Journal of Privacy and Confidentiality
Vol. 12 (2) 2022

TPDP 2021

Submitted Jan 2022
Published Nov 2022

CONSISTENT SPECTRAL CLUSTERING OF NETWORK BLOCK MODELS UNDER LOCAL DIFFERENTIAL PRIVACY

JONATHAN HEHIR, ALEKSANDRA SLAVKOVIĆ, AND XIAOYUE NIU

Department of Statistics, Penn State University, University Park, PA, USA
e-mail address: jh@psu.edu

Department of Statistics, Penn State University, University Park, PA, USA
e-mail address: sesa@psu.edu

Department of Statistics, Penn State University, University Park, PA, USA
e-mail address: xiaoyue@psu.edu

Abstract. The stochastic block model (SBM) and degree-corrected block model (DCBM) are network models often selected as the fundamental setting in which to analyze the theoretical properties of community detection methods. We consider the problem of spectral clustering of SBM and DCBM networks under a local form of edge differential privacy. Using a randomized response privacy mechanism called the edge-flip mechanism, we develop theoretical guarantees for differentially private community detection, demonstrating conditions under which this strong privacy guarantee can be upheld while achieving spectral clustering convergence rates that match the known rates without privacy. We prove the strongest theoretical results are achievable for dense networks (those with node degree linear in the number of nodes), while weak consistency is achievable under mild sparsity (node degree greater than $\sqrt{n}$). We empirically demonstrate our results on a number of network examples.

1. Introduction

In the field of network data analysis, a common problem is community detection, the algorithmic discovery of clusters or communities of dense connection. There exist numerous parametric network models that exhibit community structure. Two fundamental models used in the analysis of community detection algorithms are the stochastic block model (SBM), first introduced in Holland et al. (1983), and the degree-corrected block model (DCBM) of Karrer and Newman (2011), a popular SBM extension that allows for networks with more realistic heterogeneity in node degree. In addition to theoretical interest, this class of models has proven useful in modeling patterns of interaction in various settings ranging from social networks and political science to the natural sciences (e.g., Wang and Wong, 1987; Daudin et al., 2008; Decelle et al., 2011; Karrer and Newman, 2011; Kim et al., 2018; Lee and Wilkinson, 2019, and references therein).

Key words and phrases: community detection, differential privacy, spectral clustering, stochastic block model.

JPrivacy Confidentiality
www.journalprivacyconfidentiality.org
DOI:10.29012/jpc.811

We consider the problem of estimating latent community structure in networks using SBM and DCBM via the method of spectral clustering. In the absence of privacy, this problem has gained popularity, as spectral clustering has been shown to be a computationally tractable method for group estimation with satisfying theoretical guarantees (McSherry, 2001; Ng et al., 2002; Rohe et al., 2011; Qin and Rohe, 2013; Lei and Rinaldo, 2015; Joseph and Yu, 2016; Binkiewicz et al., 2017; Abbe, 2018; Abbe et al., 2021). In this setting, the clusters of a given network are informed by the relationships contained within that network. These relationships are represented by the edges of the network, which may reflect friendships, partnerships, collaborations, communications, financial transactions, similarities, or other interactions between some set of entities, say, people or businesses.

In many cases, these relationships may constitute sensitive or confidential information, and so we may desire a clustering that reveals high-level structures in the network while preserving the privacy of low-level relationships. To provide such a privacy guarantee, we turn to a form of local differential privacy (DP) for networks (see, e.g., Imola et al. (2021) for networks, and Duchi et al. (2013) for more general ideas on local DP). We assume that the nodes of the network are known but that their relationships are sensitive. We protect these relationships through a randomized-response mechanism applied to the edges of the network. Previous works, such as Mülle et al. (2015) and Karwa et al. (2017), have used similar techniques to obtain a synthetic network that satisfies ε-DP (of Dwork et al., 2006) with respect to the edges of the network. Recent works (Qin et al., 2017; Imola et al., 2021) have emphasized that these synthetic networks can be constructed in a distributed fashion, requiring that no single party has knowledge of the full set of true network relationships. We construct our synthetic networks in such a manner, then apply a modified spectral clustering algorithm to obtain consistent estimates of group membership for SBM and DCBM networks.

The problem of performing DP community detection has previously been given empirical consideration (Mülle et al., 2015; Nguyen et al., 2016; Qin et al., 2017), but to our knowledge, our analysis of the cost of privacy in this setting is the first theoretical treatment of the subject.[1] Our work generalizes the non-private results of Lei and Rinaldo (2015) to the DP setting, and we demonstrate certain conditions under which our DP estimator matches the known non-private convergence rates. We attribute these results to certain desirable properties of the edge-flip mechanism that we explore in some detail. While this method performs well for dense networks, the magnitude of error introduced by the local DP mechanism results in a slowing of the convergence rate for sparse networks and a requirement that node degree grow faster than $\sqrt{n}$. We derive these results in the context of more general finite-sample and asymptotic bounds on misclassification rates with privacy.

The paper is structured as follows: We first introduce the network models and notation in Section 2. In Section 3, we review differential privacy and define the edge-flip mechanism, and we demonstrate that the edge-flip mechanism can be viewed as a mixture distribution (Lemma 3.7), leading to two corollaries: a closure property for SBM and a more general post-processing step that allows us to preserve the expected spectral embedddings of edge-flipped networks. In Section 4, we propose a modified spectral clustering algorithm based on this method—with concentration bounds for the estimation of spectral embeddings of SBM and DCBM networks proven in Supplementary Material, Lemma A.1. Our main theoretical results on finite-sample bounds on misclassification and asymptotic convergence rates for

[1] Related theoretical developments include the study of differentially private graph cuts (e.g., Gupta et al., 2012; Blocki et al., 2012), as spectral clustering may be seen as a relaxation of the problem of finding minimal graph cuts (Von Luxburg, 2007).

private spectral clustering are presented in Section 5. Proofs of these results are deferred to Supplementary Material, Section A. In Section 6, we evaluate the performance of private spectral clustering on both simulated and observed networks. Finally, we conclude with a discussion of limitations and open questions in Section 7.

2. Network Models and Notation

SBM and DCBM are models for binary, undirected networks without covariates. Edges in both networks occur as independent Bernoulli random variables. Each of the n nodes is assigned to one of k communities, or *blocks*, and these memberships are denoted in the parameter $\boldsymbol{\theta} \in [k]^n$ (where $[k] = \{1, \ldots, k\}$). In SBM, the probability of a given edge occurring between nodes i and j depends only on the blocks to which i and j belong, with those block-to-block edge probabilities recorded in the symmetric matrix $B \in (0,1]^{k \times k}$. In DCBM, edge probabilities further depend on a parameter $\boldsymbol{\psi} \in (0,1]^n$ that determines the relative expected node degree for each node. If we let Y be the symmetric $n \times n$ adjacency matrix from a DCBM, the elements of Y are then distributed as

$$Y_{ij} \overset{\text{ind}}{\sim} \begin{cases} \text{Bernoulli}(\boldsymbol{\psi}_i \boldsymbol{\psi}_j B_{\boldsymbol{\theta}_i \boldsymbol{\theta}_j}), & i < j \\ 0, & i = j \\ Y_{ji}, & i > j. \end{cases}$$

For such a network, we will write $Y \sim \text{DCBM}(\boldsymbol{\theta}, \boldsymbol{\psi}, B)$. The stochastic block model can be regarded as a special case of the DCBM, where $\boldsymbol{\psi} = \mathbf{1}_n$, a vector of n ones, i.e.,

$$\text{SBM}(\boldsymbol{\theta}, B) \overset{D}{=} \text{DCBM}(\boldsymbol{\theta}, \mathbf{1}_n, B), \tag{2.1}$$

where $\overset{D}{=}$ indicates equality in joint distribution of network edges.

We let $C_j = \{i : \boldsymbol{\theta}_i = j\}$ denote the set of nodes in the j-th community. We denote the size of the j-th community in an SBM or DCBM (i.e., the number of nodes in a block) as $n_j = |C_j|$, the smallest of which we denote by $n_{\min} = \min_{j \in [k]} n_j$, the largest by $n_{\max} = \max_{j \in [k]} n_j$, and the second-largest as $n'_{\max}$. As in Lei and Rinaldo (2015), we denote the effective size of the j-th block in a DCBM as $\tilde{n}_j = \sum_{i \in C_j} \boldsymbol{\psi}_i^2$ and the largest effective size of a block as $\tilde{n}_{\max} = \max_{j \in [k]} \tilde{n}_j$. As a measure of heterogeneity within the j-th block, we use $\nu_j = n_j^{-2} (\sum_{i \in C_j} \boldsymbol{\psi}_i^{-2})(\sum_{i \in C_j} \boldsymbol{\psi}_i^2) \in [1, \infty)$. When $\nu_j = 1$, all nodes in the j-th community have the same expected degree, and larger values of ν_j denote greater heterogeneity.

We use λ_B to denote the smallest absolute nonzero eigenvalue of the matrix B. The largest entry in B is denoted $\max B = \max_{ij} B_{ij}$. We represent networks via adjacency matrices, e.g., $Y \in \{0,1\}^{n \times n}$. The i-th row of the matrix Y is denoted Y_{i*}. We use $\|\mathbf{x}\|_p$ to denote the ℓ_p norm of a vector $\mathbf{x}$, $\|A\|_F$ to denote the Frobenius norm of a matrix, and $\|A\|$ to denote the operator norm of the matrix A, i.e., $\|A\| = \sup_{\|\mathbf{x}\|_2 = 1} \|A\mathbf{x}\|_2$.

