure, a collection of SBMs is generated on the same vertex set with the same latent community labels. Several variants have been explored (Han et al., 2015; Arroyo et al., 2021; Paul and Chen, 2020, 2021; Lei et al., 2019; Ali et al., 2019; Bhattacharyya and Chatterjee, 2020; Chen et al., 2020), but typically the layers are *conditionally independent* given the community labels. This is a major difference compared to the setting we consider, where the graphs are correlated through the formation of edges. Moreover, the node labels are assumed to be known in the multi-layer setting, which completely removes the need for graph matching.

The recent works (Mayya and Reeves, 2019; Ma and Nandy, 2021) jointly consider multi-layer networks and node-level information that is correlated with the latent community memberships, thus synthesizing these two strands of literature.

**Graph matching: correlated Erdős-Rényi model.** Arguably the simplest probabilistic generative model of correlated graphs is to consider two correlated Erdős-Rényi random graphs. Consequently, this model, introduced by Pedarsani and Grossglauser (2011), has been the focus of the theoretical literature on graph matching. The information-theoretic limits for recovering the latent vertex correspondence $\pi_*$ have been determined for exact recovery (Cullina and Kiyavash, 2016, 2018; Wu et al., 2021) and almost exact recovery (Cullina et al., 2020), and significant progress has been made for weak recovery as well (Ganassali and Massoulié, 2020; Hall and Massoulié, 2020; Ganassali et al., 2021; Wu et al., 2021).

In particular, we highlight the work of Cullina et al. (2020), which is central to this paper and which we extend to correlated SBMs. They showed that the so-called $k$-*core matching* achieves almost exact recovery when the average degree of the intersection graph diverges, and moreover, with high probability, all nodes in this partial matching are known to be correctly matched. This latter property is very useful, especially for downstream tasks such as combining a partial matching with community recovery steps. This directly motivates our choice of using a $k$-core matching in the algorithm that proves Theorem 1.

The quest for efficient algorithms for graph matching has led to numerous algorithmic advances (Mossel and Xu, 2019; Barak et al., 2019; Ding et al., 2021; Fan et al., 2020; Mao et al., 2021b), culminating in the recent work of Mao et al. (2021a), who demonstrated an efficient algorithm for exact recovery in the constant noise regime.

**Graph matching: beyond Erdős-Rényi.** A growing literature studies graph matching in models going beyond Erdős-Rényi, including correlated SBMs (Onaran et al., 2016; Cullina et al., 2016; Lyzinski, 2018; Rácz and Sridhar, 2021; Shirani et al., 2021) and more (Korula and Lattanzi, 2014; Rácz and Sridhar, 2022; Yu et al., 2021). Closest to our work is that of Rácz and Sridhar (2021), who determined the information-theoretic limits for exact graph matching in correlated SBMs, and subsequently leveraged this for exact community recovery. Our main contribution, discussed in detail in Sections 2 and 3, is to go beyond exact graph matching and to understand when exact community recovery is possible in the regime where exact graph matching is impossible.

## 6. Discussion and future work

Our work leaves open several important avenues for future work, which we now outline.

- **Efficient algorithms.** In the parameter regime where exact community recovery is possible from $(G_1, G_2)$ (see Theorem 1), it is important to understand whether this is possible efficiently (in time polynomial in $n$). The algorithm that we developed to prove Theorem 1 is not efficient; specifically, the $k$-core matching step is inefficient, while the other steps are efficient. Finding efficient algorithms for graph matching has been the motivating force behind several recent works (e.g., (Mossel and Xu, 2019; Barak et al., 2019; Ding et al., 2021; Fan et al., 2020; Mao et al., 2021b)), culminating in the recent breakthrough work of Mao et al. (2021a), who developed an efficient algorithm for graph matching in correlated Erdős-Rényi random graphs with constant noise. This promisingly suggests that efficient algorithms exist in the setting of the current paper as well.

  We note, however, that using a $k$-core matching in this paper was a careful choice motivated by the desirable property that, with high probability, all nodes in this partial matching are known to be correctly matched. This raises the possibility that developing efficient algorithms for the full regime of Theorem 1 may require significant new ideas beyond extending the work of Mao et al. (2021a) to correlated SBMs (which, in itself, is an interesting open problem).

- **Three or more correlated graphs.** What happens in the case of several correlated SBMs? Achieving exact community recovery down to the threshold in Theorem 1 requires carefully passing information between the two correlated graphs. It would be interesting to understand how this generalizes to three or more graphs.

- **Beyond exact community recovery.** While here we focus on exact community recovery, it is of great interest to understand how multiple correlated SBMs can help with recovering communities in other parameter regimes. We conjecture that synthesizing information from a second, correlated graph can help in all settings.

  For instance, when only almost exact community recovery is possible from $(G_1, G_2)$, we conjecture that the optimal error rate is of smaller order than if only $G_1$ were known. Similarly, in the partial recovery regime, we conjecture that a larger fraction of nodes can be recovered when given $(G_1, G_2)$, as compared to when only $G_1$ is given. Finally, we conjecture that the threshold for community detection decreases in the case of multiple correlated SBMs, compared to a single SBM. Understanding all of these regimes quantitatively is an important direction for future work.

- **General correlated stochastic block models.** We focused here on the simplest setting of the SBM with two balanced communities. A natural future direction is to extend our results to more general SBMs with multiple communities, which are understood well in the single graph setting (Abbe, 2017).

## Acknowledgments

## References

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

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 (FOCS)*, pages 670–688, 2015.

Emmanuel Abbe, Afonso S. Bandeira, and Georgina Hall. Exact recovery in the stochastic block model. *IEEE Transactions on Information Theory*, 62(1):471–487, 2016.

Emmanuel Abbe, Jianqing Fan, and Kaizheng Wang. An $\ell_p$ theory of PCA and spectral clustering. Preprint available at https://arxiv.org/abs/2006.14062, 2020a.

Emmanuel Abbe, Jianqing Fan, Kaizheng Wang, and Yiqiao Zhong. Entrywise eigenvector analysis of random matrices with low expected rank. *Annals of Statistics*, 48(3):1452–1474, 2020b.

Hafiz Tiomoko Ali, Sijia Liu, Yasin Yilmaz, Romain Couillet, Indika Rajapakse, and Alfred Hero. Latent heterogeneous multilayer community detection. In *2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 8142–8146, 2019.

Jesús Arroyo, Avanti Athreya, Joshua Cape, Guodong Chen, Carey E. Priebe, and Joshua T. Vogelstein. Inference for multiple heterogeneous networks with a common invariant subspace. *Journal of Machine Learning Research*, 22(142):1–49, 2021.

Boaz Barak, Chi-Ning Chou, Zhixian Lei, Tselil Schramm, and Yueqi Sheng. (Nearly) Efficient Algorithms for the Graph Matching Problem on Correlated Random Graphs. In *Advances in Neural Information Processing Systems (NeurIPS)*, pages 9190–9198, 2019.

Sharmodeep Bhattacharyya and Shirshendu Chatterjee. Consistent Recovery of Communities from Sparse Multi-relational Networks: A Scalable Algorithm with Optimal Recovery Conditions. In *Complex Networks XI*, pages 92–103, 2020.

Norbert Binkiewicz, Joshua T Vogelstein, and Karl Rohe. Covariate-assisted spectral clustering. *Biometrika*, 104(2):361–377, 2017.

Ravi B. Boppana. Eigenvalues and graph bisection: An average-case analysis. In *28th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 280–285, 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 (FOCS)*, pages 1347–1357, 2015.

Cécile Bothorel, Juan David Cruz, Matteo Magnani, and Barbora Micenková. Clustering attributed graphs: models, measures and methods. *Network Science*, 3(3):408–444, 2015.

Thang Bui, S. Chaudhuri, T. Leighton, and M. Sipser. Graph Bisection Algorithms With Good Average Case Behavior. In *25th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 181–192, 1984.

Shuxiao Chen, Sifan Liu, and Zongming Ma. Global and Individualized Community Detection in Inhomogeneous Multilayer Networks. Preprint available at https://arxiv.org/abs/2012.00933, 2020.

Daniel Cullina and Negar Kiyavash. Improved Achievability and Converse Bounds for Erdős-Rényi Graph Matching. In *ACM SIGMETRICS*, volume 44, pages 63–72, 2016.

Daniel Cullina and Negar Kiyavash. Exact alignment recovery for correlated Erdős-Rényi graphs. Preprint available at https://arxiv.org/abs/1711.06783, 2018.

Daniel Cullina, Kushagra Singhal, Negar Kiyavash, and Prateek Mittal. On the simultaneous preservation of privacy and community structure in anonymized networks. Preprint available at https://arxiv.org/abs/1603.08028, 2016.

Daniel Cullina, Negar Kiyavash, Prateek Mittal, and H. Vincent Poor. Partial Recovery of Erdős-Rényi Graph Alignment via k-Core Alignment. *SIGMETRICS Perform. Eval. Rev.*, 48(1):99–100, July 2020.

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.

Yash Deshpande, Subhabrata Sen, Andrea Montanari, and Elchanan Mossel. Contextual Stochastic Block Models. In *Advances in Neural Information Processing Systems (NeurIPS)*, pages 8581–8593, 2018.

Jian Ding, Zongming Ma, Yihong Wu, and Jiaming Xu. Efficient random graph matching via degree profiles. *Probability Theory and Related Fields*, 179(1):29–115, 2021.

M.E Dyer and A.M Frieze. The solution of some random NP-hard problems in polynomial expected time. *Journal of Algorithms*, 10(4):451–489, 1989.

Zhou Fan, Cheng Mao, Yihong Wu, and Jiaming Xu. Spectral graph matching and regularized quadratic relaxations: Algorithm and theory. In *Proceedings of the 37th International Conference on Machine Learning (ICML)*, volume 119 of *Proceedings of Machine Learning Research*, pages 2985–2995. PMLR, 13–18 Jul 2020.

Luca Ganassali and Laurent Massoulié. From tree matching to sparse graph alignment. In *Proceedings of the Thirty Third Conference on Learning Theory (COLT)*, volume 125 of *Proceedings of Machine Learning Research*, pages 1633–1665. PMLR, 09–12 Jul 2020.

Luca Ganassali, Laurent Massoulié, and Marc Lelarge. Impossibility of Partial Recovery in the Graph Alignment Problem. In *Conference on Learning Theory*, pages 2080–2102. PMLR, 2021.

Georgina Hall and Laurent Massoulié. Partial Recovery in the Graph Alignment Problem. Preprint available at https://arxiv.org/abs/2007.00533, 2020.

Qiuyi Han, Kevin Xu, and Edoardo Airoldi. Consistent estimation of dynamic and multi-layer block models. In *International Conference on Machine Learning (ICML)*, pages 1511–1520. PMLR, 2015.

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

Varun Kanade, Elchanan Mossel, and Tselil Schramm. Global and Local Information in Clustering Labeled Block Models. *IEEE Transactions on Information Theory*, 62(10):5906–5917, 2016.

Nitish Korula and Silvio Lattanzi. An efficient reconciliation algorithm for social networks. *Proceedings of the VLDB Endowment*, 7(5):377–388, 2014.

Jing Lei, Kehui Chen, and Brian Lynch. Consistent community detection in multi-layer network data. *Biometrika*, 107(1):61–73, 12 2019.

Chen Lu and Subhabrata Sen. Contextual stochastic block model: Sharp thresholds and contiguity. Preprint available at https://arxiv.org/abs/2011.09841, 2020.

Tomasz Łuczak. Size and connectivity of the k-core of a random graph. *Discrete Mathematics*, 91 (1):61–68, 1991.

Vince Lyzinski. Information Recovery in Shuffled Graphs via Graph Matching. *IEEE Transactions on Information Theory*, 64(5):3254–3273, 2018.

Zongming Ma and Sagnik Nandy. Community Detection with Contextual Multilayer Networks. Preprint available at https://arxiv.org/abs/2104.02960, 2021.

Cheng Mao, Mark Rudelson, and Konstantin Tikhomirov. Exact matching of random graphs with constant correlation. Preprint available at https://arxiv.org/abs/2110.05000, 2021a.

Cheng Mao, Mark Rudelson, and Konstantin Tikhomirov. Random Graph Matching with Improved Noise Robustness. In *Proceedings of the 34th Conference on Learning Theory (COLT)*, pages 3296–3329. PMLR, 2021b.

Laurent Massoulié. Community detection thresholds and the weak Ramanujan property. In *Proceedings of the 46th Annual ACM Symposium on Theory of Computing (STOC)*, pages 694–703. ACM, 2014.

Vaishakhi Mayya and Galen Reeves. Mutual information in community detection with covariate information and correlated networks. In *2019 57th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pages 602–607, 2019.

Elchanan Mossel and Jiaming Xu. Local Algorithms for Block Models with Side Information. In *Proceedings of the 2016 ACM Conference on Innovations in Theoretical Computer Science (ITCS)*, pages 71–80, 2016.

Elchanan Mossel and Jiaming Xu. Seeded graph matching via large neighborhood statistics. In *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 1005–1014, 2019.

Elchanan Mossel, Joe Neeman, and Allan Sly. Reconstruction and estimation in the planted partition model. *Probability Theory and Related Fields*, 162, 07 2014.

Elchanan Mossel, Joe Neeman, and Allan Sly. Consistency thresholds for the planted bisection model. *Electronic Journal of Probability*, 21(none):1 – 24, 2016.

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

Efe Onaran, Siddharth Garg, and Elza Erkip. Optimal de-anonymization in random graphs with community structure. In *2016 50th Asilomar Conference on Signals, Systems and Computers*, pages 709–713. IEEE, 2016.

Subhadeep Paul and Yuguo Chen. Spectral and matrix factorization methods for consistent community detection in multi-layer networks. *The Annals of Statistics*, 48(1):230 – 250, 2020.

Subhadeep Paul and Yuguo Chen. Null Models and Community Detection in Multi-Layer Networks. *Sankhya A*, pages 1–55, 2021.

Pedram Pedarsani and Matthias Grossglauser. On the privacy of anonymized networks. In *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, pages 1235–1243, 2011.

H. Vincent Poor. *An Introduction to Signal Detection and Estimation (2nd Ed.)*. Springer-Verlag, Berlin, Heidelberg, 1994. ISBN 0387941738.

Miklós Z. Rácz and Anirudh Sridhar. Correlated Stochastic Block Models: Exact Graph Matching with Applications to Recovering Communities. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2021.

Miklós Z. Rácz and Anirudh Sridhar. Correlated randomly growing graphs. *Annals of Applied Probability*, 32(2), 2022.

Hussein Saad and Aria Nosratinia. Recovering a single community with side information. *IEEE Transactions on Information Theory*, 66(12):7939–7966, 2020.

Farhad Shirani, Siddharth Garg, and Elza Erkip. A concentration of measure approach to correlated graph matching. *IEEE Journal on Selected Areas in Information Theory*, 2(1):338–351, 2021.

Yihong Wu, Jiaming Xu, and Sophie H. Yu. Settling the Sharp Reconstruction Thresholds of Random Graph Matching. Preprint available at https://arxiv.org/abs/2102.00082, 2021.

Bowei Yan and Purnamrita Sarkar. Covariate regularized community detection in sparse graphs. *Journal of the American Statistical Association*, 116(534):734–745, 2021.

Liren Yu, Jiaming Xu, and Xiaojun Lin. The Power of $D$-hops in Matching Power-Law Graphs. *Proceedings of the ACM on Measurement and Analysis of Computing Systems*, 5(2):1–43, 2021.

Yuan Zhang, Elizaveta Levina, and Ji Zhu. Community detection in networks with node features. *Electronic Journal of Statistics*, 10(2):3153–3178, 2016.

## Appendix A. Organization

The rest of the paper is devoted to the proofs of Theorems 1 and 2 and is structured as follows. First, after collecting notation in Section B, we describe the recovery algorithm in detail in Section C. Section D contains preliminary lemmas which are useful throughout, followed by three sections where the three main steps of the algorithm are analyzed: Section E contains the analysis of the $k$-core estimator, Section F proves the correctness of the estimated community labels for the matched vertices, and Section G deals with classifying the remaining vertices. The different steps are combined into a proof of Theorem 1 in Section H. Finally, Section I contains the proof of the impossibility result, Theorem 2.

## Appendix B. Notation

Recall that the underlying vertex set is $V = [n] := \{1, 2, \ldots, n\}$. We denote by $\mathcal{S}_n$ the set of permutations of $[n]$. Recall that $V^+ := \{i \in [n] : \sigma_*(i) = +1\}$ and $V^- := \{i \in [n] : \sigma_*(i) = -1\}$ denote the vertices in the two communities. To emphasize the different vertex labels in $G_1$ and $G_2$, we define $\boldsymbol{\sigma}_*^{\mathbf{1}} := \boldsymbol{\sigma}_*$ and $\boldsymbol{\sigma}_*^{\mathbf{2}} := \boldsymbol{\sigma}_* \circ \pi_*^{-1}$, which are the community labels in $G_1$ and $G_2$, respectively. Accordingly, we define $V_1^+ := V^+$ and $V_1^- := V^-$, as well as $V_2^+ := \{i \in [n] : \sigma_*^2(i) = +1\}$ and $V_2^- := \{i \in [n] : \sigma_*^2(i) = -1\}$, to denote the two communities in the two graphs.

Let $\binom{[n]}{2} := \{\{i, j\} : i, j \in [n], i \neq j\}$ denote the set of all unordered vertex pairs. We use $(i, j)$, $(j, i)$, and $\{i, j\}$ interchangeably to denote the unordered pair consisting of $i$ and $j$. Given a community labeling $\boldsymbol{\sigma}$, we define the sets $\mathcal{E}^+(\boldsymbol{\sigma}) := \left\{(i, j) \in \binom{[n]}{2} : \sigma(i)\sigma(j) = +1\right\}$ and $\mathcal{E}^-(\boldsymbol{\sigma}) := \left\{(i, j) \in \binom{[n]}{2} : \sigma(i)\sigma(j) = -1\right\}$. In words, $\mathcal{E}^+(\boldsymbol{\sigma})$ is the set of *intra-community* vertex pairs, and $\mathcal{E}^-(\boldsymbol{\sigma})$ is the set of *inter-community* vertex pairs. Note that $\mathcal{E}^+(\boldsymbol{\sigma})$ and $\mathcal{E}^-(\boldsymbol{\sigma})$ partition $\binom{[n]}{2}$.

Let $A$ be the adjacency matrix of $G_1$, let $B$ be the adjacency matrix of $G_2$, and let $B'$ be the adjacency matrix of $G_2'$. Note that, by construction, we have that $B'_{i,j} = B_{\pi_*(i), \pi_*(j)}$ for every $i, j$. By the construction of the correlated SBMs, we have the following probabilities for every $(i, j) \in \binom{[n]}{2}$:

$$\mathbb{P}\left((A_{i,j}, B'_{i,j}) = (1, 1) \,\middle|\, \boldsymbol{\sigma}_*\right) = \begin{cases} s^2 p & \text{if } \sigma_*(i) = \sigma_*(j), \\ s^2 q & \text{if } \sigma_*(i) \neq \sigma_*(j); \end{cases}$$

$$\mathbb{P}\left((A_{i,j}, B'_{i,j}) = (1, 0) \,\middle|\, \boldsymbol{\sigma}_*\right) = \begin{cases} s(1-s)p & \text{if } \sigma_*(i) = \sigma_*(j), \\ s(1-s)q & \text{if } \sigma_*(i) \neq \sigma_*(j); \end{cases}$$

$$\mathbb{P}\left((A_{i,j}, B'_{i,j}) = (0, 1) \,\middle|\, \boldsymbol{\sigma}_*\right) = \begin{cases} s(1-s)p & \text{if } \sigma_*(i) = \sigma_*(j), \\ s(1-s)q & \text{if } \sigma_*(i) \neq \sigma_*(j); \end{cases}$$

$$\mathbb{P}\left((A_{i,j}, B'_{i,j}) = (0, 0) \,\middle|\, \boldsymbol{\sigma}_*\right) = \begin{cases} 1 - p(2s - s^2) & \text{if } \sigma_*(i) = \sigma_*(j), \\ 1 - q(2s - s^2) & \text{if } \sigma_*(i) \neq \sigma_*(j). \end{cases}$$

For brevity, for $i, j \in \{0, 1\}$ we write

$$p_{ij} := \mathbb{P}\left((A_{1,2}, B'_{1,2}) = (i, j) \,\middle|\, \boldsymbol{\sigma}_*\right) \qquad \text{if } \sigma_*(1) = \sigma_*(2)$$

and

$$q_{ij} := \mathbb{P}\left(\left(A_{1,2}, B'_{1,2}\right) = (i,j) \,\middle|\, \boldsymbol{\sigma}_*\right) \qquad \text{if } \sigma_*(1) \neq \sigma_*(2).$$

We also utilize some common notation for general graphs $G$. If $R$ is a subset of the vertex set of $G$, we let $G\{R\}$ denote the induced subgraph of $G$ corresponding to $R$. For a vertex $i$ in $G$, we let $\mathcal{N}_G(i)$ be the set of neighbors of $i$ in $G$. We abbreviate $\mathcal{N}_{G_1}(i)$ as $\mathcal{N}_1(i)$, and similarly $\mathcal{N}_{G_2}(i)$ as $\mathcal{N}_2(i)$. We also let $\deg_G(i) = |\mathcal{N}_G(i)|$ denote the degree of $i$ in $G$.

For an event $\mathcal{A}$, we denote by $\mathbf{1}(\mathcal{A})$ the indicator of $\mathcal{A}$, which is 1 if $\mathcal{A}$ occurs and 0 otherwise. Given a function $f$ and a subset $M$ of its domain, we let $f\{M\}$ denote the restriction of $f$ to $M$. We also let $f(M)$ denote the image of $M$ under $f$.

Throughout the paper we use standard asymptotic notation and all limits are as $n \to \infty$.

## Appendix C. The recovery algorithm

Our recovery algorithm begins by forming a matching between a subset of the vertices in $G_1$ and a subset of the vertices in $G_2$. Formally, we have the following definitions of a matching and a $k$-core matching.

**Definition 3** *Let $G_1$ and $G_2$ be two graphs with vertex set $[n]$. The pair $(M, \mu)$ is a matching between $G_1$ and $G_2$ if*

- $M \subseteq [n]$,

- $\mu : M \to [n]$, *and*

- $\mu$ *is injective.*

Given a matching $(M, \mu)$, we introduce the following related notation. We let $G_1 \vee_\mu G_2$ be the *union graph*, whose vertex set is $M$, and whose edge set is $\{\{i,j\} : i,j \in M, A_{ij} + B_{\mu(i),\mu(j)} \geq 1\}$. In other words, the union graph contains edges that appear in either graph, relative to the matching $\mu$. Similarly, let $G_1 \wedge_\mu G_2$ be the *intersection graph*, whose vertex set is $M$, and whose edge set is $\{\{i,j\} : i,j \in M, A_{ij} = B_{\mu(i),\mu(j)} = 1\}$. The intersection graph contains all edges appearing in both graphs, relative to the matching $\mu$. Let $G_1 \setminus_\mu G_2$ be the graph whose vertex set is $M$, and whose edge set is $\{\{i,j\} : i,j \in M, A_{ij} = 1, B_{\mu(i),\mu(j)} = 0\}$. In other words, $G_1 \setminus_\mu G_2$ contains edges appearing in $G_1$ but not $G_2$, again, relative to the matching. Finally, let $G_2 \setminus_\mu G_1$ be the graph whose vertex set is $M$, and whose edge set is $\{\{i,j\} : i,j \in M, A_{ij} = 0, B_{\mu(i),\mu(j)} = 1\}$. Note that all four definitions use vertex numbering relative to $G_1$. If $\pi : [n] \to [n]$ is a permutation, then the notation $G_1 \vee_\pi G_2$, $G_1 \wedge_\pi G_2$, $G_1 \setminus_\pi G_2$, and $G_2 \setminus_\pi G_1$ is defined according to the matching $([n], \pi)$.

In order to introduce our matching algorithm, we require the following definition. We let $d_{\min}(G)$ be the minimal degree in a graph $G$.

**Definition 4** *A matching $(M, \mu)$ is a $k$-core matching of $(G_1, G_2)$ if $d_{\min}(G_1 \wedge_\mu G_2) \geq k$ (i.e., for every $i \in M$, the degree of $i$ in the graph $G_1 \wedge_\mu G_2$ is at least $k$). A matching $(M, \mu)$ is called a maximal $k$-core matching if it involves the greatest number of vertices, among all $k$-core matchings.*

The term *$k$-core matching* comes from the notion of a *$k$-core*. The $k$-core of a graph $G$ is the maximal subgraph with minimum degree at least $k$. Our first step is to produce a maximal $k$-core matching of the graphs $(G_1, G_2)$; see Figure 4 for an illustration.

---

**Algorithm 1** $k$-core matching

---

**Input:** Pair of graphs $(G_1, G_2)$ on $n$ vertices, $k \in [n]$.
**Output:** A matching $(\widehat{M}, \widehat{\mu})$ of $(G_1, G_2)$.
 1: By enumerating all possible matchings, find the maximal $k$-core matching $(\widehat{M}, \widehat{\mu})$ of $G_1$ and $G_2$.

---

Let $(\widehat{M}, \widehat{\mu})$ be the matching found by Algorithm 1 with $k = 13$. We will show that $\widehat{M}$ coincides with the maximal $k$-core of $G_1 \wedge_{\pi_*} G_2$, with high probability (see Lemma 21). Furthermore, we will show that a vanishing fraction of the vertices is excluded from this matching. Specifically, let $F := [n] \setminus \widehat{M}$ be the set of vertices which are excluded from the matching; we will show that $|F| \leq n^{1-s^2 \mathrm{T_c}(\alpha,\beta)+o(1)}$ with high probability (see Lemma 26).

Our next step is to leverage the correctly matched region in order to find the correct communities. In particular, the goal of the following algorithm is to perfectly recover the communities in the subset of vertices found by the $k$-core matching (with $k = 13$). The algorithm involves several intermediate steps which we detail below.

---

**Algorithm 2** Labeling the $k$-core

---

**Input:** Pair of graphs $(G_1, G_2)$ on $n$ vertices, a $k$-core matching $(M, \mu)$, parameters $\alpha, \beta, s, \epsilon > 0$.
**Output:** A labeling of $G_1\{M\}$.
  **Finding a preliminary labeling**
 1: Apply Algorithm 3 to the graph $G_1$ and parameters $(s\alpha, s\beta, \epsilon)$, obtaining a labeling $\widehat{\boldsymbol{\sigma}}_1$.

  **Expanding** $F := [n] \setminus M$
 2: Let $F := [n] \setminus M$. Apply Algorithm 4 to $(G_2, [n] \setminus \mu(M))$, obtaining the set $F' \supseteq [n] \setminus \mu(M)$. Let $\overline{F} := [n] \setminus \mu^{-1}([n] \setminus F')$, and note that $\overline{F} \supseteq F$.

