Local Spectral Clustering of Density Upper Level Sets

Alden Green Sivaraman Balakrishnan Ryan J. Tibshirani

Department of Statistics and Data Science
Carnegie Mellon University

{ajgreen,sbalakri,ryantibs}@stat.cmu.edu

November 21, 2019

Abstract

We analyze the Personalized PageRank (PPR) algorithm, a local spectral method for clustering, which extracts clusters using locally-biased random walks around a given seed node. In contrast to previous work, we adopt a classical statistical learning setup, where we obtain samples from an unknown distribution, and aim to identify connected regions of high-density (density clusters). We prove that PPR, run on a neighborhood graph, extracts sufficiently salient density clusters, that satisfy a set of natural geometric conditions. We also show a converse result, that PPR can fail to recover geometrically poorly-conditioned density clusters, even asymptotically. Finally, we provide empirical support for our theory.

1 Introduction

In this paper, we consider the problem of clustering: splitting a given data set into groups that satisfy some notion of within-group similarity and between-group difference. Our particular focus is on spectral clustering methods, a family of powerful nonparametric clustering algorithms. Roughly speaking, a spectral algorithm first constructs a geometric graph G, where vertices correspond to samples, and edges correspond to proximities between samples. The algorithm then estimates a feature embedding based on (an appropriate) Laplacian matrix of G, and applies a simple clustering technique (like k-means clustering) in the embedded feature space.

When applied to geometric graphs built from a large number of samples, global spectral clustering methods can be computationally cumbersome and insensitive to the local geometry of the underlying distribution [25, 28]. This has led to increased interest in *local* spectral clustering algorithms, which leverage locally-biased spectra computed using random walks around some user-specified seed node. A popular local clustering algorithm is the Personalized PageRank (PPR) algorithm, first introduced by Haveliwala [16], and then further developed by several others [4, 28, 40, 42, 51].

Local spectral clustering techniques have been practically very successful [5, 13, 25, 28, 50], which has led many authors to develop supporting theory [3, 12, 41, 51] that gives worst-case guarantees on traditional graph-theoretic notions of cluster quality (such as conductance). In this paper, we adopt a classical statistical viewpoint, and examine what the output of local clustering on a data set reveals about the underlying density f of the samples. In particular, we examine the ability of PPR to recover *density clusters* of f, defined as the connected components of the upper level set $\{x \in \mathbb{R}^d : f(x) \geq \lambda\}$ for some $\lambda > 0$ (an object of central interest in the statistical clustering literature, dating back to the work of Hartigan [15]).

2 Background and related work

We begin by providing some standard background on the PPR algorithm and the density clustering setup, before turning our attention to related work and a detailed summary of our contributions.

2.1 PPR on a neighborhood graph

Let $X = \{x_1, \ldots, x_n\}$ be a sample drawn i.i.d. from a distribution $\mathbb{P}$ on $\mathbb{R}^d$, with density f. For a radius $r > 0$, we define $G_{n,r} = (V, E)$ to be the *r-neighborhood graph* of X, an unweighted, undirected graph with vertices $V = X$, and an edge $(x_i, x_j) \in E$ if and only if $\|x_i - x_j\| \leq r$, where $\|\cdot\|$ is the Euclidean norm. We denote by $A \in \mathbb{R}^{n \times n}$ the adjacency matrix, with entries $A_{uv} = 1$ if $(u, v) \in E$ and 0 otherwise. We also denote by D the diagonal degree matrix, with entries $D_{uu} := \sum_{v \in V} A_{uv}$, and by I the $n \times n$ identity matrix.

First, we define the PPR vector $p_v = p(v, \alpha; G_{n,r})$, based on a given seed node $v \in V$ and a teleportation parameter $\alpha \in [0, 1]$, to be the solution of the following linear system:

$$p_v = \alpha e_v + (1 - \alpha) p_v W, \tag{1}$$

where $W = (I + D^{-1}A)/2$ is the lazy random walk matrix over $G_{n,r}$ and e_v is the indicator vector for node v (that has a 1 in position v and 0 elsewhere).

Next, we define a *β-sweep cut* of $p_v = (p_v(u))_{u \in V}$, for a given level $\beta > 0$, as

$$S_{\beta,v} := \left\{ u \in V : \frac{p_v(u)}{D_{uu}} > \beta \right\}. \tag{2}$$

We will use the normalized cut metric to determine which sweep cut S_β is the best cluster estimate. For a set $S \subseteq V$ with complement $S^c = V \setminus S$, we define $\mathrm{cut}(S; G_{n,r}) := \sum_{u \in S, v \in S^c} A_{uv}$, and $\mathrm{vol}(S; G_{n,r}) := \sum_{u \in S} D_{uu}$. We then define the *normalized cut* of S as

$$\Phi(S; G_{n,r}) := \frac{\mathrm{cut}(S; G_{n,r})}{\min\{\mathrm{vol}(S; G_{n,r}), \mathrm{vol}(S^c; G_{n,r})\}}. \tag{3}$$

Having computed sweep cuts S_β over $\beta \in (L, U)$ (where the range (L, U) is user-specified), we output the cluster estimate $\widehat{C} = S_{\beta^*}$ with minimum normalized cut. For concreteness, the PPR algorithm is summarized in Algorithm 1.

Algorithm 1 PPR on a neighborhood graph

Input: data $X = \{x_1, \ldots, x_n\}$, radius $r > 0$, teleportation parameter $\alpha \in [0, 1]$, seed $v \in X$, sweep cut range (L, U).
Output: cluster estimate $\widehat{C} \subseteq V$.

1: Form the neighborhood graph $G_{n,r}$.
2: Compute the PPR vector $p_v = p(v, \alpha; G_{n,r})$ as in (1).
3: For $\beta \in (L, U)$, compute sweep cuts S_β as in (2).
4: Return the cluster $\widehat{C} = S_{\beta^*}$, where

$$\beta^* = \underset{\beta \in (L,U)}{\arg\min}\ \Phi(S_\beta; G_{n,r}).$$

2.2 Estimation of density clusters

Let $\mathbb{C}_f(\lambda)$ denote the connected components of the density upper level set $\{x \in \mathbb{R}^d : f(x) > \lambda\}$. For a given density cluster $\mathcal{C} \in \mathbb{C}_f(\lambda)$, we call $\mathcal{C}[X] = \mathcal{C} \cap X$ the *empirical density cluster*. The size of the symmetric set difference between estimated and empirical cluster is a commonly used metric to quantify cluster estimation error [22, 32, 33]. We will consider a related metric, the volume of the symmetric set difference, which weights points according to their degree in $G_{n,r}$. To keep things simple, for a given set $S \subseteq X$ we write $\mathrm{vol}_{n,r}(S) := \mathrm{vol}(S; G_{n,r})$.

Definition 1. For an estimator $\widehat{C} \subseteq X$ and a set $\mathcal{S} \subseteq \mathbb{R}^d$, their symmetric set difference is

$$\widehat{C} \vartriangle \mathcal{S}[X] := \big(\widehat{C} \setminus \mathcal{S}[X]\big) \cup \big(\mathcal{S}[X] \setminus \widehat{C}\big).$$

Furthermore, we denote the volume of the symmetric set difference by

$$\Delta(\widehat{C}, \mathcal{S}[X]) := \mathrm{vol}_{n,r}(\widehat{C} \vartriangle \mathcal{S}[X]).$$

Note that the symmetric set difference does not measure whether $\widehat{C}$ can (perfectly) distinguish any two distinct clusters $\mathcal{C}, \mathcal{C}' \in \mathbb{C}_f(\lambda)$. We therefore also study a second notion of cluster estimation, first introduced by Hartigan [15], and defined asymptotically.

Definition 2. For an estimator $\widehat{C} \subseteq X$ and cluster $\mathcal{C} \in \mathbb{C}_f(\lambda)$, we call $\widehat{C}$ *consistent* for $\mathcal{C}$ if for all $\mathcal{C}' \in \mathbb{C}_f(\lambda)$ with $\mathcal{C} \neq \mathcal{C}'$, the following holds as $n \to \infty$:

$$\mathcal{C}[X] \subseteq \widehat{C} \quad \text{and} \quad \widehat{C} \cap \mathcal{C}'[X] = \emptyset, \tag{4}$$

with probability tending to 1.

Consistent cluster recovery roughly ensures that, for a given threshold $\lambda > 0$, the estimated cluster $\widehat{C}$ contains all points in a true density cluster $\mathcal{C} \in \mathbb{C}_f(\lambda)$, and simultaneously does not contain any points in any other density cluster $\mathcal{C}' \in \mathbb{C}_f(\lambda)$.

With these definitions in place, our broad goal will be to understand the extent to which the PPR algorithm is able to recover a cluster which either guarantees a low symmetric set difference to a true density cluster, or which consistently estimates a true density cluster.

2.3 Related work

In addition to the background on local spectral clustering given above, a few related lines of work are worth highlighting. In the stochastic block model (SBM), arguably one of the simplest models of network formation, edges between nodes independently occur with probability based on a latent community membership. In the SBM, the ability of spectral algorithms to perform clustering—or community detection—is well-understood, dating back to McSherry [29] who gives conditions under which the entire community structure can be recovered. In more recent work, Rohe et al. [34] upper bound the fraction of nodes misclassified by a spectral algorithm for the high-dimensional (large number of blocks) SBM, and Lei and Rinaldo [24] extend these results to the sparse (low average degree) regime. Relatedly, Balakrishnan et al. [6], Clauset et al. [10], Li et al. [26], analyze the misclassification rate when the block model exhibits some hierarchical structure. The framework we consider, in which nodes correspond to data points sampled from an underlying density, and edges between nodes are formed based on geometric proximity, is quite different than the SBM, and therefore so is our analysis.

In general, the study of spectral algorithms on neighborhood graphs has been focused on establishing asymptotic convergence of eigenvalues and eigenvectors of certain sample objects to the eigenvalues and eigenfunctions of corresponding limiting operators. Koltchinskii and Gine [21] establish convergence of spectral projections of the adjacency matrix to a limiting integral operator, with similar results obtained using simplified proofs in Rosasco et al. [35]. von Luxburg et al. [49] studies convergence of eigenvectors of the Laplacian matrix for a neighborhood of fixed radius. Belkin and Niyogi [8] and Trillos and Slepcev [46] extend these results to the regime where the radius $r \to 0$ as $n \to \infty$.

These results are of fundamental importance; however, the behavior of the spectra of these continuum operators can in general be hard to grasp. Therefore, further work relating this spectra to the geometry of the underlying distribution $\mathbb{P}$ is of interest. In this spirit, Schiebinger et al. [36], Shi et al. [38], Trillos et al. [45] examine the ability of spectral algorithms to recover the latent labels in certain geometrically well-conditioned nonparametric mixture models. Their results focus on global rather than local methods, and thus impose global rather than local conditions on the nature of the density. Moreover, they do not in general guarantee recovery of density clusters, which is the focus in our work. Perhaps most importantly, these works rely on general cluster saliency conditions, which implicitly depend on many distinct geometric aspects of the cluster $\mathcal{C}$ under consideration. We make this dependence more explicit, and in doing so expose the role each geometric condition plays in the clustering problem.

Our analysis naturally builds on a few of the aforementioned theoretical analyses of PPR. For an arbitrary graph G and subset $S \subseteq G$, Andersen et al. [4] relate the quality of the PPR cluster $\widehat{C}$ to the normalized cut functional $\Phi(S; G)$. While this analysis is tight in a worst-case sense, it fails to account for possible improvements when the cluster S is additionally assumed to be internally well-connected, which is an intuitively more favorable case for clustering. Building on this intuition, Zhu et al. [51] assume that the subgraph $G[S]$ is internally well-connected—as measured by a functional such as mixing time of a random walk over $G[S]$—and prove upper bounds on $\mathrm{vol}(\widehat{C} \vartriangle S; G)$. Both of these analyses also hold with respect to an approximate form of PPR (aPPR), which can be efficiently computed.

We apply these results to our setting by carefully analyzing the normalized cut and mixing time functionals in the particular case of $G = G_{n,r}$ and $S = \mathcal{C}[X]$. One of our main challenges is to prove an upper bound on the mixing time of a random walk run only over the subset of nodes in $G_{n,r}$ which fall within a density cluster $\mathcal{C}$. To do so, we rely on a series of seminal works upper bounding the mixing time of *geometric random walks* (see Vempala [48] for a comprehensive review.) This study was initiated by Dyer et al. [11], who used geometric random walks as a fundamental subroutine to efficiently compute volumes of high-dimensional convex bodies. These results are improved in Kannan et al. [18, 20], Lovász and Simonovits [27], who show, inter alia, that the bounds on mixing time can be sharpened by avoiding so-called "start-penalties". As we discuss further in what follows, these improvements are crucial to our work. Following the work of Abbasi-Yadkori [1], Abbasi-Yadkori et al. [2], we extend these results to hold for Lipschitz deformations of convex sets. Additionally, we relate the mixing time of these (continuous-space) geometric random walks to the mixing time of random walks over (discrete) neighborhood graphs.

Finally, it is worth mentioning that density clustering and level set estimation are themselves very well-studied problems in statistics. Polonik [32], Rigollet and Vert [33] study density clustering under the symmetric set difference metric, Singh et al. [39], Tsybakov [47] describe minimax optimal level-set estimators under Hausdorff loss and Balakrishnan et al. [7], Chaudhuri and Dasgupta [9], Hartigan [15], Kpotufe and von Luxburg [23] consider consistent estimation of the cluster tree. We emphasize that our goal is not to improve on these results, nor to offer a better algorithm for level set estimation; indeed, seen as a density clustering algorithm, PPR has none of the optimality guarantees found in the aforementioned works. Instead, our motivation is to start with a widely-used local spectral method, PPR, and to better understand and characterize the distinctions between those density clusters which are well-conditioned for PPR, and those which are not.

2.4 Summary of results

A summary of our results (and an outline for this paper) is as follows.

