Local Differential Privacy-Preserving Spectral Clustering for General Graphs

Anonymous authors
Paper under double-blind review

Abstract

Spectral clustering is a widely used algorithm to find clusters in networks. Several researchers have studied the stability of spectral clustering under local differential privacy with the additional assumption that the underlying networks are generated from the stochastic block model (SBM). However, we argue that this assumption is too restrictive since social networks do not originate from the SBM. Thus, we delve into an analysis for general graphs in this work. Our primary focus is the edge flipping method – a common technique for protecting local differential privacy. We show that, when the edges of an n-vertex graph satisfying some reasonable well-clustering assumptions are flipped with a probability of $O(\log n/n)$, the clustering outcomes are largely consistent. Empirical tests further corroborate these theoretical findings. Conversely, although clustering outcomes have been stable for dense and well-clustered graphs produced from the SBM, we show that in general, spectral clustering may yield highly erratic results on certain dense and well-clustered graphs when the flipping probability is $\omega(\log n/n)$. This indicates that the best privacy budget obtainable for general graphs is $\Theta(\log n)$.

1 Introduction

As the demand for trustworthy artificial intelligence grows, the need to protect user privacy becomes more crucial. Several methods have been proposed to address this concern. Among these, differential privacy is the most common. Introduced by Dwork (2008), differential privacy measures the amount of privacy a system leaks by using a metric called the privacy budget. This method involves corrupting users' information, then processing the corrupted data to obtain statistical conclusions while still maintaining privacy. Developing algorithms that can accurately provide statistical conclusions from the corrupted information is a topic of interest among many researchers such as Zhu et al. (2017).

In this work, we are interested in a variant of differential privacy called local differential privacy introduced by Kasiviswanathan et al. (2011). Unlike traditional differential privacy, local differential privacy does not allow the collection of all users' information before it is corrupted. Instead, it requires users to corrupt their data at their local devices before sending them to central servers. This ensures that users' information is not leaked during transmission. As discussed by Erlingsson et al. (2014) and Apple's Differential Privacy Team (2017), local differential privacy is used by companies for their services. One of the advantages of the local differential privacy notion is that the information revealed from the users' sensitive information is quantified by a term called *privacy budget*. We say that an algorithm protects users' information when its privacy budget is small.

We focus on algorithms for social networks. In a social network, each user is represented by a node, and their relationships with other users are represented by edges. One technique for protecting user privacy under the local differential privacy notion is randomized response or edge flipping, which is a technique considered in Warner (1965), Mangat (1994), and Wang et al. (2016). In this technique, before sending their adjacency vector (which represents their friend list) to the central server, each bit in the adjacency vector is flipped with a specified probability p. We obtain a local differential privacy with the budget of $\Theta(\log 1/p)$ by the flipping.

Several algorithms, including those introduced by Imola et al. (2022; 2021), have been proposed for processing social networks of which edges are flipped. These include graph clustering algorithms such as Ji et al. (2020); Mohamed et al. (2022); Fu et al. (2023). One of the most widely used and scalable graph clustering algorithms – spectral clustering – has also received a lot of attention in this context. Many analyses such as Hehir et al. (2022) have been recently done for this algorithm. However, all of these analyses assume that the input social networks are generated from the stochastic block model (SBM).

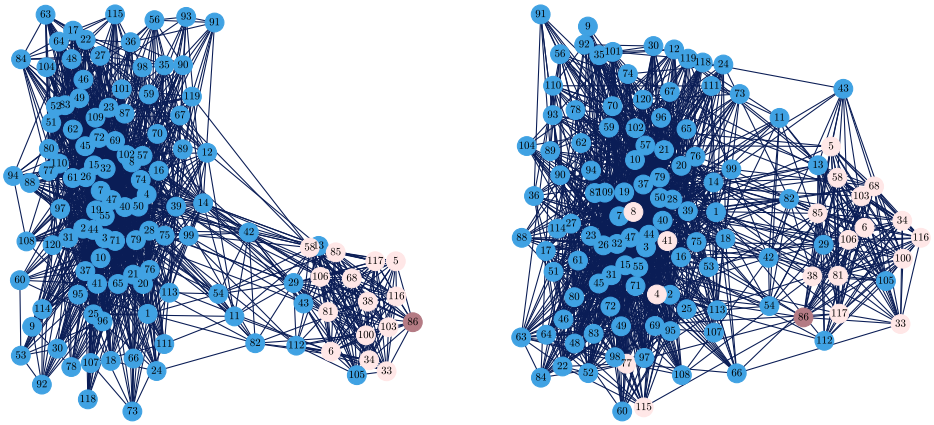
1.1 Our Contribution

We argue that assuming that the input graph is generated from the SBM is too restrictive. Thus, in this study, we consider the robustness of spectral clustering for general graphs. In what follows, let G be an n-vertex input graph. Our main contribution of Section 3 can be summarized by the following theorem:

Theorem 1.1 (Informal version of Theorem 3.1)**.** *Let G' be obtained from G via the edge flipping mechanism with probability $p = O(\log n/n)$. Then, under some reasonable assumptions, the number of vertices misclassified by the spectral clustering algorithm by running it on G' instead of G is $O(\eta(G) \cdot n)$ with probability $1 - o(1)$, where $\eta(G)$ is a small constant.*

In simpler terms, we demonstrate that:

$$\text{Spectral clustering is robust against edge flipping or the randomized response method with probability } p = O(\log n/n), \text{ or privacy budget } \epsilon = \Omega(\log n). \tag{1.1}$$



(a) (b)

Figure 1.1: A part of the Facebook network detailed in Leskovec & Mcauley (2012) before and after flipping edges with a probability of 0.005. The neighborhood (colored light pink) of node 86 changes a lot after the flipping.

Our findings are depicted in Figure 1.1. This figure focuses on a segment of the Facebook network in the Stanford network analysis project (SNAP) described in Leskovec & Mcauley (2012), as sourced from the "0.edges" file, called as Facebook0 in this paper. We have modified each relationship in this network with a 0.005 probability of flipping, and the figure displays the network both before and after these changes. Observations from the figure reveal significant additions and removals of edges, suggesting that we can protect user information via edge-flipping. For instance, the connectivity (or degree) of node 86 is notably higher after the flipping. However, it is worth noting that the clustering characteristics remain largely unchanged before and after the network modification.

One of the results of Mohamed et al. (2022) proves (1.1), assuming that the input social networks are generated from the SBM. We make much weaker assumptions in our work. The only two assumptions we require are: 1) the social network has a sufficient cluster structure and 2) its maximum degree is sufficiently large. The SBM's considered by Mohamed et al. (2022) also satisfy these assumptions (see Remark 2.11).

We use some ideas from the work of Peng & Yoshida (2020), who have studied the sensitivity of spectral clustering algorithms. However, their work focuses on scenarios where each edge is *removed* with a specific probability. In contrast, local differential privacy not only removes edges but also adds edges to social networks. Furthermore, the number of edges added is often much greater than those removed, especially for sparse networks. Thus, we can only incorporate their concepts in limited sections of our proof, with the core components (such as Section 3.3) being original.

The work detailed in Hehir et al. (2022) demonstrates that stable results from graphs produced by SBM are unattainable with a privacy budget of $o(\log n)$. This suggests that having such a privacy budget for general graphs is also implausible. Because it has been proven in the paper that a constant privacy budget can be achieved for dense, well-clustered graphs generated by SBM, one might anticipate a similar outcome for general graphs. Regrettably, in Section 4 of this paper, we present a dense, well-clustered graph on which spectral clustering results significantly shift when edges are flipped at a probability of $\omega(\log n/n)$. This indicates that even within this regime, securing a smaller privacy budget is not feasible.

