On the Price of Differential Privacy for Spectral Clustering over Stochastic Block Models

Antti Koskela*, Mohamed Seif*, Andrea J. Goldsmith

***Abstract*—We investigate privacy-preserving spectral clustering for community detection within stochastic block models (SBMs). Specifically, we focus on edge differential privacy (DP) and propose private algorithms for community recovery. Our work explores the fundamental trade-offs between the privacy budget and the accurate recovery of community labels. Furthermore, we establish information-theoretic conditions that guarantee the accuracy of our methods, providing theoretical assurances for successful community recovery under edge DP.**

***Index Terms*—Differential Privacy, Graphs, Stochastic Block Model, Perturbation, Community Detection, Spectral Clustering.**

I. Introduction

Community detection within networks is a pivotal challenge in graph mining and unsupervised learning [1]. The primary objective is to identify divisions (or communities) in a graph where connections are densely concentrated inside communities and sparsely distributed between them. The Stochastic Block Model (SBM) is widely utilized to represent the structural patterns of networks [2]. In the SBM framework, nodes are assigned to specific communities, and the probability of connections between any two nodes is based on their community memberships. Specifically, nodes within the same community are more likely to be connected than those in different communities. This variation in connection probabilities is fundamental to the challenge of detecting communities. Research aimed at studying and enhancing community detection methods using the SBM approach has been highly active, with numerous advancements and discoveries detailed in comprehensive reviews such as the one by Abbe et al. [3].

Network data, such as the connections found in social networks, often contain sensitive information. Therefore, protecting individual privacy during data analysis is essential. Differential Privacy (DP) [4] has become the standard method for providing strong privacy guarantees. DP ensures that the inclusion or exclusion of any single user's data in a dataset has only a minimal effect on the results of statistical queries.

In the realm of network or graph data, both edge and node privacy models have been investigated. As discussed in [5], two primary privacy concepts have been introduced for analyzing graph data: (1) Edge DP, which aims to safeguard individual relationships (edges) within a graph by utilizing randomized algorithms to minimize the impact of any specific edge's presence or absence during analysis, and (2) Node DP, which focuses on protecting the privacy of nodes and their associated connections (edges). Edge DP is better suited for private community detection, as it focuses on protecting individual relationships, which are central to defining and identifying community labels. Additionally, DP algorithms have been adapted to address specific network analysis tasks, such as counting stars, triangles, cuts, dense subgraphs, and communities, as well as generating synthetic graphs [6], [7], [8], [9]. More recently, in our previous work, we explored community detection in SBMs under the edge privacy model for various settingsand established sufficient conditions for recoverability thresholds using ML-based estimators and their semidefinite relaxations [10], [11], [12].

In the existing literature, efficient algorithms for community detection in SBMs have been developed using spectral methods, such as those detailed in [13], [14], as well as through semidefinite programming (SDP) approaches [15]. While these spectral methods have demonstrated significant effectiveness in terms of the computational complexity in identifying community structures, there remains a limited understanding of their performance under privacy constraints. Specifically, there is a notable gap in knowledge regarding private spectral methods for various community recovery requirements, including partial and exact recovery. This highlights an important area for future research aimed at ensuring that community detection techniques can both preserve privacy and maintain high accuracy across different recovery requirements (e.g., exact or partial recovery) within the SBM framework.

Related Work. The closest related work to our study is [16], in which the authors analyze the consistency of privacy–preserving spectral clustering under the Stochastic Block Model (SBM). While insightful, their results stop short of deriving *explicit separation conditions* that tie together the SBM parameters (e.g., block edge probabilities, number of communities) and the privacy budget ε. Pinpointing these conditions is essential for understanding *when* private spectral methods can provably recover communities and *how* the privacy constraint degrades the signal–to–noise ratio required for success.

Our earlier efforts [10], [11] addressed this question from an optimization standpoint by casting community detection as a semidefinite program (SDP). Although the SDP approach delivers strong statistical guarantees, its polynomial-time complexity renders it impractical for the massive graphs that arise in modern social-network or e-commerce platforms—often containing tens of millions of nodes and billions of edges.

Key Outstanding Challenge. A rigorous privacy–utility

*A. Koskela and M. Seif contributed equally to this work.
A. Koskela is with Nokia Bell Labs.
M. Seif and A. J. Goldsmith are with Princeton University.
The work was supported by the AFOSR award #002484665, Huawei Intelligent Spectrum grant, and NSF award CNS-2147631. The authors thank Iraj Saniee for facilitating this collaboration.

trade-off analysis is still needed—one that links scalable private algorithms to explicit separation thresholds expressed in terms of SBM parameters and the privacy budget ε. Closing this theoretical gap would pave the way for deployable, high-accuracy, privacy-preserving community detection that (i) **scales** to multi-million-node graphs, (ii) **respects** user-level differential-privacy guarantees, and (iii) **achieves** the full spectrum of recovery objectives—exact, partial, or weak consistency—studied in the SBM literature.

Contributions. We make the following key contributions to privacy-preserving spectral clustering under edge DP for community detection on the symmetric binary SBM[1]:

1) **Graph Perturbation-Based Mechanism**: We apply the randomized response technique to perturb the adjacency matrix of the graph. Subsequently, a spectral clustering algorithm is executed on the perturbed graph to recover community structures. This approach inherently satisfies ϵ-DP for any $\epsilon > 0$ due to the post-processing property of differential privacy.
2) **Subsampling Stability-Based Mechanism**: Inspired by the work of [4], we introduce the subsampling stability-based estimator, which involves generating multiple correlated subgraphs by randomly sampling edges with probability q_s. A non-private clustering algorithm is then applied to each subgraph, and the resulting community labels are aggregated into a histogram. The stability of this histogram, influenced by parameters such as (p, q), (ϵ, δ), q_s, and the number of subgraphs m, ensures accurate recovery of the original community labels.
3) **Noisy Power Iteration Method**: We execute the power method while injecting carefully calibrated Gaussian noise at every matrix–vector multiplication to obfuscate each edge's contribution. Subsequently, the normalized noisy eigenvectors are used to form the clustering embedding. This approach inherently satisfies (ϵ, δ)-edge DP for any $\epsilon > 0$ (with suitably small δ) via the Gaussian mechanism and iterative composition.
4) **Tradeoff Analysis and Theoretical Guarantees**: We investigate the fundamental tradeoffs between the privacy budget and the accuracy of community recovery. Additionally, we provide theoretical guarantees by establishing information-theoretic conditions that ensure successful community detection under edge DP. Furthermore, we provide a lower bound on the overlap rate between the estimated labels and the ground truth labels of the communities.

Notation. Boldface uppercase letters denote matrices (e.g., $\mathbf{A}$), while boldface lowercase letters are used for vectors (e.g., $\mathbf{a}$). We use $\mathrm{Bern}(p)$ to denote a Bernoulli random variable with success probability p. For asymptotic analysis, we say the function $f(n) = o(g(n))$ when $\lim_{n\to\infty} f(n)/g(n) = 0$. Also, $f(n) = O(g(n))$ means there exist some constant $C > 0$ such that $|f(n)/g(n)| \le C$, $\forall n$, and $f(n) = \Omega(g(n))$ means there exists some constant $c > 0$ such that $|f(n)/g(n)| \ge c$, $\forall n$.

[1]Generalizing the privacy mechanisms to multiple communities is a sufficiently interesting direction and is left for future work.

II. Problem Statement & Preliminaries

We consider an undirected graph $G = (V, E)$ consisting of n vertices, where the vertices are divided into two equally sized communities C_1 and C_2, $V = C_1 \cup C_2$ and E is the edge set. The community label for vertex i is denoted by $\sigma_i^* \in \{-1, +1\}, \forall i \in [n]$. Further, we assume that the graph G is generated through an SBM, where the edges within the same community are generated independently with probability p, and the edges across the communities C_1 and C_2 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 follows:

$$A_{i,j} \sim \begin{cases} \mathrm{Bern}(p), & i < j, \quad \sigma_i^* = \sigma_j^*, \\ \mathrm{Bern}(q), & i < j, \quad \sigma_i^* \neq \sigma_j^* \end{cases}$$

with $A_{i,i} = 0$ and $A_{i,j} = A_{j,i}$ for $i > j$.

Definition 1 (Laplacian Matrix)**.** Let $G = (V, E)$ be undirected graph. The Laplacian $\mathbf{L} \in \mathbb{R}^{n\times n}$ of G is the matrix defined as

$$\mathbf{L} = \mathbf{D} - \mathbf{A},$$

where $\mathbf{D} \in \mathbb{R}^{n\times n}$ is the degree matrix, which is a diagonal matrix where each diagonal entry d_{ii} is defined as $d_{ii} = \sum_{j=1}^{n} A_{ij}$, and all off-diagonal entries of $\mathbf{D}$ are zero.

Community Detection via Spectral Method. Spectral clustering partitions the vertices of a graph G into communities by leveraging its spectral properties. To accomplish this, we perform an eigen decomposition of the Laplacian matrix $\mathbf{L}$, obtaining its eigenvalues $0 = \lambda_1 \le \lambda_2 \le \cdots \le \lambda_n$ and their corresponding eigenvectors $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n$. In the case of dividing the graph into two communities, spectral clustering specifically utilizes the eigenvector associated with the second smallest eigenvalue λ_2 of $\mathbf{L}$. This eigenvector effectively captures the essential structure needed to separate the graph's vertices into distinct communities based on the graph's connectivity [17].

Definition 2 ((β, η)-Accurate Recovery)**.** A community recovery alghorithm $\hat{\boldsymbol{\sigma}}(G) = \{\hat{\sigma}_1, \hat{\sigma}_2, \cdots, \hat{\sigma}_n\}$ achieves (β, η)-accurate recovery (up to a global flip) if

$$\Pr\Big(\mathrm{err\,rate}\big(\hat{\boldsymbol{\sigma}}(G), \boldsymbol{\sigma}^*\big) \le \beta\Big) \ge 1 - \eta, \tag{1}$$

where the probability is taken over both the randomness of the graph G (drawn according to an SBM) and the randomness of the algorithm. Here, the error rate (up to a global flip) is defined via the Hamming distance as

$$\mathrm{err\,rate}\big(\hat{\boldsymbol{\sigma}}(G), \boldsymbol{\sigma}^*\big) \;=\; \frac{1}{n} \cdot \min_{s\in\{+1,-1\}} \mathrm{Ham}\big(\hat{\boldsymbol{\sigma}}(G), s\boldsymbol{\sigma}^*\big).$$

Definition 3 ((ϵ, δ)-edge DP)**.** A (randomized) community estimator $\hat{\boldsymbol{\sigma}}$ as a function of G satisfies (ϵ, δ)-edge DP for some $\epsilon \in \mathbb{R}^+$ and $\delta \in [0, 1]$, if for all pairs of adjacency matrices G and G' that differ in *one* edge, and any measurable subset $\mathcal{S} \subseteq \mathrm{Range}(\hat{\boldsymbol{\sigma}})$, we have

$$\Pr(\hat{\boldsymbol{\sigma}}(G) \in \mathcal{S}) \le e^{\epsilon}\Pr(\hat{\boldsymbol{\sigma}}(G') \in \mathcal{S}) + \delta,$$

Algorithm 1 Spectral Clustering Algorithm

1: **Input:** $G(\mathcal{V}, E)$
2: **Output:** Labeling vector $\hat{\boldsymbol{\sigma}}(\mathbf{A})$
3: Compute Laplacian: $\mathbf{L} = \mathbf{D} - \mathbf{A}$
4: Eigen decomposition of $\mathbf{L}$: obtain $\lambda_i, \mathbf{u}_i$
5: Select Fiedler vector $\mathbf{u}_2$ for λ_2
6: Assign communities:

$$\hat{\sigma}(v) = \begin{cases} 1 & \text{if } u_{2,v} \leq 0 \\ -1 & \text{otherwise} \end{cases}$$

7: **Optional:** Flip labels to minimize clustering error
8: **Return:** $\hat{\boldsymbol{\sigma}}(\mathbf{A})$

where the probabilities are computed only over the randomness in the estimation process. The setting when $\delta = 0$ is referred as pure ϵ-edge DP

III. Main Results & Discussions

We first establish a lower bound for all DP community recovery algorithms applied to graphs generated from binary SBMs, utilizing packing arguments under DP [18]. We then consider three different DP community recovery algorithms: graph perturbation-based mechanism, subsampling stability-based mechanism and a noisy power method applied on the adjacency matrix.