In asymptotic notation, we write $a_n = o(b_n)$ when $|a_n/b_n| \to 0$ as $n \to \infty$; $a_n = \omega(b_n)$ when $|a_n/b_n| \to \infty$ as $n \to \infty$; $a_n = O(b_n)$ when $|a_n/b_n| \leq C$ for some $C > 0$ and all n; $a_n = \Omega(b_n)$ when $|a_n/b_n| \geq C$ for some $C > 0$ and all n; and $a_n = \Theta(b_n)$ when $a_n = O(b_n)$ and $a_n = \Omega(b_n)$. Finally, we write $X_n = O_P(b_n)$ if for any $\alpha > 0$ there exists a constant C such that $P(|X_n/b_n| > C) < \alpha$ for all large n.

3. Privacy Mechanism as a Mixture Distribution

3.1. **Defining Privacy.** The framework of differential privacy offers more than just a formal privacy definition: it offers *many* privacy definitions that build on each other. To fully appreciate the privacy implications of a given definition—and thereby to justify our specific choice of definition—we briefly review the core concepts of DP and discuss the nuanced distinctions between privacy definitions that could be applied to the problem at hand. Consider first the typical definition of DP:

Definition 3.1 (Differential Privacy (Dwork et al., 2006))**.** Suppose that $\varepsilon > 0, \delta \in [0,1)$. Let $\mathcal{Y}, \mathcal{Z}$ be sets, and let $d : \mathcal{Y} \times \mathcal{Y} \to \mathbb{Z}_{\geq 0}$ be an integer-valued metric. Let $\mathcal{M} : \mathcal{Y} \to \mathcal{Z}$ be a randomized algorithm. We say $\mathcal{M}$ satisfies (ε, δ)–**differential privacy** with respect to d if for any $y, y' \in \mathcal{Y}$ satisfying $d(y, y') = 1$ and any $E \subseteq \mathcal{Z}$,

$$P(\mathcal{M}(y) \in E) \leq e^{\varepsilon} \cdot P(\mathcal{M}(y') \in E) + \delta. \tag{3.1}$$

If $\delta = 0$, we say $\mathcal{M}$ satisfies ε–**differential privacy**.

Such an algorithm $\mathcal{M}$ is called a *privacy mechanism*, and the elements y and y' are often referred to as *databases*. In non-network settings, the metric d is usually chosen to be the Hamming distance, such that (3.1) can be interpreted thus: changing any one record in an arbitrary database $y \in \mathcal{Y}$ does not significantly affect the distribution of $\mathcal{M}(y)$.[2] Consequently, one can gain virtually no information about a single record based on the private output $\mathcal{M}(y)$. The parameters ε and δ quantify the strength of the privacy guarantee, with smaller values conferring a stronger guarantee. We focus on the case where $\delta = 0$, a setting often referred to as *pure differential privacy.* The parameter ε is commonly called the *privacy-loss budget.*

In the network setting, there are two primary choices for the metric d that are frequently employed, and these correspond to the notions of edge DP and node DP (Hay et al., 2009; Kasiviswanathan et al., 2013):

Definition 3.2 (Edge DP)**.** Let Y, Y' be two networks with n nodes. Then $d_{\text{edge}}(Y, Y')$ is given by the total number of edges that differ between Y and Y'. If $\mathcal{M}$ satisfies (ε, δ)–DP with respect to d_{edge}, then we say $\mathcal{M}$ satisfies (ε, δ)–**edge differential privacy**.

Definition 3.3 (Node DP)**.** Let Y, Y' be two networks with n nodes. Then $d_{\text{node}}(Y, Y')$ is given by the minimum number of nodes in Y whose incident edges can be modified in order to obtain Y'. If $\mathcal{M}$ satisfies (ε, δ)–DP with respect to d_{node}, then we say $\mathcal{M}$ satisfies (ε, δ)–**node differential privacy**.

Since $d_{\text{edge}}(Y, Y') \geq d_{\text{node}}(Y, Y')$, the definition of node DP is more strict than the definition of edge DP. Indeed, node DP implies edge DP, but the reverse is not true. In many cases, the strictness of node DP precludes meaningful analysis (Kasiviswanathan et al., 2013). This is the case, for example, in the problem we wish to solve. Suppose $\hat{\boldsymbol{\theta}}(Y)$ is an ε–DP estimator of the group membership $\boldsymbol{\theta}$ for the network Y with respect to a distance measure d. This implies that when we look at any given node's estimated label, $[\hat{\boldsymbol{\theta}}(\cdot)]_i$,

[2] Note that the databases y, y' are not presented as random quantities in this definition. One can think of $\mathcal{M}(y)$ as a distribution conditioned on the observed data, as in Wasserman and Zhou (2010).

across two networks Y, Y' satisfying $d(Y, Y') = 1$, it must be that

$$\frac{P([\hat{\boldsymbol{\theta}}(Y)]_i = \ell)}{P([\hat{\boldsymbol{\theta}}(Y')]_i = \ell)} \in [e^{-\varepsilon}, e^{\varepsilon}] \quad \forall \ell.$$

For edge DP ($d = d_{\text{edge}}$), this means that changing any single edge in the network cannot significantly affect the distribution of the estimated label for a given node. For node DP ($d = d_{\text{node}}$), however, we can take any given node and change any subset (or even all) of its incident edges without significantly affecting the distribution of its estimated label. Since our goal is to infer labels from precisely these edges, this notion of privacy is too strict for our purposes. For this reason, we will focus on edge-based definitions of DP, which aim to protect the privacy of individual relationships within the network.[3]

So far, the DP definitions we have considered all fall under the umbrella of *central DP*. In central DP, we assume that one party, often called the *trusted curator*, has complete knowledge of the true database y. This arrangement is not always desirable, which motivates the concept of *local DP* (Kasiviswanathan et al., 2011; Duchi et al., 2013). In local DP, records in the database are held by a number of distributed parties (e.g., users of a social network), and only these parties require true knowledge of their records. To facilitate some central analysis of the database, a randomized algorithm is independently applied to each record to produce a *local differentially private view* of that record. These local DP views may then be shared with a central processor for further analysis. Thus, in contrast with central DP, where a single mechanism is applied at the database level by a single database owner, local DP applies privacy mechanisms at the record level, eliminating the need for a trusted curator, while still allowing for centralized analysis of the data.

Applying local DP to the edge setting, we have further choices yet. Possibly the most widely known definition of local DP in the edge setting is *edge local DP* of Qin et al. (2017), but we will instead focus on *relationship DP* as defined in Imola et al. (2021). This definition is more tailored to the setting of undirected networks, where it provides a stronger privacy guarantee.[4] In relationship DP (as in edge local DP), each node reports a private view (or summary) of its *neighbor list*, the set of nodes with which it shares an edge. The formal definition is given below.

Definition 3.4 (Relationship DP (Imola et al., 2021))**.** Let $\varepsilon > 0$, and let $\mathcal{M}_1, \dots, \mathcal{M}_n$ be randomized algorithms with domain $\{0,1\}^n$. The algorithm $\mathcal{M}(Y) = (\mathcal{M}_1(Y_{1*}), \dots, \mathcal{M}_n(Y_{n*}))$ is said to satisfy ε**–relationship DP** if for each $E \subseteq \text{Range}(\mathcal{M})$ and networks Y, Y' satisfying $d_{\text{edge}}(Y, Y') = 1$,

$$P(\mathcal{M}(Y) \in E) \leq e^{\varepsilon} \cdot P(\mathcal{M}(Y') \in E).$$

3.2. **The Edge-Flip Mechanism.**

The specific privacy mechanism we employ is a simple randomized-response mechanism that produces an ε-relationship-DP synthetic copy $\mathcal{M}_{\varepsilon}(Y)$ of the original network Y by randomly flipping edges in Y. This can be performed locally by assigning each node to flip and self-report a subset of its edges. In this way, no single party ever needs full knowledge of the true adjacency matrix Y. Since differential privacy is preserved under post-processing (Dwork et al., 2006), any analysis performed on the

[3] Edge-based definitions of DP implicitly assume that the only sensitive information in the network is encoded in these relationships. For example, in a community detection setting, we assume that the identities of nodes are not sensitive but that their relationships are.

[4] More precisely, in an undirected network, ε–edge local DP implies 2ε–relationship DP (Imola et al., 2021).

synthetic network $\mathcal{M}_\varepsilon(Y)$ preserves ε–relationship DP. Edge-flipping mechanisms have been utilized in several earlier papers and interpreted under various edge DP definitions, including (central) edge DP (Mülle et al., 2015; Karwa and Slavković, 2016; Karwa et al., 2017), edge local DP (Qin et al., 2017), and relationship DP (Imola et al., 2021). We include here an explicit formulation of the edge-flipping procedure from Imola et al. (2021), which we term the *symmetric edge-flip mechanism.*

Definition 3.5 (Symmetric Edge-Flip Mechanism)**.** Suppose that $\varepsilon > 0$. For $i \in [n]$, let $\mathcal{M}_i : \{0,1\}^n \to \{0,1\}^n$ be such that

$$[\mathcal{M}_i(\mathbf{x})]_j \overset{\text{ind}}{=} \begin{cases} 0 & i \geq j \\ 1 - \mathbf{x}_j & i < j \quad \text{w.p.} \quad \frac{1}{1+e^\varepsilon} \\ \mathbf{x}_j & i < j \quad \text{w.p.} \quad \frac{e^\varepsilon}{1+e^\varepsilon}, \end{cases}$$

and let

$$T(Y) = \begin{bmatrix} \mathcal{M}_1(Y_{1*}) \\ \vdots \\ \mathcal{M}_n(Y_{n*}) \end{bmatrix}.$$

Then the $n \times n$ **symmetric edge-flip mechanism** is the mechanism $\mathcal{M}_\varepsilon(Y) = T(Y) + [T(Y)]^T$.