1. In Section 3, we introduce a set of natural geometric conditions on the density cluster $\mathcal{C}$ and show that if Algorithm 1 is properly initialized, then the size of the symmetric set difference between $\widehat{C}$ and a thickened version of the density cluster $\mathcal{C}_\sigma$ can be bounded in a meaningful way based on the geometric parameters.

2. We further show in Section 3 that if the density cluster $\mathcal{C}$ is particularly well-conditioned, then Algorithm 1 will consistently estimate a density cluster in the strong sense of (4).

3. In Section 4, we detail some of the analysis required to prove our main results, and expose the roles that various geometric quantities play in the difficulty of the clustering problem.

4. In Section 5, we provide an accompanying lower bound, which demonstrates that when the cluster $\mathcal{C}$ is sufficiently poorly conditioned, it will not be recovered by Algorithm 1.

5. In Section 6, we empirically investigate the tightness of our analysis, and provide examples showing how violations of our geometric conditions impact density cluster recovery by PPR.

One of our main takeaways can be paraphrased as follows: PPR, run on a neighborhood graph, recovers only *geometrically compact* high-density clusters. Our theoretical results make this takeaway precise, and provide a concrete way of quantifying the geometric compactness of a density cluster.

3 Main results

In this section, we present our main results on accuracy of the PPR algorithm for recovering density clusters. We begin by formally introducing various geometric conditions, and use these to define a condition number $\kappa(\mathcal{C})$, which measures the difficulty PPR will have in estimating a density cluster $\mathcal{C}$. With this condition in place our first main result (Theorem 1) provides a bound on the symmetric set difference between the estimated cluster $\widehat{C}$, obtained by an appropriately initialized version of the PPR algorithm, and the target cluster $\mathcal{C}$. Our next main result (Theorem 2) shows that for sufficiently well-conditioned target clusters $\mathcal{C}$, the output of the PPR algorithm $\widehat{C}$ is consistent in the sense of Definition 2.

3.1 Preliminaries

At a high level, for PPR to be successful, the underlying density cluster must be geometrically well-conditioned. A basic requirement is that we need to avoid clusters which contain arbitrarily thin bridges or spikes. As in the work of Chaudhuri and Dasgupta [9], we consider a thickened version of $\mathcal{C} \in \mathbb{C}_f(\lambda)$ defined as $\mathcal{C}_\sigma := \left\{x \in \mathbb{R}^d : \mathrm{dist}(x, \mathcal{C}) \leq \sigma\right\}$, which we call the *$\sigma$-expansion* of $\mathcal{C}$. Here $\mathrm{dist}(x, \mathcal{C}) := \inf_{y \in \mathcal{C}} \|y - x\|$. We now list our conditions on $\mathcal{C}_\sigma$.

(A1) *Bounded density within cluster:* There exist constants $0 < \lambda_\sigma < \Lambda_\sigma < \infty$ such that

$$\lambda_\sigma \leq \inf_{x \in \mathcal{C}_\sigma} f(x) \leq \sup_{x \in \mathcal{C}_\sigma} f(x) \leq \Lambda_\sigma.$$

(A2) *Low noise density:* There exists constants $c_0 > 0$ and $\gamma \in [0, 1]$ such that for any $x \in \mathbb{R}^d$ with $0 < \mathrm{dist}(x, \mathcal{C}_\sigma) \leq \sigma$,

$$\inf_{x' \in \mathcal{C}_\sigma} f(x') - f(x) \geq c_0 \cdot \mathrm{dist}(x, \mathcal{C}_\sigma)^\gamma.$$

Roughly, this assumption ensures that the density decays sufficiently quickly as we move away from the target cluster $\mathcal{C}_\sigma$, and is a standard assumption in the level-set estimation literature (see for instance Singh et al. [39]).

(A3) *Lipschitz embedding:* There exists $g : \mathbb{R}^d \to \mathbb{R}^d$ with the following properties:

(a) we have $\mathcal{C}_\sigma = g(\mathcal{K})$, for a convex set $\mathcal{K} \subseteq \mathbb{R}^d$ with $\mathrm{diam}(\mathcal{K}) = \sup_{x,y \in \mathcal{K}} \|x - y\| =: \rho < \infty$;

(b) $\det(\nabla g(x)) = 1$ for all $x \in \mathcal{C}_\sigma$, where $\nabla g(x)$ is the Jacobian of g evaluated at x; and

(c) for some $L \geq 1$,

$$\|g(x) - g(y)\| \leq L \|x - y\| \text{ for all } x, y \in \mathcal{K}.$$

Succinctly, we assume that $\mathcal{C}_\sigma$ is the image of a convex set with finite diameter under a measure preserving, Lipschitz transformation.

(A4) *Bounded volume:* For a set $\mathcal{S} \subseteq \mathbb{R}^d$, define a $\mathbb{P}$-weighted volume of $\mathcal{S}$ to be

$$\mathrm{vol}_{\mathbb{P},r}(\mathcal{S}) := \int_{\mathcal{S}} \mathbb{P}(B(x, r)) f(x)\, dx. \tag{5}$$

where $B(x, r)$ is the closed ball of radius r centered at x. We assume that the radius of the neighborhood graph $0 < r \leq \sigma/(2d)$ is chosen such that

$$\mathrm{vol}_{\mathbb{P},r}(\mathcal{C}_\sigma) \leq \frac{1}{2} \mathrm{vol}_{\mathbb{P},r}(\mathbb{R}^d).$$

To motivate these conditions, we now give a brief high-level sketch of our analysis (which we will return to more formally in Section 4). Zhu et al. [51] show that for an arbitrary graph $G = (V, E)$ and subset of vertices $S \subseteq V$, the PPR algorithm (properly initialized within S) will output an estimate $\widehat{C}$ of S satisfying, for a constant $M > 0$,

$$\mathrm{vol}(\widehat{C} \vartriangle S; G) \leq M \cdot \Phi(S, G) \cdot \tau_\infty(G[S]), \tag{6}$$

where $\Phi(S; G)$ is the normalized cut of S (as defined in (3)), and $\tau_\infty(G[S])$ is the *mixing time* of a random walk over the induced subgraph $G[S]$ (to be defined precisely later, in (15)). The left-hand side in (6) is one of our principle metrics of interest, the volume of the symmetric set difference, and our main goal will therefore be to upper bound the graph functionals Φ and τ_∞. Towards this goal, as we will show in Section 4, the conditions (A1)–(A4) allow us to upper bound the normalized cut $\Phi(\mathcal{C}_\sigma[X]; G_{n,r})$, and the mixing time $\tau_\infty(G_{n,r}[\mathcal{C}_\sigma[X]])$. Specifically, assumption (A2) yields an upper bound on $\mathrm{cut}(\mathcal{C}_\sigma[X]; G_{n,r})$, and (A1) yields a lower bound on $\mathrm{vol}_{n,r}(\mathcal{C}_\sigma[X])$; together with (A4), this gives an upper bound on the normalized cut. On the other hand, (A1) and (A3) preclude bottlenecks in the induced subgraph $G_{n,r}[\mathcal{C}_\sigma[X]]$, and combined with the upper bound on diameter in (A3), this leads to an upper bound on the mixing time over this subgraph.

3.1.1 Condition number

We will define the condition number $\kappa(\mathcal{C})$ of a cluster $\mathcal{C}$ in terms of a suitable upper bound—expressed in terms of the geometric parameters from (A1)–(A4)—on the product of normalized cut and mixing time, $\Phi(\mathcal{C}_\sigma[X]; G_{n,r}) \cdot \tau_\infty(G_{n,r}[\mathcal{C}_\sigma[X]])$. Following (6), we see that the smaller the condition number $\kappa(\mathcal{C})$ is, the more success PPR will have in recovering the target $\mathcal{C}$.

Definition 3. For $\lambda > 0$ and $\mathcal{C} \in \mathbb{C}_f(\lambda)$, let $\mathcal{C}$ satisfy (A1)–(A4) for some $\sigma > 0$. Then, for universal constants $c_1, c_2, c_3 > 0$ to be specified later (in Theorems 3 and 4), define

$$\Phi_u(\mathcal{C}) := c_1 r \frac{d}{\sigma} \frac{\lambda}{\lambda_\sigma} \frac{(\lambda_\sigma - c_0 \frac{r^\gamma}{\gamma+1})}{\lambda_\sigma}, \quad \tau_u(\mathcal{C}) := c_2 \frac{\Lambda_\sigma^4 d^3 \rho^2 L^2}{\lambda_\sigma^4 r^2} \log^2\left(\frac{\Lambda_\sigma}{\lambda_\sigma^2 r}\right) + c_3. \tag{7}$$

Letting $\kappa(\mathcal{C}) := \Phi_u(\mathcal{C}) \cdot \tau_u(\mathcal{C})$, we call $\kappa(\mathcal{C})$ the *condition number* of $\mathcal{C}$. We also call the set $\mathcal{C}$ a *κ-well-conditioned* density cluster.

The condition number $\kappa(\mathcal{C})$ succinctly captures the role of the various geometric parameters. We note in passing that $\Phi_u(\mathcal{C})$ and $\tau_u(\mathcal{C})$ are exactly the upper bounds on $\Phi(\mathcal{C}_\sigma[X]; G_{n,r})$ and $\tau_\infty(G_{n,r}[\mathcal{C}_\sigma[X]])$ that we derive in our analysis later, in Section 4.

3.1.2 Well-initialized algorithm

As is typical in the local clustering literature, our algorithmic results will be stated with respect to specific ranges of each of the user-specified parameters. In particular, for a well-conditioned density cluster $\mathcal{C}$, we require that some of the tuning parameters of Algorithm 1 are chosen to fall within specific ranges,

$$\begin{aligned} 0 < r \leq \frac{\sigma}{2d}, \quad \alpha \in [\tfrac{1}{10}, \tfrac{1}{9}) \cdot \frac{1}{\tau_u(\theta)}, \\ (L, U) \subseteq (\tfrac{1}{50}, \tfrac{1}{5}) \cdot \frac{1}{2\binom{n}{2}\mathrm{vol}_{\mathbb{P},r}(\mathcal{C}_\sigma)}. \end{aligned} \tag{8}$$

Definition 4. If the input parameters to Algorithm 1 satisfy (8) (for some well-conditioned density cluster $\mathcal{C}$), then we say the algorithm is *well-initialized*.

In practice of course, it is not feasible to set tuning parameters based on the underlying (unknown) density f. Typically, one runs PPR over some range of tuning parameter values and selects the cluster which has the smallest normalized cut.

3.2 Cluster recovery in symmetric set difference

We now present our first main result: a bound on the volume of the symmetric set difference between the estimated cluster $\widehat{C}$ and empirical cluster $\mathcal{C}_\sigma[X]$. In this theorem, and hereafter, we let $c, c_i > 0$ denote universal constants, and $b, b_i > 0$ denote constants which may depend on $\mathbb{P}, \lambda, r, d$ and so on, but *not* on the sample size n.

Theorem 1. *Fix $\lambda, \sigma > 0$, let $\mathcal{C} \in \mathbb{C}_f(\lambda)$ be a κ-well-conditioned density cluster, and assume Algorithm 1 is well-initialized with respect to $\mathcal{C}$. Then for any*

$$n \geq b_1 (\log n)^{\max\{\frac{3}{d}, 1\}},$$

there exists a set $\mathcal{C}_\sigma[X]^g \subseteq \mathcal{C}_\sigma[X]$ of large volume, $\mathrm{vol}_{n,r}(\mathcal{C}_\sigma[X]^g) \geq \mathrm{vol}_{n,r}(\mathcal{C}_\sigma[X])/2$, such that the following holds: if Algorithm 1 is run with any seed node $v \in \mathcal{C}_\sigma[X]^g$, then the PPR estimated cluster $\widehat{C}$ satisfies

$$\Delta(\widehat{C}, \mathcal{C}_\sigma[X]) \leq c \cdot \kappa(\mathcal{C}) \cdot \mathrm{vol}_{n,r}(\mathcal{C}_\sigma[X]), \tag{9}$$

with probability at least $1 - b_2/n$.

The proof of Theorem 1, as with all results in this paper, is deferred to the appendix. We reiterate that the primary technical work involved in proving Theorem 1 involves showing that $\Phi_u(\mathcal{C})$ and $\tau_u(\mathcal{C})$ in (7) are valid upper bounds on the normalized cut and mixing time; once this has been established, the result follows more or less straightforwardly from Zhu et al. [51].

3.3 Consistent cluster recovery

The bound on symmetric set difference (9) does not imply consistent density cluster estimation in the sense of (4). This notion of consistency requires a uniform bound over the PPR vector p_v: as an example, suppose that we were able to show that for all $\mathcal{C}' \in \mathbb{C}_f(\lambda), \mathcal{C}' \neq \mathcal{C}$, and each $u \in \mathcal{C}, w \in \mathcal{C}'$,

$$\frac{p_v(w)}{D_{ww}} \leq \frac{1}{100\binom{n}{2}\mathrm{vol}_{\mathbb{P},r}(\mathcal{C}_\sigma)} < \frac{1}{10\binom{n}{2}\mathrm{vol}_{\mathbb{P},r}(\mathcal{C}_\sigma)} \leq \frac{p_v(u)}{D_{uu}}. \tag{10}$$

Then, any (L, U) satisfying (8) and any sweep cut S_β for $\beta \in (L, U)$ would fulfill both conditions laid out in (4). In Theorem 2, we show that a sufficiently small upper bound on $\kappa(\mathcal{C})$ ensures that with high probability the uniform bound (10) is satisfied, and hence implies $\widehat{C}$ will be a consistent estimator. We will need one additional regularity condition, to preclude arbitrarily low degree vertices for points $x \in \mathcal{C}'[X]$.

(A5) *Bounded density in other clusters:* Letting σ, λ_σ be as in (A1), for each $\mathcal{C}' \in \mathbb{C}_f(\lambda)$ and for all $x \in \mathcal{C}'_\sigma$, it holds that $\lambda_\sigma \leq f(x)$.

