Differentially Private Community Detection for Stochastic Block Models

Mohamed Seif [*1] **Dung Nguyen** [*2 3] **Anil Vullikanti** [2 3] **Ravi Tandon** [1]

Abstract

The goal of community detection over graphs is to recover underlying labels/attributes of users (e.g., political affiliation) given the connectivity between users. There has been significant recent progress on understanding the fundamental limits of community detection when the graph is generated from a stochastic block model (SBM). Specifically, sharp information theoretic limits and efficient algorithms have been obtained for SBMs as a function of p and q, which represent the intra-community and inter-community connection probabilities. In this paper, we study the community detection problem while preserving the privacy of the individual connections between the vertices. Focusing on the notion of (ϵ, δ)-edge differential privacy (DP), we seek to understand the fundamental tradeoffs between (p, q), DP budget (ϵ, δ), and computational efficiency for exact recovery of community labels.

To this end, we present and analyze the associated information-theoretic tradeoffs for three differentially private community recovery mechanisms: a) stability based mechanism; b) sampling based mechanisms; and c) graph perturbation mechanisms. Our main findings are that stability and sampling based mechanisms lead to a superior tradeoff between (p, q) and the privacy budget (ϵ, δ); however this comes at the expense of higher computational complexity. On the other hand, albeit low complexity, graph perturbation mechanisms require the privacy budget ϵ to scale as $\Omega(\log(n))$ for exact recovery.

[*]Equal contribution [1]Department of Electrical and Computer Engineering, University of Arizona. [2]Biocomplexity Institute and Initiative, University of Virginia. [3]Department of Computer Science, University of Virginia. Correspondence to: Mohamed Seif <mseif@email.arizona.edu>, Dung Nguyen <dungn@virginia.edu>, Ravi Tandon <tandonr@email.arizona.edu>.

Proceedings of the 39^{th} International Conference on Machine Learning, Baltimore, Maryland, USA, PMLR 162, 2022. Copyright 2022 by the author(s).

1. Introduction

Community detection in networks is a fundamental problem in the area of graph mining and machine learning, with many interesting applications such as social networks, image segmentation, and biological networks (see, e.g., the survey by (Fortunato, 2010)). The main goal is to partition the network into communities that are "well-connected"; no standard definition for communities exists, and a large number of methods have been proposed, e.g., (Blondel et al., 2008; Girvan & Newman, 2002; Holland et al., 1983), but, in general, there is a limited theoretical basis for the performance of these methods. One exception is the stochastic block model (SBM) (Holland et al., 1983), which is a probabilistic generative model for generating networks with underlying communities, providing a rigorous framework for detection algorithms. In the simplest canonical form of an SBM, the n vertices are partitioned into r communities, and a pair of vertices connect with probability p within communities and with probability q across communities, where $p > q$. "Recovering" communities in a graph generated from an SBM (defined formally in Section 2) has been a very active area of research, e.g., (Condon & Karp, 2001; Arias-Castro & Verzelen, 2014; Abbe et al., 2015; Hajek et al., 2016a). The exact conditions for recoverability are well understood in terms of the scaling of p and q (more specifically the difference between p and q). In particular, in the dense regime (the focus of this paper), with $p = a\log(n)/n$ and $q = b\log(n)/n$, for some constants $a > b > 0$, it is known that exact recovery is possible *if and only if* $\sqrt{a} - \sqrt{b} > \sqrt{r}$ (see (Abbe, 2017) for a comprehensive survey). Efficient algorithms for recovering communities have been developed using spectral methods and semi-definite programming (SDP) (Boppana, 1987; McSherry, 2001; Abbe et al., 2015; Massoulié, 2014; Gao et al., 2017; Hajek et al., 2016a; Abbe et al., 2020; Wang et al., 2020).

In many applications, e.g., healthcare, social networks, and finance, network data is often private and sensitive, and there is a risk of revealing private information through adversarial queries. Differential Privacy (DP) (Dwork et al., 2014) is the *de facto* standard notion for providing rigorous privacy guarantees. DP ensures that each user's presence in the dataset has minimal statistical influence (measured by the privacy budget ϵ) on the output of queries. Within the con-

	MLE-Stability	SDP-Stability	Bayesian	Exponential	RR + SDP
ϵ	$\mathcal{O}(1)$	$\mathcal{O}(1)$	$\Omega(\log(a/b))$	$\mathcal{O}(1)$	$\Omega(\log(n))$
δ	$1/n^2$	$1/n^2$	0	0	0
$\sqrt{a}-\sqrt{b}\geq$	$\sqrt{2}\cdot\sqrt{1+3/2\epsilon}$	$4\sqrt{2}\cdot(1+\sqrt{3}/\sqrt{2\epsilon})$	Theorem 3.6	Theorem 3.7	Theorem 3.8
Time complexity	$\mathcal{O}(\exp(n))$	$n^{(\mathcal{O}(\log(n)))}$	$\mathcal{O}(\exp(n))$	$\mathcal{O}(\exp(n))$	$\mathcal{O}(\text{poly}(n))$

Table 1. Summary of the recovery threshold(s), complexity and (ϵ, δ)-edge DP for $r = 2$ communities.

text of network/graph data, two privacy models have been considered— edge and node privacy, and DP algorithms have also been developed for a few network problems, e.g., the number of subgraphs, such as stars and triangles, cuts, dense subgraphs, and communities, and releasing synthetic graphs (Kasiviswanathan et al., 2013; Blocki et al., 2013; Mülle et al., 2015; Nguyen et al., 2016; Qin et al., 2017; Imola et al., 2021; Blocki et al., 2013); most of them focus on edge privacy models, especially when the output is not a count. Finally, there has been very little work on community detection with privacy. (Nguyen et al., 2016) consider communities based on the modularity. Very recently, (Hehir et al., 2021; Ji et al., 2019) consider community detection in the SBM models subject to edge privacy constraints (also see related work Section 1.2); however, neither provides any rigorous bounds on the accuracy or the impact of edge privacy on the recovery threshold.

1.1. Contributions

In this paper, *we present the first differentially private algorithms for community detection in SBMs with rigorous bounds on recoverability*, under the edge privacy model. Informally, a community recovery algorithm satisfies edge privacy if the output has similar distribution irrespective of the presence or absence of an edge between any two vertices in the network (see Definition 2.2). Edge DP is the most natural privacy notion for community detection, as it involves outputting the partition of the nodes into communities. Our focus is on characterizing the recoverability threshold under edge DP, i.e., how much does the difference between p and q have to change in order to ensure recoverability with privacy. We analyze three classes of mechanisms for this problem.

1. Stability based mechanisms. We show that the stability mechanism (Thakurta & Smith, 2013) gives (ϵ, δ)-DP algorithms for our problem. The main idea is to determine if a non-private community recovery estimator is stable with respect to graph G, i.e., the estimate of community structure does not change if a few edges are perturbed; if the estimator is stable, the non-private estimate of community labels can be released; otherwise, we release a random label. We analyze stability based mechanism for two estimators—the maximum likelihood estimator (MLE), which involves solving a min-bisection problem, and an SDP based estimator. We also derive sufficient conditions for exact recovery for $r = 2$ and $r > 2$ communities for both these types of algorithms—these require a slightly larger separation between p and q as a function of the privacy budget ϵ; further, the threshold converges to the well known non-private bound as ϵ becomes large. The SDP based stability mechanism can be implemented in quasi-polynomial time.

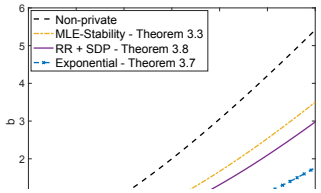
Stability based mechanisms are less common in the DP literature, compared to other mechanisms, e.g., exponential or randomized response, since proving stability turns out to be very challenging, in general, and is one of our important technical contributions. Stability of the MLE scheme requires showing that the optimum bisection does not change when $k = \mathcal{O}(\log n)$ edges are perturbed, with high probability. This becomes even harder for the SDP based algorithm, which doesn't always produce an optimum solution. (Hajek et al., 2016a) construct a "certificate" for proving optimality of the SDP solution, with high probability. A technical contribution is to identify a new condition that makes the certificate *deterministic*—this is crucial in our stability analysis.

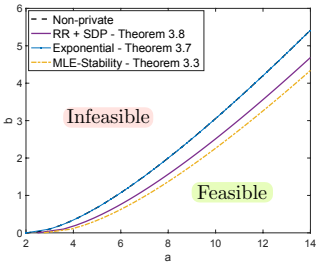
2. Sampling based mechanisms. In the second approach, we design two different sampling based mechanisms: (1) Bayesian Estimation and (2) Exponential mechanism. We show that these algorithms are differentially private (with constant ϵ for Bayesian Estimation and arbitrary small ϵ for the Exponential mechanism) and guarantee exact recovery under certain regimes of ϵ, a, b; note that, in contrast to the stability based mechanisms, we have $\delta = 0$.

3. Randomized Response (RR) based mechanism. We also study and analyze a baseline approach, in which one can use a randomized response (RR) technique to perturb the adjacency matrix, and subsequently run an SDP based algorithm for community recovery on the perturbed graph. Due to the post-processing properties of DP, this mechanism satisfies ϵ-DP for any $\epsilon > 0$. We show that in contrast to stability and sampling based methods, the baseline RR approach requires $\epsilon = \Omega(\log(n))$ for exact recovery.

4. Empirical evaluation. We also present simulation results on both synthetic and real-world graphs to validate our theoretical findings (Section 4). We observe that the stability based mechanism generally outperforms the others in terms of the error, which is quite small even for fairly small ϵ. Interestingly, the error is low even in real world networks.

Comparison between different mechanisms. We summarize our theoretical results for differentially private community recovery in Table 1, which shows the tradeoffs between





(b) $\epsilon = 4$. (*low* privacy regime)

Figure 1. Exact Recovery Threshold as a function of (a, b), and the privacy budget ϵ for $r = 2$ communities.

(a, b), (ϵ, δ) as well as the computational complexity of the mechanisms for $r = 2$ communities. Note that none of the mechanisms is redundant— each is the best in some part of the complex space consisting of the parameters a, b, ϵ, δ, and the running time. To further illustrate these tradeoffs, we plot the recovery threshold conditions for these mechanisms in Fig. 1. From Fig. 1(a), we observe that for the high privacy regime (smaller ϵ), MLE based Stability mechanism requires the least separation between a and b compared to all other algorithms. In the low privacy regime (larger ϵ), as shown in Fig. 1(b), we can see that exponential mechanism tends to overlap with the non-private recovery threshold (Abbe et al., 2015), whereas stability-based and RR based mechanisms require more separation between a and b. Complete proofs are presented in the Appendix.

1.2. Related Work

We first summarize a few of the main results on the complexity of different recoverability algorithms and then discuss some relevant work on SBMs with DP. The seminal work of (Abbe et al., 2015) showed that the optimal reconstruction of graph partitions is achieved by the maximum likelihood (ML) estimator, which is computationally intractable. (Boppana, 1987; McSherry, 2001) designed polynomial time algorithms for exact recovery; however, they did not achieve the optimal information theoretic bound, i.e., $\sqrt{a} - \sqrt{b} > \sqrt{r}$. (Abbe et al., 2015) showed the first computationally efficient algorithm that achieves the information theoretic limit. This algorithm has two phases: the first phase performs partial recovery via the algorithm of (Massoulié, 2014). The second phase uses a local improvement to refine the recovery. (Hajek et al., 2016a) showed that an SDP based rounding algorithm achieves the optimal recovery threshold in polynomial time, and settled the conjecture of (Abbe et al., 2015). Recently, there have been different computationally efficient recovery algorithms (Gao et al., 2017; Hajek et al., 2016a; Abbe et al., 2020; Wang et al., 2020) proposed that achieve the optimal recovery threshold in polynomial time or quasi-linear time for different settings, e.g., multiple communities with different sizes.

As mentioned earlier, there has been little work on community detection with differential privacy. (Nguyen et al., 2016) consider the problem of finding communities by modularity maximization. (Qin et al., 2017) design heuristics for models which are related to SBM. Other related work is on estimating parameters of graphons, which are generalizations of SBMs. (Borgs et al., 2015) developed an exponential time algorithm for estimating properties in the node DP model, and derived optimal information theoretic error bounds. (Sealfon & Ullman, 2019) improved this and designed a polynomial time algorithm. (Hehir et al., 2021) study the problem of privacy-preserving community detection on SBMs using a simple spectral method (Lei & Rinaldo, 2015) for multiple communities. They generalized the convergence rate analysis of the spectral algorithm and showed the impact of the privacy parameters on the misclassification rate between the ground truth labels and the estimated labels for the algorithm. (Ji et al., 2019) propose a DP gradient based community detection algorithm. However, neither of these results analyze the thresholds for recoverability, which has remained an open problem (under edge DP constraints) till now.

2. Problem Statement & Preliminaries

We consider an undirected graph $G = (\mathcal{V}, E)$ consisting of n vertices (vertices), where vertices are divided into r communities with $\frac{n}{r}$ vertices in each community. The community label for vertex i is denoted by $\sigma_i^* \in \{1, 2, \cdots, r\}, \forall i \in [n]$. We focus on the setting when the graph G is generated through a Stochastic block model (SBM), where the edges within the classes are generated independently with probability p and the edges between the classes are generated independently with probability q. The connections between vertices are represented by an adjacency matrix $\mathbf{A} \in \{0, 1\}^{n \times n}$, where the elements in $\mathbf{A}$ are drawn as:

$$A_{i,j} \sim \begin{cases} \text{Bern}(p), & i < j, \ \ \sigma_i = \sigma_j, \\ \text{Bern}(q), & i < j, \ \ \sigma_i \neq \sigma_j. \end{cases} \tag{1}$$

with $A_{i,i} = 0$ and $A_{i,j} = A_{j,i}$. For the scope of this paper, we focus on the so called "dense" connectivity regime, where $p = \frac{a \log(n)}{n}$ and $q = \frac{b \log(n)}{n}$, and $a, b \geq 0$ are fixed constants. Note that one can consider other regimes for p and q such as the "sparse" regime (Decelle et al., 2011), i.e., $p = \frac{a}{n}$ and $q = \frac{b}{n}$, however, in this regime

exact recovery is not possible since the graph with high probability is not connected. On the other hand, in the dense regime one can still exactly recover the labels of the graph with high probability. The goal of community detection problem is to design a (stochastic) estimator $\hat{\boldsymbol{\sigma}} : \mathbf{A} \to \{1, 2, \cdots, r\}^n$ for community recovery (i.e, the true label vector $\boldsymbol{\sigma}^* = \{\sigma_1^*, \sigma_2^*, \cdots, \sigma_n^*\}$) upon observing the adjacency matrix. We next define the notion of exact asymptotic recovery as a measure of performance of an estimator.

Definition 2.1 (Exact Recovery)**.** *An estimator $\hat{\boldsymbol{\sigma}} = \{\hat{\sigma}_1, \hat{\sigma}_2, \cdots, \hat{\sigma}_n\}$ satisfies exact recovery (upto a global permutation of the community labels) if the probability of error behaves as*

$$\Pr(\hat{\boldsymbol{\sigma}} \neq \boldsymbol{\sigma}^*) = o(1), \tag{2}$$

where the probability is taken over both the randomness of the graph G as well as the stochastic estimation process.

In addition to exact recovery, we require that the recovery algorithm for community detection also protects the individual relationships (i.e., the edges in the graph G) in the network. Specifically, we adopt the notion of (ϵ, δ)-edge differential privacy (DP) (Karwa et al., 2011), defined next.

Definition 2.2 ((ϵ, δ)-edge DP)**.** *An estimator $\hat{\boldsymbol{\sigma}}$ satisfies (ϵ, δ)-edge DP for some $\epsilon \in \mathbb{R}^+$ and $\delta \in (0, 1]$, if for any pair of adjacency matrices $\mathbf{A}$ and $\mathbf{A}'$ that differ in one edge, we have*

$$\Pr(\hat{\boldsymbol{\sigma}}(\mathbf{A}) = \boldsymbol{\sigma}) \leq e^{\epsilon} \Pr(\hat{\boldsymbol{\sigma}}(\mathbf{A}') = \boldsymbol{\sigma}) + \delta. \tag{3}$$

For privacy constraints in (3), the probabilities are computed only over the randomness in the estimation process. The case of $\delta = 0$ is called pure ϵ-edge DP.

2.1. Prior results on exact recovery without privacy

The optimal maximum likelihood (ML) estimator for community detection, given by $\hat{\boldsymbol{\sigma}}_{\text{ML}} = \arg\max_{\boldsymbol{\sigma}} p(\mathbf{A}|\boldsymbol{\sigma})$ has been recently analyzed in a series of papers (Boppana, 1987; McSherry, 2001; Choi et al., 2012; Abbe et al., 2015; Mossel et al., 2015). It has been shown that for SBMs with "dense" regime, i.e., $p = \frac{a\log(n)}{n}$ and $q = \frac{b\log(n)}{n}$, exact recovery is possible if and only if $\sqrt{a} - \sqrt{b} > \sqrt{r}$ (often referred to as the phase transition boundary or exact recovery threshold). Even for $r = 2$ communities, the ML estimator is equivalent to finding the minimum bisection of the graph, which is known to be NP-hard (Abbe et al., 2015). Specifically, the ML estimator of $\boldsymbol{\sigma}^*$ is the solution of the following optimization problem:

$$\hat{\boldsymbol{\sigma}}_{\text{ML}} = \arg\max_{\boldsymbol{\sigma}} \{\boldsymbol{\sigma}^T \mathbf{A} \boldsymbol{\sigma} : \mathbf{1}^T \boldsymbol{\sigma} = 0, \sigma_i = \pm 1\}. \tag{4}$$

Subsequently, several works have studied if polynomial time algorithms can still achieve the exact recovery threshold. For instance, it has been shown (Hajek et al., 2016a), (Hajek et al., 2016b) that SDP relaxation of the ML estimator can also achieve the same recovery threshold. Recently, Abbe et.al. (Abbe et al., 2020) have analyzed the spectral clustering estimator (Lei & Rinaldo, 2015), and showed that it achieves the same recovery threshold as ML for $r = 2$.

3. Main Results & Discussions

In this section, we present three different approaches for the design of community detection algorithms for exact recovery while satisfying edge differential privacy. In the first approach, we analyze the stability property of ML based and SDP based algorithms. For MLE based algorithm, the stability property of the min-bisection hinges on the concentration properties of SBMs in terms of the intra and inter communities edges. For SDP based algorithm, we introduce a concept of concentration that both (1) provides sufficient conditions for the dual certificate of the SDP and (2) persists under certain degrees of connection perturbation. In the second approach, we study and analyze sampling based mechanisms, which release a differentially private estimate of the community labels via sampling. In the third approach, we perturb the adjacency matrix $\mathbf{A}$ to satisfy DP (using randomized response (RR)), and perform the estimation of community labels using the perturbed graph by using computationally efficient SDP relaxation of the maximum-likelihood estimator. In Table 1, we summarize our main results for the case of $r = 2$ communities, where we show the constraints on the privacy budget (ϵ, δ) and sufficient conditions on (a, b) for exact recovery.

3.1. Stability-based Mechanisms

The basic idea behind stability based mechanisms is as follows: Let us consider a non-private estimator for community detection $\hat{\boldsymbol{\sigma}}$. We first *privately* compute the stability of this estimator with respect to a graph G, which essentially is the *minimum* number of edge modifications on G, so that the estimator output on the modified graph G' differs from that on G, i.e., $\hat{\boldsymbol{\sigma}}(G) \neq \hat{\boldsymbol{\sigma}}(G')$. If the graph G is stable enough (i.e., if the estimate of stability is larger than a threshold, which depends on (ϵ, δ)), then we release the non-private estimate $\hat{\boldsymbol{\sigma}}(G)$, otherwise we release a random label vector. The key intuition is that from the output of a stable estimator, one cannot precisely infer the presence or absence of a single edge (thereby providing edge DP guarantee). Before presenting the general stability mechanism, we formally define $d_{\hat{\sigma}}(G)$, which quantifies the stability of an estimator $\hat{\boldsymbol{\sigma}}$ with respect to a graph G.

Definition 3.1 (Stability of $\hat{\boldsymbol{\sigma}}$)**.** *The stability of an estimator $\hat{\boldsymbol{\sigma}}$ with respect to a graph G is defined as follows:*

$$d_{\hat{\boldsymbol{\sigma}}}(G) = \{\min k : \exists G', \mathit{dist}(G, G') \leq k, \hat{\boldsymbol{\sigma}}(G) \neq \hat{\boldsymbol{\sigma}}(G')\}. \tag{5}$$

We now present the general stability based mechanism in

Algorithm 1. We first state the following claim about the

Algorithm 1 $\mathcal{M}^{\hat{\boldsymbol{\sigma}}}_{\text{Stability}}(G)$: Stability Based Mechanism

1: **Input:** $G(\mathcal{V}, E) \in \mathcal{G}$
2: **Output:** labelling vector $\hat{\boldsymbol{\sigma}}_{\text{Private}}$.
3: $d_{\hat{\boldsymbol{\sigma}}}(G) \leftarrow$ stability of $\hat{\boldsymbol{\sigma}}$ with respect to graph G
4: $\tilde{d} \leftarrow d_{\hat{\boldsymbol{\sigma}}}(G) + \text{Lap}(1/\epsilon)$
5: **if** $\tilde{d} > \frac{\log 1/\delta}{\epsilon}$ **then**
6: Output $\hat{\boldsymbol{\sigma}}(G)$
7: **else**
8: Output $\perp$ (random label)

privacy guarantee of the above mechanism (Dwork et al., 2014).

Lemma 3.2. *For any community detection algorithm $\hat{\boldsymbol{\sigma}}$, $\mathcal{M}^{\hat{\boldsymbol{\sigma}}}_{\text{Stability}}(G)$ satisfies (ϵ, δ)-edge DP.*

In the above algorithm, Step 4 ensures that the stability is computed privately and Step 5 ensures that the non-private estimate is released only if the estimator is stable enough (i.e., $\tilde{d} > \frac{\log 1/\delta}{\epsilon}$).