A. General Lower Bound for ϵ-edge DP Community Recovery Algorithms

We establish a rigorous lower bound for all differentially private community recovery algorithms operating on graphs generated from SBMs. Our methodology closely follows the frameworks outlined in [18], [19], focusing on the notion of edge DP. Precisely speaking, we define the classification error rate as

$$\begin{aligned} &\operatorname{err\,rate}(\hat{\boldsymbol{\sigma}}(\mathbf{A}), \boldsymbol{\sigma}^*) \\ &= \frac{1}{n} \cdot \min\{\operatorname{Ham}(\hat{\boldsymbol{\sigma}}(\mathbf{A}), \boldsymbol{\sigma}^*), \operatorname{Ham}(-\hat{\boldsymbol{\sigma}}(\mathbf{A}), \boldsymbol{\sigma}^*)\}. \end{aligned}$$

Let us consider a series of pairwise disjoint sets $\mathcal{S}_i, i \in [m]$. Each set $\mathcal{S}_i$ contains vectors $\boldsymbol{u} \in \{\pm 1\}^n$, where n is the vector dimension. A vector $\boldsymbol{u}$ is included in $\mathcal{S}_i$ if $\operatorname{err\,rate}(\boldsymbol{u}, \boldsymbol{\sigma}^i)$ with a fixed vector $\boldsymbol{\sigma}^i$ does not exceed the threshold β. This is formally expressed as:

$$\mathcal{S}_i = \{\boldsymbol{u} \in \{\pm 1\}^n : \operatorname{err\,rate}(\boldsymbol{u}, \boldsymbol{\sigma}^i) \leq \beta\}, \tag{2}$$

$i = 1, 2, \ldots, m$, where $\mathcal{S}_i$'s are pairwise disjoint sets.

We next derive the necessary conditions for

$$\Pr(\hat{\boldsymbol{\sigma}}(\mathbf{A}) \in \mathcal{S}_i) \geq 1 - \eta \tag{3}$$

as a function of the SBM parameters and the privacy budget. Note that the randomness here is taken over the randomness of graph G that is generated from $\operatorname{SBM}(\boldsymbol{\sigma}^i, n, p, q)$.

Lemma III.1. *Let $\boldsymbol{\sigma}^i$ be a fixed vector and the set $\mathcal{S}_i$ be defined as in Eq.* (2) *for some $\beta > 0$. Suppose the condition of Eq.* (3) *holds. Then,*

$$\begin{aligned} (1-\eta)^2 \leq &\left(\mathbb{E}_{\mathbf{A},\mathbf{A}' \sim \Pi(\mathbf{A},\mathbf{A}')} \left[e^{2\epsilon \operatorname{Ham}(\mathbf{A},\mathbf{A}')}\right]\right) \\ &\times \Pr(\hat{\boldsymbol{\sigma}}(\mathbf{A}') \in \mathcal{S}_i), \end{aligned}$$

Proof. Without loss of generality, let us consider a graph $\mathbf{A} \sim \operatorname{SBM}(\boldsymbol{\sigma}^1, n, p, q)$ that is generated from the ground truth labeling vector $\boldsymbol{\sigma}^* = \boldsymbol{\sigma}^1$. Further, for this case, we want to derive the necessary conditions for any ϵ-edge DP recovery algorithms that

$$\Pr(\hat{\boldsymbol{\sigma}}(\mathbf{A}) \in \mathcal{S}_1) \geq 1 - \eta$$

which implies that

$$\sum_{i=2}^{m} \Pr(\hat{\boldsymbol{\sigma}}(\mathbf{A}) \in \mathcal{S}_i) \leq \eta, \tag{4}$$

where $\mathbf{A} \sim \operatorname{SBM}(\boldsymbol{\sigma}^1, n, p, q)$.

We next individually lower bound each term in Eqn. (4). To do so, we first invoke the group privacy property of DP [4] and show that for any two adjacency matrices $\mathbf{A}$ and $\mathbf{A}'$, we have

$$\Pr(\hat{\boldsymbol{\sigma}}(\mathbf{A}) \in \mathcal{S}) \leq e^{\epsilon \operatorname{Ham}(\mathbf{A},\mathbf{A}')} \Pr(\hat{\boldsymbol{\sigma}}(\mathbf{A}') \in \mathcal{S}),$$

for any measurable set $\mathcal{S} \subseteq \{\pm 1\}^n$. For each $i = 1, 2, \cdots, m$, taking the expectation with respect to the coupling distribution $\Pi(\mathbf{A}, \mathbf{A}')$ between $\mathbf{A}$ and $\mathbf{A}'$ and setting $\mathcal{S} = \mathcal{S}_i$, yields the following:

$$\begin{aligned} &\mathbb{E}_{\mathbf{A},\mathbf{A}' \sim \Pi(\mathbf{A},\mathbf{A}')} \left[\Pr(\hat{\boldsymbol{\sigma}}(\mathbf{A}) \in \mathcal{S}_i)\right] \\ &\leq \mathbb{E}_{\mathbf{A},\mathbf{A}' \sim \Pi(\mathbf{A},\mathbf{A}')} \left[e^{\epsilon \operatorname{Ham}(\mathbf{A},\mathbf{A}')} \Pr(\hat{\boldsymbol{\sigma}}(\mathbf{A}') \in \mathcal{S}_i)\right] \end{aligned}$$

which implies that

$$\begin{aligned} &\Pr(\hat{\boldsymbol{\sigma}}(\mathbf{A}) \in \mathcal{S}_i) \\ &\leq \mathbb{E}_{\mathbf{A},\mathbf{A}' \sim \Pi(\mathbf{A},\mathbf{A}')} \left[e^{\epsilon \operatorname{Ham}(\mathbf{A},\mathbf{A}')} \Pr(\hat{\boldsymbol{\sigma}}(\mathbf{A}') \in \mathcal{S}_i)\right] \\ &\overset{(a)}{\leq} \left(\mathbb{E}_{\mathbf{A},\mathbf{A}' \sim \Pi(\mathbf{A},\mathbf{A}')} \left[e^{2\epsilon \operatorname{Ham}(\mathbf{A},\mathbf{A}')}\right]\right)^{1/2} \\ &\quad \times \left(\mathbb{E}_{\mathbf{A},\mathbf{A}' \sim \Pi(\mathbf{A},\mathbf{A}')} \left[\Pr^2(\hat{\boldsymbol{\sigma}}(\mathbf{A}') \in \mathcal{S}_i)\right]\right)^{1/2} \\ &\leq \left(\mathbb{E}_{\mathbf{A},\mathbf{A}' \sim \Pi(\mathbf{A},\mathbf{A}')} \left[e^{2\epsilon \operatorname{Ham}(\mathbf{A},\mathbf{A}')}\right]\right)^{1/2} \\ &\quad \times \left(\Pr(\hat{\boldsymbol{\sigma}}(\mathbf{A}') \in \mathcal{S}_i)\right)^{1/2}, \end{aligned}$$

from which we further get that

$$\begin{aligned} 1 - \eta \overset{(b)}{\leq} &\left(\mathbb{E}_{\mathbf{A},\mathbf{A}' \sim \Pi(\mathbf{A},\mathbf{A}')} \left[e^{2\epsilon \operatorname{Ham}(\mathbf{A},\mathbf{A}')}\right]\right)^{1/2} \\ &\times \left(\Pr(\hat{\boldsymbol{\sigma}}(\mathbf{A}') \in \mathcal{S}_i)\right)^{1/2}, \end{aligned}$$

and

$$\begin{aligned} (1-\eta)^2 \leq &\left(\mathbb{E}_{\mathbf{A},\mathbf{A}' \sim \Pi(\mathbf{A},\mathbf{A}')} \left[e^{2\epsilon \operatorname{Ham}(\mathbf{A},\mathbf{A}')}\right]\right) \\ &\times \Pr(\hat{\boldsymbol{\sigma}}(\mathbf{A}') \in \mathcal{S}_i), \end{aligned} \tag{5}$$

where step (a) follows from applying Cauchy-Schwartz inequality. In step (b), we invoked the condition in Eqn. (3). □

We next focus on computing the term $\mathbb{E}_{\mathbf{A},\mathbf{A}'\sim\Pi(\mathbf{A},\mathbf{A}')}\left[e^{2\epsilon\,\mathrm{Ham}(\mathbf{A},\mathbf{A}')}\right]$ in the expression of Lemma III.1. This together with Lemma III.1 leads to the general lower bound for the number of vertices in the graph.

It is worthwhile mentioning that the Hamming distance between two labeling vectors $\boldsymbol{\sigma}$ and $\boldsymbol{\sigma}'$ (each of size n) directly determines how many rows in the adjacency matrices $\mathbf{A}$ and $\mathbf{A}'$ are generated from the same versus different distributions. More precisely, we have two cases: case (1): $\mathrm{Ham}(\boldsymbol{\sigma},\boldsymbol{\sigma}')$ rows in $\mathbf{A}$ and $\mathbf{A}'$ have elements generated from different distributions, and case (2): $n-\mathrm{Ham}(\boldsymbol{\sigma},\boldsymbol{\sigma}')$ rows have elements from the same distribution.

Case (1): In this case, the probability the corresponding elements in the two matrices $\mathbf{A}$ and $\mathbf{A}'$ are different is $\bar{q}=1-(q\cdot p+p\cdot q)=1-2pq$.

Case (2): In this case, the probability the corresponding elements in the two matrices $\mathbf{A}$ and $\mathbf{A}'$ are same is $\bar{p}=1-(p^2+q^2)$.

Guided by these insights, we are ready to prove our general lower bound.

Theorem III.1 (Necessary Condition)**.** *Define* $\Delta \triangleq e^{2\epsilon}+\left(1-e^{2\epsilon}\right)\left(p^2+q^2\right)-1$. *Suppose there exists an ϵ-edge DP mechanism such that, for any ground truth labeling vector $\boldsymbol{\sigma}^*$ and for $G\sim\mathrm{SBM}(\boldsymbol{\sigma}^*,n,p,q)$, the mechanism outputs $\hat{\boldsymbol{\sigma}}$ satisfying the (β,η)-accurate recovery condition* (1)*. Then, a necessary condition is that n must satisfy*

$$n\geq\frac{\beta A+\sqrt{\beta^2A^2+8\left(1-8\beta\right)\Delta B}}{8\beta\left(1-8\beta\right)\Delta}.$$

where $A=\log\left(\frac{1}{8e^{\beta}}\right)$ *and* $B=\log\left(\frac{1}{\eta}\right)$.