Remark 1.2. *For many readers, it may seem counter-intuitive that the privacy budget increases with the number of users, given that differential privacy tends to be more effective with larger databases. This can be explained by considering the nature of the data being protected. In relational databases or general graph differential privacy, there are n pieces of information to protect. However, for local edge differential privacy, the protection extends to $O(n^2)$ bits of edge information.*

Remark 1.3. *Spectral clustering analysis under local differential privacy is a relatively recent area of exploration. However, there is a substantial body of work on graph clustering with differential privacy, as evidenced by studies like Mohamed et al. (2022) and Wang et al. (2013). Notably, a recent study by Chen et al. (2023) provides both upper and lower limits for privacy budgets pertaining to dense graphs generated from the SBM.*

2 Preliminaries

2.1 Notation

Edge-subsets. For the remainder of the paper, we assume that $G = (V, E(G))$ is a graph of n vertices. For any subset $F \subseteq \binom{V}{2}$, we denote by $G \triangle F$ the graph $(V, E(G) \triangle F)$. By $F \sim_p \binom{V}{2}$, we mean a subset F is taken uniformly from $\binom{V}{2}$ with probability p.

Cuts. For a subset $S \subseteq V$ of vertices, we denote by $\overline{S}$ the complement set $V \setminus S$. Further, given two subsets $A, B \subseteq V$ with $A \cap B = \varnothing$, let $e_G(A, B)$ denote the number of edges of G with one endpoint in A and one in B. For any two sets of nodes $S, S' \subseteq V$, $d_{\text{size}}(S, S')$ is given by $d_{\text{size}}(S, S') = \min\left(|S \triangle S'| + |\overline{S} \triangle \overline{S'}|, |S \triangle \overline{S'}| + |\overline{S} \triangle S'|\right)$. As $|S \triangle T| = |\overline{S} \triangle \overline{T}|$, we can equivalently write $d_{\text{size}}(S, S') = 2|S \triangle S'|$. A cut $(S, \overline{S})$ is similar to $(S', \overline{S'})$ if $d_{\text{size}}(S, S')$ is small.

Spectral Graph Theory. Any $n \times n$ real symmetric matrix A has n real eigenvalues. We denote the i-th smallest eigenvalue of A as $\lambda_i(A)$, i.e. $\lambda_1(A) \leq \lambda_2(A) \leq \cdots \leq \lambda_n(A)$. For any graph G, the Laplacian matrix L_G is given by $D_G - A_G$, where D_G is the diagonal degree matrix with $(D_G)_{ii} = \deg_G(i)$ and A_G is the adjacency matrix of G.

In this work, we define the *spectral robustness* of the graph G as $\eta(G) := \frac{\Delta(G)\lambda_2(L_G)}{\lambda_3(L_G)^2}$, where $\Delta(G)$ denotes the maximum degree of any vertex of G.

2.2 Edge Differential Privacy under Randomized Response

The notion of ϵ-edge differential privacy is defined as follows:

Definition 2.1 (ϵ-edge differential privacy; Nissim et al. (2007))**.** *Let G be a social network and let Y be a randomized mechanism that outputs $Y(G)$ from the social network G. For any $\varepsilon > 0$, any possible output of the mechanism Y denoted by y, and any two social networks $G^{(1)} = (V, E^{(1)}(G))$ and $G^{(2)} = (V, E^{(2)}(G))$ that differ by one edge, we say that Y is ε-edge differentially private if $e^{-\varepsilon} \leq \frac{\Pr[Y(G^{(1)})=y]}{\Pr[Y(G^{(2)})=y]} \leq e^{\varepsilon}$. We refer to the value of ε as the privacy budget of Y.*

Intuitively, a lower value of ε results in better privacy protection. In this research, for $0 \leq p \leq 0.5$, we investigate a randomized mechanism Y_p that seeks to generate a result highly similar to spectral clustering outcomes, using randomized response. The mechanism Y_p is defined as $Y_p = \mathcal{SC} \circ \mathcal{F}_p$, where $\mathcal{F}_p$ represents a randomized function that modifies the relationship between each node pair with a probability of p, and $\mathcal{SC}$ is a function for computing spectral clustering. In other words, the randomized mechanism performs spectral clustering on $G \Delta F$, in which $(u, v) \in F$ with a probability of p for every $u, v \in V$. The following theorem is shown in Wang et al. (2016).

Theorem 2.2 (Wang et al. (2016))**.** *The publication Y_p is ε-edge differentially private whenever $\frac{1-p}{p} \leq e^{\varepsilon}$.*

Theorem 2.2 implies that Y_p is ε-edge differential private for $\varepsilon \geq \ln(1-p) - \ln p$. When p is small, we have $\ln(1-p) \approx 0$, and therefore the privacy budget of the publication Y_p is $\Omega(\log 1/p)$.

2.3 Spectral Clustering

For a graph G, the general goal of clustering techniques is to find a good cut $(S, \overline{S})$ such that $e_G(S, \overline{S})$ is small, and most of the edges of G are either concentrated in S or $\overline{S}$. In order to avoid trivial cuts (for example where S comprises of a single vertex), it is customary to instead define the *cut-ratio* $\alpha_G(S) = \frac{e_G(S,\overline{S})}{|S||\overline{S}|}$ and find cuts that minimize $\alpha_G(S)$ (see Wei & Cheng (1989); Hagen & Kahng (1992)). $\alpha(G) = \min_{\varnothing \subsetneq S \subsetneq V} \alpha_G(S)$ is defined as the cut-ratio of G. Unless otherwise specified, we shall denote by S^* the cut that achieves $\alpha_G(S^*) = \alpha(G)$.

Another widely used way of defining the cut-ratio is $\alpha'_G(S) = \frac{e_G(S,\overline{S})}{\min(|S|,|\overline{S}|)}$. This definition is used in Peng & Yoshida (2020), Guattery & Miller (1995), and Kwok et al. (2013). We observe that these two definitions are related:

Lemma 2.3. $\frac{1}{2} \cdot n\alpha_G(S) \leq \alpha'_G(S) \leq n\alpha_G(S)$.

Proof. For the left side of the inequality, note that $\frac{n}{2} \cdot \alpha_G(S) = \frac{|S|+|\overline{S}|}{2} \cdot \frac{e_G(S,\overline{S})}{|S||\overline{S}|} = \frac{1}{2}\left(\frac{e_G(S,\overline{S})}{|S|} + \frac{e_G(S,\overline{S})}{|\overline{S}|}\right) \leq \max\left\{\frac{e_G(S,\overline{S})}{|S|}, \frac{e_G(S,\overline{S})}{|\overline{S}|}\right\} = \alpha'_G(S)$. On the other hand, $\alpha'_G(S) = \alpha_G(S) \cdot \max(|S|, |\overline{S}|) \leq n \cdot \alpha_G(S)$. □

Lemma 2.3 will be useful in converting results formulated using α'_G to those using our cut-ratio α_G.

Now we describe the spectral clustering algorithm. Spectral clustering uses the eigenvalues and eigenvectors of L_G to compute a cut of S. Let us denote by $\mathcal{SC}_2$ the following algorithm that clusters a given graph G into two clusters:

- Compute (or approximate) the second smallest eigenvector $\vec{v} = (v_1, \ldots, v_n)^\intercal$ of L_G, and reorder the vertices of G such that $v_1 \leq \cdots \leq v_n$.
- Return the cut $(S, \overline{S})$, where $S = \{v_1, \ldots, v_{i_0}\}$ and $i_0 = \arg\min_{1 \leq i \leq n} \alpha_G(v_1, \ldots, v_i)$.