Next we give our main result on consistent cluster recovery by PPR.

Theorem 2. *Under the assumptions of Theorem 1, additionally assume (A5), and*

$$\kappa(\mathcal{C}) \leq c\frac{(\lambda_\sigma r^d \nu_d)^2}{\mathrm{vol}_{\mathbb{P},r}(\mathcal{C}_\sigma)}. \tag{11}$$

Then for any

$$n \geq b_1 (\log n)^{\max\{\frac{3}{d}, 1\}},$$

there exists a set $\mathcal{C}_\sigma[X]^g \subseteq \mathcal{C}_\sigma[X]$ of large volume, $\mathrm{vol}_{n,r}(\mathcal{C}_\sigma[X]^g) \geq \mathrm{vol}_{n,r}(\mathcal{C}_\sigma[X])/2$, such that if Algorithm 1 is run with any seed node $v \in \mathcal{C}_\sigma[X]^g$, then the PPR estimated cluster $\widehat{C}$ satisfies (4) *with probability at least $1 - b_2/n$.*

Some remarks are in order.

Remark 1. We note that the restriction on $\kappa(\mathcal{C})$ imposed by (11) results in an upper bound on the symmetric set difference metric $\Delta(\widehat{C}, \mathcal{C}_\sigma[X])$ on the order of r^d. In plain terms, we are able to recover a density cluster $\mathcal{C}$ in the strong sense of (4) only when we can guarantee a very small fraction of points will be misclassified. This strong condition is the price we pay in order to obtain the uniform bound in (10).

Remark 2. Letting the radius of the neighborhood graph shrink, $r \to 0$ as $n \to \infty$, would be computationally attractive (it would ensure that the graph $G_{n,r}$ is sparse), but the presence of a factor of $\log^2(1/r)/r$ in $\kappa(\mathcal{C})$ prevents us from making claims about the behavior of PPR in this regime. Although the restriction to a kernel function fixed in n is common in spectral clustering theory [36, 49], it is an interesting question whether PPR exhibits some degeneracy over neighborhood graphs as $r \to 0$, or if this is merely looseness in our upper bounds.

3.4 Approximate PPR vector

In practice, exactly solving the system of equations (1) to compute the PPR vector may be too computationally expensive. To address this limitation, Andersen et al. [4] introduced the ε-*approximate* PPR vector (aPPR), which we will denote by $p^{(\varepsilon)}$. We refer the curious reader to Andersen et al. [4] for a formal algorithmic definition of the aPPR vector, and limit ourselves to highlighting a few salient points: the aPPR vector can be computed in order $\mathcal{O}(1/(\varepsilon\alpha))$ time, while satisfying the following uniform error bound:

$$\text{for all } u \in V, \quad p(u) - \varepsilon D_{uu} \leq p^{(\varepsilon)}(u) \leq p(u). \tag{12}$$

For a sufficiently small choice of ε, the application of (12) within the proofs of Theorems 1 and 2 leads to analogous results which hold for $p^{(\varepsilon)}$.

Corollary 1. *Consider instead of Algorithm 1 using the approximate PPR vector from Andersen et al. [4] satisfying* (12)*, and forming the corresponding cluster estimate* $\widehat{C}$ *in the same manner. Then provided we take*

$$\varepsilon = \frac{1}{25}\mathrm{vol}_{\mathbb{P},r}(\mathcal{C}_\sigma), \tag{13}$$

under the assumptions of Theorem 1 the upper bound on symmetric set difference in (9) *still holds, and under the assumptions of Theorem 2 the set inclusion and disjointedness statements in* (4) *still hold, each with probability at least* $1 - b_2/n$ *(under possibly different choices of the constants* c_i *and* b_i*).*

4 Analysis overview

The primary technical contribution in our work is to show that the geometric assumptions (A1)–(A4) translate to meaningful bounds on the normalized cut and mixing time of $\mathcal{C}_\sigma[X]$ in the neighborhood graph $G_{n,r}$. In doing so, we elucidate how these geometric conditions contribute to the difficulty of the clustering problem.

4.1 Upper bound on normalized cut

We start with a finite-sample upper bound on the normalized cut (3) of $\mathcal{C}_\sigma[X]$. For simplicity, we write $\Phi_{n,r}(\mathcal{C}_\sigma[X]) := \Phi(\mathcal{C}_\sigma[X]; G_{n,r})$.

Theorem 3. *Fix* $\lambda, \sigma > 0$*, and assume* $\mathcal{C} \in \mathbb{C}_f(\lambda)$ *satisfies Assumptions (A1), (A2), and (A4). Then*

$$\frac{\Phi_{n,r}(\mathcal{C}_\sigma[X])}{r} \leq c_1 \frac{d}{\sigma}\frac{\lambda}{\lambda_\sigma}\frac{(\lambda_\sigma - c_0\frac{r^\gamma}{\gamma+1})}{\lambda_\sigma}, \tag{14}$$

with probability at least $1 - 3\exp\{-bn\}$.

Remark 3. Observe that the diameter ρ is absent from Theorem 3, in contrast to the ultimate bound in Theorem 1 where the diameter enters through the condition number $\kappa(\mathcal{C})$, which worsens (increases) as ρ increases. This reflects (what may be regarded as) established wisdom regarding spectral partitioning algorithms more generally [14, 17], albeit newly applied to the density clustering setting: if the diameter ρ is large, then PPR may fail to recover $\mathcal{C}_\sigma[X]$ even when $\mathcal{C}$ is sufficiently well-conditioned to ensure that $\mathcal{C}_\sigma[X]$ has a small normalized cut in $G_{n,r}$. This will be supported by simulations in Section 6.

4.2 Upper bound on mixing time

For $S \subseteq V$, denote by $G[S] = (S, E_S)$ the subgraph induced by S (where $E_S = E \cap (S \times S)$). Let W_S be the lazy random walk matrix over $G[S]$, and write

$$q_v^{(t)}(u) = e_v W_S^t e_u$$

for the t-step transition probability of the random walk over $G[S]$ originating at $v \in V$. Also let $\pi = (\pi(u))_{u \in S}$ be the stationary distribution of this random walk. (As W_S is the transition matrix of a lazy random walk, it is well-known that a unique stationary distribution exists and is given by $\pi(u) = \deg(u; G[S])/\mathrm{vol}(S; G[S])$, where we write $\deg(u; G[S]) = \sum_{w \in S} \mathbb{1}((u,w) \in E_S)$ for the degree of u in $G[S]$.) We define the *mixing time* of $G[S]$ as

$$\tau_\infty(G[S]) = \min\left\{t : \frac{\pi(u) - q_v^{(t)}(u)}{\pi(u)} \leq \frac{1}{4}, \text{ for } u, v \in V\right\}. \tag{15}$$

Next, we give an asymptotic (in the number of samples n) upper bound on $\tau_\infty(G_{n,r}[\mathcal{C}_\sigma[X]])$.

Theorem 4. *Fix $\lambda, \sigma > 0$, and assume $\mathcal{C} \in \mathbb{C}_f(\lambda)$ satisfies Assumptions (A1) and (A3). Also assume that $0 < r \leq \sigma/(2\sqrt{d})$. Then for any*

$$n \geq b_1 (\log n)^{\max\{\frac{3}{d}, 1\}},$$

the mixing time satisfies

$$\tau_\infty(G_{n,r}[\mathcal{C}_\sigma[X]]) \leq c_2 \frac{\Lambda_\sigma^4 d^3 \rho^2 L^2}{\lambda_\sigma^4 r^2} \log^2\left(\frac{\Lambda_\sigma}{\lambda_\sigma^2 r}\right) + c_3, \tag{16}$$

with probability at least $1 - b_2/n$.

The proof of Theorem 4 relies heavily on analogous mixing time bounds developed for a continuous-space "ball walk" over convex sets. To the best of our knowledge, our result is the first bound on the mixing time of random walks over neighborhood graphs that is independent of n, the number of vertices.

Remark 4. The embedding assumption (A3) and Lipschitz parameter L play an important role in proving the upper bound in Theorem 4. There is some interdependence between L and σ, ρ, which might lead one to hope that (A3) is non-essential. However, it is not possible to eliminate condition (A3) without incurring an additional factor of at least $(\rho/\sigma)^d$ in (16), achieved, for instance, when $\mathcal{C}_\sigma$ is a dumbbell-like set consisting of two balls of diameter ρ linked by a cylinder of radius σ. Abbasi-Yadkori [1], Abbasi-Yadkori et al. [2] develop theory regarding Lipschitz deformations of convex sets, wherein it is observed that star-shaped sets as well as half-moon shapes of the type we consider in Section 6 both satisfy (A3) for reasonably small values of L.

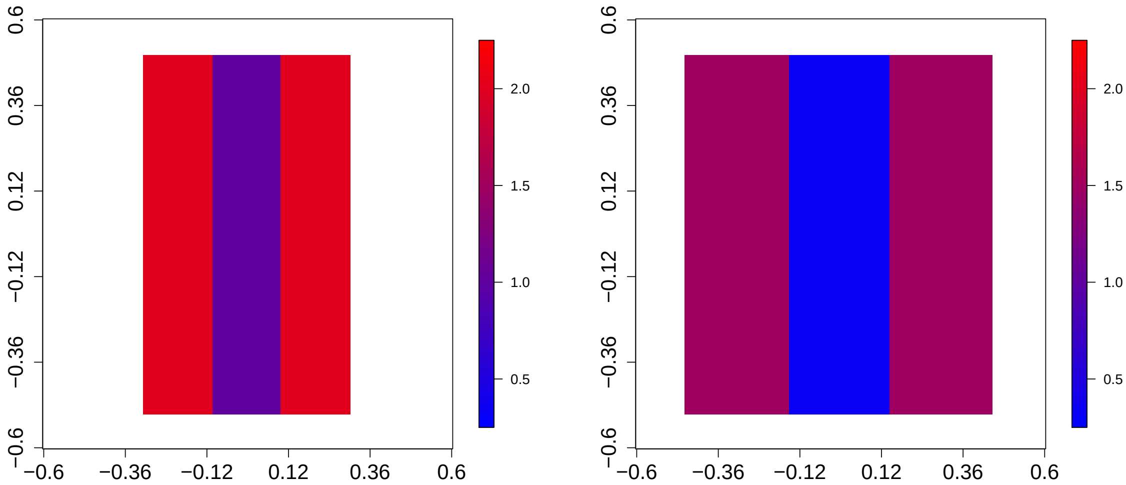


Figure 1. *The density f in (17), for $\rho = 1$, and two different choices of ϵ and σ. Left: $\epsilon = 0.3$ and $\sigma = 0.1$; right: $\epsilon = 0.2$ and $\sigma = 0.2$.*

5 Negative result

In this section, we exhibit a hard case for density clustering using PPR, that is, a distribution $\mathbb{P}$ for which PPR is unlikely to recover a density cluster. Let $\mathcal{C}^{(0)}, \mathcal{C}^{(1)}, \mathcal{C}^{(2)}$ be rectangles in $\mathbb{R}^2$,

$$\mathcal{C}^{(0)} = \left[-\frac{\sigma}{2}, \frac{\sigma}{2}\right] \times \left[-\frac{\rho}{2}, \frac{\rho}{2}\right], \quad \mathcal{C}^{(1)} = \mathcal{C}^{(0)} - \{(\sigma, 0)\}, \quad \mathcal{C}^{(2)} = \mathcal{C}^{(0)} + \{(\sigma, 0)\},$$

where $0 < \sigma < \rho$, and let $\mathbb{P}$ be the mixture distribution over $\mathcal{X} = \mathcal{C}^{(0)} \cup \mathcal{C}^{(1)} \cup \mathcal{C}^{(2)}$ given by

$$\mathbb{P} = \frac{1-\epsilon}{2}\Psi_1 + \frac{1-\epsilon}{2}\Psi_2 + \frac{\epsilon}{2}\Psi_0,$$

where Ψ_k is the uniform distribution over $\mathcal{C}^{(k)}$ for $k = 0, 1, 2$. The density function f of $\mathbb{P}$ is simply

$$f(x) = \frac{1}{\rho\sigma}\left(\frac{1-\epsilon}{2}\mathbb{1}(x \in \mathcal{C}^{(1)}) + \frac{1-\epsilon}{2}\mathbb{1}(x \in \mathcal{C}^{(2)}) + \frac{\epsilon}{2}\mathbb{1}(x \in \mathcal{C}^{(0)})\right), \tag{17}$$

so that for any $\epsilon < \lambda < (1-\epsilon)/2$, we have $\mathbb{C}_f(\lambda) = \left\{\mathcal{C}^{(1)}, \mathcal{C}^{(2)}\right\}$. Figure 1 visualizes the density f for two different choices of ϵ, σ, ρ.