Proof. We next focus in calculating the term $M_{\mathrm{Ham}(\mathbf{A},\mathbf{A}')}(2\epsilon)\triangleq\mathbb{E}_{\mathbf{A},\mathbf{A}'\sim\Pi(\mathbf{A},\mathbf{A}')}\left[e^{2\epsilon\,\mathrm{Ham}(\mathbf{A},\mathbf{A}')}\right]$ with the following set of steps:

$$\begin{aligned}&M_{\mathrm{Ham}(\mathbf{A},\mathbf{A}')}(2\epsilon)\\&=\left(M_{\mathrm{Ham}(\mathbf{A},\mathbf{A}')\text{:same dist.}}(2\epsilon)\right)^{(n-\mathrm{Ham}(\boldsymbol{\sigma},\boldsymbol{\sigma}'))}\\&\quad\times\left(M_{\mathrm{Ham}(\mathbf{A},\mathbf{A}')\text{:different dist.}}(2\epsilon)\right)^{\mathrm{Ham}(\boldsymbol{\sigma},\boldsymbol{\sigma}')},\end{aligned}$$

where,

$$\begin{aligned}M_{\mathrm{Ham}(\mathbf{A},\mathbf{A}')\text{:same dist.}}(2\epsilon)&=e^{2\epsilon}\bar{p}+(1-\bar{p}),\\M_{\mathrm{Ham}(\mathbf{A},\mathbf{A}')\text{:different dist.}}(2\epsilon)&=e^{2\epsilon}\bar{q}+(1-\bar{q}).\end{aligned}$$

We then can readily show that,

$$\begin{aligned}&M_{\mathrm{Ham}(\mathbf{A},\mathbf{A}')}(2\epsilon)\\&\leq\left(M_{\mathrm{Ham}(\mathbf{A},\mathbf{A}')\text{:same dist.}}(2\epsilon)\right)^{(n-\mathrm{Ham}(\boldsymbol{\sigma},\boldsymbol{\sigma}'))\cdot\mathrm{Ham}(\boldsymbol{\sigma},\boldsymbol{\sigma}')}\\&=(e^{2\epsilon}+(1-e^{2\epsilon})(p^2+q^2))^{(n-\mathrm{Ham}(\boldsymbol{\sigma},\boldsymbol{\sigma}'))\cdot\mathrm{Ham}(\boldsymbol{\sigma},\boldsymbol{\sigma}')}.\end{aligned}$$

Plugging (6) in (5) yields the following:

$$\Pr(\hat{\boldsymbol{\sigma}}(\mathbf{A}')\in\mathcal{S}_i)\geq\frac{(1-\eta)^2}{M_{\mathrm{Ham}(\mathbf{A},\mathbf{A}')}(2\epsilon)}.$$

Finally, we lower bound the packing number m with respect to err rate. Building upon the framework established in [19], we can readily demonstrate that

$$m\geq\frac{1}{2}\cdot\frac{|\mathcal{B}_{\mathrm{Ham}}(\boldsymbol{\sigma}^*,4\beta n)|}{|\mathcal{B}_{\mathrm{Ham}}(\boldsymbol{\sigma}^*,2\beta n)|},$$

where $\mathcal{B}_{\mathrm{Ham}}(\boldsymbol{\sigma}^*,t\beta n)=\{\boldsymbol{\sigma}\in\{\pm1\}^n:\mathrm{Ham}(\boldsymbol{\sigma},\boldsymbol{\sigma}^*)\leq\beta\}$ and $|\mathcal{B}_{\mathrm{Ham}}(\boldsymbol{\sigma}^*,t\beta n)|$ is the number of vectors within the Hamming distance of $t\beta n$ from $\boldsymbol{\sigma}^*$ for $t>0$. It includes all vectors that can be obtained by flipping any $t\beta n$ elements of $\boldsymbol{\sigma}^*$. Thus, we can further lower bound m as

$$m\geq\frac{1}{2}\cdot\frac{\binom{n}{4\beta n}}{\binom{n}{2\beta n}}\geq\frac{1}{2}\cdot\left(\frac{1}{8e\beta}\right)^{2\beta n}.$$

Recall that, we have

$$(m-1)\cdot\frac{(1-\eta)^2}{M_{\mathrm{Ham}(\mathbf{A},\mathbf{A}')}(2\epsilon)}\leq\eta. \tag{6}$$

Taking the logarithm for both sides of (6), we have

$$\begin{aligned}&8\beta n(n-8\beta n)\cdot\log(e^{2\epsilon}+(1-e^{2\epsilon})(p^2+q^2))\\&\geq2\beta n\cdot\log\left(\frac{1}{8e\beta}\right)+\log\left(\frac{1}{\eta}\right)\end{aligned}$$

which implies

$$\begin{aligned}&8\beta n(n-8\beta n)\cdot\log(e^{2\epsilon}+(1-e^{2\epsilon})(p^2+q^2))\\\geq&2\beta n\cdot\log\left(\frac{1}{8e\beta}\right)+\log\left(\frac{1}{\eta}\right)\end{aligned}$$

and further

$$\begin{aligned}&\Rightarrow\log(e^{2\epsilon}+(1-e^{2\epsilon})(p^2+q^2))\\&\geq\frac{\log\left(\frac{1}{8e\beta}\right)}{4(n-8\beta n)}+\frac{\log\left(\frac{1}{\eta}\right)}{8\beta n(n-8\beta n)}\end{aligned}$$

and

$$\begin{aligned}&e^{2\epsilon}+(1-e^{2\epsilon})(p^2+q^2)-1\\&\geq\frac{\log\left(\frac{1}{8e\beta}\right)}{4(n-8\beta n)}+\frac{\log\left(\frac{1}{\eta}\right)}{8\beta n(n-8\beta n)}\end{aligned} \tag{7}$$

Solving the last inequality of Eq. (7) for n, we get the lower bound

$$n\geq\frac{\beta A+\sqrt{\beta^2A^2+8\left(1-8\beta\right)\Delta B}}{8\beta\left(1-8\beta\right)\Delta}.$$

where $A=\log\left(\frac{1}{8e^{\beta}}\right)$ and $B=\log\left(\frac{1}{\eta}\right)$.

□

We next present our three private spectral-based algorithms and outline their respective accuracy guarantees. The privacy analysis of the methods is based on standard techniques of DP [4].

B. Graph Perturbation-Based Mechanism

We next provide a novel analysis for the spectral method applied on the randomized response released edge-DP adjacency matrix.

Definition 4 (Randomized Response Perturbation Mechanism)**.** Let $\mathbf{A}$ be the adjacency matrix of the graph. Under Warner's randomized response [20] with a uniform perturbation parameter $\mu = 1/(e^{\epsilon}+1)$, the perturbed adjacency matrix is given by

$$\hat{\mathbf{A}} = \mathbf{A} + \mathbf{E},$$

where $\mathbf{E}$ is a symmetric perturbation matrix with i.i.d. entries:

$$E_{ij} = \begin{cases} 0, & \text{with probability } 1-\mu, \\ 1-2A_{ij}, & \text{with probability } \mu. \end{cases} \tag{8}$$

The noise matrix $\mathbf{E}$ defined in Eqn. (8) clearly satisfies

$$\mathbb{E}[E_{ij}] = \mu\left(1-2A_{ij}\right),$$
$$\mathrm{var}(E_{ij}) = \mu\left(1-\mu\right)\left(1-2A_{ij}\right)^2.$$

The analysis is based on decomposing the random release of the adjacency matrix to a deterministic and random terms. We first state some auxilisry results needed for the main result.

1) Auxiliary Results for Graph Perturbation Mechanism: The analysis of the graph perturbation mechanism is based on a decomposition of the error into deterministic and random terms and to this end we first need high-probability bounds for the terms $\|\mathbf{L} - \mathbb{E}[\mathbf{L}]\|$ and $\|\hat{\mathbf{L}} - \mathbf{L}\|$.

Lemma III.2 (Concentration of Laplacian Matrices [21])**.** *Let $\mathbf{L}$ be a Laplacian whose elements are drawn from a nonhomogeneous Erdös-Rényi model, where each edge (i,j) is generated independently with probability p_{ij}. Then, the following holds true with probability at least $1-\eta$,*

$$\begin{aligned}&\|\mathbf{L} - \mathbb{E}[\mathbf{L}]\| \\ \leq& C_{III.2}\left(\sqrt{n\max_{(i,j)} p_{ij}\log(n/\eta)} + \log(n/\eta)\right),\end{aligned}$$

where $C_{III.2}$ is a universal constant, and $\eta \geq n^{-10}$.

Next, we introduce two essential lemmas that underpin our main results.

Lemma III.3 (Concentration of Perturbed Laplacian Matrices via Randomized Response Mechanism)**.** *Given a Laplacian matrix $\mathbf{L}$ and its perturbed version $\hat{\mathbf{L}}$ via a randomized response mechanism. The following holds true for a universal constant $C_{III.3}$:*

$$\|\hat{\mathbf{L}} - \mathbf{L}\| \leq C_{III.3}\sqrt{(\sum_{i<j}\mu_{ij}(1-\mu_{ij}))\cdot\log(n/\eta)}$$

with probability at least $1-\eta$.

Proof. We are now interested in upper bounding the operator norm (spectral norm) $\|\hat{\mathbf{L}} - \mathbf{L}\|$.

To apply standard matrix concentration inequalities, define the centered random variables:

$$\begin{aligned}X_{ij} &= (\hat{A}_{ij} - A_{ij}) - \mathbb{E}[\hat{A}_{ij} - A_{ij}] \\ &= (\hat{A}_{ij} - A_{ij}) - \mu_{ij}(1-2A_{ij}).\end{aligned}$$

Now, X_{ij} are independent (for distinct edges) zero-mean bounded random variables with $|X_{ij}| \leq 1$.

The matrix of interest is:

$$\hat{\mathbf{A}} - \mathbf{A} = \sum_{i<j}(\hat{A}_{ij} - A_{ij})(\mathbf{E}_{ij} + \mathbf{E}_{ji}),$$

where $\mathbf{E}_{ij}$ is the matrix unit with a 1 in position (i,j) and 0 elsewhere.

Inserting the centered version

$$\hat{\mathbf{A}} - \mathbf{A} = \sum_{i<j}[X_{ij} + \mu_{ij}(1-2A_{ij})](\mathbf{E}_{ij} + \mathbf{E}_{ji})$$

we have

$$\begin{aligned}\hat{\mathbf{A}} - \mathbf{A} = &\underbrace{\sum_{i<j} X_{ij}(\mathbf{E}_{ij} + \mathbf{E}_{ji})}_{\mathbf{Z}} \\ &+ \underbrace{\sum_{i<j}\mu_{ij}(1-2A_{ij})(\mathbf{E}_{ij} + \mathbf{E}_{ji})}_{\mathbf{M}}.\end{aligned}$$

Here, $\mathbf{Z}$ is a zero-mean random symmetric matrix, and $\mathbf{M}$ is a deterministic offset matrix.

Next, consider the degree matrix changes. Since $\hat{D}_{ii} = \sum_j \hat{A}_{ij}$,

$$\hat{\mathbf{D}} - \mathbf{D} = \mathrm{diag}\left(\sum_j(\hat{A}_{ij} - A_{ij})\right).$$

This can also be expressed as a sum of independent random vectors along the diagonal. By a similar reasoning, the change in degree matrix can also be decomposed into a zero-mean part plus a deterministic part.