The cut-ratio of G can be quantified very precisely via the famous Cheeger's inequality.

Lemma 2.4 (Cheeger's Inequality; Cheeger (1971); Alon (1986))**.** $\lambda_2(L_G) \leq n\alpha(G) \leq \sqrt{8\Delta(G)\lambda_2(L_G)}$.

We shall also use the following improvement of Lemma 2.4:

Lemma 2.5 (Improved Cheeger Inequality; Kwok et al. (2013)). *Let $\mathcal{SC}_2(G)$ denote the cut given by the spectral clustering algorithm. Then, $\alpha_G(\mathcal{SC}_2(G)) \leq O\left(\frac{\lambda_2(L_G)\Delta(G)^{1/2}}{n\lambda_3(L_G)^{1/2}}\right)$.*

Lemma 2.4 and 2.5 give us a way of quantifying the quality of the cut output by $\mathcal{SC}_2$ in terms of the cut-ratio of G. Indeed, as $\frac{\lambda_2(L_G)}{n} \leq \alpha(G)$,

$$\alpha_G(\mathcal{SC}_2(G)) \leq O\left(\frac{\Delta(G)^{1/2}}{\lambda_3(L_G)^{1/2}}\right) \cdot \alpha(G). \tag{2.1}$$

Let S^* be the cut of G with the smallest cut-ratio. While equation (2.1) can be interpreted as a measure of how close $\mathcal{SC}_2(G)$ is with S^*, we shall need stability results from Peng & Yoshida (2020); Kwok et al. (2013) to precisely bound $d_{\mathsf{size}}(\mathcal{SC}_2(G), S^*)$.

Lemma 2.6 (Stability of min-cut). *Let $G = (V, E)$ be any graph with optimal min-cut S^*. Then, for any $\rho \geq 1$, if $S \subseteq V$ satisfies $\alpha_G(S) \leq \rho \cdot \alpha_G(S^*)$, then $d_{size}(S, S^*) \leq O\left(\frac{\lambda_2(L_G)\Delta(G)^{1/2}}{\lambda_3(L_G)^{3/2}} \cdot \rho\right) \cdot n$.*

Proof. (sketch) Observe that by Lemma 2.3, $\alpha_G(S) \leq \rho \cdot \alpha_G(S^*)$ implies $\alpha'_G(S) \leq 2\rho \cdot \alpha'_G(S^*)$. This lemma then follows from a direct application of Lemma 3.5 of Peng & Yoshida (2020). □

2.4 Concentration Inequalities

We also require some concentration inequalities for random variables, which we present here.

Lemma 2.7 (Hoeffding's inequality; Hoeffding (1963)). *Let $X_1, \ldots, X_n$ be independent random variables such that $a_i \leq X_i \leq b_i$ almost surely. If $S = X_1 + \cdots + X_n$, then we have $\Pr[S \leq \mathbb{E}(S) - t] \leq \exp\left(-2t^2/\sum_{i=1}^n (b_i - a_i)^2\right)$.*

Lemma 2.8 (Chernoff bound; Mitzenmacher & Upfal (2017)). *For a binomial random variable X with mean μ and $t > 0$, we have $\Pr[X \geq \mu + t] \leq \exp\left(-\frac{t^2}{2\mu + t}\right)$.*

Lemma 2.9 (Weyl's Inequality; Weyl (1912)). *For any real symmetric matrices M and H $|\lambda_i(M + H) - \lambda_i(M)| \leq \|H\|_2$, where $\|H\|_2$ denotes the spectral norm of H.*

2.5 Assumptions

In order to demonstrate the robustness of spectral clustering, we require assumptions on the social network G and the probability p of edge flipping. Recall that $F \sim_p \binom{V}{2}$ is the set of vertex pairs to be flipped.

Assumption 2.10. *We assume the following:*

1. $p < \log n/10n$,
2. *(a)* $\Delta(G) \geq 10 \log n \lambda_3(L_G)$, *(b)* $\lambda_2(L_G) \geq 1/10$, *(c)* $\eta(G) := \frac{\lambda_2(L_G)\Delta(G)}{\lambda_3(L_G)^2}$ *is small,*
   (d) $\lambda_3(L_G) \geq 10 \log n$,
3. *Let the minimum cuts of G and $G \triangle F$ be $(S^*, \overline{S^*})$ and $(S^*_F, \overline{S^*_F})$, respectively. Then each of $|S^*|, |\overline{S^*}|, |S^*_F|, |\overline{S^*_F}|$ have size at least $n/10$.*

Plausibility of Assumption 2.10:

The first assumption can be justified by our discussion in Section 2.2, where we observe that privacy can be maintained as long as p is $\Omega(1/n)$. We further note that, if G is a sparse social network with $O(n)$ edges and $p \gg \log n/n$, then as $\mathbb{E}(|F|) = \Omega(n \log n)$, $G \triangle F$ will have too much noise, and would become close to the Erdős-Rényi random graph $F \sim \mathcal{G}(n, p)$. Spectral algorithms cannot perform well for these graphs. For example, it is shown in Chung & Radcliffe (2011) that the eigenvalues of the normalized Laplacian $\mathcal{L}_F$ are close to those of the expected values. A quick calculation shows that the second and third eigenvalues of $\mathbb{E}(\mathcal{L}_F)$ are both equal (and close to 1), implying the inefficiency of spectral clustering algorithms on $\mathcal{G}(n, p)$ for p asymptotically larger than $\log n/n$.

On the other hand, one may think that values of p larger than $\log n/n$, for example $p = \Omega(1)$ is achievable by the edge flipping mechanism if the input graph G is dense. However, there are two issues with this: firstly, social networks are not dense in practice. Secondly, we demonstrate in Section 4, a well-clustered dense graph, whose sparsest cut changes drastically when introducing noise $p = \omega(\log n/n)$.

The second assumption derives from usual properties of social networks. Recall that we have $0 = \lambda_1(L_G) \leq \lambda_2(L_G) \leq \cdots \leq \lambda_n(L_G) < 2\Delta(G)$. This assumption asserts that there are big gaps between $\lambda_2(L_G)$, $\lambda_3(L_G)$ and $\Delta(G)$. First, we note that most social networks that we encounter in practice, have super-nodes (nodes of degree $\Omega(n)$), justifying our assumption (a).

Assumption (b) ensures that G is well-connected: note that disconnected graphs have $\lambda_2(L_G) = 0$ and graphs that have small edge-separators have a small $\lambda_2(L_G)$. We also make a note here that we can relax this assumption to any constant threshold $\lambda_2(L_G) \geq \Theta(1)$, by changing the value of γ_0 used in Section 3.3. Conversely, the counterexample provided in Section 4 has connectivity of $O(1/n)$, implying that for the stability of $\mathcal{SC}_2$, the algebraic connectivity of G cannot be too small.

Finally, (c) ensures that there is a gap between $\lambda_3(L_G)$ and $\lambda_2(L_G)$, which ensures that the graph has a good bi-cluster structure, which lets $\mathcal{SC}_2$ find good clusters in G.

Observe that using inequalities (a), (b) and (c), we can deduce that $\lambda_3(L_G) = \frac{\lambda_2(L_G)\Delta(G)}{\lambda_3(L_G)\eta(G)} \geq \frac{\log n}{\eta(G)}$, which implies our assumption of (d).

Our final assumption stems from the fact that usually social networks admit linearly sized clusters, and also we are usually interested in detecting clusters of larger size via the definition of the cut ratio $\alpha(G)$, for example. Moreover, our analysis in Equation 3.6 suggests that $|S_F^*|$ and $|\overline{S}_F^*|$ are close to the output $|S_F|$ and $|\overline{S}_F|$ of $\mathcal{SC}_2$, which is designed to output balanced cuts. This justifies our assumption of $|S_F^*|, |\overline{S}_F^*| \geq 0.1n$.