Theorem 3.6. *The symmetric edge-flip mechanism $\mathcal{M}_\varepsilon$ satisfies ε–relationship DP.*

Proof. The proof of this follows from Theorem 3 of Imola et al. (2021) and the corresponding discussion. For completeness, we give a formal proof in Supplementary Material, Section A.5. □

The simplicity of the edge-flip mechanism affords it several key advantages: In practice, it is easy and flexible to implement. In theory, the distribution of the resulting synthetic network is transparent and tractable. In fact, in Lemma 3.7 we show that the network generated by the edge-flip mechanism is a mixture of the original non-private network with an Erdős–Rényi network. This in turn leads to closure and statistical inference properties that are important and useful, both theoretically and practically. These properties are demonstrated in the corollaries that follow.

Lemma 3.7. *Let $Y \in \{0,1\}^{n\times n}$ be a random, undirected, binary network. Let $Z \sim G(n, 1/2)$ be an Erdős–Rényi random graph with n nodes and edge probability $1/2$, and let $U_{ij} \overset{ind}{\sim}$ $\text{Bernoulli}(2/(2e^\varepsilon+1))$ for $1 \leq i < j \leq n$. Define $\mathcal{M}'(Y)$ to be the symmetric mixture network such that for $i < j$,*

$$\mathcal{M}'(Y)_{ij} = (Z_{ij})^{U_{ij}}(Y_{ij})^{(1-U_{ij})}.$$

Then $\mathcal{M}'(Y)$ is equal in distribution to $\mathcal{M}_\varepsilon(Y)$.

Proof. It is sufficient to show that the conditional distributions of $\mathcal{M}'(Y) \,|\, Y$ and $\mathcal{M}_\varepsilon(Y) \,|\, Y$ are equivalent. Conditioned on Y, the entries of $\mathcal{M}'(Y)$ and $\mathcal{M}_\varepsilon(Y)$ are independent, binary random variables, and

$$\begin{aligned}
P(\mathcal{M}'(Y)_{ij} = 1 \,|\, Y) &= E[\mathcal{M}'(Y)_{ij} \,|\, Y] \\
&= E[\, E[\mathcal{M}'(Y)_{ij} | Y, Z, U] \,|\, Y] \\
&= E\left[U_{ij} Z_{ij} + (1 - U_{ij}) Y_{ij} \,|\, Y\right] \\
&= \frac{1}{e^{\varepsilon} + 1} + \frac{e^{\varepsilon} - 1}{e^{\varepsilon} + 1} Y_{ij} \\
&= \frac{1}{e^{\varepsilon} + 1}(1 - Y_{ij}) + \frac{e^{\varepsilon}}{e^{\varepsilon} + 1} Y_{ij} \\
&= E[\mathcal{M}_{\varepsilon}(Y)_{ij} \,|\, Y] \\
&= P(\mathcal{M}_{\varepsilon}(Y)_{ij} = 1 \,|\, Y).
\end{aligned}$$

□

Lemma 3.7 provides useful intuition about the edge-flip mechanism. As $\varepsilon \to \infty$, our synthetic network $\mathcal{M}_\varepsilon(Y)$ approaches the true network Y, and as $\varepsilon \to 0$, we approach an Erdős–Rényi network. For block models, this is a welcome property, as it means the community structure of the network is exactly preserved. For the SBM in particular, we have closure of the model family under $\mathcal{M}_\varepsilon$. The following corollary follows from a simple extension of the proof of Lemma 3.7.

Corollary 3.8. *If $Y \sim \mathrm{SBM}(\boldsymbol{\theta}, B)$, then $\mathcal{M}_\varepsilon(Y) \sim \mathrm{SBM}(\boldsymbol{\theta}, \tau_\varepsilon(B))$, where*

$$\tau_\varepsilon(B) = \frac{1}{e^{\varepsilon} + 1} \mathbf{1}_k \mathbf{1}_k^T + \frac{e^{\varepsilon} - 1}{e^{\varepsilon} + 1} B. \tag{3.2}$$

The closure property of Corollary 3.8 may be used to immediately obtain consistency results for SBM under differential privacy via application of existing non-private results. This approach, however, comes with two major limitations: First, applying the results of Lei and Rinaldo (2015) to the private network $\mathcal{M}_\varepsilon(Y)$ yields error bounds in terms of the spectrum of $\tau_\varepsilon(B)$, while for the non-private Y, we obtain bounds in terms of the spectrum of B. This makes it difficult to make meaningful, general comparisons between the resulting error bounds. Secondly, while closure holds for SBM, it does not hold for the broader DCBM family. As a result, this approach offers no theoretical insight on how to perform consistent community detection outside of the specific case of the SBM.

To facilitate comparisons between private and non-private error bounds and in consideration of the broader DCBM family, we will explore a more generally applicable framework here, relying instead on a weaker property of the edge-flip mechanism that holds for any random binary network. Via a small "downshift" transformation to an edge-flipped network, we can recover a network whose expectation matches that of the original network, up to a scaling factor.

Corollary 3.9. *Let $Y \in \{0,1\}^{n \times n}$ be a random, undirected, binary network, and let $(d_\varepsilon \circ \mathcal{M}_\varepsilon)(Y) = \mathcal{M}_\varepsilon(Y) - (e^\varepsilon + 1)^{-1}(\mathbf{1}_n \mathbf{1}_n^T - I_n)$ be the downshifted edge-flipped network. Then*

$$E(d_\varepsilon \circ \mathcal{M}_\varepsilon)(Y) = \frac{e^{\varepsilon} - 1}{e^{\varepsilon} + 1} EY.$$

Proof. This follows from a slight rewriting of the derivation in the proof of Lemma 3.7, namely

$$\begin{aligned} E[\mathcal{M}_\varepsilon(Y)] &= E[\, E[\mathcal{M}_\varepsilon(Y) \,|\, Y]\,] \\ &= \frac{1}{e^\varepsilon + 1}(\mathbf{1}_n \mathbf{1}_n^T - I_n) + \frac{e^\varepsilon - 1}{e^\varepsilon + 1} EY. \end{aligned}$$

□

Corollary 3.9 transforms a $\{0,1\}$-valued network into a weighted network in exchange for the following property: for any random, undirected, binary network Y, the matrices $E[Y]$ and $E[(d_\varepsilon \circ \mathcal{M}_\varepsilon)(Y)]$ share the same eigenvectors. This fact leads to a more general post-processing step under DP that allows us to preserve the expected spectral embeddings of edge-flipped networks.

4. Modified Clustering Methods with Concentration Bounds

Our goal is to use spectral clustering to estimate the unobserved block membership parameter $\boldsymbol{\theta}$ while satisfying ε-relationship DP. A variety of related clustering algorithms exist under the umbrella term of *spectral clustering* (Ng et al., 2002; Von Luxburg, 2007; Rohe et al., 2011; Lei and Rinaldo, 2015; Binkiewicz et al., 2017). In our setting, we focus on the method of adjacency spectral clustering with k known, following the framework of Lei and Rinaldo (2015). In adjacency spectral clustering, we place the leading k eigenvectors (by absolute value of corresponding eigenvalues) of the observed adjacency matrix Y into an $n \times k$ matrix $\hat{U}$. We consider each row of $\hat{U}$ to be the spectral embedding of the respective node in the network, and we perform a clustering process over these embeddings: for SBM, we perform simple k-means, and for DCBM, we normalize the embeddings to have unit norm before performing k-medians clustering. The resulting estimated cluster memberships serve as our estimator, $\hat{\boldsymbol{\theta}}$.

This approach is theoretically justified by demonstrating that the observed embeddings $\hat{U}$ concentrate around a set of "expected" embeddings U that satisfy appropriate geometric properties. In particular, let $Y \sim \mathrm{DCBM}(\boldsymbol{\theta}, \boldsymbol{\psi}, B)$, let P be the $n \times n$ matrix with entries $P_{ij} = \boldsymbol{\psi}_i \boldsymbol{\psi}_j B_{\boldsymbol{\theta}_i \boldsymbol{\theta}_j}$, and let U be the $n \times k$ matrix that holds the leading k eigenvectors of P. (Note that $E[Y] = P$ except on the diagonal.) In the case of SBM, it can be shown that the rows of U correspond to k distinct points, and two nodes belong to the same block if and only if their expected embeddings are equal. For general DCBM, the rows of U fall along k rays emanating from the origin, with each ray corresponding to a unique block (Jin, 2015; Lei and Rinaldo, 2015).

Unfortunately, these geometric properties are not, in general, preserved by the edge-flipping routine: the embeddings of $\mathcal{M}_\varepsilon(Y)$ do not have the same geometric properties as the embeddings of Y. This limitation prevents direct application of standard DCBM results to $\mathcal{M}_\varepsilon(Y)$ without modification. However, from Corollary 3.9, we know that the "downshifted" network $(d_\varepsilon \circ \mathcal{M}_\varepsilon)(Y)$ shares the same eigenvectors as the original Y in expectation, suggesting that the embeddings of $(d_\varepsilon \circ \mathcal{M}_\varepsilon)(Y)$ and Y will concentrate in the same locations. Since $(d_\varepsilon \circ \mathcal{M}_\varepsilon)(Y)$ is a weighted network, these concentration results require further extensions of the standard results for binary networks (e.g., Lei and Rinaldo, 2015). Indeed, we state and prove explicit concentration bounds in Supplementary Material, Lemma A.1. This is the key fact that powers the main results of Section 5.

To perform spectral clustering on the edge-flipped network, then, we need only to modify the common algorithms to use the downshifted adjacency matrix $(d_\varepsilon \circ \mathcal{M}_\varepsilon)(Y)$ in place of $\mathcal{M}_\varepsilon(Y)$. Our proposed modified algorithms for SBM and DCBM are given in Algorithms 1 and 2, respectively. The k-means and k-medians problems, as well as their approximations, are defined in Supplementary Material, Section A.1.