Our first main contribution is to analyze the performance of $\mathcal{M}^{\hat{\boldsymbol{\sigma}}}_{\text{Stability}}(G)$ and establish sharp phase transition thresholds for exact recovery as a function of (p, q) and (ϵ, δ). Specifically, we focus on two possible choices for $\hat{\boldsymbol{\sigma}}$: a) when we use the MLE estimator, i.e., $\hat{\boldsymbol{\sigma}} = \hat{\boldsymbol{\sigma}}_{\text{MLE}}$, and b) when we use the computationally efficient SDP relaxation, i.e., $\hat{\boldsymbol{\sigma}} = \hat{\boldsymbol{\sigma}}_{\text{SDP}}$.

Stability of MLE. We start by first presenting the results for MLE based approach for both $r = 2$ communities and then for $r > 2$ communities.

Theorem 3.3. *For $r = 2$ communities, $\mathcal{M}^{MLE}_{\text{Stability}}(G)$ satisfies exact recovery if*

$$\sqrt{a} - \sqrt{b} > \sqrt{2} \times \sqrt{1 + \frac{t+1}{2\epsilon}} \tag{6}$$

for any $\epsilon > 0$ and $\delta = n^{-t}$, $t > 0$.

We note two important points: *(1) In contrast to the non-private recovery threshold $\sqrt{a} - \sqrt{b} > \sqrt{2}$, the impact of edge DP shows up explicitly in the threshold condition; (2) As we relax the relax the privacy budget, namely as $\epsilon \to \infty$, the privacy constrained threshold converges to the non-private threshold.* We next generalize our results to $r > 2$ equal sized communities and present a sufficient condition on a and b for exact recovery.

Theorem 3.4. *For $r > 2$ communities, $\mathcal{M}^{MLE}_{\text{Stability}}(G)$ satisfies exact recovery if*

$$\sqrt{a} - \sqrt{b} > \sqrt{r} \times \sqrt{1 + \frac{t+1}{\epsilon} \times \left(1 + \log\left(\sqrt{\frac{a}{b}}\right)\right)} \tag{7}$$

for any $\epsilon > 0$ and $\delta = n^{-t}$, $t > 0$.

The result for $r > 2$ communities is slightly weaker compared to the case for $r = 2$ case. However, *it still converges to the non-private optimal threshold ($\sqrt{a} - \sqrt{b} > \sqrt{r}$) when the privacy budget $\epsilon \to \infty$.*

Main Ideas behind the Proof(s) of Theorems 3.3 and 3.4 and Intuition behind the private recovery threshold: Analyzing the error probability for the stability based mechanism for SBM is highly non-trivial. Specifically, there are two types of error events occur in this mechanism when estimating the true labels $\boldsymbol{\sigma}^*$: (1) When the stability mechanism outputs the ML estimate $\hat{\boldsymbol{\sigma}}_{\text{MLE}}$, then we are interested in bounding the $\Pr(\hat{\boldsymbol{\sigma}}_{\text{MLE}} \neq \boldsymbol{\sigma}^*)$. This error probability can be analyzed using existing results on exact recovery (Abbe et al., 2015), and the error vanishes as $o(1)$ if $\sqrt{a} - \sqrt{b} > \sqrt{r}$. (2) The second source of error is when the mechanism outputs a random label $\perp$, whose probability is bounded by $\Pr(\tilde{d} \leq \frac{\log 1/\delta}{\epsilon})$. The key technical challenge arises in the analysis of this probability. Specifically, we show that when the graph G is drawn from an SBM, the ML estimator is $\Omega(\log(n))$-stable with high probability. By leveraging this result, we bound the probability $\Pr(\tilde{d} \leq \frac{\log 1/\delta}{\epsilon})$, and in order to make this probability decay as $o(1)$ for exact recovery, we obtain sufficient conditions on (a, b) presented in Theorems 3.3 and 3.4.

Stability of SDP relaxation. We show that the SDP relaxation (SDP for short) method also has the stability property, i.e., a graph G generated by an SBM is $\Omega(\log n)$-stable with respect to the SDP with high probability, which gives us the following result for both $r = 2$ and $r > 2$ multiple equal-sized communities.

Theorem 3.5. *For $r \geq 2$ communities, $\mathcal{M}^{SDP}_{\text{Stability}}(G)$ satisfies exact recovery if*

$$\sqrt{a} - \sqrt{b} > \sqrt{r} \times 4\left(1 + \frac{\sqrt{t+1}}{\sqrt{2\epsilon}}\right) \tag{8}$$

for any $\epsilon > 0$ and $\delta = n^{-t}$, $t > 0$.

In contrast with the threshold condition (7), we have a larger constant in (8) for $\mathcal{M}^{\text{SDP}}_{\text{Stability}}(G)$, arising out of the concentration bounds for the SDP relaxation algorithm.

Main ideas in the proof of Theorem 3.5. The proof of the stability of $\mathcal{M}^{\text{SDP}}_{\text{Stability}}(G)$ becomes more complex than that of MLE, because SDP only takes the ground truth label as the optimal solution in some regimes; further, arguing that a solution $\hat{\boldsymbol{\sigma}} = \hat{\boldsymbol{\sigma}}_{\text{SDP}}$ is not easy (since it may not be the min bisection). (Hajek et al., 2016a) design a sophisticated "certificate" for proving that the SDP solution is indeed the optimal, and show that the certificate holds with high probability when $\sqrt{a} - \sqrt{b} > \sqrt{r}$ (note that this certificate is much more complex than the primal-dual based certificate

used earlier for $r = 2$ communities (Chen, 2018)). The high probability bound for the certificates is unfortunately not sufficient, since we need to argue about the stability for a graph G generated from the SBM deterministically, and there are $n^{\mathcal{O}(\log n)}$ graphs within distance $\mathcal{O}(\log n)$ of G. Specifically, the high probability bound for the certificate does not hold after flipping $\Omega(\log n)$ connections, which is required to maintain the stability of the optimal solution. Instead, we define a notion of "concentration", and show that if a graph is concentrated, then $SDP(G)$ is optimal at the ground truth label; *note that this holds deterministically, not with high probability*. We then use this notion of concentration to determine stability, by showing that all graphs within $\mathcal{O}(\log n)$ distance of G are also concentrated. Finally, we derive a lower bound on $\sqrt{a} - \sqrt{b}$ that is both (1) sufficient for concentration and (2) able to preserve concentration after flipping up to $\Omega(\log n)$ connections. We give more details below.

We say that a graph is (c_1, c_2, c_3, c_4)-concentrated, for constants $c_1, \ldots, c_4$, if the following four conditions hold:

- $\min_{i \in V(G)}(s_i - r_i) > c_1 \log n$, where s_i is the number of same-community neighbors of i and r_i is the maximum number of neighbors of i in one of the other communities.
- $\|\mathbf{A} - \mathbb{E}[\mathbf{A}]\|_2 \leq c_2\sqrt{\log n}$
- $\max_{k \in [r]} \frac{1}{K} \sum_{i \in C_k} r_i \leq Kq + c_3\sqrt{\log n}$, where $K = n/r$.
- $e(C_k, C_{k'}) \geq K^2 q - 3/4K\sqrt{\log n} - c_4 \log n$, where $e(C_k, C_{k'})$ is the number of inter-community edges between communities k and $k' \neq k$

Next, we prove that a graph generated by an SBM with appropriate parameters will be $(c_1, \ldots, c_4)$-concentrated w.h.p.. The concentration holds with high probability only when a and b satisfies some conditions related to r i.e., a must be large enough (relatively to b) and they will determine the exact recovery threshold of the method. Next, we prove that the concentration persists under $\Omega(\log n)$ edge perturbations, i.e., that if the original graph is concentrated under a tuple $(c_1, \ldots, c_4)$, a graph obtained by flipping up to $\Omega(\log n)$ connections of the original one is also concentrated with slightly different tuple.

We then apply the analyses of (Hajek et al., 2016b) to prove that when a graph is $(c_1, \ldots, c_4)$-concentrated for some constants c_i, the SDP relaxation (SDP for short) outputs the (1) uniquely optimal solution and (2) the optimal solution is also the ground truth community vector. Our proof differs from (Hajek et al., 2016b)'s proof in a way that (Hajek et al., 2016b)'s conditions holds with high probability and ours holds deterministically. First we note that the SDP can be presented by the following form:

$$\begin{aligned} &\text{maximize } \langle \mathbf{A}, \mathbf{Z} \rangle \\ &\text{subsect to } \mathbf{Z} \succcurlyeq 0 \\ &\qquad Z_{ii} = 1, \forall i \in [n] \\ &\qquad Z_{ij} \geq 0, \forall i, j \in [n] \\ &\qquad \mathbf{Z}\mathbf{1} = K\mathbf{1}, \end{aligned}$$

Then we provide the condition for a dual certificate (deterministically). Intuitively, if we can construct a positive semi-definite matrix S^* by the following formula without violating the two constraints below, the SDP is uniquely optimal at $\mathbf{Z}^*$ constructed by the ground truth community label (We say SDP(G) is optimal at the ground truth community label for short).

Lemma 6 of (Hajek et al., 2016b). Suppose there exists $\mathbf{D}^* = diag(d_i^*)$ with $d_i^* > 0$ for all $i, \mathbf{B}^* \in \mathcal{S}^n$ with $\mathbf{B}^* \geq \mathbf{0}$ and $B_{ij} > 0$ whenever i and j are in distinct clusters, and $\lambda^* \in \mathbb{R}^n$ such that $\mathbf{S}^* \triangleq \mathbf{D}^* - \mathbf{B}^* - \mathbf{A} + \lambda^*\mathbf{1}^T + \mathbf{1}(\lambda^*)^T$ satisfies $\mathbf{S}^* \succcurlyeq 0$ and

$$\begin{aligned} \mathbf{S}^*\xi_k^* &= 0, \forall k \in [r] \\ B_{ij}^* Z_{ij}^* &= 0, \forall i, j \in [n] \end{aligned}$$

Then SDP$(G) = \mathbf{Z}^*$ is the unique solution for the SDP.

We then prove that the concentration of the input graph implies the existence of a positive semi-definite matrix S^*, which satisfies the dual certificate above, i.e., we point out that there's always a way to construct matrices $\mathbf{D}^*, \mathbf{B}^*$ that satisfies above conditions from the concentration's conditions. We note that when such $\mathbf{S}^*$ exists, the SDP will uniquely output the ground truth community vector.

Therefore if a graph G (with size n large enough) is generated by an SBM with the ground truth community vector and G is concentrated, $SDP(G)$ will outputs $\mathbf{Z}^*$. We also know that any G' obtained by flipping up to $c \log n/\epsilon$ edges of G is also concentrated (for some constant c). It means that $SDP(G')$ also outputs $\mathbf{Z}^*$ and proves that SDP is $c \log n/\epsilon$-stable. Compose with the fact that a graph generated by such SBM will be concentrated with high probability, we conclude that SDP is $c \log n/\epsilon$-stable with high probability. The threshold for Theorem 3.5 derives from the conditions of a, b and r for which the concentration holds with high probability and choosing the constant c accordingly to δ.

Complexity of Stability Based Mechanisms. A naive implementation of $\mathcal{M}^{\hat{\sigma}}_{\text{Stability}}(G)$, which involves computing $d_{\hat{\boldsymbol{\sigma}}}(G)$ in Step 3 using (5), requires computing $\hat{\boldsymbol{\sigma}}(G')$ for all graphs G'. It can be shown that the algorithm works if we use $\min\{d_{\hat{\boldsymbol{\sigma}}}(G), \mathcal{O}(\log n)\}$, instead of $d_{\hat{\boldsymbol{\sigma}}}(G)$, for which it suffices to compute $\hat{\boldsymbol{\sigma}}(G')$ for only those graphs G' with

$d(G, G') = \mathcal{O}(\log n)$. The MLE algorithm takes exponential time, so algorithm $\mathcal{M}^{\text{MLE}}_{\text{Stability}}(G)$ still takes exponential time; however, $\mathcal{M}^{\text{SDP}}_{\text{Stability}}(G)$ can be implemented in quasi-polynomial time, i.e., $n^{(\mathcal{O}(\log(n)))}$, using the above observation.

3.2. Sampling Mechanisms

We present two sampling based approaches for private community detection. In the first approach of *Bayesian Sampling*, presented in Algorithm 2, we compute the posterior probability of label vectors given the graph G and release a label estimate by sampling from this posterior distribution.

Algorithm 2 $\mathcal{M}_{\text{Bayesian}}(G)$: Bayesian Sampling Mechanism

1: **Input:** $G(\mathcal{V}, E) \in \mathcal{G}$
2: **Output:** A labelling vector $\hat{\boldsymbol{\sigma}} \in \mathcal{L}$.
3: For every $\boldsymbol{\sigma} \in \mathcal{L}$, calculate $p(\boldsymbol{\sigma}|G) = \frac{p(\boldsymbol{\sigma}) \times p(G|\boldsymbol{\sigma})}{p(G)}$
4: Sample and output a labelling $\hat{\boldsymbol{\sigma}} \in \mathcal{L}$ with probability $\Pr(\hat{\boldsymbol{\sigma}}|G)$

Surprisingly, we show that this mechanism satisfies pure ϵ-edge DP whenever ϵ is larger than a threshold, namely, $\epsilon \geq \log(a/b)$. This is in-contrast with Stability mechanisms which achieve approximate (ϵ, δ)-edge DP, for any $\epsilon > 0$ but require $\delta = 1/n^t$, for any $t > 0$. Our main result for the Bayesian mechanism is stated in the following theorem along with the corresponding recovery threshold.

Theorem 3.6. *The mechanism $\mathcal{M}_{\text{Bayesian}}(G)$ satisfies ϵ-edge DP, $\forall \epsilon \geq \epsilon_0 = \log\left(\frac{a}{b}\right)$, and for $r = 2$ communities, satisfies exact recovery if*

$$\sqrt{a} - \sqrt{b} > \max\left[\sqrt{2}, \frac{2}{(\sqrt{2}-1)(1-e^{-\epsilon_0})}\right]. \tag{9}$$

Despite the fact that the Bayesian mechanism provides pure edge DP, one disadvantage is that it requires the knowledge of (a, b) for computing the posterior distribution. To this end, we present and analyze the exponential sampling mechanism in Algorithm 3, where we sample from a distribution over the labels which can be computed directly from the graph and does not require the knowledge of (a, b). Specifically, for any label vector $\boldsymbol{\sigma}$ (partition of the graph in two communities), the score$(\boldsymbol{\sigma}) = E_{\text{inter}}(G, \boldsymbol{\sigma})$ is defined as the set of cross-community edges in the partition $\boldsymbol{\sigma}$, the corresponding sampling probability is computed as a function of this score and the privacy budget.

Theorem 3.7. *The exponential sampling mechanism $\mathcal{M}_{\text{Expo.}}(G)$ satisfies ϵ-edge DP and for $r = 2$ communities, performs exact recovery if*

$$\sqrt{a} - \sqrt{b} > \max\left[\sqrt{2}, \frac{2}{(\sqrt{2}-1)\epsilon}\right]. \tag{10}$$

Algorithm 3 $\mathcal{M}_{\text{Expo.}}(G)$: Exponential Mechanism

1: **Input:** $G(\mathcal{V}, E) \in \mathcal{G}$
2: **Output:** A labelling vector $\hat{\boldsymbol{\sigma}} \in \mathcal{L}$.
3: For every $\boldsymbol{\sigma} \in \mathcal{L}$, calculate score$(\boldsymbol{\sigma}) = E_{\text{inter}}(G, \boldsymbol{\sigma})$
4: Sample and output a labelling $\hat{\boldsymbol{\sigma}} \in \mathcal{L}$ with probability $\exp(-\epsilon \times \text{score}(\boldsymbol{\sigma}))$

Complexity and comparison with stability based mechanisms. A key advantage of the sampling based mechanisms over stability based mechanisms is that they give ϵ-DP solutions. However, implementing the sampling step in these mechanisms takes exponential time, as no efficient algorithm is known for sampling $\boldsymbol{\sigma}$ with probability depending on its utility.

3.3. Graph Perturbation Mechanisms

In this section, we present and analyze randomized response (RR) based mechanism for private community detection. The basic idea is to perturb the edges of the random graph (i.e., the adjacency matrix $\mathbf{A}$), where each element $A_{i,j}$ is perturbed independently to satisfy ϵ-edge DP. For a graph with an adjacency matrix $\mathbf{A}$, the perturbed matrix is denoted as $\tilde{\mathbf{A}}$, where $\mu = \Pr(\tilde{A}_{i,j} = 1 | A_{i,j} = 0) = \Pr(\tilde{A}_{i,j} = 0 | A_{i,j} = 1)$. By picking $\mu = \frac{1}{e^\epsilon + 1}$, it can be readily shown that the mechanism satisfies ϵ-edge DP. One can then apply any community recovery algorithm (MLE, SDP or spectral methods) on the perturbed matrix $\tilde{\mathbf{A}}$. This mechanism is presented in Algorithm 4.

Algorithm 4 $\mathcal{M}^{\hat{\sigma}}_{\text{RR}}(G)$: Graph Perturbation Mechanism via Randomized Response

1: **Input:** $G(\mathcal{V}, E) \in \mathcal{G}$
2: **Output:** A labelling vector $\hat{\boldsymbol{\sigma}} \in \mathcal{L}$.
3: Perturb $\mathbf{A} \to \tilde{\mathbf{A}}$ via randomized response mechanism
4: Apply community detection algorithm on $\tilde{\mathbf{A}}$
5: Output $\hat{\boldsymbol{\sigma}}(\tilde{\mathbf{A}})$

From the perspective of computational complexity, this Algorithm is faster compared to the stability and sampling based approaches. However, in the next Theorem, we state our main result which shows that RR based mechanism achieves exact recovery if $\epsilon = \Omega(\log(n))$, i.e., it requires the privacy leakage to grow with n for exact recovery.

Theorem 3.8. *The mechanism $\mathcal{M}^{SDP}_{\text{RR}}(G)$ satisfies ϵ-edge DP, $\forall \epsilon \geq \epsilon_n = \Omega(\log(n))$, and for $r = 2$ communities, satisfies exact recovery if*

$$\sqrt{a} - \sqrt{b} > \sqrt{2} \times \frac{\sqrt{e^\epsilon + 1}}{\sqrt{e^\epsilon - 1}} + \frac{1}{\sqrt{e^\epsilon - 1}}. \tag{11}$$

In order to understand the intuition behind the worse privacy leakage of RR mechanism for exact recovery, it is

instructive to consider the statistics of the perturbed adjacency matrix $\tilde{\mathbf{A}}$ as a function of ϵ. Specifically, the perturbed elements in the adjacency matrix $\tilde{\mathbf{A}}$ are distributed as follows $\tilde{A}_{i,j} \sim \text{Bern}(\tilde{p}), i < j$, if $\sigma_i = \sigma_j$, and $\tilde{A}_{i,j} \sim \text{Bern}(\tilde{q}), i < j$, if $\sigma_i \neq \sigma_j$, where

$$\tilde{p} = \underbrace{\left[\frac{n}{(e^\epsilon+1)\times\log(n)} + \frac{e^\epsilon-1}{e^\epsilon+1}\times a\right]}_{a_n} \times \frac{\log(n)}{n},$$

$$\tilde{q} = \underbrace{\left[\frac{n}{(e^\epsilon+1)\times\log(n)} + \frac{e^\epsilon-1}{e^\epsilon+1}\times b\right]}_{b_n} \times \frac{\log(n)}{n}. \quad (12)$$

Note that $\tilde{p}$ and $\tilde{q}$ are the intra- and inter- community connection probabilities for the perturbed matrix. From the above, we note that if ϵ is chosen as a constant, and as n grows, then $\lim_{n\to\infty}\tilde{p} = \lim_{n\to\infty}\tilde{q}$, i.e., if we insist on constant ϵ, then asymptotically, the statistics of the inter- and intra-community edges are the same and exact recovery is impossible. The result of Theorem 3.8 shows that one can indeed get exact recovery by allowing the leakage to grow logarithmically with n.

4. Numerical Experiments

In this section, we present experimental results to assess the performance of our proposed private community detection algorithms, and the associated tradeoffs between privacy and community recovery for both synthetically generated graphs (SBMs) as well as real-world graphs. The proposed mechanisms are implemented in MATLAB 2020b and the optimization (SDP) is done through CVX solver (Grant et al., 2009). In the numerical results, we perform Monte Carlo simulations, where in each iteration we compute the normalized hamming distance between $\boldsymbol{\sigma}$ and $\hat{\boldsymbol{\sigma}}$ as an estimate for the error probability. Our numerical experiments address the following questions:

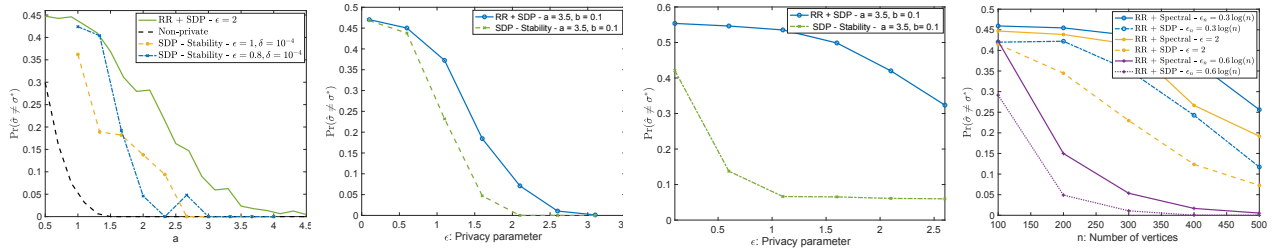
Q1: How does the error probability change with a and b? We first study community recovery on synthetic graphs (SBM) with $n = 100$ vertices, $r = 2$ communities, $b = 0.1$ and vary the parameter a. Fig. 2(a) shows the impact of increasing a on the error probability of (i) non-private recovery; (ii) SDP-stability mechanism and (iii) randomized-response SDP mechanism. For a fixed privacy budget ϵ, we observe that when the difference between a and b increases, the error probabilities for all private mechanisms decrease but are no better than the non-private case. For a fixed ϵ, the SDP-stability mechanism achieves a smaller error probability compared to RR+SDP mechanism, however, this comes at the expense of approximate edge DP guarantee.