Remark 2.11. *When dealing with SBM's with probabilities $p = a \log n/n$ inside each cluster and $q = b \log n/n$ between, the work of Deng et al. (2021) proves that $\lambda_2(L_G) \leq O(b \log n)$ and $\lambda_3(L_G) \geq b \log n$. Therefore, these SBM's satisfy Assumption 2.10, and the result for Randomized Response in Mohamed et al. (2022) follows from our work.*

3 Main Theorem

We restate and prove a formal version of Theorem 1.1 in this section.

Theorem 3.1. *Let $G = (V, E)$ be a graph and p satisfy Assumption 2.10. Let $F \sim_p \binom{V}{2}$. Then, with probability at least $1 - 5n^{-8/5}$,*

$$d_{size}\left(\mathcal{SC}_2(G), \mathcal{SC}_2(G \triangle F)\right) = O(\eta(G) \cdot n).$$

Proof Structure. Suppose S^* and S_F^* are the optimum min-cuts of G and $G \triangle F$. Denote by S and S_F the outputs of $\mathcal{SC}_2$ on G and $G \triangle F$, respectively.

The key idea is to bound $d_{\text{size}}(S, S_F)$ using triangle inequality:

$$d_{\text{size}}(S, S_F) \leq d_{\text{size}}(S, S^*) + d_{\text{size}}(S^*, S_F^*) + d_{\text{size}}(S_F^*, S_F). \tag{3.1}$$

We bound each of the terms in their own subsection below. Observe that by Equations (3.2), (3.6) and (3.8), we obtain

$$d_{\text{size}}(S, S_F) \leq O\left(\frac{\lambda_2(L_G)\Delta(G)}{\lambda_3(L_G)^2}\right) \cdot n = O(\eta(G) \cdot n)$$

with probability at least $1 - 4n^{-21/11} - n^{-8/5} \geq 1 - 5n^{-8/5}$, completing the proof. In the remainder of this section, we bound each term appearing in the right side of Equation (3.1).

3.1 The term $d_{\mathsf{size}}(S, S^*)$.

An upper bound on this term is a direct corollary of Cheeger's inequality and stability: observe that Lemma 2.6 and (2.1) give us

$$d_{\text{size}}(S, S^*) \leq O\left(\frac{\lambda_2(L_G)\Delta(G)^{1/2}}{\lambda_3(L_G)^{3/2}} \cdot \frac{\Delta(G)^{1/2}}{\lambda_3(L_G)^{1/2}}\right) \cdot \ n = O\left(\frac{\lambda_2(L_G)\Delta(G)}{\lambda_3(L_G)^2}\right) \cdot n \tag{3.2}$$

3.2 The term $d_{\mathsf{size}}(S_F^*, S_F)$.

First, we describe a lemma to compare the eigenvalues and maximum degrees of $G \triangle F$ and G.

Lemma 3.2. *Let G have n vertices, and $F \sim \mathcal{G}(n, p)$. Under Assumption 2.10, with probability at least $1 - 3n^{-21/11}$, all of the following hold:*

(a) $\lambda_2(L_{G\triangle F}) \leq \lambda_2(L_G)$, *(b)* $\lambda_3(L_{G\triangle F}) \geq \lambda_3(L_G)/10$, *(c)* $\Delta(G\triangle F) \leq 2\Delta(G)$.

Proof. Part (a). By monotonicity of λ_2, $\lambda_2(L_{G\triangle F}) \leq \lambda_2(L_{G\cup F})$. As $\lambda_2(L_{G\cup F}) \leq \lambda_2(L_G) + \lambda_2(L_{F\setminus G})$, and $p < \log n/10n$, F (and hence $F\setminus G$) is almost surely disconnected Erdős et al. (1960), implying $\lambda_2(L_{F\setminus G}) = 0$. Hence, we have $\lambda_2(L_{G\triangle F}) \leq \lambda_2(L_{G\cup F}) \leq \lambda_2(L_G)$. ■

Part (b). For this part, we shall use Weyl's Inequality as follows: suppose $F_1 = F \setminus G$ and $F_2 = G \cap F$ be subgraphs of F on the vertex set $V(G)$. By additivity of the Laplacian, $L_{G\triangle F} - L_G = L_{F_1} - L_{F_2}$. Now as $\|A\|_2 = \max_{x\in\mathbb{R}^n} x^\intercal Ax$ for any symmetric $n \times n$ matrix A, which implies

$$\|L_{G\triangle F} - L_G\|_2 = \max_{x\in\mathbb{R}^n} |x^\intercal L_{F_1}x - x^\intercal L_{F_2}x| \leq \max_{x\in\mathbb{R}^n} x^\intercal L_F x = \lambda_n(L_F) \leq 2\Delta(F).$$

By the union bound, note that for any $v \in V(G)$,

$$\Pr[\Delta(F) > \frac{9}{2}\log n] \leq n \cdot \Pr[\deg_F(v) > \frac{9}{2}\log n] \leq n \cdot \Pr\left[\deg_F(v) - p(n-1) > 4\log n\right]. \tag{3.3}$$

Using the Chernoff bound, the probability in (3.3) is at most

$$n \cdot \exp\left(-\frac{16(\log n)^2}{2(n-1)p + 4\log n}\right) < n \cdot \exp\left(-\frac{16(\log n)^2}{\frac{11}{2}\log n}\right) = n \cdot \exp\left(-\frac{32}{11}\log n\right) = n^{-21/11}, \tag{3.4}$$

Thus $\|L_{G\triangle F} - L_G\|_2 \leq 9\log n$ holds with probability at least $1 - n^{-21/11}$. By Weyl's inequality and Assumption 2.10(2),

$$\lambda_3(L_G) - \lambda_3(L_{G\triangle F}) \leq 9\log n \leq \frac{9}{10}\lambda_3(L_G),$$

finishing the proof of (b). ■

Part (c). Observe that for every vertex $v \in V(G)$, we have

$$\deg_{G\triangle F}(v) - \deg_G(v) \leq \deg_F(v) \leq \Delta(F).$$

Hence,

$$\Pr\left[\deg_{G\triangle F}(v) > \deg_G(v) + \Delta(G)\right] \leq \Pr\left[\Delta(F) > \Delta(G)\right] \leq \Pr\left[\Delta(F) > 10\log n\right].$$

By a similar calculation to (3.3) and (3.4), we conclude that $\deg_{G\triangle F}(v) > \deg_G(v) + \Delta(G)$ holds with probability at most n^{-4}. Again, by the union bound, with probability at least $1 - n^{-3}$, we have

$$\deg_{G\triangle F}(v) \leq \deg_G(v) + \Delta(G) \text{ for all } v \in V(G). \tag{3.5}$$

Taking the maximum of (3.5) over all v, we see that (c) holds with probability at least $1 - n^{-3}$, which is greater than $1 - n^{-21/11}$. ■

As the assertions of (a), (b), (c) each hold with probability at least $1 - n^{-21/11}$, all of them simultaneously hold with probability at least $1 - 3n^{-21/11}$, completing our proof of Lemma 3.2. □

Now, observe that by the same argument as (3.2) in addition with Lemma 3.2, we get that with probability at least $1 - 3n^{-21/11}$,

$$d_{\text{size}}(S_F^*, S_F) \leq O\left(\frac{\lambda_2(L_{G\triangle F})\Delta(G\triangle F)}{\lambda_3(L_{G\triangle F})^2}\right) \cdot n = O\left(\frac{\lambda_2(L_G)\Delta(G)}{\lambda_3(L_G)^2}\right) \cdot n \tag{3.6}$$

3.3 The term $d_{\text{size}}(S^*, S_F^*)$.

For the remainder of this section, let γ_0 be given by

$$\gamma_0 := 200\sqrt{\Delta(G)/\lambda_3(L_G)} > 200\sqrt{10\log n}.$$

In order to bound $d_{\text{size}}(S^*, S_F^*)$, we require the following rather technical lemma.