  **Classifying** $[n] \setminus \overline{F}$ **and** $\overline{F} \setminus F$
 3: For each $i \in [n] \setminus \overline{F}$, set $\widehat{\sigma}(i) \in \{-1, 1\}$ according to the neighborhood majority (resp., minority) of $\widehat{\boldsymbol{\sigma}}_1$ with respect to the graph $(G_1 \vee_\mu G_2)\{[n] \setminus \overline{F}\}$ if $\alpha > \beta$ (resp., $\alpha < \beta$).
 4: For each $i \in \overline{F} \setminus F$, set $\widehat{\sigma}(i) \in \{-1, 1\}$ according to the neighborhood majority (resp., minority) of $\widehat{\boldsymbol{\sigma}}$ with respect to the graph $(G_1 \setminus_\mu G_2)\{([n] \setminus \overline{F}) \cup \{i\}\}$ if $\alpha > \beta$ (resp., $\alpha < \beta$).

 5: Return $\widehat{\boldsymbol{\sigma}} : M \to \{-1, 1\}$.

---

Step 1 in Algorithm 2 results in almost exact recovery of the community labels in $G_1$. We then need to transform the almost-exact labeling to an exact labeling. We would like to do this by a neighborhood vote in the union graph $G_1 \vee_{\widehat{\mu}} G_2$; however, the contribution of edges from vertices in $\widehat{M}$ to vertices in $F$ is hard to analyze. To remedy this difficulty, we expand the set $F$ into the set $\overline{F} \supseteq F$, such that any vertex of $G_2$ outside of $\pi_*(\overline{F})$ has at most one neighbor in $\pi_*(\overline{F})$, with high probability. This expansion, which is constructed by a method of Łuczak (1991), is guaranteed to satisfy $|\overline{F}| \leq 3|F|$ with high probability (see Lemma 25). Given the expanded set, we classify all $i \in [n] \setminus \overline{F}$ according to the majority of neighborhood values of $\widehat{\boldsymbol{\sigma}}_1$ in the union graph $(G_1 \vee_{\widehat{\mu}} G_2)\{[n] \setminus \overline{F}\}$. See Figure 5 for an illustration of this procedure.

In order to show that the majority vote correctly rectifies the labels, we first show a general result on almost-exact recovery in SBMs, which relates the set of incorrectly classified vertices to the set

of vertices with weak majorities with respect to the ground-truth labeling (Lemma 27), adapted from Mossel et al. (2016). Namely, let $J$ be the set of incorrectly classified vertices of a graph $G$ drawn from the SBM, and let $I_\epsilon(G)$ be the set of vertices in $G$ which do not have $\epsilon \log(n)$ majorities with respect to the ground-truth labeling $\boldsymbol{\sigma}_*$ (see Definitions 15 and 16 for formal definitions.) We will show that with high probability, for suitable $\epsilon > 0$, it holds that $J \subseteq I_\epsilon(G)$. Given this general result, it follows that the error set of $\widehat{\boldsymbol{\sigma}}_1$ is contained within the set $I_\epsilon(G_1)$. We then show that on the graphs $G_1$ and $(G_2 \setminus_{\widehat{\mu}} G_1)\{[n] \setminus \overline{F}\}$, each vertex has few neighbors in $I_\epsilon(G_1)$ (Lemmas 30 and 34). In turn, this allows us to show that on the union graph $(G_1 \vee_\mu G_2)\{[n] \setminus \overline{F}\}$, the neighborhood labels of a given vertex $i$ with respect to $\widehat{\boldsymbol{\sigma}}_1$ are close to the neighborhood labels with respect to $\boldsymbol{\sigma}_*$. Finally, we show that each vertex in $(G_1 \vee_\mu G_2)\{[n] \setminus \overline{F}\}$ has an $\epsilon \log n$ majority with respect to $\boldsymbol{\sigma}_*$ (Lemma 33). Therefore, taking a majority with respect to $\widehat{\boldsymbol{\sigma}}_1$ on $(G_1 \vee_\mu G_2)\{[n] \setminus \overline{F}\}$ transforms the almost-exact labeling to the correct labeling.

To complete the labeling of the $k$-core, we classify vertices in $\overline{F} \setminus F$ according to the majority of neighborhood values of $\widehat{\boldsymbol{\sigma}}\{[n] \setminus \overline{F}\}$, with respect to the graph $(G_1 \setminus_{\widehat{\mu}} G_2)\{([n] \setminus \overline{F}) \cup \{i\}\}$. We are able to do so because the edges in the graph $(G_1 \setminus_{\widehat{\mu}} G_2)\{([n] \setminus \overline{F}) \cup \{i\}\}$ are (nearly) independent of the construction of $\overline{F}$, conditioned on the random partition representation of correlated SBMs (see Section D.2 for details).

We now provide the details of the subroutines used by Algorithm 2.

---

**Algorithm 3** Almost-exact community recovery (Mossel et al., 2016, Algorithm 1)

---

**Input:** A graph $G$ on $n$ vertices, parameters $\alpha, \beta, \epsilon > 0$.
**Output:** A labeling on $G$ given by $\widehat{\boldsymbol{\sigma}} : [n] \to \{-1, 1\}$.
1: Choose a positive integer $m$ satisfying $(\log(\epsilon m(200 \max\{1, \alpha, \beta\})^{-1}) - 1)\epsilon/2 > 1$. Initialize two empty sets, $W_+$ and $W_-$.
2: Using the spectral method of (Abbe et al., 2020b, Section 3.2), find a community partition of $[n]$, denoted by $(U_+, U_-)$.
3: Partition $[n]$ randomly into $\{U_1, \ldots, U_m\}$.
4: For each $i \in [m]$:

- Using the spectral method of (Abbe et al., 2020b, Section 3.2), find a community partition $(U_{i,+}, U_{i,-})$ of $G\{[n] \setminus U_i\}$. If $|U_{i,+} \Delta U_+| \geq n/2$, then swap $U_{i,+}$ and $U_{i,-}$.

- For $v \in U_i$, insert $v$ into $W_+$ or $W_-$ according to its neighborhood majority (resp., minority) in $U_{i,+} \cup U_{i,-}$ if $\alpha > \beta$ (resp., $\alpha < \beta$).

5: For $i \in W_+$, set $\widehat{\sigma}(i) = 1$, and for $i \in W_-$, set $\widehat{\sigma}(i) = -1$. Return $\widehat{\boldsymbol{\sigma}}$.

---

**Algorithm 4** Łuczak expansion

---

**Input:** A graph $G$ on $n$ vertices, a set $U \subseteq [n]$
**Output:** A set $\overline{U} \subseteq [n]$ such that $\overline{U} \supseteq U$ and for all $i \in [n] \setminus \overline{U}$, $i$ has at most one neighbor in $\overline{U}$.
1: Let $U_0 = U$. For each $i \in \{0, 1, \ldots, n\}$:

- Let $U'_{i+1}$ be the set of vertices outside $U_i$ that have at least two neighbors in $U_i$.

- If $U'_{i+1} = \emptyset$ return $U_i$.

- Else set $U_{i+1} = U_i \cup \{v\}$, where $v$ is an arbitrarily chosen vertex in $U'_{i+1}$.

---

The complete algorithm appears below. First, the $k$-core matching is found. Next, the vertices comprising the matching are labeled. Finally, the vertices excluded from the matching are labeled according to neighborhood labels in the graph $G_1$, restricted to the vertices comprising the matching.

---

**Algorithm 5** Full community recovery

---

**Input:** Pair of graphs $(G_1, G_2)$ on $n$ vertices, $k \in [n]$, and $\epsilon > 0$.
**Output:** A labeling of $G_1$ given by $\widehat{\sigma} : [n] \to \{-1, 1\}$.
1: Apply Algorithm 1 on input $(G_1, G_2, k)$, obtaining a matching $(\widehat{M}, \widehat{\mu})$.
2: Apply Algorithm 2 on input $(G_1, G_2, \widehat{M}, \widehat{\mu}, \epsilon)$, obtaining a labeling $\widehat{\sigma} : \widehat{M} \to \{-1, 1\}$.
3: For $i \in [n] \setminus \widehat{M}$, classify $i$ according to its neighborhood majority (resp., minority) in the graph $G_1\{\widehat{M} \cup \{i\}\}$ if $\alpha > \beta$ (resp., $\alpha < \beta$).

---

**Theorem 5** *Fix constants $\alpha, \beta > 0$ and $s \in [0, 1]$. Let $(G_1, G_2) \sim \mathrm{CSBM}\left(n, \frac{\alpha \log n}{n}, \frac{\beta \log n}{n}, s\right)$. Suppose that (6) and (7) hold. Let $\epsilon > 0$ satisfy*

$$\left(1 - (1 - s)^2\right) \mathrm{D}_+(\alpha, \beta) > 1 + 2\epsilon |\log(\alpha/\beta)| \qquad and \qquad 0 < \epsilon \le \frac{s \mathrm{D}_+(\alpha, \beta)}{4 |\log(\alpha/\beta)|}. \qquad (10)$$

*Then Algorithm 5 on input $(G_1, G_2, 13, \epsilon)$ correctly labels all of the vertices in $G_1$, with high probability.*

Since (6) and (7) imply the existence of $\epsilon > 0$ such that (10) holds, proving Theorem 1 reduces to proving Theorem 5. The proof of Theorem 5 can be found in Section H, with the supporting results proved in Sections E, F, and G.

## Appendix D. Preliminary results

### D.1. Binomial probabilities

Since binomial differences appear frequently in our analysis, we include two useful results here. The first can be found in the proof of (Mossel et al., 2016, Proposition 2.8).

**Lemma 6** *Suppose that $\alpha \ge \beta$. Let $Y \sim \mathrm{Bin}(m^+, \alpha \log(n)/n)$ and $Z \sim \mathrm{Bin}(m^-, \beta \log(n)/n)$ be independent. If $m^+ = (1 + o(1))n/2$ and $m^- = (1 + o(1))n/2$, then*

$$\mathbb{P}(Y < Z) = n^{-\mathrm{D}_+(\alpha, \beta) + o(1)}.$$

**Remark 7** *The precise result in (Mossel et al., 2016, Proposition 2.8) characterizes $\mathbb{P}(Y \le Z)$ when $m^+ = m^- = n/2$. It is readily seen from their proof of the result that it also holds for the slightly more general values of $m^+$ and $m^-$ that we consider and when the inequality between $Y$ and $Z$ is strict.*

The following result follows from (Abbe et al., 2020b, Lemma 8 of Supplement).

**Lemma 8** *Suppose that $\alpha > \beta$. Let $Y \sim \mathrm{Bin}\left(m_+, \frac{\alpha \log n}{n}\right)$ and $Z \sim \mathrm{Bin}\left(m_-, \frac{\beta \log n}{n}\right)$ be independent. If $m^+ = (1 + o(1))n/2$ and $m^- = (1 + o(1))n/2$, then for any $\epsilon > 0$,*

$$\mathbb{P}\left(Y - Z \le \epsilon \log n\right) \le n^{-\left(\mathrm{D}_+(\alpha, \beta) - \frac{\epsilon \log(\alpha/\beta)}{2}\right) + o(1)}.$$
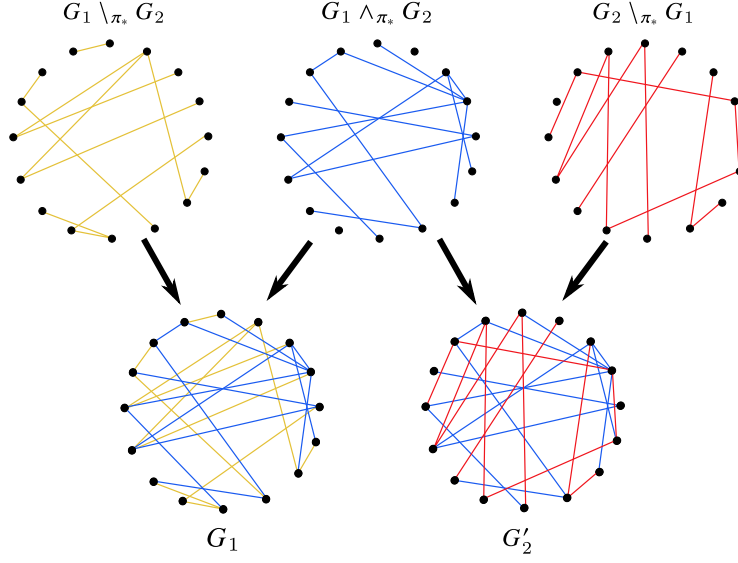
Figure 6: Decomposition of correlated $G_1$ and $G_2'$ into $G_1 \setminus_{\pi_*} G_2$, $G_1 \wedge_{\pi_*} G_2$, and $G_2 \setminus_{\pi_*} G_1$, which are conditionally independent given $\pi_*$, $\boldsymbol{\sigma_*}$, and the random partition $\{\mathcal{E}_{00}, \mathcal{E}_{10}, \mathcal{E}_{11}, \mathcal{E}_{01}\}$ (see Lemma 10).

**Remark 9** *Note that (Abbe et al., 2020b, Lemma 8 of Supplement) treats the case where $m_+ = m_- = n/2$, obtaining a tail bound of $n^{-\left(D_+(\alpha,\beta) - \frac{\epsilon \log(\alpha/\beta)}{2}\right)}$. Allowing for $m_+, m_- = (1 + o(1))n/2$ is reflected by an additional $o(1)$ in the exponent.*

### D.2. A useful construction of correlated SBMs

In this section, we detail a useful alternate construction of correlated SBMs that highlights the independent regions of $G_1$ and $G_2$.

To begin, we construct a random partition $\{\mathcal{E}_{00}, \mathcal{E}_{01}, \mathcal{E}_{10}, \mathcal{E}_{11}\}$ of $\binom{[n]}{2}$ as follows. Independently for each $\{i, j\} \in \binom{[n]}{2}$, we add $\{i, j\}$ to $\mathcal{E}_{00}$ with probability $(1 - s)^2$, to $\mathcal{E}_{01}$ with probability $s(1 - s)$, to $\mathcal{E}_{10}$ with probability $s(1 - s)$, and to $\mathcal{E}_{11}$ with probability $s^2$. Subsequently, for each pair $\{i, j\}$, we construct an edge between $i$ and $j$ with probability $p$ if the two vertices are in the same community, else with probability $q$ if the two vertices are in different communities. The graph $G_1$ is constructed using the edges formed in $\mathcal{E}_{10} \cup \mathcal{E}_{11}$ and the graph $G_2'$ is constructed using the edges formed in $\mathcal{E}_{01} \cup \mathcal{E}_{11}$. The graph $G_2$ is then generated from $G_2'$ and $\pi_*$ by relabeling the vertices of $G_2'$ according to $\pi_*$. The usefulness of this construction is that it provides an alternate way to generate correlated SBMs and highlights regions of the two graphs that are independent of each other (see Figure 6 for an illustration). This is formally stated in the following result.

**Lemma 10** *The random partition construction of correlated SBMs above is equivalent to the original construction described in Section 2. Moreover, conditioned on $\pi_*$, $\boldsymbol{\sigma_*}$, $\mathcal{E}_{00}$, $\mathcal{E}_{01}$, $\mathcal{E}_{10}$, and $\mathcal{E}_{11}$, the graphs $G_1 \setminus_{\pi_*} G_2$, $G_1 \wedge_{\pi_*} G_2$, and $G_2 \setminus_{\pi_*} G_1$ are mutually independent.*

**Proof** To prove the first statement, it suffices to check that the marginal distribution of the pair $(A_{i,j}, B_{\pi_*(i),\pi_*(j)})$ is the same under both constructions, due to the independence over vertex pairs. Under the original construction, if $\sigma_*^1(i) = \sigma_*^1(j)$, then

$$\mathbb{P}\left((A_{i,j}, B_{\pi_*(i),\pi_*(j)}) = (a, b) \,\middle|\, \pi_*, \boldsymbol{\sigma_*^1}\right) = \begin{cases} s^2 p & \text{if } (a, b) = (1, 1), \\ s(1 - s)p & \text{if } (a, b) \in \{(1, 0), (0, 1)\}, \\ 1 - \left(1 - (1 - s)^2\right) p & \text{if } (a, b) = (0, 0). \end{cases} \tag{11}$$

When $\sigma_*^1(i) = -\sigma_*^1(j)$, the joint distribution is the same as in (11), with $p$ replaced by $q$. In the second construction, if $(a, b) \neq (0, 0)$, then we have, for $i, j$ satisfying $\sigma_*^1(i) = \sigma_*^1(j)$, that

$$\mathbb{P}\left((A_{i,j}, B_{\pi_*(i),\pi_*(j)}) = (a, b) \,\middle|\, \pi_*, \boldsymbol{\sigma_*^1}\right) = \mathbb{P}\left(\{i, j\} \in \mathcal{E}_{ab} \,\middle|\, \pi_*, \boldsymbol{\sigma_*^1}\right) \cdot p$$
$$= \begin{cases} s^2 p & \text{if } (a, b) = (1, 1), \\ s(1 - s)p & \text{if } (a, b) \in \{(1, 0), (0, 1)\}. \end{cases} \tag{12}$$

When $\sigma_*^1(i) = -\sigma_*^1(j)$, the joint distribution is the same as in (12), with $p$ replaced by $q$. Since the probabilities agree when $(a, b) \neq (0, 0)$, they must also agree when $(a, b) = (0, 0)$, since all the probabilities sum to one.

To prove the second statement, notice that the graph $G_1 \setminus_{\pi_*} G_2$ is comprised of edges in $\mathcal{E}_{10}$, $G_1 \wedge_{\pi_*} G_2$ is comprised of edges in $\mathcal{E}_{11}$, and $G_2 \setminus_{\pi_*} G_1$ is comprised of edges in $\mathcal{E}_{01}$. Conditioned on $\pi_*$, $\boldsymbol{\sigma_*^1}$, and the random partition $\{\mathcal{E}_{00}, \mathcal{E}_{10}, \mathcal{E}_{11}, \mathcal{E}_{01}\}$, the formation of edges in $\mathcal{E}_{10}$, $\mathcal{E}_{11}$, and $\mathcal{E}_{01}$ are mutually independent. ∎

**Remark 11** *A useful consequence of the random partition construction is that marginally, the graphs $G_1 \setminus_{\pi_*} G_2$, $G_1 \wedge_{\pi_*} G_2$, and $G_2 \setminus_{\pi_*} G_1$ are SBMs. However, conditioned on the partition $\{\mathcal{E}_{00}, \mathcal{E}_{01}, \mathcal{E}_{10}, \mathcal{E}_{11}\}$, these graphs are* not *SBMs, since, for instance, the set of potential edges in $G_1 \setminus_{\pi_*} G_2$ is given by $\mathcal{E}_{10}$ rather than $\binom{[n]}{2}$.*

When we utilize the random partition construction in later proofs, it is useful to consider a high-probability event where the partition behaves in a nice manner. Formally, we introduce the following event.

**Definition 12 (The event $\mathcal{F}$)** *For $a, b \in \{0, 1\}$, define the constant*

$$s_{ab} := \begin{cases} s^2 & \text{if } (a, b) = (1, 1), \\ s(1 - s) & \text{if } (a, b) \in \{(1, 0), (0, 1)\}, \\ (1 - s)^2 & \text{if } (a, b) = (0, 0). \end{cases}$$

*The event $\mathcal{F}$ holds if and only if*

$$n/2 - n^{3/4} \leq |V^+|, |V^-| \leq n/2 + n^{3/4}$$

*and also the following conditions hold for all $a, b \in \{0, 1\}$ and all $i \in [n]$:*

- $s_{ab}\left(|V^{\sigma_*(i)}| - n^{3/4}\right) \leq |\{j : \{i, j\} \in \mathcal{E}_{ab} \cap \mathcal{E}^+(\boldsymbol{\sigma_*^1})\}| \leq s_{ab}\left(|V^{\sigma_*(i)}| + n^{3/4}\right);$

- $s_{ab}\left(|V^{-\sigma_*(i)}| - n^{3/4}\right) \leq |\{j : \{i,j\} \in \mathcal{E}_{ab} \cap \mathcal{E}^-(\boldsymbol{\sigma}_*^1)\}| \leq s_{ab}\left(|V^{-\sigma_*(i)}| + n^{3/4}\right).$

The first condition ensures that the two communities are balanced. The next two conditions stipulate that for each vertex $i \in [n]$, the number of potential neighbors to both communities in each component of the partition is balanced (and approximately the expected size, according to the weights $\{s_{ab}\}_{a,b\in\{0,1\}}$). The following result shows that $\mathcal{F}$ holds with high probability.

**Lemma 13** *Define $s_{\min} := \min_{a,b\in\{0,1\}} s_{ab}$. Then $\mathbb{P}(\mathcal{F}^c) \leq 100n e^{-\frac{1}{2}s_{\min}^2\sqrt{n}}$.*

**Proof** Define the event

$$\mathcal{G} := \left\{ ||V^+| - n/2|, ||V^-| - n/2| < n^{3/4} \right\}.$$

By Hoeffding's inequality, we have that

$$\mathbb{P}\left(||V^+| - n/2| \geq n^{3/4}\right), \mathbb{P}\left(||V^-| - n/2| \geq n^{3/4}\right) \leq 2e^{-\sqrt{n}}.$$

Thus, by a union bound, we have that $\mathbb{P}(\mathcal{G}^c) \leq 4e^{-\sqrt{n}}$.

We next study the remaining conditions of the event $\mathcal{F}$. We assume that $s \in (0,1)$, and hence $s_{ab} \in (0,1)$ for all $a, b \in \{0,1\}$; for $s \in \{0,1\}$ we have that $\mathcal{G} \subseteq \mathcal{F}$, so we are done. Condition on $\pi_*$ and $\boldsymbol{\sigma}_*^1$. Fix $i \in [n]$. Notice that

$$K_{ab}^+(i) := |\{j : \{i,j\} \in \mathcal{E}_{ab} \cap \mathcal{E}^+(\boldsymbol{\sigma}_*^1)\}| \sim \mathrm{Bin}\left(|V^{\sigma_*(i)}| - 1, s_{ab}\right).$$

Hoeffding's inequality implies that

$$\mathbb{P}\left(\left|K_{ab}^+(i) - s_{ab}\left(|V^{\sigma_*(i)}| - 1\right)\right| \geq \frac{s_{ab}}{2}n^{3/4} \,\middle|\, \pi_*, \boldsymbol{\sigma}_*\right)\mathbf{1}(\mathcal{G}) \leq 2\exp\left(-\frac{s_{ab}^2 n^{3/2}}{2|V^{\sigma_*(i)}|}\right)\mathbf{1}(\mathcal{G})$$
$$\leq 2e^{-(1-o(1))s_{ab}^2\sqrt{n}}.$$

Taking a union bound over $i \in [n]$ shows that

$$\mathbb{P}\left(\exists i \in [n] : |K_{ab}^+(i) - s_{ab}|V^{\sigma_*(i)}|| \geq s_{ab}n^{3/4}\right)$$
$$\leq \sum_{i=1}^n \mathbb{P}\left(|K_{ab}^+(i) - s_{ab}(|V^{\sigma_*(i)}| - 1)| \geq \frac{s_{ab}}{2}n^{3/4}\right)$$
$$\leq \sum_{i=1}^n \left(\mathbb{E}\left[\mathbb{P}\left(|K_{ab}^+(i) - s_{ab}(|V^{\sigma_*(i)}| - 1)| \geq \frac{s_{ab}}{2}n^{3/4} \,\middle|\, \pi_*, \boldsymbol{\sigma}_*\right)\mathbf{1}(\mathcal{G})\right] + \mathbb{P}(\mathcal{G}^c)\right)$$
$$\leq 6n e^{-(1-o(1))s_{ab}^2\sqrt{n}}.$$

Similarly, defining

$$K_{ab}^-(i) := |\{j : \{i,j\} \in \mathcal{E}_{ab} \cap \mathcal{E}^-(\boldsymbol{\sigma}_*^1)\}| \sim \mathrm{Bin}\left(|V^{-\sigma_*(i)}|, s_{ab}\right),$$

an identical analysis shows that

$$\mathbb{P}\left(\exists i \in [n] : |K_{ab}^-(i) - s_{ab}|V^{-\sigma_*(i)}|| \geq n^{3/4}\right) \leq 6n e^{-(1-o(1))s_{ab}^2\sqrt{n}}.$$

The conclusion then follows by a union bound. ■

## Appendix E. Analysis of the $k$-core estimator (Algorithm 1)

### E.1. General results on correctness

Recall the definition of a matching (Definition 3) and a $k$-core matching (Definition 4). We now introduce some additional definitions. Given a graph $G$ with vertex set $[n]$, let $\deg_G(i)$ be the degree of vertex $i \in [n]$ within $G$. For a matching $(M, \mu)$, define

$$f(M, \mu, G_1, G_2, \pi_*) := \sum_{i \in M : \mu(i) \neq \pi_*(i)} \deg_{G_1 \wedge_\mu G_2}(i).$$

For brevity, we sometimes write $f(\mu)$. In words, this is the sum, over vertices that are incorrectly matched by $\mu$, of degrees in the intersection graph $G_1 \wedge_\mu G_2$.

**Definition 14** *Given two graphs $(G_1, G_2)$ with vertex set $[n]$, a matching $(M, \mu)$ is a* weak $k$-core *matching of $(G_1, G_2)$ if $f(\mu) \geq k |\{i \in M : \mu(i) \neq \pi_*(i)\}|$.*

In other words, a matching $(M, \mu)$ is a weak $k$-core matching if the *average degree* of an incorrectly-matched vertex is at least $k$. Note that if $(M, \mu)$ is a $k$-core matching, then it is also a weak $k$-core matching.

Finally, we introduce the notion of a maximal matching.

**Definition 15** *A matching $(M, \mu)$ of $(G_1, G_2)$ is $\pi_*$-maximal if for every $i \in [n]$, either $i \in M$ or $\pi_*(i) \in \mu(M)$, where $\mu(M)$ is the image of $M$ under $\mu$.*

We can always extend a matching $(M, \mu)$ to a $\pi_*$-maximal matching by adding all possible input-output pairs of the form $(i, \pi_*(i))$ to $(M, \mu)$ that do not break the one-to-one property of $\mu$; see Lemma 16 for details. Moreover, this extension preserves the weak $k$-core matching property.