Algorithm 1: Edge-Flipped Spectral Clustering (k-means)

```
Data: edge-flipped adjacency matrix $A \in \{0,1\}^{n\times n}$, number of blocks $k$,
      approximation error $\gamma$, privacy budget $\varepsilon$
Result: private estimate of block membership $\hat{\boldsymbol{\theta}} \in [k]^n$
Let $A_\downarrow = \begin{cases} A & \varepsilon = \infty \\ A - (e^\varepsilon + 1)^{-1}(\mathbf{1}_n\mathbf{1}_n^T - I_n) & \varepsilon < \infty \end{cases}$ ;
Let $\hat{u}_1, \dots, \hat{u}_k \in \mathbb{R}^n$ be the $k$ leading eigenvectors (by absolute value) of $A_\downarrow$ ;
Let $\hat{\boldsymbol{\theta}}$ be a $(1+\gamma)$-approximate solution to the $k$-means problem over the rows of
  $\hat{U}_\downarrow = [\hat{u}_1 \ \dots \ \hat{u}_k] \in \mathbb{R}^{n\times k}$ ;
```

Algorithm 2: Edge-Flipped Spectral Clustering (Normalized k-medians)

```
Data: edge-flipped adjacency matrix $A \in \{0,1\}^{n\times n}$, number of blocks $k$,
      approximation error $\gamma$, privacy budget $\varepsilon$
Result: private estimate of block membership $\hat{\boldsymbol{\theta}} \in [k]^n$
Let $A_\downarrow = \begin{cases} A & \varepsilon = \infty \\ A - (e^\varepsilon + 1)^{-1}(\mathbf{1}_n\mathbf{1}_n^T - I_n) & \varepsilon < \infty \end{cases}$ ;
Let $\hat{u}_1, \dots, \hat{u}_k \in \mathbb{R}^n$ be the $k$ leading eigenvectors (by absolute value) of $A_\downarrow$ ;
Let $\hat{U}_\downarrow = [\hat{u}_1 \ \dots \ \hat{u}_k] \in \mathbb{R}^{n\times k}$ ;

// construct row-normalized matrix $\hat{U}'_\downarrow$ over non-zero rows
Let $I_+ = \{i \in [n] : \|(\hat{U}_\downarrow)_{i*}\|_2 > 0\}$, $f : [|I_+|] \to [n]$ s.t. $f(i) = (I_+)_i$ ;
for $i = 1, \dots, |I_+|$ do
    Let $(\hat{U}'_\downarrow)_{i*} = (\hat{U}_\downarrow)_{f(i)*} / \|(\hat{U}_\downarrow)_{f(i)*}\|_2$ ;
end

Let $\hat{\boldsymbol{\theta}}'$ be a $(1+\gamma)$-approximate solution to the $k$-medians problem over the rows of
  $\hat{U}'_\downarrow$ ;
Let $\hat{\boldsymbol{\theta}}_i = \begin{cases} 1 & i \notin I_+ \\ \hat{\boldsymbol{\theta}}'_{f^{-1}(i)} & i \in I_+ \end{cases}$ ,   $i \in [n]$ ;
```

To evaluate the performance of the clustering algorithms, we assume knowledge of ground truth group labels $\boldsymbol{\theta}$ and measure two forms of misclassification. The first is an overall measure of misclassification that captures the proportion of misidentified nodes, up to a permutation of labelings. Let $S_{[k]}$ denote the set of all permutations $\sigma : [k] \to [k]$, and let $\mathbb{I}(\cdot)$ denote an indicator function. Then the *overall misclassification* for the estimated groups $\hat{\boldsymbol{\theta}}$ is given by

$$L(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}) = \min_{\sigma \in S_{[k]}} \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}(\sigma(\hat{\boldsymbol{\theta}}_i) \neq \boldsymbol{\theta}_i)). \tag{4.1}$$

As in Lei and Rinaldo (2015), we also consider the *worst-case misclassification* within a single community in order to account for trivialities[5]. This error is given by

$$\tilde{L}(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}) = \max_{j \in [k]} \min_{\sigma \in S_{[k]}} \frac{1}{|C_j|} \sum_{i \in C_j} \mathbb{I}(\sigma(\hat{\boldsymbol{\theta}}_i) \neq j). \tag{4.2}$$

5. Misclassification Bounds for Private Clustering

We generalize the results of Lei and Rinaldo (2015) developed for the non-private network setting to derive finite-sample misclassification bounds for spectral clustering of SBM and DCBM under local differential privacy using the modified spectral clustering algorithms proposed in Section 4 (Algorithms 1 and 2). We argue in Section 4 and prove in Supplementary Material, Section A that the spectral embeddings of Y and the weighted network $(d_\varepsilon \circ \mathcal{M}_\varepsilon)(Y)$ will concentrate in the same locations. In Lemma A.1, we also demonstrated that the concentration weakens with a stronger privacy guarantee (i.e., smaller ε). In this section, we use these concentration results to derive misclassification bounds for private spectral clustering of SBM and DCBM; all proofs are provided in Section A of the Supplementary Material. We begin with the results for SBM.

5.1. **SBM.**

Theorem 5.1. *Let $Y \sim \mathrm{SBM}(\boldsymbol{\theta}, B)$ with $\max B \geq \log n/n$, and minimum absolute eigenvalue $\lambda_B > 0$. Let $\varepsilon \in (0, \infty]$, and let $\hat{\boldsymbol{\theta}}_\varepsilon$ be the result of Algorithm 1 on $\mathcal{M}_\varepsilon(Y)$ (where $\mathcal{M}_\infty(Y) = Y$), using an approximation error of γ. Let*

$$g_\varepsilon(B) = \begin{cases} \frac{e^\varepsilon+1}{e^\varepsilon-1}\left(\max B + \frac{1}{e^\varepsilon-1}\right), & \varepsilon < \infty \\ \max B, & \varepsilon = \infty \end{cases} \tag{5.1}$$

There exists a universal constant c_1 such that with probability at least $1 - n^{-1}$,

$$\tilde{L}(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}_\varepsilon) \leq c_1 \frac{(2+\gamma)kn}{n_{min}^2 \lambda_B^2} g_\varepsilon(B)$$

and

$$L(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}_\varepsilon) \leq c_1 \frac{(2+\gamma)kn'_{max}}{n_{min}^2 \lambda_B^2} g_\varepsilon(B),$$

provided that

$$n_{min} > \lambda_B^{-1} \sqrt{c_1(2+\gamma)kn\, g_\varepsilon(B)}. \tag{5.2}$$

[5] Consider an asymptotic regime with $k = 2$ where one community is of constant size N_1 and the other is of size $n - N_1$. In this setting, a trivial estimator such as $\hat{\boldsymbol{\theta}} = \mathbf{1}_n$ can achieve consistency in the sense that $L(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}) \to 0$.

Taking $\varepsilon = \infty$ (i.e., the non-private setting) in Theorem 5.1 recovers a result that is exactly equivalent to the results for SBM in Lei and Rinaldo (2015). Consequently, the cost of privacy on misclassification is captured by $g_\varepsilon(B)$. Noting that $g_\varepsilon(B)$ is a decreasing function in ε and that $\lim_{\varepsilon\to\infty} g_\varepsilon(B) = g_\infty(B)$, Theorem 5.1 provides a smooth interpolation between the known finite-sample misclassification bounds without privacy and the corresponding bounds under local DP.

We can use the finite-sample bounds from Theorem 5.1 to construct particular asymptotic regimes under which $\hat{\boldsymbol{\theta}}_\varepsilon$ is consistent and derive convergence rates. For a given sequence of SBM networks, it suffices to show that

$$\frac{kn}{n_{\min}^2} g_\varepsilon(B) = o(1).$$

If this holds, condition (5.2) is met for large n, and both measures of misclassification, $\tilde{L}(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}_\varepsilon)$ and $L(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}_\varepsilon)$, tend to zero.

Since $g_\varepsilon(B)$ explains the difference in classification performance between private and non-private clustering, a comparison of convergence rates can be boiled down to the differences in asymptotic behavior of $g_\varepsilon(B)$. We note that if $\varepsilon = \infty$, then $g_\varepsilon(B) = \max B$. This will serve as our baseline. On the other hand, if $\varepsilon < \infty$, then $g_\varepsilon(B) = O(\max B + \zeta_\varepsilon^{-1} + \zeta_\varepsilon^{-2})$, where $\zeta_\varepsilon = e^\varepsilon - 1$, an increasing function of ε satisfying $\zeta_0 = 0$. (See Supplementary Material, Fact A.2.) Thus,

$$\tilde{L}(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}_\varepsilon) = \begin{cases} O_P\left(\frac{kn}{n_{\min}^2 \lambda_B^2} \max B\right) & \varepsilon = \infty \\ O_P\left(\frac{kn}{n_{\min}^2 \lambda_B^2} (\max B + \zeta_\varepsilon^{-1} + \zeta_\varepsilon^{-2})\right) & \varepsilon < \infty \end{cases}$$

$$\text{and} \quad L(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}_\varepsilon) = \begin{cases} O_P\left(\frac{kn'_{\max}}{n_{\min}^2 \lambda_B^2} \max B\right) & \varepsilon = \infty \\ O_P\left(\frac{kn'_{\max}}{n_{\min}^2 \lambda_B^2} (\max B + \zeta_\varepsilon^{-1} + \zeta_\varepsilon^{-2})\right) & \varepsilon < \infty \end{cases}.$$

From this, it is clear that if ε and $\max B$ are bounded away from zero, we can obtain the same convergence rate under local DP as in the absence of privacy. For example, for a fixed privacy budget and a sequence of dense SBMs (i.e., one for which the edge probabilities do not tend to zero), we can effectively employ local DP at the same cost, asymptotically, as in the non-private setting (up to a constant). For a sparse SBM (one in which $\max B \to 0$) with fixed privacy budget, however, this does not hold, as $g_\varepsilon(B) = \Theta(1)$ for $\varepsilon < \infty$, but $g_\infty(B) = \max B = o(1)$. Thus the convergence rate for sparse SBM slows by a factor of $(\max B)^{-1}$.