Overall, we have:

$$\begin{aligned}\mathbf{E} = \hat{\mathbf{L}} - \mathbf{L} &= (\hat{\mathbf{D}} - \mathbf{D}) - (\hat{\mathbf{A}} - \mathbf{A}) \\ &= (\hat{\mathbf{D}} - \mathbf{D}) - \mathbf{Z} - \mathbf{M}.\end{aligned}$$

We will focus on the zero-mean part to apply a matrix Bernstein-type inequality. In our case, the random variables X_{ij} correspond to modifications of single entries. The matrices we sum over are rank-2 updates (like $(\mathbf{E}_{ij} + \mathbf{E}_{ji})$ and their diagonal adjustments). Each such update has operator norm at most 2.

The variance parameter σ^2 depends on sums of variances $\mu_{ij}(1-\mu_{ij})$. Summing over all edges, we get:

$$\sigma^2 = O\left(\sum_{i<j}\mu_{ij}(1-\mu_{ij})\right).$$

In a sparse or moderately dense regime, this is often $O(n)$, but depends on the distribution of p_{ij}.

Thus, with high probability,

$$\begin{aligned}\|\mathbf{Z}\|_2 &= O(\sqrt{\sigma^2\log n}) \\ &= O\left(\sqrt{(\sum_{i<j}\mu_{ij}(1-\mu_{ij}))\log n}\right).\end{aligned}$$

Recall $\mathbf{E} = \hat{\mathbf{L}} - \mathbf{L} = (\hat{\mathbf{D}} - \mathbf{D}) - \mathbf{Z} - \mathbf{M}$. A similar argument applies to $(\hat{\mathbf{D}} - \mathbf{D})$, which also can be viewed as a sum of

independent diagonal updates. The matrix Bernstein or Vector Bernstein inequality can handle these diagonal terms similarly.

The deterministic part $\mathbf{M}$ is known and can be bounded separately. The main random fluctuation is in $\mathbf{Z}$ and $(\hat{\mathbf{D}} - \mathbf{D})$. Combining these results, we obtain a concentration inequality of the form:

$$\mathbb{P}\left(\|\hat{\mathbf{L}} - \mathbf{L}\| \geq t\right) \leq 2n \exp\left(\frac{-ct^2}{\sum_{i<j} \mu_{ij}(1-\mu_{ij}) + t}\right),$$

for some absolute constant c. For t large enough, this yields a bound like:

$$\|\hat{\mathbf{L}} - \mathbf{L}\| \leq C\sqrt{\left(\sum_{i<j} \mu_{ij}(1-\mu_{ij})\right)\log(n/\eta)}$$

with probability at least $1 - \eta$. □

2) Main Result for Graph Perturbation Mechanism: We are now ready to state the main results for the graph perturbation mechanism.

Theorem III.2 (Distance to Ground Truth Labels)**.** *Let $\mathbf{u}_2$ and $\hat{\mathbf{u}}_2$ be the second eigenvectors of the unperturbed Laplacian matrix $\mathbf{L}$ and the privatized Laplacian matrix $\hat{\mathbf{L}}$ obtained via Warner's randomized response, respectively. Then, with probability at least $1 - 3\eta$, we have*

$$\begin{aligned}&\min_{s\in\{\pm 1\}} \|\hat{\mathbf{u}}_2 - s\mathbf{u}_2\|_2 \\ &\leq \frac{4\sqrt{2}}{n(p-q)} \cdot \Bigg(qn \;+\; \sqrt{8\mu(1-\mu)n\log(2/\eta)} \\ &\qquad\qquad +\; \frac{4}{3\sqrt{n}}\log(2/\eta)\Bigg).\end{aligned}$$

Proof. The proof is based on a *brief perturbation analysis* of the difference $\Delta\mathbf{L} = \hat{\mathbf{L}} - \mathbf{L}$. Starting with the definition,

$$\Delta\mathbf{L}\,\mathbf{u}_2 \;=\; (\hat{\mathbf{L}} - \mathbf{L})\,\mathbf{u}_2 \;=\; (\Delta\mathbf{D})\,\mathbf{u}_2 \;-\; \mathbf{E}\,\mathbf{u}_2,$$

observe that $\Delta\mathbf{D}$ is diagonal, and hence its action on $\mathbf{u}_2$ can be recast in terms of $\mathbf{E}$. In particular, we derive the componentwise expression

$$(\Delta\mathbf{L}\,\mathbf{u}_2)_i \;=\; \sum_{j=1}^{n} E_{ij}\,(u_{2i} - u_{2j}),$$

thereby yielding

$$\Delta\mathbf{L}\,\mathbf{u}_2 \;=\; \sum_{i=1}^{n}\sum_{j=1}^{n} E_{ij}\,(u_{2i} - u_{2j})\,\mathbf{e}_i,$$

where $\mathbf{e}_i$ is the i-th standard basis vector.

Next, we decompose E_{ij} into its mean and zero-mean parts:

$$E_{ij} \;=\; \mu_{ij}\,(1 - 2A_{ij}) \;+\; \tilde{E}_{ij}, \quad \text{where} \quad \mathbb{E}\big[\tilde{E}_{ij}\big] \;=\; 0.$$

Define the random vectors

$$\mathbf{x}_{ij} \;=\; \tilde{E}_{ij}\,(u_{2j} - u_{2i})\,(\mathbf{e}_i - \mathbf{e}_j),$$

so that the difference of Laplacians acting on $\mathbf{u}_2$ becomes

$$\Delta\mathbf{L}\,\mathbf{u}_2 \;=\; \sum_{i<j} \mathbf{x}_{ij} \;+\; \mathbf{d},$$

where $\mathbf{d}$ is a deterministic vector resulting from the means $\mu_{ij}\,(1 - 2A_{ij})$.

To control the magnitude of the random sum $\sum_{i<j} \mathbf{x}_{ij}$, we use the vector Bernstein Inequality [22]. Specifically, two crucial quantities must be bounded:

(i) The norm of each summand $\mathbf{x}_{ij}$. Since $|u_{2i} - u_{2j}| \leq D_u$ and $\tilde{E}_{ij}$ takes values in a bounded set, each summand obeys

$$\|\mathbf{x}_{ij}\|_2 \;\leq\; 2\,|\tilde{E}_{ij}|\,D_u \;\leq\; 2\,D_u \;\triangleq\; L_{ij}.$$

Hence $L_{ij} \leq 2D_u$. It is worth highlighting that we normalize the labeling vectors by a $1/\sqrt{n}$ factor. In this case, we have $D_u = 2/\sqrt{n}$.

(ii) The sum of variances. We compute

$$\sigma^2 \;=\; \sum_{i<j} \mathbf{E}\big[\|\mathbf{x}_{ij}\|_2^2\big] \;=\; 4\sum_{i<j} \mu_{ij}\,(1-\mu_{ij})\,(u_{2i} - u_{2j})^2.$$

By the vector Bernstein Inequality, for any $t > 0$,

$$\Pr\Big(\Big\|\sum_{i<j} \mathbf{x}_{ij}\Big\|_2 \;\geq\; t\Big) \;\leq\; 2\,\exp\Big(\frac{-t^2/2}{\sigma^2 + (L\,t)/3}\Big),$$

where $L = \max_{i<j} L_{ij}$. Solving in terms of a confidence parameter η, we obtain that with probability at least $1 - \eta$,

$$\Big\|\sum_{i<j} \mathbf{x}_{ij}\Big\|_2 \;\leq\; \sqrt{2\,\sigma^2\,\log(2/\eta)} \;+\; \frac{L}{3}\,\log(2/\eta).$$

Since $\Delta\mathbf{L}\,\mathbf{u}_2$ is this sum plus the deterministic term $\mathbf{d}$, we conclude:

$$\big\|\Delta\mathbf{L}\,\mathbf{u}_2\big\|_2 \;\leq\; \|\mathbf{d}\|_2 \;+\; \sqrt{2\,\sigma^2\,\log(2/\eta)} \;+\; \frac{L}{3}\,\log(2/\eta)$$

with probability at least $1 - \eta$.

Substituting back into the Generalized Davis-Kahan Theorem [23], we obtain:

$$\begin{aligned}&\|\mathbf{u}_2 - \hat{\mathbf{u}}_2\|_2 \\ &\leq \sqrt{2} \cdot \frac{\|\mathbf{d}\|_2 + \sqrt{2\sigma^2\log(2/\eta)} + \dfrac{L}{3}\log(2/\eta)}{|\hat{\lambda}_3 - \lambda_2|}.\end{aligned}$$

By Weyl's inequality [24], we can readily show the following lower bound on $\hat{\lambda}_3 - \lambda_2$:

$$\hat{\lambda}_3 - \lambda_2 \geq \frac{n(p-q)}{2} - 2\|\mathbf{L} - \mathbb{E}[\mathbf{L}]\| - \|\hat{\mathbf{L}} - \mathbf{L}\|. \quad (9)$$

Continuing from the provided inequalities and incorporating the concentration results from Lemma III.2 and Lemma III.3, we arrive at the claim.

□

The above result holds under the following separation condition between p and q. Specifically, for a universal constant $C = 4 \cdot \max(2C_{III.2}, C_{III.3})$, with $C_{III.2}$ and $C_{III.3}$ given in Lemmas III.2 and III.3, respectively, it is required that

$$n(p-q) \;\geq\; C\Big[\mathcal{T} \;+\; \frac{n}{\sqrt{2}}\sqrt{\mu(1-\mu)}\,\sqrt{\log(n/\eta)}\Big],$$

where $\mathcal{T} \triangleq \sqrt{np\,\log(n/\eta)} \;+\; \log(n/\eta)$.

This separation condition ensures that the difference $p - q$ is large enough relative to n and the parameters μ, η so that the upper bound in Eqn. (9) on the distance to the ground

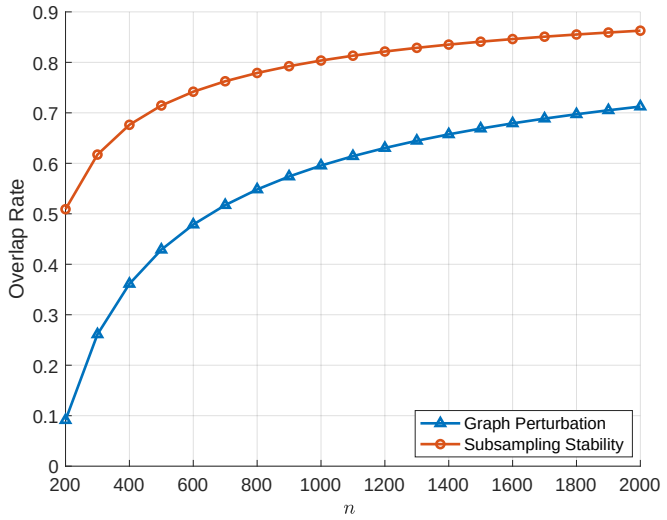

Fig. 1: Overlap rate vs n, for $\epsilon = 1$, $p = 0.25$, $q = 0.0025$, and $\delta = 10^{-6}$. A fair comparison is ensured by setting the same total failure probability $\delta_{\text{failure}} = 0.01$ for both mechanisms. For the Graph-perturbation mechanism, the confidence level η is directly set to δ_{failure}. For the Subsampling stability mechanism, η is adjusted to satisfy the condition $3m\eta = \delta_{\text{failure}}$.