**Lemma 16** *For any matching $(M, \mu)$ of $(G_1, G_2)$, there exists a $\pi_*$-maximal matching $(M', \mu')$ such that $f(\mu) \leq f(\mu')$.*

**Proof** If $(M, \mu)$ is not $\pi_*$-maximal, then the set

$$S := \{i \in [n] : i \notin M \text{ and } \pi_*(i) \notin \mu(M)\}$$

is nonempty. We can then construct $(M', \mu')$ by defining $M' := M \cup S$, and $\mu'(i) := \mu(i)$ for $i \in M$ and $\mu'(i) := \pi_*(i)$ for $i \in S$. Clearly, $(M', \mu')$ is $\pi_*$-maximal by construction. Furthermore,

$$\begin{aligned} f(\mu') &= \sum_{i \in M' : \mu'(i) \neq \pi_*(i)} \deg_{G_1 \wedge_{\mu'} G_2}(i) = \sum_{i \in M : \mu'(i) \neq \pi_*(i)} \deg_{G_1 \wedge_{\mu'} G_2}(i) \\ &\geq \sum_{i \in M : \mu(i) \neq \pi_*(i)} \deg_{G_1 \wedge_\mu G_2}(i) = f(\mu). \end{aligned}$$

Above, the second equality follows since $\mu'(i) = \pi_*(i)$ for $i \in S$, and the inequality on the last line follows since $\mu(i) = \mu'(i)$ for $i \in M$ and also $\deg_{G_1 \wedge_{\mu'} G_2}(i) \geq \deg_{G_1 \wedge_\mu G_2}(i)$, since $G_1 \wedge_\mu G_2$ is a subgraph of $G_1 \wedge_{\mu'} G_2$. ∎

We also define the set of $\pi_*$-maximal matchings with $d$ errors:

$$\mathcal{M}(d) := \{(M, \mu) : (M, \mu) \text{ is } \pi_*\text{-maximal and } |\{i \in M : \mu(i) \neq \pi_*(i)\}| = d\}.$$

**Remark 17** *While the set $\mathcal{M}(d)$ depends on $\pi_*$ (since errors are measured relative to $\pi_*$), the set $\{(M, \pi_*^{-1} \circ \mu) : (M, \mu) \in \mathcal{M}(d)\}$ does* not *depend on $\pi_*$. Therefore, certain properties of $\mathcal{M}(d)$ also do not depend on $\pi_*$. In particular, its size $|\mathcal{M}(d)|$ is a deterministic function of $n$ and $d$.*

The usefulness of defining the set $\mathcal{M}(d)$ is that it is much smaller than the set of matchings with $d$ errors (without the maximality condition). For instance, when $d = 1$, the number of matchings is at least the number of ways to choose $M$, which in turn is $2^n - 1$. The following lemma shows that, on the other hand, the size of $\mathcal{M}(d)$ is polynomial in $n$. This was previously proven in Cullina et al. (2020), and provided here for completeness since the proof is short.

**Lemma 18** *For any $1 \leq d \leq n$, we have that $|\mathcal{M}(d)| \leq n^{2d}/d!$.*

**Proof** For a $\pi_*$-maximal $(M, \mu)$, define $M' := \{i \in M : \mu(i) \neq \pi_*(i)\}$, which is the set of vertices in $G_1$ that are incorrectly matched by $\mu$. By the definition of a $\pi_*$-maximal matching, $(M, \mu)$ is fully specified by $\{(i, \mu(i))\}_{i \in M'}$. Moreover, $|M'| = |\mu(M')| = d$ for $(M, \mu) \in \mathcal{M}(d)$. The number of ways to choose $M'$ and $\mu(M')$ is $\binom{n}{d}^2$, and the number of potential bijections between $M'$ and $\mu(M')$ is *at most* $d!$. The reason why this is an upper bound (and not an equality) is that based on the choice of $M'$ and $\mu(M')$, certain bijections may not be legal (e.g., if $\mu(M') = \pi_*(M')$ then the bijection must be chosen so that $\mu(i) \neq \pi_*(i)$ for $i \in M'$). Putting everything together shows that

$$|\mathcal{M}(d)| \leq d! \binom{n}{d}^2 \leq d! \left(\frac{n^d}{d!}\right)^2 = \frac{n^{2d}}{d!}.$$

∎

Using Lemmas 16 and 18, Cullina et al. (2020) proved the following result, which provides conditions under which any $k$-core matching is guaranteed to be correct. Since the proof is short, we provide it here for completeness. We emphasize that the following result applies to any distribution over a pair of graphs $(G_1, G_2)$ on $n$ vertices; subsequently, in Section E.2, we will apply this result (and Corollary 20) to correlated SBMs.

**Lemma 19** *Let $(G_1, G_2)$ be a pair of random graphs on the vertex set $[n]$ with ground-truth matching $\pi_*$. For any positive integer $k$, define the quantity*

$$\xi := \max_{1 \leq d \leq n} \max_{(M, \mu) \in \mathcal{M}(d)} \mathbb{P}\left(f(\mu) \geq kd\right)^{1/d}.$$

*Then*

$$\mathbb{P}\left(\exists (M, \mu) : d_{\min}(G_1 \wedge_\mu G_2) \geq k \text{ and } \exists i \in M \text{ s.t. } \mu(i) \neq \pi_*(i)\right) \leq e^{n^2 \xi} - 1.$$

We note, as a continuation of Remark 17, that while $\pi_*$ appears in the definition of $\xi$, in fact, $\xi$ is a deterministic constant that does not depend on $\pi_*$.

**Proof** We can rewrite the probability of interest by segmenting the event $\{\exists i \in M \text{ s.t. } \mu(i) \neq \pi_*(i)\}$ according to the number of discrepancies between $\mu$ and $\pi_*$:

$$\mathbb{P}\left(\exists (M, \mu) : d_{\min}(G_1 \wedge_\mu G_2) \geq k \text{ and } \exists i \in M \text{ s.t. } \mu(i) \neq \pi_*(i)\right)$$

$$= \sum_{d=1}^n \mathbb{P}\left(\exists (M, \mu) : d_{\min}(G_1 \wedge_\mu G_2) \geq k \text{ and } |\{i \in M : \mu(i) \neq \pi_*(i)\}| = d\right)$$

$$\overset{(a)}{\leq} \sum_{d=1}^n \mathbb{P}\left(\exists (M, \mu) : f(\mu) \geq kd \text{ and } |\{i \in M : \mu(i) \neq \pi_*(i)\}| = d\right)$$

$$\overset{(b)}{=} \sum_{d=1}^n \mathbb{P}\left(\exists (M, \mu) \in \mathcal{M}(d) : f(\mu) \geq kd\right),$$

where $(a)$ follows since any $k$-core matching is also a weak $k$-core matching and $(b)$ follows from the dominance of $\pi_*$-maximal matchings established in Lemma 16. By a union bound we have that

$$\mathbb{P}\left(\exists (M, \mu) \in \mathcal{M}(d) : f(\mu) \geq kd\right) \leq |\mathcal{M}(d)| \max_{(M,\mu) \in \mathcal{M}(d)} \mathbb{P}\left(f(\mu) \geq kd\right) \leq \frac{(n^2 \xi)^d}{d!},$$

where in the second inequality we used Lemma 18 and the definition of $\xi$. The conclusion follows by combining the two displays above and summing over $d \in [n]$. ∎

Lemma 19 implies the following useful corollary.

**Corollary 20** *Recalling the definitions of $(\widehat{M}, \widehat{\mu})$ from Algorithm 1 and of $\xi$ from Lemma 19, we have that*

$$\mathbb{P}\left(\widehat{M} \text{ is the } k\text{-core of } G_1 \wedge_{\pi_*} G_2 \text{ and } \widehat{\mu}\{\widehat{M}\} = \pi_*\{\widehat{M}\}\right) \geq 2 - \exp(n^2 \xi).$$

**Proof** Let $M_*$ be the vertex set of the $k$-core of $G_1 \wedge_{\pi_*} G_2$ and observe that $(M_*, \pi_*\{M_*\})$ is a $k$-core matching. Let us now define the event

$$\mathcal{H} := \{\text{for all } k\text{-core matchings } (M, \mu), \mu(i) = \pi_*(i) \text{ for all } i \in M\}.$$

We first claim that $\mathcal{H}$ implies that $\widehat{M}$ is the $k$-core of $G_1 \wedge_{\pi_*} G_2$, and thus

$$\mathbb{P}\left(\widehat{M} \text{ is the } k\text{-core of } G_1 \wedge_{\pi_*} G_2 \text{ and } \widehat{\mu}\{\widehat{M}\} = \pi_*\{\widehat{M}\}\right) \geq \mathbb{P}(\mathcal{H}).$$

To show this, first note that $\mathcal{H}$ implies $\widehat{\mu}\{\widehat{M}\} = \pi_*\{\widehat{M}\}$. To show that $\widehat{M}$ is the $k$-core of $G_1 \wedge_{\pi_*} G_2$, suppose there existed a $k$-core matching $(M, \mu)$ for which $|M \setminus M_*| > 0$. On the event $\mathcal{H}$, we have that $\mu\{M\} = \pi_*\{M\}$, which implies that the subgraph of $G_1 \wedge_{\pi_*} G_2$ corresponding to $M$ has minimum degree $k$. This in turn implies that the subgraph of $G_1 \wedge_{\pi_*} G_2$ corresponding to $M \cup M_*$ has minimum degree $k$, which contradicts the maximality of $M_*$. Hence on $\mathcal{H}$, $(M_*, \pi_*\{M_*\})$ is the maximum $k$-core matching. To conclude, note that Lemma 19 implies that

$$\mathbb{P}(\mathcal{H}) \geq 1 - (\exp(n^2 \xi) - 1) = 2 - \exp(n^2 \xi).$$

∎

### E.2. Correctness of the $k$-core estimator for correlated SBMs

The main result of this section is the following lemma.

**Lemma 21** *Fix constants $\alpha, \beta > 0$ and $s \in [0,1]$. Let $(G_1, G_2) \sim \mathrm{CSBM}\left(n, \frac{\alpha \log n}{n}, \frac{\beta \log n}{n}, s\right)$. Let $M_*$ be the set of vertices of the $13$-core in the graph $G_1 \wedge_{\pi_*} G_2$. Let $(\widehat{M}, \widehat{\mu})$ be the output of Algorithm 1, with $k = 13$. Then $\mathbb{P}\left((\widehat{M}, \widehat{\mu}) = (M_*, \pi_*\{M_*\})\right) = 1 - o(1)$.*

**Remark 22** *Therefore, any result that holds with high probability for $(M_*, \pi_*\{M_*\})$ also holds with high probability for $(\widehat{M}, \widehat{\mu})$, so we can effectively replace $(\widehat{M}, \widehat{\mu})$ by $(M_*, \pi_*\{M_*\})$ in any analysis.*

In light of Corollary 20, the strategy to show that the $k$-core estimator matches the $k$-core of $G_1 \wedge_{\pi_*} G_2$ is to bound $\xi$ for correlated SBMs. Previously, Cullina et al. (2020) proved a version of Lemma 21 for correlated Erdős–Rényi graphs with constant average degree. Their methods provided a fairly tight characterization of $\xi$ in that setting using generating functions corresponding to $f(\mu)$. While it is natural to expect that it is possible to generalize their arguments to correlated SBMs, our proof presents a simpler, looser approach that avoids dealing with complicated generating functions. The proof of Lemma 21 relies on the following result, which bounds $\mathbb{P}(f(\mu) \geq kd)$ for SBMs with general parameters.

**Lemma 23** *Fix $n > 0$. Given parameters $p, q, s \in [0,1]$, let $(G_1, G_2) \sim \mathrm{CSBM}(n, p, q, s)$. Let $d, k \in [n]$. For any matching $(M, \mu) \in \mathcal{M}(d)$ and any $\theta > 0$, we have that*

$$\mathbb{P}\left(f(\mu) \geq kd\right) \leq 3 \exp\left[d\left(-\theta k + \frac{1}{2} s^2 \gamma \left(e^{2\theta} - 1\right) + n(s\gamma)^2 \left(e^{6\theta} - 1\right)\right)\right],$$

*where $\gamma = \max\{p, q\}$.*

**Proof** Define the following sets:

$$\mathcal{A}(\mu) := \left\{(i,j) \in [n]^2 : i \in M \text{ and } \mu(i) \neq \pi_*(i)\right\};$$
$$\mathcal{B}(\mu) := \left\{(i,j) \in [n]^2 : \mu(i) = \pi_*(j) \text{ and } \mu(j) = \pi_*(i)\right\};$$
$$\mathcal{C}(\mu) := \mathcal{A}(\mu) \setminus \mathcal{B}(\mu).$$

In words, $\mathcal{A}(\mu)$ is the set of vertex pairs where one of the endpoints is mismatched by $\mu$; $\mathcal{B}(\mu)$ is the set of vertex pairs in $\mathcal{A}(\mu)$ where the vertices are transposed by $\mu$, resulting in a *correctly* matched edge; and $\mathcal{C}(\mu)$ is the set of remaining vertex pairs in $\mathcal{A}(\mu)$ that induce *incorrectly* matched edges. It is useful to note that for $(i,j) \in \mathcal{B}(\mu)$, both $i$ and $j$ are one of the $d$ misclassified vertices, so $|\mathcal{B}(\mu)| \leq d$. Furthermore, a simple counting argument shows that $|\mathcal{A}(\mu)| \leq dn$.

Using the sets $\mathcal{A}(\mu)$, $\mathcal{B}(\mu)$, and $\mathcal{C}(\mu)$, we can write

$$f(\mu) = \sum_{i \in M : \mu(i) \neq \pi_*(i)} \deg_{G_1 \wedge_\mu G_2}(i) = \sum_{(i,j) \in \mathcal{A}(\mu)} A_{ij} B_{\mu(i)\mu(j)}$$
$$= \sum_{(i,j) \in \mathcal{B}(\mu)} A_{ij} B_{\mu(i)\mu(j)} + \sum_{(i,j) \in \mathcal{C}(\mu)} A_{ij} B_{\mu(i)\mu(j)}$$
$$= 2 \sum_{(i,j) \in \mathcal{B}(\mu) : i < j} A_{ij} B_{\mu(i)\mu(j)} + \sum_{(i,j) \in \mathcal{C}(\mu)} A_{ij} B_{\mu(i)\mu(j)}.$$

28

Recall that $\mathcal{E}^+(\boldsymbol{\sigma}_*)$ is the set of intra-community vertex pairs and $\mathcal{E}^-(\boldsymbol{\sigma}_*)$ is the set of inter-community vertex pairs. For $(i, j) \in \mathcal{B}(\mu)$, we have, conditionally on $\boldsymbol{\sigma}_*$, that

$$A_{ij} B_{\mu(i)\mu(j)} = A_{ij} B_{\pi_*(i)\pi_*(j)} \sim \begin{cases} \text{Bern}\left(s^2 p\right) & \text{if } (i, j) \in \mathcal{E}^+(\boldsymbol{\sigma}_*), \\ \text{Bern}\left(s^2 q\right) & \text{if } (i, j) \in \mathcal{E}^-(\boldsymbol{\sigma}_*). \end{cases}$$

Recalling that $\gamma = \max\{p, q\}$, let $X_{\mathcal{B}} \sim \text{Bin}\left(\frac{1}{2}|\mathcal{B}(\mu)|, s^2\gamma\right)$. We then have the stochastic domination

$$\sum_{(i,j)\in\mathcal{B}(\mu):i<j} A_{ij} B_{\mu(i)\mu(j)} \overset{\text{st}}{\preceq} X_{\mathcal{B}}.$$

The analysis of the summation over $\mathcal{C}(\mu)$ is more challenging since the terms in the summation are correlated. To handle the correlation, we split the summation into three parts, where within each part, all terms are independent. To this end, consider two unordered pairs $\{i, j\}$ and $\{\ell, m\}$ such that $\{i, j\} \neq \{\ell, m\}$. Note that the variables $A_{i,j} B_{\mu(i),\mu(j)}$ and $A_{\ell,m} B_{\mu(\ell),\mu(m)}$ are dependent if and only if

$$\{\mu(i), \mu(j)\} = \{\pi_*(\ell), \pi_*(m)\} \quad \text{or} \quad \{\mu(\ell), \mu(m)\} = \{\pi_*(i), \pi_*(j)\}. \tag{13}$$

We can now construct a graph $H$ with vertex set $\{\{i, j\} : (i, j) \in \mathcal{C}(\mu)\}$ which captures dependencies between terms of the form $A_{i,j} B_{\mu(i),\mu(j)}$, so there is an edge between $\{i, j\}$ and $\{\ell, m\}$ if and only if (13) holds. Note that each unordered pair $\{i, j\}$ may correspond to one or both of the ordered pairs $(i, j)$ and $(j, i)$. By construction, $A_{i,j} B_{\mu(i),\mu(j)}$ and $A_{\ell,m} B_{\mu(\ell),\mu(m)}$ are independent if and only if there is no edge between $\{i, j\}$ and $\{\ell, m\}$ in $H$. Furthermore, if $S$ is a subset of the vertices of $H$ such that no two elements of $S$ are connected by an edge, then the collection of random variables $\{A_{i,j} B_{\mu(i),\mu(j)} : \{i, j\} \in S\}$ are mutually independent.

Note further that any unordered pair $\{i, j\}$ has at most two neighbors in $H$, so $H$ is 3-colorable. Let $\{\mathcal{C}_1, \mathcal{C}_2, \mathcal{C}_3\}$ be the partition of $\{\{i, j\} : (i, j) \in \mathcal{C}(\mu)\}$ induced by the coloring. Since there are no edges between any two elements of $\mathcal{C}_1$, we can express the sum corresponding to $\mathcal{C}_1$ as a sum of two independent binomial random variables:

$$X_{\mathcal{C}_1} := \sum_{\{i,j\}\in\mathcal{C}_1} A_{i,j} B_{\mu(i),\mu(j)} \overset{d}{=} \text{Bin}\left(|\mathcal{C}_1 \cap \mathcal{E}^+(\boldsymbol{\sigma}_*)|, (sp)^2\right) + \text{Bin}\left(|\mathcal{C}_1 \cap \mathcal{E}^-(\boldsymbol{\sigma}_*)|, (sq)^2\right)$$

$$\overset{\text{st}}{\preceq} \text{Bin}\left(|\mathcal{C}_1|, (s\gamma)^2\right) \overset{\text{st}}{\preceq} \text{Bin}\left(|\mathcal{C}(\mu)|, (s\gamma)^2\right).$$

Similarly, we have that

$$X_{\mathcal{C}_2} := \sum_{\{i,j\}\in\mathcal{C}_2} A_{i,j} B_{\mu(i),\mu(j)} \overset{\text{st}}{\preceq} \text{Bin}\left(|\mathcal{C}(\mu)|, (s\gamma)^2\right),$$

$$X_{\mathcal{C}_3} := \sum_{\{i,j\}\in\mathcal{C}_3} A_{i,j} B_{\mu(i),\mu(j)} \overset{\text{st}}{\preceq} \text{Bin}\left(|\mathcal{C}(\mu)|, (s\gamma)^2\right).$$

Accounting for the fact that each vertex $\{i, j\}$ in $H$ may correspond to one or two ordered pairs in $\mathcal{C}(\mu)$, we obtain

$$\sum_{(i,j)\in\mathcal{C}(\mu)} A_{ij} B_{\mu(i)\mu(j)} \overset{\text{st}}{\preceq} 2\left(X_{\mathcal{C}_1} + X_{\mathcal{C}_2} + X_{\mathcal{C}_3}\right),$$

Putting everything together, using a union bound we can write

$$\mathbb{P}(f(\mu) \geq kd) = \mathbb{P}\left(2\left(X_{\mathcal{B}} + X_{\mathcal{C}_1} + X_{\mathcal{C}_2} + X_{\mathcal{C}_3}\right) \geq kd\right) \leq \sum_{i=1}^{3} \mathbb{P}\left(2X_{\mathcal{B}} + 6X_{\mathcal{C}_i} \geq kd\right). \quad (14)$$

The terms in the summation can be handled with a Chernoff bound. For $i \in \{1,2,3\}$ and any $\theta > 0$,

$$\mathbb{P}\left(2X_{\mathcal{B}} + 6X_{\mathcal{C}_i} \geq kd\right) = \mathbb{P}\left(e^{\theta\left(2X_{\mathcal{B}}+6X_{\mathcal{C}_i}\right)} \geq e^{\theta kd}\right)$$

$$\leq e^{-\theta kd}\mathbb{E}\left[e^{\theta\left(2X_{\mathcal{B}}+6X_{\mathcal{C}_i}\right)}\right]$$

$$= e^{-\theta kd}\mathbb{E}\left[e^{2\theta X_{\mathcal{B}}}\right]\mathbb{E}\left[e^{6\theta X_{\mathcal{C}_i}}\right].$$

If $Z \sim \text{Bin}(m,p)$, then $\mathbb{E}\left[e^{tZ}\right] = \left(1 + p(e^t - 1)\right)^m \leq \exp\left(mp(e^t - 1)\right)$. Using this fact, along with the bounds $|\mathcal{B}(\mu)| \leq d$ and $|\mathcal{C}(\mu)| \leq |\mathcal{A}(\mu)| \leq dn$, we obtain that

$$\mathbb{P}\left(2X_{\mathcal{B}} + 6X_{\mathcal{C}_i} \geq kd\right) \leq \exp\left(-\theta kd + \frac{1}{2}|\mathcal{B}(\mu)|s^2\gamma\left(e^{2\theta} - 1\right) + |\mathcal{C}(\mu)|(s\gamma)^2\left(e^{6\theta} - 1\right)\right)$$

$$\leq \exp\left[d\left(-\theta k + \frac{1}{2}s^2\gamma\left(e^{2\theta} - 1\right) + n(s\gamma)^2\left(e^{6\theta} - 1\right)\right)\right].$$

Substituting into (14), we obtain the desired result. ∎

**Proof** [Proof of Lemma 21] Fix $d \in [n]$ and a matching $(M, \mu) \in \mathcal{M}(d)$. By Lemma 23, we have that

$$\mathbb{P}\left(f(\mu) \geq 13d\right) \leq 3\exp\left[d\left(-13\theta + \frac{1}{2}s^2\gamma\left(e^{2\theta} - 1\right) + n(s\gamma)^2\left(e^{6\theta} - 1\right)\right)\right],$$

where $\gamma = \max\{\alpha, \beta\}\frac{\log n}{n}$. Setting $\theta = c\log n$ for a constant $c \in \left(\frac{2}{13}, \frac{1}{6}\right)$, we thus have that

$$\mathbb{P}\left(f(\mu) \geq 13d\right) \leq 3\exp\left[d\left(-13c\log n + o(1)\right)\right].$$

Therefore, $\xi \leq 3\exp\left(-13c\log n + o(1)\right) \leq 4n^{-13c}$ for $n$ sufficiently large. Since $13c > 2$, this implies that $n^2\xi = o(1)$. The proof is complete by Corollary 20. ∎

### E.3. The size of the $k$-core of $G_1 \wedge_{\pi_*} G_2$

We follow the method of Łuczak (1991) in upper bounding $|F|$, the number of vertices outside of the $k$-core of an SBM. In particular, we refine the results of Łuczak (1991) for random graphs in the logarithmic degree regime. In the remainder of this section, unless otherwise stated, we assume that $G \sim \text{SBM}(n, \alpha\log(n)/n, \beta\log(n)/n)$, where $\alpha, \beta > 0$ are treated as generic parameters. The following result shows that it is unlikely that there exist well-connected small subgraphs of $G$.

**Lemma 24** *Let $a > 1$ and $\epsilon > 0$ be fixed. Then for $n$ sufficiently large,*

$$\mathbb{P}\left(\text{there exists } \mathcal{S} \subset [n] \text{ such that } |\mathcal{S}| \leq n^{1-\epsilon} \text{ and } G\{\mathcal{S}\} \text{ has at least } a|\mathcal{S}| \text{ edges}\right) \leq \frac{2}{\log n}.$$

**Proof** Let $\mathcal{S}$ be an $m$-vertex subset of $[n]$, and let $X_{\mathcal{S}}$ be the indicator variable that is 1 if the subgraph induced by $\mathcal{S}$ has at least $am$ edges. Set $\gamma := \max\{\alpha, \beta\}$. We can bound

$$\mathbb{E}[X_{\mathcal{S}}] \leq \binom{\binom{m}{2}}{am} \left(\gamma \frac{\log n}{n}\right)^{am} \leq \binom{m^2}{am} \left(\gamma \frac{\log n}{n}\right)^{am} \leq \left(\frac{m}{n} \cdot \frac{\gamma e \log n}{a}\right)^{am}.$$

Summing over all possible $m$-vertex subsets, we have that

$$\mathbb{E}\left[\sum_{\mathcal{S} \subset [n]:|\mathcal{S}|=m} X_{\mathcal{S}}\right] \leq \binom{n}{m}\left(\frac{m}{n} \cdot \frac{\gamma e \log n}{a}\right)^{am} \leq \left(\left(\frac{\gamma e^{1+1/a} \log n}{a}\right)^{a} \left(\frac{m}{n}\right)^{a-1}\right)^{m}.$$

Next, note that if $m \leq n^{1-\epsilon}$ then for $n$ sufficiently large,

$$\left(\frac{\gamma e^{1+1/a} \log n}{a}\right)^{a} \left(\frac{m}{n}\right)^{a-1} \leq \frac{1}{\log n}.$$

Hence, for $n$ sufficiently large we have that

$$\mathbb{E}\left[\sum_{\mathcal{S} \subset [n]:|\mathcal{S}|\leq n^{1-\epsilon}} X_{\mathcal{S}}\right] \leq \sum_{m=1}^{n^{1-\epsilon}} \left(\frac{1}{\log n}\right)^{m} \leq \frac{1}{\log n}\left(1 - \frac{1}{\log n}\right)^{-1} \leq \frac{2}{\log n}.$$

Finally, Markov's inequality implies the claim. ∎

Our next result uses Lemma 24 to study Algorithm 4.

**Lemma 25 (Łuczak expansion)** *Let $G \sim \mathrm{SBM}(n, \alpha \log(n)/n, \beta \log(n)/n)$ for fixed $\alpha, \beta > 0$, and fix $c \in (0, 1)$. Then, for $n$ sufficiently large, the following holds with probability $1 - o(1)$. For every $U \subseteq [n]$ such that $|U| \leq n^{c}$, denoting by $\overline{U}$ the output of Algorithm 4 on input $(G, U)$, we have that $|\overline{U}| \leq 3|U|$ and for all $v \in [n] \setminus \overline{U}$, $v$ has at most one neighbor in $\overline{U}$ with respect to $G$.*

A version of Lemma 25 specialized to the analysis of $k$-core sizes was previously proven in Łuczak (1991). Since the proof is short and the Łuczak expansion appears as a subroutine in our community recovery algorithm, we provide the full proof of Lemma 25 here.

**Proof** [Proof of Lemma 25] Let $U_m = \overline{U}$ be the final set produced by Algorithm 4 on input $(G, U)$. From the construction, it is clear that $U_m \supseteq U_0$. Moreover, if $v \in [n] \setminus U_m$, then $v$ cannot have more than one neighbor in $U_m$, else we could construct another set $U_{m+1}$, contradicting the maximality of $U_m$. This holds for any $U \subseteq [n]$.