Q2: What is the impact of ϵ on the error probability? In Figs. 2 (b) and (c), we fix $n = 200$, $a = 3.5$, $b = 0.1$ and study the impact of privacy budget ϵ on the error probability for the case of $r = 2$ and $r = 3$ communities. Specifically, for $r = 2$, we observe that the SDP-stability mechanism (with $\delta = 10^{-5}$) outperforms RR+SDP; furthermore, as ϵ increases beyond a certain threshold, error probability for both converge to 0. For $r = 3$ communities, we can observe that the difference in performance between SDP-stability and RR+SDP is even more pronounced. In this setting, however, we do not expect the error probability to converge to 0 even if $\epsilon \to \infty$ since the chosen values ($a = 3.5, b = 0.1$) do not satisfy the exact recovery threshold ($\sqrt{a} - \sqrt{b} > \sqrt{r}$).

Q3: What is the impact of the problem size on the accuracy (SDP-Stability, RR+SDP, RR+Spectral)? In Fig. 2 (d), we compare the performance of SDP relaxation based recovery versus spectral method proposed in (Hehir et al., 2021), both under randomized response for $a = 3.5, b = 0.1$ and $r = 2$ communities. We can observe that RR+SDP has less probability of error as a function of n compared with the RR-Spectral method; however, RR+SDP has more computational complexity. In Fig. 2(e), we show the error probability behavior as a function of n, the number of vertices for $r = 2$ communities and different privacy levels. From the figure, we observe that for the RR based approach, the privacy level should scale as $\Omega(\log(n))$ to achieve exact recovery, which is consistent with our theoretical findings. On the other hand, the stability based mechanisms can still provide exact recovery for finite ϵ. We can draw similar conclusions for the case of $r = 3$ communities in Fig. 2(f).

Q4: How do the *private* community detection mechanisms perform on real-world datasets? We now discuss our results for two real-world datasets (shown in Figs. 2 (g) & (h)): (1) Zachary's Karate Club dataset which contains a social network of friendships between 34 members of a karate club at a US university in the 1970s. (Girvan & Newman, 2002) and (2) The Political Blogosphere dataset (Adamic & Glance, 2005) which consists of 1490 political blogs captured during 2004 US elections. Each blog is classified as left/liberal or right/conservative, i.e., $r = 2$ and links between blogs were automatically extracted from a crawl of the front page of the blog. For the smaller size Karate club dataset ($n = 34$), we observe from Fig. 2(g), we can observe the impact of choosing δ on SDP-stability mechanism. Specifically, when $\delta = 34^{-2} \approx 10^{-3}$, then RR+SDP has lower error probability for smaller ϵ compared to SDP-stability. For the larger Political Blogosphere dataset ($n = 1490$), the SDP-stability mechanism outperforms RR+SDP for all values of ϵ and $\delta = 1490^{-2} \approx 4 \times 10^{-7}$. We can observe that SDP-Stability performs better than RR+SDP for both datasets.

Q5: How tight are the obtained bounds? We have plotted the phase transition behavior for both RR+SDP and SDP-stability mechanisms (see Fig. 3). We observe that our theoretical bound (red line) is quite tight, and the threshold

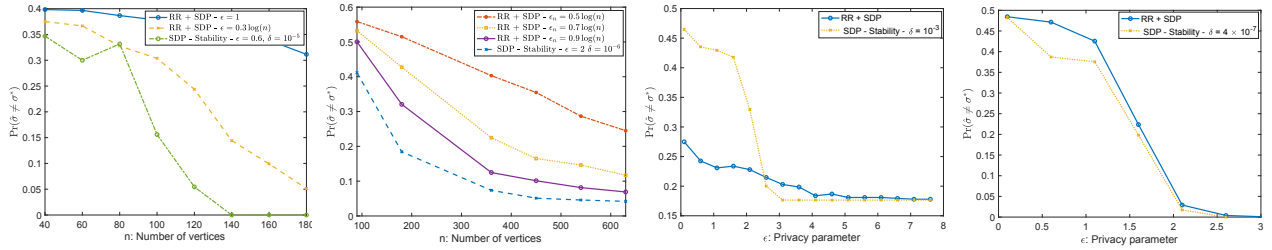


(a) Impact of changing a where $r = 2$ and $n = 100$ vertices.

(b) Impact of ϵ where $r = 2$ and $n = 200$ vertices.

(c) Impact of ϵ where $r = 3$ and $n = 200$ vertices.

(d) SDP vs. Spectral method for $r = 2$ communities.



(e) Comparison between stability and RR + SDP where $r = 2$.

(f) Comparison between stability and RR + SDP where $r = 3$.

(g) Impact of ϵ for Karate Club dataset where $r = 2$.

(h) Impact of ϵ for Political Blogosphere dataset where $r = 2$.

fixed b, ϵ. (b)-(c) Show the impact of ϵ for $r = 2, 3$, erformance on real-world datasets.

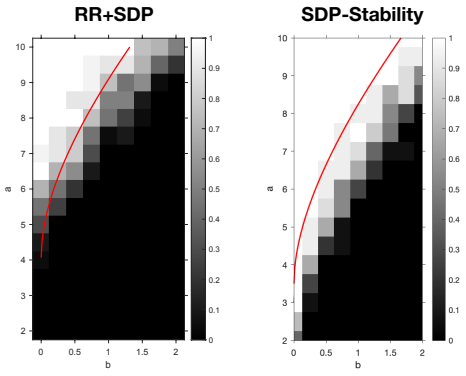


Figure 3. Phase transition for Randomized response (RR)-SDP and SDP-Stability mechanisms: darker pixels represent lower empirical probability of success for $n = 50$ nodes, $\epsilon = 2$, $\delta = 4 \times 10^{-4}$, and $r = 2$ communities. The red lines represent the theoretical bounds presented in Theorems 3.3 and 3.8.

region obtained from empirical success probability is close to this bound.

5. Conclusion

In this paper, we studied the problem of community detection for SBMs subject to edge differential privacy. We presented and analyzed three classes of privacy-preserving mechanisms (stability, sampling, and randomized response) and studied the tradeoffs for exact recovery as a function of the connectivity (a, b) and privacy parameters (ϵ, δ). From our results, we deduce the following conclusions: the stability based mechanisms can achieve (ϵ, δ)-DP for any $\epsilon > 0$ and require $\delta = n^{-t}$, $t > 0$. The sampling based mechanisms can instead achieve pure $(\epsilon, 0)$-DP; however, they require a larger separation between (a, b) for exact recovery compared to the stability based methods. Among the three mechanisms, the randomized response mechanism, while least computationally complex, requires ϵ to scale as $\Omega(\log(n))$ for exact recovery. We also presented simulation results on both synthetic and real-world graphs to validate our theoretical findings. There are several interesting open problems: a) obtaining converse results (necessary conditions) to assess the optimality (or gap to optimality) of the proposed mechanisms; b) generalization of the mechanisms and the associated analysis to degree-corrected SBMs; c) studying the impact of privacy on other recovery notions (such as weak recovery (Abbe, 2017)); and d) design and analysis of efficient algorithms for stability based mechanisms.

Acknowledgements

We thank the anonymous ICML reviewers for their insightful suggestions. The work of M. Seif and R. Tandon was supported by NSF grants CAREER 1651492, CNS 1715947, CCF 2100013 and CNS 2209951. The work of D. Nguyen and A. Vullikanti was partially supported by NSF grants CCF-1918656, IIS-1931628, IIS-1955797, and NIH grant R01GM109718.

References

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

Abbe, E., Bandeira, A. S., and Hall, G. Exact recovery in the stochastic block model. *IEEE Transactions on Information Theory*, 62(1):471–487, 2015.

Abbe, E., Fan, J., Wang, K., and Zhong, Y. Entrywise eigenvector analysis of random matrices with low expected rank. *Annals of statistics*, 48(3):1452, 2020.

Adamic, L. A. and Glance, N. The political blogosphere and the 2004 us election: divided they blog. In *Proceedings of the 3rd international workshop on Link discovery*, pp. 36–43, 2005.

Arias-Castro, E. and Verzelen, N. Community detection in dense random networks. *The Annals of Statistics*, 42(3): 940–969, 2014.

Blocki, J., Blum, A., Datta, A., and Sheffet, O. Differentially private data analysis of social networks via restricted sensitivity. In *Proceedings of the 4th Conference on Innovations in Theoretical Computer Science*, ITCS '13, pp. 87–96, New York, NY, USA, 2013. Association for Computing Machinery. ISBN 9781450318594. doi: 10.1145/2422436.2422449. URL https://doi.org/10.1145/2422436.2422449.

Blondel, V. D., Guillaume, J.-L., Lambiotte, R., and Lefebvre, E. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008:10008, 2008.

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

Borgs, C., Chayes, J. T., and Smith, A. Private graphon estimation for sparse graphs. *arXiv preprint arXiv:1506.06162*, 2015.

Chen, A. Lecture 7: Stochastic block model 2. *Beyond Worst Case Analysis*, 2018. URL https://ugtcs.berkeley.edu/src/bwca-fa18/scribe-notes-7.pdf.

Choi, D. S., Wolfe, P. J., and Airoldi, E. M. Stochastic blockmodels with a growing number of classes. *Biometrika*, 99(2):273–284, 2012.

Condon, A. and Karp, R. M. Algorithms for graph partitioning on the planted partition model. *Random Structures & Algorithms*, 18(2):116–140, 2001.

Decelle, A., Krzakala, F., Moore, C., and Zdeborová, L. Asymptotic analysis of the stochastic block model for modular networks and its algorithmic applications. *Physical Review E*, 84(6):066106, 2011.

Dwork, C., McSherry, F., Nissim, K., and Smith, A. Calibrating noise to sensitivity in private data analysis. In *Theory of cryptography conference*, pp. 265–284. Springer, 2006.

Dwork, C., Roth, A., et al. The algorithmic foundations of differential privacy. *Foundations and Trends in Theoretical Computer Science*, 9(3-4):211–407, 2014.

Fortunato, S. Community detection in graphs. *Physics reports*, 486(3-5):75–174, 2010.

Gao, C., Ma, Z., Zhang, A. Y., and Zhou, H. H. Achieving optimal misclassification proportion in stochastic block models. *The Journal of Machine Learning Research*, 18 (1):1980–2024, 2017.

Girvan, M. and Newman, M. E. Community structure in social and biological networks. *Proceedings of the national academy of sciences*, 99(12):7821–7826, 2002.

Grant, M., Boyd, S., and Ye, Y. cvx users' guide. *online: http://www. stanford. edu/˜ boyd/software. html*, 2009.

Hajek, B., Wu, Y., and Xu, J. Achieving exact cluster recovery threshold via semidefinite programming. *IEEE Transactions on Information Theory*, 62(5):2788–2797, 2016a.

Hajek, B., Wu, Y., and Xu, J. Achieving exact cluster recovery threshold via semidefinite programming: Extensions. *IEEE Transactions on Information Theory*, 62(10):5918–5937, 2016b.

Hehir, J., Slavkovic, A., and Niu, X. Consistency of privacy-preserving spectral clustering under the stochastic block model. *arXiv preprint arXiv:2105.12615*, 2021.

Holland, P. W., Laskey, K. B., and Leinhardt, S. Stochastic blockmodels: First steps. *Social networks*, 5(2):109–137, 1983.

Imola, J., Murakami, T., and Chaudhuri, K. Locally differentially private analysis of graph statistics. In *30th USENIX Symposium on Security*, 2021.

Ji, T., Luo, C., Guo, Y., Ji, J., Liao, W., and Li, P. Differentially private community detection in attributed social networks. In *Asian Conference on Machine Learning*, pp. 16–31. PMLR, 2019.

Karwa, V., Raskhodnikova, S., Smith, A., and Yaroslavtsev, G. Private analysis of graph structure. *Proceedings of the VLDB Endowment*, 4(11):1146–1157, 2011.

Kasiviswanathan, S. P., Nissim, K., Raskhodnikova, S., and Smith, A. Analyzing graphs with node differential privacy. In *Proceedings of the 10th Theory of Cryptography Conference on Theory of Cryptography*, TCC'13, pp. 457–476, Berlin, Heidelberg, 2013. Springer-Verlag. ISBN 978-3-642-36593-5. doi: 10.1007/978-3-642-36594-2_26. URL http://dx.doi.org/10.1007/978-3-642-36594-2_26.

Lei, J. and Rinaldo, A. Consistency of spectral clustering in stochastic block models. *The Annals of Statistics*, 43(1): 215–237, 2015.

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

McSherry, F. Spectral partitioning of random graphs. In *Proceedings 42nd IEEE Symposium on Foundations of Computer Science (FOCS)*, pp. 529–537, 2001.

Mossel, E., Neeman, J., and Sly, A. Consistency thresholds for the planted bisection model. In *Proceedings of the forty-seventh annual ACM symposium on Theory of computing*, pp. 69–75, 2015.

Mülle, Y., Clifton, C., and Böhm, K. Privacy-integrated graph clustering through differential privacy. In *EDBT/ICDT Workshops*, volume 157, 2015.

Nguyen, H. H., Imine, A., and Rusinowitch, M. Detecting communities under differential privacy. In *Proceedings of the 2016 ACM on Workshop on Privacy in the Electronic Society*, pp. 83–93, 2016.

Qin, Z., Yu, T., Yang, Y., Khalil, I., Xiao, X., and Ren, K. Generating synthetic decentralized social graphs with local differential privacy. In *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security (CCS)*, pp. 425–438, 2017.

Sealfon, A. and Ullman, J. Efficiently estimating erdos-renyi graphs with node differential privacy. *arXiv preprint arXiv:1905.10477*, 2019.

Thakurta, A. G. and Smith, A. Differentially private feature selection via stability arguments, and the robustness of the Lasso. In *Conference on Learning Theory*, pp. 819–850. PMLR, 2013.

Wang, P., Zhou, Z., and So, A. M.-C. A nearly-linear time algorithm for exact community recovery in stochastic block model. In *International Conference on Machine Learning (ICML)*, pp. 10126–10135. PMLR, 2020.

A. Appendix

In this Appendix, we provide the proofs of our results presented in Section 3. We provide auxiliary results that are used throughout the proofs at end of this document. We summarize the notations and symbols used in Table 2.

n	Number of vertices
$\mathbf{A}$	Adjacency matrix
$\tilde{\mathbf{A}}$	Perturbed adjacency matrix
ϵ	Privacy budget
$\boldsymbol{\sigma}^*$	Ground truth labels
$\hat{\boldsymbol{\sigma}}$	Estimated labels
$\mathcal{L}$	Set of all possible labels
$E_{\text{inter}}(G, \boldsymbol{\sigma})$	set of cross-community edges of graph G
$E(C_1, C_2)$	Number of edges between C_1 and C_2
$\mathbf{I}$	Identity matrix
$\mathbf{J}$	All ones matrix
$\mathbf{1}$	All ones vector

Table 2. List of Symbols.

B. Proof of Lemma 3.2 ($\mathcal{M}^{\text{MLE}}_{\text{stability}}(G)$ satisfies (ϵ, δ)-edge DP)

Algorithm 5 $\mathcal{M}^{\hat{\boldsymbol{\sigma}}}_{\text{Stability}}(G)$: Stability Based Mechanism

Input: $G(\mathcal{V}, E) \in \mathcal{G}$
Output: labelling vector $\hat{\boldsymbol{\sigma}}_{\text{Private}}$.
$d_{\hat{\boldsymbol{\sigma}}}(G) \leftarrow$ stability of $\hat{\boldsymbol{\sigma}}$ with respect to graph G
$\tilde{d} \leftarrow d_{\hat{\boldsymbol{\sigma}}}(G) + \text{Lap}(1/\epsilon)$
if $\tilde{d} > \frac{\log 1/\delta}{\epsilon}$ **then**
    Output $\hat{\boldsymbol{\sigma}}(G)$
else
    Output $\perp$ (random label)

The proof that the stability based mechanism satisfies $(\epsilon, \delta))$-edge DP follows directly from (Dwork et al., 2014), and we include the proof here for the sake of completeness by adapting it to the community detection problem. Given a pair of neighbor graphs $G \sim G'$, we denote $d(G)$ is the distance from G to its nearest unstable instance and $d(G')$ is the distance from G' to its nearest unstable instance. Due to the triangle inequality, $|d(G) - d(G')| \leq 1$, hence the sensitivity of d: $\Delta_d = 1$. Adding a Laplacian noise of magnitude of $1/\epsilon$ guarantees ϵ-differential privacy for $\tilde{d}$. In order to verify (ϵ, δ)-edge DP for the overall mechanism, we consider two scenarios: the first one, when the output of the mechanism is $\perp$. In this case, we have:

$$\begin{aligned}
\Pr(\mathcal{M}_{\text{Stability}}(G) = \perp) &= \Pr\left[\tilde{d}(G) \leq \frac{\log 1/\delta}{\epsilon}\right] \\
&\overset{(a)}{\leq} e^{\epsilon} \Pr\left[\tilde{d}(G') \leq \frac{\log 1/\delta}{\epsilon}\right] \\
&= e^{\epsilon} \Pr(\mathcal{M}_{\text{Stability}}(G') = \perp). \qquad (13)
\end{aligned}$$

where (a) follows from the fact that $\tilde{d}$ satisfies ϵ-DP. For the second scenario, when the output of the mechanism is some label vector $\boldsymbol{\sigma}$, we have to analyze two cases. Let us denote $\hat{\boldsymbol{\sigma}}$ as the output of the estimator when the graph is G, and $\hat{\boldsymbol{\sigma}}'$ as the output of the estimator when the graph is G'.

- Case 1. $\hat{\boldsymbol{\sigma}} = \hat{\boldsymbol{\sigma}}' = \boldsymbol{\sigma}$:

$$\begin{aligned}\Pr(\mathcal{M}_{\text{Stability}}(G) = \boldsymbol{\sigma}) &= \Pr\left[\tilde{d}(G) > \frac{\log 1/\delta}{\epsilon}\right] \\ &\leq e^{\epsilon} \Pr\left[\tilde{d}(G') > \frac{\log 1/\delta}{\epsilon}\right] \\ &= e^{\epsilon} \Pr(\mathcal{M}_{\text{Stability}}(G') = \boldsymbol{\sigma}). \end{aligned} \tag{14}$$

- Case 2. $\hat{\boldsymbol{\sigma}} \neq \hat{\boldsymbol{\sigma}}'$: In this case $d(G) = d(G') = 0$, then we have the following:

$$\Pr(\mathcal{M}_{\text{Stability}}(G) = \hat{\boldsymbol{\sigma}}') = 0 \leq e^{\epsilon} \Pr(\mathcal{M}_{\text{Stability}}(G') = \hat{\boldsymbol{\sigma}}'). \tag{15}$$

Also,

$$\begin{aligned}\Pr(\mathcal{M}_{\text{Stability}}(G') = \hat{\boldsymbol{\sigma}}') &\leq e^{\epsilon} \Pr(\mathcal{M}_{\text{Stability}}(G) = \hat{\boldsymbol{\sigma}}') + \Pr(\mathcal{M}_{\text{Stability}}(G') = \hat{\boldsymbol{\sigma}}') \\ &\leq e^{\epsilon} \Pr(\mathcal{M}_{\text{Stability}}(G) = \hat{\boldsymbol{\sigma}}') + \delta, \end{aligned} \tag{16}$$

where $\Pr(\mathcal{M}_{\text{Stability}}(G) = \boldsymbol{\sigma}') = 0$ and $\delta > 0$. The remaining two sub-cases (switching the roles of G and G') follow similarly and this completes the proof of Lemma 3.2.

C. Proof of Theorem 3.3 (Exact Recovery Threshold for $\mathcal{M}^{\text{MLE}}_{\text{Stability}}(G)$ for $r = 2$)

The error probability for the stability mechanism can be expressed as

$$\begin{aligned}\Pr(\mathcal{M}_{\text{Stability}}(G) \neq \boldsymbol{\sigma}^*) &= \Pr(\mathcal{M}_{\text{Stability}}(G) = \perp) \times \Pr(\mathcal{M}_{\text{Stability}}(G) \neq \boldsymbol{\sigma}^* | \mathcal{M}_{\text{Stability}}(G) = \perp) \\ &\quad + \Pr(\mathcal{M}_{\text{Stability}}(G) = \hat{\boldsymbol{\sigma}}_{\text{ML}}) \times \Pr(\mathcal{M}_{\text{Stability}}(G) \neq \boldsymbol{\sigma}^* | \hat{\boldsymbol{\sigma}}_{\text{stb.}} = \hat{\boldsymbol{\sigma}}_{\text{ML}}) \\ &\leq \Pr(\hat{\boldsymbol{\sigma}}_{\text{Stability}} = \perp) \times 1 + 1 \times \Pr(\hat{\boldsymbol{\sigma}}_{\text{Stability}} \neq \boldsymbol{\sigma}^* | \mathcal{M}_{\text{Stability}}(G) = \hat{\boldsymbol{\sigma}}_{\text{ML}}) \\ &\leq \Pr(\mathcal{M}_{\text{Stability}}(G) = \perp) + \Pr(\hat{\boldsymbol{\sigma}}_{\text{ML}} \neq \boldsymbol{\sigma}^*), \end{aligned} \tag{17}$$

where the probability is taken over the randomness of the Laplacian mechanism and over the randomness graph generated from SBM. We further upper bound the first term in the above equation as follows:

$$\begin{aligned}\Pr(\mathcal{M}_{\text{Stability}}(G) = \perp) &= \Pr\left[\tilde{d}(G) < \frac{\log(1/\delta)}{\epsilon}\right] \\ &= \Pr\left[d(G) + \text{Lap}(1/\epsilon) < \frac{\log(1/\delta)}{\epsilon}\right] \\ &= \Pr\left[\text{Lap}(1/\epsilon) < \frac{\log(1/\delta)}{\epsilon} - d(G)\right] \\ &\leq \Pr\left[d(G) < f_{n,\epsilon}\right] \times \Pr\left[\text{Lap}(1/\epsilon) < \frac{\log(1/\delta)}{\epsilon}\right] \\ &\quad + \Pr\left[d(G) \geq f_{n,\epsilon}\right] \times \Pr\left[\text{Lap}(1/\epsilon) < \frac{\log(1/\delta)}{\epsilon} - f_{n,\epsilon}\right] \\ &\leq \underbrace{\Pr\left[d(G) < f_{n,\epsilon}\right]}_{\text{Term 1}} + \underbrace{\Pr\left[\text{Lap}(1/\epsilon) < \frac{\log(1/\delta)}{\epsilon} - f_{n,\epsilon}\right]}_{\text{Term 2}}. \end{aligned} \tag{18}$$

Bounding Term 2: By picking $f_{n,\epsilon} = (t+1)\log(n)/\epsilon$ and $\delta = n^{-t}$ for any positive t, it can be readily shown that $\Pr\left[\text{Lap}(1/\epsilon) < \frac{\log(1/\delta)}{\epsilon} - f_{n,\epsilon} = -\log(n)/\epsilon\right] = o(1)$. To upper bound Term 1, we introduce an intermediate lemma which gives a lower bound on $d(G)$ as follows.