This slowing of convergence for sparse SBM limits the extent of sparsity that can be handled under this local DP mechanism. Consider a regime in which k is fixed, communities grow proportionally ($n_{\min} = \Theta(n)$), and B changes with n only via some scaling parameter $\alpha_n \to 0$, i.e., $B = \alpha_n B_0$ for some fixed matrix B_0. In this case, we have k constant, $\max B = \Theta(\alpha_n), \lambda_B = \Theta(\alpha_n)$, and so the worst-case misclassification is

$$\tilde{L}(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}_\varepsilon) = \begin{cases} O_P\left(\frac{1}{n\alpha_n}\right) & \varepsilon = \infty \\ O_P\left(\frac{\alpha_n + \zeta_\varepsilon^{-1} + \zeta_\varepsilon^{-2}}{n\alpha_n^2}\right) & \varepsilon < \infty. \end{cases} \tag{5.3}$$

Theorem 5.1 allows us to choose α_n as small as $\log n/n$, so we have consistency for non-private clustering down to $\log n/n$. For private clustering with a fixed privacy budget, however, we need $\alpha_n = \omega(n^{-1/2})$. The only way to achieve greater sparsity is to allow the privacy budget to grow arbitrarily large. For example, choosing $\varepsilon = \log(1+\alpha_n^{-1})$, we have $\zeta_\varepsilon^{-1} = \alpha_n$, and thus we can recover the same convergence rate and accommodate the same level of sparsity as without privacy—but at the cost of allowing unbounded privacy loss for large n.

5.2. **DCBM.** Generalizing the theoretical results for SBM to DCBM requires a bit of care. As described in Section 3, in the more general setting of DCBM, the expected embeddings for a given block fall along distinct rays emanating from the origin—in contrast with distinct points in the case of SBM. It is for this reason that Algorithm 2 is used for DCBM. The key theorethical result from this more general treatment is given below.

Theorem 5.2. *Let $Y \sim \mathrm{DCBM}(\boldsymbol{\theta}, \boldsymbol{\psi}, B)$ with $\max_{i\in C_j} \boldsymbol{\psi}_i = 1$ for $j \in [k]$, $\max B \geq \log n/n$, and minimum absolute eigenvalue $\lambda_B > 0$. Let $\varepsilon \in (0, \infty]$, and let $\hat{\boldsymbol{\theta}}_\varepsilon$ be the result of Algorithm 2 on $\mathcal{M}_\varepsilon(Y)$ (where $\mathcal{M}_\infty(Y) = Y$), using an approximation error of γ. Let $g_\varepsilon(B)$ as in Theorem 5.1. There exists a universal constant c_2 such that with probability at least $1 - n^{-1}$,*

$$L(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}_\varepsilon) \leq c_2 \frac{(2.5+\gamma)}{\tilde{n}_{min}\lambda_B} \sqrt{\frac{k}{n}\left(\sum_{j=1}^{k} n_j^2 \nu_j\right) g_\varepsilon(B)},$$

provided that

$$n_{min} > \frac{c_2(2.5+\gamma)\sqrt{kn\, g_\varepsilon(B) \sum_{j=1}^{k} n_j^2 \nu_j}}{\tilde{n}_{min}\lambda_B}. \tag{5.4}$$

As was the case with SBM, the results of Theorem 5.2 for $\varepsilon = \infty$ (non-private setting) match the original results of Lei and Rinaldo (2015). The cost of privacy is once again determined by the function $g_\varepsilon(B)$. In this case, however, the bound changes with the square root of $g_\varepsilon(B)$.

Comparing the results of Theorems 5.1 and 5.2, the more general result involves additional parameters, of course, but also generally provides a weaker result. Ignoring constants, condition (5.4) is more stringent than (5.2), as $\sqrt{\sum_{j=1}^{k} n_j^2 \nu_j}/\tilde{n}_{\min} \geq \sqrt{k}$. The resulting upper bound on $L(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}_\varepsilon)$ also results in weaker convergence rates. In keeping with Lei and Rinaldo (2015), Theorem 5.2 offers no bound on $\tilde{L}(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}_\varepsilon)$, as its proof bounds only the total number of misclassified nodes in the network. A trivial bound for $\tilde{L}$ can be obtained by observing that:

$$\tilde{L}(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}_\varepsilon) \leq (n/n_{\min}) L(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}_\varepsilon). \tag{5.5}$$

To simplify matters, we illustrate an asymptotic regime in which communities grow proportionally ($n_j = \Theta(n/k)$ for all $j \in [k]$) and assume that there exists a global lower bound $0 < a \leq \boldsymbol{\psi}_i$ for $i \in [n]$. Under these conditions, we can state a simple asymptotic result.

Lemma 5.3. *Suppose Y satisfies the conditions of Theorem 5.2, and suppose further that $0 < a \leq \boldsymbol{\psi}_i \leq 1$ for all $i \in [n]$ and $n_j = \Theta(n/k)$ for all $j \in [k]$. Then*

$$\frac{k^2\sqrt{g_\varepsilon(B)}}{a^3\lambda_B\sqrt{n}} = o(1) \implies L(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}_\varepsilon) = O_P\left(\frac{k\sqrt{g_\varepsilon(B)}}{a^3\lambda_B\sqrt{n}}\right).$$

To compare this to the results seen in Eq. (5.3), consider again the case when $B = \alpha_n B_0$ for some fixed matrix B_0 and sequence $\alpha_n \to 0$. Then since k is fixed, $n/n_{\min} = \Theta(1)$, and so combining Lemma 5.3 with Eq. (5.5) yields a convergence rate of:

$$\tilde{L}(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}_\varepsilon) = \begin{cases} O_P\left(\frac{1}{a^3\sqrt{n\alpha_n}}\right) & \varepsilon = \infty \\ O_P\left(\frac{\sqrt{\alpha_n + \zeta_\varepsilon^{-1} + \zeta_\varepsilon^{-2}}}{a^3 \alpha_n \sqrt{n}}\right) & \varepsilon < \infty. \end{cases}$$

For SBM, where $a = 1$—or more generally for any regime in which a is bounded away from zero—this bound is precisely the square root of the bound obtained in Eq. (5.3). Once again, we see that the convergence rate for sparse networks slows under private methods, and while some sparsity is attainable with privacy through this method, to accommodate the same level of sparsity with privacy as without, we would again need to allow the privacy-loss budget ε to grow arbitrarily large.

6. Empirical Evaluations

6.1. **Simulation Studies.** The theoretical results above suggest that the edge-flip privacy mechanism can be used to achieve convergence for spectral clustering of SBM and DCBM networks that is similar in rate for dense networks, while slower for sparse networks. However, existing work suggests that the established non-private bounds from which we developed our theory may not be tight (Lei and Rinaldo, 2015; Athreya et al., 2016). Here we evaluate these results through simulation to assess how empirical performance compares with theoretical. To facilitate this analysis, we consider two special cases of the SBM and DCBM.

Definition 6.1 (Symmetric SBM)**.** The **symmetric SBM** is an SBM network consisting of n nodes, k equal-sized blocks, and $B = pI_k + r\mathbf{1}_k\mathbf{1}_k^T$. For such a network, we write $Y \sim \mathrm{SSBM}(n, k, p, r)$.

Definition 6.2 (Symmetric DCBM)**.** The **symmetric DCBM** is a DCBM network consisting of n nodes, k equal-sized blocks, $B = pI_k + r\mathbf{1}_k\mathbf{1}_k^T$, and $\boldsymbol{\psi} \in [a, 1]^n$ taking value 1 for the first node of a given block and values randomly drawn from $\mathrm{Uniform}(a, 1)$ elsewhere. For such a network, we write $Y \sim \mathrm{SDCBM}(n, k, p, r, a)$.

Starting with SBM, we simulate two regimes following the symmetric SBM. In the first setting, we use a dense symmetric SBM. We consider 16 values of n ranging from $n = 30$ to $n = 12000$ and $\varepsilon \in \{0.5, 0.75, 1, 1.5, 2, 3, 4, \infty\}$, where $\varepsilon = \infty$ represents the original network (i.e., no privacy). For each n, ε pair, we draw 100 networks $Y^{(i)} \sim \mathrm{SSBM}(n, k = 3, p = 0.2, r = 0.05)$. We then apply Algorithm 1 to $\mathcal{M}_\varepsilon(Y^{(i)})$, and report the mean overall misclassification rate $L(\boldsymbol{\theta}^{(i)}, \hat{\boldsymbol{\theta}}^{(i)})$ of Eq. (4.1). The results of this experiment are plotted on log–log axes in the left panel of Figure 1. The bounds from Section 5 suggest polynomial convergence rates in this setting that should be approximately the same without or without privacy. Indeed, on a log–log scale, there are no clear differences in these curves, although all seem to exhibit downward curvature suggesting faster than polynomial convergence.