Now let $\epsilon := (1 - c)/2$ and let $n$ be sufficiently large such that $3n^c \leq n^{1-\epsilon}$. Let $\mathcal{H}$ denote the event described in Lemma 24 with this $\epsilon$ and $a = 4/3$. Since $\mathbb{P}(\mathcal{H}) = 1 - o(1)$, to prove the claim it suffices to show that on the event $\mathcal{H}$ we have that $|\overline{U}| \leq 3|U|$ for every $U \subseteq [n]$ such that $|U| \leq n^c$.

Suppose that for some $U \subseteq [n]$ such that $|U| \leq n^c$ we have that $|U_m| > 3|U|$. Then, since exactly one vertex is added in each step of the construction, there must exist $1 \leq \ell \leq m$ such that $|U_\ell| = 3|U|$. Let $E_\ell$ denote the number of edges in $G\{U_\ell\}$. Since each vertex added in the construction introduces at least two new edges, we have that

$$E_\ell \geq 2\ell = 2(|U_\ell| - |U|) = \frac{4}{3}|U_\ell|.$$

However, on the event $\mathcal{H}$ this is not possible, since $|U_\ell| \le 3n^c \le n^{1-\epsilon}$. ∎

Let $F$ be the set of vertices outside the $k$-core of $G$. A straightforward application of Lemma 25 allows us to bound $|F|$, as was done in Łuczak (1991).

**Lemma 26** *Let $G \sim \mathrm{SBM}(n, \alpha \log(n)/n, \beta \log(n)/n)$ for fixed $\alpha, \beta > 0$. Fix $k \ge 1$. With probability $1 - o(1)$, we have that $|F| \le n^{1-(\alpha+\beta)/2+o(1)}$.*

**Proof** Define $U$ to be the set of vertices with degree at most $k$ in $G$, and let $\overline{U} \supseteq U$ be the set produced by Algorithm 4. We claim that $G\{[n] \setminus \overline{U}\}$ has minimum degree at least $k$. To see why, note that if $v \in [n] \setminus \overline{U}$, then $v \notin U$, which implies that $v$ has degree at least $k+1$ in $G$. However $v$ can have at most one neighbor in $\overline{U}$ by construction, so $v$ must have at least $k$ neighbors in $[n] \setminus \overline{U}$, and the claim follows.

Since the $k$-core is the vertex set corresponding to the *largest* induced subgraph with minimum degree at least $k$, we have the bound $|F| \le |\overline{U}|$. Furthermore, by Lemma 25 we have, with high probability, that $|\overline{U}| \le 3|U|$.

It thus remains to bound $|U|$. Define the event $\mathcal{H} := \{n/2 - n^{3/4} \le |V^+|, |V^-| \le n/2 + n^{3/4}\}$. Bounding the expectation of $|U|\mathbf{1}(\mathcal{H})$ and letting $\gamma := \max\{\alpha, \beta\}$, we have

$$
\mathbb{E}[|U|\mathbf{1}(\mathcal{H})]
$$
$$
\le n \sum_{i=0}^{k} \sum_{j=0}^{i} \binom{(1+o(1))\frac{n}{2}}{j} \binom{(1+o(1))\frac{n}{2}}{i-j} p^j (1-p)^{(1-o(1))\frac{n}{2}-j} q^{i-j}(1-q)^{(1-o(1))\frac{n}{2}-i+j}
$$
$$
\le 2n \left(1 - \frac{\alpha \log n}{n}\right)^{(1-o(1))\frac{n}{2}} \left(1 - \frac{\beta \log n}{n}\right)^{(1-o(1))\frac{n}{2}} \sum_{i=0}^{k} \sum_{j=0}^{i} \left(\frac{(1+o(1))n}{2}\right)^i \left(\frac{\gamma \log n}{n}\right)^i
$$
$$
\le 2n^{1-(\alpha+\beta)/2+o(1)} \sum_{i=0}^{k} (i+1) \left(\frac{(1+o(1))\gamma \log n}{2}\right)^i
$$
$$
\le 2(k+1)^2 \left(\frac{\gamma \log n}{2}\right)^k n^{1-(\alpha+\beta)/2+o(1)}.
$$

Turning to the probability of interest, we obtain that

$$
\mathbb{P}\left(|U| > (\log n)\mathbb{E}[|U|\mathbf{1}(\mathcal{H})]\right) \le \mathbb{P}\left(|U|\mathbf{1}(\mathcal{H}) > (\log n)\mathbb{E}[|U|\mathbf{1}(\mathcal{H})]\right) + \mathbb{P}(\mathcal{H}^c)
$$
$$
\le \frac{1}{\log n} + o(1) = o(1),
$$

where we have used Markov's inequality to bound the first term on the right hand side in the first line, and Lemma 13 to bound $\mathbb{P}(\mathcal{H}^c)$. Putting everything together, we have shown that with probability $1 - o(1)$, we have that $|F| \le 3|U| \le n^{1-(\alpha+\beta)/2+o(1)}$. ∎

## Appendix F. Labeling the $k$-core matching (Algorithm 2)

### F.1. Almost exact recovery in a single SBM

For the purposes of this section, we assume that $G \sim \mathrm{SBM}(n, \alpha \log(n)/n, \beta \log(n)/n)$, where $\alpha, \beta > 0$ are generic parameters and $G$ has communities of size $(1 + o(1))n/2$. It is well known

that in the logarithmic degree regime, almost exact recovery of communities is always possible from a single SBM (see, e.g., Abbe (2017)). In our case, it is of further interest to precisely control the size of the error set—in particular, we require the error set to be as small as possible in an information-theoretic sense—as well as the *geometry* of the error set, since it will be subsequently used in later parts of Algorithm 2. To properly control the size and geometry of the error set, we leverage the community recovery algorithm of Mossel et al. (2016) (Algorithm 3). The purpose of the refinements in Algorithm 3 is to characterize the geometry of the error set, which is what the main result of this subsection describes. Before stating the result, we define some useful notation. Let $\gamma = \max\{\alpha, \beta\}$, and define, for a vertex $v$ in $G$, the quantity

$$\mathsf{maj}_G(v) := |\mathcal{N}(v) \cap V^{\sigma_*(v)}| - |\mathcal{N}(v) \cap V^{-\sigma_*(v)}|.$$

If $\alpha > \beta$, define the set

$$I_\epsilon(G) := \{v \in [n] : \mathsf{maj}_G(v) \leq \epsilon \log n \text{ or } |\mathcal{N}(v)| \geq 100 \max\{1, \gamma\} \log n\}. \tag{15}$$

If $\alpha < \beta$, let

$$I_\epsilon(G) := \{v \in [n] : -\mathsf{maj}_G(v) \leq \epsilon \log n \text{ or } |\mathcal{N}(v)| \geq 100 \max\{1, \gamma\} \log n\}. \tag{16}$$

If $\alpha > \beta$, we say that a vertex $v$ has a *$k$-majority* if $\mathsf{maj}_G(v) \geq k$. If $\alpha < \beta$, we say that a vertex $v$ has a $k$-majority if $-\mathsf{maj}_G(v) \geq k$.

**Lemma 27** *Let $\epsilon > 0$. With probability $1 - o(1)$, Algorithm 3 on input $(G, \alpha, \beta, \epsilon)$ correctly classifies all vertices in $[n] \setminus I_\epsilon(G)$.*

**Proof** Our proof is adapted from (Mossel et al., 2016, Proposition 4.3), with minor changes due to the more general parameter regime we consider. We will assume that $\alpha > \beta$; the case $\alpha < \beta$ is similar, and full details can be found in Mossel et al. (2016).

Fix $i \in [m]$ and $v \in (U_i \cap V^+) \setminus I_\epsilon$, and define $k_+ := |\mathcal{N}(v) \cap V^+|$ and $k_- := |\mathcal{N}(v) \cap V^-|$ to be the number of same-community and different-community neighbors of $v$, respectively. We also define $k_{+,i} := |\mathcal{N}(v) \cap V^+ \cap U_i|$ and $k_{-,i} := |\mathcal{N}(v) \cap V^- \cap U_i|$ to be the number of same-community and different-community neighbors of $v$ in $U_i$, respectively. Finally, we define $k_{+,\neg i} := |(\mathcal{N}(v) \cap V^+) \setminus U_i| = k_+ - k_{+,i}$ and $k_{-,\neg i} := |(\mathcal{N}(v) \cap V^-) \setminus U_i| = k_- - k_{-,i}$ to be the number of same-community and different-community neighbors of $v$ in $[n] \setminus U_i$, respectively. Note in particular that conditioned on $k_+$ and $k_-$, we have that $k_{+,i} \sim \mathrm{Bin}(k_+, 1/m)$ and $k_{-,i} \sim \mathrm{Bin}(k_-, 1/m)$.

We first show that $k_{+,\neg i} - k_{-,\neg i} \geq \frac{\epsilon}{2} \log n$ with high probability. To this end, we have the lower bound

$$k_{+,\neg i} - k_{-,\neg i} = k_+ - k_- - (k_{+,i} - k_{-,i}) \geq \epsilon \log n - (k_{+,i} + k_{-,i}),$$

where the final inequality uses $k_+ - k_- \geq \epsilon \log n$ for $v \notin I_\epsilon$. It therefore suffices to show that $k_{+,i} + k_{-,i} \leq \frac{\epsilon}{2} \log n$. Noticing that $k_{+,i} + k_{-,i} \sim \mathrm{Bin}(|\mathcal{N}(v)|, 1/m)$, which is stochastically dominated by $\mathrm{Bin}(100 \max\{1, \gamma\} \log n, 1/m)$ for $v \notin I_\epsilon$, a Chernoff bound implies that

$$\mathbb{P}\left(k_{+,i} + k_{-,i} \geq \frac{\epsilon}{2} \log n\right) \leq \inf_{\theta > 0} \left\{ e^{-\theta \frac{\epsilon}{2} \log n} \left(1 + \frac{1}{m}(e^\theta - 1)\right)^{100 \max\{1, \gamma\} \log n} \right\}$$

$$\leq \exp\left(-\sup_{\theta > 0} \left\{\frac{\theta \epsilon}{2} - \frac{100 \max\{1, \gamma\}}{m}(e^\theta - 1)\right\} \log n\right)$$

$$\leq \exp\left(-\left(\log\left(\frac{\epsilon m}{200 \max\{1, \gamma\}}\right) - 1\right) \frac{\epsilon}{2} \log n\right) = o(n^{-1}).$$

In the display above, the first inequality on the third line uses $\theta = \log(\epsilon m/(200 \max\{1, \gamma\}))$ and the final equality follows from our choice of $m$ in Algorithm 3. It therefore holds that $k_{+,\neg i} - k_{-,\neg i} \geq \frac{\epsilon}{2} \log n$ for all $v \notin I_\epsilon$ with probability $1 - o(1)$.

From this point, the proof of the lemma is identical to the one provided in Mossel et al. (2016); we state a brief overview of the proof here and defer the interested reader to (Mossel et al., 2016, Proposition 4.3) for the details. Consider the random variables $X^+ := |\mathcal{N}(v) \cap U_{i,+} \cap V^-|$ and $X^- := |\mathcal{N}(v) \cap U_{i,-} \cap V^+|$, which correspond to the number of neighbors of $v$ in $[n] \setminus U_i$ that are misclassified by the partition $(U_{i,+}, U_{i,-})$. We can write

$$|\mathcal{N}(v) \cap U_{i,+}| = k_{+,\neg i} - X^+ + X^-,$$
$$|\mathcal{N}(v) \cap U_{i,-}| = k_{-,\neg i} - X^- + X^+.$$

It follows that $v$ is correctly classified if $|X^+ - X^-| < |k_{+,\neg i} - k_{-,\neg i}|/2$. Moving forward, our goal is to show that this inequality holds with high probability.

Following Mossel et al. (2016), we let $E^- := |U_{i,-} \cap V^+|$ and $E^+ := |U_{i,+} \cap V^-|$ denote the total number of vertices of each type misclassified by the partition $(U_{i,+}, U_{i,-})$ that are potential neighbors of $v$. In particular, $E^+ = o(n)$ and $E^- = o(n)$. As this is a partition corresponding to the structure of $G\{[n] \setminus U_i\}$, the neighbors of $v$ in $[n] \setminus U_i$ are independent of the partition. Moreover, the vertices in $U_{i,-} \cap V^+$ are equally likely to be neighbors of $v$, with the same holding for vertices in $U_{i,+} \cap V^-$. We can therefore generate $X^-$ by randomly choosing $k_{+,\neg i}$ vertices of $V^+ \setminus U_i$ without replacement, with $X^-$ being equal to the number of sampled vertices in $U_{i,-} \cap V^+$. Conditioned on $E^-$, $E^+$, $k_{+,\neg i}$, and $k_{-,\neg i}$, we thus have the distributional representations

$$X^- \stackrel{d}{=} \mathrm{HyperGeom}(|V^+ \setminus U_i|, k_{+,\neg i}, E^-) \quad \text{and} \quad X^+ \stackrel{d}{=} \mathrm{HyperGeom}(|V^- \setminus U_i|, k_{-,\neg i}, E^+).$$

From here, one can utilize properties of Hypergeometric random variables to show that with high probability,

$$\left| \mathbb{E}\left[ X^- \,\middle|\, |V^+ \setminus U_i|, k_{+,\neg i}, E^- \right] - \mathbb{E}\left[ X^+ \,\middle|\, |V^- \setminus U_i|, k_{-,\neg i}, E^+ \right] \right| = o(1)|k_{+,\neg i} - k_{-,\neg i}|.$$

Finally, the lemma follows from showing that $X^+$ and $X^-$ concentrate around their means and by taking a union bound over all vertices. An identical analysis holds for the case where $v \in (U_i \cap V^-) \setminus I_\epsilon(G)$. ∎

The next few results further characterize the set $I_\epsilon(G)$.

**Lemma 28** *Denote $\gamma := \max\{\alpha, \beta\}$. Then for every $\boldsymbol{\sigma}_*$ we have that*

$$\mathbb{P}\left(\forall i \in [n], |\mathcal{N}(i)| \leq 100 \max\{1, \gamma\} \log(n) \,\middle|\, \boldsymbol{\sigma}_*\right) \geq 1 - n^{-99}.$$

**Proof** Fix $i \in [n]$, and let $X \sim \mathrm{Bin}\left(n, \gamma \log(n)/n\right)$. Then for every $\boldsymbol{\sigma}_*$ we have that

$$\mathbb{P}\left(|\mathcal{N}(i)| \geq 100 \max\{1, \gamma\} \log n \,\middle|\, \boldsymbol{\sigma}_*\right) \leq \mathbb{P}(X \geq 100 \max\{1, \gamma\} \log n)$$
$$\leq \exp\left(-\frac{(100 \max\{1, \gamma\} \log n)^2}{2\gamma \log n + \frac{200}{3} \max\{1, \gamma\} \log n}\right)$$
$$\leq \exp\left(-100 \max\{1, \gamma\} \log n\right) \leq n^{-100}.$$

The second line uses Bernstein's inequality and the third uses $2\gamma + \frac{200}{3} \max\{1, \gamma\} \leq 100 \max\{1, \gamma\}$. We conclude by taking a union bound. ∎

**Lemma 29** *Suppose that* $\mathrm{D}_+(\alpha, \beta) < 99$. *Then for every* $\epsilon > 0$ *we have that*

$$\mathbb{E}[|I_\epsilon(G)|] \le 3n^{1-\mathrm{D}_+(\alpha,\beta)+\epsilon|\log(\alpha/\beta)|}.$$

**Proof** Assume that $\alpha > \beta$. Let $\mathcal{G}$ be the event that $n/2 - n^{3/4} \le |V^+|, |V^-| \le n/2 + n^{3/4}$ and let $\mathcal{H}$ be the event defined in Lemma 28. For any $i \in [n]$, we have that

$$\mathbb{P}\left(\{i \in I_\epsilon(G)\} \cap \mathcal{H} \,|\, \boldsymbol{\sigma}_*\right) \le \mathbb{P}\left(\mathsf{maj}_G(i) \le \epsilon \log n \,|\, \boldsymbol{\sigma}_*\right).$$

Notice that $\mathsf{maj}_G(i) \overset{d}{=} Y - Z$, where $Y, Z$ are independent with

$$Y \sim \mathrm{Bin}\left(|V^{\sigma_*(i)}| - 1, \alpha\frac{\log n}{n}\right) \qquad \text{and} \qquad Z \sim \mathrm{Bin}\left(|V^{-\sigma_*(i)}|, \beta\frac{\log n}{n}\right).$$

Since $|V^+| = (1 - o(1))n/2$ and $|V^-| = (1 - o(1))n/2$ on the event $\mathcal{G}$, Lemma 8 implies that

$$\mathbb{P}\left(\mathsf{maj}_G(i) \le \epsilon \log n \,|\, \boldsymbol{\sigma}_*\right) \mathbf{1}(\mathcal{G}) \le n^{-\mathrm{D}_+(\alpha,\beta)+\frac{\epsilon \log(\alpha/\beta)}{2}+o(1)} \le n^{-\mathrm{D}_+(\alpha,\beta)+\epsilon \log(\alpha/\beta)},$$

where the final inequality holds for $n$ sufficiently large. Putting everything together, we can bound the probability that $i \in I_\epsilon(G)$ as

$$\begin{aligned}
\mathbb{P}(i \in I_\epsilon(G)) &\le \mathbb{P}(\{i \in I_\epsilon(G)\} \cap \mathcal{G} \cap \mathcal{H}) + \mathbb{P}(\mathcal{G}^c) + \mathbb{P}(\mathcal{H}^c) \\
&\le \mathbb{E}\left[\mathbb{P}\left(\{i \in I_\epsilon(G)\} \cap \mathcal{H} \,|\, \boldsymbol{\sigma}_*\right) \mathbf{1}(\mathcal{G})\right] + 2n^{-99} \\
&\le n^{-\mathrm{D}_+(\alpha,\beta)+\epsilon \log(\alpha/\beta)} + 2n^{-99} \le 3n^{-\mathrm{D}_+(\alpha,\beta)+\epsilon \log(\alpha/\beta)},
\end{aligned}$$

where the inequality on the second line uses Lemmas 13 and 28 and the final inequality follows from $\mathrm{D}_+(\alpha, \beta) - \epsilon \log(\alpha/\beta) < 99$ under our assumption that $\mathrm{D}_+(\alpha, \beta) < 99$. Finally, to bound the expectation of $|I_\epsilon(G)|$, we can write

$$\mathbb{E}[|I_\epsilon(G)|] = \sum_{i \in [n]} \mathbb{P}(i \in I_\epsilon(G)) \le 3n^{1-\mathrm{D}_+(\alpha,\beta)+\epsilon \log(\alpha/\beta)}.$$

The case $\alpha < \beta$ follows from identical arguments. ∎

**Lemma 30** *If* $0 < \epsilon < \frac{\mathrm{D}_+(\alpha,\beta)}{2|\log(\alpha/\beta)|}$, *then*

$$\mathbb{P}\left(\forall i \in [n], |\mathcal{N}(i) \cap I_\epsilon(G)| \le 2\left\lceil \mathrm{D}_+(\alpha, \beta)^{-1} \right\rceil \,\Big|\, \boldsymbol{\sigma}_*\right) = 1 - o(1).$$

**Proof** Let us assume $\alpha > \beta$, let $\mathcal{G}$ be the event where $n/2 - n^{3/4} \le |V^+|, |V^-| \le n/2 + n^{3/4}$, and let $\mathcal{H}$ be the event defined in Lemma 28. Fix $S \subset [n]$, and assume that $|S|$ is of constant size with respect to $n$. In the calculations below, we will use the representation

$$\mathsf{maj}_G(j) = \sigma_*(j) \sum_{k \in \mathcal{N}(j)} \sigma_*(k).$$

35

We can then write

$$\mathbb{P}\left(\{S \subset \mathcal{N}(i) \cap I_\epsilon(G)\} \cap \mathcal{H} \,|\, \boldsymbol{\sigma}_*\right)$$

$$\leq \mathbb{P}\left(\forall j \in S, i \sim j \text{ and } \sigma_*(j) \sum_{k \in \mathcal{N}(j)} \sigma_*(k) \leq \epsilon \log n \,\middle|\, \boldsymbol{\sigma}_*\right)$$

$$\leq \mathbb{P}\left(\forall j \in S, i \sim j \text{ and } \sigma_*(j) \sum_{k \in \mathcal{N}(j) \setminus (S \cup \{i\})} \sigma_*(k) \leq 2\epsilon \log n \,\middle|\, \boldsymbol{\sigma}_*\right),$$

where the inequality on the second line is due to $|S \cup \{i\}| \leq \epsilon \log n$ for $n$ sufficiently large. To simplify the right hand side, let us introduce the events

$$E_j := \left\{ i \sim j \text{ and } \sigma_*(j) \sum_{k \in \mathcal{N}(j) \setminus (S \cup \{i\})} \sigma_*(k) \leq 2\epsilon \log n \right\}, \qquad j \in S.$$

Notice in particular that conditioned on $\boldsymbol{\sigma}_*$, the event $E_j$ depends on the neighbors of $j$ outside of the set $S$, hence the $E_j$'s are conditionally independent with respect to $\boldsymbol{\sigma}_*$. We can therefore bound

$$\mathbb{P}\left(\{S \subset \mathcal{N}(i) \cap I_\epsilon(G)\} \cap \mathcal{H} \,|\, \boldsymbol{\sigma}_*\right) \leq \prod_{j \in S} \mathbb{P}\left(E_j \,|\, \boldsymbol{\sigma}_*\right). \tag{17}$$

To bound $\mathbb{P}\left(E_j \,|\, \boldsymbol{\sigma}_*\right)$, notice that the events $\{i \sim j\}$ and $\{\sigma_*(j) \sum_{k \in \mathcal{N}(j) \setminus (S \cup \{i\})} \sigma_*(k) \leq 2\epsilon \log n\}$ are independent conditioned on $\boldsymbol{\sigma}_*$. Furthermore, on $\mathcal{G}$ we have the distributional representation

$$\sigma_*(j) \sum_{k \in \mathcal{N}(j) \setminus (S \cup \{i\})} \sigma_*(k) \stackrel{d}{=} Y - Z,$$

where, since $|S| = o(n)$, $Y$ and $Z$ are independent with

$$Y \sim \mathrm{Bin}\left((1 - o(1))\frac{n}{2} - 1, \alpha\frac{\log n}{n}\right), \qquad Z \sim \mathrm{Bin}\left((1 - o(1))\frac{n}{2}, \beta\frac{\log n}{n}\right).$$

Using $\mathbb{P}\left(i \sim j \,|\, \boldsymbol{\sigma}_*\right) \leq \gamma \log(n)/n$, as well as Lemma 8, shows that

$$\mathbb{P}\left(E_j \,|\, \boldsymbol{\sigma}_*\right)\mathbf{1}(\mathcal{G}) \leq \frac{\gamma \log n}{n} \cdot n^{-\mathrm{D}_+(\alpha,\beta)+\epsilon\log(\alpha/\beta)+o(1)} = n^{-1-\mathrm{D}_+(\alpha,\beta)+\epsilon\log(\alpha/\beta)+o(1)}.$$

Combining the bound in the display above with (17) shows that

$$\mathbb{P}\left(\{S \subset \mathcal{N}(i) \cap I_\epsilon(G)\} \cap \mathcal{H} \,|\, \boldsymbol{\sigma}_*\right)\mathbf{1}(\mathcal{G}) \leq n^{-|S|(1+\mathrm{D}_+(\alpha,\beta)-\epsilon\log(\alpha/\beta)-o(1))}.$$

We may now take a union bound over subsets $S$ of size $m$ to obtain

$$\mathbb{P}\left(\{|\mathcal{N}(i) \cap I_\epsilon(G)| \geq m\} \cap \mathcal{H} \,|\, \boldsymbol{\sigma}_*\right)\mathbf{1}(\mathcal{G}) \leq \sum_{S \subset [n]:|S|=m} \mathbb{P}\left(\{S \subset \mathcal{N}(i) \cap I_\epsilon(G)\} \cap \mathcal{H} \,|\, \boldsymbol{\sigma}_*\right)\mathbf{1}(\mathcal{G})$$

$$\leq n^{-m(\mathrm{D}_+(\alpha,\beta)-\epsilon\log(\alpha/\beta)-o(1))}.$$

The final inequality uses the fact that the number of $m$-element subsets of $[n]$ is at most $n^m$. Taking a union bound over $i \in [n]$ now shows that

$$\mathbb{P}\left(\{\exists i \in [n] : |\mathcal{N}(i) \cap I_\epsilon(G)| \geq m\} \cap \mathcal{H} \mid \boldsymbol{\sigma}_*\right) \mathbf{1}(\mathcal{G}) \leq n^{1-m(\mathrm{D}_+(\alpha,\beta)-\epsilon\log(\alpha/\beta)-o(1))}$$
$$\leq n^{1-m\mathrm{D}_+(\alpha,\beta)/2+o(1)}, \tag{18}$$

where the final inequality uses the assumption that $\epsilon \leq \frac{\mathrm{D}_+(\alpha,\beta)}{2\log(\alpha/\beta)}$. Moreover, the right hand side is $o(1)$ provided $m > 2\mathrm{D}_+(\alpha,\beta)^{-1}$, so we may set $m := \lceil 2\mathrm{D}_+(\alpha,\beta)^{-1}\rceil + 1$ for all of the arguments to hold. The desired result follows from the bound

$$\mathbb{P}(\exists i \in [n] : |\mathcal{N}(i) \cap I_\epsilon(G)| \geq m)$$
$$\leq \mathbb{P}(\{\exists i \in [n] : |\mathcal{N}(i) \cap I_\epsilon(G)| \geq m\} \cap \mathcal{G} \cap \mathcal{H}) + \mathbb{P}(\mathcal{G}^c) + \mathbb{P}(\mathcal{H}^c)$$
$$\leq \mathbb{E}\left[\mathbb{P}\left(\{\exists i \in [n] : |\mathcal{N}(i) \cap I_\epsilon(G)| \geq m\} \cap \mathcal{H} \mid \boldsymbol{\sigma}_*\right)\mathbf{1}(\mathcal{G})\right] + o(1)$$
$$= o(1),$$

where the inequality on the first line is due to a union bound, the inequality on the second line uses the tower rule as well as Lemmas 28 and 13, and the final inequality is due to (18) as well as our assumption on $\epsilon$. The case $\alpha < \beta$ follows from identical arguments. ∎

### F.2. From almost exact to exact recovery in $[n] \setminus \overline{F}$

In this subsection, our main result is the following.