Algorithm 2 Subsampling Stability Mechanism

Input: Graph $G = (\mathcal{V}, E)$, privacy budget ϵ, δ, graph structure properties.
Output: Private labelling vector $\hat{\boldsymbol{\sigma}}$.
Compute the base sampling probability $q_s \leftarrow \min(1, \epsilon/(32\log(n)))$.
Compute $m \leftarrow \lceil \log(n/\delta)/q_s^2 \rceil$.
Subsample m subgraphs $\{G_1, G_2, \ldots, G_m\}$ using q_s.
Compute the label vectors $\bar{\boldsymbol{\sigma}} = (\hat{\boldsymbol{\sigma}}(G_1), \ldots, \hat{\boldsymbol{\sigma}}(G_m))$.
Aggregate the label vectors via majority voting and compute the stability score:

$$\hat{d} \leftarrow \frac{\text{count}_{(1)} - \text{count}_{(2)}}{4mq_s} - 1.$$

Add Laplace noise to ensure privacy:

$$\tilde{d} \leftarrow \hat{d} + \text{Lap}(0, 1/\epsilon).$$

if $\tilde{d} > \log(1/\delta)/\epsilon$ **then**
  Output $\hat{\boldsymbol{\sigma}}_{\text{final}} = \text{mode}(\bar{\boldsymbol{\sigma}})$.
else
  Output $\perp$ (a random label vector).
end if

truth labels (denoted as $C_1(\epsilon, \eta)$) is meaningful and captures the success of the privatized spectral method.

We next translate the norm difference bound into a statement about the overlap (fraction of correctly identified labels). This connection allows us to interpret the effects of the privacy mechanism directly in terms of the clustering accuracy.

Lemma III.4 (Overlap Rate for Graph Perturbation Mechanism)**.** *Consider the graph perturbation mechanism, which satisfies ϵ-edge DP. Define the overlap rate between the ground truth labels $\boldsymbol{\sigma}^*$ and the estimated labels $\hat{\boldsymbol{\sigma}}(G)$ obtained via the private spectral method as*

$$\text{overlap rate}\big(\hat{\boldsymbol{\sigma}}(G), \boldsymbol{\sigma}^*\big) \triangleq 1 - \text{err rate}\big(\hat{\boldsymbol{\sigma}}(G), \boldsymbol{\sigma}^*\big).$$

Under the conditions of Theorem III.2, we have

$$\text{overlap rate}\big(\hat{\boldsymbol{\sigma}}(G), \boldsymbol{\sigma}^*\big) \geq 1 - \frac{C_{III.2}(\epsilon, \eta)}{8}$$

with probability at least $1 - 3\eta$, where $C_{III.2}(\epsilon, \eta)$ is determined by the right hand side of the inequality of Theorem III.2.

C. Subsampling Stability-Based Mechanism

The key idea in this algorithm is to create m *correlated* subgraphs $\{G_1, G_2, \ldots, G_m\}$ of the original graph G where each subgraph G_ℓ is generated by randomly subsampling with replacement of the edges in G with probability q_s. We then apply our *non-private* spectral method $\hat{\boldsymbol{\sigma}}(G_\ell)$ on each subgraph G_ℓ. The labeling vectors $\hat{\boldsymbol{\sigma}}(G_\ell)$ are then represented on a histogram. Now, define $\text{count}(\boldsymbol{\sigma}) \triangleq |\{k \in [m] : \hat{\boldsymbol{\sigma}}(G_k) = \boldsymbol{\sigma}\}|$. As shown in [4], the stability of the histogram is proportional to the difference between the most frequent bin (i.e., the mode) and the second most frequent bin. In other words, the most frequent outcome of the histogram agrees with the outcome of the original graph with high probability under an appropriate choice of the SBM parameters (p, q), the privacy budget (ϵ, δ), the edge sampling probability q_s, and the number of subsampled weighted graphs m. For further details on the mechanism and stability of community detection algorithms over SBMs, please refer to our previous work in [11]. We summarize the mechanism in Algorithm 2.

1) Auxiliary Result for the Subsampling Stability Mechanism: The proof of the main theorem for the Subsampling Stability Mechanism is based on the same perturbation analysis as the analysis of the graph perturbation mechanism presented in the previous subsection. For this analysis, we will need the following concentration bound for the Laplacian matrix released by the subsampling stability mechanism.

Lemma III.5 (Concentration of Subsampled Laplacian Matrices)**.** *Consider an undirected graph with n nodes and edge set E, represented by the Laplacian matrix $\mathbf{L}$. Let $\hat{\mathbf{L}}$ be the Laplacian matrix of a subsampled graph, obtained by independently including each edge $(i, j) \in E$ with probability*

$$q_{s,ij} = q_s \cdot (1 - r(i, j)),$$

where q_s is a base sampling probability, $r(i, j) \in [0, 1]$ is a removal probability determined by the edge-specific subsampling mechanism. The following concentration bound holds for a universal constant $C_{III.5}$:

$$\|\hat{\mathbf{L}} - \mathbf{L}\| \leq C_{III.5} \sqrt{\Big(\sum_{i<j} q_{s,ij}(1 - q_{s,ij})\Big) \cdot \log(n/\eta)},$$

with probability at least $1 - \eta$.

2) Main Result for the Subsampling Stability Mechanism:

Theorem III.3 (Distance to Ground Truth Labels)**.** *Let $\mathbf{u}_2$ and $\hat{\mathbf{u}}_2$ be the second eigenvectors of the unperturbed Laplacian matrix $\mathbf{L}$ and the Laplacian matrix $\hat{\mathbf{L}}$ of the subsampled graph*

G_ℓ obtained via the subsampling mechanism, respectively. Then, with probability at least $1-3\eta$, we have

$$\min_{s\in\{\pm 1\}} \|\hat{\mathbf{u}}_2 - s\mathbf{u}_2\|_2 \leq \frac{4\sqrt{2}}{n(p-q)} \cdot \left(\|\mathbf{d}\|_2 + \sqrt{2\sigma^2 \log(2/\eta)} + \frac{L}{3}\log(2/\eta) \right) \triangleq C_{III.3}(\epsilon,\eta),$$

where, $\mathbf{d}$ is a deterministic part of the perturbation, derived from the adjusted edge sampling mechanism where $\|\mathbf{d}\|_2 = \frac{2q_s}{\sqrt{n}}\sqrt{\frac{q}{p+q}\cdot|E|}$, $|E|$ represents the total number of edges in the graph, and $\sigma^2 = \frac{16}{n}\cdot q_s(1-q_s)|E_{inter}|$, where $|E_{inter}|$ denotes the number of inter-community edges. This is the variance of the edge sampling noise, with q_s which can be further upper bounded as $\sigma^2 \leq \frac{4}{n}\cdot\frac{q}{p+q}\cdot|E|$, and $L = \max_{(i,j)\in E}\|\mathbf{x}_{ij}\|_2 \leq 4/\sqrt{n}$ is the upper bound on the norm of the noise components.

Proof. The proof is based on the same perturbation analysis as the analysis of the graph perturbation mechanism given in the proof of Thm. III.2, with the bound of Lemma III.5 used instead for the high-probability bound for the term $\|\hat{\mathbf{L}} - \mathbf{L}\|$. □

The above result holds under the following separation condition between p and q. Specifically, for a universal constant $C = 4\cdot\max(2C_{III.2}, C_{III.5})$, with $C_{III.2}$ and $C_{III.5}$ given in Lemmas III.2 and III.5, respectively, it is required that

$$n(p-q) \;\geq\; C\Big[\mathcal{T} \;+\; \frac{n}{\sqrt{2}}\sqrt{q_s(1-q_s)}\sqrt{\log(n/\eta)}\Big],$$

where $\mathcal{T} \triangleq \sqrt{np\log(n/\eta)} \;+\; \log(n/\eta)$.

Remark III.1 (Impact of q_s on the Distance Bound). The edge sampling probability q_s plays a crucial role in the trade-off between privacy and spectral accuracy. As q_s decreases:

- The deterministic perturbation $\|\mathbf{d}\|_2$ and the variance σ^2 decrease, reflecting reduced magnitudes of the subsampling-induced noise.
- However, the spectral gap $\hat{\lambda}_3 - \lambda_2$ (refer to Eqn. (9)) also decreases significantly due to the increased instability of the subsampled graph, particularly when critical edges are removed.

Here, $\hat{\lambda}_3$ is the third smallest eigenvalue of the subsampled graph G_ℓ. This reduction in the spectral gap dominates the bound, leading to an overall increase in the upper bound on the distance $\|\hat{\mathbf{u}}_2 - s\mathbf{u}_2\|_2$.

Lemma III.6 (Overlap Rate for Subsampling Stability Mechanism)**.** *Consider the subsampling stability-based mechanism, which satisfies $(2\epsilon,\delta)$-edge DP. The overlap rate between the ground truth labels $\boldsymbol{\sigma}^*$ and the final estimated labels $\hat{\boldsymbol{\sigma}}_{final}$ obtained via majority voting on m correlated subgraphs $\{G_1,\ldots,G_m\}$ satisfies:*

$$\text{overlap rate}(\hat{\boldsymbol{\sigma}}_{final}, \boldsymbol{\sigma}^*) \geq 1 - \frac{C_{III.3}(\epsilon,\eta)}{4} - O\left(\frac{1}{\sqrt{m}}\right),$$

with probability at least $1-3m\eta$, where $C_{III.3}(\epsilon,\eta)$ is the error contribution from the non-private spectral method for individual subsampled Graph G_ℓ.

Proof Sketch. For a single subsampled graph G_ℓ, the labeling $\hat{\boldsymbol{\sigma}}(G_\ell)$ satisfies: $\text{overlap rate}(\hat{\boldsymbol{\sigma}}(G_\ell), \boldsymbol{\sigma}^*) \geq 1 - C_{III.3}$, implying that at most $C_{III.3}n$ nodes are mislabeled in each subgraph. Let $S_\ell \subseteq [n]$ denote the set of mislabeled nodes for subgraph G_ℓ, so $|S_\ell| \leq C_{III.3}n$. The final labeling $\hat{\boldsymbol{\sigma}}_{\text{final}}$ is obtained by taking a majority vote over the m subgraphs. A node i is mislabeled in $\hat{\boldsymbol{\sigma}}_{\text{final}}$ if it is mislabeled in more than $m/2$ subgraphs.

Let X_i denote the number of subgraphs in which node i is mislabeled. Since $\mathbb{E}[X_i] = mC_{III.3}$, the probability that $X_i \geq m/2$ can be bounded using Hoeffding's inequality:

$$\Pr(X_i \geq m/2) \leq \exp\left(-2\left(\frac{m}{2} - mC_{III.3}\right)^2/m\right).$$

Aggregating the probability of mislabeling across all nodes, the fraction of mislabeled nodes (denoted as $\mathcal{M}$) in the final labeling satisfies that $\frac{|\mathcal{M}|}{n} \leq 2C_{III.3} + O(1/\sqrt{m})$, where $2C_{III.3}$ accounts for the worst-case overlap between mislabeled nodes across subgraphs.

D. Noisy Power Iteration Method

As motivation, we build on the approach of [25], which applies power iteration to the centered adjacency matrix

$$\mathbf{B} = \;\mathbf{A} - \rho\,\mathbf{1}\mathbf{1}^\top, \quad \rho = \mathbf{1}^\top\mathbf{A}\mathbf{1}/n^2.$$

We tailor this procedure to the differential-privacy setting by replacing the standard power iteration with the *noisy power method* of [26]:

$$\mathbf{x}_t = \mathbf{B}\mathbf{y}_{t-1} + \mathbf{z}_t, \quad \mathbf{y}_{t-1} = \mathbf{x}_{t-1}/\|\mathbf{x}_{t-1}\|_2,$$

where $\mathbf{z}_t \sim \mathcal{N}(0, C^2\sigma^2\mathbf{I}_n)$ and C limits the 2-norm sensitivity of the product $\mathbf{B}\mathbf{y}_{t-1}$ with respect to change of a single edge in the graph. The DP guarantees are then obtained from a composition analysis of the Gaussian mechanism. To this end, we need to bound the sensitivty of the multiplication $\mathbf{B}\mathbf{y}_{t-1}$ w.r.t. to change of a single edge, as formalized in the following lemma.