Lemma C.1. *Let (R^{ML}, B^{ML}) be the output of MLE(G). Let $d(G)$ be the distance from G to the nearest unstable instance, then $d(G)$ is lower bounded by*

$$d(G) \geq \min_{(R,B) \neq (R^{ML}, B^{ML})} E^{(G)}(R, B) - E^{(G)}(R^{ML}, B^{ML}). \tag{19}$$

where $E^{(G)}(R, B)$ denotes the number of edges between partitions R and B of the graph G.

Proof. W.l.o.g, let's consider two equal sized communities R and B. We prove this Lemma by contradiction. For a fixed graph G, let us assume that we have $\min_{(R,B)\neq(R^{\text{ML}},B^{\text{ML}})} E^{(G)}(R,B) - E^{(G)}(R^{\text{ML}},B^{\text{ML}}) = \tilde{d} > d$. Let $\tilde{G}$ is the nearest graph from G that we have $\text{MLE}(G) \neq \text{MLE}(\tilde{G})$. We know that $\text{dist}(G,\tilde{G}) = d$ so we can obtain $\tilde{G}$ by adding m_{add} edges to and removing m_{remove} edges from G and $m_{\text{add}} + m_{\text{remove}} = d$. For any labelling $(R,B) \neq (R^{\text{ML}},B^{\text{ML}})$, we have

$$\begin{aligned}
E^{(\tilde{G})}(R^{\text{ML}},B^{\text{ML}}) &\leq E^{(G)}(R^{\text{ML}},B^{\text{ML}}) + m_{\text{add}} \\
&\leq \min_{(R,B)\neq(R^{\text{ML}},B^{\text{ML}})} E^{(G)}(R,B) - \tilde{d} + m_{\text{add}} \\
&\leq \min_{(R,B)\neq(R^{\text{ML}},B^{\text{ML}})} E^{(G)}(R,B) + m_{\text{remove}} + m_{\text{add}} - \tilde{d} \\
&\leq \min_{(R,B)\neq(R^{\text{ML}},B^{\text{ML}})} E^{(G)}(R,B) + d - \tilde{d} \\
&< \min_{(R,B)\neq(R^{\text{ML}},B^{\text{ML}})} E^{(G)}(R,B) + d
\end{aligned} \tag{20}$$

which implies $(R^{ML},B^{ML}) = \text{MLE}(\tilde{G})$, which contradicts that $\text{MLE}(G) \neq \text{MLE}(\tilde{G})$. □

With this Lemma, we now return to analyze the first term in (18) as follows.

Bounding Term 1: By expanding the probability by law of total of probability theorem, we get the following sequences of steps:

$$\begin{aligned}
&\Pr\left[d(G) < f_{n,\epsilon}\right] \\
&\overset{(a)}{\leq} \Pr\left[E(R^{\min},B^{\min}) - E(R^{\text{ML}},B^{\text{ML}}) < f_{n,\epsilon}\right] \\
&= \Pr\left[E(R^{\min},B^{\min}) - E(R^{\text{ML}},B^{\text{ML}}) < f_{n,\epsilon} | E(R^{\text{ML}},B^{\text{ML}}) = E(R^*,B^*)\right] \times \Pr\left[E(R^{\text{ML}},B^{\text{ML}}) = E(R^*,B^*)\right] \\
&\quad + \Pr\left[E(R^{\min},B^{\min}) - E(R^{\text{ML}},B^{\text{ML}}) < f_{n,\epsilon} | E(R^{\text{ML}},B^{\text{ML}}) \neq E(R^*,B^*)\right] \times \Pr\left[E(R^{\text{ML}},B^{\text{ML}}) \neq E(R^*,B^*)\right] \\
&\overset{(b)}{\leq} \Pr\left[E(R^{\min},B^{\min}) - E(R^*,B^*) < f_{n,\epsilon}\right] \times 1 + 1 \times o(1) \\
&= \Pr\left[E(R^{\min},B^{\min}) - E(R^*,B^*) < f_{n,\epsilon}\right] + o(1),
\end{aligned} \tag{21}$$

where (a) follows from Lemma 2.1, and we have defined $(R^{\min},B^{\min})$ as the solution of the minimization $\min_{(R,B)\neq(R^{\text{ML}},B^{\text{ML}})} E^{(G)}(R,B)$; in step (b), $\Pr\left[E(R^{\text{ML}},B^{\text{ML}}) \neq E(R^*,B^*)\right] = o(1)$, when $\sqrt{a} - \sqrt{b} > \sqrt{2}$ (Abbe et al., 2015). Note that any two communities $(R,B) \neq (R^*,B^*)$, (R,B) can be expressed as

$$\begin{aligned}
B &= B^* - S_2 + S_1, \\
R &= R^* - S_1 + S_2,
\end{aligned} \tag{22}$$

where S_1 and S_2 are the set of mis-classified labels with respect to R^* and B^*, respectively. By the construction of symmetric communities, we have $|S_1| = |S_2|$. We can further write the $E(R^{\min},B^{\min}) - E(R^*,B^*)$ as

$$\begin{aligned}
E(R^{\min},B^{\min}) - E(R^*,B^*) = E(S_1, R^*\backslash S_1) + E(S_2, B^*\backslash S_2) \\
- E(S_1, B^*\backslash S_2) - E(S_2, R^*\backslash S_1).
\end{aligned} \tag{23}$$

Given S_1, S_2 both of size k, $k \in [1, \frac{n}{2}]$, we have

$$\Pr(E(R^{\min},B^{\min}) - E(R^*,B^*) \leq f_{n,\epsilon} || S_1| = k, |S_2| = k) = \Pr(\tilde{m}_1^{(k)} - \tilde{m}_2^{(k)} < f_{n,\epsilon}), \tag{24}$$

where $\tilde{m}_1^{(k)} \sim \text{Bin}(2k(n-k),p)$, and $\tilde{m}_2^{(k)} \sim \text{Bin}(2k(n-k),q)$. Applying Chernoff's bounds, we get

$$\begin{aligned}
\Pr(\tilde{m}_1^{(k)} - \tilde{m}_2^{(k)} < f_{n,\epsilon}) &\leq \min_{\lambda>0} e^{\lambda f_{n,\epsilon}} \times \mathbb{E}\left[e^{-\lambda(m_1^{(k)} - m_2^{(k)})}\right] \\
&= \min_{\lambda>0} e^{\lambda f_{n,\epsilon}} \times \mathbb{E}\left[e^{-\lambda m_1^{(k)}}\right] \times \mathbb{E}\left[e^{\lambda m_2^{(k)}}\right] \\
&= \min_{\lambda>0} e^{\lambda f_{n,\epsilon}} \times (1 - p(1-e^{-\lambda}))^{2k(n-k)} \times (1 - q(1-e^{\lambda}))^{2k(n-k)} \\
&\leq n^{-g},
\end{aligned} \tag{25}$$

where $g = (a+b) \times \frac{2k(n-k)}{n} - \gamma - \frac{\alpha}{2}\log\left[\frac{(\gamma-\alpha)a}{(\gamma+\alpha)b}\right]$ and $\gamma = \sqrt{\frac{(t+1)^2}{\epsilon^2} + 4 \times \frac{4k^2(n-k)^2}{n^2}ab}$ and $\alpha = (t+1)/\epsilon$. The upper bound in (25) is invoked from an existing result (Hajek et al., 2016b) (stated in Lemma H.4). We further lower bound g as follows:

$$
\begin{aligned}
g &= (a+b) \times \frac{2k(n-k)}{n} - \gamma - \frac{\alpha}{2}\log\left[\frac{(\gamma-\alpha)a}{(\gamma+\alpha)b}\right] \\
&\geq (a+b) \times \frac{2k(n-k)}{n} - \gamma - \frac{\alpha}{2}\log\left(\frac{a}{b}\right) \\
&= (a+b) \times \frac{2k(n-k)}{n} - \sqrt{\frac{(t+1)^2}{\epsilon^2} + 4 \times \frac{4k^2(n-k)^2}{n^2}ab} - \frac{(t+1)}{2\epsilon}\log\left(\frac{a}{b}\right) \\
&= (a+b) \times \frac{2k(n-k)}{n} - \frac{2k(n-k)}{n} \times \sqrt{\frac{n^2}{4k^2(n-k)^2} \times \frac{(t+1)^2}{\epsilon^2} + 4ab} - \frac{t+1}{2\epsilon}\log\left(\frac{a}{b}\right) \\
&\overset{(a)}{\geq} (a+b) \times \frac{2k(n-k)}{n} - \frac{2k(n-k)}{n} \times \sqrt{\frac{(t+1)^2}{\epsilon^2} + 4ab} - \frac{t+1}{2\epsilon}\log\left(\frac{a}{b}\right) \\
&= (a+b) \times \frac{2k(n-k)}{n} - \frac{2k(n-k)}{n} \times 2 \times \sqrt{\frac{(t+1)^2}{4\epsilon^2} + ab} - \frac{(t+1)}{2\epsilon}\log\left(\frac{a}{b}\right) \\
&= \frac{4k(n-k)}{n} \times \underbrace{\left[\frac{a+b}{2} - \sqrt{\frac{(t+1)^2}{4\epsilon^2} + ab}\right]}_{C_{a,b}} - \frac{t+1}{2\epsilon}\log\left(\frac{a}{b}\right),
\end{aligned} \tag{26}
$$

where in step (a), we have the term $\frac{n^2}{k^2(n-k)^2} \leq 4, \forall k \in [1 : \frac{n}{2}]$. Applying the union bound and assuming that $C_{a,b} > 1$, we have

$$
\begin{aligned}
&\Pr(E(R^{\min}, B^{\min}) - E(R^*, B^*) < f_{n,\epsilon}) \leq \sum_{k=1}^{\frac{n}{2}} \binom{n}{k}^2 \times \exp\left[-\frac{\log(n)}{n} \times 4k(n-k)C_{a,b} + \frac{t+1}{2\epsilon}\log\left(\frac{a}{b}\right)\right] \\
&\leq \sum_{k=1}^{\frac{n}{2}} \left(\frac{ne}{k}\right)^{2k} \times \exp\left[-\frac{\log(n)}{n} \times 4k(n-k)C_{a,b} + \frac{t+1}{2\epsilon}\log\left(\frac{a}{b}\right)\right] \\
&= \sum_{k=1}^{\frac{n}{2}} \exp\left[2k\left(\log\left(\frac{n}{k}\right) + 1\right) - \frac{\log(n)}{n} \times 4k(n-k)C_{a,b} + \frac{t+1}{2\epsilon}\log\left(\frac{a}{b}\right)\right] \\
&= \sum_{k=1}^{\frac{n}{2}} \exp\left[2k\left(\log(n) - \log(k) + 1 - \left(2 - \frac{2k}{n}\right)C_{a,b}\log(n)\right) + \frac{t+1}{2\epsilon}\log\left(\frac{a}{b}\right)\right]
\end{aligned}
$$

$$
\begin{aligned}
&\leq \sum_{k=1}^{\frac{n}{2}} \exp\left[2k\bigg(\log(n)-\log(k)+1-2\times\left(1-\frac{k}{n}\right)\times 1\times\log(n)\bigg)+\frac{t+1}{2\epsilon}\log\left(\frac{a}{b}\right)\right]\\
&= \sum_{k=1}^{\frac{n}{2}} \exp\left[2k\bigg(-\log(n)-\log(k)+1+\frac{2k}{n}\log(n)\bigg)+\frac{t+1}{2\epsilon}\log\left(\frac{a}{b}\right)\right]\\
&= \sum_{k=1}^{\frac{n}{2}} \exp\left[2k\bigg(-\log(k)+1+\frac{2k}{n}\log(n)\bigg)\right]\times\exp(-2k\log(n))\times\exp\left[\frac{t+1}{2\epsilon}\log\left(\frac{a}{b}\right)\right]\\
&< n^{-2}\times\left(\frac{a}{b}\right)^{(t+1)/2\epsilon}\times\sum_{k=1}^{\frac{n}{2}} \exp\left[2k\bigg(-\log(k)+1+\frac{2k}{n}\log(n)\bigg)\right]\\
&= n^{-2}\times\left(\frac{a}{b}\right)^{(t+1)/2\epsilon}\times\sum_{k=1}^{\frac{n}{2}} \exp\left[-2k\bigg(\log(k)-\frac{2k}{n}\log(n)-1\bigg)\right]\\
&\overset{(a)}{\leq} n^{-2}\times\left(\frac{a}{b}\right)^{(t+1)/2\epsilon}\times\sum_{k=1}^{\frac{n}{2}} \exp\left[-2k\bigg(\frac{1}{3}\log(k)-1\bigg)\right]\\
&\overset{(b)}{=} o(1),
\end{aligned}
$$

where in step (a), we have that $\log(k)-\frac{2k}{n}\log(n)>\frac{1}{3}\log(k)$ for sufficiently large n (Abbe et al., 2015). In step (b), we have that $\left(\frac{a}{b}\right)^{(t+1)/2\epsilon}\times\sum_{k=1}^{\frac{n}{2}}\exp\left[-2k\left(\frac{1}{3}\log(k)-1\right)\right]=O(1)$. To this end, we have the following conditions on a and b:

$$
a-b>\frac{t+1}{\epsilon}\Rightarrow\sqrt{a-b}>\frac{\sqrt{t+1}}{\sqrt{\epsilon}}\overset{(a)}{\Rightarrow}\sqrt{a}-\sqrt{b}>\frac{\sqrt{t+1}}{\sqrt{\epsilon}}, \tag{27}
$$

$$
\frac{a+b}{2}-\sqrt{\frac{(t+1)^2}{4\epsilon^2}+ab}>1\overset{(b)}{\Rightarrow}\sqrt{a}-\sqrt{b}>\sqrt{2}\times\sqrt{1+\frac{t+1}{2\epsilon}}, \tag{28}
$$

where in (a), we have that $\sqrt{a-b}>\sqrt{a}-\sqrt{b}$ where $a>b$, while in (b), we used the fact that $\sqrt{a+b}<\sqrt{a}+\sqrt{b}$. Therefore, a sufficient condition to make Term 1 behave as $o(1)$ will be

$$
\begin{aligned}
\sqrt{a}-\sqrt{b}&>\sqrt{2}\times\max\left[\frac{\sqrt{t+1}}{\sqrt{2\epsilon}},\sqrt{1+\frac{t+1}{2\epsilon}}\right]\\
&=\sqrt{2}\times\sqrt{1+\frac{t+1}{2\epsilon}}.
\end{aligned} \tag{29}
$$

This completes the proof of Theorem 3.3.

D. Proof of Theorem 3.4 (Threshold condition for $\mathcal{M}^{\mathbf{MLE}}_{\text{Stability}}(G)$ for $r>2$)

The proof steps follow on similar lines as the $r=2$ case. Specifically, the error probability boils down to establishing an upper bound on $\Pr\left(d(G)\leq\frac{(t+1)\log(n)}{\epsilon}\right)$ (similar to Term 1 in the proof of Theorem 3.3) as follows.

$$
\Pr\left(d(G)\leq\frac{(t+1)\log(n)}{\epsilon}\right)\overset{(a)}{\leq} rn\times\Pr\left(\text{Bin}\left(\frac{n}{r},p\right)-\text{Bin}\left(\frac{n}{r},q\right)\leq\frac{(t+1)\log(n)}{\epsilon}\right)\leq n^{-g}, \tag{30}
$$

where step (a) follows from applying the union bound. In order to further upper bound (30), we invoke Lemma H.4. Define $\gamma = \sqrt{\frac{(t+1)^2}{\epsilon^2} + 4 \times \frac{ab}{r^2}}$ and $\alpha = \frac{t+1}{\epsilon}$. The function g is lower bounded as follows:

$$\begin{aligned}
g &= \frac{a+b}{r} - \gamma - \frac{\alpha}{2}\log\left[\frac{(\gamma-\alpha)a}{(\gamma+\alpha)b}\right] \\
&\geq \frac{a+b}{r} - \gamma - \frac{\alpha}{2}\log\left(\frac{a}{b}\right) \\
&= \frac{a+b}{r} - \sqrt{\frac{(t+1)^2}{\epsilon^2} + 4 \times \frac{ab}{r^2}} - \frac{t+1}{2\epsilon}\log\left(\frac{a}{b}\right) \\
&= \frac{a+b}{r} - \frac{2}{r}\sqrt{\frac{r^2(t+1)^2}{4\epsilon^2} + ab} - \frac{t+1}{2\epsilon}\log\left(\frac{a}{b}\right) \\
&= \frac{1}{r} \times \left[a+b-2\sqrt{\frac{r^2(t+1)^2}{4\epsilon^2} + ab} - \frac{r(t+1)}{2\epsilon}\log\left(\frac{a}{b}\right)\right] \\
&\geq \frac{1}{r} \times \left[a+b-2\sqrt{ab} - \frac{(t+1)r}{\epsilon} - \frac{r(t+1)}{2\epsilon}\log\left(\frac{a}{b}\right)\right] \\
&\geq \frac{1}{r} \times \left[a+b-2\sqrt{ab} - \frac{r}{\epsilon} \times \left(t+1+\frac{t+1}{2}\log\left(\frac{a}{b}\right)\right)\right] \\
&= \frac{1}{r} \times \left[(\sqrt{a}-\sqrt{b})^2 - \frac{r}{\epsilon} \times \left(t+1+\frac{t+1}{2}\log\left(\frac{a}{b}\right)\right)\right] \\
&= \frac{(\sqrt{a}-\sqrt{b})^2}{r} - \frac{1}{\epsilon} \times \left(t+1+\frac{t+1}{2}\log\left(\frac{a}{b}\right)\right) \qquad (31)
\end{aligned}$$

We have the following conditions on a and b:

$$\begin{aligned}
&\frac{(t+1)\log(n)}{\epsilon} \leq \frac{n}{r} \times (a-b) \times \frac{\log(n)}{n} \\
&\Rightarrow a - b \geq \frac{(t+1)r}{\epsilon} \\
&\Rightarrow \sqrt{a} - \sqrt{b} \geq \frac{\sqrt{t+1}}{\sqrt{\epsilon}} \times \sqrt{r}. \qquad (32)
\end{aligned}$$

Also, we require that

$$\begin{aligned}
&1 - \frac{(\sqrt{a}-\sqrt{b})^2}{r} + \frac{1}{\epsilon} \times \left(t+1+\frac{t+1}{2}\log\left(\frac{a}{b}\right)\right) < 0 \\
&\Rightarrow \frac{(\sqrt{a}-\sqrt{b})^2}{r} > 1 + \frac{1}{\epsilon} \times \left(t+1+\frac{t+1}{2}\log\left(\frac{a}{b}\right)\right) \\
&\Rightarrow \sqrt{a} - \sqrt{b} > \sqrt{r} \times \sqrt{1 + \frac{1}{\epsilon} \times \left(t+1+\frac{t+1}{2}\log\left(\frac{a}{b}\right)\right)}, \qquad (33)
\end{aligned}$$

where (33) leads to a sufficient condition on a and b for exact recovery. This completes the proof of Theorem 3.4.

Before delving into proving Theorem 3.5, we first present the optimization problem for SDP relaxation for $r = 2$ and $r > 2$ communities as follows.

SDP Relaxation Recovery Algorithm:

Let us first define $\mathbf{Y} = \boldsymbol{\sigma}\boldsymbol{\sigma}^T$, where $Y_{i,i} = 1, \forall i \in [n]$, and $\mathbf{J}$ as all ones matrix. Our goal is to solve the following optimization problem:

$$\begin{aligned}\hat{\mathbf{Y}}_{\text{SDP}} = \max_{\mathbf{Y}} \quad & \text{tr}(\mathbf{AY}) \\ \text{s.t.} \quad & \mathbf{Y} \succcurlyeq \mathbf{0} \\ & Y_{i,i} = 1, \forall i \in [n] \\ & \text{tr}(\mathbf{JY}) = 0. \end{aligned} \tag{34}$$

It has been shown that if $\sqrt{a} - \sqrt{b} > \sqrt{2}$, then $\Pr(\hat{\mathbf{Y}}_{\text{SDP}} = \mathbf{Y}^*) = 1 - o(1)$. For r communities each of size $\frac{n}{r}$, the ML estimator (Hajek et al., 2016a) is given as:

$$\begin{aligned}\hat{\mathbf{Z}}_{\text{SDP}} = \max_{\mathbf{Z}} \quad & \text{tr}(\mathbf{AZ}) \\ \text{s.t.} \quad & \mathbf{Z} \succcurlyeq \mathbf{0} \\ & Z_{i,i} = 1, \forall i \in [n] \\ & Z_{i,j} \geq 0, i, j \in [n] \\ & \mathbf{Z1} = \frac{n}{r}\mathbf{1}, \end{aligned} \tag{35}$$

where $\mathbf{Z}^* = \sum_{k=1}^{r} \xi_k^* (\xi_k^*)^T$, and ξ_k^* is a binary vector that is an indicator function for community k, such that $\xi_k(i) = 1$ if vertex i is in community k and $\xi_k(i) = 0$, otherwise. It has been shown that if $\sqrt{a} - \sqrt{b} > \sqrt{r}$, then $\Pr(\hat{\mathbf{Z}}_{\text{SDP}} = \mathbf{Z}^*) = 1 - o(1)$.