**Lemma 31** *Suppose that*

$$\left(1 - (1-s)^2\right)\mathrm{D}_+(\alpha,\beta) > 1 + 2\epsilon|\log(\alpha/\beta)| \qquad \text{and} \qquad 0 < \epsilon \leq \frac{s\mathrm{D}_+(\alpha,\beta)}{4|\log(\alpha/\beta)|}.$$

*Then, with probability $1 - o(1)$, Algorithm 2 run on the input $(G_1, G_2, (\widehat{M}, \widehat{\mu}), \alpha, \beta, s, \epsilon)$ correctly labels all vertices in $[n] \setminus \overline{F}$.*

The proof of Lemma 31 follows from several intermediate steps. First, we show in Lemmas 32 and 33 that all vertices in $(G_1 \vee_{\widehat{\mu}} G_2)\{[n] \setminus \overline{F}\}$ have an $\epsilon \log n$ majority that is aligned with the ground-truth community label.

**Lemma 32** *If $\left(1 - (1-s)^2\right)\mathrm{D}_+(\alpha,\beta) > 1 + \epsilon|\log(\alpha/\beta)|$, then with probability $1 - o(1)$, we have for all $i \in [n]$ that $\mathsf{maj}_{G_1 \vee_{\pi_*} G_2}(i) \geq \epsilon \log n$.*

**Proof** The proof follows directly from Lemma 8 and a union bound. ∎

**Lemma 33** *Suppose that $(1 - (1-s)^2)\mathrm{D}_+(\alpha,\beta) > 1 + 2\epsilon|\log(\alpha/\beta)|$. Then with probability $1 - o(1)$, all vertices in $[n] \setminus \overline{F}$ have an $\epsilon \log n$ majority in $(G_1 \vee_{\widehat{\mu}} G_2)\{[n] \setminus \overline{F}\}$.*

**Proof** Recall that $F_* := [n] \setminus M_*$ is the set of vertices outside the 13-core of $G_1 \wedge_{\pi_*} G_2$, and that $\overline{F}_*$ is the Łuczak expansion of $F_*$ with respect to $G_2$. In light of Lemma 21 and Remark 22, it suffices to replace $\overline{F}$ with $\overline{F}_*$ and $G_1 \vee_{\widehat{\mu}} G_2$ with $(G_1 \vee_{\pi_*} G_2)\{[n] \setminus F_*\}$ in our analysis. For brevity, we denote $H := (G_1 \vee_{\pi_*} G_2)\{[n] \setminus \overline{F}_*\}$ in the remainder of this proof.

To begin, notice that we can lower bound the neighborhood majority of $i \in [n] \setminus \overline{F}_*$ in $H$ as

$$\mathsf{maj}_H(i) = \sigma_*(i) \sum_{j \in \mathcal{N}_H(i)} \sigma_*(j) \geq \sigma_*(i) \sum_{j \in \mathcal{N}_{G_1 \vee_{\pi_*} G_2}(i)} \sigma_*(j) - |\mathcal{N}_{G_1 \vee_{\pi_*} G_2}(i) \cap \overline{F}_*|$$

$$= \mathsf{maj}_{G_1 \vee_{\pi_*} G_2}(i) - |\mathcal{N}_{G_1 \setminus_{\pi_*} G_2}(i) \cap \overline{F}_*| - |\mathcal{N}_{G_2}(i) \cap \overline{F}_*|. \tag{19}$$

Through identical arguments, we also have the upper bound

$$\mathsf{maj}_H(i) \leq \mathsf{maj}_{G_1 \vee_{\pi_*} G_2}(i) + |\mathcal{N}_{G_1 \setminus_{\pi_*} G_2}(i) \cap \overline{F}_*| + |\mathcal{N}_{G_2}(i) \cap \overline{F}_*|. \tag{20}$$

Combining (19) and (20), we obtain

$$\left| \mathsf{maj}_H(i) - \mathsf{maj}_{G_1 \vee_{\pi_*} G_2}(i) \right| \leq |\mathcal{N}_{G_1 \setminus_{\pi_*} G_1}(i) \cap \overline{F}_*| + |\mathcal{N}_{G_2}(i) \cap \overline{F}_*|. \tag{21}$$

In particular, since $|\mathsf{maj}_{G_1 \vee_{\pi_*} G_2}(i)| \geq 2\epsilon \log n$ for all $i \in [n]$ with probability $1 - o(1)$ under the condition $(1 - (1-s)^2)\mathrm{D}_+(\alpha, \beta) > 1 + 2\epsilon|\log(\alpha/\beta)|$ by Lemma 32, it is ensured that $\mathsf{maj}_H(i) \geq \epsilon \log n$ provided the right hand side of (21) is at most $\epsilon \log n$ for all $i \in [n]$. We show that this is indeed the case by bounding the two terms on the right hand side of (21) separately.

We start by bounding $|\mathcal{N}_{G_1 \setminus_{\pi_*} G_2}(i) \cap \overline{F}_*|$. Conditioned on $\pi_*$, $\boldsymbol{\sigma}_*$, and $\boldsymbol{\mathcal{E}} := \{\mathcal{E}_{00}, \mathcal{E}_{01}, \mathcal{E}_{10}, \mathcal{E}_{11}\}$, the graph $G_1 \setminus_{\pi_*} G_2$ is independent of $\overline{F}_*$ by Lemma 10, since $\overline{F}_*$ depends only on $G_1 \wedge_{\pi_*} G_2$ and $G_2$. Thus we can stochastically dominate $|\mathcal{N}_{G_1 \setminus_{\pi_*} G_2}(i) \cap \overline{F}_*|$ by a Poisson random variable $X$ with mean $\lambda_n$ given by

$$\lambda_n := \gamma \frac{\log n}{n} \left| \{j \in \overline{F}_* : \{i, j\} \in \mathcal{E}_{10}\} \right| \leq \gamma \frac{\log n}{n} |\overline{F}_*|,$$

where $\gamma := \max\{\alpha, \beta\}$. Next, for a fixed $\delta > 0$, define the event

$$\mathcal{G} := \left\{ |\overline{F}_*| \leq n^{1 - s^2 \mathrm{T_c}(\alpha, \beta) + \delta} \right\},$$

and notice that on $\mathcal{G}$, $\lambda_n \leq n^{-s^2 \mathrm{T_c}(\alpha, \beta) + \delta + o(1)}$. Thus for any positive integer $m$ we have that

$$\mathbb{P}\left( \{|\mathcal{N}_{G_1 \setminus_{\pi_*} G_2}(i) \cap \overline{F}_*| \geq m\} \cap \mathcal{G} \right) \leq \mathbb{P}(\{X \geq m\} \cap \mathcal{G})$$

$$= \mathbb{E}\left[ \mathbb{P}\left( X \geq m \,\middle|\, |\overline{F}_*|, \boldsymbol{\mathcal{E}}, \boldsymbol{\sigma}_*, \pi_* \right) \mathbf{1}(\mathcal{G}) \right]$$

$$\leq \mathbb{E}\left[ \left( \inf_{\theta > 0} e^{-\theta m + \lambda_n(e^\theta - 1)} \right) \mathbf{1}(\mathcal{G}) \right]$$

$$\leq \mathbb{E}[e\lambda_n^m \mathbf{1}(\mathcal{G})] \leq n^{-m(s^2 \mathrm{T_c}(\alpha, \beta) - \delta - o(1))}.$$

Above, the equality on the second line is due to the tower rule and since $\mathcal{G}$ is measurable with respect to $\overline{F}_*$; the inequality on the third line is due to a Chernoff bound; the inequality on the fourth line follows from setting $\theta = \log(1/\lambda_n)$ (which is valid since $\lambda_n = o(1)$ if $\mathcal{G}$ holds); the final inequality uses the upper bound for $\lambda_n$ on $\mathcal{G}$. Taking a union bound over $i \in [n]$ shows that

$$\mathbb{P}\left( \{\exists i \in [n] : |\mathcal{N}_{G_1 \setminus_{\pi_*} G_2}(i) \cap \overline{F}_*| \geq m\} \cap \mathcal{G} \right) \leq n^{1 - m(s^2 \mathrm{T_c}(\alpha, \beta) - \delta - o(1))}.$$

In particular, for $\delta$ taken to be sufficiently small, the right hand side is $o(1)$ if $m > (s^2 T_c(\alpha, \beta))^{-1}$ (it suffices to set $m = \lceil (s^2 T_c(\alpha, \beta))^{-1} \rceil + 1$). Finally, since $\mathbb{P}(\mathcal{G}) = 1 - o(1)$ in light of Lemmas 26 and 25, we obtain

$$\mathbb{P}\left( \forall i \in [n], |\mathcal{N}_{G_1 \setminus_{\pi_*} G_2}(i) \cap \overline{F}_*| \le \lceil (s^2 T_c(\alpha, \beta))^{-1} \rceil \right) = 1 - o(1). \tag{22}$$

Turning to the second term on the right hand side of (21), notice that by the definition of the Łuczak expansion, every $i \in [n] \setminus \overline{F}_*$ can have at most one neighbor in $\overline{F}_*$ with respect to the graph $G_2$, hence $|\mathcal{N}_{G_2}(i) \cap \overline{F}_*| \le 1$. This observation coupled with (22) shows that with probability $1 - o(1)$, it holds for all $i \in [n] \setminus \overline{F}_*$ that $|\mathsf{maj}_H(i) - \mathsf{maj}_{G_1 \vee_{\pi_*} G_2}| \le \epsilon \log n$. ∎

Next, in Lemma 34, we show that with probability $1 - o(1)$, each vertex in $G_2 \setminus_{\pi_*} G_1$ (and therefore also $G_2 \setminus_{\widehat{\mu}} G_1$) has a small number of neighbors in $I_\epsilon(G_1)$.

**Lemma 34** *If* $0 < \epsilon \le \frac{s D_+(\alpha, \beta)}{4|\log(\alpha/\beta)|}$, *then*

$$\mathbb{P}\left( \forall i \in [n], |\mathcal{N}_{G_2 \setminus_{\pi_*} G_1}(i) \cap I_\epsilon(G_1)| \le 2 \lceil (s D_+(\alpha, \beta))^{-1} \rceil \right) = 1 - o(1).$$

**Proof** Since $I_\epsilon(G_1)$ depends on $G_1$ alone, it follows that $I_\epsilon(G_1)$ and $G_2 \setminus_{\pi_*} G_1$ are conditionally independent given $\pi_*$, $\boldsymbol{\sigma_*}$, and $\mathcal{E}$. Hence we can stochastically dominate $|\mathcal{N}_{G_2 \setminus_{\pi_*} G_1}(i) \cap I_\epsilon(G_1)|$ by a Poisson random variable $X$ with mean $\lambda_n$ given by

$$\lambda_n := \gamma \frac{\log n}{n} |\{j \in I_\epsilon(G_1) : \{i, j\} \in \mathcal{E}_{01}\}| \le \gamma \frac{\log n}{n} |I_\epsilon(G_1)|.$$

Next, define the event

$$\mathcal{G} := \left\{ |I_\epsilon(G_1)| \le n^{1 - s D_+(\alpha, \beta) + 2\epsilon |\log(\alpha/\beta)|} \right\}.$$

Notice that $\mathbb{P}(\mathcal{G}) = 1 - o(1)$ by Lemma 29 and Markov's inequality, provided $s D_+(\alpha, \beta) < 99$. We may assume so without loss of generality, since if $s D_+(\alpha, \beta) > 1$ we could simply recover all communities from $G_1$ or $G_2$ alone. Following identical arguments as the proof of Lemma 33, we arrive at

$$\mathbb{P}\left( \exists i \in [n] : |\mathcal{N}_{G_2 \setminus_{\pi_*} G_1}(i) \cap I_\epsilon(G_1)| \ge m \right) = o(1),$$

provided $m > (s D_+(\alpha, \beta) - 2\epsilon|\log(\alpha/\beta)|)^{-1}$. In particular, if $0 < \epsilon \le \frac{s D_+(\alpha, \beta)}{4|\log(\alpha/\beta)|}$ then it suffices to set $m = 2\lceil (s D_+(\alpha, \beta))^{-1} \rceil + 1$. ∎

Putting all of our intermediate results together shows that for each vertex in $(G_2 \setminus_{\widehat{\mu}} G_1)\{[n] \setminus \overline{F}\}$, the number of incorrectly-classified neighbors is small, hence Algorithm 2 succeeds with probability $1 - o(1)$. We prove this formally below.

**Proof** [Proof of Lemma 31] In light of Lemma 21, it suffices to prove the claim with $G_1 \vee_{\widehat{\mu}} G_2$ replaced by $(G_1 \vee_{\pi_*} G_2)\{[n] \setminus \overline{F}_*\}$. As a shorthand, we denote the latter graph by $H$. We also set $H_1 := G_1\{[n] \setminus \overline{F}_*\}$ and $H_{2 \setminus 1} := (G_2 \setminus_{\pi_*} G_1)\{[n] \setminus \overline{F}_*\}$. Let $\widehat{\sigma}_1$ be the almost-exact community labels for $G_1$ produced by Algorithm 3, and notice that in light of Lemma 27, the set of errors are contained in $I_\epsilon(G_1)$ with probability $1 - o(1)$. Moving forward, we may therefore condition on this event.

We can now compare the neighborhood majorities in $H$ corresponding to $\widehat{\sigma}_1$ with the true majority in $H$ as follows. For any $i \in [n] \setminus \overline{F}_*$, we have that

$$\left| \sigma_*(i) \sum_{j \in \mathcal{N}_H(i)} \widehat{\sigma}_1(j) - \mathsf{maj}_H(i) \right| = \left| \sum_{j \in \mathcal{N}_H(i)} \sigma_*(i) \left( \widehat{\sigma}_1(j) - \sigma_*(j) \right) \right| \leq |\mathcal{N}_H(i) \cap I_\epsilon(G_1)|$$

$$= |\mathcal{N}_{H_1}(i) \cap I_\epsilon(G_1)| + |\mathcal{N}_{H_{2 \setminus 1}}(i) \cap I_\epsilon(G_1)|$$

$$\leq |\mathcal{N}_{G_1}(i) \cap I_\epsilon(G_1)| + |\mathcal{N}_{G_2 \setminus \pi_* G_1}(i) \cap I_\epsilon(G_1)| \leq \frac{\epsilon}{2} \log n.$$

Above, the first inequality follows since $\widehat{\sigma}_1(j) \neq \sigma_*(j)$ implies that $j \in I_\epsilon(G_1)$; the next inequality follows since $H_1$ is a subgraph of $G_1$ and $H_{2 \setminus 1}$ is a subgraph of $G_2 \setminus_{\pi_*} G_1$; and the final inequality is due to Lemmas 30 and 34. To conclude the proof, notice that since $\mathsf{maj}_H(i) \geq \epsilon \log n$ for $i \in [n] \setminus \overline{F}_*$ by Lemma 33, it follows that the sign of all neighborhood majorities are equal to the ground truth community label for all vertices in $H$, with probability $1 - o(1)$. ∎

## F.3. Classifying $\overline{F} \setminus F$

The remaining set to classify is $\overline{F} \setminus F$, which we can do via a simple majority vote.

**Lemma 35** *Suppose that $\alpha, \beta, \epsilon > 0$ satisfy the following conditions:*

$$(1 - (1 - s)^2) D_+(\alpha, \beta) > 1 + 2\epsilon |\log(\alpha/\beta)|, \qquad 0 < \epsilon \leq \frac{s D_+(\alpha, \beta)}{4 |\log(\alpha/\beta)|},$$

$$s^2 T_c(\alpha, \beta) + s(1 - s) D_+(\alpha, \beta) > 1.$$

*Then, with high probability, Step 4 of Algorithm 2 with input $(G_1, G_2, (\widehat{M}, \widehat{\mu}), \alpha, \beta, s, \epsilon)$ correctly labels all vertices in $\overline{F} \setminus F$.*

**Proof** We provide a proof for the case $\alpha > \beta$; the case $\alpha < \beta$ follows from identical arguments. For $i \in [n]$, define the graphs $\widehat{H}_i := (G_1 \setminus_{\widehat{\mu}} G_2)\{\widehat{M} \cup \{i\}\}$ and $H_i := (G_1 \setminus_{\pi_*} G_2)\{M_* \cup \{i\}\}$. We also define $E_i$ to be the event where $i$ has a majority of at most $\epsilon' \log n$ in $\widehat{H}_i$ with respect to the community labeling $\widehat{\sigma}$. Our goal is to upper-bound $\mathbb{P}\left( \cup_i (\{i \in \overline{F}\} \cap E_i) \right)$.

To study this probability, it will be useful to define a "nice" event based on our previous results on Algorithm 2. Let $F_* := [n] \setminus M_*$, and let $\overline{F}'_*$ be the result of applying Algorithm 4 to $(G_2, \pi_*(F_*))$. Let $\overline{F}_* := \pi_*^{-1}(\overline{F}'_*)$. That is, $\overline{F}_*$ is the analogue of $\overline{F}$. Finally, let $\widehat{\sigma}$ be the labeling produced by Step 3 of Algorithm 2. For a fixed $\delta > 0$, we now define the event $\mathcal{H}$, which holds if and only if:

- $\overline{F} = \overline{F}_*$;

- $G_1 \setminus_{\widehat{\mu}} G_2 = (G_1 \setminus_{\pi_*} G_2)\{M_*\}$;

- $\widehat{\sigma}(i) = \sigma_*(i)$ for all $i \in [n] \setminus \overline{F}_*$;

- The event $\mathcal{F}$ holds (see Definition 12);

- $|\overline{F}_*| \le n^{1-s^2 \mathrm{T_c}(\alpha,\beta)+\delta}$.

By Lemmas 13, 21, 25, 26, and 31, we have that $\mathbb{P}(\mathcal{H}) = 1 - o(1)$. Furthermore, if we define the event $E_{*,i} := \{\mathsf{maj}_{H_i}(i) \le \epsilon' \log n\}$, we have that

$$
\mathbb{P}\left(\bigcup_{i\in[n]} (\{i \in \overline{F}\} \cap E_i)\right) \le \mathbb{P}\left(\left(\bigcup_{i\in[n]} (\{i \in \overline{F}\} \cap E_i)\right) \cap \mathcal{H}\right) + \mathbb{P}(\mathcal{H}^c)
$$

$$
\le \mathbb{P}\left(\bigcup_{i\in[n]} (\{i \in \overline{F}_*\} \cap E_{*,i} \cap \{|\overline{F}_*| \le n^{1-s^2 \mathrm{T_c}(\alpha,\beta)+\delta}\} \cap \mathcal{F})\right) + o(1)
$$

$$
\le \sum_{i=1}^{n} \mathbb{P}\left(\{i \in \overline{F}_*\} \cap E_{*,i} \cap \{|\overline{F}_*| \le n^{1-s^2 \mathrm{T_c}(\alpha,\beta)+\delta}\} \cap \mathcal{F}\right) + o(1).
$$

$$(23)$$

Using the tower rule, we may rewrite the terms in the summation on the right hand side as

$$
\mathbb{E}\left[\mathbb{P}\left(E_{*,i} \,\middle|\, \pi_*, \boldsymbol{\sigma}_*, \mathcal{E}, \overline{F}_*\right) \mathbf{1}(i \in \overline{F}_*) \mathbf{1}(\{|\overline{F}_*| \le n^{1-s^2 \mathrm{T_c}(\alpha,\beta)+\delta}\} \cap \mathcal{F})\right]. \tag{24}
$$

An important goal is to therefore understand $\mathbb{P}\left(E_{*,i} \,\middle|\, \pi_*, \boldsymbol{\sigma}_*, \mathcal{E}, \overline{F}_*\right)$. To this end, notice that conditionally on $\pi_*$, $\boldsymbol{\sigma}_*$, and $\mathcal{E}$, we have the distributional representation $\mathsf{maj}_{H_i}(i) \overset{d}{=} Y - Z$, where $Y$ and $Z$ are independent with

$$
Y \sim \mathrm{Bin}\left(|\{j \in M_* : \{i,j\} \in \mathcal{E}_{10} \cap \mathcal{E}^+(\boldsymbol{\sigma}_*^1)\}|, \alpha\frac{\log n}{n}\right),
$$

$$
Z \sim \mathrm{Bin}\left(|\{j \in M_* : \{i,j\} \in \mathcal{E}_{10} \cap \mathcal{E}^-(\boldsymbol{\sigma}_*^1)\}|, \beta\frac{\log n}{n}\right).
$$

Fixing a sufficiently small $\delta > 0$, it holds on the event $\{|\overline{F}_*| \le n^{1-s^2 \mathrm{T_c}(\alpha,\beta)+\delta}\} \cap \mathcal{F}$ (see Definition 12 for a definition of the event $\mathcal{F}$) that $|\{j \in M_* : \{i,j\} \in \mathcal{E}_{10} \cap \mathcal{E}^+(\boldsymbol{\sigma}_*^1)\}| = (1 - o(1))s(1 - s)n/2$ and $|\{j \in M_* : \{i,j\} \in \mathcal{E}_{10} \cap \mathcal{E}^-(\boldsymbol{\sigma}_*^1)\}| = (1 - o(1))s(1 - s)n/2$. Lemma 8 therefore implies

$$
\mathbb{P}\left(E_{*,i} \,\middle|\, \pi_*, \boldsymbol{\sigma}_*, \mathcal{E}, \overline{F}_*\right) \mathbf{1}(\{|\overline{F}_*| \le n^{1-s^2 \mathrm{T_c}(\alpha,\beta)+\delta}\} \cap \mathcal{F}) \le n^{-s(1-s)\mathrm{D_+}(\alpha,\beta)+\frac{\epsilon'}{2}\log(\alpha/\beta)+o(1)}.
$$

Utilizing the expression in (24) and taking a union bound over $1 \le i \le n$, we obtain

$$
\sum_{i=1}^{n} \mathbb{P}(\{i \in \overline{F}_*\} \cap E_{*,i} \cap \{|\overline{F}_*| \le n^{1-s^2 \mathrm{T_c}(\alpha,\beta)+\delta}\} \cap \mathcal{F})
$$

$$
\le n^{-s(1-s)\mathrm{D_+}(\alpha,\beta)+\frac{\epsilon'}{2}\log(\alpha/\beta)+o(1)} \sum_{i=1}^{n} \mathbb{E}[\mathbf{1}(i \in \overline{F}_*)\mathbf{1}(|\overline{F}_*| \le n^{1-s^2 \mathrm{T_c}(\alpha,\beta)+\delta})]
$$

$$
= n^{-s(1-s)\mathrm{D_+}(\alpha,\beta)+\frac{\epsilon'}{2}\log(\alpha/\beta)+o(1)} \mathbb{E}[|\overline{F}_*|\mathbf{1}(|\overline{F}_*| \le n^{1-s^2 \mathrm{T_c}(\alpha,\beta)+\delta})]
$$

$$
\le n^{1-s(1-s)\mathrm{D_+}(\alpha,\beta)-s^2 \mathrm{T_c}(\alpha,\beta)+\frac{\epsilon'}{2}\log(\alpha/\beta)+\delta+o(1)}.
$$

Under the condition $s(1-s)\mathrm{D_+}(\alpha, \beta) + s^2 \mathrm{T_c}(\alpha, \beta) > 1$, we can choose $\epsilon'$ and $\delta$ to be small enough so that the bound on the right hand side is $o(1)$. In light of (23), this proves the lemma. ∎

## Appendix G. Classifying the rest (Algorithm 5)

What remains is to classify the set of vertices outside of the $k$-core matching. Again, this can be handled through a simple majority vote.

**Lemma 36** *Suppose that $\alpha, \beta, \epsilon$ satisfy the same conditions stated in Lemma 35. Then Step 3 in Algorithm 5 on input $(G_1, G_2, 13, \epsilon)$ correctly labels all $i \in F$ with high probability.*

**Proof** For $\epsilon' > 0$, let $E_i$ be the event that $i$ does not have an $\epsilon' \log n$ neighborhood majority in the graph $(G_1 \setminus_{\pi_*} G_2)\{([n] \setminus F) \cup \{i\}\}$. By nearly identical arguments to the proof of Lemma 35, we can show that

$$\mathbb{P}\left(\cup_{i \in [n]}(\{i \in F\} \cap E_i)\right) \leq n^{1 - s^2 \mathrm{T_c}(\alpha, \beta) - s(1-s)\mathrm{D_+}(\alpha, \beta) + O(\epsilon')} + o(1).$$

For $\epsilon'$ sufficiently small, the right hand side is $o(1)$ by the assumption on the parameters.

Next, due to the maximality of the 13-core, any $i \in F$ has at most 12 neighbors in the graph $(G_1 \wedge_{\pi_*} G_2)\{([n] \setminus F) \cup \{i\}\}$. Therefore, any $i \in F$ has an $(\epsilon' \log n - 12)$ majority in $G_1\{([n] \setminus F) \cup \{i\}\}$, with high probability. ∎

## Appendix H. Proof of Theorem 5

Finally, we prove Theorem 5, from which Theorem 1 directly follows.

**Proof** [Proof of Theorem 5] By Lemma 21, the matching $(\widehat{M}, \widehat{\mu})$ obtained in Step 1 coincides with the matching $(M_*, \pi_*\{M_*\})$, with high probability. Subsequently, Lemmas 31, 35, and 36 respectively show that the vertices in the sets $[n] \setminus \overline{F}, \overline{F} \setminus F$, and $F$ are correctly labeled with high probability. ∎

## Appendix I. Impossibility of Exact Community Recovery

In this section we prove Theorem 2, which states the conditions under which exact community recovery is impossible. Since impossibility under the condition $\left(1 - (1-s)^2\right)\mathrm{D_+}(\alpha, \beta) < 1$ was already proven in Rácz and Sridhar (2021), here we prove impossibility in the regime

$$s^2 \mathrm{T_c}(\alpha, \beta) + s(1-s)\mathrm{D_+}(\alpha, \beta) < 1. \tag{25}$$

To do so, we study the performance of the maximum a posteriori (MAP) estimator for the communities in $G_1$. We will show that, even when supplied with additional information, such as all the correct community labels in $G_2$ and most of the true vertex matching $\pi_*$, the MAP estimator fails to exactly recover communities with probability bounded away from zero if the condition (25) holds. Since the MAP estimator is *optimal* in the sense that it maximizes the probability of correctness over all estimators (see, e.g., (Poor, 1994, Chapter 4)), the result of Theorem 2 follows.

### I.1. Notation

We briefly review and introduce some notation. Let $\boldsymbol{\sigma_*^1}$ be the ground-truth community labels in $G_1$, so $\sigma_*^1(i)$ is the ground-truth community label of $i \in [n]$. Correspondingly, we define the ground-truth community partition $(V_1^+, V_1^-)$, where

$$V_1^+ := \{i \in [n] : \sigma_*^1(i) = +1\} \qquad \text{and} \qquad V_1^- := \{i \in [n] : \sigma_*^1(i) = -1\}.$$

We define $\boldsymbol{\sigma_*^2}$, $V_2^+$, and $V_2^-$ to be the analogous quantities for $G_2$. Since $\sigma_*^2(i) = \sigma_*^1(\pi_*^{-1}(i))$, we have that $\sigma_*^2(\pi_*(i)) = \sigma_*^1(i)$, and also that $V_2^+ = \pi_*(V_1^+)$ and $V_2^- = \pi_*(V_1^-)$. Finally, for a given $\pi \in \mathcal{S}_n$ and a set $K \subset [n]$, let $\pi\{K\} := \{(i, \pi(i))\}_{i \in K}$ denote the restriction of $\pi$ to the set $K$.