Lemma III.7. *Let $\mathbf{A}$ and $\mathbf{A}'$ differ in a single element, and let $\mathbf{y}\in\mathbb{R}^n$ with $\|\mathbf{y}\|_2 = 1$. Then,*

$$\|\mathbf{B}\mathbf{y} - \mathbf{B}'\mathbf{y}\|_2 \leq \|\mathbf{y}\|_\infty + \frac{1}{n}.$$

Proof. Let $\mathbf{A}$ and $\mathbf{A}'$ differ in (i,j)th element. As $\mathbf{B} = \mathbf{A} - \rho\mathbf{1}\mathbf{1}^T$, where $\rho = \mathbf{1}^T\mathbf{A}\mathbf{1}/n^2$ (and similarly $\mathbf{B}' = \mathbf{A}' - \rho'\mathbf{1}\mathbf{1}^T$

TABLE I: Comparison of error bounds and summarization of the practical methods: Graph Perturbation vs. Noisy Power Iteration.

Method	Graph Perturbation Mechanism	Noisy Power Iteration
Error in the Second Eigenvector	$\min_{s\in\{\pm1\}} \|\hat{\mathbf{u}}_2 - s\mathbf{u}_2\|_2 = O\left(\frac{q}{p-q} + \frac{1}{e^{\epsilon/2}(p-q)\sqrt{n}}\right)$	$\min_{s\in\{\pm1\}} \|\mathbf{y}_N - s\mathbf{u}_2\|_2 = O\left(\frac{\sqrt{\log n}}{\epsilon(p-q)\sqrt{n}}\right)$
Noise Source	Randomized graph perturbation	Added noise in iterative updates
Time Complexity	$O(n^3)$ with full eigendecomposition $\tilde{O}(n^2)$ (leaving log factors) with power or Lanczos iteration	Dense SBM: $O\big((n\log n)N\big) = O(n(\log n)^2)$ Sparse SBM: $O(nN) = O(n\log n)$
Space Complexity	$O(n^2)$	Dense SBM: $O(n\log n)$, Sparse SBM: $O(n)$

Algorithm 3 Noisy Power Iteration

1: **Input:** Adjacency matrix $\mathbf{A} \in \{0,1\}^{n\times n}$, positive integer N
2: **Output:** Private labelling vector $\hat{\sigma}$.
3: Set $\rho = \mathbf{1}^T\mathbf{A}\mathbf{1}/n^2$, $\mathbf{B} = \mathbf{A} - \rho\mathbf{1}\mathbf{1}^T$.
4: Draw $\mathbf{y}_0$ randomly from the unit sphere $\mathbb{B}^{n-1}$.
5: **for** $t = 1$ to N **do**
6: $\quad \mathbf{x}_t = \mathbf{B}\mathbf{y}_{t-1} + \mathbf{z}_t$, $\mathbf{z}_t \sim \mathcal{N}\left(0, (\|\mathbf{y}_{t-1}\|_\infty + \frac{1}{n})^2\sigma^2\mathbf{I}_n\right)$.
7: $\quad \mathbf{y}_t = \mathbf{x}_t/\|\mathbf{x}_t\|_2$.
8: **end for**
9: $\hat{\boldsymbol{\sigma}} = \mathbf{y}_t/|\mathbf{y}_t|$.

for $\rho' = \mathbf{1}^T\mathbf{A}'\mathbf{1}/n^2$), we have that:

$$\begin{aligned}
\|\mathbf{B}\mathbf{y} - \mathbf{B}'\mathbf{y}\|_2 &\le \|\mathbf{A}\mathbf{y} - \mathbf{A}'\mathbf{y}\|_2 \\
&\quad + \left\|(\mathbf{1}^T\mathbf{A}\mathbf{1}/n^2 - \mathbf{1}^T\mathbf{A}'\mathbf{1}/n^2)\mathbf{1}\mathbf{1}^T\mathbf{y}\right\|_2 \\
&= \|y_j\mathbf{e}_i\|_2 + \frac{1}{n^2}\left\|(\mathbf{1}^T(\mathbf{A}-\mathbf{A}')\mathbf{1})\mathbf{1}\mathbf{1}^T\mathbf{y}\right\| \\
&= |y_j| + \frac{1}{n^2}\left\|\mathbf{1}\mathbf{1}^T\mathbf{y}\right\|_2 \\
&= |y_j| + \frac{1}{n^2}\|\mathbf{1}\|_2\,|\mathbf{1}^T\mathbf{y}| \\
&= |y_j| + \frac{|\sum_{i=1}^n y_i|}{n^{\frac{3}{2}}} \\
&\le \max_j |y_j| + \frac{\sum_{i=1}^n |y_i|}{n^{\frac{3}{2}}} \\
&= \|\mathbf{y}\|_\infty + \frac{\|\mathbf{y}\|_1}{n^{\frac{3}{2}}} \\
&\le \|\mathbf{y}\|_\infty + \frac{\sqrt{n}\,\|\mathbf{y}\|_2}{n^{\frac{3}{2}}} \\
&= \|\mathbf{y}\|_\infty + \frac{1}{n},
\end{aligned}$$

where $\mathbf{e}_i$ is the i-th standard basis vector. In the last inequality we have used the relation $\|\mathbf{y}\|_1 \le \sqrt{n}\|\mathbf{y}\|_2$ which holds for all $\mathbf{y} \in \mathbb{R}^n$, and in the last equality the assumption $\|\mathbf{y}\|_2 = 1$. This completes the proof of the lemma. □

With the above $\mathbf{y}$-adaptive sensitivity bound, we get the noisy power method depicted in Algorithm 3.

Remark III.2. The privacy guarantee of Algorithm 3 follows from the facts that analyzing an N-wise composition of Gaussian mechanisms, each with noise ratio σ, is equivalent to analyzing a Gaussian mechanism with noise ratio $\sigma/\sqrt{N}$ and by applying standard tail bounds for the Gaussian distribution [4]. In particular, choosing $\sigma = \frac{1}{\epsilon}\sqrt{4N\log(1/\delta)}$ ensures that the entire sequence $\mathbf{x}_1, \ldots, \mathbf{x}_N$ is (ϵ, δ)-differentially private. Note that Step 9 involves only post-processing of data-dependent intermediate values and therefore incurs no additional privacy cost.

1) Auxiliary Results for Noisy Power Iteration Method: The utility analysis of Algorithm 3 follows directly from [26, Thm. 1.3]. The general result of [26, Thm 1.3] is stated for a block matrix iteration. By carefully following the proof of [26, Thm 1.3], we observe that it can be adapted to yield a "$1-\eta$" high-probability bound by replacing [26, Lemma A.2] with the following result, which is specifically tailored to the noisy power vector iteration and applied using the sensitivity bound of Lemma III.7.

Lemma III.8. *Let $\mathbf{u} \in \mathbb{R}^n$ be a unit vector, and let $\mathbf{g}_1, \ldots, \mathbf{g}_L \sim \mathcal{N}(0, \sigma^2 I_n)$ be independent Gaussian vectors. Then, for any $\eta \in (0,1)$, with probability at least $1-\eta$, we have simultaneously*

$$\max_{\ell\in[N]} |\mathbf{u}^\top\mathbf{g}_\ell| \le \sigma\sqrt{2\log\left(\frac{2N}{\eta}\right)} \quad \textit{and}$$
$$\max_{\ell\in[N]} \|\mathbf{g}_\ell\| \le \sigma\left(\sqrt{n} + \sqrt{2\log\left(\frac{2N}{\eta}\right)}\right).$$

Proof. Since $\mathbf{u}$ is a unit vector, each $\mathbf{u}^\top\mathbf{g}_\ell$ is distributed as a one-dimensional Gaussian $\mathcal{N}(0,\sigma^2)$. Standard Gaussian concentration inequalities imply that for $t \ge 0$,

$$\Pr\left(|\mathbf{u}^\top\mathbf{g}_\ell| \ge \sigma t\right) \le 2e^{-t^2/2} \quad \text{and}$$
$$\Pr\left(\|\mathbf{g}_\ell\| \ge \sigma(\sqrt{n}+t)\right) \le e^{-t^2/2}.$$

Applying a union bound over the $2N$ events and setting $t = \sqrt{2\log(2N/\eta)}$, the claim follows. □

Using Lemma III.8 in the proof of [26, Thm 1.3] instead of [26, Lemma A.2], we directly have the following high-probability version of [26, Thm 1.3].

Lemma III.9. *If we choose $\sigma = \epsilon^{-1}\sqrt{4N\log(1/\delta)}$, then the Noisy Power Iteration of Algorithm3 with N iterations satisfies (ϵ,δ)-edge DP. Moreover, after $N = O\left(\frac{\lambda_1}{\lambda_1-\lambda_2}\log n\right)$ iterations we have with probability at least $1-\eta$ that*

$$\left\|(\mathbf{I}-\mathbf{u}_2\mathbf{u}_2^T)\mathbf{y}_N\right\|_2 \leq \frac{\sigma(\max_{t\in[N]}\|\mathbf{y}_t\|_\infty+\frac{1}{n})\left(\sqrt{n}+\sqrt{2\log(2N/\eta)}\right)}{\lambda_1-\lambda_2}, \quad (10)$$

where $\lambda_1 \geq \lambda_2$ denote the two largest eigenvalues of the matrix $\mathbf{B}$ and $\mathbf{u}_2$ denotes the eigenvector corresponding to the eigenvalue λ_1.

For lower bounding the spectral gap $\lambda_1-\lambda_2$ for the shifted matrix $\mathbf{B}$, have the following lemma from [25].

Lemma III.10. *Let p and q be parametrized as*

$$p=\frac{\alpha\log n}{n}, \quad q=\frac{\beta\log n}{n} \quad (11)$$

for some constant $\alpha>\beta>0$. Let $\lambda_1\geq\lambda_2\geq\ldots\geq\lambda_n$ be the eigenvalues of the matrix $\mathbf{B}=\mathbf{A}-\rho\mathbf{1}\mathbf{1}^T$, where $\rho=\mathbf{1}^T\mathbf{A}\mathbf{1}/n^2$. Then, for sufficiently large n, for some constants c_1 and c_2, it holds with probability at least $1-2n^{-\frac{1}{2(\alpha+\beta+1)}}-c_2n^{-3}$ that

$$\lambda_1\geq\frac{\alpha-\beta}{3}\log n$$

and

$$|\lambda_i|\leq 2c_1\sqrt{\log n}, \quad i=2,\ldots,n.$$

We directly get the following corollary from Lemma III.10.

Corollary III.1. *Suppose $p=\frac{\alpha\log n}{n}$ and $q=\frac{\beta\log n}{n}$ for constants $\alpha>\beta>0$, and let $\lambda_1\geq\lambda_2\geq\cdots\geq\lambda_n$ be the eigenvalues of the matrix $\mathbf{B}=\mathbf{A}-\rho\mathbf{E}_n$, with $\rho=\frac{1}{n^2}\mathbf{1}_n^\top\mathbf{A}\mathbf{1}_n$-*

Then, for all sufficiently large n, with probability at least $1-2n^{-\frac{1}{2(\alpha+\beta+1)}}-c_2n^{-3}$, it holds that

$$\frac{1}{\lambda_1-\lambda_2}\leq\frac{1}{\frac{1}{3}(p-q)n-2c_1\sqrt{\log n}},$$

where c_1 and c_2 are the constants from Lemma III.10.