For the second setting, we use a sparse symmetric SBM. Here we consider 13 values of n ranging from $n = 10$ to $n = 12800$ and the same eight values of ε as in the first setting. For each n, ε pair, we draw 100 networks $Y^{(i)} \sim \mathrm{SSBM}(n, k = 2, p = 1.5n^{-.3}, r = .15n^{-.3})$, then apply Algorithm 1 to $\mathcal{M}_\varepsilon(Y^{(i)})$, as in the first setting. The results are plotted on log–log

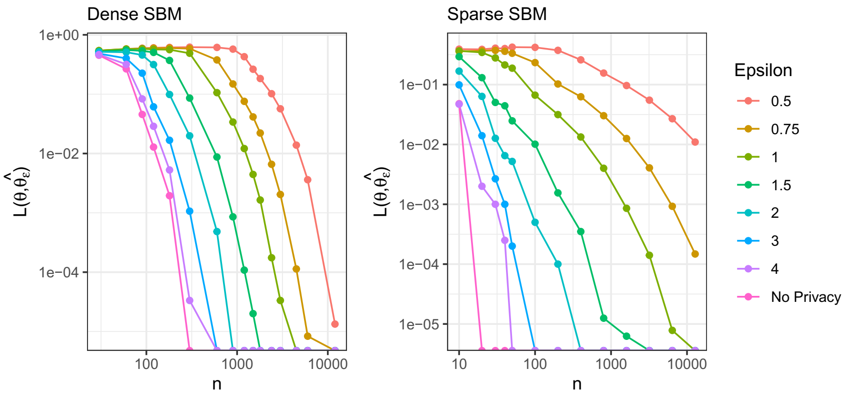

Figure 1: Proportion of misclassified nodes in simulated SBM networks for various values of ε, n. Left (dense): $Y \sim \text{SSBM}(n, k = 3, p = 0.2, r = 0.05)$, Right (sparse): $Y \sim \text{SSBM}(n, k = 2, p = 1.5n^{-.3}, r = .15n^{-.3})$

axes in the right panel of Figure 1. Here, the curves corresponding to each value of ε are no longer parallel: for smaller values of ε, the convergence rate slows.

For DCBM, we simulate two additional regimes, one dense and one sparse, in a manner similar to the SBM simulations. In the dense setting, we use a dense symmetric DCBM with the same values of n, ε as used in the dense SBM simulations. For each n, ε pair, we draw 100 networks $Y^{(i)} \sim \text{SDCBM}(n, k = 3, p = 0.4, r = 0.05, a = 0.3)$. In the sparse setting, we use the same ε values as in the other settings and 12 values of n ranging from $n = 20$ to $n = 12800$. Then we draw 100 networks $Y^{(i)} \sim \text{SDCBM}(n, k = 2, p = 2n^{-.25}, r = 0.1n^{-.25}, a = 0.3)$. In both DCBM settings, we report the mean overall misclassification rate after applying Algorithm 2. The results of these simulations are plotted on log–log axes in Figure 2. In contrast with the SBM simulations, the DCBM simulations generally exhibit slower convergence. Similar to the SBM simulations, we see a clear slowing of convergence rate for smaller values of ε in the sparse setting.

While the results of these simulations are generally consistent with the theory of Section 5, our simulations suggest the bounds given here and in Lei and Rinaldo (2015) are not tight, in the sense that the observed convergence rates appear to beat the theoretical guarantees, sometimes by considerable order. This appears consistent with the empirical findings of Athreya et al. (2017, Section 5), where a related method is evaluated in the non-private setting. In the private, sparse setting, the precise nature in which convergence slows with ε is not fully captured by our theoretical results. Specifically, for larger values of ε, the convergence properties appear more similar to the non-private setting than suggested by the theoretical bounds.

6.2. **Performance on Observed Networks.** To assess the practicality of these methods, we applied the edge-flip mechanism and Algorithm 2 to several real-world datasets, then compared the performance of spectral clustering on the private networks over various privacy budgets. We show the trade-offs of privacy loss and accuracy.

The first dataset we considered is a small, classical network used in the SBM literature: Hansell's friendship data (Hansell, 1984), which were used in Wang and Wong (1987);

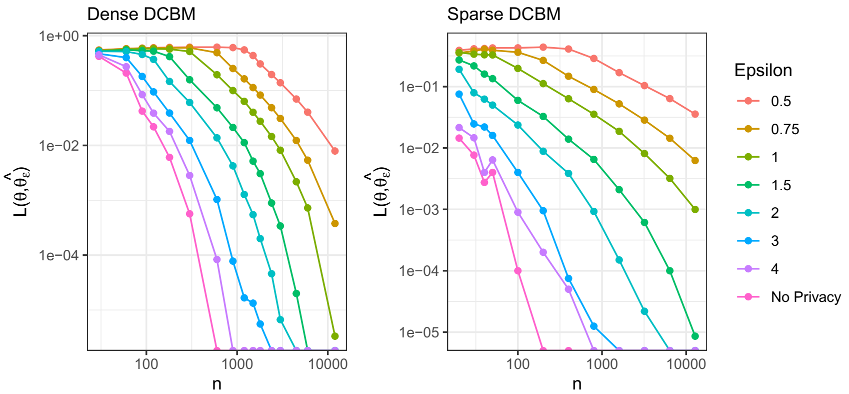

Figure 2: Proportion of misclassified nodes in simulated DCBM networks for various values of ε, n. Left (dense): $Y \sim \text{SDCBM}(n, k = 3, p = 0.4, r = 0.05, a = 0.3)$, Right (sparse): $Y \sim \text{SDCBM}(n, k = 2, p = 2n^{-.25}, r = 0.1n^{-.25}, a = 0.3)$

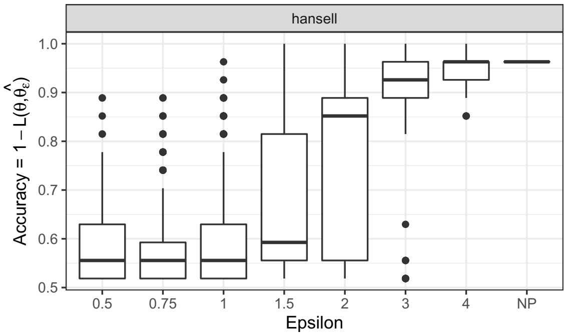

Figure 3: Trade-off of privacy-loss and accuracy of private spectral clustering on a small dataset, Hansell's friendship data (Hansell, 1984), with varying privacy-loss ε, and "NP" denoting non-private data.

Snijders and Nowicki (1997). This is a directed network of $n = 27$ (13 male, 14 female) students, and the presence of an edge (i, j) indicates that student i considers student j to be a friend. We symmetrized the network by setting $Y_{ij} = \max\{Y_{ij}, Y_{ji}\}$, and we used the students' sexes as ground-truth labels.

For this small network, we considered $\varepsilon \in \{0.5, 0.75, 1, 1.5, 2, 3, 4, \infty\}$, and for each value of ε, we applied the privacy mechanism and spectral clustering algorithm 500 times. Box plots showing the overall classification accuracy, $1 - L(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})$ from Eq. (4.1), for each value of ε are given in Figure 3. Accuracy remains high for values of $\varepsilon > 2$; for such a small network, this is about as good as we can expect.

Next, we considered two datasets from the DCBM literature. The first network consists of $n = 1137$ Facebook users from Simmons College (Traud et al., 2012; Chen et al., 2018), and edges represent friendships. Each student belongs to one of $k = 4$ class years (2006 to 2009). These class years are used as ground-truth group memberships, which are inferred through their friendships. The second dataset is the well-known network of political blogs

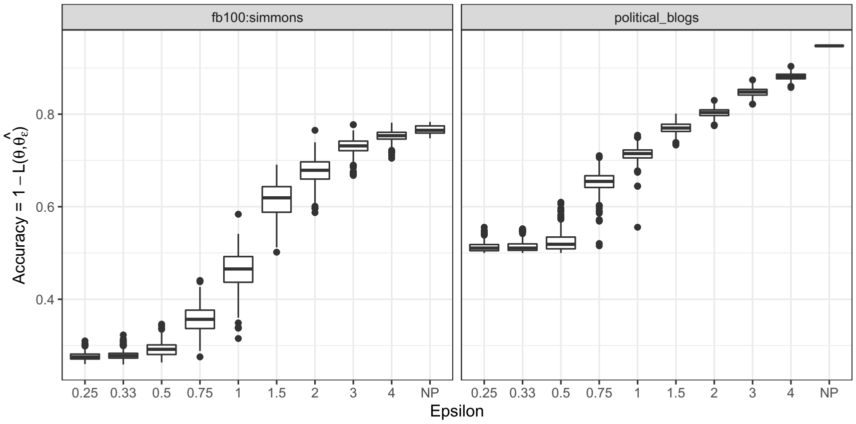

Figure 4: Trade-off of privacy-loss and accuracy of private spectral clustering on well-known DCBM datasets, with varying privacy-loss ε, and "NP" denoting non-private data. Left: Facebook friendships of Simmons College students (Traud et al., 2012), Right: political blogs network (Adamic and Glance, 2005)

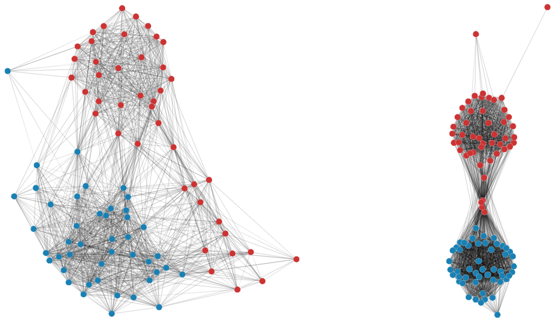

Figure 5: Visualization of congressional voting networks for 70th U.S. Senate (left) and 110th U.S. Senate (right). Democrats are depicted as blue nodes and Republican as red.

from Adamic and Glance (2005), which has been widely used in the DCBM literature (Karrer and Newman, 2011; Jin, 2015; Chen et al., 2018; Abbe, 2018). This network consists of $n = 1222$ blogs on U.S. politics, which have been categorized as either left-leaning or right-leaning ($k = 2$ groups). An edge in this network represents the existence of a hyperlink between the two blogs, and these hyperlinks are used to infer the left- or right-leaning label for each node. The performance of the private spectral clustering methods on these networks is shown in Figure 4. In both of these datasets, we see a steady decline in performance as ε decreases, without a clear inflection point.