### I.2. The MAP estimator

We begin by defining the *singleton set* of a permutation $\pi$ with respect to the adjacency matrices $A$ and $B$ to be

$$R(\pi, A, B) := \left\{ i \in [n] : \forall j \in [n], A_{i,j} B_{\pi(i),\pi(j)} = 0 \right\}.$$

In words, $R(\pi, A, B)$ is the set of singletons in $G_1 \wedge_\pi G_2$, with respect to the labeling in $G_1$. For brevity, we also write $R_\pi := R(\pi, A, B)$ and $R_* := R(\pi_*, A, B)$. We will also consider a pruned version of $R_\pi$ that is formally described below.

**Definition 37 (The set $S(\pi, A, B)$)** *We have that $i \in S(\pi, A, B)$ if and only if the following conditions hold:*

1. *We have that $i \in R_\pi$.*

2. *Vertex $i$ is a singleton in $G_1\{R_\pi\}$ (equivalently, $A_{i,j} = 0$ for all $j \in R_\pi$).*

3. *If $j \in \mathcal{N}_1(i)$, then $\pi(j) \notin \mathcal{N}_2(\pi(R_\pi))$ (equivalently, if $A_{i,j} = 1$, then $B_{k,\pi(j)} = 0$ for all $k \in \pi(R_\pi)$).*

*For brevity, we also write $S_\pi := S(\pi, A, B)$ and $S_* := S(\pi_*, A, B)$. We remark that if we define the set $\overline{R}_\pi := R_\pi \cup \pi^{-1}(\mathcal{N}_2(\pi(R_\pi)))$, then conditions #2 and #3 above can be succinctly stated as $A_{i,j} = 0$ for all $j \in \overline{R}_\pi$. As per our conventions, when $\pi = \pi_*$ we write $\overline{R}_*$.*

At a high level, $S_*$ is a set that contains no overlapping information across $G_1$ and $G_2$ under the assumption that $\pi_* = \pi$. As we shall see through a detailed study of the posterior distribution of $\pi_*$ and $\boldsymbol{\sigma_*^1}$, information from $G_2$ provides no non-trivial information about the community structure of $S_*$ in $G_1$.

Moving forward, our goal is to study the MAP estimate equipped with the additional knowledge of $\boldsymbol{\sigma_*^2}$, $S_*$, and $\pi_*\{[n] \setminus S_*\}$. It is useful to note that several important quantities can be readily obtained from this revealed information, as we detail next.

- For $i \in [n] \setminus S_*$, the community label $\sigma_*^1(i)$ can be deduced via $\sigma_*^1(i) = \sigma_*^2(\pi_*(i))$.

- The image of $S_*$ under $\pi_*$ can be found via $\pi_*(S_*) = [n] \setminus \pi_*([n] \setminus S_*)$.

- We may compute $|S_* \cap V_1^+|$ via the following formula:

$$|S_* \cap V_1^+| = |\{i \in [n] : \sigma_*^1(i) = +1\}| - |\{i \in [n] \setminus S_* : \sigma_*^1(i) = +1\}|$$
$$= |\{i \in [n] : \sigma_*^2(i) = +1\}| - |\{i \in [n] \setminus S_* : \sigma_*^2(\pi_*(i)) = +1\}|,$$

where in the second equality we used that the communities have the same sizes in $G_1$ and $G_2$.

- We may compute $|S_* \cap V_1^-|$ similarly, or also via $|S_* \cap V_1^-| = |S_*| - |S_* \cap V_1^+|$.

- Finally, define

$$\mathsf{maj}(i) := \sum_{j \in \mathcal{N}_1(i)} \sigma_*^1(j),$$

and note that for $i \in S_*$ the quantity $\mathsf{maj}(i)$ can also be computed using the revealed information. This follows from condition #2 of Definition 37: if $i \in S_*$, then $\mathcal{N}_1(i) \subseteq [n] \setminus S_*$, and we have already noted that $\sigma_*^1(j)$ is known when $j \in [n] \setminus S_*$.

With these observations made, we are now ready to describe the MAP estimator given this additional information.

**Theorem 38 (MAP estimator)** *Let $A$, $B$, $\boldsymbol{\sigma}_*^2$, $S_*$, and $\pi_*\{[n] \setminus S_*\}$ be given, and let $\widehat{\boldsymbol{\sigma}}_{\mathrm{MAP}}$ denote the MAP estimator given this information. For $i \in [n] \setminus S_*$, we have that $\widehat{\sigma}_{\mathrm{MAP}}(i) = \sigma_*^2(\pi_*(i))$. For the vertices in $S_*$, the MAP estimator depends on whether $\alpha$ or $\beta$ is larger, as follows.*

- *If $\alpha > \beta$, then the MAP estimator assigns the label $+1$ to the vertices corresponding to the largest $|S_* \cap V_1^+|$ values in the collection $\{\mathsf{maj}(i)\}_{i \in S_*}$ (breaking ties arbitrarily), and it assigns the label $-1$ to the remaining vertices in $S_*$.*

- *If $\alpha < \beta$, then it does the opposite. That is, the MAP estimator assigns the label $+1$ to the vertices corresponding to the smallest $|S_* \cap V_1^+|$ values in the collection $\{\mathsf{maj}(i)\}_{i \in S_*}$ (breaking ties arbitrarily), and it assigns the label $-1$ to the remaining vertices in $S_*$.*

The structure of the MAP estimator highlights the intuition that $G_2$ contains no relevant information about communities in $S_*$: indeed, the assignment of communities depends only on $\mathsf{maj}$, which is a function of $G_1$ only. In the following corollary, we provide a simple condition for the failure of the MAP estimator.

**Corollary 39** *If $\alpha > \beta$ and there exist $i \in S_* \cap V_1^+, j \in S_* \cap V_1^-$ such that $\mathsf{maj}(i) < \mathsf{maj}(j)$, then the MAP estimator fails (i.e., $\widehat{\boldsymbol{\sigma}}_{\mathrm{MAP}} \neq \boldsymbol{\sigma}_*$). Similarly, if $\alpha < \beta$ and there exist $i \in S_* \cap V_1^+, j \in S_* \cap V_1^-$ such that $\mathsf{maj}(i) > \mathsf{maj}(j)$, then the MAP estimator fails.*

**Proof** Suppose that $\alpha > \beta$. If the MAP estimator correctly classifies $i$ as $+1$, it follows from Theorem 38 that the MAP estimator also classifies $j$ as $+1$, which is incorrect. A similar argument holds for the case $\alpha < \beta$. ∎

### I.3. Analysis of the MAP estimator: Proof of Theorem 2

We start by defining a few useful quantities. First, we define the sigma algebra $\mathcal{I}$, which contains the information we condition on to study the performance of the MAP estimator.

**Definition 40 (The sigma algebra $\mathcal{I}$)** *We let $\mathcal{I}$ be the sigma algebra induced by the random variables $B$, $\pi_*$, $\boldsymbol{\sigma}_*^1$, $\mathcal{E}_{00}$, $\mathcal{E}_{01}$, $\mathcal{E}_{10}$, $\mathcal{E}_{11}$, and $R_*$.*

An important consequence of Definition 40 is that $\overline{R}_* = R_* \cup \pi_*^{-1}(\mathcal{N}_2(\pi_*(R_*)))$ is $\mathcal{I}$-measurable. Since $\boldsymbol{\sigma}_*^1$ is $\mathcal{I}$-measurable, it follows that $\overline{R}_* \cap V_1^+$ and $\overline{R}_* \cap V_1^-$ are $\mathcal{I}$-measurable as well.

Next, we describe a useful $\mathcal{I}$-measurable event which concerns the number of vertices of each community in $R_*$ and $\overline{R}_*$.

**Definition 41 (The event $\mathcal{G}_\delta$)** *Let $\delta > 0$. We say that the event $\mathcal{G}_\delta$ holds if and only if*

$$n^{1-s^2 \mathrm{T_c}(\alpha,\beta)-\delta} \leq |R_* \cap V_1^+|, |R_* \cap V_1^-|, |\overline{R}_* \cap V_1^+|, |\overline{R}_* \cap V_1^-| \leq n^{1-s^2 \mathrm{T_c}(\alpha,\beta)+\delta}.$$

Our next result shows that $\mathcal{G}_\delta$ holds with high probability. Since the proof is straightforward but tedious, we defer it to Section I.6.

**Lemma 42** *For any fixed $\delta > 0$, we have that $\mathbb{P}(\mathcal{G}_\delta) = 1 - o(1)$ as $n \to \infty$.*

We now turn to the proof of our impossibility result. For $i \in R_*$, define the indicator variable

$$W_i := \begin{cases} \mathbf{1}(i \in S_*, \mathsf{maj}(i) < 0) & \text{if } i \in R_* \cap V_1^+, \\ \mathbf{1}(i \in S_*, \mathsf{maj}(i) > 0) & \text{if } i \in R_* \cap V_1^-. \end{cases}$$

Our strategy is to show that $\sum_{i \in R_* \cap V_1^+} W_i > 0$ and $\sum_{i \in R_* \cap V_1^-} W_i > 0$ with high probability. This guarantees the existence of $i \in S_* \cap V_1^+$ and $j \in S_* \cap V_1^-$ such that $\mathsf{maj}(i) < 0 < \mathsf{maj}(j)$, which in turn implies that the MAP estimator fails in light of Corollary 39. To carry out these ideas formally, we will study $\sum_{i \in R_* \cap V_1^+} W_i$ and $\sum_{i \in R_* \cap V_1^+} W_i$ through a second moment method. To this end, the following intermediate results establish useful bounds on the first and second moments, on the event that $\mathcal{F} \cap \mathcal{G}_\delta$ holds (see Definition 12 for a formal definition of $\mathcal{F}$).

**Lemma 43 (First moment estimate)** *Fix $\delta > 0$ and denote $\theta := 1 - s^2 \mathrm{T_c}(\alpha,\beta) - s(1-s)\mathrm{D}_+(\alpha,\beta)$. Then*

$$\mathbb{E}\left[\sum_{i \in R_* \cap V_1^+} W_i \,\middle|\, \mathcal{I}\right] \mathbf{1}(\mathcal{F} \cap \mathcal{G}_\delta) \geq \left(1 - n^{-s^2 \mathrm{T_c}(\alpha,\beta)+2\delta}\right) n^{\theta-\delta-o(1)} \mathbf{1}(\mathcal{F} \cap \mathcal{G}_\delta)$$

*and*

$$\mathbb{E}\left[\sum_{i \in R_* \cap V_1^-} W_i \,\middle|\, \mathcal{I}\right] \mathbf{1}(\mathcal{F} \cap \mathcal{G}_\delta) \geq \left(1 - n^{-s^2 \mathrm{T_c}(\alpha,\beta)+2\delta}\right) n^{\theta-\delta-o(1)} \mathbf{1}(\mathcal{F} \cap \mathcal{G}_\delta).$$

**Lemma 44 (Second moment estimate)** *Fix $\delta > 0$ and let $\theta := 1 - s^2 \mathrm{T_c}(\alpha, \beta) - s(1-s)\mathrm{D_+}(\alpha, \beta)$.*
*If $\theta > 0$, it holds for sufficiently small $\delta$ and all $n$ large enough that*

$$\mathrm{Var}\left(\sum_{i \in R_* \cap V_1^+} W_i \,\middle|\, \mathcal{I}\right) \mathbf{1}(\mathcal{F} \cap \mathcal{G}_\delta) \leq n^{2\theta - 3\delta} \mathbf{1}(\mathcal{F} \cap \mathcal{G}_\delta)$$

*and*

$$\mathrm{Var}\left(\sum_{i \in R_* \cap V_1^-} W_i \,\middle|\, \mathcal{I}\right) \mathbf{1}(\mathcal{F} \cap \mathcal{G}_\delta) \leq n^{2\theta - 3\delta} \mathbf{1}(\mathcal{F} \cap \mathcal{G}_\delta).$$

We defer the proofs of Lemmas 43 and 44 to Section I.7. We now utilize the first and second moment estimates to prove Theorem 2.

**Proof** [Proof of Theorem 2] We start by proving that $\sum_{i \in R_* \cap V_1^+} W_i > 0$ holds with high probability. To establish a lower bound for this event, we may apply the second moment method (specifically, the Cauchy–Schwarz inequality) to obtain that

$$\mathbb{P}\left(\sum_{i \in R_* \cap V_1^+} W_i > 0 \,\middle|\, \mathcal{I}\right) \geq \frac{\mathbb{E}\left[\sum_{i \in R_* \cap V_1^+} W_i \,\middle|\, \mathcal{I}\right]^2}{\mathbb{E}\left[\left(\sum_{i \in R_* \cap V_1^+} W_i\right)^2 \,\middle|\, \mathcal{I}\right]}$$

$$= \frac{\mathbb{E}\left[\sum_{i \in R_* \cap V_1^+} W_i \,\middle|\, \mathcal{I}\right]^2}{\mathbb{E}\left[\sum_{i \in R_* \cap V_1^+} W_i \,\middle|\, \mathcal{I}\right]^2 + \mathrm{Var}\left(\sum_{i \in R_* \cap V_1^+} W_i \,\middle|\, \mathcal{I}\right)} \geq 1 - \frac{\mathrm{Var}\left(\sum_{i \in R_* \cap V_1^+} W_i \,\middle|\, \mathcal{I}\right)}{\mathbb{E}\left[\sum_{i \in R_* \cap V_1^+} W_i \,\middle|\, \mathcal{I}\right]^2}.$$

To obtain a lower bound on the *unconditional* probability, we can thus write, for a fixed $\delta > 0$ satisfying $\delta < s^2 \mathrm{T_c}(\alpha, \beta)/2$, that

$$\mathbb{P}\left(\sum_{i \in R_* \cap V_1^+} W_i > 0\right) \geq \mathbb{P}\left(\left\{\sum_{i \in R_* \cap V_1^+} W_i > 0\right\} \cap \mathcal{F} \cap \mathcal{G}_\delta\right)$$

$$= \mathbb{E}\left[\mathbb{P}\left(\sum_{i \in R_* \cap V_1^+} W_i > 0 \,\middle|\, \mathcal{I}\right) \mathbf{1}(\mathcal{F} \cap \mathcal{G}_\delta)\right]$$

$$\geq \mathbb{E}\left[\left(1 - \frac{\mathrm{Var}\left(\sum_{i \in R_* \cap V_1^+} W_i \,\middle|\, \mathcal{I}\right)}{\mathbb{E}\left[\sum_{i \in R_* \cap V_1^+} W_i \,\middle|\, \mathcal{I}\right]^2}\right) \mathbf{1}(\mathcal{F} \cap \mathcal{G}_\delta)\right]$$

$$\geq \left(1 - (1 - o(1)) n^{2\theta - 3\delta - (2\theta - 2\delta) + o(1)}\right) \mathbb{P}(\mathcal{F} \cap \mathcal{G}_\delta) = 1 - o(1),$$

where the equality on the second line follows since the events $\mathcal{F}$ and $\mathcal{G}_\delta$ are $\mathcal{I}$-measurable, the inequality on the fourth line uses Lemmas 43 and 44, and the final equality is due to Lemmas 13 and 42. An identical analysis shows that $\mathbb{P}(\sum_{i \in R_* \cap V_1^-} W_i > 0) = 1 - o(1)$. In light of Corollary 39, it follows that the MAP estimator fails with probability $1 - o(1)$. $\blacksquare$

### I.4. Properties of $S_\pi$

In this section, we prove some properties of the set $S_\pi$ that will be useful in studying the MAP estimator. To begin, we define the set

$$\mathcal{A}(S_*, \pi_*\{[n] \setminus S_*\}) := \{\pi \in \mathcal{S}_n : S_\pi = S_* \text{ and } \pi\{[n] \setminus S_\pi\} = \pi_*\{[n] \setminus S_*\}\}.$$

For brevity, we sometimes write $\mathcal{A}_*$ instead. The following result highlights the important property that for $\pi \in \mathcal{A}_*$, the structure of $G_1 \wedge_\pi G_2$ is invariant.

**Lemma 45** *For any $\pi \in \mathcal{A}_*$, we have that $A_{i,j} B_{\pi(i),\pi(j)} = A_{i,j} B_{\pi_*(i),\pi_*(j)}$. Moreover, if $i \in S_*$ or $j \in S_*$, then $A_{i,j} B_{\pi(i),\pi(j)} = A_{i,j} B_{\pi_*(i),\pi_*(j)} = 0$.*

**Proof** If $i, j \in [n] \setminus S_*$, then $\pi(i) = \pi_*(i)$ and $\pi(j) = \pi_*(j)$, hence $A_{i,j} B_{\pi(i),\pi(j)} = A_{i,j} B_{\pi_*(i),\pi_*(j)}$. On the other hand, if $i \in S_*$ or $j \in S_*$, then $A_{i,j} B_{\pi(i),\pi(j)} = A_{i,j} B_{\pi_*(i),\pi_*(j)} = 0$ by the definition of $S_\pi = S_*$ (see Definition 37, in particular condition #1). ∎

The following lemma shows that $\mathcal{A}_*$ is closed under permutations of $S_*$. For a more formal discussion, it will be useful to introduce the following definition.

**Definition 46 (The permutation $P_{\pi,\rho}$)** *Let $\rho$ be a permutation of $S_*$. The permutation $P_{\pi,\rho}$ is given by*

$$P_{\pi,\rho}(i) := \begin{cases} \pi(i) & \text{if } i \in [n] \setminus S_*, \\ \pi(\rho(i)) & \text{if } i \in S_*. \end{cases}$$

**Lemma 47** *Let $\rho$ be a permutation of $S_*$. Then $P_{\pi_*,\rho} \in \mathcal{A}_*$.*

**Proof** We abbreviate $P = P_{\pi_*,\rho}$. It is clear from the construction of $P$ that $P\{[n] \setminus S_*\} = \pi_*\{[n] \setminus S_*\}$. To prove $S_P = S_*$, it is sufficient to show that $R_P = R_*$. Indeed, once this has been established, it follows readily that $\overline{R}_P = \overline{R}_*$, and thus $S_P = S_*$ by Definition 37.

We start by proving $R_* \subseteq R_P$; we do so by proving separately that $R_* \setminus S_* \subseteq R_P$ and $S_* \subseteq R_P$.

To prove the first claim, let $i \in R_* \setminus S_*$ and notice that $P(i) = \pi_*(i)$. Fix $j \in S_*$. Then, by condition #2 of Definition 37, $A_{i,j} = 0$, so $A_{i,j} B_{P(i),P(j)} = 0$. On the other hand, if $j \in [n] \setminus S_*$, then $P(j) = \pi_*(j)$, so $A_{i,j} B_{P(i),P(j)} = A_{i,j} B_{\pi_*(i),\pi_*(j)} = 0$. Since we have shown that $A_{i,j} B_{P(i),P(j)} = 0$ for all $i \in R_* \setminus S_*$ and for all $j \in [n]$, it follows that $R_* \setminus S_* \subseteq R_P$ as desired.

We now prove that $S_* \subseteq R_P$. Let $i \in S_*$. If $j \in R_*$, then by condition #2 of Definition 37 we have $A_{i,j} = 0$, hence $A_{i,j} B_{P(i),P(j)} = 0$. On the other hand if $j \in [n] \setminus R_*$ then by condition #3 of Definition 37, $A_{i,j} = 1$ implies $B_{P(i),P(j)} = 0$ since $\pi_*(j) = P(j)$ and $P(i) \in \pi_*(R_*)$; hence $A_{i,j} B_{P(i),P(j)} = 0$. Since we have shown that $A_{i,j} B_{P(i),P(j)} = 0$ for all $i \in S_*$ and $j \in [n]$, it follows that $S_* \subseteq R_P$. Putting both results together shows that $R_* \subseteq R_P$ as desired.

Next, we prove that $R_P \subseteq R_*$. Suppose by way of contradiction that there exists $i \in R_P \setminus R_*$. Then there must exist $j \in [n]$ such that $A_{i,j} B_{P(i),P(j)} = 0$ and $A_{i,j} B_{\pi_*(i),\pi_*(j)} = 1$, which in turn implies that $(A_{i,j}, B_{P(i),P(j)}, B_{\pi_*(i),\pi_*(j)}) = (1, 0, 1)$. We consider two possibilities for $j$. If $j \in R_*$, then $A_{i,j} = 1$ implies $B_{\pi_*(i),\pi_*(j)} = 0$ by condition #3 of Definition 37, which is a contradiction. On the other hand, if $j \in [n] \setminus R_* \subseteq [n] \setminus S_*$, it follows that $P(j) = \pi_*(j)$. Since $i \in [n] \setminus R_* \subseteq [n] \setminus S_*$ as well, $P(i) = \pi_*(i)$ as well. However, this contradicts $(B_{P(i),P(j)}, B_{\pi_*(i),\pi_*(j)}) = (0, 1)$. Since all cases for $j$ lead to a contradiction, we have that $R_P \subseteq R_*$. ∎

A useful consequence of Lemma 47 is that the elements of $\mathcal{A}_*$ can be parametrized by permutations of $S$. This is captured in the following corollary.

**Corollary 48** *We have the representation*

$$\mathcal{A}_* = \{P_{\pi_*,\rho} : \rho \text{ is a permutation of } S_*\}. \tag{26}$$

**Proof** Lemma 47 shows that the right hand side of (26) is a subset of the left hand side of (26). Moreover, from the definition of the set $\mathcal{A}_*$, if $\pi \in \mathcal{A}_*$, then $\pi$ and $\pi_*$ can only disagree on inputs from $S_*$. This implies that we can find a permutation $\rho$ on $S_*$ such that $\pi = P_{\pi_*,\rho}$. Hence the set on the left hand side of (26) is a subset of the right hand side of (26). ∎

### I.5. Deriving the MAP estimator: Proof of Theorem 38

I.5.1. THE POSTERIOR DISTRIBUTION OF $\pi_*$

We start by defining some notation. For a given permutation $\pi \in \mathcal{S}_n$, define

$$\mu^+(\pi)_{ab} := \sum_{\{\pi(i),\pi(j)\}\in\mathcal{E}^+(\boldsymbol{\sigma}_*^2)} \mathbf{1}((A_{i,j}, B_{\pi(i),\pi(j)}) = (a,b)), \qquad \text{for } a,b \in \{0,1\},$$

$$\mu^-(\pi)_{ab} := \sum_{\{\pi(i),\pi(j)\}\in\mathcal{E}^-(\boldsymbol{\sigma}_*^2)} \mathbf{1}((A_{i,j}, B_{\pi(i),\pi(j)}) = (a,b)), \qquad \text{for } a,b, \in \{0,1\},$$

$$\nu^+(\pi) := \sum_{\{\pi(i),\pi(j)\}\in\mathcal{E}^+(\boldsymbol{\sigma}_*^2)} A_{i,j},$$

$$\nu^-(\pi) := \sum_{\{\pi(i),\pi(j)\}\in\mathcal{E}^-(\boldsymbol{\sigma}_*^2)} A_{i,j}.$$

In words, $\mu^+(\pi)_{ab}$ and $\mu^-(\pi)_{ab}$ capture the empirical joint distribution of correlated edges assuming $\pi_* = \pi$, and $\nu^+(\pi)$ and $\nu^-(\pi)$ count the number of intra-community and inter-community edges in $G_1$, respectively. Using these quantities, we can derive an exact expression for the posterior distribution of $\pi_*$ given $A$, $B$, and $\boldsymbol{\sigma}_*^2$. The proof is nearly identical to (Rácz and Sridhar, 2021, Lemma 3.1), but we include it here for completeness since there are a few small changes (e.g., we condition on the community labels in $G_2$ rather than in $G_1$) and the proof is short.