Lemma 3.3. *Let S^* denote the minimum cut of G and S_F^* denote the minimum cut of $G\triangle F$. Suppose $n/2 \geq |S^*|, |S_F^*| \geq \epsilon n$ for some $1/2 > \epsilon > 0$. Further, suppose $\alpha_G(S_F^*) \geq \gamma_0\alpha_G(S^*)$. Then,*

$$\Pr\left(\gamma_0\alpha_{G\triangle F}(S_F^*) - \alpha_{G\triangle F}(S^*) < 0\right) < \exp\left(-\frac{4\left(\gamma_0^2 - 1\right)^2}{25\gamma_0^2} \cdot \alpha_G(S^*)^2\epsilon^2 n^2\right). \tag{3.7}$$

As the proof is involved, we defer it to the end of this section.

First, we demonstrate the bound on $d_{\text{size}}(S^*, S_F^*)$ using Lemma 3.3. We consider two cases:

- **Case 1.** $\alpha_G(S_F^*) \leq \gamma_0\alpha_G(S^*)$: In this case, Lemma 2.6 directly gives us $d_{\text{size}}(S^*, S_F^*) \leq O\left(\frac{\gamma_0\lambda_2(L_G)\Delta(G)^{1/2}}{\lambda_3(L_G)^{3/2}}\right) \cdot n = O\left(\frac{\lambda_2(L_G)\Delta(G)}{\lambda_3(L_G)^2}\right) \cdot n$.
- **Case 2.** $\alpha_G(S_F^*) > \gamma_0\alpha_G(S^*)$: In this case, setting $\epsilon = 1/10$ in Lemma 3.3, we note that the probability that $\alpha_{G\triangle F}(S^*) > \gamma_0\alpha_{G\triangle F}(S_F^*)$ is at most:

$$\begin{aligned}\exp\left(-\frac{(2\gamma_0^2 - 2)^2\alpha_G(S^*)^2 n^2}{2500\gamma_0^2}\right) < \exp\left(-\frac{\gamma_0^2}{2500} \cdot (\alpha_G(S^*) \cdot n)^2\right) &< \exp\left(-160\log n \cdot \lambda_2(L_G)^2\right) \\ &< \exp\left(-160\log n \cdot \frac{1}{100}\right) = n^{-8/5}\end{aligned}$$

  The last line follows from Assumption 2.10(2). Hence, with probability at least $1 - n^{-8/5}$, $\alpha_{G\triangle F}(S^*) \leq \gamma_0\alpha_{G\triangle F}(S_F^*)$ holds. By Lemma 2.6, this implies $d_{\text{size}}(S^*, S_F^*) \leq O\left(\frac{\gamma_0\lambda_2(L_{G\triangle F})\Delta(G\triangle F)^{1/2}}{\lambda_3(L_{G\triangle F})^{3/2}}\right) \cdot n$. Together with Lemma 3.2, we obtain that with probability at least $1 - n^{-8/5} - 3n^{-21/11}$,

$$d_{\text{size}}(S^*, S_F^*) \leq O\left(\frac{\Delta(G)^{1/2}}{\lambda_3(L_G)^{1/2}} \cdot \frac{\lambda_2(L_{G\triangle F})\Delta(G\triangle F)^{1/2}}{\lambda_3(L_{G\triangle F})^{3/2}}\right) \cdot n = O\left(\frac{\lambda_2(L_G)\Delta(G)}{\lambda_3(L_G)^2}\right) \cdot n, \tag{3.8}$$

  finishing our upper bound on $d_{\text{size}}(S^*, S_F^*)$. ■

We now present our proof of Lemma 3.3.

Proof. (Lemma 3.3). The main idea behind the proof is as follows: first, we show that Lemma 3.3 holds with S_F^* replaced with any fixed subset A. Then, we use the fact that

$$\begin{aligned}\Pr\left(\gamma_0\alpha_{G\triangle F}(S_F^*) < \alpha_{G\triangle F}(S^*)\right) &= \Pr\left(\gamma_0\alpha_{G\triangle F}(S_F^*) < \alpha_{G\triangle F}(S^*) \mid \alpha_G(S_F^*) > \gamma_0\alpha_G(S^*)\right) \\ &= \sum_{A:\alpha_G(A) > \gamma_0\alpha_G(S^*)} \Pr(S_F^* = A) \cdot \Pr\left(\gamma_0\alpha_{G\triangle F}(S_F^*) < \alpha_{G\triangle F}(S^*) \mid S_F^* = A\right) \\ &\leq \max_{A:\alpha_G(A) > \gamma_0\alpha_G(S^*)} \Pr\left(\gamma_0\alpha_{G\triangle F}(A) < \alpha_{G\triangle F}(S^*)\right),\end{aligned} \tag{3.9}$$

as $\sum_A \Pr(S_F^* = A) = 1$.

Now, we bound $\Pr\left(\gamma_0 \alpha_{G\triangle F}(A) < \alpha_{G\triangle F}(S^*)\right)$ for any fixed A.

Claim 3.4. *Let S^* denote the minimum cut of G. Suppose $\frac{n}{2} \geq |S^*| \geq \epsilon n$ for some $\frac{1}{2} > \epsilon > 0$. Then, for any $\gamma > 1$ and $\frac{n}{2} \geq |A| \geq \epsilon n$,*

$$\Pr\left(\gamma \alpha_{G\triangle F}(A) - \alpha_{G\triangle F}(S^*) < 0\right) < \exp\left(-\frac{4\left(\gamma\alpha_G(A) - \alpha_G(S^*)\right)^2}{25\gamma^2} \cdot \epsilon^2 n^2\right). \tag{3.10}$$

Proof of Claim 3.4. Let $Y_A := \gamma\alpha_{G\triangle F}(A) - \alpha_{G\triangle F}(S^*)$. We wish to show that $Y_A \geq 0$ with high probability.

For any tuple $(x, y) \in V \times V$, define $X_{(x,y)}$ as the boolean random variable such that $X_{(x,y)} = 1$ if $xy \in E(G\triangle F)$ and $X_{(x,y)} = 0$ otherwise. As $X_{(x,y)} = X_{(y,x)}$, we abuse notation and write X_{xy} as a shorthand for both these variables. Note that X_{xy} are all mutually independent, and

$$\mathbb{E}(X_{(x,y)}) = \Pr(xy \in E(G\triangle F)) = \begin{cases} p, & \text{if } e \notin E(G), \\ 1-p, & \text{if } e \in E(G). \end{cases} \tag{3.11}$$

Further, for any subset $A \subseteq V$, by definition

$$\alpha_{G\triangle F}(A) = \frac{e_{G\triangle F}(A, \overline{A})}{|A||\overline{A}|} = \frac{1}{|A||\overline{A}|} \cdot \sum_{(x,y)\in A\times\overline{A}} X_{(x,y)},$$