5.1 Lower bound on symmetric set difference

As the following theorem demonstrates, even when Algorithm 1 is reasonably initialized, if the density cluster $\mathcal{C}^{(1)}$ is sufficiently geometrically ill-conditioned (in words, tall and thin) the cluster estimator $\widehat{C}$ will fail to recover $\mathcal{C}^{(1)}$. Let

$$\mathcal{L} = \{(x_1, x_2) \in \mathcal{X} : x_2 < 0\}. \tag{18}$$

Theorem 5. *Assume Algorithm 1 is initialized using inputs $r < \min\left\{\frac{1}{40}\rho, \frac{1}{4}\sigma\right\}$, $\alpha = 65\Phi_{\mathbb{P}}(\mathcal{L})$, and $(L, U) = (0, 1)$. Then, for any*

$$n \geq \max\left\{\frac{64}{\epsilon^2\rho\sigma\pi r^2}, \frac{8}{\epsilon}\right\},$$

there exists a set $\mathcal{C}[X]^g$ *of large volume,* $\mathrm{vol}_{n,r}(\mathcal{C}[X]^g \cap \mathcal{C}^{(1)}[X]) \geq \mathrm{vol}_{n,r}(\mathcal{C}^{(1)}[X]; G_{n,r})/10$*, such that for any seed node* $v \in \mathcal{C}[X]^g$*, the PPR estimated cluster* $\widehat{C}$ *satisfies*

$$\frac{\sigma\rho}{r^2 n^2} \cdot \mathrm{vol}_{n,r}(\widehat{C} \vartriangle \mathcal{C}^{(1)}[X]) \geq \frac{1}{4} - c\frac{\sqrt{\sigma/\rho}}{\epsilon^2}\sqrt{\log\left(\frac{\rho\sigma}{\epsilon^2 r^2}\right)\frac{\sigma}{r}}, \tag{19}$$

with probability at least $1 - b_1 n \exp\{-b_2 n\}$*. Consequently, if*

$$\epsilon^2 > \frac{c}{8}\sqrt{\frac{\sigma}{\rho}} \cdot \sqrt{\log\left(\frac{\rho\sigma}{\epsilon^2 r^2}\right)\frac{\sigma}{r}},$$

then with high probability $\frac{\sigma\rho}{r^2 n^2} \cdot \mathrm{vol}_{n,r}(\widehat{C} \vartriangle \mathcal{C}^{(1)}[X]) \geq 1/8$*.*

Note that $\mathrm{vol}_{n,r}(\mathcal{C}^{(1)}[X])$ for large enough n will be of the order $n^2 r^2/(\sigma\rho)$, and therefore the quantity $\frac{\sigma\rho}{r^2 n^2} \cdot \mathrm{vol}_{n,r}(\widehat{C} \vartriangle \mathcal{C}^{(1)}[X])$ in (19) is comparable to $\mathrm{vol}_{n,r}(\widehat{C} \vartriangle \mathcal{C}^{(1)}[X])/\mathrm{vol}_{n,r}(\mathcal{C}^{(1)}[X])$, which corresponds to the quantity we upper bound in Theorem 1.

Theorem 5 is stated with respect to a particular hard case, where the density clusters are rectangular subsets of $\mathbb{R}^2$. We chose this setting to make the theorem simple to state, and our results are generalizable to $\mathbb{R}^d$ and to non-rectangular clusters. Moreover, although we state our lower bound with respect to PPR run on a neighborhood graph, the conclusion is likely to hold for a much broader class of spectral clustering algorithms. In the proof of Theorem 5, we rely heavily on the fact that when ϵ^2 is sufficiently greater than σ/ρ, the normalized cut of $\mathcal{C}^{(1)}$ will be much larger than that of $\mathcal{L}$. In this case, not merely PPR but any algorithm that approximates the minimum normalized cut is unlikely to recover $\mathcal{C}^{(1)}$. In particular, local spectral clustering algorithms based on truncated random walks Spielman and Teng [41], global spectral clustering algorithms Shi and Malik [37], and p-Laplacian based spectral embeddings Hein and Bühler [17] all have provable upper bounds on the normalized cut of cluster they output, and thus we expect that they would all fail to estimate $\mathcal{C}^{(1)}$.

5.2 Comparison with previous upper bound

To better digest the implications of Theorem 5, we translate the results of our upper bound in Theorem 1 to the density f given in (17). Observe that $\mathcal{C}^{(1)}$ satisfies each of the Assumptions (A1)–(A4):

(A1) The density $f(x) = \frac{1-\epsilon}{2\rho\sigma}$ for all $x \in \mathcal{C}^{(1)}$.

(A2) The density $f(x) = \frac{\epsilon}{\rho\sigma}$ for all x such that $0 < \mathrm{dist}(x, \mathcal{C}^{(1)}) \leq \sigma$. Therefore for all such x,

$$\inf_{x' \in \mathcal{C}^{(1)}} f(x') - f(x) = \left\{\frac{1-\epsilon}{2} - \epsilon\right\}\frac{1}{\rho\sigma},$$

which meets the decay requirement with exponent $\gamma = 0$.

(A3) The set $\mathcal{C}^{(1)}$ is itself convex, and has diameter ρ.

(A4) By symmetry, $\mathrm{vol}_{\mathbb{P},r}(\mathcal{C}^{(1)}) = \mathrm{vol}_{\mathbb{P},r}(\mathcal{C}^{(2)})$, and therefore $\mathrm{vol}_{\mathbb{P},r}(\mathcal{C}^{(1)}) \leq \frac{1}{2}\mathrm{vol}_{\mathbb{P},r}(\mathbb{R}^d)$.

Remark 5. Technically, the rectangles $\mathcal{C}^{(0)}, \mathcal{C}^{(1)}, \mathcal{C}^{(2)}$ are not σ-expansions due to their sharp corners. To fix this, one can simply modify these sets to have appropriately rounded corners, and our lower bound arguments do not need to change significantly, subject to some additional bookkeeping. Thus we ignore this technicality in our subsequent discussion.

If the user-specified parameters are initialized according to (8), we may apply Theorem 1. This implies that there exists a set $\mathcal{C}^{(1)}[X] \subseteq \mathcal{C}^{(1)}$ with $\mathrm{vol}_{n,r}(\mathcal{C}[X]^g) \geq \frac{1}{2}\mathrm{vol}_{n,r}(\mathcal{C}[X])$ such that for any seed node $v \in \mathcal{C}^{(1)}[X]$, and for large enough n, the PPR estimated cluster $\widehat{C}$ satisfies with high probability

$$\mathrm{vol}_{n,r}(\widehat{C} \vartriangle \mathcal{C}^{(1)}[X]) \leq c \cdot \kappa(\mathcal{C}^{(1)}) \cdot \mathrm{vol}_{n,r}(\mathcal{C}^{(1)}[X]),$$

where the condition number may be taken to be

$$\kappa(\mathcal{C}^{(1)}) = c_1 \frac{\epsilon}{\sigma}\left(\frac{\rho^2}{r}\log^2\left(\frac{1}{r}\right)\right) + c_2.$$

for universal constants $c_1, c_2 > 0$. To facilitate comparisons between our upper and lower bounds, assume $\sigma/4 \leq \rho/40$ and set $r = \sigma/4$. Then the following statements each hold with high probability.

- If the user-specified parameters satisfy (8), and for some $c > 0$

$$\epsilon < c\left(\frac{\sigma}{\rho\log(1/\sigma)}\right)^2,$$

  then $\Delta(\widehat{C}, \mathcal{C}^{(1)}[X]) \leq c \cdot \mathrm{vol}_{n,r}(\mathcal{C}^{(1)}[X])$.

- If the user-specified parameters are as in Theorem 5, and for some $c > 0$

$$\epsilon > c\left(\frac{\sigma}{\rho}\log^2\left(\frac{\rho}{\epsilon^2\sigma}\right)\right)^{1/4},$$

  then $\Delta(\widehat{C}, \mathcal{C}^{(1)}[X]) \geq \frac{1}{8}\mathrm{vol}_{n,r}(\mathcal{C}^{(1)}[X])$.

Jointly, these upper and lower bounds give a relatively precise characterization of what it means for a density cluster to be well- or poorly-geometrically conditioned for recovery using PPR.

Remark 6. It is worth pointing out that the above conclusions are reliant on specific (albeit reasonable) ranges and choices of input parameters, which in some instances differ between the upper and lower bounds. We suspect that our lower bound continues to hold even when choosing input parameters as dictated by our upper bound, but do not pursue the details.

Remark 7. It is not hard to show that, in the example under consideration, classical plug-in density cluster estimators can consistently recover the σ-expansion $\mathcal{C}_\sigma$ of a density cluster $\mathcal{C}$, even if ϵ is large compared to σ/ρ. That PPR has trouble recovering density clusters here (where standard plug-in approaches do not) is not meant to be a knock on PPR. Rather, it simply reflects that while classical density clustering approaches are specifically designed to identify high-density regions regardless of their geometry, PPR relies on geometry as well as density when forming the output cluster.

6 Experiments

We provide numerical experiments to investigate the tightness of our bounds on the normalized cut and mixing time of $\mathcal{C}_\sigma[X]$, and examine the performance of PPR on the "two moons" dataset. We defer details of the experimental settings to the appendix.

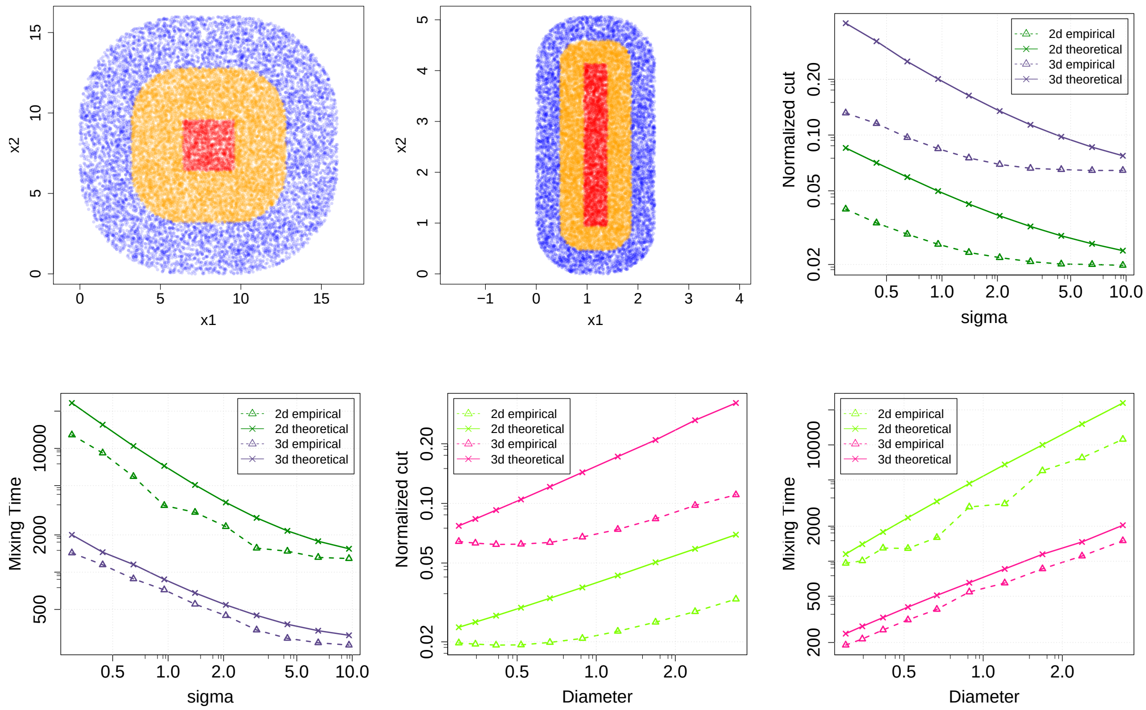


Figure 2. *Top left and top middle: samples from a geometrically well- and poor-conditioned cluster. The points in $\mathcal{C}$ are colored in red, points in $\mathcal{C}_\sigma \setminus \mathcal{C}$ are colored in yellow, and the remaining points in blue. Other panels: empirical normalized cut and mixing time, as a function of σ or ρ, versus their theoretical upper bounds.*

Validating theoretical bounds. We investigate the tightness of Theorems 3 and 4 via simulation. Figure 2 compares our upper bounds with the actual empirically-computed quantities (3) and (15), as we vary the diameter ρ and thickness σ of a cluster $\mathcal{C}$. The top left and top middle panels display the resulting empirical clusters for two different values of ρ, σ.

The bottom left and bottom right panels assure that our mixing time upper bounds track closely the empirical mixing time, in both 2 and 3 dimensions.[1] This provides empirical evidence that Theorem 4 has the right dependency on both expansion parameter σ and diameter ρ. The story for the normalized cut panels is less obvious. We remark that while, broadly speaking, the trends do not appear to match, this gap between theory and empirical results seems largest when σ and ρ are approximately equal. As the ratio ρ/σ grows, the slopes of empirical and theoretical curves become more similar.

Empirical behavior of PPR. In Figure 3, to drive home the implications of Theorems 1 and 2, we show the behavior of PPR, normalized cut, and the density clustering algorithm of Chaudhuri and Dasgupta [9] on the well-known "two moons" dataset (with added 2d Gaussian noise), considered a prototypical success story for spectral clustering algorithms. The first column shows the empirical density clusters $\mathcal{C}[X]$ and $\mathcal{C}'[X]$ for a particular threshold λ of the density function; the second column shows the cluster recovered by PPR; the third column shows the global minimum normalized cut, computed according to the algorithm of Szlam

[1] We rescaled all values of theoretical upper bounds by a constant, to mask the effect of large universal constants in these bounds. Therefore only the comparison of slopes, rather than intercepts, is meaningful.