**Lemma 49** *Let $\pi \in \mathcal{S}_n$. There is a constant $C_1 = C_1(A, B, \boldsymbol{\sigma}_*^2)$ such that*

$$\mathbb{P}\left(\pi_* = \pi \,\big|\, A, B, \boldsymbol{\sigma}_*^2\right) = C_1 \left(\frac{p_{00}p_{11}}{p_{01}p_{10}}\right)^{\mu^+(\pi)_{11}} \left(\frac{q_{00}q_{11}}{q_{01}q_{10}}\right)^{\mu^-(\pi)_{11}} \left(\frac{p_{10}}{p_{00}}\right)^{\nu^+(\pi)} \left(\frac{q_{10}}{q_{00}}\right)^{\nu^-(\pi)}.$$

**Proof** By Bayes' rule, we have that

$$\mathbb{P}\left(\pi_* = \pi \,\big|\, A, B, \boldsymbol{\sigma}_*^2\right) = \frac{\mathbb{P}\left(A, B \,\big|\, \pi_* = \pi, \boldsymbol{\sigma}_*^2\right) \mathbb{P}\left(\pi_* = \pi \,\big|\, \boldsymbol{\sigma}_*^2\right)}{\mathbb{P}\left(A, B, \boldsymbol{\sigma}_*^2\right)}.$$

Recall that in the construction of the correlated pair of SBMs $(G_1, G_2)$, the permutation $\pi_*$ is chosen independently of everything else, including the community labeling $\boldsymbol{\sigma}_*^2$. Hence we can write

$$\mathbb{P}\left(\pi_* = \pi \,\big|\, A, B, \boldsymbol{\sigma}_*^2\right) = c_1\left(A, B, \boldsymbol{\sigma}_*^2\right) \cdot \mathbb{P}\left(A, B \,\big|\, \pi_* = \pi, \boldsymbol{\sigma}_*^2\right), \tag{27}$$

where $c_1\left(A, B, \boldsymbol{\sigma}_*^2\right) := \left(n! \mathbb{P}\left(A, B, \boldsymbol{\sigma}_*^2\right)\right)^{-1}$. We now analyze $\mathbb{P}\left(A, B \mid \pi_* = \pi, \boldsymbol{\sigma}_*^2\right)$. Recall that, given $\boldsymbol{\sigma}_*^2$ and $\pi_*$, the edge formation process in $G_1$ and $G_2$ is mutually independent across all vertex pairs. Hence we have that

$$\mathbb{P}\left(A, B \mid \pi_* = \pi, \boldsymbol{\sigma}_*^2\right) = \left(p_{00}^{\mu^+(\pi)_{00}} p_{01}^{\mu^+(\pi)_{01}} p_{10}^{\mu^+(\pi)_{10}} p_{11}^{\mu^+(\pi)_{11}}\right) \left(q_{00}^{\mu^-(\pi)_{00}} q_{01}^{\mu^-(\pi)_{01}} q_{10}^{\mu^-(\pi)_{10}} q_{11}^{\mu^-(\pi)_{11}}\right).$$
(28)

To simplify (28), we can write

$$\mu^+(\pi)_{01} = \sum_{\{\pi(i), \pi(j)\} \in \mathcal{E}^+(\boldsymbol{\sigma}_*^2)} (1 - A_{i,j}) B_{\pi(i), \pi(j)} = \sum_{\{i,j\} \in \mathcal{E}^+(\boldsymbol{\sigma}_*^2)} B_{i,j} - \mu^+(\pi)_{11};$$

$$\mu^+(\pi)_{10} = \sum_{\{\pi(i), \pi(j)\} \in \mathcal{E}^+(\boldsymbol{\sigma}_*^2)} A_{i,j} (1 - B_{\pi(i), \pi(j)}) = \nu^+(\pi) - \mu^+(\pi)_{11};$$

$$\mu^+(\pi)_{00} = \sum_{\{\pi(i), \pi(j)\} \in \mathcal{E}^+(\boldsymbol{\sigma}_*^2)} (1 - A_{i,j})(1 - B_{\pi(i), \pi(j)})$$

$$= |\mathcal{E}^+(\boldsymbol{\sigma}_*^2)| - \sum_{(i,j) \in \mathcal{E}^+(\boldsymbol{\sigma}_*^2)} B_{i,j} - \nu^+(\pi) + \mu^+(\pi)_{11}.$$

In particular, note that the quantities $\sum_{\{i,j\} \in \mathcal{E}^+(\boldsymbol{\sigma}_*^2)} B_{i,j}$ and $|\mathcal{E}^+(\boldsymbol{\sigma}_*^2)|$ are measurable with respect to $B$ and $\boldsymbol{\sigma}_*^2$. Hence we can write

$$p_{00}^{\mu^+(\pi)_{00}} p_{01}^{\mu^+(\pi)_{01}} p_{10}^{\mu^+(\pi)_{10}} p_{11}^{\mu^+(\pi)_{11}} = c_2^+ \left(\frac{p_{00} p_{11}}{p_{01} p_{10}}\right)^{\mu^+(\pi)_{11}} \left(\frac{p_{10}}{p_{00}}\right)^{\nu^+(\pi)},$$
(29)

where

$$c_2^+ := p_{00}^{|\mathcal{E}^+(\boldsymbol{\sigma}_*^2)|} \left(\frac{p_{01}}{p_{00}}\right)^{\sum_{\{i,j\} \in \mathcal{E}^+(\boldsymbol{\sigma}_*^2)} B_{i,j}}.$$

Replicating these arguments for $\mu^-(\pi)_{ab}$, we have that

$$q_{00}^{\mu^-(\pi)_{00}} q_{01}^{\mu^-(\pi)_{01}} q_{10}^{\mu^-(\pi)_{10}} q_{11}^{\mu^-(\pi)_{11}} = c_2^- \left(\frac{q_{00} q_{11}}{q_{01} q_{10}}\right)^{\mu^-(\pi)_{11}} \left(\frac{q_{10}}{q_{00}}\right)^{\nu^-(\pi)},$$
(30)

where

$$c_2^- := q_{00}^{|\mathcal{E}^-(\boldsymbol{\sigma}_*^2)|} \left(\frac{q_{01}}{q_{00}}\right)^{\sum_{\{i,j\} \in \mathcal{E}^-(\boldsymbol{\sigma}_*^2)} B_{i,j}}.$$

Combining (27), (28), (29), and (30) proves the statement of the lemma, with $C_1 := c_1 c_2^+ c_2^-$. ∎

Next, we build on Lemma 49 to obtain the posterior distribution of $\pi_*$ given not only $A$, $B$, and $\boldsymbol{\sigma}_*^2$, but also $S_*$ and $\pi_*\{[n] \setminus S_*\}$. To show this formally, we recall the definition of the set $\mathcal{A}_*$ from Section I.4.

**Lemma 50 (Posterior distribution)**  *There is a constant $C_2 = C_2(A, B, \boldsymbol{\sigma}_*^2, S_*, \pi_*\{[n] \setminus S_*\})$ such that*

$$\mathbb{P}\left(\pi_* = \pi \mid A, B, \boldsymbol{\sigma}_*^2, S_*, \pi_*\{[n] \setminus S_*\}\right) = C_2 \left(\sqrt{\frac{p_{10} q_{00}}{p_{00} q_{10}}}\right)^{\nu^+(\pi) - \nu^-(\pi)} \mathbf{1}(\pi \in \mathcal{A}_*).$$

**Proof** By Bayes' rule we have that

$$
\mathbb{P}\left(\pi_* = \pi \,\middle|\, A, B, \boldsymbol{\sigma}_*^2, S_*, \pi_*\{[n] \setminus S_*\}\right)
$$

$$
= \frac{\mathbb{P}\left(\pi_* = \pi \,\middle|\, A, B, \boldsymbol{\sigma}_*^2\right) \mathbb{P}\left(S_*, \pi_*\{[n] \setminus S_*\} \,\middle|\, \pi_* = \pi, A, B, \boldsymbol{\sigma}_*^2\right)}{\mathbb{P}\left(S_*, \pi_*\{[n] \setminus S_*\} \,\middle|\, A, B, \boldsymbol{\sigma}_*^2\right)}
$$

$$
= \frac{\mathbb{P}\left(\pi_* = \pi \,\middle|\, A, B, \boldsymbol{\sigma}_*^2\right)}{\mathbb{P}\left(S_*, \pi_*\{[n] \setminus S_*\} \,\middle|\, A, B, \boldsymbol{\sigma}_*^2\right)} \mathbf{1}(\pi \in \mathcal{A}_*).
$$

The probability in the denominator is a function of $A$, $B$, $\boldsymbol{\sigma}_*^2$, $S_*$, and $\pi_*\{[n] \setminus S_*\}$. Furthermore, by Lemma 45, $\mu^+(\pi)_{11}$ and $\mu^-(\pi)_{11}$ are constant over $\pi \in \mathcal{A}_*$. Hence by Lemma 49 we can write

$$
\mathbb{P}\left(\pi_* = \pi \,\middle|\, A, B, \boldsymbol{\sigma}_*^2, S_*, \pi_*\{[n] \setminus S_*\}\right) = c_1 \left(\frac{p_{10}}{p_{00}}\right)^{\nu^+(\pi)} \left(\frac{q_{10}}{q_{00}}\right)^{\nu^-(\pi)} \mathbf{1}(\pi \in \mathcal{A}_*), \qquad (31)
$$

where

$$
c_1 = \frac{C_1}{\mathbb{P}\left(S_*, \pi_*\{[n] \setminus S_*\} \,\middle|\, A, B, \boldsymbol{\sigma}_*^2\right)} \left(\frac{p_{00}p_{11}}{p_{01}p_{10}}\right)^{\mu^+(\pi_*)_{11}} \left(\frac{q_{00}q_{11}}{q_{01}q_{10}}\right)^{\mu^-(\pi_*)_{11}}.
$$

Above, $C_1$ is the same constant as in Lemma 49. In particular, $c_1$ depends only on $A$, $B$, $\boldsymbol{\sigma}_*^2$, $S_*$, and $\pi_*\{[n] \setminus S_*\}$. To simplify the right hand side of (31) further, we can first write

$$
\left(\frac{p_{10}}{p_{00}}\right)^{\nu^+(\pi)} \left(\frac{q_{10}}{q_{00}}\right)^{\nu^-(\pi)} = \left(\sqrt{\frac{p_{10}q_{10}}{p_{00}q_{00}}}\right)^{\nu^+(\pi)+\nu^-(\pi)} \left(\sqrt{\frac{p_{10}q_{00}}{p_{00}q_{10}}}\right)^{\nu^+(\pi)-\nu^-(\pi)}. \qquad (32)
$$

Thus a simplification can be obtained by noting that $\nu^+(\pi) + \nu^-(\pi)$ depends only on $A$. Indeed,

$$
\nu^+(\pi) + \nu^-(\pi) = \sum_{\{\pi(i),\pi(j)\}\in\mathcal{E}^+(\boldsymbol{\sigma}_*^2)} A_{i,j} + \sum_{\{\pi(i),\pi(j)\}\in\mathcal{E}^-(\boldsymbol{\sigma}_*^2)} A_{i,j} = \sum_{\{i,j\}\in\binom{[n]}{2}} A_{i,j}.
$$

The desired result now follows with

$$
C_2 := c_1 \left(\sqrt{\frac{p_{10}q_{10}}{p_{00}q_{00}}}\right)^{\sum_{\{i,j\}\in\binom{[n]}{2}} A_{i,j}}.
$$

■

### I.5.2. THE POSTERIOR DISTRIBUTION OF THE COMMUNITY PARTITION

Our next few results will allow us to translate our characterization of the posterior distribution of $\pi_*$ to the posterior distribution of the community labeling. We proceed by defining some further notation. For a community partition $\mathbf{X} = (X^+, X^-)$ of $[n]$ in $G_1$, define the set

$$
\mathcal{B}(\mathbf{X}) := \left\{\pi \in \mathcal{A}_* : \pi(X^+) = V_2^+, \pi(X^-) = V_2^-\right\}.
$$

In words, $\mathcal{B}(\mathbf{X})$ is the set of permutations in $\mathcal{A}_*$ which induce the community partition $\mathbf{X}$ in $G_1$. Notice that if $\mathcal{B}(\mathbf{X}) \neq \emptyset$, then the partition $\mathbf{X}$ must be compatible with $A$, $B$, $\boldsymbol{\sigma}_*^2$, $S_*$, and $\pi_*\{[n] \setminus S_*\}$. In particular, if $\boldsymbol{\sigma}_\mathbf{X}$ denotes the community memberships associated with $\mathbf{X}$, the following must hold:

- $\sigma_{\mathbf{X}}(i) = \sigma_*^2(\pi_*(i)) = \sigma_*^1(i)$ for $i \in [n] \setminus S_*$;

- $|S_* \cap X^+| = |S_* \cap V_1^+|$ and $|S_* \cap X^-| = |S_* \cap V_1^-|$.

The first condition must hold since we know the true vertex correspondence—and therefore the true community labels—outside of the set $S_*$. The second condition must hold since the number of vertices of each community in $S_*$ can be deduced by examining the community labels of $\pi_*(S_*)$ with respect to $\boldsymbol{\sigma}_*^2$.

We proceed by establishing a few more useful results related to $\mathcal{B}(\mathbf{X})$. The following lemma shows that the size of $\mathcal{B}(\mathbf{X})$ does not depend on the specific choice of $\mathbf{X}$.

**Lemma 51** *If $\mathcal{B}(\mathbf{X})$ is nonempty, then $|\mathcal{B}(\mathbf{X})| = |S_* \cap V_1^+|! |S_* \cap V_1^-|!$.*

**Proof** Suppose that $\pi_0, \pi_1 \in \mathcal{B}(\mathbf{X})$. By Corollary 48, we can write $\pi_1 = P_{\pi_0, \rho}$ for some permutation $\rho$ on $S_*$. Notice that if $i \in S_* \cap X^+$, then $\rho(i) \in S_* \cap X^+$, and if $i \in S_* \cap X^-$, then $\rho(i) \in S_* \cap X^-$; otherwise, $\pi_0(i)$ and $\pi_1(i) = \pi_0(\rho(i))$ would have different community labels with respect to $\boldsymbol{\sigma}_*^2$, which would violate our assumption that $\pi_0, \pi_1 \in \mathcal{B}(\mathbf{X})$. We can therefore decompose $\rho$ into two disjoint permutations $\rho^+$ and $\rho^-$, where $\rho^+$ is a permutation of $S_* \cap X^+$ and $\rho^-$ is a permutation of $S_* \cap X^-$. Since there are $|S_* \cap X^+|! = |S_* \cap V_1^+|!$ choices for $\rho^+$ and $|S_* \cap X^-|! = |S_* \cap V_1^-|!$ choices for $\rho^-$, the desired result follows. ∎

Our next result shows that $\nu^+(\pi) - \nu^-(\pi)$ is invariant over elements of $\mathcal{B}(\mathbf{X})$, where recall that

$$\nu^+(\pi) := \sum_{\{\pi(i), \pi(j)\} \in \mathcal{E}^+(\boldsymbol{\sigma}_*^2)} A_{i,j} \quad \text{and} \quad \nu^-(\pi) := \sum_{\{\pi(i), \pi(j)\} \in \mathcal{E}^-(\boldsymbol{\sigma}_*^2)} A_{i,j}.$$

We briefly recall some relevant notation. Let $\boldsymbol{\sigma}_{\mathbf{X}}$ be the community labels induced by $\mathbf{X}$, so that $\sigma_{\mathbf{X}}(i) = +1$ if $i \in X^+$ and $\sigma_X(i) = -1$ if $i \in X^-$. We also recall that

$$\mathsf{maj}(i) = \sum_{j \in \mathcal{N}_1(i)} \sigma_*^1(j) = \sum_{j \in [n]} A_{i,j} \sigma_*^1(j).$$

While $\mathsf{maj}(i)$ is generally not measurable with respect to $\{A, B, \boldsymbol{\sigma}_*^2, S_*, \pi_*\{[n] \setminus S_*\}\}$, it *is* measurable if $i \in S_*$. The reason is that for $j \in [n] \setminus S_*$, $\pi_*(j)$ is known so we can deduce $\sigma_*^1(j) = \sigma_*^2(\pi_*(j))$. In addition, if $i \in S_*$, then all neighbors of $i$ in $G_1$ are in $[n] \setminus S_*$, so $\mathsf{maj}(i)$ is indeed measurable.

**Lemma 52** *For all $\pi \in \mathcal{B}(\mathbf{X})$, we have that*

$$\nu^+(\pi) - \nu^-(\pi) = C_3 + \sum_{i \in S_*} \sigma_{\mathbf{X}}(i) \mathsf{maj}(i),$$

*where $C_3$ depends on $A$, $B$, $\boldsymbol{\sigma}_*^2$, $S_*$, and $\pi_*\{[n] \setminus S_*\}$, but not on $\mathbf{X}$.*

**Proof** Noting that $\sigma_{\mathbf{X}}(i)\sigma_{\mathbf{X}}(j) = 1$ for $\{\pi(i), \pi(j)\} \in \mathcal{E}^+(\boldsymbol{\sigma}_*^2)$ and that $\sigma_{\mathbf{X}}(i)\sigma_{\mathbf{X}}(j) = -1$ for $\{\pi(i), \pi(j)\} \in \mathcal{E}^-(\boldsymbol{\sigma}_*^2)$, we can write

$$\nu^+(\pi) - \nu^-(\pi) = \sum_{\{i,j\} \in \binom{[n]}{2}} A_{i,j} \sigma_{\mathbf{X}}(i)\sigma_{\mathbf{X}}(j).$$

Since $\pi \in \mathcal{A}_*$, we have $\pi(i) = \pi_*(i)$ for $i \in [n] \setminus S_*$, and hence $\sigma_{\mathbf{X}}(i) = \sigma_*^2(\pi_*(i)) = \sigma_*^1(i)$. Defining

$$C_3 := \sum_{\substack{\{i,j\} \in \binom{[n]}{2}: \\ i,j \in [n] \setminus S_*}} A_{i,j} \sigma_{\mathbf{X}}(i) \sigma_{\mathbf{X}}(j) = \sum_{\substack{\{i,j\} \in \binom{[n]}{2}: \\ i,j \in [n] \setminus S_*}} A_{i,j} \sigma_*^1(i) \sigma_*^1(j),$$

it is clear that $C_3$ depends only on $A$, $\boldsymbol{\sigma}_*^2$, $S_*$, and $\pi_*\{[n] \setminus S_*\}$, and not on $\mathbf{X}$. Noting further that $A_{i,j} = 0$ when $i, j \in S_*$ by the construction of $S_*$, we have that

$$\nu^+(\pi) - \nu^-(\pi) = C_3 + \sum_{i \in S_*, j \in [n] \setminus S_*} A_{i,j} \sigma_{\mathbf{X}}(i) \sigma_{\mathbf{X}}(j) = C_3 + \sum_{i \in S_*} \sigma_{\mathbf{X}}(i) \sum_{j \in [n] \setminus S_*} A_{i,j} \sigma_*^1(j)$$

$$= C_3 + \sum_{i \in S_*} \sigma_{\mathbf{X}}(i) \mathsf{maj}(i).$$

∎

We can now put everything together to derive the posterior probability of a given community partition.

**Lemma 53**  *If $\mathcal{B}(\mathbf{X})$ is nonempty, then*

$$\mathbb{P}\left((V_1^+, V_1^-) = (X^+, X^-) \,\middle|\, A, B, \boldsymbol{\sigma}_*^2, S_*, \pi_*\{[n] \setminus S_*\}\right) = C_4 \left(\frac{p_{10} q_{00}}{p_{00} q_{10}}\right)^{\frac{1}{2} \sum_{i \in S_*} \sigma_{\mathbf{X}}(i) \mathsf{maj}(i)},$$

*where $C_4$ is a constant depending on $A$, $B$, $\boldsymbol{\sigma}_*^2$, $S_*$, and $\pi_*\{[n] \setminus S_*\}$, but* not *on the partition $\mathbf{X}$.*

**Proof**  To compute the posterior probability of a community partition, we may equivalently compute the posterior probability of the set of permutations that generate the community partition under consideration. We thus have that

$$\mathbb{P}\left((V_1^+, V_1^-) = (X^+, X^-) \,\middle|\, A, B, \boldsymbol{\sigma}_*^2, S_*, \pi_*\{[n] \setminus S_*\}\right)$$

$$= \sum_{\pi \in \mathcal{B}(\mathbf{X})} \mathbb{P}\left(\pi_* = \pi \,\middle|\, A, B, \boldsymbol{\sigma}_*^2, S_*, \pi_*\{[n] \setminus S_*\}\right) = \sum_{\pi \in \mathcal{B}(\mathbf{X})} C_2 \left(\frac{p_{10} q_{00}}{p_{00} q_{10}}\right)^{(\nu^+(\pi) - \nu^-(\pi))/2},$$

where the last equality follows by Lemma 50 and $C_2$ is the constant appearing in Lemma 50 (and note that $\pi \in \mathcal{A}_*$ for every $\pi \in \mathcal{B}(\mathbf{X})$). Now by Lemma 52, this sum is equal to

$$C_2 \left(\frac{p_{10} q_{00}}{p_{00} q_{10}}\right)^{C_3/2} \sum_{\pi \in \mathcal{B}(\mathbf{X})} \left(\frac{p_{10} q_{00}}{p_{00} q_{10}}\right)^{\frac{1}{2} \sum_{i \in S_*} \sigma_{\mathbf{X}}(i) \mathsf{maj}(i)}.$$

Since the summand in the display above does not depend on $\pi$, and $|\mathcal{B}(\mathbf{X})| = |S_* \cap V_1^+|! |S_* \cap V_1^-|!$ by Lemma 51, the desired result follows with

$$C_4 := C_2 \left(\frac{p_{10} q_{00}}{p_{00} q_{10}}\right)^{C_3/2} |S_* \cap V_1^+|! |S_* \cap V_1^-|!.$$

■

The characterization of the MAP estimator follows as a corollary.

**Proof** [Proof of Theorem 38] Suppose that $A$, $B$, $\boldsymbol{\sigma}_*^2$, $S_*$, and $\pi_*\{[n] \setminus S_*\}$ are given. First of all, we have that $\widehat{\sigma}_{\mathrm{MAP}}(i) = \sigma_*^2(\pi_*(i))$ for all $i \in [n] \setminus S_*$ (since for any community labeling not satisfying this, the posterior probability is 0). For vertices in $S_*$, first note that

$$\frac{p_{10}q_{00}}{p_{00}q_{10}} = \frac{\alpha}{\beta}(1 + o(1)).$$

Thus, by Lemma 53, the MAP estimator for the community partition of $G_1$ maximizes (resp., minimizes) $\sum_{i \in S_*} \sigma_{\mathbf{X}}(i)\mathsf{maj}(i)$ if $\alpha > \beta$ (resp., $\alpha < \beta$), while respecting the constraint that $|S_* \cap X^+| = |S_* \cap V_1^+|$ and $|S_* \cap X^-| = |S_* \cap V_1^-|$. The maximum is obtained by setting $\sigma_{\mathbf{X}}(i) = +1$ for $i \in S_*$ corresponding to the $|S_* \cap V_1^+|$ largest values of $\{\mathsf{maj}(i)\}_{i \in S_*}$ (breaking ties arbitrarily), and setting $\sigma_{\mathbf{X}}(i) = -1$ for the rest. The minimum is obtained by setting $\sigma_{\mathbf{X}}(i)$ in an opposite manner. ■

### I.6. Bounding the size of "bad" sets: Proof of Lemma 42

**Proof** [Proof of Lemma 42] For a given vertex $i \in [n]$, let $E_i$ be the event that $i$ is a singleton in $G_1 \wedge_{\pi_*} G_2$. Throughout the proof, we assume that the communities are approximately balanced; specifically, we assume that the event $\mathcal{H} := \{n/2 - n^{3/4} \leq |V^+|, |V^-| \leq n/2 + n^{3/4}\}$ holds. Note that $\mathbb{P}(\mathcal{H}) = 1 - o(1)$ by Lemma 13.

Conditioning on $\boldsymbol{\sigma}_*^1$, if $i \in V_1^+$, then we have that

$$
\begin{aligned}
\mathbb{P}\left(E_i \,\middle|\, \boldsymbol{\sigma}_*^1\right) \mathbf{1}(\mathcal{H}) &= \left(1 - s^2\alpha\frac{\log n}{n}\right)^{|V_1^+| - 1} \left(1 - s^2\beta\frac{\log n}{n}\right)^{|V_1^-|} \mathbf{1}(\mathcal{H}) \\
&\leq \exp\left(-s^2\frac{\log n}{n}(\alpha(|V_1^+| - 1) + \beta|V_1^-|)\right) \mathbf{1}(\mathcal{H}) \\
&= \exp\left(-(1 - o(1))s^2 \left(\frac{\alpha + \beta}{2}\right)\log n\right) \mathbf{1}(\mathcal{H}) = n^{-s^2(\alpha+\beta)/2 + o(1)}\mathbf{1}(\mathcal{H}),
\end{aligned}
\tag{33}
$$

where in the inequality on the second line we used that $1 - x \leq e^{-x}$, and subsequently we used that $|V_1^+|, |V_1^-| = (1 + o(1))n/2$ on the event $\mathcal{H}$. The bound in (33) implies that

$$
\begin{aligned}
\mathbb{E}\left[|R_* \cap V_1^+| \,\middle|\, \boldsymbol{\sigma}_*^1\right] \mathbf{1}(\mathcal{H}) &= \sum_{i \in V_1^+} \mathbb{P}\left(E_i \,\middle|\, \boldsymbol{\sigma}_*^1\right) \mathbf{1}(\mathcal{H}) \leq |V_1^+| n^{-s^2(\alpha+\beta)/2 + o(1)}\mathbf{1}(\mathcal{H}) \\
&\leq n^{1 - s^2(\alpha+\beta)/2 + o(1)}.
\end{aligned}
$$

Markov's inequality now implies that

$$\mathbb{P}\left(|R_* \cap V_1^+| \geq n^{1 - s^2(\alpha+\beta)/2 + \delta} \,\middle|\, \boldsymbol{\sigma}_*^1\right) \mathbf{1}(\mathcal{H}) \leq n^{-\delta + o(1)} = o(1), \tag{34}$$

which in turn implies the unconditional bound $\mathbb{P}(|R_* \cap V_1^+| \geq n^{1 - s^2(\alpha+\beta)/2 + \delta}) = o(1)$ as well.