Which, by (3.11), implies

$$\begin{aligned} \mathbb{E}(\alpha_{G\triangle F}(A)) &= \frac{1}{|A||\overline{A}|} \cdot \left(\sum_{e\in E_G(A,\overline{A})} \mathbb{E}(X_e) + \sum_{e\in A\times\overline{A}\setminus E_G(A,\overline{A})} \mathbb{E}(X_e)\right) \\ &= \frac{1}{|A||\overline{A}|} \cdot \left(e_G(A,\overline{A}) \cdot (1-p) + |A||\overline{A}| \cdot p - e_G(A,\overline{A}) \cdot p\right) \\ &= (1-2p) \cdot \alpha_G(A) + p. \end{aligned} \tag{3.12}$$

Let μ denote the expectation of Y_A. By linearity and (3.12),

$$\mu = \mathbb{E}(Y_A) = (1-2p) \cdot \left(\gamma\alpha_G(A) - \alpha_G(S^*)\right) + p \cdot (\gamma - 1) > \frac{4}{5} \cdot \left(\gamma\alpha_G(A) - \alpha_G(S^*)\right), \tag{3.13}$$

As $\gamma > 1$ and $p < 1/10$. We also have $\mu > 0$, and $\Pr(Y_A < 0) = \Pr(Y_A - \mu < -\mu)$. Now we shall use Hoeffding's inequality to provide an upper bound on $\Pr(Y_A < 0)$. To that end, Y_A has to be rewritten as a sum of independent random variables. However,

$$Y_A = \frac{\gamma}{|A||\overline{A}|} \cdot \sum_{e\in A\times\overline{A}} X_e - \frac{1}{|S^*||\overline{S^*}|} \cdot \sum_{e\in S^*\times\overline{S^*}} X_e. \tag{3.14}$$

As the two summations in Y_A have overlapping terms, we separate them as follows. Let $Z_1 = S^* \setminus A$, $Z_2 = S^* \cap A$, $Z_3 = A \setminus S^*$, $Z_4 = \overline{S^* \cup A}$. Observe then,

$$\begin{aligned} S^* \times \overline{S^*} &= (Z_1 \times Z_3) \sqcup (Z_1 \times Z_4) \sqcup (Z_2 \times Z_3) \sqcup (Z_2 \times Z_4) \\ A \times \overline{A} &= (Z_3 \times Z_1) \sqcup (Z_3 \times Z_4) \sqcup (Z_2 \times Z_1) \sqcup (Z_2 \times Z_4) \end{aligned} \tag{3.15}$$

This lets us break each sum in (3.14) into four parts, and using $X_{(x,y)} = X_{(y,x)}$, we can write Y as

$$Y_A = \sum_{e\in(Z_1\times Z_3)\sqcup(Z_2\times Z_4)} \left(\frac{\gamma}{|A||\overline{A}|} - \frac{1}{|S^*||\overline{S^*}|}\right) X_e + \sum_{e\in(Z_1\times Z_4)\sqcup(Z_2\times Z_3)} \frac{\gamma X_e}{|A||\overline{A}|} - \sum_{e\in(Z_3\times Z_4)\sqcup(Z_1\times Z_2)} \frac{X_e}{|S^*||\overline{S^*}|}. \tag{3.16}$$

Note that all summands in (3.16) are independent of each other. For simplicity, let us denote $z_i := |Z_i|$ for $i = 1, \dots, 4$.

Since $-|c| \le cX_e \le |c|$ for any constant $c \in \mathbb{R}$, we can use Hoeffding's inequality to get $\Pr(Y < 0) = \Pr(Y_A - \mu < -\mu) \le \exp(-\frac{2\mu^2}{D})$, where $D = 4(z_1z_3 + z_2z_4)\left(\frac{\gamma}{(z_2+z_3)(z_1+z_4)} - \frac{1}{(z_1+z_2)(z_3+z_4)}\right)^2 + \frac{4\gamma^2(z_1z_4+z_2z_3)}{(z_2+z_3)^2(z_1+z_4)^2} + \frac{4(z_3z_4+z_1z_2)}{(z_1+z_2)^2(z_3+z_4)^2}$. After some calculations, this leads to

$$D = \frac{4\gamma^2(z_1+z_2)(z_3+z_4)}{(z_2+z_3)^2(z_1+z_4)^2} + \frac{4(z_2+z_3)(z_1+z_4)}{(z_1+z_2)^2(z_3+z_4)^2} - \frac{8\gamma(z_1z_3+z_2z_4)}{(z_1+z_2)(z_3+z_4)(z_2+z_3)(z_1+z_4)} \tag{3.17}$$

$$< 4\gamma^2\left(\frac{(z_1+z_2)(z_3+z_4)}{(z_2+z_3)^2(z_1+z_4)^2} + \frac{(z_2+z_3)(z_1+z_4)}{(z_1+z_2)^2(z_3+z_4)^2}\right) \tag{3.18}$$

$$= 4\gamma^2\left(\frac{|S^*||\overline{S^*}|}{|A|^2|\overline{A}|^2} + \frac{|A||\overline{A}|}{|S^*|^2|\overline{S^*}|^2}\right) \le 4\gamma^2 \cdot 2 \cdot \frac{n^2/4}{\epsilon^2 n^4/4} = \frac{8\gamma^2}{\epsilon^2 n^2}. \tag{3.19}$$

Here (3.19) follows from the fact that $n^2/4 \ge |S^*||\overline{S^*}|, |A||\overline{A}| \ge \epsilon n \cdot n/2$. Therefore, in conjunction with (3.13), we obtain $\Pr(Y_A < 0) \le \exp\left(-\frac{2\mu^2}{D}\right) < \exp\left(-\frac{4(\gamma\alpha_G(A)-\alpha_G(S^*))^2}{25\gamma^2} \cdot \epsilon^2 n^2\right)$, as desired. ■

Now we return to our proof of Lemma 3.3. For any set $A \subseteq V$ with $\frac{n}{2} \ge |A| \ge \epsilon n$ and $\alpha_G(A) > \gamma_0\alpha_G(S^*)$, we have $\Pr(\gamma_0\alpha_{G\triangle F}(A) < \alpha_{G\triangle F}(S^*)) < \exp\left(-\frac{4(\gamma_0\alpha_G(A)-\alpha_G(S^*))^2}{25\gamma_0^2} \cdot \epsilon^2 n^2\right) \le \exp\left(-\frac{4(\gamma_0^2-1)^2}{25\gamma_0^2} \cdot \alpha_G(S^*)^2\epsilon^2 n^2\right)$. When we plug this back into Equation (3.9), it gives our desired bound.

□

4 Instability of spectral clustering when $p = \omega(\log n/n)$

We now construct a dense graph G whose sparsest cut drastically changes under edge flipping with $p = \omega(\log n/n)$.

Let $\delta > 0$ be a small constant, and $p = \omega(\log n/n)$. Consider a graph G on $(1+\delta)n$ vertices with vertex set $A \cup B \cup C$, where $|A| = \delta n$, $|B| = |C| = n/2$. Add all $\binom{|A|}{2}$ edges in A and $\binom{|C|}{2}$ edges in C. Finally, add B–B and B–C edges each with probability $\log n/n$, and A–B and A–C edges each with probability $1/10n$. A visual representation of this construction is shown in Figure 4.1.