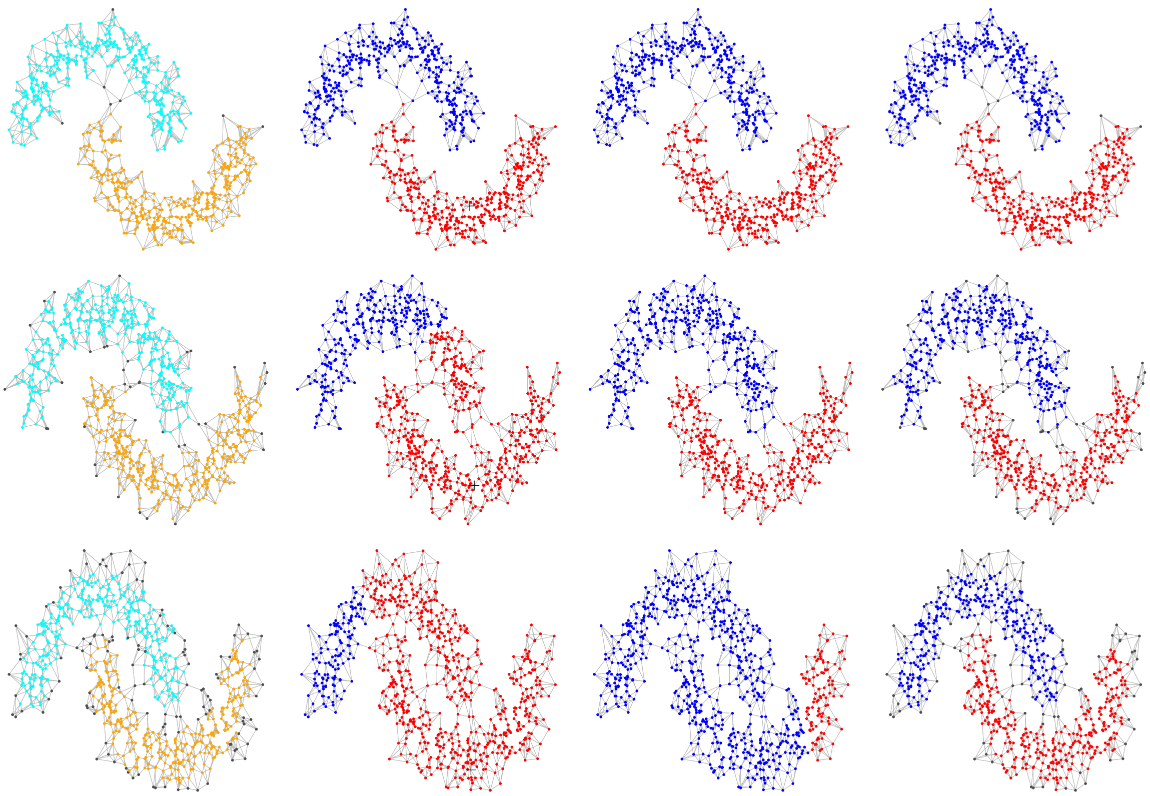


Figure 3. *True density (column 1), PPR (column 2), normalized cut (column 3) and estimated density (column 4) clusters for 3 different simulated data sets. Seed node for PPR denoted by a black cross.*

and Bresson [43]; and the last column shows a cut of the density cluster tree estimator of Chaudhuri and Dasgupta [9]. We can see the degrading ability of PPR to recover density clusters as the two moons become less well-separated. Of particular interest is the fact that PPR fails to recover one of the moons even when normalized cut still succeeds in doing so. Additionally, we note that the Chaudhuri-Dasgupta algorithm succeeds even when both PPR and normalized cut fail. This supports our main message, which is that PPR recovers only geometrically well-conditioned density clusters.

7 Discussion

There are an almost limitless number of ways to define what the "right" clustering is. In this paper, we have considered one such notion—density upper level sets—and have detailed a set of natural geometric criteria which, when appropriately satisfied, translate to provable bounds on estimation of the cluster by PPR. We have also exhibited a hard case, showing that when a density cluster is sufficiently geometrically ill-conditioned, PPR can fail to recover it. Finally, we have empirically demonstrated the tightness of our analysis for reasonable sample sizes.

Acknowledgements

SB is grateful to Peter Bickel, Martin Wainwright, and Larry Wasserman for helpful and inspiring conversations. This work was supported in part by the NSF grant DMS-1713003.

References

[1] Yasin Abbasi-Yadkori. Fast mixing random walks and regularity of incompressible vector fields. *arXiv preprint arXiv:1611.09252*, 2016.

[2] Yasin Abbasi-Yadkori, Peter Bartlett, Victor Gabillon, and Alan Malek. Hit-and-Run for Sampling and Planning in Non-Convex Spaces. In *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics*, pages 888–895, 2017.

[3] Reid Andersen and Yuval Peres. Finding sparse cuts locally using evolving sets. In *Proceedings of the Forty-first Annual ACM Symposium on Theory of Computing*, STOC '09, pages 235–244, New York, NY, USA, 2009. ACM.

[4] Reid Andersen, Fan Chung, and Kevin Lang. Local graph partitioning using pagerank vectors. In *Proceedings of the 47th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 475–486, 2006.

[5] Reid Andersen, David F Gleich, and Vahab Mirrokni. Overlapping clusters for distributed computation. In *Proceedings of the fifth ACM international conference on Web search and data mining*, pages 273–282. ACM, 2012.

[6] Sivaraman Balakrishnan, Min Xu, Akshay Krishnamurthy, and Aarti Singh. Noise thresholds for spectral clustering. In *Advances in Neural Information Processing Systems 24*. Curran Associates, Inc., 2011.

[7] Sivaraman Balakrishnan, Srivatsan Narayanan, Alessandro Rinaldo, Aarti Singh, and Larry Wasserman. Cluster trees on manifolds. In *Advances in Neural Information Processing Systems 26*, pages 2679–2687, USA, 2013. Curran Associates, Inc.

[8] Mikhail Belkin and Partha Niyogi. Convergence of laplacian eigenmaps. In *Advances in Neural Information Processing Systems 19*, pages 129–136. Curran Associates, Inc., 2007.

[9] Kamalika Chaudhuri and Sanjoy Dasgupta. Rates of convergence for the cluster tree. In *Advances in Neural Information Processing Systems 23*, pages 343–351. Curran Associates, Inc., 2010.

[10] Aaron Clauset, Cristopher Moore, and MEJ Newman. Hierarchical structure and the prediction of missing links in networks. *Nature*, 453(7191):98–102, 2008.

[11] Martin Dyer, Alan Frieze, and Ravi Kannan. A random polynomial-time algorithm for approximating the volume of convex bodies. *Journal of the ACM (JACM)*, 38(1):1–17, 1991.

[12] Shayan Oveis Gharan and Luca Trevisan. Approximating the expansion profile and almost optimal local graph clustering. In *Proceedings of the 2012 IEEE 53rd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 187–196. IEEE, 2012.

[13] David F Gleich and C Seshadhri. Vertex neighborhoods, low conductance cuts, and good seeds for local community methods. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 597–605. ACM, 2012.

[14] Stephen Guattery and Gary L Miller. On the performance of spectral graph partitioning methods. In *SODA*, volume 95, pages 233–242, 1995.

[15] John A. Hartigan. Consistency of single-linkage for high-density clusters. *Journal of the American Statistical Association*, 1981.

[16] Taher H Haveliwala. Topic-sensitive pagerank: A context-sensitive ranking algorithm for web search. *IEEE transactions on knowledge and data engineering*, 15(4):784–796, 2003.

[17] Matthias Hein and Thomas Bühler. An inverse power method for nonlinear eigenproblems with applications in 1-spectral clustering and sparse pca. In *Advances in Neural Information Processing Systems 23*, pages 847–855, 2010.

[18] Ravi Kannan, László Lovász, and Miklós Simonovits. Random walks and an o*(n5) volume algorithm for convex bodies. *Random Structures & Algorithms*, 11(1):1–50, 1997.

[19] Ravi Kannan, Santosh Vempala, and Adrian Vetta. On clusterings: Good, bad and spectral. *Journal of the ACM (JACM)*, 51(3):497–515, May 2004.

[20] Ravi Kannan, László Lovász, and Ravi Montenegro. Blocking conductance and mixing in random walks. *Combinatorics, Probability and Computing*, 15(4):541–570, 2006.

[21] Vladimir Koltchinskii and Evarist Gine. Random matrix approximation of spectra of integral operators. *Bernoulli*, 6(1):113–167, 02 2000.

[22] Aleksandr P. Korostelev and Alexandre B. Tsybakov. *Minimax theory of image reconstruction.* Springer, 1993.

[23] Samory Kpotufe and Ulrike von Luxburg. Pruning nearest neighbor cluster trees. In *Proceedings of the 28th International Conference on International Conference on Machine Learning*, ICML'11, pages 225–232, 2011.

[24] Jing Lei and Alessandro Rinaldo. Consistency of spectral clustering in stochastic block models. *Ann. Statist.*, 43(1):215–237, 02 2015.

[25] Jure Leskovec, Kevin J. Lang, and Michael Mahoney. Empirical comparison of algorithms for network community detection. In *Proceedings of the 19th International Conference on World Wide Web*, 2010.

[26] Tianxi Li, Lihua Lei, Sharmodeep Bhattacharyya, Purnamrita Sarkar, Peter J Bickel, and Elizaveta Levina. Hierarchical community detection by recursive partitioning. *arXiv preprint arXiv:1810.01509*, 2018.

[27] László Lovász and Miklós Simonovits. The mixing rate of markov chains, an isoperimetric inequality, and computing the volume. In *Proceedings of the 31st annual symposium on foundations of computer science (FOCS)*, pages 346–354. IEEE, 1990.

[28] Michael W. Mahoney, Lorenzo Orecchia, and Nisheeth K. Vishnoi. A local spectral method for graphs: with applications to improving graph partitions and exploring data graphs locally. *Journal of Machine Learning Research*, 13:2339–2365, 2012.

[29] Frank McSherry. Spectral partitioning of random graphs. In *Proceedings of the 42nd IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 529–537, 2001.

[30] Ravi Montenegro. *Faster mixing by isoperimetric inequalities.* PhD thesis, Yale University, 2002.

[31] Ben Morris and Yuval Peres. Evolving sets, mixing and heat kernel bounds. *Probability Theory and Related Fields*, 133(2):245–266, 2005.

[32] Wolfgang Polonik. Measuring mass concentrations and estimating density contour clusters-an excess mass approach. *Ann. Statist.*, 23(3):855–881, 1995.

[33] Philippe Rigollet and Régis Vert. Optimal rates for plug-in estimators of density level sets. *Bernoulli*, 15(4):1154–1178, 2009.

[34] Karl Rohe, Sourav Chatterjee, and Bin Yu. Spectral clustering and the high-dimensional stochastic blockmodel. *Ann. Statist.*, 39(4):1878–1915, 08 2011.

[35] Lorenzo Rosasco, Mikhail Belkin, and Ernesto De Vito. On learning with integral operators. *Journal of Machine Learning Research*, 11(Feb):905–934, 2010.

[36] Geoffrey Schiebinger, Martin J. Wainwright, and Bin Yu. The geometry of kernelized spectral clustering. *Ann. Statist.*, 43(2):819–846, 04 2015.

[37] Jianbo Shi and Jitendra Malik. Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8), 2000.

[38] Tao Shi, Mikhail Belkin, and Bin Yu. Data spectroscopy: Eigenspaces of convolution operators and clustering. *Ann. Statist.*, 37(6B):3960–3984, 12 2009.

[39] Aarti Singh, Clayton Scott, and Robert Nowak. Adaptive hausdorff estimation of density level sets. *Ann. Statist.*, 37(5B):2760–2782, 10 2009.

[40] Daniel A Spielman and Shang-Hua Teng. Spectral sparsification of graphs. *SIAM Journal on Computing*, 40(4):981–1025, 2011.

[41] Daniel A Spielman and Shang-Hua Teng. A local clustering algorithm for massive graphs and its application to nearly linear time graph partitioning. *SIAM Journal on Computing*, 42(1):1–26, 2013.

[42] Daniel A Spielman and Shang-Hua Teng. Nearly linear time algorithms for preconditioning and solving symmetric, diagonally dominant linear systems. *SIAM Journal on Matrix Analysis and Applications*, 35(3):835–885, 2014.

[43] Arthur Szlam and Xavier Bresson. Total variation and cheeger cuts. In *Proceedings of the 27th International Conference on International Conference on Machine Learning*, ICML'10, pages 1039–1046, 2010.

[44] Nicolás García Trillos, Dejan Slepčev, James Von Brecht, Thomas Laurent, and Xavier Bresson. Consistency of cheeger and ratio graph cuts. *Journal of Machine Learning Research*, 17(1):6268–6313, 2016.

[45] Nicolas Garcia Trillos, Franca Hoffmann, and Bamdad Hosseini. Geometric structure of graph laplacian embeddings. *arXiv preprint arXiv:1901.10651*, 2019.

[46] Nicolás García Trillos and Dejan Slepcev. A variational approach to the consistency of spectral clustering. *Applied and Computational Harmonic Analysis*, 45(2):239–281, 2018.

[47] Alexandre B Tsybakov. On nonparametric estimation of density level sets. *Ann. Statist.*, 25(3):948–969, 1997.

[48] Santosh Vempala. Geometric random walks: a survey. *Combinatorial and computational geometry*, 52(2), 2005.

[49] Ulrike von Luxburg, Mikhail Belkin, and Olivier Bousquet. Consistency of spectral clustering. *Ann. Statist.*, 36(2):555–586, 04 2008.

[50] Xiao-Ming Wu, Zhenguo Li, Anthony M. So, John Wright, and Shih fu Chang. Learning with partially absorbing random walks. In *Advances in Neural Information Processing Systems 25*, pages 3077–3085. Curran Associates, Inc., 2012.

[51] Zeyuan Allen Zhu, Silvio Lattanzi, and Vahab S Mirrokni. A local algorithm for finding well-connected clusters. In *Proceedings of the 30th International Conference on International Conference on Machine Learning*, pages 396–404, 2013.

A Proof of Theorem 3

To ease notation, letting $S \subseteq X$ and $\mathcal{S} \subseteq \mathbb{R}^d$, we write

$$\mathrm{cut}_{n,r}(S) = \mathrm{cut}(\mathcal{C}_\sigma[X]; G_{n,r}),\ \mathrm{cut}_{\mathbb{P},r}(\mathcal{S}) = \frac{\mathbb{E}[\mathrm{cut}_{n,r}(\mathcal{S})]}{2\binom{n}{2}}$$

for the random variable and mean of cut size, respectively.

With this notation in place, the goal of Theorem 3 is to show that for a universal constant $c_1 > 0$,

$$\Phi_{n,r}(\mathcal{C}_\sigma[X]) := \frac{\mathrm{cut}_{n,r}(\mathcal{C}_\sigma[X])}{\min\{\mathrm{vol}_{n,r}(\mathcal{C}_\sigma[X]), \mathrm{vol}_{n,r}((\mathbb{R}^d \setminus \mathcal{C}_\sigma)[X])\}} \leq c_1 \frac{d}{\sigma} \frac{\lambda}{\lambda_\sigma} \frac{(\lambda_\sigma - c_0 \frac{r^\gamma}{\gamma+1})}{\lambda_\sigma}$$

with probability at least $1 - 3\exp\{-nb\}$.