E. Proof of Theorem 3.5 (Threshold condition for $\mathcal{M}^{\textbf{SDP}}_{\text{Stability}}(G)$ for $r \geq 2$)

Lemma E.1. *Given any function $f : \mathcal{G} \to \mathcal{R}$, the f-based Stability algorithm $\mathcal{M}^f_{\text{Stability}}$ with $\delta = n^{-t}$ for any positive t outputs $f(G)$ with probability at least $1 - \mathcal{O}(n^{-1})$, if G is $\frac{(t+1)\log n}{\epsilon}$-stable under f with probability at least $1 - \mathcal{O}(n^{-1})$.*

Proof. Because G is $\frac{(t+1)\log n}{\epsilon}$-stable under f with probability at least $1 - \mathcal{O}(n^{-1})$, $d(G) \geq \frac{(t+1)\log n}{\epsilon}$ with probability at least $1 - \mathcal{O}(n^{-1})$. We drop the parameter (G) when the context is clear. The probability that $\mathcal{M}^f_{Stability}(G)$ does not output $f(G)$ is:

$$\Pr\left[\mathcal{M}^f_{\text{Stability}}(G) \neq f(G)\right] = \Pr\left[\tilde{d} \leq \frac{\log 1/\delta}{\epsilon}\right] \tag{36}$$

$$= \Pr\left[d + \text{Lap}(1/\epsilon) \leq \frac{\log 1/\delta}{\epsilon}\right] \tag{37}$$

$$= \Pr\left[\text{Lap}(1/\epsilon) \leq \frac{\log 1/\delta}{\epsilon} - d\right] \tag{38}$$

$$\begin{aligned} &= \Pr\left[\text{Lap}(1/\epsilon) \leq \frac{\log 1/\delta}{\epsilon} - d \middle| d > \frac{(t+1)\log n}{\epsilon}\right] \Pr\left[d > \frac{(t+1)\log n}{\epsilon}\right] \\ &\quad + \Pr\left[\text{Lap}(1/\epsilon) \leq \frac{\log 1/\delta}{\epsilon} - d \middle| d \leq \frac{(t+1)\log n}{\epsilon}\right] \Pr\left[d \leq \frac{(t+1)\log n}{\epsilon}\right] \end{aligned} \tag{39}$$

$$\leq \Pr\left[\text{Lap}(1/\epsilon) \leq \frac{\log 1/\delta - (t+1)\log n}{\epsilon}\right] \Pr\left[d > \frac{(t+1)\log n}{\epsilon}\right] + \Pr\left[d < \frac{(t+1)\log n}{\epsilon}\right] \tag{40}$$

$$\tag{41}$$

$$\leq \Pr\left[\mathrm{Lap}(1/\epsilon) \leq \frac{t\log n - (t+1)\log n}{\epsilon}\right] + \mathcal{O}(n^{-1}) \tag{42}$$

$$\leq \Pr\left[\mathrm{Lap}(1/\epsilon) \leq \frac{-\log n}{\epsilon}\right] + \mathcal{O}(n^{-1}) \tag{43}$$

$$\leq \Pr\left[|\mathrm{Lap}(1/\epsilon)| \geq \frac{\log n}{\epsilon}\right] + \mathcal{O}(n^{-1}) \tag{44}$$

$$\leq n^{-1} + \mathcal{O}(n^{-1}) \tag{45}$$

$$= \mathcal{O}(n^{-1}) \tag{46}$$

Finally, we have that $\mathcal{M}^{f}_{\text{Stability}}(G)$ outputs $f(G)$ with probability at least $1 - \mathcal{O}(n^{-1})$. □

Proving the optimality of the SDP based algorithm is very challenging; (Hajek et al., 2016a) use a sophisticated dual certificate, and use it to show that the SDP solution is optimal, with high probability (Theorem 4 of (Hajek et al., 2016b)). However, this probability (which is $1 - 1/n^{O(1)}$) is not high enough for a union bound to ensure stability for all graphs within $\mathcal{O}(\log n)$ distance. Our main technical contribution in this analysis is a slightly different certificate, which ensures that the SDP solution is optimal with probability **1**; we refer to this certificate as "concentration".

Definition E.2. *A graph G is called (c_1, c_2, c_3, c_4)-concentrated if it satisfies all four (4) conditions below*

- $min_{i \in V(G)}(s_i - r_i) > c_1 \log n$
- $\|A - \mathbb{E}[A]\|_2 \leq c_2\sqrt{\log n}$
- $max_{k\in[r]} \frac{1}{K}\sum_{i\in C_k} r_i \leq Kq + c_3\sqrt{\log n}$, *where* $K = n/r$,
- $e(C_k, C_{k'}) \geq K^2 q - 3/4K\sqrt{\log n} - c_4 \log n$

in which s_i is the number of same-community neighbors of node i and r_i is the maximum number of neighbors of i in one of the other communities; $e(C_k, C_{k'})$ is the number of inter-community edges between communities k and $k' \neq k$.

We note that the bound in the first condition is stronger than the one in Lemma 4 of (Hajek et al., 2016b).

Next we prove that in some regimes of the SBM, the concentration of a graph G generated by the SBM holds with high probability. (Hajek et al., 2016b) proves the second and the third conditions holds with probability at least $1 - 1/poly(n)$. We will prove that the first and the last condition will hold with probability at least $1 - O(1/n)$ to complete the Lemma.

Lemma E.3. *A graph G generated by an SBM is (c_1, c_2, c_3, c_4)-concentrated with probability at least $1 - \mathcal{O}(n^{-1})$ for some constants (c_1, c_2, c_3, c_4) with $c_1 > 0$, if there exists a constant $\gamma \in (0, 1)$ such that $\frac{a-b}{r} > \frac{4}{\gamma^2}$.*

Proof. The second and the third condition have been proved to hold with high probability by (Hajek et al., 2016a). We need to prove the first and last conditions will hold with high probability to complete the lemma.

The first condition's proof. Let $r_i(k)$ be the number of cross-community neighbors of node i in community k (i is not in community k). We have $\forall k : r_i(k) \leq r_i$.

For any node i and community k that does not contain i:

$$\mathbb{E}[s_i - r_i(k)] = \frac{n}{r}p - \frac{n}{r}q \tag{47}$$

$$= \frac{(a-b)n\log n}{rn} \tag{48}$$

$$= \frac{(a-b)}{r}\log n \tag{49}$$

$$\tag{50}$$

There exists a constant $\gamma \in (0,1)$ such that $\frac{a-b}{r} > \frac{4}{\gamma^2}$ and:

$$\Pr\left[s_i - r_i(k) < (1-\gamma)\frac{a-b}{r}\log n\right] \leq \exp(-\frac{\gamma^2(a-b)\log n}{2r}) \tag{51}$$

$$\leq n^{-2} \tag{52}$$

Taking the union bound over all r communities, we have:

$$\Pr\left[s_i - r_i < (1-\gamma)\frac{a-b}{r}\log n\right] \leq rn^{-2} \tag{53}$$

Taking the union bound on all node i, the lemma follows that $\Pr\left[\forall i \in V : s_i - r_i < c_1 \log n\right] < rn^{-1}$ and $c_1 = (1-\gamma)\frac{a-b}{r}$.

The last condition's proof. We first notice that $e(C_k, C_{k'}) \sim Binom(K^2, q)$, since there are K^2 pairs of nodes and the probability of edges between each pair is q. By Chernoff's bound, we have:

$$\Pr\left[e(C_k, C_{k'}) < (1-\alpha)K^2 q\right] \leq e^{-\alpha^2 K^2 q/2} \tag{54}$$

Set $\alpha = \frac{(3/4)K\sqrt{\log n} + c_4 \log n}{K^2 q}$. We notice that $|c_4 \log n| \ll (K/4)\sqrt{\log n}$, therefore $\alpha \geq \frac{(K/2)\sqrt{\log n}}{K^2 q} = \frac{\sqrt{\log n}}{2Kq}$

Therefore we have:

$$\Pr\left[e(C_k, C_{k'}) < \left(1 - \frac{(3/4)K\sqrt{\log n} + c_4 \log n}{K^2 q}\right) K^2 q\right] \leq e^{-\left(\frac{(3/4)K\sqrt{\log n} + c_4 \log n}{K^2 q}\right)^2 K^2 q/2} \tag{55}$$

$$\Pr\left[e(C_k, C_{k'}) < K^2 q - (3/4)K\sqrt{\log n} - c_4 \log n\right] \leq e^{-(\frac{\sqrt{\log n}}{2Kq})^2 K^2 q/2} \tag{56}$$

$$= e^{-\frac{\log n}{8q}} \tag{57}$$

$$= n^{-\frac{1}{8q}} \tag{58}$$

Taking the union bound over all k and k', we have that $e(C_k, C_{k'}) \geq K^2 q - (3/4)K\sqrt{\log n} - c_4 \log n$ with probability at least $1 - r^2 n^{-\mathcal{O}(\frac{n}{\log n})} > 1 - \mathcal{O}(n^{-1})$.

Taking the union bound over all four conditions, the Lemma follows. □

Next we prove that the concentration persists under $\Omega(\log n)$ edge perturbations. Specifically, we prove that if a graph is concentrated, a graph obtained by flipping up to $\Omega(\log n)$ connections of the original one is also concentrated, albeit with slightly different tuple of constants.

Lemma E.4. *If a graph G is (c_1, c_2, c_3, c_4)-concentrated, all graphs G' at distance at most $c \log n/\epsilon$ are (c_1', c_2', c_3', c_4')-concentrated with $c_1' = c_1 - c/\epsilon, c_2' = c_2 + \sqrt{2c/\epsilon}, c_3' = c_3 + c/\epsilon, c_4' = c_4 + c/\epsilon$.*

Proof. For the first condition:

$$\min_{i \in V(G)} (s_i - r_i) \geq \min_{i \in V(G)} \left(s_i - r_i - \frac{c \log n}{\epsilon}\right) \tag{59}$$

$$\geq \min_{i \in V(G)} (s_i - r_i) - \frac{c \log n}{\epsilon} \tag{60}$$

$$\geq (c_1 - c/\epsilon)\log n \tag{61}$$

$$= c_1' \log n \tag{62}$$

Second, let $\mathbf{A}'$ be the adjacency matrix of graph G'. We have $\mathbb{E}[\mathbf{A}'] = \mathbb{E}[\mathbf{A}] = \bar{\mathbf{A}}$ with the assumption that both G and G' are generated by the same SBM. We have:

$$\begin{aligned}
\|\mathbf{A}' - \bar{\mathbf{A}}\|_2 &\leq \|\mathbf{A} + (\mathbf{A}' - \mathbf{A}) - \bar{\mathbf{A}}\|_2 && (63)\\
&\leq \|\mathbf{A} - \bar{\mathbf{A}}\|_2 + \|\mathbf{A}' - \mathbf{A}\|_2 && (64)\\
&\leq c_2\sqrt{\log n} + \|\mathbf{A}' - \mathbf{A}\|_F && (65)\\
&\leq c_2\sqrt{\log n} + \sqrt{\frac{2c\log n}{\epsilon}} && (66)\\
&= (c_2 + \sqrt{2c/\epsilon})\sqrt{\log n} && (67)\\
&= c_2'\sqrt{\log n} && (68)
\end{aligned}$$

Third, the third condition is:

$$\begin{aligned}
\max_{k\in[r]} \frac{1}{K}\sum_{i\in C_k} r_i' &\leq \max_{k\in[r]}(\frac{1}{K}\sum_{i\in C_k} r_i + \frac{c\log n}{\epsilon}) && (69)\\
&\leq \max_{k\in[r]} \frac{1}{K}\sum_{i\in C_k} r_i + \frac{1}{K}\frac{c\log n}{\epsilon} && (70)\\
&\leq Kq + c_3\sqrt{\log n} + \frac{cr\log n}{n\epsilon} && (71)\\
&\leq Kq + (c_3 + c/\epsilon)\sqrt{\log n} && (72)\\
&= Kq + c_3'\sqrt{\log n} && (73)
\end{aligned}$$

Finally, the last condition is:

$$\begin{aligned}
e(C_k', C_{k'}') &\geq e(C_k, C_{k'}) - \frac{c\log n}{\epsilon} && (74)\\
&= K^2 q - (3/4)K\sqrt{\log n} - c_4\log n - \frac{c\log n}{\epsilon} && (75)\\
&= K^2 q - (3/4)K\sqrt{\log n} - (c_4 + c/\epsilon)\log n && (76)\\
&= K^2 q - (3/4)K\sqrt{\log n} - c_4'\log n && (77)
\end{aligned}$$

□

We follow (Hajek et al., 2016b) to prove that when a graph is (c_1, c_2, c_3, c_4)-concentrated for some constants c_i, the SDP relaxation (SDP for short) outputs the optimal ground truth community vector. First we note that the SDP can be presented by the following form:

$$\begin{aligned}
\text{maximize } &\langle \mathbf{A}, \mathbf{Z}\rangle && (78)\\
\text{subsect to } &\mathbf{Z} \succcurlyeq 0 && (79)\\
&Z_{ii} = 1, \forall i \in [n] && (80)\\
&Z_{ij} \geq 0, \forall i, j \in [n] && (81)\\
&\mathbf{Z}\mathbf{1} = K\mathbf{1}, && (82)
\end{aligned}$$

with $K = n/r$ is the size of each community.

Then the following Lemma provides the condition for a dual certificate (deterministically). Intuitively, if we can construct a positive semi-definite matrix S^* by the following formula without violating the two constraints below, the SDP is uniquely optimal at $\mathbf{Z}^*$ constructed by the ground truth community label (We say SDP(G) is optimal at the ground truth community label for short).

Lemma E.5. ***Lemma 6 of (Hajek et al., 2016b).*** *Suppose there exists $\mathbf{D}^* = diag(d_i^*)$ with $d_i^* > 0$ for all $i, \mathbf{B}^* \in \mathcal{S}^n$ with $\mathbf{B}^* \geq \mathbf{0}$ and $\mathbf{B^*}_{ij} > 0$ whenever i and j are in distinct clusters, and $\lambda^* \in \mathbb{R}^n$ such that $\mathbf{S}^* \triangleq \mathbf{D}^* - \mathbf{B}^* - \mathbf{A} + \lambda^* \mathbf{1}^T + \mathbf{1}(\lambda^*)^T$ satisfies $\mathbf{S}^* \succcurlyeq 0$ and*

$$\mathbf{S}^* \xi_k^* = 0, \forall k \in [r] \tag{83}$$
$$B_{ij}^* Z_{ij}^* = 0, \forall i, j \in [n] \tag{84}$$

Then SDP$(G) = \mathbf{Z}^*$ is the unique solution for the SDP.

In the following statement, we claim that when the concentration holds, the SDP outputs the uniquely optimal ground truth community deterministically. We closely follow the proof of Theorem 4 of (Hajek et al., 2016b) but replacing their high probability bounds by our concentration conditions. We prove that the concentration of the input graph implies the existence of a positive semi-definite matrix S^* satisfies Lemma E.5.

Lemma E.6. *When a graph G is (c_1, c_2, c_3, c_4)-concentrated for some constants c_j and $c_1 > 0$, the SDP outputs optimal ground truth community.*

Proof. By the result of Lemma E.5, we will prove that the exists a $\mathbf{S}^* \succcurlyeq 0$ and satisfies the two contraints above the Lemma follows. The main idea is to specify a way to construct $\mathbf{B}^*$, $\mathbf{D}^*$, and λ^* that satisfy all properties above. Theorem 4 of (Hajek et al., 2016b) defines $\mathbf{B}^*$, $\mathbf{D}^*$, and λ^* as follows:

$$u_{kk'} = \frac{1}{2K}\left(\frac{e(C_k, C_{k'})}{K} - Kq + \sqrt{\log n}\right) \tag{85}$$
$$y_{kk'}^*(i) = \frac{1}{K}(r_i - e(i, C_{k'})) + u_{kk'} \tag{86}$$
$$z_{kk'}^*(j) = \frac{1}{K}(r_j - e(j, C_k)) + u_{kk'} \tag{87}$$
$$\mathbf{B}^*_{\mathbf{C_k} \times \mathbf{C_{k'}}}(i, j) = y_{kk'}^*(i) + z_{kk'}^*(j), \forall, 1 \leq k \leq k' \leq r \tag{88}$$
$$\alpha_k = \frac{1}{2}(Kq - \sqrt{\log n}) \tag{89}$$
$$d_i^* = e(i, C_k) - r_i + 2\alpha_k - \frac{1}{K}\sum_{i \in C_k} r_i \tag{90}$$
$$\lambda_i^* = \frac{1}{K}(r_i - \alpha_k)\text{ , for } i \in C_k \tag{91}$$

We follow the proof of Theorem 4 of (Hajek et al., 2016b) to prove $\mathbf{S}^* \succcurlyeq 0$. We note that the main difference between our proof and Theorem 4 of (Hajek et al., 2016b) is that our proof proves the statement (about $\mathbf{S}^*$) is always true under the concentration condition while (Hajek et al., 2016b) proves the statement is true with high probability over the SBM. We also utilize a different bound on $\min_{i \in V}(s_i - r_i)$ from (Hajek et al., 2016b) to tolerate the change of up to $c \log n/\epsilon$ connections later.

Let E be the subspace spanned by vectors $\{\xi_k^*\}_k \in [r]$, i.e., $E = \text{span}(\xi_k^* : k \in [r])$. We show that

$$x^T \mathbf{S}^* x > 0 \; \forall x \perp E, \|x\|_2 = 1 \tag{92}$$

when the input graph G is (c_1, c_2, c_3, c_4)-concentrated and n is large enough.

Note that $\mathbb{E}[\mathbf{A}] = (p - q)\mathbf{Z}^* + q\boldsymbol{J} - p\boldsymbol{I}$ and $\mathbf{Z}^* = \sum_{k \in [r]} \xi_k^* (\xi_k^*)^T$. For any x such that $x \perp E$ and $\|x\|_2 = 1$,

$$x^T\mathbf{S}^*x = x^T\mathbf{D}^*x - x^T\mathbb{E}[\mathbf{A}]x - x^T\mathbf{B}^*x + 2x^T\lambda^*\mathbf{1}^Tx - x^T(\mathbf{A}-\mathbb{E}[\mathbf{A}])x \tag{93}$$
$$= x^T\mathbf{D}^*x - (p-q)x^T\mathbf{Z}^*x - qx^T\mathbf{J}x + p - x^T\mathbf{B}^*x - x^T(\mathbf{A}-\mathbb{E}[\mathbf{A}])x \tag{94}$$
$$= x^T\mathbf{D}^*x + p - x^T\mathbf{B}^*x - x^T(\mathbf{A}-\mathbb{E}[\mathbf{A}])x \tag{95}$$
$$\geq x^T\mathbf{D}^*x + p - x^T\mathbf{B}^*x - \|\mathbf{A}-\mathbb{E}[\mathbf{A}]\|_2 \tag{96}$$
$$\geq x^T\mathbf{D}^*x + p - x^T\mathbf{B}^*x - c_2\sqrt{\log n}, \tag{97}$$

where the second equality holds because $\mathbb{E}[\mathbf{A}] = (p-q)\mathbf{Z}^* + q\boldsymbol{J} - p\boldsymbol{I}$ and $x \perp \mathbf{1}$; and the third equality holds because $\langle x, \xi_k^*\rangle = 0$ and $x \perp \mathbf{1}$.

Theorem 4 of (Hajek et al., 2016b) shows that $\mathbf{B}^*$ can be chosen such that for any $x \perp E$, we have $x^T\mathbf{B}^*x = 0$ where both constraints of Lemma E.5 are satisfied.

From the definition of D^*, we have

$$\lambda_{\min}(D^*) \geq \min_i d_i^* \tag{98}$$
$$= \min_i e(i, C_k) - r_i + 2\alpha_k - \frac{1}{K}\sum_{i\in C_k} r_i \tag{99}$$
$$= \min_i e(i, C_k) - r_i + Kq - \sqrt{\log n} - \frac{1}{K}\sum_{i\in C_k} r_i \tag{100}$$
$$= \min_i e(i, C_k) - r_i - \sqrt{\log n} - \left(\frac{1}{K}\sum_{i\in C_k} r_i - Kq\right) \tag{101}$$
$$\geq \min_i (e(i, C_k) - r_i) - (c_3+1)\sqrt{\log n} \tag{102}$$
$$\geq c_1 \log n - (c_3+1)\sqrt{\log n}, \tag{103}$$

where the second inequality holds because from the third condition of concentration, $\frac{1}{K}\sum_{i\in C_k} r_i - Kq \leq c_3\sqrt{\log n}$; and the last inequality holds because from the first condition of concentration, $e(i, C_k) - r_i) \geq c_1 \log n$. We then have:

$$x^T\mathbf{D}^*x \geq \lambda_{\min}(\mathbf{D}^*)\|x\|_2^2 \tag{104}$$
$$\geq \lambda_{\min}(\mathbf{D}^*) \tag{105}$$
$$\geq \min_i d_i^* \tag{106}$$
$$\geq c_1 \log n - (c_3+1)\sqrt{\log n}. \tag{107}$$

With that, we simplify $x^T\mathbf{S}^*x$ with the assumption that the graph is (c_1, c_2, c_3, c_4)-concentrated:

$$x^T\mathbf{S}^*x \geq c_1 \log n - (c_2+c_3+1)\sqrt{\log n} \tag{108}$$
$$\geq 0 \text{ when } n \text{ is large enough.} \tag{109}$$

Finally, to guarantee that $\mathbf{B}^*{}_{ij} > 0$ whenever i and j are in distinct clusters, we will prove $y_{kk'}^*(i)$ and $z_{kk}^*(j) > 0$. From their definitions, we see that $r_i - e(i, C_{k'}) \geq 0$ and $r_j - e(j, C_k) \geq 0$, therefore we need to prove $u_{kk'} > 0$ for all k and $k' \neq k$ to complete the proof, i.e., to prove $e(C_k, C_{k'}) > K^2q - K\sqrt{\log n}$. The fourth condition of (c_1, c_2, c_3, c_4)-concentration says that there is a constant c_4 such that $e(C_k, C_{k'}) \geq K^2q - (3/4)K\sqrt{\log n} - c_4 \log n$ for all k, k'. Since $c_4 \log n \ll (K/4)\sqrt{\log n}$ for n large enough, it confirms that $e(C_k, C_{k'}) > K^2q - K\sqrt{\log n}$, $\mathbf{B}^*$ satisfies above conditions.