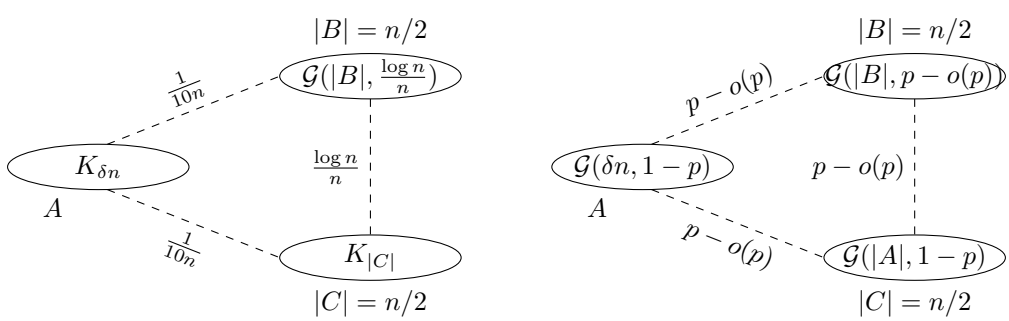


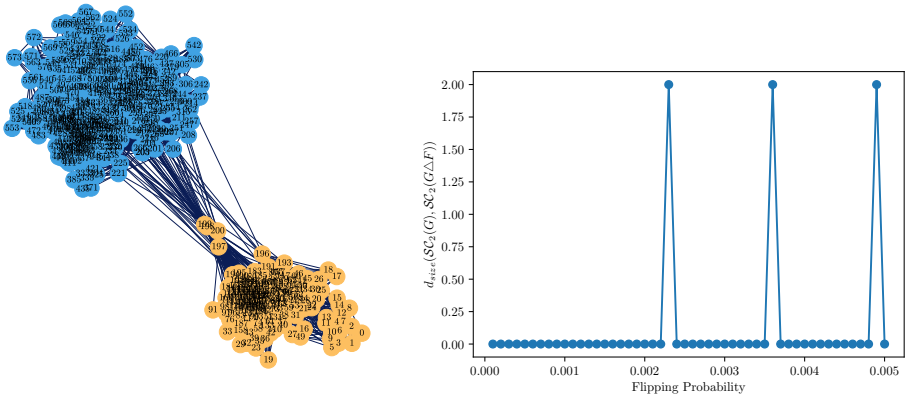
Figure 4.1: The (dense) graph G (left) and the graph $G \triangle F$ where $F \sim_p \binom{V}{2}$ (right). Dashed lines represent probabilistic edges between the parts A, B and C.

It can be seen that for G, the cut $(A, B \cup C)$ is the sparsest with high probability, as $\alpha_G(A) \approx \frac{1}{10n}$. However, if G' is the graph obtained from G after edge flipping with probability $p = \omega(\log n/n)$, then in G', the sets A and C only become slightly less dense, and every A–B, B–B, B–C and A–C edge exists with probability $p - o(p)$. Hence, while A and C would be on different parts of the sparsest cut, any cut $(A \cup B', B - B' \cup C)$ with $B' \subseteq B$ would attain the minimum cut-ratio of p, and spectral clustering will choose a cut different from $(A, B \cup C)$ since it would be unbalanced due to small δ. In particular, this implies the instability of spectral clustering on G', and leads to large $d_{\text{size}}(\mathcal{SC}_2(G), \mathcal{SC}_2(G'))$ with high probability.

5 Experiments

We conduct experiments on real social networks to verify our theoretical results. In this work, we mainly use the network called "Social circles: Facebook" obtained from the Stanford network analysis project (SNAP), detailed in Leskovec & Mcauley (2012). We found that spectral clustering cannot find plausible results for some of those networks, due to the fact that there are many connected components, and also small sets of nodes with only one to two edges to the rest of the graph. Those small sets usually form a cluster in the outcomes of spectral clustering, which makes the outcomes undesirable. We therefore decided to eliminate all small node sets that have at most 10 outgoing edges.

We examine the graphs defined in the files "0.edges" and "1684.edges." We call the graphs as Facebook0 and Facebook1684. After removing nodes of small degree, there are $n = 120$ left in the first graph and $n = 574$ left in the second. As illustrated in Figure 1.1, the social network Facebook0 is composed of two clusters, both of which are quite sizable. Alternatively, the social network Facebook1684 is divided into three clusters. The first is a small yellow cluster including nodes 0–15, followed by a bigger cluster of yellow nodes, and a dense cluster of blue nodes. For ease of reference, these clusters will be subsequently named Cluster A, Cluster B, and Cluster C, respectively. This graph possesses the attributes necessary for Assumption 2.10.



(a) (b)

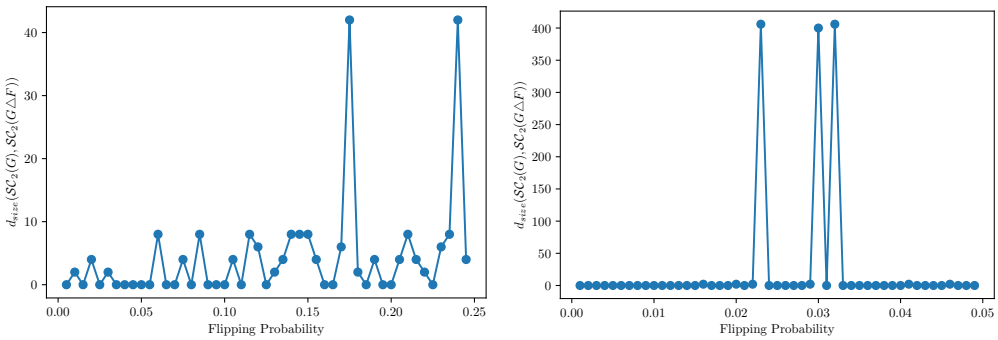
Figure 5.1: (a): The social network Facebook1684 obtained from SNAP after pruning. Each node was assigned a color based on the spectral clustering outcomes. (b): We generated 100 graphs from the first graph Facebook0 (Figure 1.1), and plotted the worst discrepancy d_{size} between the outputs of the spectral clustering of the original and perturbed graphs for these 100 random runs.

Our main theorem ensures that the clustering outcomes remain mostly consistent when edges are flipped with a probability $p < \frac{\log n}{10n}$. The upper bound is about 0.004 for the first graph and about 0.001 in the second. We examine $p \in \{0.0001q : 1 \leq q \leq 50\}$. For each probability p and graph, we create 100 random graphs F with the given probability. Note that the original graph is represented by G. We then compute the difference between the clustering results of G (represented by $\mathcal{SC}_2(G)$) and that of $G \triangle F$ (represented by $\mathcal{SC}_2(G \triangle F)$).

The chart in Figure 5.1b shows the result we obtain from the first graph. The chart demonstrates the difference between the clustering outputs, represented as $d_{\text{size}}(\mathcal{SC}_2(G), \mathcal{SC}_2(G \triangle F))$. The values are the worst value derived from the 100 random graphs for each probability. This illustration reveals that, across all considered probabilities, the clustering outcomes remain consistent in every random graph. In each instance, when comparing the original graph to the graph with flipped edges, a minimum of 116 nodes are assigned to the same clusters. Only a maximum of four nodes out of 120 experience a change in their cluster placement.

For the second graph, the result is even more robust. For all the probabilities we have conducted the experiment, there were no change in the clustering results by the edge flipping. These two experiments suggest that the clustering results exhibit strong resilience to edge flipping.

5.1 Results on Larger Flipping Probability



(a) (b)

Figure 5.2: The robustness results of the social networks upon the introduction of a flipping probability that exceeds the value specified in Assumption 2.10.

In Figure 5.2, the flipping probability is raised above the level outlined in Assumption 2.10. Our experiments with the network derived from the Facebook0 network demonstrate that clustering outcomes remain stable as long as the flipping probability does not exceed 0.15.

In contrast, with the network derived from Facebook1684, the stability of the results is preserved only when the flipping probability remains under 0.04. Utilizing spectral clustering on the original graph divides it into two segments: one combining clusters A and B, and another comprising cluster C. However, exceeding a flipping probability of 0.04 occasionally alters the spectral clustering outcome to one group consisting of cluster A and another combining clusters B and C. This variation seems reasonable, as the latter grouping also yields a low conductance.