The proof of this theorem follows essentially from two technical Lemmas. Lemma 36 relates the terms in the numerator and denominator of $\Phi_{n,r}(\mathcal{C}_\sigma[X])$ to their expected values. We restate the conclusions of this Lemma: for any $\delta > 0$,

$$\frac{\mathrm{cut}_{n,r}(\mathcal{C}_\sigma[X])}{2\binom{n}{2}} \leq (1+\delta)\mathrm{cut}_{\mathbb{P},r}(\mathcal{C}_\sigma), \quad \frac{\mathrm{vol}_{n,r}(\mathcal{C}_\sigma[X])}{2\binom{n}{2}} \geq (1-\delta)\mathrm{vol}_{\mathbb{P},r}(\mathcal{C}_\sigma)$$
$$\frac{\mathrm{vol}_{n,r}((\mathbb{R}^d \setminus \mathcal{C}_\sigma)[X])}{2\binom{n}{2}} \geq (1-\delta)\mathrm{vol}_{\mathbb{P},r}(\mathbb{R}^d \setminus \mathcal{C}_\sigma) \tag{20}$$

with probability at least

$$1 - \exp\left\{-n\delta^2(\mathrm{cut}_{\mathbb{P},r}(\mathcal{C}_\sigma))^2\right\} - \exp\left\{-n\delta^2(\mathrm{vol}_{\mathbb{P},r}(\mathcal{C}_\sigma))^2\right\} - \exp\left\{-n\delta^2(\mathrm{vol}_{\mathbb{P},r}(\mathbb{R}^d \setminus \mathcal{C}_\sigma))^2\right\}$$
$$\geq 1 - 3\exp\left\{-n\delta^2(\mathrm{cut}_{\mathbb{P},r}(\mathcal{C}_\sigma))^2\right\}.$$

We note that as a consequence of (A4) we have that $\mathrm{vol}_{\mathbb{P},r}(\mathbb{R}^d \setminus \mathcal{C}_\sigma) \geq \mathrm{vol}_{\mathbb{P},r}(\mathcal{C}_\sigma)$, so it will suffice to lower bound $\mathrm{vol}_{\mathbb{P},r}(\mathcal{C}_\sigma)$ (since a lower bound for $\mathrm{vol}_{\mathbb{P},r}(\mathbb{R}^d \setminus \mathcal{C}_\sigma)$ follows). The following result provides upper and lower bounds on the expected values $\mathrm{cut}_{\mathbb{P},r}(\mathcal{C}_\sigma)$ and $\mathrm{vol}_{\mathbb{P},r}(\mathcal{C}_\sigma)$ respectively:

Lemma 1. *Under the setup and conditions of Theorem 3, and for any* $0 < r \leq \sigma/(2d)$,

$$\mathrm{cut}_{\mathbb{P},r}(\mathcal{C}_\sigma) \leq \frac{4d\nu_d r^{d+1}\lambda}{\sigma}\left(\lambda_\sigma - c_0\frac{r^\gamma}{\gamma+1}\right)\nu(\mathcal{C}_\sigma), \tag{21}$$

$$\mathrm{vol}_{\mathbb{P},r}(\mathcal{C}_\sigma) \geq \frac{12}{25}\lambda_\sigma^2\nu_d r^d \nu(\mathcal{C}_\sigma). \tag{22}$$

Taking Lemma 1 and (20) as given we can now complete the proof of the theorem. We lower bound $\Phi_{n,r}(\mathcal{C}_\sigma[X])$ as follows:

$$\Phi_{n,r}(\mathcal{C}_\sigma[X]) \leq \frac{(1+\delta)\mathrm{cut}_{\mathbb{P},r}(\mathcal{C}_\sigma)}{(1-\delta)\mathrm{vol}_{\mathbb{P},r}(\mathcal{C}_\sigma)} \leq \frac{25(1+\delta)d\nu_d r\lambda\left(\lambda_\sigma - c_0\frac{r^\gamma}{\gamma+1}\right)}{3(1-\delta)\sigma\lambda_\sigma^2}.$$

Plugging in $\delta = \frac{1}{2}$, the theorem is satisfied by choosing constants $c_1 = 50$ and $b = \frac{1}{4}(\mathrm{cut}_{\mathbb{P},r}(\mathcal{C}_\sigma))^2$.

A.1 Proof of Lemma 1

We write $\mathbb{P}(\mathcal{A}) = \int_{\mathcal{A}} f(x)dx$ for measurable $\mathcal{A} \subseteq \mathbb{R}^d$. We let $\mathcal{C}_{\sigma,\sigma+r} := \{x : 0 < \mathrm{dist}(x, \mathcal{C}_\sigma) < r\}$, where $\mathcal{C}_\sigma$ is as in Theorem 3. Our goal will be to upper bound $\mathrm{cut}_{\mathbb{P},r}(\mathcal{C}_\sigma)$ by a term that depends on the probability mass $\mathbb{P}(\mathcal{C}_{\sigma,\sigma+r})$, and the bulk of our technical effort will be devoted to showing the following upper bound on $\mathbb{P}(\mathcal{C}_{\sigma,\sigma+r})$:

Lemma 2. *Under the conditions of Theorem 3, and for any* $0 < r \leq \sigma/(2d)$,

$$\mathbb{P}(\mathcal{C}_{\sigma,\sigma+r}) \leq \frac{2dr}{\sigma}\left(\lambda_\sigma - c_0\frac{r^\gamma}{\gamma+1}\right)\nu(\mathcal{C}_\sigma)$$

Define the *uniform local conductance* $\ell_{\nu,r}(u)$ to be

$$\ell_{\nu,r}(u) = \nu\big(\mathcal{C}_\sigma \cap B(u,r)\big).$$

In order to lower bound $\mathrm{vol}_{\mathbb{P},r}(\mathcal{C}_\sigma)$ we will show the following lower bound on the uniform local conductance:

Lemma 3. *Let $u \in \mathcal{C}_\sigma$. Then, for any* $0 < r \leq \frac{\sigma}{2\sqrt{d}}$,

$$\ell_{\nu,r}(u) \geq \frac{6}{25}\nu_d r^d.$$

Taking these two results as given we can now prove each of the two claims (21) and (22) in turn.

Proof of Claim (21): For each i, j such that $i \neq j$, we can write

$$\mathrm{cut}_{\mathbb{P},r}(\mathcal{C}_\sigma) = \mathbb{P}(x_i \notin \mathcal{C}_\sigma, x_j \in \mathcal{C}_\sigma, \|x_i - x_j\| \leq r).$$

Writing this as an integral, we have

$$\begin{aligned}\mathrm{cut}_{\mathbb{P},r}(\mathcal{C}_\sigma) &= \int_{\mathbb{R}^d\setminus\mathcal{C}_\sigma} f(x)\mathbb{P}\big(B(x,r)\cap\mathcal{C}_\sigma\big)\,dx \\ &= \int_{\mathcal{C}_{\sigma,\sigma+r}} f(x)\mathbb{P}\big(B(x,r)\cap\mathcal{C}_\sigma\big)\,dx \\ &\leq \nu_d r^d\lambda\int_{\mathcal{C}_{\sigma,\sigma+r}} f(x)\,dx = \nu_d r^d\lambda\mathbb{P}(\mathcal{C}_{\sigma,\sigma+r}).\end{aligned}$$

where the inequality follows from (A2), which implies $f(x) \leq \lambda$ for $x \in \mathcal{C}_\sigma \setminus \mathcal{C}$. Then, upper bounding the integral using Lemma 2 gives the final result.

Proof of Claim (22): For each i, j such that $i \neq j$, we can write

$$\mathrm{vol}_{\mathbb{P},r}(\mathcal{C}_\sigma) = \mathbb{P}(x_i \in \mathcal{C}_\sigma, x_j \in B(x_i, r))$$

Writing this as an integral, we have

$$\begin{aligned}\mathrm{vol}_{\mathbb{P},r}(\mathcal{C}_\sigma) &= 2\int_{\mathcal{C}_\sigma} f(x)\mathbb{P}(B(x,r))\,dx \\ &\geq 2\int_{\mathcal{C}_\sigma} f(x)\mathbb{P}(B(x,r)\cap \mathcal{C}_\sigma)\,dx\end{aligned}$$

whence the claim then follows by Lemma 3. It remains to prove Lemmas 2 and 3, and we turn our attention to this next.

A.2 Proof of Lemma 2

The proof of this Lemma relies on certain volume estimates whose statement and proof we defer to Appendix A.4. We partition $\mathcal{C}_{\sigma,\sigma+r}$ into slices based on distance from $\mathcal{C}_\sigma$ as follows: for $k \in \mathbb{N}$,

$$\mathcal{T}_{i,k} = \left\{x \in \mathcal{C}_{\sigma,\sigma+r} : t_{i,k} < \frac{\mathrm{dist}(x, \mathcal{C}_\sigma)}{r} \leq t_{i+1,k}\right\}, \quad \mathcal{C}_{\sigma,\sigma+r} = \bigcup_{i=0}^{k-1} \mathcal{T}_{i,k},$$

where $t_{i,k} = i/k$ for $i = 0, \ldots, k-1$. As a result, for any $k \in \mathbb{N}$,

$$\mathbb{P}(\mathcal{C}_{\sigma,\sigma+r}) = \int_{\mathcal{C}_{\sigma,\sigma+r}} f(x)\,dx = \sum_{i=0}^{k-1}\int_{\mathcal{T}_{i,k}} f(x)\,dx \leq \sum_{i=0}^{k-1} \nu(\mathcal{T}_{i,k}) \max_{x\in\mathcal{T}_{i,k}} f(x). \tag{23}$$

Assumptions (A1) and (A2) imply the upper bound

$$\max_{x\in\mathcal{T}_{i,k}} f(x) \leq \lambda_\sigma - c_0(rt_{i,k})^\gamma,$$

and writing

$$\nu(\mathcal{T}_{i,k}) = \nu(\mathcal{C}_\sigma + rt_{i+1,k}B) - \nu(\mathcal{C}_\sigma + rt_{i,k}B) =: \nu_{i+1,k} - \nu_{i,k},$$

we have

$$\begin{aligned}\sum_{i=0}^{k-1} \nu(\mathcal{T}_{i,k}) \max_{x\in\mathcal{T}_{i,k}} f(x) &\leq \sum_{i=0}^{k-1}\left\{\nu_{i+1,k} - \nu_{i,k}\right\}\left(\lambda_\sigma - c_0(rt_{i,k})^\gamma\right) \\ &= \underbrace{\sum_{i=1}^{k} \nu_{i,k}\left([\lambda_\sigma - c_0(rt_{i-1,k})^\gamma] - [\lambda_\sigma - c_0(rt_{i,k})^\gamma]\right)}_{:=\Sigma_k} + \underbrace{\left(\nu_{k,k}\left[\lambda_\sigma - c_0 r^\gamma\right] - \nu_{1,k}\lambda_\sigma\right)}_{:=\xi}\end{aligned} \tag{24}$$

where the second equality comes from rearranging terms in the sum.

We first consider the term Σ_k. $\mathcal{C}$ has finite diameter by Assumption (A1), as otherwise $\int_{\mathcal{C}_\sigma} f(x)dx = \infty$. Letting $\overline{\mathcal{C}}$ be the closure of $\mathcal{C}$, we observe that $\overline{\mathcal{C}_\sigma} = \overline{\mathcal{C}} + \sigma B$, and moreover for any $\delta > 0$, $\nu(\overline{\mathcal{C}_\sigma} + \delta B) = \nu(\mathcal{C}_\sigma + \delta B)$ (as the boundary $\partial(\mathcal{C}_\sigma + \delta B)$ is Lipschitz and therefore has measure zero). As a result, for each $t_{i,k}, i = 1, \ldots, k$ we may apply Lemma 4 to $\overline{\mathcal{C}}$ and obtain

$$\nu_{i,k} = \nu(\mathcal{C}_\sigma + rt_{i,k}B) \le \nu(\mathcal{C}_\sigma)\left(1 + \frac{rt_{i,k}}{\sigma}\right)^d \tag{25}$$

which in turn gives

$$\begin{aligned}\Sigma_k &\le c_0\nu(\mathcal{C}_\sigma)r^\gamma \sum_{i=1}^{k}\left(1 + \frac{rt_{i,k}}{\sigma}\right)^d \left((t_{i,k})^\gamma - (t_{i-1,k})^\gamma\right) \\ &= c_0\nu(\mathcal{C}_\sigma)r^\gamma \sum_{i=1}^{k}\left(1 + \frac{ru_{i,k}^{1/\gamma}}{\sigma}\right)^d (u_{i,k} - u_{i,k-1}). \qquad \text{(substituting } u_{i,k} := t_{i,k}^\gamma\text{)}\end{aligned} \tag{26}$$

The expression in (26) is a Riemann sum, and taking the limit as $k \to \infty$ we obtain

$$\begin{aligned}\lim_{k\to\infty} c_0\nu(\mathcal{C}_\sigma)r^\gamma \sum_{i=1}^{k}\left(1 + \frac{ru_{i,k}^{1/\gamma}}{\sigma}\right)^d (u_{i,k} - u_{i,k-1}) &= c_0\nu(\mathcal{C}_\sigma)r^\gamma \int_0^1 \left(1 + \frac{ru^{1/\gamma}}{\sigma}\right)^d du \\ &\overset{\text{(i)}}{\le} c_0\nu(\mathcal{C}_\sigma)r^\gamma \int_0^1 \left(1 + \frac{2dru^{1/\gamma}}{\sigma}\right) du \\ &= c_0\nu(\mathcal{C}_\sigma)r^\gamma \left(1 + \gamma\frac{2dr}{(\gamma+1)\sigma}\right).\end{aligned} \tag{27}$$

where (i) follows from the upper bound in Lemma 6 in light of the fact $r \le \sigma/(2d)$.