The theoretical and simulated results from earlier suggest that these methods are best paired with suitably large or dense networks. A relevant collection of networks can be found, for example, in Andris et al. (2015), where networks of Democrat and Republican members of the U.S. House of Representatives are constructed based on voting similarity. Using their

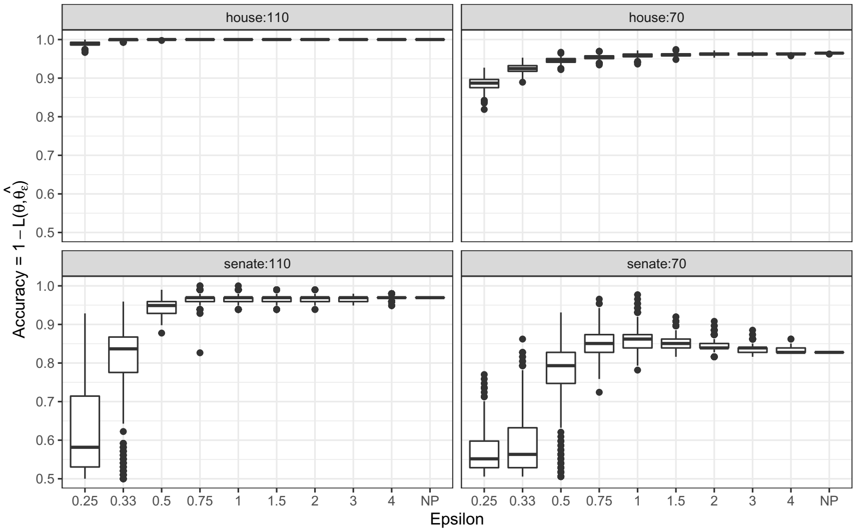

Figure 6: Trade-off of privacy-loss and accuracy of private spectral clustering on U.S. Congressional voting networks. Clockwise from top left: 110th House of Representatives, 70th House of Representatives, 70th Senate, 110th Senate

methods and data from Lewis et al. (2021), we reconstructed these networks for a number of sessions of the U.S. House of Representatives and U.S. Senate (Hehir, 2022b).[6]

Changes over time in the patterns of voting among members of Congress offer us a range of networks that span nearly complete separation of the parties (particularly in recent years) to networks that exhibit greater levels of connectivity across parties. Figure 5 depicts two such networks, with the 70th U.S. Senate on the left and the 110th U.S. Senate on the right. In the 70th Senate, 13% of cross-party pairs are connected, while 76% and 59% of Democrat and Republican pairs are connected (respectively). In the 110th Senate, 9% of cross-party pairs are connected, while 98% and 88% of Democrat and Republican pairs are connected (respectively).

For the 70th Senate ($n = 87$), 110th Senate ($n = 98$), 70th House ($n = 425$), and 110th House ($n = 423$), we applied the private spectral clustering methods 50 times for each value of $\varepsilon \in \{0.25, 0.33, 0.5, 0.75, 1, 1.5, 2, 3, 4, \infty\}$. The results are depicted in Figure 6. The high density (and relatively low variability of degree) in these networks enables impressive clustering performance: for the House networks, clustering performance is solid for all ε considered; in the smaller Senate networks, we see a drop in performance for values of $\varepsilon < 0.5$.

7. Discussion

Just as SBM and DCBM are fundamental network models featuring community structure, the edge-flip mechanism is a fundamental network privacy mechanism providing local differential privacy on networks at the edge level. A great advantage of the edge-flip mechanism is

[6] Although these networks are produced from public information, networks involving public figures like members of Congress illustrate a case where edge privacy is particularly relevant, since the set of nodes is public knowledge, but their interactions or relations may be sensitive.

that it produces a synthetic network with clear distributional properties. In particular, the explicit mixture distribution of $\mathcal{M}_\varepsilon(Y)$ described in Lemma 3.7—in contrast with the unclear distributional properties of some alternative mechanisms—greatly facilitates the process of accounting for privacy in various statistical procedures. Moreover, by nature of producing a synthetic network, any number of statistical procedures can be applied to the single output $\mathcal{M}_\varepsilon(Y)$ with a fixed privacy budget ε.

Our work demonstrates how to use the clear distribution of $\mathcal{M}_\varepsilon(Y)$ to extend the theory of consistent community detection to the local DP setting. This is, to our knowledge, the first attempt to address this important problem. Algorithms 1 and 2 represent modest extensions to the algorithms originally proposed in Lei and Rinaldo (2015), requiring only the addition of a small "downshift" transformation to unlock a powerful privacy guarantee with a clear accounting of the cost to clustering performance. In fact, Corollary 3.8 suggests that for SBM networks in particular, this downshift transformation is optional, as $\mathcal{M}_\varepsilon(Y)$ is itself an SBM whose community structure is identical to Y.

Focusing on the specific case of SBM, we may ask whether it is preferable to perform clustering for SBM on $(d_\varepsilon \circ \mathcal{M}_\varepsilon)(Y)$, as prescribed by Algorithm 1, or to proceed with ordinary spectral clustering on $\mathcal{M}_\varepsilon(Y)$ per the closure property of Corollary 3.8. By combining the closure property with the $\varepsilon = \infty$ case of Theorem 5.1, one can obtain performance bounds in terms of the matrix $\tau_\varepsilon(B)$ of Eq. (3.2). In some cases, these bounds may be superior to the bounds resulting from Theorem 5.1 alone.[7] Nonetheless, we have opted to present the results for the downshifted procedure for the sake of generality. This allows us to keep the SBM misclassification bounds in terms of the smallest absolute eigenvalue of B instead of that of $\tau_\varepsilon(B)$, where a general relationship between these two quantities is elusive.

The benefits of the edge-flip mechanism's tractable distribution are, of course, not limited to the network models considered in this paper, nor to the field of network clustering. For example, Karwa et al. (2017) have previously shown how to adjust inferential procedures for exponential random graph models using the edge-flip mechanism. These examples are likely only scratching the surface of the theory that can be extended to accommodate privacy under the edge-flip mechanism.

The primary drawback to the edge-flip mechanism is its relationship to sparse networks. We posit that this limitation is inherent to working with a local-DP synthetic network, in which the magnitude of noise required for privacy begins to dwarf the signal present in a sparse network. Looking at the edge-flip mechanism in particular, for a sparse network Y, the mixture $\mathcal{M}_\varepsilon(Y)$ will be considerably more dense, and a disproportionate number of its edges will result from the privacy mechanism, not the original network of interest. Although these changes preserve key properties of SBM and DCBM networks, reductions in sparsity dilute the signal-to-noise ratio considerably. While the theory supports the claim that these methods are consistent for SBM and DCBM networks exhibiting mild sparsity, empirical performance for sparse networks indicates a considerable utility loss when a very strong privacy guarantee is applied to a sparse network. Other privacy mechanisms preserving the sparsity of a network, especially non-local mechanisms, could be considered for greater performance in these cases.

[7] For example, if B is positive definite, then one can show that $g_\infty(\tau_\varepsilon(B))/\lambda^2_{\tau_\varepsilon(B)} < g_\varepsilon(B)/\lambda^2_B$ via Weyl's inequality.

The trade-off of privacy-loss and accuracy that we observe for label recovery on dense networks such as the Congressional voting networks, where considerable accuracy is maintained at $\varepsilon = 0.5$, suggests that the methods employed here are of more than just theoretical interest. By comparison, in their work with exponential random graph models and the edge-flip mechanism, Karwa et al. (2017) use values of ε ranging from 3 to 6 on a network of similar size. Nonetheless, it would be useful to develop theoretical results for alternative privacy mechanisms that can accommodate greater sparsity.

Lastly, while the results given here focus on spectral clustering of SBM and DCBM, some of the insights from these techniques are suggestive of further theoretical developments in related fields. For example, concentration bounds for the spectral embeddings of the downshifted private network could potentially be extended to the more general family of random dot product graphs (Athreya et al., 2017). Doing this would open a wide range of avenues for future work.

Acknowledgment

Computations for this research were performed on the Pennsylvania State University's Institute for Computational and Data Sciences' Roar supercomputer. This work was supported in part by NIAID/NIH R01-AI136664 and NSF Award No. SES-1853209 to The Pennsylvania State University.

Code and Data

Complete code and data required to replicate the results of this paper can be found in Hehir (2022a).

References

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

E. Abbe, J. Fan, and K. Wang. An ℓ_p theory of PCA and spectral clustering, 2021. http://arxiv.org/abs/2006.14062.

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*, pages 36–43, 2005. https://doi.org/10.1145/1134271.1134277.

C. Andris, D. Lee, M. J. Hamilton, M. Martino, C. E. Gunning, and J. A. Selden. The rise of partisanship and super-cooperators in the us house of representatives. *PLOS One*, 10 (4):e0123507, 2015. https://doi.org/10.1371/journal.pone.0123507.

A. Athreya, C. E. Priebe, M. Tang, V. Lyzinski, D. J. Marchette, and D. L. Sussman. A limit theorem for scaled eigenvectors of random dot product graphs. *Sankhya A*, 78(1): 1–18, 2016. ISSN 0976-836X. https://doi.org/10.1007/s13171-015-0071-x.

A. Athreya, D. E. Fishkind, M. Tang, C. E. Priebe, Y. Park, J. T. Vogelstein, K. Levin, V. Lyzinski, and Y. Qin. Statistical inference on random dot product graphs: a survey. *The Journal of Machine Learning Research*, 18(1):8393–8484, 2017. http://jmlr.org/papers/v18/17-448.html.

N. Binkiewicz, J. T. Vogelstein, and K. Rohe. Covariate-assisted spectral clustering. *Biometrika*, 104(2):361–377, 2017. ISSN 0006-3444. https://doi.org/10.1093/biomet/asx008.