We now focus on deriving a probabilistic *lower* bound for $|R_* \cap V_1^+|$. To this end, we start by deriving a matching lower bound for $\mathbb{P}\left(E_i \,\big|\, \boldsymbol{\sigma}_*^1\right)$. A useful fact we shall use is that by Taylor's theorem, $\log(1-x) \geq -(1+\epsilon)x$ provided $0 < x < \epsilon/(1+\epsilon)$. Thus for $\epsilon = \epsilon_n$ suitably small,

$$\left(\log \mathbb{P}\left(E_i \,\big|\, \boldsymbol{\sigma}_*^1\right)\right)\mathbf{1}(\mathcal{H}) \geq \left(|V_1^+|\log\left(1 - s^2\alpha\frac{\log n}{n}\right) + |V_1^-|\log\left(1 - s^2\beta\frac{\log n}{n}\right)\right)\mathbf{1}(\mathcal{H})$$

$$\geq -(1+\epsilon)s^2\frac{\log n}{n}\left(\alpha|V_1^+| + \beta|V_2^-|\right)\mathbf{1}(\mathcal{H})$$

$$= -(1-o(1))(1+\epsilon)s^2\left(\frac{\alpha+\beta}{2}\right)(\log n)\mathbf{1}(\mathcal{H}).$$

Since we used Taylor's theorem for $x = O(\log(n)/n)$, the inequality holds for $\epsilon = n^{-1/2}$. Hence

$$\mathbb{E}\left[|R_* \cap V_1^+| \,\big|\, \boldsymbol{\sigma}_*^1\right]\mathbf{1}(\mathcal{H}) = \sum_{i \in V_1^+} \mathbb{P}\left(E_i \,\big|\, \boldsymbol{\sigma}_*^1\right)\mathbf{1}(\mathcal{H}) \geq n^{1-s^2(\alpha+\beta)/2-o(1)}\mathbf{1}(\mathcal{H}). \tag{35}$$

We next upper bound the variance of $|R_* \cap V_1^+|$. For distinct $i, j \in V_1^+$, we have that

$$\mathrm{Cov}\left(\mathbf{1}(E_i), \mathbf{1}(E_j) \,\big|\, \boldsymbol{\sigma}_*^1\right) = \mathbb{P}\left(E_i \cap E_j \,\big|\, \boldsymbol{\sigma}_*^1\right) - \mathbb{P}\left(E_i \,\big|\, \boldsymbol{\sigma}_*^1\right)\mathbb{P}\left(E_j \,\big|\, \boldsymbol{\sigma}_*^1\right)$$

$$= \left(1 - s^2\alpha\frac{\log n}{n}\right)^{2|V^+|-3}\left(1 - s^2\beta\frac{\log n}{n}\right)^{2|V^-|}\left(1 - \left(1 - s^2\alpha\frac{\log n}{n}\right)\right).$$

Notice that on the event $\mathcal{H}$, we have that $2|V^+| - 3 = (1 + o(1))n$ and $2|V^-| = (1 + o(1))n$. Hence, using the inequality $1 - x \leq e^{-x}$, we have that

$$\mathrm{Cov}\left(\mathbf{1}(E_i), \mathbf{1}(E_j) \,\big|\, \boldsymbol{\sigma}_*^1\right)\mathbf{1}(\mathcal{H}) \leq s^2\alpha\frac{\log n}{n}\exp\left(-(1-o(1))s^2(\alpha+\beta)\log n\right)\mathbf{1}(\mathcal{H})$$

$$= n^{-1-s^2(\alpha+\beta)+o(1)}\mathbf{1}(\mathcal{H}). \tag{36}$$

It follows that

$$\mathrm{Var}\left(|R_* \cap V_1^+| \,\big|\, \boldsymbol{\sigma}_*^1\right)\mathbf{1}(\mathcal{H}) = \mathrm{Var}\left(\sum_{i \in V_1^+} \mathbf{1}(E_i) \,\Bigg|\, \boldsymbol{\sigma}_*^1\right)\mathbf{1}(\mathcal{H})$$

$$= \sum_{i,j \in V_1^+} \mathrm{Cov}\left(\mathbf{1}(E_i), \mathbf{1}(E_j) \,\big|\, \boldsymbol{\sigma}_*^1\right)\mathbf{1}(\mathcal{H})$$

$$\leq \sum_{i \in V_1^+} \mathbb{P}\left(E_i \,\big|\, \boldsymbol{\sigma}_*^1\right)\mathbf{1}(\mathcal{H}) + \sum_{i,j \in V_1^+ : i \neq j} \mathrm{Cov}\left(\mathbf{1}(E_i), \mathbf{1}(E_j) \,\big|\, \boldsymbol{\sigma}_*^1\right)\mathbf{1}(\mathcal{H})$$

$$\leq \left(n^{1-s^2(\alpha+\beta)/2+o(1)} + n^{1-s^2(\alpha+\beta)+o(1)}\right)\mathbf{1}(\mathcal{H})$$

$$= n^{1-s^2(\alpha+\beta)/2+o(1)}\mathbf{1}(\mathcal{H}). \tag{37}$$

Above, we have used (33) and (36) to bound the terms of the summations in the second line. The bounds in (35) and (37), along with the Paley–Zygmund inequality, imply that

$$
\mathbb{P}\left(|R_* \cap V_1^+| \geq n^{1-s^2(\alpha+\beta)/2-\delta} \,\Big|\, \boldsymbol{\sigma}_*^1\right) \mathbf{1}(\mathcal{H})
$$

$$
\geq \left(1 - n^{-\delta+o(1)}\right)^2 \frac{\mathbb{E}\left[|R_* \cap V_1^+| \,\big|\, \boldsymbol{\sigma}_*^1\right]^2}{\mathbb{E}\left[|R_* \cap V_1^+|^2 \,\big|\, \boldsymbol{\sigma}_*^1\right]} \mathbf{1}(\mathcal{H})
$$

$$
\geq \left(1 - n^{-\delta+o(1)}\right)^2 \left(1 - \frac{\mathrm{Var}\left(|R_* \cap V_1^+| \,\big|\, \boldsymbol{\sigma}_*^1\right)}{\mathbb{E}\left[|R_* \cap V_1^+| \,\big|\, \boldsymbol{\sigma}_*^1\right]^2}\right) \mathbf{1}(\mathcal{H})
$$

$$
\geq \left(1 - n^{-\delta+o(1)}\right)^2 \left(1 - n^{-\left(1-s^2(\alpha+\beta)/2\right)+o(1)}\right) \mathbf{1}(\mathcal{H})
$$

$$
= (1 - o(1))\mathbf{1}(\mathcal{H}). \tag{38}
$$

Together, (34) and (38), along with $\mathbb{P}(\mathcal{H}) = 1 - o(1)$, show that

$$
\mathbb{P}\left(n^{1-s^2(\alpha+\beta)/2-\delta} \leq |R_* \cap V_1^+| \leq n^{1-s^2(\alpha+\beta)/2+\delta}\right) = 1 - o(1).
$$

Identical arguments show that the same holds when $|R_* \cap V_1^+|$ is replaced with $|R_* \cap V_1^-|$.

We now study the size of $|\overline{R}_* \cap V_1^+|$. Since $\overline{R}_* \supseteq R_*$, with probability $1 - o(1)$ we have the lower bound $|\overline{R}_* \cap V_1^+| \geq |R_* \cap V_1^+| \geq n^{1-s^2(\alpha+\beta)/2-\delta}$. To establish an upper bound, we use the bound

$$
|\overline{R}_* \cap V_1^+| \leq |\overline{R}_*| \leq \sum_{i \in R_*} (1 + |\mathcal{N}_2(\pi_*(i))|) \leq |\overline{R}_*| \left(1 + \max_{j \in [n]} |\mathcal{N}_2(j)|\right).
$$

In light of Lemma 28, we have that $\max_{j \in [n]} |\mathcal{N}_2(j)| \leq 100s \max\{\alpha, \beta\} \log n$ with probability at least $1 - o(1)$. Hence, with probability at least $1 - o(1)$, we have that

$$
|\overline{R}_* \cap V_1^+| \leq (1 + 100s \max\{\alpha, \beta\} \log n) |R_*|
$$

$$
\leq 2(1 + 100s \max\{\alpha, \beta\} \log n)n^{1-s^2(\alpha+\beta)/2+\delta} \leq n^{1-s^2(\alpha+\beta)/2+2\delta}.
$$

Identical steps show that $|\overline{R}_* \cap V_1^-| \leq n^{1-s^2(\alpha+\beta)/2+2\delta}$ with probability $1 - o(1)$ as well. ∎

## I.7. First and second moment estimates for $W_i$: Proofs of Lemmas 43 and 44

We first prove some useful intermediate results. Our first result establishes some useful conditional independence properties given the sigma algebra $\mathcal{I}$.

**Lemma 54** *The following hold:*

1. *The sets $\overline{R}_*$, $\overline{R}_* \cap V_1^+$, and $\overline{R}_* \cap V_1^-$ are $\mathcal{I}$-measurable.*

2. *Let $i \in R_*$. Conditioned on $\mathcal{I}$, $\{A_{i,j} : \{i,j\} \in \mathcal{E}_{10}\}$ is a collection of mutually independent random variables where*

$$
A_{i,j} \sim \begin{cases} \mathrm{Bern}\left(\alpha \frac{\log n}{n}\right) & \text{if } \sigma_*^1(i) = \sigma_*^1(j), \\ \mathrm{Bern}\left(\beta \frac{\log n}{n}\right) & \text{if } \sigma_*^1(i) = -\sigma_*^1(j). \end{cases} \tag{39}
$$

3. *The random variables* $\mathbf{1}(i \in S_*)$ *and* $\sum_{j \in [n] \setminus \overline{R}_*} A_{i,j} \sigma_*^1(j)$ *are conditionally independent given* $\mathcal{I}$.

**Proof** From the formula $\overline{R}_* = R_* \cup \pi_*^{-1}(\mathcal{N}_2(\pi_*(R_*)))$ provided in Definition 37, it is clear that $\overline{R}_*$ is $\mathcal{I}$-measurable since it depends only on $R_*$, $\pi_*$, and $B$. The sets $\overline{R}_* \cap V_1^+$ and $\overline{R}_* \cap V_1^-$ can be readily obtained from $\overline{R}_*$ and $\boldsymbol{\sigma}_*^1$.

We now prove Item #2. Notice that the sets $R_*$ and $\overline{R}_*$ depend only on $G_1 \wedge_{\pi_*} G_2$, $G_2$, and $\pi_*$. Thus, by Lemma 10, $G_1 \setminus_{\pi_*} G_2$ is conditionally independent of $R_*$ and $\overline{R}_*$ given $\pi_*$, $\boldsymbol{\sigma}_*$, and the partition $\{\mathcal{E}_{00}, \mathcal{E}_{01}, \mathcal{E}_{10}, \mathcal{E}_{11}\}$. In particular, the collection $\{A_{i,j} : \{i,j\} \in \mathcal{E}_{10}\}$ is conditionally independent of $\mathcal{I}$, and (39) follows.

Finally, we prove Item #3. By Definition 37, we can write

$$\mathbf{1}(i \in S_*) = \mathbf{1}(i \in R_*)\mathbf{1}(A_{i,j} = 0 \text{ for all } j \in \overline{R}_*).$$

In particular, $\mathbf{1}(i \in S_*)$ is measurable with respect to the sigma-algebra generated by $\mathcal{I}$ and the collection $\mathcal{C}_1 := \{A_{i,j} : j \in \overline{R}_* \text{ and } \{i,j\} \in \mathcal{E}_{10}\}$. On the other hand, since $\overline{R}_*$ is $\mathcal{I}$-measurable, $\sum_{j \in [n] \setminus \overline{R}_*} A_{i,j} \sigma_*^1(j)$ is measurable with respect to the sigma-algebra generated by $\mathcal{I}$ and the collection $\mathcal{C}_2 := \{A_{i,j} : j \in [n] \setminus \overline{R}_* \text{ and } \{i,j\} \in \mathcal{E}_{10}\}$. Since $\mathcal{C}_1 \cap \mathcal{C}_2 = \emptyset$, Item #2 implies that the two random variables are conditionally independent given $\mathcal{I}$. ∎

Our next result shows that, with high probability, vertices in $R_*$ are also in $S_*$.

**Lemma 55** *For any $\delta > 0$ and $i \in R_*$, it holds for sufficiently large $n$ that*

$$\mathbb{P}\left(i \in S_* \mid \mathcal{I}\right)\mathbf{1}(\mathcal{G}_\delta) \geq \left(1 - n^{-s^2 \mathrm{T_c}(\alpha,\beta) + 2\delta}\right)\mathbf{1}(\mathcal{G}_\delta).$$

**Proof** For $i \in R_*$, define the following $\mathcal{I}$-measurable random sets:

$$\mathcal{C}^+(i) := \left\{j \in \overline{R}_* : \{i,j\} \in \mathcal{E}_{10} \cap \mathcal{E}^+(\boldsymbol{\sigma}_*^1)\right\},$$
$$\mathcal{C}^-(i) := \left\{j \in \overline{R}_* : \{i,j\} \in \mathcal{E}_{10} \cap \mathcal{E}^-(\boldsymbol{\sigma}_*^1)\right\}.$$

By Definition 37, $i \in S_*$ if and only if $i \in R_*$ and $A_{i,j} = 0$ for all $j \in \overline{R}_*$. Phrased differently, $i \in S_*$ if and only if $i \in R_*$ and $A_{i,j} = 0$ for all $j \in \mathcal{C}^+(i) \cup \mathcal{C}^-(i)$. Item #2 of Lemma 54 then implies, for $i \in R_*$, that

$$\mathbb{P}\left(i \in S_* \mid \mathcal{I}\right) = \mathbb{P}\left(A_{i,j} = 0 \text{ for all } j \in \mathcal{C}^+(i) \cup \mathcal{C}^-(i) \mid \mathcal{I}\right)$$
$$= \left(1 - \alpha\frac{\log n}{n}\right)^{|\mathcal{C}^+(i)|}\left(1 - \beta\frac{\log n}{n}\right)^{|\mathcal{C}^-(i)|}$$
$$\geq \left(1 - \alpha|\mathcal{C}^+(i)|\frac{\log n}{n}\right)\left(1 - \beta|\mathcal{C}^-(i)|\frac{\log n}{n}\right) \geq 1 - (\alpha|\mathcal{C}^+(i)| + \beta|\mathcal{C}^-(i)|)\frac{\log n}{n}.$$

Above, the first inequality is due to Bernoulli's inequality. To further simplify the lower bound, note that $|\mathcal{C}^+(i)|, |\mathcal{C}^-(i)| \leq |\overline{R}_*|$, which is at most $2n^{1 - s^2 \mathrm{T_c}(\alpha,\beta) + \delta}$ on the event $\mathcal{G}_\delta$. Thus

$$\mathbb{P}\left(i \in S_* \mid \mathcal{I}\right)\mathbf{1}(\mathcal{G}_\delta) \geq \left(1 - 2(\alpha + \beta)(\log n)n^{-s^2 \mathrm{T_c}(\alpha,\beta) + \delta}\right)\mathbf{1}(\mathcal{G}_\delta).$$

Since $2(\alpha + \beta) \log n \le n^\delta$ for sufficiently large $n$, the desired result follows. ∎

Next, for $i \in R_*$, we define the random variable

$$
X_i := \begin{cases} \mathbb{P}\left(\displaystyle\sum_{j \in [n] \setminus \overline{R}_*} A_{i,j} \sigma_*^1(j) < 0 \,\middle|\, \mathcal{I}\right) & \text{if } i \in R_* \cap V_1^+, \\[4mm] \mathbb{P}\left(\displaystyle\sum_{j \in [n] \setminus \overline{R}_*} A_{i,j} \sigma_*^1(j) > 0 \,\middle|\, \mathcal{I}\right) & \text{if } i \in R_* \cap V_1^-. \end{cases}
$$

As we shall see in the proofs of Lemmas 43 and 44, we can bound the first and second moments of $\sum_{i \in R_* \cap V_1^+} W_i$ and $\sum_{i \in R_* \cap V_1^-} W_i$ by functions of the $X_i$'s. Our next result characterizes the behavior of the $X_i$'s on the high-probability $\mathcal{I}$-measurable event $\mathcal{F} \cap \mathcal{G}_\delta$.

**Lemma 56** *For $i \in R_*$, we have that*

$$
X_i \mathbf{1}(\mathcal{F} \cap \mathcal{G}_\delta) = n^{-s(1-s)\mathrm{D}_+(\alpha,\beta)+o(1)} \mathbf{1}(\mathcal{F} \cap \mathcal{G}_\delta).
$$

**Proof** Let $i \in R_* \cap V_1^+$. Since $\overline{R}_*$ and $\boldsymbol{\sigma}_*^1$ are $\mathcal{I}$-measurable, Item #2 of Lemma 54 implies that

$$
\sum_{j \in [n] \setminus \overline{R}_*} A_{i,j} \sigma_*^1(j) \stackrel{d}{=} Y - Z, \tag{40}
$$

where $Y \sim \mathrm{Bin}(m^+, p)$ and $Z \sim \mathrm{Bin}(m^-, q)$ are independent, with $p = \alpha \log(n)/n$, $q = \beta \log(n)/n$, and

$$
m^+ := \left|\left\{j \in [n] \setminus \overline{R}_* : \{i,j\} \in \mathcal{E}_{10} \cap \mathcal{E}^+(\boldsymbol{\sigma}_*^1)\right\}\right|,
$$
$$
m^- := \left|\left\{j \in [n] \setminus \overline{R}_* : \{i,j\} \in \mathcal{E}_{10} \cap \mathcal{E}^-(\boldsymbol{\sigma}_*^1)\right\}\right|.
$$

We use the shorthand $s_{10} := s(1-s)$. On the event $\mathcal{F} \cap \mathcal{G}_\delta$, we have the upper bound

$$
m^+ \le \left|\left\{j : \{i,j\} \in \mathcal{E}_{10} \cap \mathcal{E}^+(\boldsymbol{\sigma}_*^1)\right\}\right| \le s_{10}\left(\frac{n}{2} + 2n^{3/4}\right) = (1 + o(1))s_{10}\frac{n}{2}
$$

and the matching lower bound

$$
\begin{aligned}
m^+ &\ge \left|\left\{j : \{i,j\} \in \mathcal{E}_{10} \cap \mathcal{E}^+(\boldsymbol{\sigma}_*^1)\right\}\right| - |\overline{R}_*| \\
&\ge s_{10}\left(\frac{n}{2} - 2n^{3/4}\right) - n^{1-s^2\mathrm{T_c}(\alpha,\beta)+\delta} = (1 - o(1))s_{10}\frac{n}{2}.
\end{aligned}
$$

Similarly, we can show that on the event $\mathcal{F} \cap \mathcal{G}_\delta$, we have that

$$
(1 - o(1))s_{10}\frac{n}{2} \le m^- \le (1 + o(1))s_{10}\frac{n}{2}.
$$

Noting that $p = (1 + o(1))s_{10}\alpha \log(s_{10}n)/(s_{10}n)$ and $q = (1 + o(1))s_{10}\beta \log(s_{10}n)/(s_{10}n)$, Lemma 6 and (40) now imply that for $i \in R_* \cap V_1^+$, we have that

$$
\mathbb{P}\left(\sum_{j \in [n] \setminus \overline{T}_*} A_{i,j} \sigma_*^1(j) < 0 \,\middle|\, \mathcal{I}\right) \mathbf{1}(\mathcal{F} \cap \mathcal{G}_\delta)
$$
$$
= \mathbb{P}\left(Y < Z \,\middle|\, \mathcal{I}\right) \mathbf{1}(\mathcal{F} \cap \mathcal{G}_\delta) = n^{-s_{10}\mathrm{D}_+(\alpha,\beta)+o(1)} \mathbf{1}(\mathcal{F} \cap \mathcal{G}_\delta). \tag{41}
$$

57

An identical analysis can be done for $i \in R_* \cap V_1^-$. ∎

We now turn to the proofs of the moment estimates.

**Proof** [Proof of Lemma 43] As a shorthand, denote $\gamma := s^2 \mathrm{T_c}(\alpha, \beta) - 2\delta$. Let $i \in R_* \cap V_1^+$. Then,

$$
\mathbb{E}\left[W_i \,\middle|\, \mathcal{I}\right] \mathbf{1}(\mathcal{F} \cap \mathcal{G}_\delta) = \mathbb{P}\left(i \in S_* \text{ and } \sum_{j \in [n] \setminus \overline{R}_*} A_{i,j} \sigma_*^1(j) < 0 \,\middle|\, \mathcal{I}\right) \mathbf{1}(\mathcal{F} \cap \mathcal{G}_\delta)
$$

$$
= \mathbb{P}\left(i \in S_* \,\middle|\, \mathcal{I}\right) \cdot \mathbb{P}\left(\sum_{j \in [n] \setminus \overline{R}_*} A_{i,j} \sigma_*^1(j) < 0 \,\middle|\, \mathcal{I}\right) \mathbf{1}(\mathcal{F} \cap \mathcal{G}_\delta),
$$

where the first equality follows since $\mathsf{maj}(i) = \sum_{j \in [n]} A_{i,j} \sigma_*^1(j)$ and $A_{i,j} = 0$ when $i \in S_*$ and $j \in \overline{R}_*$; the second equality is due to Item #3 of Lemma 54. Now using Lemma 55, the definition of $X_i$, and Lemma 56, we have that

$$
\mathbb{E}\left[W_i \,\middle|\, \mathcal{I}\right] \mathbf{1}(\mathcal{F} \cap \mathcal{G}_\delta) \geq (1 - n^{-\gamma}) X_i \mathbf{1}(\mathcal{F} \cap \mathcal{G}_\delta) \tag{42}
$$

$$
\geq (1 - n^{-\gamma}) n^{-s(1-s)\mathrm{D}_+(\alpha,\beta)+o(1)} \mathbf{1}(\mathcal{F} \cap \mathcal{G}_\delta). \tag{43}
$$

Summing over $i \in R_* \cap V_1^+$, we obtain that

$$
\mathbb{E}\left[\sum_{i \in R_* \cap V_1^+} W_i \,\middle|\, \mathcal{I}\right] \mathbf{1}(\mathcal{F} \cap \mathcal{G}_\delta) \geq (1 - n^{-\gamma}) \left(|R_* \cap V_1^+| n^{-s(1-s)\mathrm{D}_+(\alpha,\beta)+o(1)}\right) \mathbf{1}(\mathcal{F} \cap \mathcal{G}_\delta)
$$

$$
\geq (1 - n^{-\gamma}) n^{1 - s^2 \mathrm{T_c}(\alpha,\beta) - s(1-s)\mathrm{D}_+(\alpha,\beta) - \delta + o(1)} \mathbf{1}(\mathcal{F} \cap \mathcal{G}_\delta),
$$

where the inequality on the second line uses the fact that $|R_* \cap V_1^+| \geq n^{1 - s^2 \mathrm{T_c}(\alpha,\beta) - \delta}$ on $\mathcal{G}_\delta$. Identical arguments yield the same lower bound for $\mathbb{E}\left[\sum_{i \in R_* \cap V_1^-} W_i \,\middle|\, \mathcal{I}\right] \mathbf{1}(\mathcal{F} \cap \mathcal{G}_\delta)$. ∎

**Proof** [Proof of Lemma 44] Fix $i, j \in R_* \cap V_1^+$, where $i \neq j$. We have that

$$
\mathbb{E}\left[W_i W_j \,\middle|\, \mathcal{I}\right] = \mathbb{P}\left(i, j \in S_* \text{ and } \sum_{k \in [n] \setminus \overline{R}_*} A_{i,k} \sigma_*^1(k), \sum_{k \in [n] \setminus \overline{R}_*} A_{j,k} \sigma_*^1(k) < 0 \,\middle|\, \mathcal{I}\right)
$$

$$
\leq \mathbb{P}\left(\sum_{k \in [n] \setminus \overline{R}_*} A_{i,k} \sigma_*^1(k), \sum_{k \in [n] \setminus \overline{R}_*} A_{j,k} \sigma_*^1(k) < 0 \,\middle|\, \mathcal{I}\right)
$$

$$
= \mathbb{P}\left(\sum_{k \in [n] \setminus \overline{R}_*} A_{i,k} \sigma_*^1(k) < 0 \,\middle|\, \mathcal{I}\right) \mathbb{P}\left(\sum_{k \in [n] \setminus \overline{R}_*} A_{j,k} \sigma_*^1(k) < 0 \,\middle|\, \mathcal{I}\right)
$$

$$
= X_i X_j.
$$

Above, the equality on the first line follows since $A_{i,k} = A_{j,k} = 0$ for $i, j \in S_*$ and $k \in \overline{R}_*$. The equality on the third line follows from noting that $\sum_{k \in [n] \setminus \overline{R}_*} A_{i,k} \sigma_*^1(k)$ is a measurable function

of $\mathcal{I}$ and $\{A_{i,k}\}_{k\in[n]\setminus\overline{R}_*}$, $\sum_{k\in[n]\setminus\overline{R}_*} A_{j,k}\sigma_*^1(k)$ is a measurable function of $\mathcal{I}$ and $\{A_{j,k}\}_{k\in[n]\setminus\overline{R}_*}$, and the collections $\{A_{i,k}\}_{k\in[n]\setminus\overline{R}_*}$ and $\{A_{j,k}\}_{k\in[n]\setminus\overline{R}_*}$ are conditionally independent given $\mathcal{I}$ (by Item #2 of Lemma 54). On the other hand, for the case $i = j$ we have through similar arguments that

$$\mathbb{E}\left[W_i^2\,\middle|\,\mathcal{I}\right] = \mathbb{E}\left[W_i\,\middle|\,\mathcal{I}\right] = \mathbb{P}\left(i \in S_* \text{ and } \sum_{k\in[n]\setminus\overline{R}_*} A_{i,k}\sigma_*^1(k) < 0\,\middle|\,\mathcal{I}\right) \le X_i.$$

We can then write

$$\mathbb{E}\left[\left(\sum_{i\in R_*\cap V_1^+} W_i\right)^2\,\middle|\,\mathcal{I}\right] = \sum_{i\in R_*\cap V_1^+} \mathbb{E}\left[W_i\,\middle|\,\mathcal{I}\right] + \sum_{i,j\in R_*\cap V_1^+:i\neq j} \mathbb{E}\left[W_iW_j\,\middle|\,\mathcal{I}\right]$$

$$\le \sum_{i\in R_*\cap V_1^+} X_i + \sum_{i,j\in R_*\cap V_1^+:i\neq j} X_iX_j$$

$$\le \sum_{i\in R_*\cap V_1^+} X_i + \left(\sum_{i\in R_*\cap V_1^+} X_i\right)^2. \tag{44}$$

In light of the lower bound in (42), we also have that

$$\mathbb{E}\left[\sum_{i\in R_*\cap V_1^+} W_i\,\middle|\,\mathcal{I}\right]^2 \mathbf{1}(\mathcal{F}\cap\mathcal{G}_\delta) \ge \left(1 - n^{-\gamma}\right)^2 \left(\sum_{i\in R_*\cap V_1^+} X_i\right)^2 \mathbf{1}(\mathcal{F}\cap\mathcal{G}_\delta), \tag{45}$$

where we recall that $\gamma := s^2 \mathrm{T_c}(\alpha,\beta) - 2\delta$. On the event $\mathcal{F}\cap\mathcal{G}_\delta$, we also have the upper bound

$$\left(\sum_{i\in R_*\cap V_1^+} X_i\right)\mathbf{1}(\mathcal{F}\cap\mathcal{G}_\delta) = |R_*\cap V_1^+|n^{-s(1-s)\mathrm{D}_+(\alpha,\beta)+o(1)}\mathbf{1}(\mathcal{F}\cap\mathcal{G}_\delta) \le n^{\theta+\delta+o(1)}\mathbf{1}(\mathcal{F}\cap\mathcal{G}_\delta),$$

$$\tag{46}$$

where the first equality is due to Lemma 56 and we recall that $\theta := 1 - s^2\mathrm{T_c}(\alpha,\beta) - s(1-s)\mathrm{D}_+(\alpha,\beta)$. Combining (44) and (45) and using $(1-n^{-\gamma})^2 \ge 1 - 2n^{-\gamma}$ shows that

$$\mathrm{Var}\left(\sum_{i\in R_*\cap V_1^+} W_i\,\middle|\,\mathcal{I}\right)\mathbf{1}(\mathcal{F}\cap\mathcal{G}_\delta) \le \left(\sum_{i\in R_*\cap V_1^+} X_i + 2n^{-\gamma}\left(\sum_{i\in R_*\cap V_1^+} X_i\right)^2\right)\mathbf{1}(\mathcal{F}\cap\mathcal{G}_\delta)$$

$$\le \left(n^{\theta+\delta+o(1)} + 2n^{2\theta+2\delta-\gamma+o(1)}\right)\mathbf{1}(\mathcal{F}\cap\mathcal{G}_\delta).$$

To simplify the bound on the right hand side further, notice that if $\delta < \theta/4$ and $\delta \le \gamma/6$ (i.e., $\delta \le s^2\mathrm{T_c}(\alpha,\beta)/8$), then we have $\theta + \delta + o(1) < 2\theta - 3\delta$ and $2\theta + 2\delta - \gamma + o(1) < 2\theta - 3\delta$ for $n$ large enough. Hence

$$\mathrm{Var}\left(\sum_{i\in R_*\cap V_1^+} W_i\,\middle|\,\mathcal{I}\right)\mathbf{1}(\mathcal{F}\cap\mathcal{G}_\delta) \le n^{2\theta-3\delta}\mathbf{1}(\mathcal{F}\cap\mathcal{G}_\delta).$$

Through identical steps, the same result holds for $\sum_{i\in R_*\cap V_1^-} W_i$. $\blacksquare$