An upper bound on ξ follows from largely the same logic, although it does not involve integration:

$$\begin{aligned}\xi &\overset{\text{(ii)}}{\le} \nu(\mathcal{C}_\sigma)\left\{\left(1 + \frac{r}{\sigma}\right)^d (\lambda_\sigma - c_0 r^\gamma) - \lambda_\sigma\right\} \\ &\overset{\text{(iii)}}{\le} \nu(\mathcal{C}_\sigma)\left\{\left(1 + \frac{2dr}{\sigma}\right)(\lambda_\sigma - c_0 r^\gamma) - \lambda_\sigma\right\} = \nu(\mathcal{C}_\sigma)\left\{\frac{2dr}{\sigma}(\lambda_\sigma - c_0 r^\gamma) - c_0 r^\gamma\right\}.\end{aligned} \tag{28}$$

where (ii) follows from (25), and (iii) from Lemma 6. As the bounds in (23) and (24) hold for all k, these along with (27) and (28) imply the desired result.

A.3 Proof of Lemma 3

The proof of this result relies on estimates of the volume of spherical caps which we defer to Appendix A.5. Since $u \in \mathcal{C}_\sigma$ there exists $x \in \mathcal{C}$ such that $u \in B(x, \sigma)$, and as $B(x, \sigma) \subseteq \mathcal{C}_\sigma$,

$$\nu\big(B(u, r) \cap B(x, \sigma)\big) \le \nu\big(B(u, r) \cap \mathcal{C}_\sigma\big)$$

Without loss of generality, let $\|u - x\| = \sigma$; it is not hard to see that if $\|u - x\| < \sigma$, the volume of the overlap will only grow. Then, since $\|u - x\| = \sigma$, $B(u, r) \cap B(x, \sigma)$ contains a spherical cap of radius r and height

$$h = r - r^2/2\sigma = r\left(1 - \frac{r}{2\sigma}\right)$$

which by Lemma 7 has volume

$$\nu_{\text{cap}} = \frac{1}{2}\nu_d r^d I_{1-\alpha}\left(\frac{d+1}{2}, \frac{1}{2}\right)$$

with $\alpha = 1 - \frac{2rh-h^2}{r^2} = \frac{r^2}{4\sigma^2} \leq \frac{1}{16d}$. By Lemmas 8 (applied with $t = 1$) and 9,

$$
\begin{aligned}
I_{1-\alpha}\left(\frac{d+1}{2}, \frac{1}{2}\right) &\geq 1 - \frac{\Gamma\left(\frac{d}{2}+1\right)}{\Gamma\left(\frac{d+1}{2}\right)\Gamma\left(\frac{1}{2}\right)} \frac{3}{4\sqrt{d}} \\
&\geq 1 - \frac{3}{4}\sqrt{\frac{d+2}{\pi d}} \geq 1 - \frac{3}{4}\sqrt{\frac{3}{2\pi}}.
\end{aligned}
$$

A.4 Volume Estimates

We begin by recalling some notation. We let $\mathcal{A} \subseteq \mathbb{R}^d$, and for $\sigma \geq 0$, write $\sigma B := B(0,\sigma) = \left\{x \in \mathbb{R}^d : \|x\| \leq \sigma\right\}$ for the closed ball of radius σ centered at the origin (and let $B^\circ(0,\sigma)$ denote the corresponding open ball). Let $\mathcal{A}_\sigma = \mathcal{A} + \sigma B$ be the direct sum of $\mathcal{A}$ and σB, $\mathcal{A}_\sigma = \{z = x + y : x \in \mathcal{A}, y \in \sigma B\}$. Recall that we use ν for Lebesgue measure, and $\nu_d = \nu(B)$ for $B = (0,1)$.

Lemma 4 provides a bound on the ratio $\nu(\mathcal{C}_\sigma + rB)/\nu(\mathcal{C}_\sigma)$, an important intermediate quantity in bounding the ratio $\mathrm{cut}(\mathcal{C}_\sigma[X]; G_{n,r})/\mathrm{vol}(\mathcal{C}_\sigma[X]; G_{n,r})$.

Lemma 4. *If $\mathcal{A}$ is closed and bounded, then for any $\delta > 0$,*

$$
\nu(\mathcal{A}_\sigma + \delta B) \leq \left(1 + \frac{\delta}{\sigma}\right)^d \nu(\mathcal{A}_\sigma). \tag{29}
$$

Proof. We will show that for any $\epsilon > 0$,

$$
\frac{\nu(\mathcal{A}_\sigma + \delta B)}{\nu(\mathcal{A}_\sigma)} \leq \frac{(\sigma + \delta + \epsilon)^d}{\sigma^d} \tag{30}
$$

Taking the limit as $\epsilon \to 0$ results in (29).

Fix $\epsilon > 0$. Our first goal is to find a finite collection $x_1, \ldots, x_N \in \mathbb{R}^d$ (where N is a finite number that may implicitly depend on ϵ) such that

$$
\bigcup_{i=1}^N B(x_i, \sigma) \subseteq \mathcal{A}_\sigma \subset \bigcup_{i=1}^N B(x_i, \sigma + \epsilon).
$$

Note that $\mathcal{A}_\sigma$ is the direct sum of two compact sets, and is therefore itself compact. Moreover, for any $\epsilon > 0$,

$$
\mathcal{A}_\sigma \subset \bigcup_{x \in \mathcal{A}} B^\circ(x, \sigma + \epsilon)
$$

so by compactness there exists a finite subcover $x_1, \ldots, x_N \in \mathcal{A}$ such that

$$
\mathcal{A}_\sigma \subset \bigcup_{i=1}^N B^\circ(x_i, \sigma + \epsilon). \tag{31}
$$

As a direct consequence of (31), $\mathcal{A}_\sigma + \delta B \subset \bigcup_{i=1}^N B^\circ(x_i, \sigma + \epsilon + \delta)$, and by definition for every $x_i \in \mathcal{A}$, $B(x_i, \sigma) \in \mathcal{A}_\sigma$. Summarizing our findings, we have

$$
\bigcup_{i=1}^N B(x_i, \sigma) \subseteq \mathcal{A}_\sigma, \ \mathcal{A}_\sigma + \delta B \subset \bigcup_{i=1}^N B^\circ(x_i, \sigma + \delta + \epsilon). \tag{32}
$$

We next show a lower bound on $\nu(\mathcal{A}_\sigma)$. Partition $\mathcal{A}_\sigma$ using the balls $B(x_i, \sigma)$, meaning let $\mathcal{A}_\sigma^{(1)} := B(x_1, \sigma)$, $\mathcal{A}_\sigma^{(2)} := B(x_2, \sigma) \setminus B(x_1, \sigma)$, and so on, so that

$$\mathcal{A}_\sigma^{(\mathrm{i})} := B(x_i, \sigma) \setminus \bigcup_{j=1}^{i-1} \mathcal{A}_\sigma^{(j)}. \qquad (i = 1, \ldots, N)$$

Observe that $\bigcup_{i=1}^N \mathcal{A}_\sigma^{(\mathrm{i})} = \bigcup_{i=1}^N B(x_i, \sigma)$, so by (31) $\mathcal{A}_\sigma \supseteq \bigcup_{i=1}^N \mathcal{A}_\sigma^{(\mathrm{i})}$. As $\mathcal{A}_\sigma^{(1)}, \ldots, \mathcal{A}_\sigma^{(N)}$ are non-overlapping,

$$\begin{aligned} \nu(\mathcal{A}_\sigma) &\geq \sum_{i=1}^N \nu(\mathcal{A}_\sigma^{(\mathrm{i})}) \\ &= \sigma^d \nu_d \sum_{i=1}^N \frac{\nu(\mathcal{A}_\sigma^{(\mathrm{i})})}{\nu(B(x_i, \sigma))} \end{aligned}$$

We turn to proving an analogous upper bound on $\nu(\mathcal{A}_\sigma + \delta B)$. Let $\mathcal{A}_{\sigma+\epsilon+\delta}^{(1)} := B(x_1, \sigma + \delta + \epsilon)$ and

$$\mathcal{A}_{\sigma+\delta+\epsilon}^{(\mathrm{i})} := B(x_i, \sigma + \delta + \epsilon) \setminus \bigcup_{j=1}^{i-1} \mathcal{A}_{\sigma+\delta+\epsilon}^{(j)}. \qquad (i = 2, \ldots, N)$$

As $\bigcup_{i=1}^N \mathcal{A}_{\sigma+\delta+\epsilon}^{(\mathrm{i})} = \bigcup_{i=1}^N B(x_i, \sigma + \delta + \epsilon)$, by (31)

$$\mathcal{A}_\sigma + \delta B \subset \bigcup_{i=1}^N \mathcal{A}_{\sigma+\delta+\epsilon}^{(\mathrm{i})}$$

and therefore

$$\begin{aligned} \nu(\mathcal{A}_\sigma + \delta B) &\leq \sum_{i=1}^N \nu\big(\mathcal{A}_{\sigma+\delta+\epsilon}^{(\mathrm{i})}\big) \\ &= \sum_{i=1}^N \nu_d (\sigma + \delta + \epsilon)^d \frac{\nu(\mathcal{A}_{\sigma+\delta+\epsilon}^{(\mathrm{i})})}{\nu(B(x_i, \sigma + \delta + \epsilon))} \\ &\leq \nu_d (\sigma + \delta + \epsilon)^d \sum_{i=1}^N \frac{\nu(\mathcal{A}_\sigma^{(\mathrm{i})})}{\nu(B(x_i, \sigma))} \end{aligned}$$

where the last inequality follows from Lemma 5. We have shown (30), and thus the claim. □

We require Lemma 5 to prove Lemma 4.

Lemma 5. *For $i = 1, \ldots, N$ and $\mathcal{A}_\sigma^{(i)}, \mathcal{A}_{\sigma+\delta+\epsilon}^{(i)}$ as in Lemma 4,*

$$\frac{\nu(\mathcal{A}_{\sigma+\delta+\epsilon}^{(i)})}{\nu(B(x_i, \sigma + \delta + \epsilon))} \leq \frac{\nu(\mathcal{A}_\sigma^{(i)})}{\nu(B(x_i, \sigma))}$$

Proof. Let $\delta' := \delta + \epsilon$. It will be sufficient to show that

$$\left(\mathcal{A}_{\sigma+\delta'}^{(\mathrm{i})} - \{x_i\}\right) \subseteq \left(1 + \frac{\delta'}{\sigma}\right) \cdot \left(\mathcal{A}_\sigma^{(\mathrm{i})} - \{x_i\}\right)$$

since then

$$\nu(\mathcal{A}^{(\mathrm{i})}_{\sigma+\delta'}) \leq \left(1+\frac{\delta'}{\sigma}\right)^d \nu(\mathcal{A}^{(\mathrm{i})}_{\sigma}) = \frac{\nu(B(x_i,\sigma+\delta'))}{\nu(B(x_i,\sigma))}\nu(\mathcal{A}^{(\mathrm{i})}_{\sigma}).$$

Assume without loss of generality that $x_i = 0$, and let $x \in \mathcal{A}^{(\mathrm{i})}_{\sigma+\delta'}$, meaning

$$\|x\| \leq \sigma+\delta', \;\; \|x-x_j\| > \sigma+\delta' \text{ for } j=1,\dots,i-1. \tag{33}$$

Letting $x' = \frac{\sigma}{\sigma+\delta'}x$, since $\|x\| \leq \sigma+\delta'$, $\|x'\| \leq \sigma$ and therefore $x' \in B(0,\sigma)$. Additionally observe that for any $j=1,\dots,i-1$, by the triangle inequality

$$\left\|x'-x_j\right\| \geq \|x-x_j\| - \left\|x-x'\right\| > \sigma+\delta' - \frac{\delta'}{\sigma+\delta'}\|x\| \geq \sigma$$

and therefore $x' \notin B(x_j,\sigma)$ for any $j=1,\dots,i-1$. So $x' \in \mathcal{A}^{(\mathrm{i})}_{\sigma}$. □

We will need to carefully control the volume of expansion sets using the estimate in Lemma 4; Lemma 6 serves this purpose (see also, Lemma 23 in [7]).

Lemma 6. *For any* $0 \leq x \leq 1/2d$*,*

$$\begin{aligned}(1+x)^d &\leq 1+2\,dx\\ (1-x)^d &\geq 1-2\,dx.\end{aligned}$$

Proof. We take the binomial expansion of $(1+x)^d$:

$$\begin{aligned}(1+x)^d &= \sum_{k=0}^{d}\binom{d}{k}x^k\\ &= 1+\,dx+\,dx\left(\sum_{k=2}^{d}\frac{\binom{d}{k}x^{k-1}}{d}\right)\\ &\leq 1+\,dx+\,dx\left(\sum_{k=2}^{d}\frac{\binom{d}{k}}{(2d)^{k-1}d}\right) && \left(\text{since } x \leq \tfrac{1}{2d}\right)\\ &\leq 1+\,dx+\,dx\left(\sum_{k=2}^{d}\frac{1}{2^{k-1}}\right) \leq 1+2\,dx.\end{aligned}$$

The proof for the corresponding lower bound on $(1-x)^d$ is symmetric. □

A.5 Spherical Caps and Associated Estimates

In this section, we state a result for the volume of a spherical cap and derive some useful upper bounds.

Lemma 7. *Let* $\mathrm{cap}_r(h)$ *denote a spherical cap of radius* r *and height* h*. Then,*

$$\nu\big(\mathrm{cap}_r(h)\big) = \frac{1}{2}\nu_d r^d I_{1-\alpha}\left(\frac{d+1}{2};\frac{1}{2}\right)$$

where

$$\alpha := 1-\frac{2rh-h^2}{r^2}$$

and

$$I_{1-\alpha}(z,w) = \frac{\Gamma(z+w)}{\Gamma(z)\Gamma(w)} \int_0^{1-\alpha} u^{z-1}(1-u)^{w-1} du.$$

is the cumulative distribution function of a $\mathrm{Beta}(z,w)$ *distribution, evaluated at* $1-\alpha$.