Section 4 demonstrates that in certain networks, clustering outcomes are unstable when the probability exceeds the level specified in Assumption 2.10. However, our experiments indicate that this threshold may be higher for particular graph types. In future work, we plan to develop theoretical results for these specific graphs.

6 Concluding Remarks

In this manuscript, we demonstrate and empirically verify that under some assumptions, the spectral clustering algorithm is robust under the randomized response method. While our primary objective is its use in local differential privacy, our validation also confirms the robustness of spectral clustering against social networks containing inaccurate adjacency information. We demonstrate that the outcomes are robust when $p < \log n/10n$, but also acknowledge that the results can undergo significant alterations for larger p values. This occurs because randomized response introduces an excessive number of edges to the graph in such cases. We are aiming to examine the robustness of other local differential privacy approaches (e.g., as in Adhikari et al. (2020)) that do not add as many edges as the randomized response method.

Although in Peng & Yoshida (2020), there are results for spectral clustering with k clusters, we cannot use ideas from those results in this work. Indeed, because we also consider edge addition, we have to demonstrate many additional theoretical results including Lemma 2.3 and Lemma 3.3. These analyses cannot be directly extended to the case when $k > 2$. We, anyway, believe that such an extension would be interesting future work.

References

Maneesh Babu Adhikari, Vorapong Suppakitpaisarn, Arinjita Paul, and C Pandu Rangan. Two-stage framework for accurate and differentially private network information publication. In *CSoNet 2020*, pp. 267–279. Springer, 2020.

Noga Alon. Eigenvalues and expanders. *Combinatorica*, 6(2):83–96, 1986.

Apple's Differential Privacy Team. Learning with privacy at scale. *Apple Machine Learning Research*, 2017.

Jeff Cheeger. *A Lower Bound for the Smallest Eigenvalue of the Laplacian*, pp. 195–200. Princeton University Press, Princeton, 1971. ISBN 978-1-4008-6931-2. doi: doi:10.1515/9781400869312-013. URL `https://doi.org/10.1515/9781400869312-013`.

Hongjie Chen, Vincent Cohen-Addad, Tommaso d'Orsi, Alessandro Epasto, Jacob Imola, David Steurer, and Stefan Tiegel. Private estimation algorithms for stochastic block models and mixture models. *arXiv preprint arXiv:2301.04822*, 2023.

Fan Chung and Mary Radcliffe. On the spectra of general random graphs. *The Electronic Journal of Combinatorics*, pp. P215–P215, 2011.

Shaofeng Deng, Shuyang Ling, and Thomas Strohmer. Strong consistency, graph laplacians, and the stochastic block model. *The Journal of Machine Learning Research*, 22(1):5210–5253, 2021.

Cynthia Dwork. Differential privacy: A survey of results. In *Theory and Applications of Models of Computation: 5th International Conference, TAMC 2008, Xi'an, China, April 25-29, 2008. Proceedings 5*, pp. 1–19. Springer, 2008.

Paul Erdős, Alfréd Rényi, et al. On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci*, 5 (1):17–60, 1960.

Úlfar Erlingsson, Vasyl Pihur, and Aleksandra Korolova. RAPPOR: Randomized aggregatable privacy-preserving ordinal response. In *SIGSAC 2014*, pp. 1054–1067, 2014.

Nan Fu, Weiwei Ni, Sen Zhang, Lihe Hou, and Dongyue Zhang. GC-NLDP: A graph clustering algorithm with local differential privacy. *Computers & Security*, 124:102967, 2023.

Stephen Guattery and Gary L. Miller. On the performance of spectral graph partitioning methods. In *SODA 1995*, pp. 233–242, 1995. ISBN 0898713498.

Lars Hagen and Andrew B Kahng. New spectral methods for ratio cut partitioning and clustering. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 11(9):1074–1085, 1992.

Jonathan Hehir, Aleksandra Slavkovic, and Xiaoyue Niu. Consistent spectral clustering of network block models under local differential privacy. *Journal of Privacy and Confidentiality*, 12(2), 2022.

Wassily Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58(301):13–30, 1963. doi: 10.1080/01621459.1963.10500830. URL `https://www.tandfonline.com/doi/abs/10.1080/01621459.1963.10500830`.

Jacob Imola, Takao Murakami, and Kamalika Chaudhuri. Locally differentially private analysis of graph statistics. In *USENIX Security 2021*, pp. 983–1000, 2021.

Jacob Imola, Takao Murakami, and Kamalika Chaudhuri. Communication-Efficient Triangle Counting under Local Differential Privacy. In *USENIX Security 2022*, pp. 537–554, 2022.

Tianxi Ji, Changqing Luo, Yifan Guo, Qianlong Wang, Lixing Yu, and Pan Li. Community detection in online social networks: A differentially private and parsimonious approach. *IEEE Transactions on Computational Social Systems*, 7(1):151–163, 2020.

Shiva Prasad Kasiviswanathan, Homin K. Lee, Kobbi Nissim, Sofya Raskhodnikova, and Adam Smith. What can we learn privately? *SIAM Journal on Computing*, 40(3):793–826, 2011.

Tsz Chiu Kwok, Lap Chi Lau, Yin Tat Lee, Shayan Oveis Gharan, and Luca Trevisan. Improved Cheeger's inequality: Analysis of spectral partitioning algorithms through higher order spectral gap. In *STOC 2013*, pp. 11–20, 2013.

Jure Leskovec and Julian Mcauley. Learning to discover social circles in ego networks. *NIPS 2012*, 25, 2012.

Naurang S. Mangat. An improved randomized response strategy. *Journal of the Royal Statistical Society: Series B (Methodological)*, 56(1):93–95, 1994.

Michael Mitzenmacher and Eli Upfal. *Probability and computing: Randomization and probabilistic techniques in algorithms and data analysis*. Cambridge University Press, 2017.

Mohamed S Mohamed, Dung Nguyen, Anil Vullikanti, and Ravi Tandon. Differentially private community detection for stochastic block models. In *International Conference on Machine Learning*, pp. 15858–15894. PMLR, 2022.

Kobbi Nissim, Sofya Raskhodnikova, and Adam Smith. Smooth sensitivity and sampling in private data analysis. In *STOC 2007*, pp. 75–84, 2007.

Pan Peng and Yuichi Yoshida. Average sensitivity of spectral clustering. In *KDD 2020*, pp. 1132–1140, 2020.

Yue Wang, Xintao Wu, and Leting Wu. Differential privacy preserving spectral graph analysis. In *PAKDD'13*, pp. 329–340. Springer, 2013.

Yue Wang, Xintao Wu, and Donghui Hu. Using randomized response for differential privacy preserving data collection. In *EDBT/ICDT Workshops*, 2016.

Stanley L. Warner. Randomized response: A survey technique for eliminating evasive answer bias. *Journal of the American Statistical Association*, 60(309):63–69, 1965.

Yen-Chuen Wei and Chung-Kuan Cheng. Towards efficient hierarchical designs by ratio cut partitioning. In *ICCAD 1989*, pp. 298–301, 1989. doi: 10.1109/ICCAD.1989.76957.

Hermann Weyl. Das asymptotische verteilungsgesetz der eigenwerte linearer partieller differentialgleichungen (mit einer anwendung auf die theorie der hohlraumstrahlung). *Mathematische Annalen*, 71(4):441–479, 1912.

Tianqing Zhu, Gang Li, Wanlei Zhou, and S Yu Philip. Differentially private data publishing and analysis: A survey. *IEEE Transactions on Knowledge and Data Engineering*, 29(8):1619–1638, 2017.