Apply the result of Lemma E.5, the Lemma follows. □

Next we prove that if a graph is appropriately concentrated, it is also stable under the SDP relaxation.

Lemma E.7. *When a graph G is (c_1, c_2, c_3, c_4)-concentrated and $c_1 - c/\epsilon > 0$, it is also $\frac{c \log n}{\epsilon}$-stable.*

Proof. When a graph G is (c_1, c_2, c_3, c_4)-concentrated, all graph G' at distance at most $\frac{c \log n}{\epsilon}$, formally $d(G', G) \leq \frac{c \log n}{\epsilon}$, are (c_1', c_2', c_3', c_4')-concentrated with $c_1' = c_1 - c/\epsilon > 0$ as the result of Lemma E.4. Hence, in the radius of $c \log n/\epsilon$ from G, all graph G' has that $SDP(G')$ is unique and optimal at $\mathbf{Z}^*$ constructed by the ground truth communities with n large enough.

From that, for all graphs G' such that $d(G, G') \leq \frac{c \log n}{\epsilon}$, we have SDP$(G)$ = SDP(G') and the lemma follows. □

Lemma E.8. *A graph G generated by an SBM is $\Omega(\frac{c \log n}{\epsilon})$-stable with respect to SDP function with probability at least $1 - \mathcal{O}(n^{-1})$ if there exists a constant $\gamma \in (0, 1)$ such that $\frac{a-b}{r} \geq \frac{4}{\gamma^2}$ and $(1-\gamma)\frac{a-b}{r} \geq \frac{c}{\epsilon}$*

Proof. By the result of Lemma E.3 and Lemma E.7, the lemma follows. We note that in Lemma E.3, we select $c_1 = (1-\gamma)\frac{a-b}{r}$, hence we need $(1-\gamma)\frac{a-b}{r} > \frac{c}{\epsilon}$ to satisfy the condition of Lemma E.7. □

Finally, we prove that our mechanism outputs the ground-truth community label with high probability if the SBM satisfies two conditions for the stability. The first condition allows the concentration to holds with high probability and the second condition makes the concentration to persist under edge perturbation (of up to $\Omega(\log n)$ connections).

Theorem E.9. *Given a graph G is generated by an r-community SBM model with $\frac{a-b}{r} \geq \frac{4}{\gamma^2}$ and $(1-\gamma)\frac{a-b}{r} \geq \frac{t+1}{\epsilon}$ for some constant $\gamma \in (0, 1)$, $\mathcal{M}^{SDP}_{\text{Stability}}$ with $\delta = n^{-t}$ outputs $\mathbf{Z}^*$ constructed by the ground-truth community vector w.h.p., i.e. $\Pr[\mathcal{M}_{Stability\_SDP}(G) \neq \mathbf{Z}^*] = o(1)$.*

Proof. Lemma E.8 states that a graph G generated by the r-community SBM is $(c \log n/\epsilon)$-stable under SDP w.h.p.. By applying Lemma E.1, substituting the generic function f by SDP and $c = t + 1$, $\mathcal{M}^{SDP}_{\text{Stability}}$ outputs $SDP(G)$ w.h.p.. Given that $\Pr[SDP(G) \neq \mathbf{Z}^*] = o(1)$(Hajek et al., 2016a), the Theorem follows. Finally, setting $C = \gamma/2$ and using $\sqrt{a-b} > \sqrt{a} - \sqrt{b}$,we loosen the two conditions above as:

$$\sqrt{a} - \sqrt{b} \geq \frac{\sqrt{r}}{C} \text{ and} \tag{110}$$

$$\sqrt{a} - \sqrt{b} \geq \sqrt{\frac{(t+1)r}{(1-2C)\epsilon}} \tag{111}$$

Set $C = 1/4$, we (loosely) simplify the conditions above to:

$$\sqrt{a} - \sqrt{b} \geq \sqrt{r} \times 4 \times \left(1 + \frac{\sqrt{t+1}}{\sqrt{2\epsilon}}\right). \tag{112}$$

. □

F. Proof of Theorem 3.6 (Threshold condition for $\mathcal{M}_{\text{Bayesian}}(G)$ for $r = 2$)

We first prove that $\mathcal{M}_{\text{Bayesian}}(G)$ satisfies ϵ-edge DP for $\epsilon \geq \log(a/b)$. For a fixed graph G, w.l.o.g., let us assume $G = G' + e$, where e is an edge. We define $E_{\text{intra}}(G, \boldsymbol{\sigma})$ as the set of same-community edges in graph G and $E_{\text{inter}}(G, \boldsymbol{\sigma})$ is the set of cross-community edges of graph G, with respect to a labeling $\boldsymbol{\sigma}$. We analyze two cases: (1) $e \in E_{\text{intra}}(G, \boldsymbol{\sigma})$ (1) and $e \in E_{\text{inter}}(G, \boldsymbol{\sigma})$. For each of them, we analyze the likelihood ratio of releasing a label vector $\boldsymbol{\sigma}$ if the input graph is G or G'. We start with the first case:

- **Case** 1**:** $e \in E_{\textbf{intra}}(G, \boldsymbol{\sigma})$

$$\begin{aligned} R_1 &= \frac{\Pr(\boldsymbol{\sigma}|G')}{\Pr(\boldsymbol{\sigma}|G)} = \frac{\Pr(G'|\boldsymbol{\sigma})}{\Pr(G|\boldsymbol{\sigma})} \times \frac{\Pr(G)}{\Pr(G')} \\ &= \frac{1-p}{p} \times \frac{\Pr(G)}{\Pr(G')} \\ &= \frac{1-p}{p} \times \frac{\sum_{\boldsymbol{\sigma}'} \Pr(\boldsymbol{\sigma}') \Pr(G|\boldsymbol{\sigma}')}{\sum_{\boldsymbol{\sigma}'} \Pr(\boldsymbol{\sigma}') \Pr(G'|\boldsymbol{\sigma}')} \\ &\overset{(a)}{=} \frac{1-p}{p} \times \frac{\sum_{\boldsymbol{\sigma}'} \Pr(\boldsymbol{\sigma}') \Pr(G'|\boldsymbol{\sigma}') \times \frac{p}{1-p}}{\sum_{\boldsymbol{\sigma}'} \Pr(\boldsymbol{\sigma}') \Pr(G'|\boldsymbol{\sigma}')} = 1, \end{aligned} \tag{113}$$

where step (a) follows from the fact that $\Pr(G|\boldsymbol{\sigma}) = \frac{p}{1-p} \times \Pr(G'|\boldsymbol{\sigma})$. Note that, the distribution $p(G|\sigma)$ is given as

$$\begin{aligned} p(G|\sigma) &= \prod_{i<j} p(A_{i,j}|\sigma) \overset{(a)}{=} \prod_{i<j} p(A_{i,j}|\sigma_i, \sigma_j) \\ &\overset{(b)}{=} \prod_{i<j} \left[\frac{P(A_{i,j}) + Q(A_{i,j})}{2} + \frac{P(A_{i,j}) - Q(A_{i,j})}{2} \sigma_i \sigma_j \right], \end{aligned}$$

where in step (a), the random variable $A_{i,j}$ only depends on the labels σ_i and σ_j. In step (b), we have $P(A_{i,j}) = p^{A_{i,j}}(1-p)^{1-A_{i,j}}$, and $Q(A_{i,j}) = q^{A_{i,j}}(1-q)^{1-A_{i,j}}$, where $A_{i,j} \in \{0, 1\}$. The distribution $p(\sigma) = (\frac{1}{r})^n$, while $p(G) = \sum_{\sigma'} p(G|\sigma')p(\sigma') = (\frac{1}{r})^n \sum_{\sigma'} \prod_{i<j} \left[\frac{P(A_{i,j})+Q(A_{i,j})}{2} + \frac{P(A_{i,j})-Q(A_{i,j})}{2} \sigma'_i \sigma'_j \right]$.

We next analyze the other ratio as follows.

$$\begin{aligned} R_2 &= \frac{\Pr(\boldsymbol{\sigma}|G)}{\Pr(\boldsymbol{\sigma}|G')} = \frac{\Pr(G|\boldsymbol{\sigma})}{\Pr(G'|\boldsymbol{\sigma})} \times \frac{\Pr(G')}{\Pr(G)} \\ &= \frac{p}{1-p} \times \frac{\Pr(G')}{\Pr(G)} \\ &= \frac{p}{1-p} \times \frac{\sum_{\boldsymbol{\sigma}'} \Pr(\boldsymbol{\sigma}') \Pr(G'|\boldsymbol{\sigma}')}{\sum_{\boldsymbol{\sigma}'} \Pr(\boldsymbol{\sigma}') \Pr(G|\boldsymbol{\sigma}')} \\ &= \frac{p}{1-p} \times \left[\frac{\sum_{\boldsymbol{\sigma}': e \in E_{\text{intra}}(\boldsymbol{\sigma}')} \Pr(\boldsymbol{\sigma}') \Pr(G'|\boldsymbol{\sigma}') + \sum_{\boldsymbol{\sigma}': e \in E_{\text{inter}}(L')} \Pr(\boldsymbol{\sigma}') \Pr(G'|\boldsymbol{\sigma}')}{\sum_{\boldsymbol{\sigma}'} \Pr(\boldsymbol{\sigma}') \Pr(G|\boldsymbol{\sigma}')} \right] \\ &= \frac{p}{1-p} \times \left[\frac{\sum_{\boldsymbol{\sigma}': e \in E_{\text{intra}}(\boldsymbol{\sigma}')} \Pr(\boldsymbol{\sigma}') \Pr(G|\boldsymbol{\sigma}') \times \frac{1-p}{p} + \sum_{\boldsymbol{\sigma}': e \in E_{\text{inter}}(\boldsymbol{\sigma}')} \Pr(\boldsymbol{\sigma}') \Pr(G|\boldsymbol{\sigma}') \times \frac{1-q}{q}}{\sum_{\boldsymbol{\sigma}'} \Pr(\boldsymbol{\sigma}') \Pr(G|\boldsymbol{\sigma}')} \right] \\ &\overset{(a)}{\leq} \frac{p}{1-p} \times \left[\frac{\sum_{\boldsymbol{\sigma}': e \in E_{\text{intra}}(\boldsymbol{\sigma}')} \Pr(\boldsymbol{\sigma}') \Pr(G|\boldsymbol{\sigma}') \times \frac{1-q}{q} + \sum_{\boldsymbol{\sigma}': e \in E_{\text{inter}}(\boldsymbol{\sigma}')} \Pr(\boldsymbol{\sigma}') \Pr(G|\boldsymbol{\sigma}') \times \frac{1-q}{q}}{\sum_{\boldsymbol{\sigma}'} \Pr(\boldsymbol{\sigma}') \Pr(G|\boldsymbol{\sigma}')} \right] \\ &= \frac{p(1-q)}{q(1-p)} = e^{\epsilon_0}, \end{aligned} \tag{114}$$

where step (a) follows that $\frac{1-q}{q} > \frac{1-p}{p}$ for $p > q$. Now, we analyze the second case as follows.

- **Case 2:** $e \in E_{\textbf{inter}}(G, \boldsymbol{\sigma})$

$$
\begin{aligned}
R_1 &= \frac{\Pr(\boldsymbol{\sigma}|G')}{\Pr(\boldsymbol{\sigma}|G)} \\
&= \frac{1-q}{q} \times \frac{\sum_{\boldsymbol{\sigma}'} \Pr(\boldsymbol{\sigma}') \Pr(G|\boldsymbol{\sigma}')}{\sum_{\boldsymbol{\sigma}'} \Pr(\boldsymbol{\sigma}') \Pr(G'|\boldsymbol{\sigma}')} \\
&\leq \frac{1-q}{q} \times \frac{\sum_{\boldsymbol{\sigma}'} \Pr(\boldsymbol{\sigma}') \Pr(G'|\boldsymbol{\sigma}') \times \frac{p}{1-p}}{\sum_{\boldsymbol{\sigma}'} \Pr(\boldsymbol{\sigma}') \Pr(G'|\boldsymbol{\sigma}')} \\
&= \frac{p(1-q)}{q(1-p)}.
\end{aligned} \tag{115}
$$

We next analyze the other ratio as follows:

$$
\begin{aligned}
R_2 &= \frac{\Pr(\boldsymbol{\sigma}|G)}{\Pr(\boldsymbol{\sigma}|G')} \\
&= \frac{q}{1-q} \times \frac{\sum_{\boldsymbol{\sigma}'} \Pr(\boldsymbol{\sigma}') \Pr(G'|\boldsymbol{\sigma}')}{\sum_{\boldsymbol{\sigma}'} \Pr(\boldsymbol{\sigma}'] \Pr(G|\boldsymbol{\sigma}')} \\
&= \frac{q}{1-q} \times \left[\frac{\sum_{\boldsymbol{\sigma}':e\in E_{\text{intra}}(\boldsymbol{\sigma}')} \Pr(\boldsymbol{\sigma}') \Pr(G'|\boldsymbol{\sigma}') + \sum_{\boldsymbol{\sigma}':e\in E_{\text{inter}}(\boldsymbol{\sigma}')} \Pr(\boldsymbol{\sigma}') \Pr(G'|\boldsymbol{\sigma}')}{\sum_{\boldsymbol{\sigma}'} \Pr(\boldsymbol{\sigma}') \Pr(G|\boldsymbol{\sigma}')} \right] \\
&= \frac{q}{1-q} \times \left[\frac{\sum_{\boldsymbol{\sigma}':e\in E_{\text{intra}}(\boldsymbol{\sigma}')} \Pr(\boldsymbol{\sigma}') \Pr(G|\boldsymbol{\sigma}') \times \frac{1-p}{p} + \sum_{\boldsymbol{\sigma}':e\in E_{\text{inter}}(\boldsymbol{\sigma}')} \Pr(\boldsymbol{\sigma}') \Pr(G|\boldsymbol{\sigma}') \times \frac{1-q}{q}}{\sum_{\boldsymbol{\sigma}'} \Pr(\boldsymbol{\sigma}') \Pr(G|\boldsymbol{\sigma}')} \right] \\
&\leq \frac{q}{1-q} \times \left[\frac{\sum_{\boldsymbol{\sigma}':e\in E_{\text{intra}}(\boldsymbol{\sigma}')} \Pr(\boldsymbol{\sigma}') \Pr(G|\boldsymbol{\sigma}') \times \frac{1-q}{q} + \sum_{\boldsymbol{\sigma}':e\in E_{\text{inter}}(\boldsymbol{\sigma}')} \Pr(\boldsymbol{\sigma}') \Pr(G|\boldsymbol{\sigma}') \times \frac{1-q}{q}}{\sum_{\boldsymbol{\sigma}'} \Pr(\boldsymbol{\sigma}') \Pr(G|\boldsymbol{\sigma}')} \right] = 1.
\end{aligned} \tag{116}
$$

From the above cases, we conclude that the Bayesian sampling mechanism satisfies ϵ-edge DP for all $\epsilon \geq \log(\frac{p}{q}) + \log(\frac{1-q}{1-p}) \geq \log(\frac{p}{q}) = \log(\frac{a}{b}) = \epsilon_0$.

We next analyze the error probability of the Bayesian mechanism. For a fixed graph G, our goal is to show that

$$
\begin{aligned}
\frac{\Pr(\hat{\boldsymbol{\sigma}}_{\text{Bayesian}}(G) \neq \boldsymbol{\sigma}^*)}{\Pr(\boldsymbol{\sigma}^*|G)} &= \frac{\sum_{\boldsymbol{\sigma}\neq\boldsymbol{\sigma}^*} \Pr(\boldsymbol{\sigma}|G)}{\Pr(\boldsymbol{\sigma}^*|G)} \\
&= \frac{\sum_{\boldsymbol{\sigma}\neq\boldsymbol{\sigma}^*} \Pr(G|\boldsymbol{\sigma})}{\Pr(G|\boldsymbol{\sigma}^*)} \leq o(1)
\end{aligned} \tag{117}
$$

which is equivalent to $\Pr(\hat{\boldsymbol{\sigma}}_{\text{Bayesian}}(G) \neq \boldsymbol{\sigma}^*) \leq o(1)$. Given the ground truth partitions (R^*, B^*), let us define the following variables:

$$
\begin{aligned}
m_1 &\triangleq E(R^* \backslash S_1, S_1), \\
m_2 &\triangleq E(B^* \backslash S_2, S_1), \\
m_3 &\triangleq E(B^* \backslash S_2, S_2), \\
m_4 &\triangleq E(R^* \backslash S_1, S_2),
\end{aligned} \tag{118}
$$

where $S_1 \subseteq R^*$, $S_2 \subseteq B^*$ that represent the mis-classified nodes in both communities (R, B). Given S_1, S_2 both of size k, $k \in [1, \frac{n}{2}]$, we have

$$
R^{(k)} \triangleq \frac{\Pr(G|\boldsymbol{\sigma}, |S_1| = |S_2| = k)}{\Pr(G|\boldsymbol{\sigma^*}, |S_1| = |S_2| = k)} = \underbrace{\left(\frac{q}{p}\right)^{m_1} \left(\frac{1-q}{1-p}\right)^{k(n-k)-m_1} \times \left(\frac{p}{q}\right)^{m_2} \left(\frac{1-p}{1-q}\right)^{k(n-k)-m_2}}_{=P_1^{(k)}} \tag{119}
$$

$$
\times \underbrace{\left(\frac{q}{p}\right)^{m_3} \left(\frac{1-q}{1-p}\right)^{k(n-k)-m_3} \times \left(\frac{p}{q}\right)^{m_4} \left(\frac{1-p}{1-q}\right)^{k(n-k)-m_4}}_{=P_2^{(k)}}. \tag{120}
$$

Note that,

$$\frac{\Pr(\hat{\boldsymbol{\sigma}}_{\text{Bayesian}}(G) \neq \boldsymbol{\sigma}^*)}{\Pr(\boldsymbol{\sigma}^*|G)} \leq \sum_{k=1}^{\frac{n}{2}} \binom{n}{k}^2 R^{(k)}. \tag{121}$$

In order to bound the above ratio, we will first bound $R^{(k)}$ individually. To this end, let us define $\alpha = \frac{p}{q}, \beta = \frac{1-q}{1-p}$. We now simplify each term individually (i.e., (119) & (120)) as follows:

$$\begin{aligned}
P_1^{(k)} &= \alpha^{-m_1} \times \beta^{k(n-k)-m_1} \times \alpha^{m_2} \times \beta^{-(k(n-k)-m_2)} \\
&= \alpha^{m_2-m_1} \times \beta^{m_2-m_1} \\
&= \left(\frac{p}{q} \times \frac{1-q}{1-p}\right)^{m_2-m_1} \\
&= \left(\frac{q}{p} \times \frac{1-p}{1-q}\right)^{m_1-m_2} \\
&\overset{(a)}{=} (1-x)^{m_1-m_2}, \\
&\overset{(b)}{\leq} \exp\left[-x(m_1-m_2)\right],
\end{aligned} \tag{122}$$

where $x \triangleq 1 - \frac{q(1-p)}{p(1-q)}$ where $x \leq 1$. Step (b) follows that $(1-x) \leq e^{-x}, \forall x > 0$. Similarly, we have

$$P_2^{(k)} \leq \exp\left[-x(m_3-m_4)\right]. \tag{123}$$

Therefore, we have

$$\begin{aligned}
R^{(k)} &\leq \exp\left[-x(m_1+m_3-(m_2+m_4))\right] \\
&= \exp\left[-x(\tilde{m}_1^{(k)} - \tilde{m}_2^{(k)})\right],
\end{aligned} \tag{124}$$

where $\tilde{m}_1^{(k)} \sim \mathrm{Bin}(2k(n-k), p)$, and $\tilde{m}_2^{(k)} \sim \mathrm{Bin}(2k(n-k), q)$. For all $k \in [1 : \frac{n}{2}]$, we have that w.h.p.