Proof. Lemma III.10 directly gives the following lower bound for the spectral gap:

$$\lambda_1-\lambda_2\geq\frac{\alpha-\beta}{3}\log n-2c_1\sqrt{\log n}.$$

Taking reciprocals yields:

$$\frac{1}{\lambda_1-\lambda_2}\leq\frac{1}{\frac{\alpha-\beta}{3}\log n-2c_1\sqrt{\log n}}.$$

Substituting p andn q into the expression above:

$$\frac{1}{\lambda_1-\lambda_2}\leq\frac{1}{\frac{1}{3}\cdot\frac{(p-q)n}{\log n}\cdot\log n-2c_1\sqrt{\log n}}=\frac{1}{\frac{1}{3}(p-q)n-2c_1\sqrt{\log n}}.$$

This completes the proof of Corollary III.1. □

2) Main Result for Noisy Power Iteration: We are now ready to prove the main convergence theorem for the noisy power method. Notice that we can always bound $\max_{t\in[N]}\|\mathbf{y}_t\|_\infty$ by 1 since y_t has unit norm for all $t\in[N]$.

Theorem III.4. *If we choose $\sigma=\epsilon^{-1}\sqrt{4N\log(1/\delta)}$ and $C=\|\mathbf{y}\|_\infty+\frac{1}{n}$, the Noisy Power Iteration with N iterations satisfies (ϵ,δ)-DP. Moreover, with $N=O\left(\frac{\lambda_1}{\lambda_1-\lambda_2}\log n\right)$ iterations, for some constants c_1 and c_2, it holds with probability at least $1-2n^{-\frac{1}{2(\alpha+\beta+1)}}-c_2n^{-3}$ that*

$$\min_{s\in\{\pm1\}}\|\mathbf{y}_N-s\mathbf{u}_2\|_2\leq\frac{\sqrt{2}\sigma(1+\frac{1}{n})\left(\sqrt{n}+\sqrt{2\log\left(\frac{2N}{\eta}\right)}\right)}{\frac{1}{3}(p-q)n-2c_1\sqrt{\log n}}.$$

Proof. By simple linear algebra, we first derive a lower bound for the left-hand side of the inequality (10). Since $\mathbf{u}_2$ and $\mathbf{y}_N$ are of unit norm, we have that

$$\begin{aligned}\left\|(\mathbf{I}-\mathbf{u}_2\mathbf{u}_2^T)\mathbf{y}_N\right\|_2^2&=\|\mathbf{y}_N\|_2^2-2(\mathbf{u}_2^T\mathbf{y}_N)^2+(\mathbf{u}_2^T\mathbf{y}_N)^2\\&=1-(\mathbf{u}_2^T\mathbf{y}_N)^2\end{aligned}$$

and furthermore, since $\|\mathbf{y}_N\|_2=\|\mathbf{u}_2\|_2=1$,

$$\begin{aligned}\min_{s\in\{\pm1\}}\|\mathbf{y}_N-s\mathbf{u}_2\|_2^2&=\min_{s\in\{\pm1\}}\left(2-2s(\mathbf{u}_2^T\mathbf{y}_N)\right)\\&=2-2|\mathbf{u}_2^T\mathbf{y}_N|\\&\leq 2\cdot\left(1-(\mathbf{u}_2^T\mathbf{y}_N)^2\right)\\&=2\left\|(\mathbf{I}-\mathbf{u}_2\mathbf{u}_2^T)\mathbf{y}_N\right\|_2^2\end{aligned}$$

which implies

$$\min_{s\in\{\pm1\}}\|\mathbf{y}_N-s\mathbf{u}_2\|_2\leq\sqrt{2}\left\|(\mathbf{I}-\mathbf{u}_2\mathbf{u}_2^T)\mathbf{y}_N\right\|_2. \quad (12)$$

The claim follows then from the inequality (12), Lemma III.9 and Cor. III.1. □

Corollary III.2. *Under the assumptions of Thm. III.4, we have that with probability at least $1-2n^{-\frac{1}{2(\alpha+\beta+1)}}-c_2n^{-3}-\eta$,*

$$\min_{s\in\{\pm1\}}\|\mathbf{y}_N-s\mathbf{u}_2\|_2=O\left(\frac{\sqrt{\log 1/\delta}}{\epsilon(p-q)}\sqrt{\frac{\log n}{n}}\right).$$

Proof. We also have that

$$\begin{aligned}\frac{\lambda_1}{\lambda_1-\lambda_2}&=1+\frac{\lambda_2}{\lambda_1-\lambda_2}\\&=1+O\left(\frac{\sqrt{\log n}}{(p-q)n}\right)=O(1)\end{aligned}$$

and therefore $N=O\left(\frac{\lambda_1}{\lambda_1-\lambda_2}\log n\right)=O(\log n)$.

Substituting σ and N into eqn. (10) and neglecting a $\log N=O(\log\log n)$ factor, we have that with probability at least $1-2n^{-\frac{1}{2(\alpha+\beta+1)}}-c_2n^{-3}-\eta$,

$$\|\mathbf{u}_2-\mathbf{y}_N\|_2=O\left(\frac{\sqrt{\log 1/\delta}}{\epsilon(p-q)}\sqrt{\frac{\log n}{n}}\right).$$

This completes the proof of the corollary. □

Remark III.3. We remark that, analogous to the graph perturbation-based mechanism, the error bound provided in Theorem III.4 for the noisy power iteration method can be

directly translated into a lower bound on the overlap rate using Lemma III.4. We formalize this in the following lemma.

Lemma III.11 (Overlap Rate for Noisy Power Iteration)**.** *Consider the private spectral method based on noisy power iteration that estimates the second eigenvector* $\mathbf{u}_2$ *of the unperturbed graph Laplacian. Let* $\mathbf{y}_N$ *denote the final estimate after* N *iterations, and suppose that*

$$\min_{s\in\{\pm 1\}} \|\mathbf{y}_N - s\mathbf{u}_2\|_2 \leq \Delta,$$

with probability at least $1-\tilde{\eta}$*. Here,* $\tilde{\eta}$ *is taken directly from Theorem III.4, and* Δ *corresponds to the upper bound on the Euclidean distance given in the same theorem. Then, the overlap rate between the estimated labels* $\hat{\boldsymbol{\sigma}} = \operatorname{sign}(\mathbf{y}_N)$ *and the true labels* $\boldsymbol{\sigma}^*$ *satisfies*

$$\text{overlap rate}(\hat{\boldsymbol{\sigma}}, \boldsymbol{\sigma}^*) \geq 1 - \frac{\Delta^2}{8},$$

with probability at least $1-\eta$*.*

3) Tighter Privacy Analysis for Noisy Power Iteration: To further optimize the privacy-utility trade-offs for the noisy power iteration, we consider a tighter privacy analysis via so-called dominating pairs of distributions [27].

The noisy power iteration algorithm with N steps can be see as an adaptive composition of the form

$$\mathcal{M}^{(N)}(G) = \big(\mathcal{M}_1(G), \mathcal{M}_2(\mathcal{M}_1(G), G), \ldots, \mathcal{M}_N(\mathcal{M}_1(G), \ldots, \mathcal{M}_{N-1}(G), G)\big).$$

We obtain accurate (ϵ, δ)-differential privacy guarantees for adaptive compositions by leveraging dominating pairs of distributions, as introduced in [27]. Due to the scaling of the noise in Algorithm 3, the privacy loss at each step is dominated by that of the Gaussian mechanism with sensitivity 1 and noise standard deviation σ. Accordingly, the dominating pair of distributions for each step is given by (P, Q), where $P = \mathcal{N}(1, \sigma^2)$ and $Q = \mathcal{N}(0, \sigma^2)$.

Applying standard composition results [27], [28] to these dominating pairs, it follows that the N-fold adaptive composition of Algorithm 3 is itself dominated by the pair (P, Q), where $P = \mathcal{N}(1, N\sigma^2)$ and $Q = \mathcal{N}(0, N\sigma^2)$. Thus, the resulting (ϵ, δ)-DP guarantee corresponds to that of the Gaussian mechanism with sensitivity 1 and noise parameter $\sqrt{N}\sigma$. The analytical expression from [29, Thm. 8] then yields the following bound.

Lemma III.12. *Algorithm 3 is* $\big(\epsilon, \delta(\epsilon)\big)$*-DP for*

$$\delta(\epsilon) = \Phi\left(-\frac{\epsilon\sigma}{\sqrt{T}} + \frac{\sqrt{N}}{2\sigma}\right) - e^{\epsilon}\Phi\left(-\frac{\epsilon\sigma}{\sqrt{N}} - \frac{\sqrt{N}}{2\sigma}\right),$$

where Φ *denotes the CDF of the standard univariate Gaussian distribution.*

To compute ϵ as a function σ and δ or σ as a function of ϵ and δ, the expression of Lemma III.12 can be inverted using, e.g., bisection method.

IV. Experimental Comparisons

Two of out of the three discussed algorithms perform well in practive: the graph perturtbation based mechanism and the noisy power method. Table I summarized the derived error bounds for these two methods. We notice that the error bounds are similar except for the additional constant term in the bound for the graph perturbation mechanism.

A. Synthetic SBM Graphs

These differences are also reflected in the experimental results shown in Fig. 2, 3, and 4 where we empirically measure the overlap rate for synthetic SBM graphs. Fixing the probabilities p and q, we see that the noisy power method becomes the better one as the n (number of nodes) increases.

In Fig. 2, 3, and 4, each point describes the mean of 1000 experiments. The required σ-values for the noisy power method were computed using the expression given in Lemma III.12, based on ϵ, δ and the number of iteration N which was fixed to 8 for all experiments.

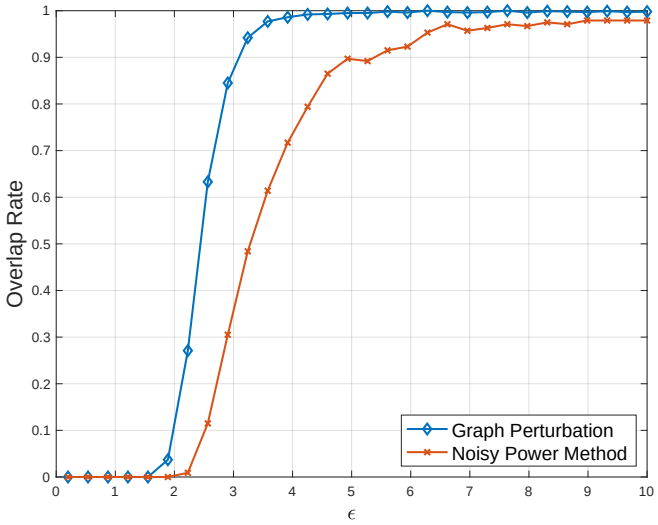

Fig. 2: Overlap vs ϵ, when $n = 200$, $p = 0.2$, $q = 0.02$, $\delta = n^{-2}$.