J. Blocki, A. Blum, A. Datta, and O. Sheffet. The Johnson-Lindenstrauss transform itself preserves differential privacy. In *2012 IEEE 53rd Annual Symposium on Foundations of Computer Science*, volume abs/1204.2136, pages 410–419. IEEE, IEEE, 10 2012. https://doi.org/10.1109/focs.2012.67.

Y. Chen, X. Li, J. Xu, et al. Convexified modularity maximization for degree-corrected stochastic block models. *Annals of Statistics*, 46(4):1573–1602, 2018. ISSN 0090-5364. https://doi.org/10.1214/17-AOS1595.

J.-J. Daudin, F. Picard, and S. Robin. A mixture model for random graphs. *Statistics and Computing*, 18(2):173–183, 2008. ISSN 0960-3174. https://doi.org/10.1007/s11222-007-9046-7.

A. Decelle, F. Krzakala, C. Moore, and L. Zdeborová. Asymptotic analysis of the stochastic block model for modular networks and its algorithmic applications. *Physical Review E*, 84 (6):066106, 12 2011. ISSN 1539-3755. https://doi.org/10.1103/physreve.84.066106.

J. C. Duchi, M. I. Jordan, and M. J. Wainwright. Local privacy and statistical minimax rates. In *2013 IEEE 54th Annual Symposium on Foundations of Computer Science*, volume abs/1302.3203, pages 429–438. IEEE, IEEE, 10 2013. https://doi.org/10.1109/focs.2013.53.

C. Dwork, F. McSherry, K. Nissim, and A. Smith. Calibrating noise to sensitivity in private data analysis. In S. Halevi and T. Rabin, editors, *Theory of Cryptography Conference*, volume 3876, pages 265–284. Springer, Springer Berlin Heidelberg, 2006. ISBN 9783540327318. https://doi.org/10.1007/11681878_14.

A. Gupta, A. Roth, and J. Ullman. Iterative constructions and private data release. In R. Cramer, editor, *Theory of Cryptography Conference*, volume abs/1107.3731, pages 339–356. Springer, Springer Berlin Heidelberg, 2012. ISBN 9783642289132. https://doi.org/10.1007/978-3-642-28914-9_19.

S. Hansell. Cooperative groups, weak ties, and the integration of peer friendships. *Social Psychology Quarterly*, 47:316–328, 12 1984. ISSN 0190-2725. https://doi.org/10.2307/3033634.

M. Hay, C. Li, G. Miklau, and D. Jensen. Accurate estimation of the degree distribution of private networks. In W. Wang, H. Kargupta, S. Ranka, P. S. Yu, and X. Wu, editors, *2009 Ninth IEEE International Conference on Data Mining*, pages 169–178. IEEE, IEEE, 12 2009. https://doi.org/10.1109/icdm.2009.11.

J. Hehir. private-spectral-clustering: Spectral Clustering with Edge-Flip Differential Privacy. Software, Zenodo, Sept. 2022a. https://doi.org/10.5281/zenodo.7098162.

J. Hehir. jonhehir/congress-voting-networks: v2022.01. Software, Zenodo, Jan. 2022b. https://doi.org/10.5281/zenodo.5838420.

P. W. Holland, K. B. Laskey, and S. Leinhardt. Stochastic blockmodels: First steps. *Social Networks*, 5(2):109–137, 1983. ISSN 0378-8733. https://doi.org/10.1016/0378-8733(83)90021-7.

J. Imola, T. Murakami, and K. Chaudhuri. Locally differentially private analysis of graph statistics. In M. Bailey and R. Greenstadt, editors, *30th USENIX Security Symposium (USENIX Security 21)*, pages 983–1000. USENIX Association, 10 2021. https://www.usenix.org/conference/usenixsecurity21/presentation/imola.

J. Jin. Fast community detection by SCORE. *Annals of Statistics*, 43(1):57–89, 2015. ISSN 0090-5364. https://doi.org/10.1214/14-AOS1265.

A. Joseph and B. Yu. Impact of regularization on spectral clustering. *Annals of Statistics*, 44(4):1765–1791, 2016. https://doi.org/10.1109/ita.2014.6804241.

B. Karrer and M. E. Newman. Stochastic blockmodels and community structure in networks. *Physical Review E*, 83(1):016107, 2011. ISSN 1539-3755. https://doi.org/10.1103/physreve.83.016107.

V. Karwa and A. Slavković. Inference using noisy degrees: Differentially private β-model and synthetic graphs. *Annals of Statistics*, 44(1):87–112, February 2016. ISSN 0090-5364. https://doi.org/10.1214/15-AOS1358.

V. Karwa, P. N. Krivitsky, and A. B. Slavković. Sharing social network data: differentially private estimation of exponential family random-graph models. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 66(3):481–500, April 2017. ISSN 0035-9254. https://doi.org/10.1111/rssc.12185.

S. P. Kasiviswanathan, H. K. Lee, K. Nissim, S. Raskhodnikova, and A. Smith. What can we learn privately? *SIAM Journal on Computing*, 40(3):793–826, 10 2011. https://doi.org/10.1109/focs.2008.27.

S. P. Kasiviswanathan, K. Nissim, S. Raskhodnikova, and A. Smith. Analyzing graphs with node differential privacy. In A. Sahai, editor, *Theory of Cryptography*, volume 7785, pages 457–476. Springer Berlin Heidelberg, Berlin, Heidelberg, 2013. ISBN 9783642365935. https://doi.org/10.1007/978-3-642-36594-2_26.

B. Kim, K. H. Lee, L. Xue, and X. Niu. A review of dynamic network models with latent variables. *Statistics Surveys*, 12:105, 2018. ISSN 1935-7516. https://doi.org/10.1214/18-SS121.

C. Lee and D. J. Wilkinson. A review of stochastic block models and extensions for graph clustering. *Applied Network Science*, 4(1):1–50, 12 2019. ISSN 2364-8228. https://doi.org/10.1007/s41109-019-0232-2.

J. Lei and A. Rinaldo. Consistency of spectral clustering in stochastic block models. *Annals of Statistics*, 43(1):215–237, 2015. ISSN 0090-5364. https://doi.org/10.1214/14-AOS1274.

J. B. Lewis, K. Poole, H. Rosenthal, A. Boche, A. Rudkin, and L. Sonnet. Voteview: Congressional roll-call votes database, 2021. https://voteview.com/.

F. McSherry. Spectral partitioning of random graphs. In *Proceedings 42nd IEEE Symposium on Foundations of Computer Science*, pages 529–537. IEEE, IEEE, 2001. https://doi.org/10.1109/sfcs.2001.959929.

Y. Mülle, C. Clifton, and K. Böhm. Privacy-integrated graph clustering through differential privacy. In P. M. Fischer, G. Alonso, M. Arenas, and F. Geerts, editors, *EDBT/ICDT Workshops*, volume 1330, pages 247–254, 2015. http://ceur-ws.org/Vol-1330/paper-39.pdf.

A. Y. Ng, M. I. Jordan, and Y. Weiss. On spectral clustering: Analysis and an algorithm. In T. G. Dietterich, S. Becker, and Z. Ghahramani, editors, *Advances in Neural Information Processing Systems*, pages 849–856. MIT Press, 2002. https://proceedings.neurips.cc/paper/2001/hash/801272ee79cfde7fa5960571fee36b9b-Abstract.html.

H. H. Nguyen, A. Imine, and M. Rusinowitch. Detecting communities under differential privacy. In E. R. Weippl, S. K. 0001, and S. D. C. d. Vimercati, editors, *Proceedings of the 2016 ACM on Workshop on Privacy in the Electronic Society*, pages 83–93. ACM, 10 2016. https://doi.org/10.1145/2994620.2994624.

T. Qin and K. Rohe. Regularized spectral clustering under the degree-corrected stochastic blockmodel. In C. J. C. Burges, L. Bottou, Z. Ghahramani, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems*, pages 3120–3128, 2013. https://proceedings.neurips.cc/paper/2013/hash/0ed9422357395a0d4879191c66f4faa2-Abstract.html.

Z. Qin, T. Yu, Y. Yang, I. Khalil, X. Xiao, and K. Ren. Generating synthetic decentralized social graphs with local differential privacy. In B. M. Thuraisingham, D. Evans, T. Malkin, and D. Xu, editors, *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security—CCS '17*, pages 425–438, Dallas, Texas, USA, 10 2017. ACM Press. https://doi.org/10.1145/3133956.3134086.

K. Rohe, S. Chatterjee, and B. Yu. Spectral clustering and the high-dimensional stochastic blockmodel. *Annals of Statistics*, 39(4):1878–1915, 2011. ISSN 0090-5364. https://doi.org/10.1214/11-AOS887.

T. A. Snijders and K. Nowicki. Estimation and prediction for stochastic blockmodels for graphs with latent block structure. *Journal of Classification*, 14(1):75–100, January 1997. ISSN 0176-4268. https://doi.org/10.1007/s003579900004.

A. L. Traud, P. J. Mucha, and M. A. Porter. Social structure of Facebook networks. *Physica A: Statistical Mechanics and its Applications*, 391(16):4165–4180, 2012. ISSN 0378-4371. https://doi.org/10.1016/j.physa.2011.12.021.

U. Von Luxburg. A tutorial on spectral clustering. *Statistics and Computing*, 17(4):395–416, 12 2007. ISSN 0960-3174. https://doi.org/10.1007/s11222-007-9033-z.

Y. J. Wang and G. Y. Wong. Stochastic blockmodels for directed graphs. *Journal of the American Statistical Association*, 82(397):8–19, 1987. ISSN 0162-1459. https://doi.org/10.1080/01621459.1987.10478385.

L. Wasserman and S. Zhou. A statistical framework for differential privacy. *Journal of the American Statistical Association*, 105(489):375–389, March 2010. ISSN 0162-1459. https://doi.org/10.1198/jasa.2009.tm08651.