$$\begin{aligned}
\frac{\Pr(\hat{\boldsymbol{\sigma}}_{\text{Bayesian}}(G) \neq \boldsymbol{\sigma}^*)}{\Pr(\boldsymbol{\sigma}^*|G)} &= \sum_{k=1}^{\frac{n}{2}} \binom{n}{k}^2 R^{(k)} \\
&\leq \sum_{k=1}^{\frac{n}{2}} \left(\frac{ne}{k}\right)^{2k} R^{(k)} \\
&= \sum_{k=1}^{\frac{n}{2}} \exp\left[2k(\log(n) - \log(k) + 1) - x(\tilde{m}_1^{(k)} - \tilde{m}_2^{(k)})\right] \\
&= \sum_{k=1}^{\frac{n}{2}} \exp\left[2k\bigg(\log(n) - \log(k) + 1 - \frac{x}{2k}(\tilde{m}_1^{(k)} - \tilde{m}_2^{(k)})\bigg)\right] \\
&= \sum_{k=1}^{\frac{n}{2}} \exp\left[-2k\bigg(-\log(n) + \log(k) - 1 + \frac{x}{2k}(\tilde{m}_1^{(k)} - \tilde{m}_2^{(k)})\bigg)\right] \\
&= \sum_{k=1}^{\frac{n}{2}} \exp\left[-2k\bigg(\log(k) - 1\bigg)\right] \times \exp\left[-2k\bigg(-\log(n) + \frac{x}{2k}(\tilde{m}_1^{(k)} - \tilde{m}_2^{(k)})\bigg)\right] \\
&\leq \sum_{k=1}^{\frac{n}{2}} \exp\left[-2k\bigg(\log(k) - 1\bigg)\right] \qquad (125) \\
&\times \exp\left[-2k\bigg(-\log(n) + \frac{x}{2k} \times (1-\tilde{\delta}) \times 2k(n-k)(a-b)\frac{\log(n)}{n}\bigg)\right] \\
&\leq \sum_{k=1}^{\frac{n}{2}} \exp\left[-2k\bigg(\log(k) - 1\bigg)\right] \times \exp\left[-2k\bigg(-\log(n) + (1-\tilde{\delta}) \times \frac{x}{2} \times (a-b)\log(n)\bigg)\right] \\
&\overset{(a)}{=} \sum_{k=1}^{\frac{n}{2}} \exp\left[-2k\bigg(\log(k) - 1\bigg)\right] \times \exp\left[-2k\bigg(\frac{x}{2} \times (1-\tilde{\delta}) \times (a-b) - 1\bigg)\log(n)\right] = o(1), \qquad (126)
\end{aligned}$$

where in step (a), we have

$$\frac{x}{2}(1-\tilde{\delta})(a-b) > 1 \Rightarrow \sqrt{a} - \sqrt{b} > \frac{\sqrt{2}}{\sqrt{x(1-\tilde{\delta})}}. \qquad (127)$$

To this end, the error probability of the Bayesian mechanism is

$$\begin{aligned}
\Pr(\hat{\boldsymbol{\sigma}}_{\text{Bayesian}}(G) \neq \boldsymbol{\sigma}^*) &= \Pr(\hat{\boldsymbol{\sigma}}_{\text{Bayesian}}(G) \neq \boldsymbol{\sigma}^* | E_{S_1,S_2}) \times \Pr(E_{S_1,S_2}) \\
&\quad + \Pr(\hat{\boldsymbol{\sigma}}_{\text{Bayesian}}(G) \neq \boldsymbol{\sigma}^* | E^c_{S_1,S_2}) \times \Pr(E^c_{S_1,S_2}) \\
&\leq o(1) \times 1 + 1 \times n^{-\delta'}, \qquad (128)
\end{aligned}$$

where $E_{S_1,S_2} \triangleq \{\tilde{m}_1 - \tilde{m}_2 \geq c' \log(n)\}$ and $E^c_{S_1,S_2}$ denotes its complement, and $x = 1 - \frac{b(n-a\log(n))}{a(n-b\log(n))} > 1 - \frac{b}{a}$. In order to make the error probability behave as $o(1)$, we have the following condition on a and b:

$$\sqrt{a} - \sqrt{b} > \sqrt{2} \times \max\left[\frac{\sqrt{2}}{\tilde{\delta}}, \frac{1}{\sqrt{x(1-\tilde{\delta})}}\right]. \qquad (129)$$

Note that we showed the Bayesian mechanism satisfies $\epsilon \geq \epsilon_0 = \log\left(\frac{a}{b}\right)$, therefore, we have

$$\frac{b}{a} = e^{-\epsilon_0} \qquad (130)$$

We pick $\tilde{\delta}$ as $\tilde{\delta} = (\sqrt{2}-1)\epsilon$ so that the lower bound in (129) is minimized when the two arguments are equal. Plugging the value of $\tilde{\delta}$ into (129) yields the following

$$\sqrt{a} - \sqrt{b} > \frac{2}{(1-e^{-\epsilon_0})(\sqrt{2}-1)}. \tag{131}$$

This completes the proof of Theorem 3.6.

G. Proof of Theorem 3.7 (Threshold condition for $\mathcal{M}_{\text{Expo.}}(G)$ for $r = 2$)

The privacy analysis of $\mathcal{M}_{\text{Expo.}}(G)$ is straightforward and follows on similar lines as in (Dwork et al., 2006).

We next analyze the error probability of $\mathcal{M}_{\text{Expo.}}(G)$. The error probability analysis follows on similar lines as the Bayesian mechanism.

$$\begin{aligned}
R^{(k)} &= \frac{\exp(-\epsilon \times E_{\text{inter}}(G, \boldsymbol{\sigma}))}{\exp(-\epsilon \times E_{\text{inter}}(G, \boldsymbol{\sigma}^*))} \\
&= \frac{\exp(-\epsilon \times (E_{\text{inter}}(G, \boldsymbol{\sigma}^*) + m_1 + m_3 - m_2 - m_4))}{\exp(-\epsilon \times E_{\text{inter}}(G, \boldsymbol{\sigma}^*))} \\
&= \frac{\exp(-\epsilon \times E_{\text{inter}}(G, \boldsymbol{\sigma}^*)) \times \exp(-\epsilon(m_1 + m_3 - m_2 - m_4))}{\exp(-\epsilon \times E_{\text{inter}}(G, \boldsymbol{\sigma}^*))} \\
&= \exp(-\epsilon(m_1 + m_3 - m_2 - m_4)) \\
&= \exp(-\epsilon(\tilde{m}_1^{(k)} - \tilde{m}_2^{(k)})),
\end{aligned} \tag{132}$$

where $\tilde{m}_1^{(k)} \sim \text{Bin}(2k(n-k), p)$, and $\tilde{m}_2^{(k)} \sim \text{Bin}(2k(n-k), q)$. Now, we have

$$\frac{\Pr(\hat{\boldsymbol{\sigma}}_{\text{Expo.}}(G) \neq \boldsymbol{\sigma}^*)}{\Pr(\boldsymbol{\sigma}^*|G)} \leq \sum_{k=1}^{n/2} \binom{n}{k}^2 R^{(k)}. \tag{133}$$

Similarly, in order to make the error probability behaves as $o(1)$, we have

$$\sqrt{a} - \sqrt{b} > \sqrt{2} \times \max\left[\frac{\sqrt{2}}{\tilde{\delta}}, \frac{1}{\sqrt{\epsilon(1-\tilde{\delta})}}\right]. \tag{134}$$

We pick $\tilde{\delta} = (\sqrt{2}-1)\epsilon$ and this yields

$$\sqrt{a} - \sqrt{b} > \frac{2}{(\sqrt{2}-1)\epsilon}. \tag{135}$$

This completes the proof of Theorem 3.7.

H. Proof of Theorem 3.8 (Threshold condition for $\mathcal{M}_{\text{RR}}(G)$ for $r = 2$)

Error probability analysis of SDP recovery algorithm:

For the ease of exposition, let us consider a graph G with $2n$ vertices. The Lagrangian function is written as follows:

$$\mathcal{L}(\tilde{\mathbf{A}}, \mathbf{Y}, \mathbf{S}, \mathbf{D}, \lambda) = \text{tr}(\tilde{\mathbf{A}}\mathbf{Y}) + \text{tr}(\mathbf{S}\mathbf{Y}) - \text{tr}(\mathbf{D}(\mathbf{Y} - \mathbf{I})) - \lambda \,\text{tr}(\mathbf{J}\mathbf{Y}), \tag{136}$$

where $\mathbf{S} \succcurlyeq \mathbf{0}$, $\mathbf{D} = \text{diag}(d_i)$ and $\lambda \in \mathbb{R}$. Then,

$$\nabla_{\mathbf{Y}} \mathcal{L} = \tilde{\mathbf{A}} + \mathbf{S} - \mathbf{D} - \lambda \mathbf{J} = \mathbf{0}. \tag{137}$$

In order to satisfy the first order stationery condition, we have

$$\mathbf{S}^* = \mathbf{D}^* - \tilde{\mathbf{A}} + \lambda^* \mathbf{J}. \tag{138}$$

From the KKT conditions, we have the following:

$$\operatorname{tr}(\mathbf{D}, \mathbf{Y} - \mathbf{I}) = 0 \Rightarrow \mathbf{D}^*\mathbf{Y}^* = \mathbf{D}^*\mathbf{I}, \tag{139}$$

$$\lambda^* \operatorname{tr}(\mathbf{J}\mathbf{Y}^*) = 0, \tag{140}$$

$$\operatorname{tr}(\mathbf{S}^*\mathbf{Y}^*) = 0 \Rightarrow \mathbf{S}\boldsymbol{\sigma}^* = \mathbf{0}, \tag{141}$$

where in eqn. (141), $\boldsymbol{\sigma}^*$ is the null space of $\mathbf{S}$. In order to ensure that $\mathbf{Y}^*$ is the unique solution, we require that $\lambda_2(\mathbf{S}^*) > 0$, i.e., the second smallest eigenvalues of $\mathbf{S}$. This comes from the rank-nullity Theorem, i.e.,

$$\text{rank}(\mathbf{S}^*) + \text{Null}(\mathbf{S}^*) = 2n \Rightarrow \text{rank}(\mathbf{S}^*) = 2n - 1. \tag{142}$$

To this end, we have the following:

$$\begin{aligned}\operatorname{tr}(\tilde{\mathbf{A}}\mathbf{Y}) \leq \mathcal{L}(\tilde{\mathbf{A}}, \mathbf{Y}^*, \mathbf{S}^*, \mathbf{D}^*, \lambda^*) &= \operatorname{tr}(\mathbf{S}^* - \mathbf{D}^* + \tilde{\mathbf{A}} - \lambda^*\mathbf{J}Y) + \operatorname{tr}(\mathbf{D}^*\mathbf{I}) \overset{(a)}{=} \operatorname{tr}(\mathbf{D}^*\mathbf{I}) \\ &= \operatorname{tr}(\mathbf{D}^*\mathbf{Y}^*) \\ &= \operatorname{tr}(\mathbf{S}^* + \tilde{\mathbf{A}} - \lambda^*\mathbf{J}, \mathbf{Y}^*) = \operatorname{tr}(\tilde{\mathbf{A}}\mathbf{Y}^*).\end{aligned} \tag{143}$$

Now, our goal is to prove that w.h.p. $\mathbf{S}^* \succcurlyeq 0$ with $\lambda_2(\mathbf{S}^*) > 0$. More specifically, we want to show that

$$\Pr\left[\inf_{\mathbf{x}:\|\mathbf{x}\|=1, \mathbf{x}\perp\boldsymbol{\sigma}^*} \mathbf{x}^T\mathbf{S}^*\mathbf{x} > 0\right] \geq 1 - o(1). \tag{144}$$

Alternatively,

$$\Pr\left[\inf_{\mathbf{x}:\|\mathbf{x}\|=1, \mathbf{x}\perp\boldsymbol{\sigma}^*} \mathbf{x}^T\mathbf{S}^*\mathbf{x} \leq 0\right] \leq o(1). \tag{145}$$

Before we proceed, we note that

$$\mathbb{E}[\tilde{\mathbf{A}}] = \frac{\tilde{p} - \tilde{q}}{2}\mathbf{Y}^* + \frac{\tilde{p} + \tilde{q}}{2}\mathbf{J} - \tilde{p}\mathbf{I}. \tag{146}$$

Now, for any $\mathbf{x}$ such that $\|\mathbf{x}\| = 1, \mathbf{x} \perp \boldsymbol{\sigma}^*$ (i.e., $\boldsymbol{x}^T\boldsymbol{\sigma}^* = 0$), and $\lambda^* \leq (\tilde{p} + \tilde{q})/2$,

$$\begin{aligned}\mathbf{x}^T\mathbf{S}^*\mathbf{x} &= \mathbf{x}^T\mathbf{D}^*\mathbf{x} - \mathbf{x}^T\mathbb{E}[\tilde{\mathbf{A}}]\mathbf{x} + \lambda^*\mathbf{x}^T\mathbf{J}\mathbf{x} - \mathbf{x}^T(\tilde{\mathbf{A}} - \mathbb{E}[\tilde{\mathbf{A}}])\mathbf{x} \\ &= \mathbf{x}^T\mathbf{D}^*\mathbf{x} - \frac{\tilde{p} - \tilde{q}}{2}\mathbf{x}^T\mathbf{Y}^*\mathbf{x} + \left(\lambda^* - \frac{\tilde{p} + \tilde{q}}{2}\right)\mathbf{x}^T\mathbf{J}\mathbf{x} + \tilde{p} - \mathbf{x}^T(\tilde{\mathbf{A}} - \mathbb{E}[\tilde{\mathbf{A}}])\mathbf{x} \\ &\leq \mathbf{x}^T\mathbf{D}^*\mathbf{x} - \frac{\tilde{p} - \tilde{q}}{2}\mathbf{x}^T\mathbf{Y}^*\mathbf{x} + \tilde{p} - \|\tilde{\mathbf{A}} - \mathbb{E}[\tilde{\mathbf{A}}]\| \\ &= \mathbf{x}^T\mathbf{D}^*\mathbf{x} - \frac{\tilde{p} - \tilde{q}}{2}\mathbf{x}^T\boldsymbol{\sigma}^*\boldsymbol{\sigma}^{*T}\mathbf{x} + \tilde{p} - \|\tilde{\mathbf{A}} - \mathbb{E}[\tilde{\mathbf{A}}]\| \\ &= \mathbf{x}^T\mathbf{D}^*\mathbf{x} + \tilde{p} - \|\tilde{\mathbf{A}} - \mathbb{E}[\tilde{\mathbf{A}}]\| \\ &\leq \mathbf{x}^T\mathbf{D}^*\mathbf{x} + \tilde{p} = \sum_{i\in[2n]} d_i^* + \tilde{p}.\end{aligned} \tag{147}$$

Note that from (143), we have $\operatorname{tr}(\mathbf{D}^*\mathbf{I}) = \operatorname{tr}(\tilde{\mathbf{A}}\mathbf{Y}^*)$. Therefore,

$$d_i^* = \sum_{j=1}^{2n} \tilde{A}_{i,j}\sigma_i^*\sigma_j^*. \tag{148}$$

Also, note that each d_i^* is equal in distribution to $X - Y$, where $X \sim \text{Bin}\big(n-1, a_n\frac{\log(n)}{n}\big)$ and $Y \sim \text{Bin}\big(n, b_n\frac{\log(n)}{n}\big)$. Applying the union bound, our goal now is to derive conditions on a_n and b_n such that

$$\sum_{i=1}^{2n} \Pr(d_i^* \leq 0) \leq n \times \Pr(Y \geq X) = o(1). \tag{149}$$

Similar as before, the above probability will be of order $o(1)$ holds if

$$\begin{aligned}&\sqrt{a}-\sqrt{b}-\sqrt{\frac{1}{e^{\epsilon}-1}}>\sqrt{2}\times\sqrt{\frac{e^{\epsilon}+1}{e^{\epsilon}-1}}\\ \Rightarrow&\sqrt{a}-\sqrt{b}>\sqrt{2}\times\sqrt{\frac{e^{\epsilon}+1}{e^{\epsilon}-1}}+\frac{1}{\sqrt{e^{\epsilon}-1}}.\end{aligned}\tag{150}$$

To this end, we conclude that

$$\Pr(\hat{\boldsymbol{\sigma}}\neq\boldsymbol{\sigma})\leq\Pr(\hat{\mathbf{Y}}\neq\mathbf{Y})=o(1).\tag{151}$$

We are ready now to analyze the error probability $\Pr(\hat{\boldsymbol{\sigma}}\neq\boldsymbol{\sigma})$. Denote $Y\sim\mathrm{Bin}(n,q_n)$ and $X\sim\mathrm{Bin}(n,p_n)$. Using union bound, we have that

$$\Pr(\hat{\boldsymbol{\sigma}}\neq\boldsymbol{\sigma})\leq 2n\times\Pr(Y\geq X).\tag{152}$$

We next expand $\Pr(Y\geq X)$ using law of total probability Theorem. We first define $Z=X+Y$ and $c_n=a_n+b_n$, then we have

$$\begin{aligned}\Pr(Y\geq X)&=\sum_{k=0}^{2n}\Pr(Y\geq X|Z=k)\Pr(Z=k)\\ &\leq\sum_{k=0}^{10c_n\log(n)}\Pr(Y\geq X|Z=k)\Pr(Z=k)+\Pr(Z\geq 10c_n\log(n))\\ &=\Pr(Y\geq X|Z=0)\Pr(Z=0)+\sum_{k=1}^{10c_n\log(n)}\Pr(Y\geq X|Z=k)\Pr(Z=k)+\Pr(Z\geq 10c_n\log(n))\\ &=\Pr(Y=X=0)+\sum_{k=1}^{10c_n\log(n)}\Pr(Y\geq X|Z=k)\Pr(Z=k)+\Pr(Z\geq 10c_n\log(n))\\ &=\Pr(Y=0)\times\Pr(X=0)+\sum_{k=1}^{10c_n\log(n)}\Pr(Y\geq X|Z=k)\Pr(Z=k)+\Pr(Z\geq 10c_n\log(n))\\ &\overset{(a)}{\leq} n^{-c_n}+\sum_{k=1}^{10c_n\log(n)}\Pr(Y\geq X|Z=k)\Pr(Z=k)+n^{-10c_n}\\ &\leq 2n^{-c_n}+\sum_{k=1}^{10c_n\log(n)}\Pr(Y\geq X|Z=k)\Pr(Z=k),\end{aligned}\tag{153}$$

where in step (a), we have that $\Pr(X=0)=(1-p_n)^n\leq e^{-np_n}=n^{-a_n}$. Similarly, $\Pr(Y=0)\leq n^{-b_n}$. Also, we have $\Pr(Z\geq 10c_n\log(n))\leq n^{-10c_n}$ using Bernstein's inequality. We next upper bound $\Pr(Z=k)$ as follows:

$$\Pr(Z=k)=\sum_{i=0}^{k}\Pr(Y=i)\times\Pr(X=k-i),\tag{154}$$

where,

$$\begin{aligned}
\Pr(Y=i) &= \binom{n}{i} \times p_n^i \times (1-p_n)^{n-i} \\
&= \frac{n!}{i!(n-i)!} \times p_n^i \times (1-p_n)^{n-i} \\
&= \frac{n!}{i!(n-i)!} \times \frac{(b_n \log(n))^i}{n^i} \times \left(1 - \frac{b_n \log(n)}{n}\right)^{n-i} \\
&= \frac{(b_n \log(n))^i}{i!} \times \frac{n!}{n^i \times (n-i)!} \times e^{-\frac{b_n \log(n)}{n} \times (n-i)} \\
&= \frac{(b_n \log(n))^i}{i!} \times \frac{n!}{n^i \times (n-i)!} \times e^{-b_n \log(n) \times (1-i/n)} \\
&= \frac{(b_n \log(n))^i}{i!} \times \frac{n!}{n^i \times (n-i)!} \times n^{-b_n} \times n^{b_n \times i/n} \\
&= \frac{(b_n \log(n))^i}{i!} \times \frac{n \times (n-1) \times \cdots \times (n-i+1)}{n^i} \times n^{-b_n} \times n^{b_n \times i/n} \\
&\overset{(a)}{\leq} \frac{(b_n \log(n))^i}{i!} \times 1 \times n^{-b_n} \times n^{b_n \times i/n},
\end{aligned} \tag{155}$$

where in step (a) follows that $\prod_{j=0}^{i-1} \left(1 - \frac{j}{n}\right) \leq 1$. Similarly, we upper bound $\Pr(X = k-i)$ as

$$\Pr(X = k-i) \leq \frac{(a_n \log(n))^{k-i}}{(k-i)!} \times n^{-a_n} \times n^{a_n \times (k-i)/n}. \tag{156}$$