B. Political Blogs Dataset

We consider a symmetrized version of the Political Blogs dataset [30], originally a directed network of hyperlinks between U.S. political blogs collected in 2005. In this version, the direction of links is ignored, resulting in an undirected graph with 1,490 nodes and 16,718 edges. Each node represents a blog and is annotated with a label in $\{1, -1\}$ based on its content. Edge weights reflect the number of mutual hyperlinks between blog pairs, capturing the strength of their connection.

This dataset turns out to be more challenging for the noisy power method, as the eigenpair used for the deflation is not as good approximation of the leading eigenpair of the adjacency matrix $\mathbf{A}$ as in case of SBMs. We experimentally observe that only some of the randomly drawn initial vectors $\mathbf{y}_0$ for Algorithm 3 converge towards the second eigenvector of $\mathbf{A}$. To this end, we consider a variant, where we first privately

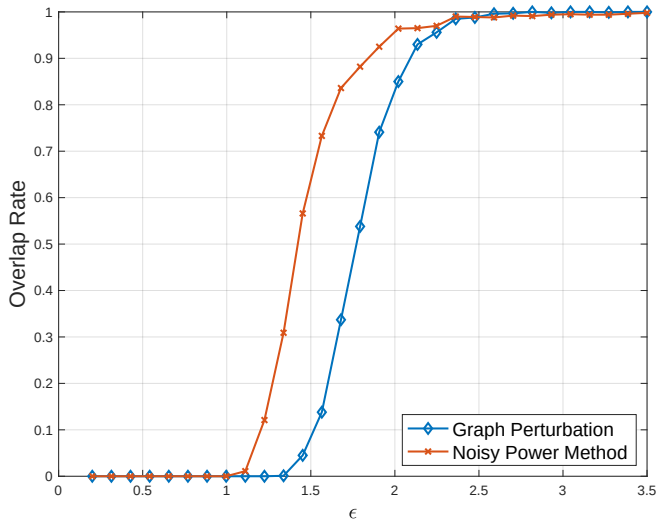

Fig. 3: Overlap vs ϵ, when $n = 400$, $p = 0.2$, $q = 0.02$, $\delta = n^{-2}$.

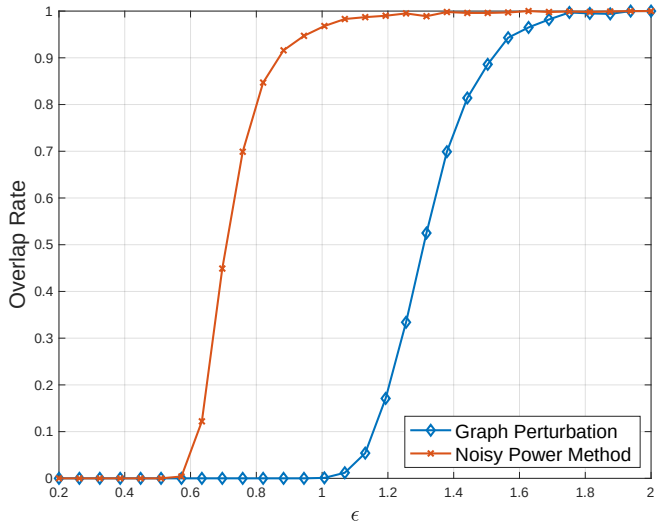

Fig. 4: Overlap vs ϵ, when $n = 800$, $p = 0.2$, $q = 0.02$, $\delta = n^{-2}$.

search for a suitable initial vector by adding symmetric normally distributed σ^2 variance noise to $\mathbf{A}$ and setting $\mathbf{y}_0$ of Algorithm 3 to be the second eigenvector of the resulting noisy matrix. As a result, the total privacy guarantee can be seen as a $(N+1)$-wise decomposition of Gaussian mechanisms, each with noise scale σ, and we again get the (ϵ, δ)-DP guarantees using Lemma III.12.

Fig. 5 shows the performance of the graph perturbation method and the noisy power method as a function of ϵ, when $\delta = 1/n^2$ for the noisy power method. In addition to the results for the fully private method where we privately initialize $\mathbf{y}_0$ for Algorithm 3, we show the convergence with a randomly chosen initial value that converges to the second eigenvector of $\mathbf{A}$.

In Fig. 5, each point describes the mean of 100 experiments. The number of iterations N was fixed to 3 for the noisy power method.

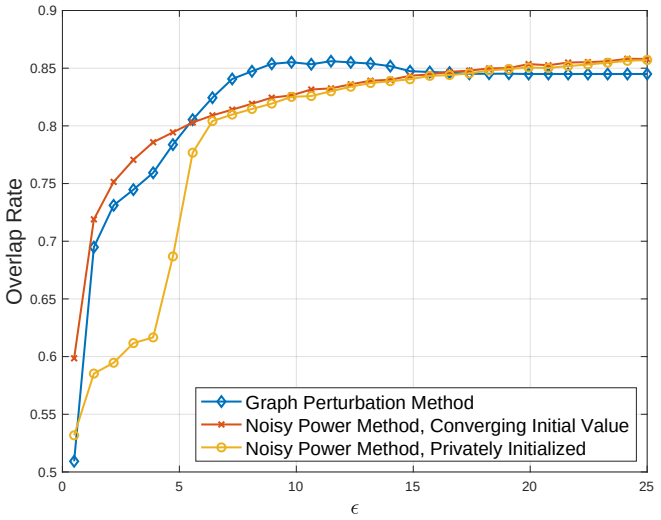

Fig. 5: Overlap vs ϵ for the Political Blogs dataset.

V. Conclusion

We developed privacy-preserving spectral clustering methods for community detection over the binary symmetric SBMs under edge DP. Our approaches include (1) a Graph Perturbation-Based Mechanism, which perturbs the adjacency matrix using either randomized response, followed by spectral clustering, (2) a Subsampling Stability-Based Mechanism, which leverages subsampling and aggregation for accurate recovery, and (3) an edge DP power method that adds carefully calibrated Gaussian noise to each matrix–vector multiplication, guaranteeing edge DP for every intermediate eigenvector estimate while still converging to the true leading eigenvectors. We also analyzed the tradeoff between privacy and accuracy, providing theoretical guarantees. Future work will generalize these ideas to (i) SBMs with more than two communities and (ii) graphs exhibiting degree heterogeneity.

References

[1] S. Fortunato, "Community detection in graphs," *Physics reports*, vol. 486, no. 3-5, pp. 75–174, 2010.

[2] E. Abbe, "Community detection and stochastic block models: recent developments," *Journal of Machine Learning Research*, vol. 18, no. 177, pp. 1–86, 2018.

[3] ——, "Community detection and stochastic block models: recent developments," *Journal of Machine Learning Research*, vol. 18, no. 1, pp. 6446–6531, 2017.

[4] C. Dwork, A. Roth *et al.*, "The algorithmic foundations of differential privacy." *Foundations and Trends in Theoretical Computer Science*, vol. 9, no. 3-4, pp. 211–407, 2014.

[5] V. Karwa, S. Raskhodnikova, A. Smith, and G. Yaroslavtsev, "Private analysis of graph structure," *Proceedings of the VLDB Endowment*, vol. 4, no. 11, pp. 1146–1157, 2011.

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

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

[8] J. Imola, T. Murakami, and K. Chaudhuri, "Locally differentially private analysis of graph statistics," in *Proceedings of the 30th USENIX Symposium on Security*, 2021.

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

[10] M. Seif, D. Nguyen, A. Vullikanti, and R. Tandon, "Differentially private community detection for stochastic block models," in *Proceedings of the 2022 International Conference on Machine Learning (ICML)*. PMLR, 2022, pp. 15 858–15 894.

[11] M. Seif, Y. Chen, A. J. Goldsmith, and H. V. Poor, "Differentially private sketch-and-solve for community detection via semidefinite programming," *IEEE Journal on Selected Areas in Information Theory*, 2024.

[12] M. Seif, L. Xie, A. J. Goldsmith, and H. V. Poor, "Private online community detection for censored block models," *arXiv preprint arXiv:2405.05724*, 2024.

[13] S. Dhara, J. Gaudio, E. Mossel, and C. Sandon, "The power of two matrices in spectral algorithms," arXiv preprint arXiv:2210.05893, 2022.

[14] ——, "Spectral recovery of binary censored block models," in *Proceedings of the 2022 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*. SIAM, 2022, pp. 3389–3416.

[15] B. Hajek, Y. Wu, and J. Xu, "Achieving exact cluster recovery threshold via semidefinite programming," *IEEE Transactions on Information Theory*, vol. 62, no. 5, pp. 2788–2797, 2016.

[16] J. Hehir, A. Slavković, and X. Niu, "Consistent spectral clustering of network block models under local differential privacy," *The Journal of privacy and confidentiality*, vol. 12, no. 2, 2022.

[17] U. von Luxburg, "A tutorial on spectral clustering," *Statistics and Computing*, vol. 17, no. 4, pp. 395–416, 2007.

[18] S. Vadhan, "The complexity of differential privacy," *Tutorials on the Foundations of Cryptography: Dedicated to Oded Goldreich*, pp. 347–450, 2017.

[19] H. Chen, V. Cohen-Addad, T. d'Orsi, A. Epasto, J. Imola, D. Steurer, and S. Tiegel, "Private estimation algorithms for stochastic block models and mixture models," *Proceedings of the 2023 Advances in Neural Information Processing Systems (NeurIPS)*, vol. 36, pp. 68 134–68 183, 2023.

[20] S. L. Warner, "Randomized response: A survey technique for eliminating evasive answer bias," *Journal of the American Statistical Association*, vol. 60, no. 309, pp. 63–69, 1965.

[21] C. M. Le, E. Levina, and R. Vershynin, "Concentration and regularization of random graphs," *Random Structures & Algorithms*, vol. 51, no. 3, pp. 538–561, 2017.

[22] M. Mitzenmacher and E. Upfal, *Probability and Computing: Randomization and Probabilistic Techniques in Algorithms and Data Analysis*. Cambridge University Press, 2017.

[23] C. Davis and W. M. Kahan, "The rotation of eigenvectors by a perturbation. iii," *SIAM Journal on Numerical Analysis*, vol. 7, no. 1, pp. 1–46, 1970.

[24] R. A. Horn and C. R. Johnson, *Matrix analysis*. Cambridge university press, 2012.

[25] P. Wang, Z. Zhou, and A. M.-C. So, "A nearly-linear time algorithm for exact community recovery in stochastic block model," in *Proceedings of the 2020 International Conference on Machine Learning (ICML)*. PMLR, 2020, pp. 10 126–10 135.

[26] M. Hardt and E. Price, "The noisy power method: A meta algorithm with applications," *Proceedings of the 2014 Advances in neural information processing systems*, vol. 27, 2014.

[27] Y. Zhu, J. Dong, and Y.-X. Wang, "Optimal accounting of differential privacy via characteristic function," in *Proceedings of the 2022 International Conference on Artificial Intelligence and Statistics*. PMLR, 2022, pp. 4782–4817.

[28] S. Gopi, Y. T. Lee, and L. Wutschitz, "Numerical composition of differential privacy," *Proceedings of the 2021 Advances in Neural Information Processing Systems*, vol. 34, pp. 11 631–11 642, 2021.

[29] B. Balle and Y.-X. Wang, "Improving the gaussian mechanism for differential privacy: Analytical calibration and optimal denoising," in *Proceedings of the 2018 International Conference on Machine Learning (ICML)*. PMLR, 2018, pp. 394–403.

[30] L. A. Adamic and N. Glance, "The Political Blogosphere and the 2004 US Election: Divided They Blog," in *Proceedings of the 3rd international workshop on Link discovery*, 2005, pp. 36–43.