To this end, we get

$$\begin{aligned}
\Pr(Z=k) &\overset{(a)}{\leq} n^{-c_n} \times n^{a_n k/n} \sum_{i=0}^{k} n^{-(a_n - b_n) \times i/n} \times \frac{(b_n \log(n))^i}{i!} \times \frac{(a_n \log(n))^{k-i}}{(k-i)!} \\
&\leq n^{-c_n} \times n^{a_n k/n} \sum_{i=0}^{k} \frac{(b_n \log(n))^i}{i!} \times \frac{(a_n \log(n))^{k-i}}{(k-i)!} \\
&\leq n^{-c_n} \times n^{a_n k/n} \times \frac{1}{k!} \sum_{i=0}^{k} k! \times \frac{(b_n \log(n))^i}{i!} \times \frac{(a_n \log(n))^{k-i}}{(k-i)!} \\
&\leq n^{-c_n} \times n^{a_n k/n} \times \frac{1}{k!} \sum_{i=0}^{k} \binom{k}{i} \times (b_n \log(n))^i \times (a_n \log(n))^{k-i} \\
&= n^{-c_n} \times n^{a_n k/n} \times \frac{(c_n \log(n))^k}{k!},
\end{aligned} \tag{157}$$

where in step (a), we have that $a_n \geq b_n$ and $n^{-(a_n - b_n) \times i/n} \leq 1$. Similarly, we upper bound $\Pr(Y \geq X | Z = k)$ as follows:

$$\begin{aligned}
\Pr(Y \geq X | Z=k) &= \sum_{i=\frac{k}{2}}^{k} \Pr(Y=i) \times \Pr(X = k-i) \\
&\leq n^{-c_n} \times n^{a_n k/n} \times \frac{1}{k!} \sum_{i=\frac{k}{2}}^{k} \binom{k}{i} \times (b_n \log(n))^i \times (a_n \log(n))^{k-i} \\
&= n^{-c_n} \times n^{a_n k/n} \times \frac{(c_n \log(n))^k}{k!} \sum_{i=\frac{k}{2}}^{k} \binom{k}{i} \times \eta_n^i \times (1-\eta_n)^{k-i},
\end{aligned} \tag{158}$$

where $\eta_n = \frac{b_n}{a_n+b_n} < 1/2$. For a fixed k where $k \leq 10c_n \log(n)$, we have

$$\begin{aligned}
\Pr(Z=k) \times \Pr(Y \geq X | Z=k) &\leq n^{-2c_n} \times n^{2a_n k/n} \times \frac{(c_n \log(n))^{2k}}{(k!)^2} \times \Pr(\mathrm{Bin}(k,\eta_n) \geq \frac{k}{2}) \\
&\overset{(a)}{\leq} \frac{1-\eta_n}{1-2\eta_n} \times n^{-2c_n} \times \frac{(c_n n^{a_n/n} \log(n))^{2k}}{(k!)^2} \times \Pr(\mathrm{Bin}(k,\eta_n) = \frac{k}{2}) \\
&= \frac{1-\eta_n}{1-2\eta_n} \times n^{-2c_n} \times \frac{(c_n n^{a_n/n} \log(n))^{2k}}{(k!)^2} \times \frac{k!}{\left(\frac{k}{2}!\right)^2} \times \theta_n^k \\
&= \frac{1-\eta_n}{1-2\eta_n} \times n^{-2c_n} \times \frac{(c_n n^{a_n/n} \log(n))^{2k}}{k!} \times \frac{1}{\left(\frac{k}{2}!\right)^2} \times \theta_n^k \\
&= \frac{1-\eta_n}{1-2\eta_n} \times n^{-2c_n} \times \frac{(c_n \log(n))^{k}}{k!} \times (c_n n^{2a_n/n} \log(n))^k \times \frac{1}{\left(\frac{k}{2}!\right)^2} \times \theta_n^k \\
&\overset{(b)}{\leq} \frac{1}{\sqrt{2\pi}} \times \frac{1-\eta_n}{1-2\eta_n} \times n^{-2c_n} \times \frac{(c_n e \log(n))^{k}}{k^{k+1/2}} \times (c_n n^{2a_n/n} \log(n))^k \times \frac{1}{\left(\frac{k}{2}!\right)^2} \times \theta_n^k \\
&\overset{(c)}{\leq} \frac{1}{\sqrt{2\pi}} \times \frac{1-\eta_n}{1-2\eta_n} \times n^{-2c_n} \times n^{c_n} \times (c_n n^{2a_n/n} \log(n))^k \times \frac{1}{\left(\frac{k}{2}!\right)^2} \times \frac{\theta_n^k}{\sqrt{k}} \\
&= \frac{1}{\sqrt{2\pi}} \times \frac{1-\eta_n}{1-2\eta_n} \times n^{-c_n} \times \frac{(\theta_n c_n n^{2a_n/n} \log(n))^{k}}{\left(\frac{k}{2}!\right)^2 \sqrt{k}} \\
&\overset{(d)}{\leq} \frac{2}{(2\pi^2)^{3/2}} \times \frac{1-\eta_n}{1-2\eta_n} \times n^{-c_n} \times \frac{(c_n n^{2a_n/n} e \log(n))^{k}}{k^{k+1}} \times \frac{(2\theta_n)^k}{\sqrt{k}} \\
&= \frac{1}{\sqrt{2}\pi^3} \times \frac{1-\eta_n}{1-2\eta_n} \times n^{-c_n} \times \frac{(2\theta_n c_n n^{2a_n/n} e \log(n))^{k}}{k^{k+3/2}} \\
&= \frac{1}{\sqrt{2}\pi^3} \times n^{2a_n k/n} \times \frac{1-\eta_n}{1-2\eta_n} \times n^{-c_n} \times \frac{(2\theta_n c_n e \log(n))^{k}}{k^{k+3/2}} \\
&\leq \frac{1}{\sqrt{2}\pi^3} \times n^{20 a_n c_n \log(n)/n} \times \frac{1-\eta_n}{1-2\eta_n} \times n^{-c_n} \times \frac{(2\theta_n c_n e \log(n))^{k}}{k^{k+3/2}}, \qquad (159)
\end{aligned}$$

where $\theta_n = \sqrt{\eta_n(1-\eta_n)} = \frac{\sqrt{a_n b_n}}{a_n+b_n}$. In step (a), we used Lemma H.1, while in steps (c) and (d) we used the following lower bound on $k!$, i.e., $k! \geq \sqrt{2\pi} k^{k+1/2} e^{-k}$. In step (b), it can be readily shown that

$$\frac{(c_n e \log(n))^k}{k^k} \leq n^{c_n}.$$

We next upper bound $\frac{(2\theta_n c_n e \log(n))^k}{k^{k+3/2}} \triangleq e^{f(k)}$. By taking $\log(\cdot)$ for this term, we have

$$f(k) = k \times \log(2\theta_n c_n e \log(n)) - (k+3/2) \times \log(k). \qquad (160)$$

We next take the derivative of $f(k)$,

$$\begin{aligned}
& f'(k) = \log(2\theta_n c_n e \log(n)) - (1+3/2k) - \log(k) = 0. \qquad (161) \\
\Rightarrow & \log(2\theta_n c_n e \log(n)) - (1+3/2k) - \log(k) = 0, \\
\Rightarrow & \log(e^{-1} 2\theta_n c_n e \log(n)) = 3/2k + \log(k), \\
\Rightarrow & e^{-1} 2\theta_n c_n e \log(n) = k \times e^{3/2k}, \\
\Rightarrow & 2\theta_n c_n \log(n) = k \times e^{3/2k}. \qquad (162)
\end{aligned}$$

Therefore, the optimal solution can be written as

$$k^* = 2\theta_n c_n \log(n) e^{-3/2k^*}. \qquad (163)$$

To this end, we have

$$\begin{aligned} e^{f(k^*)} &= \frac{n^{2\theta_n c_n e^{-3/2k^*}}}{2\theta_n c_n \log(n) e^{-3/2k^*}} \\ &\leq \frac{n^{2\theta_n c_n e^{-3/20 c_n \log(n)}}}{2\theta_n c_n \log(n)} \times e^{3/2k^*} \\ &\leq \frac{n^{2\theta_n c_n}}{2\theta_n c_n \log(n)} \times e^{3/2}. \end{aligned} \tag{164}$$

The second term will be upper bounded by

$$\begin{aligned} n^{-c_n} \times 10 c_n \log(n) \times \frac{n^{2\theta_n c_n}}{2\theta_n c_n \log(n)} \times e^{3/2} &= \frac{5e^{3/2}}{\theta_n} \times n^{-c_n + 2\theta_n c_n} \\ &= \frac{5e^{3/2}}{\theta_n} \times n^{-2[c_n/2 - \theta_n c_n]}. \end{aligned} \tag{165}$$

In order to achieve exact recovery, we require that a_n and b_n:

$$\frac{c_n}{2} - \theta_n c_n > 1 \Rightarrow \frac{a_n + b_n}{2} - \sqrt{a_n b_n} > 1. \tag{166}$$

Plugging the values of θ_n, c_n, we get

$$\Pr(\hat{\boldsymbol{\sigma}} \neq \boldsymbol{\sigma}) \leq 2n \times \left[2n^{-(a_n + b_n)} + \zeta_n \times n^{-2\left[(a_n+b_n)/2 - \sqrt{a_n b_n}\right]} \right], \tag{167}$$

where $\zeta_n = \frac{5 \times e^{3/2}}{\sqrt{2\pi^3}} \times e^{20 a_n (a_n + b_n) \log^2(n)/n} \times \frac{a_n}{a_n - b_n} \times \frac{a_n + b_n}{\sqrt{a_n b_n}}$. We upper bound the term $e^{20 a_n (a_n+b_n) \log^2(n)/n}$ such that

$$\begin{aligned} e^{20 a_n (a_n + b_n) \log^2(n)/n} &\leq e^{40 a_n^2 \log^2(n)/n} \leq n^{\alpha} = e^{\alpha \log(n)} \\ \Rightarrow a_n^2 &\leq \frac{\alpha}{40} \times \frac{n}{\log(n)}, \end{aligned} \tag{168}$$

where $\alpha < a_n + b_n - 2\sqrt{a_n b_n} - 1$. Plugging the value of a_n, we get

$$e^{\epsilon} > \frac{n}{\log(n)} \times \frac{1}{\sqrt{\frac{\alpha n}{40 \log(n)}} - a}. \tag{169}$$

Based on the previous condition, we have two cases: When (1) $a > \sqrt{\frac{\alpha n}{40 \log(n)}}$, in this case, a sufficient condition will be $\epsilon > \log(n) - \log(\log(n))$, and (2) $a < \sqrt{\frac{\alpha n}{40 \log(n)}}$, in this case we require that $\epsilon > \log(n) - \log(\log(n)) - \log\left(\sqrt{\frac{\alpha n}{40 \log(n)}} - a \right)$.

Derivation of Recovery Threshold Condition:

The randomized response mechanism $\mathcal{M}_{\text{RR}}(G)$ can be expressed as

$$\tilde{A}_{i,j} = (A_{i,j} + N_{i,j}) \bmod 2, \forall i \neq j, \tag{170}$$

where $N_{i,j} \sim \text{Bern}(1-\mu)$, $\mu = \frac{1}{e^{\epsilon}+1}$, $N_{i,j} = N_{j,i}$, and the operation $\bmod 2$ ensures that the released output is bounded, i.e., $\tilde{A}_{i,j} \in \{0, 1\}$. If nodes i and j belong to the same community, we have the following:

$$\begin{aligned} \tilde{p} &= \Pr(\tilde{A}_{i,j} = 1) = \Pr(A_{i,j} = 1) \times \Pr(N_{i,j} = 0 | A_{i,j} = 1) + \Pr(A_{i,j} = 0) \times \Pr(N_{i,j} = 0 | A_{i,j} = 1) \\ &\overset{(a)}{=} \Pr(A_{i,j} = 1) \times \Pr(N_{i,j} = 0) + \Pr(A_{i,j} = 0) \times \Pr(A_{i,j} = 1) \\ &= p \times (1 - \mu) + (1 - p) \times \mu \triangleq p \circledast \mu, \end{aligned} \tag{171}$$

where in step (a), the perturbation mechanism is independent of the $A_{i,j}$'s. Similarly, if nodes i and j belong to different communities, we have

$$\tilde{q} = \Pr(\tilde{A}_{i,j} = 1) = q \circledast \mu. \tag{172}$$

Plugging the expression of μ, p and q into the previous equations (171) and (172), we get

$$\tilde{p} = \left[\frac{n}{(e^\epsilon+1)\times\log(n)} + \frac{e^\epsilon-1}{e^\epsilon+1}\times a\right] \times \frac{\log(n)}{n} \triangleq a_n \times \frac{\log(n)}{n}. \tag{173}$$

Similarly,

$$\tilde{q} = \left[\frac{n}{(e^\epsilon+1)\times\log(n)} + \frac{e^\epsilon-1}{e^\epsilon+1}\times b\right] \times \frac{\log(n)}{n} \triangleq b_n \times \frac{\log(n)}{n}. \tag{174}$$

We next derive a necessary threshold condition for randomized response mechanism. From eqn. (167), in order to ensure exact recovery we require that

$$\frac{a_n+b_n}{2} - \sqrt{a_n b_n} > 1. \tag{175}$$

Plugging the expressions for a_n and b_n into (175), we get the following:

$$\begin{aligned}
&\frac{1}{2}\left[\frac{2}{e^\epsilon+1} + \frac{e^\epsilon-1}{e^\epsilon+1}\times(a+b)\right] - \sqrt{\frac{1}{e^\epsilon+1}+\frac{e^\epsilon-1}{e^\epsilon+1}\times a}\times\sqrt{\frac{1}{e^\epsilon+1}+\frac{e^\epsilon-1}{e^\epsilon+1}\times b} > 1\\
\Rightarrow &\frac{1}{e^\epsilon+1} + \frac{e^\epsilon-1}{e^\epsilon+1}\times\frac{a+b}{2} - \frac{1}{e^\epsilon+1}\times\sqrt{1+(e^\epsilon-1)\times a}\times\sqrt{1+(e^\epsilon-1)\times b} > 1\\
\Rightarrow &\frac{e^\epsilon-1}{e^\epsilon+1}\times\frac{a+b}{2} - \frac{1}{e^\epsilon+1}\times\sqrt{1+(e^\epsilon-1)\times a}\times\sqrt{1+(e^\epsilon-1)\times b} > \frac{e^\epsilon}{e^\epsilon+1}\\
\Rightarrow &\frac{a+b}{2} - \frac{1}{e^\epsilon-1}\times\sqrt{1+(e^\epsilon-1)\times a}\times\sqrt{1+(e^\epsilon-1)\times b} > \frac{e^\epsilon}{e^\epsilon-1}\\
\Rightarrow &\frac{a+b}{2} - \sqrt{\left(\frac{1}{e^\epsilon-1}+a\right)\left(\frac{1}{e^\epsilon-1}+b\right)} > \frac{e^\epsilon}{e^\epsilon-1}.
\end{aligned} \tag{176}$$

We can further simplify the above equation as follows:

$$\begin{aligned}
&a+b-2\sqrt{\left(\frac{1}{e^\epsilon-1}+a\right)\left(\frac{1}{e^\epsilon-1}+b\right)} > \frac{2e^\epsilon}{e^\epsilon-1}\\
\Rightarrow &-\frac{2}{e^\epsilon-1} + \frac{1}{e^\epsilon-1} + a + \frac{1}{e^\epsilon-1} + b - 2\sqrt{\left(\frac{1}{e^\epsilon-1}+a\right)\left(\frac{1}{e^\epsilon-1}+b\right)} > \frac{2e^\epsilon}{e^\epsilon-1}\\
\Rightarrow &\left(\sqrt{\frac{1}{e^\epsilon-1}+a} - \sqrt{\frac{1}{e^\epsilon-1}+b}\right)^2 > \frac{2e^\epsilon}{e^\epsilon-1}.
\end{aligned} \tag{177}$$

To this end, we get

$$\sqrt{\frac{1}{e^\epsilon-1}+a} - \sqrt{\frac{1}{e^\epsilon-1}+b} > \sqrt{2}\times\sqrt{\frac{e^\epsilon+1}{e^\epsilon-1}}. \tag{178}$$

A more stringent condition is

$$\begin{aligned}
&\sqrt{a} - \sqrt{b} - \sqrt{\frac{1}{e^\epsilon-1}} > \sqrt{2}\times\sqrt{\frac{e^\epsilon+1}{e^\epsilon-1}}\\
\Rightarrow &\sqrt{a} - \sqrt{b} > \sqrt{2}\times\sqrt{\frac{e^\epsilon+1}{e^\epsilon-1}} + \frac{1}{\sqrt{e^\epsilon-1}}.
\end{aligned} \tag{179}$$

This completes the proof of Theorem 3.8. Note that the threshold condition matches the non-private case when $\epsilon = \infty$.

Auxiliary Results:

Lemma H.1. *Suppose $X \sim \mathrm{Bin}(k,p)$, then for $p < 1/2$, we have*

$$\Pr(X \geq \frac{k}{2}) \leq \frac{1-p}{1-2p} \times \Pr(X = \frac{k}{2}). \tag{180}$$

Proof. Our goal is to upper bound the following ratio:

$$\frac{\Pr(X \geq \frac{k}{2})}{\Pr(X = \frac{k}{2})} \leq \frac{\Pr(X = \frac{k}{2}) + \Pr(X = \frac{k}{2}+1) + \cdots + \Pr(X = k)}{\Pr(X = \frac{k}{2})} \tag{181}$$

For any m and i, we have the following:

$$\begin{aligned}
\frac{\Pr(X = m+i)}{\Pr(X = m)} &= \frac{\binom{k}{m+i} p^{m+i}(1-p)^{k-m-i}}{\binom{k}{m} p^m (1-p)^{k-m}} \\
&= \frac{\binom{k}{m+i}}{\binom{k}{m}} \times \left(\frac{p}{1-p}\right)^i \\
&= \frac{m!}{(m+i)!} \times \frac{(k-m)!}{(k-m-i)!} \times \left(\frac{p}{1-p}\right)^i \\
&\leq \left(\frac{k-m}{m+1}\right)^i \delta^i,
\end{aligned} \tag{182}$$

where $\delta = \frac{p}{1-p} < 1, \forall p < 1/2$. For $m = \frac{k}{2}$, we have

$$\frac{\Pr(X = \frac{k}{2}+i)}{\Pr(X = \frac{k}{2})} \leq \left(\frac{k/2}{k/2+1}\right)^i \delta^i \leq \delta^i. \tag{183}$$

To this end,

$$\begin{aligned}
\frac{\Pr(X \geq \frac{k}{2})}{\Pr(X = \frac{k}{2})} &\leq \sum_{i=0}^{k/2} \delta^i \leq \sum_{i=0}^{\infty} \delta^i \\
&= \frac{1}{1-\delta} = \frac{1}{1-\frac{p}{1-p}} = \frac{1-p}{1-2p}.
\end{aligned} \tag{184}$$

□

Lemma H.2. *Consider a random variable Z as a sum of $2n$ independent random variables, i.e., $Z = \sum_{i=1}^{n}(X_i + Y_i)$, where $X_i \sim \mathrm{Bern}(p_n)$, $Y_i \sim \mathrm{Bern}(q_n)$ and $\mu_Z = c_n \log(n)$. Then, we have*

$$\Pr(Z \geq 10 c_n \log(n)) \leq n^{-10c_n}. \tag{185}$$

Proof. By direct application of Bernstein's inequality, it is straight forward to show that $|X_i| \leq 1$, $|Y_i| \leq 1$ and $\sum_{i=1}^{n} \mathbb{E}\left[(X_i - \mu_{X_i})^2\right] + \sum_{i=1}^{n} \mathbb{E}\left[(Y_i - \mu_{Y_i})^2\right] \leq n(p_n + q_n) = c_n \log(n)$. Then, we have the following:

$$\begin{aligned}
\Pr(Z \geq 10 c_n \log(n)) &= \Pr(Z - c_n \log(n) \geq 9 c_n \log(n)) \\
&= \Pr(Z - \mu_Z \geq 9 c_n \log(n)) \\
&\leq \exp\left[-\frac{\frac{1}{2} \times 81 \times \times c_n^2 \log^2(n)}{c_n \log(n) + \frac{1}{3} \times 9 c_n \log(n)}\right] \\
&\leq \exp\left[-\frac{40 c_n^2 \log^2(n)}{c_n \log(n) + 3 c_n \log(n)}\right] \\
&= \exp\left(-10 c_n \log(n)\right) = n^{-10c_n}.
\end{aligned} \tag{186}$$

□

Lemma H.3. *(Chernoeff-Hoeffding bound) Let $X = \sum_{i\in[n]} X_i$, where X_i's are indentically and independently distributed over the support $\{0,1\}$. Then, for any $\gamma \in (0,1]$, we have*

$$\Pr\left[X \notin [(1-\gamma)\mathbb{E}(X), (1+\gamma)\mathbb{E}(X)]\right] \leq 2e^{-\frac{\gamma^2\mathbb{E}(X)}{3}}. \tag{187}$$

Lemma H.4. *(Tail bounds on the difference of two Binomial R.V.s (Hajek et al., 2016b)) Let X and R be independent R.V.s with $X \sim \mathrm{Bin}(m_1, \frac{a\log(n)}{n})$ and $R \sim \mathrm{Bin}(m_2, \frac{b\log(n)}{n})$, where $m_1, m_2 \in \mathbb{N}$, such that $f_{n,\epsilon} \leq (m_1 a - m_2 b)\frac{\log(n)}{n}$, then*

$$\Pr(X - R \leq f_{n,\epsilon}) \leq n^{-g(m_1/n, m_2/n, a, b, f_{n,\epsilon}/\log(n))}, \tag{188}$$

where,

$$g(m_1/n, m_2/n, a, b, f_{n,\epsilon}/\log(n)) = a \times \frac{m_1}{n} + b \times \frac{m_2}{n} - \gamma - \frac{\alpha}{2} \times \log\left[\frac{(\gamma-\alpha)am_1}{(\gamma+\alpha)bm_2}\right], \tag{189}$$

where $\alpha = f_{n,\epsilon}/\log(n)$, and $\gamma = \sqrt{\alpha^2 + 4\frac{m_1 m_2}{n^2}ab}$.

Definition H.5 (Multiplicative Chernoff Bound)**.** *Given $\tilde{m}_1^{(k)} \sim \mathrm{Bin}(2k(n-k), p)$, and $\tilde{m}_2^{(k)} \sim \mathrm{Bin}(2k(n-k), q)$, we have*

$$\Pr\left[\tilde{m}_1^{(k)} - \tilde{m}_2^{(k)} < (1-\tilde{\delta})\mu^{(k)} || S_1| = k, |S_2| = k\right] \leq \exp(-\tilde{\delta}^2\mu^{(k)}/2), \tag{190}$$

where,

$$\mu^{(k)} = 2k(n-k)(a-b)\frac{\log(n)}{n}. \tag{191}$$

Applying the union bound for possible values of $k \in [1 : \frac{n}{2}]$, it yields

$$\begin{aligned}
\Pr\left[E^c_{S_1,S_2}\right] &= \sum_{k=1}^{\frac{n}{2}} \binom{n}{k}^2 \times \exp\left[-\frac{\tilde{\delta}^2}{2} \times 2k(n-k) \times (a-b) \times \frac{\log(n)}{n}\right] \\
&\leq \sum_{k=1}^{\frac{n}{2}} \left(\frac{ne}{k}\right)^{2k} \times \exp\left[-\frac{\tilde{\delta}^2}{2} \times k \times (a-b) \times \log(n)\right] \\
&= \sum_{k=1}^{\frac{n}{2}} \exp\left[2k\bigg(\log(n) - \log(k) + 1 - \frac{\tilde{\delta}^2}{4} \times (a-b) \times \log(n)\bigg)\right] \\
&= \sum_{k=1}^{\frac{n}{2}} \exp\left[-2k\bigg(\log(k) - 1 + \bigg(\frac{\tilde{\delta}^2}{4} \times (a-b) - 1\bigg)\log(n)\bigg)\right].
\end{aligned} \tag{192}$$

In order to make the probability decays with n, we require that

$$\tilde{\delta}^2(a-b) > 4 \Rightarrow a - b > \frac{4}{\tilde{\delta}^2} \Rightarrow \sqrt{a} - \sqrt{b} > \frac{\sqrt{2}}{\tilde{\delta}/\sqrt{2}}. \tag{193}$$