The following result provides a lower bound on the Beta integral, and the result in Lemma 9 provides an upper bound on the ratio of Gamma functions.

Lemma 8. *For any* $0 \le t \le 1$ *and* $\alpha \le \frac{t^2}{16d}$,

$$\int_0^{1-\alpha} u^{(d-1)/2}(1-u)^{-1/2} du \ge \frac{\Gamma\left(\frac{d+1}{2}\right)\Gamma\left(\frac{1}{2}\right)}{\Gamma\left(\frac{d}{2}+1\right)} - \frac{3t}{4\sqrt{d}}$$

Proof. We can write

$$\int_0^{1-\alpha} u^{(d-1)/2}(1-u)^{-1/2} du = \int_0^{1} u^{(d-1)/2}(1-u)^{-1/2} du - \int_{1-\alpha}^{1} u^{(d-1)/2}(1-u)^{-1/2} du$$

The first integral is simply the beta function, with

$$B\left(\frac{d+1}{2}, \frac{1}{2}\right) := \frac{\Gamma\left(\frac{d+1}{2}\right)\Gamma\left(\frac{1}{2}\right)}{\Gamma\left(\frac{d}{2}+1\right)}.$$

To upper bound the second integral, we expand $(1-u)^{-1/2}$ around $u = 1-\alpha$, obtaining

$$(1-u)^{-1/2} \le \alpha^{-1/2} + \max_{u\in(1-\alpha,1)} \frac{\alpha}{2}(1-u)^{-3/2} = \frac{3}{2}\alpha^{-1/2}.$$

As a result,

$$\begin{aligned}
\int_{1-\alpha}^{1} u^{(d-1)/2}(1-u)^{-1/2} du &\le \frac{3}{2}\alpha^{-1/2} \int_{1-\alpha}^{1} u^{(d-1)/2} du \\
&= \frac{3}{d+1}\alpha^{-1/2}\left(1-(1-\alpha)^{(d+1)/2}\right) \\
&\overset{\text{(iv)}}{\le} \frac{3}{(d+1)}\alpha^{-1/2}(\alpha(d+1)) \\
&= 3\alpha^{1/2}.
\end{aligned}$$

where (iv) follows from Lemma 6, and the fact $\alpha \le \frac{1}{16d}$. The result follows from the condition $\alpha \le \frac{t^2}{16d}$. □

Lemma 9.

$$\frac{\Gamma\left(\frac{d}{2}+1\right)}{\Gamma\left(\frac{d+1}{2}\right)\Gamma\left(\frac{1}{2}\right)} \le \sqrt{\frac{d+2}{2\pi}}$$

The proof of Lemma 9 is straightforward and follows from the fact that $\Gamma(1/2) = \sqrt{\pi}$ and the upper bound $\Gamma(x+1)/\Gamma(x+s) \le (x+1)^{1-s}$ for $s \in [0,1]$.

B Proof of Theorem 4

Let $\widetilde{G}_{n,r} := G_{n,r}[\mathcal{C}_\sigma[X]]$; in general, we will use tilde notation to refer to quantities computed over $\mathcal{C}_\sigma$ or over the induced subgraph $\widetilde{G}_{n,r}$. The goal of Theorem 4 is to show that with high probability,

$$\tau_\infty(\widetilde{G}_{n,r}) \leq c_2 \frac{\Lambda_\sigma^4 d^3 \rho^2 L^2}{\lambda_\sigma^4 r^2} \log^2\left(\frac{\Lambda_\sigma}{\lambda_\sigma^2 r}\right) + c_3.$$

We follow a two-step approach typically used to establish upper bounds on the mixing time of Markov chains.

In the first step, we relate the mixing time $\tau_\infty(G)$ of an arbitrary undirected graph $G = (V, E)$ to expansion properties of subsets $U \subseteq V$. We now build to a formal definition of these expansion properties. First, we recall the the *cut* and *volume* functionals over a graph, and introduce the *degree* functional as well. For $u \in V$, $S \subseteq V$, and $S^c = V \setminus S$,

$$\mathrm{cut}(S;G) = \sum_{u \in S} \sum_{v \in S^c} \mathbb{1}((u,v) \in E), \quad \deg(u;G) = \sum_{v \in V} \mathbb{1}((u,v) \in E), \quad \mathrm{vol}(S;G) = \sum_{u \in S} \deg(u;G).$$

Additionally, we recall the *normalized cut* $\Phi(S;G)$, defined (as in (3)) as

$$\Phi(S;G) = \frac{\mathrm{cut}(S;G)}{\min\{\mathrm{vol}(S;G), \mathrm{vol}(S^c;G)\}}.$$

We can now formally define the expansion parameters *local spread* $s(G)$ and *conductance* $\Phi(G)$,

$$s(G) := \frac{9}{10} \cdot \min_{u \in V}\{\deg(u;G)\} \cdot \min_{u \in V}\{\pi(v)\}, \quad \Phi(G) := \min_{S \subseteq V} \Phi(S;G).$$

The following proposition establishes an upper bound on the mixing time $\tau_\infty(G)$ in terms of the local spread $s(G)$ and conductance $\Phi(G)$.

Proposition 1. *Assume* $\min_{u \in V} \deg(u;G) \geq 10$. *Then,*

$$\tau_\infty(G) \leq \frac{2}{\Phi^2(G)} \log\left(\frac{1440}{s(G)}\right) \log\left(\frac{14}{s(G)}\right) + 3\log\left(\frac{14}{s(G)}\right) + 3$$

The second step in our approach is to lower bound the local spread and conductance over the neighborhood graph $\widetilde{G}_{n,r}$. In the following result we give lower bounds for both quantities.

Proposition 2. *Under the setup and conditions of Theorem 4, there exist constants* b_3,b_4, *and* b_5 *independent of* n *such that the following statement holds true: for any* n *such that*

$$n \geq \max\left\{(\log n)^{\max\{\frac{3}{d},1\}}\left(\frac{1}{b_3}\right)^d, b_4\right\}$$

the following inequalities:

$$\min_{u \in \mathcal{C}_\sigma[X]} \deg(u; \widetilde{G}_{n,r}) \geq 10, \tag{34}$$

and

$$s(\widetilde{G}_{n,r}) \geq \frac{\lambda_\sigma^2 \nu_d r^d}{20\Lambda_\sigma}, \tag{35}$$

and

$$\Phi(\widetilde{G}_{n,r}) \geq \frac{\lambda_\sigma^2 r}{\Lambda_\sigma^2 2^{14} \rho L \sqrt{d}}, \tag{36}$$

hold with probability at least $1 - \frac{b_5}{n} - 2n\exp\left\{-\frac{2\lambda_\sigma \nu_d r^d n}{1875}\right\} - 2\exp\left\{-\frac{2}{25} n (\widetilde{\mathrm{vol}}_{\mathbb{P},r}(\mathcal{C}_\sigma))^2\right\}$.

Taking these results as given, the proof of Theorem 4 is more or less complete. Since the condition $\min_{u\in\mathcal{C}_\sigma[X]} \deg(u;\widetilde{G}_{n,r}) \geq 10$ is satisfied, we may apply Proposition 1 and obtain

$$\begin{aligned}
\tau_\infty(\widetilde{G}_{n,r}) &\leq \frac{2}{\Phi^2(\widetilde{G}_{n,r})} \log\left(\frac{1440}{s(\widetilde{G}_{n,r})}\right) \log\left(\frac{14}{s(\widetilde{G}_{n,r})}\right) + 3\log\left(\frac{14}{s(\widetilde{G}_{n,r})}\right) + 3 \\
&\leq \frac{2^{29}\Lambda_\sigma^4\rho^2L^2d^3}{\lambda_\sigma^4 r^2} \log^2\left(\frac{28800\Lambda_\sigma}{\lambda_\sigma^2 r}\right) + 3\log^2\left(\frac{280\Lambda_\sigma}{\lambda_\sigma^2 r}\right) + 3 \\
&\leq \frac{2^{34}\Lambda_\sigma^4\rho^2L^2d^3}{\lambda_\sigma^4 r^2} \log^2\left(\frac{\Lambda_\sigma}{\lambda_\sigma^2 r}\right) + 3
\end{aligned}$$

whereupon the theorem follows after an appropriate choice of universal constants $c_2 = 2^{34}, c_3 = 3$, and constants $b_2 = b_5 + \frac{7500}{\lambda_\sigma \nu_d r^d} + 25(\widetilde{\mathrm{vol}}_{\mathbb{P},r}(\mathcal{C}_\sigma))^2$ and $b_1 = \frac{1}{b_3^d} + b_4$. In the rest of this section we proceed to proving Propositions 1 and 2.

B.1 Proof of Proposition 1

We first generalize some notation from the main text. Let A be the adjacency matrix of an undirected graph $G = (V, E)$, and D the associated diagonal degree matrix. The lazy random walk over G is the Markov chain with transition probabilities given by $W := \frac{I + D^{-1}A}{2}$ and stationary distribution π with elements $\pi_u = D_{uu}/\mathrm{vol}(V;G)$. Denote the t-step probability distribution of this random walk, originating from a vertex $v \in V$, as $q^{(t)} : V \times V \to [0,1]$, $q^{(t)}(v,u) = e_v W^t e_u$.

Consider the uniform distance[2] between the distributions $q_v^{(t)} := q^{(t)}(v,\cdot)$ and π, given by

$$d_{\mathrm{unif}}(q_v^{(t)}, \pi) = \max_{u\in V}\left\{\frac{\pi(u) - q_v^{(t)}(u)}{\pi(u)}\right\}.$$

Our goal is to demonstrate that for a sufficiently large t, $d_{\mathrm{unif}}(q_v^{(t)}, \pi) \leq \frac{1}{4}$ for every $v \in V$ (see (15) to recall the definition of $\tau_\infty(G)$). To do so, we introduce the *total variation distance* between the distributions $q_v^{(t)}$ and π, given by

$$\left\|q_v^{(t)} - \pi\right\|_{\mathrm{TV}} = \sum_{u\in V}\left|q_v^{(t)}(u) - \pi(u)\right|.$$

As we will see, an upper bound on the uniform distance can be obtained from an analogous upper bound on the total variation distance. First, however, we upper bound the total variation distance $||q_v^{(t)} - \pi||_{\mathrm{TV}}$ as a function of the local spread $s(G)$, the conductance $\Phi(G)$ and the number of steps t.

Lemma 10. *For any $v \in V$, and any $0 < a \leq 1/4$,*

$$\begin{aligned}
\left\|q_v^{(t+3)} - \pi\right\|_{\mathrm{TV}} \leq \max&\left\{as(G), \frac{1}{8} + \frac{9a}{20} + \frac{1}{2\min_{u\in V}\deg(u;G)}\right\} \\
&+ \left(\frac{1}{1-2as(G)} + \frac{1}{2as(G)}\right)\left(1 - \frac{\Phi^2(G)}{2}\right)^t
\end{aligned} \quad (37)$$

[2]Note d_{unif} is not formally a distance as it is not symmetric.

As mentioned, a bound on the total variation distance implies a corresponding bound on the uniform distance d_{unif}, given by the following result:

Lemma 11. *Let* $\left\|q_v^{(t)} - \pi\right\|_{\text{TV}} \le \frac{s(G)}{14}$. *Then,*

$$d_{\text{unif}}(q_v^{(t+3)}, \pi) \le \frac{1}{4}$$

Taking these Lemmas as given, we proceed that show that for

$$\tau_1 = \frac{2}{\Phi^2(G)} \log\left(\frac{1440}{s(G)}\right) \log\left(\frac{14}{s(G)}\right) + 3\log\left(\frac{14}{s(G)}\right) + 3$$

the uniform distance $d_{\text{unif}}(q_v^{(\tau_1+3)}, \pi) \le \frac{1}{4}$, which proves the claim of Proposition 1. Fix $a = \frac{1}{18}$, and let $\tau_0 = \frac{2}{\Phi^2(G)} \log\left(\frac{80}{as(G)}\right)$, so that

$$\left(1 - \frac{\Phi^2(G)}{2}\right)^{\tau_0} \le \exp(-\tau_0 \Phi^2(G)/2) \le \frac{as(G)}{80}.$$

Recall that by assumption, $\min_{u \in V} \deg(u; G) \ge 10$. Moreover note that for $a \le 1/4$, $2as(G) < 1/2 < 1 - 2as(G)$, since $s(G) \le 9/10 < 1$. By Lemma 10 we therefore obtain

$$\left\|q_v^{(\tau_0+3)} - \pi\right\|_{\text{TV}} \le \max\left\{\frac{1}{20}, \frac{1}{8} + \frac{1}{20} + \frac{1}{20}\right\} + \left(\frac{1}{as(G)}\right)\frac{as(G)}{80} \le \frac{1}{4}.$$

It is a well-known fact (see e.g. [30]) that if $\left\|q_v^{(t)} - \pi\right\|_{\text{TV}} \le \frac{1}{4}$, then for any $0 < \epsilon < 1$, $\left\|q_v^{(t \log_2(1/\epsilon))} - \pi\right\|_{\text{TV}} \le \epsilon$. Therefore, letting $\tau_1 = (\tau_0 + 3)\log(\frac{14}{s(G)})$,

$$\left\|q_v^{(\tau_1)} - \pi\right\|_{\text{TV}} \le \frac{s(G)}{14}$$

and so by Lemma 11, $d_{\text{unif}}(q_v^{(\tau_1+3)}, \pi) \le \frac{1}{4}$. The proof of Proposition 1 is therefore complete once we have proved Lemmas 10 and 11.

B.2 Proof of Lemma 10

We generalize the notation of the previous subsection. For a starting distribution z to be specified later, and for $t \ge 0$ an integer, let $q_z^{(t)}$ be the t-step probability distribution of the lazy random walk with starting distribution z.[3]

We will find it useful to introduce the *Lovasz-Simonovits curve*, defined for any $t \in \mathbb{N}$ and starting distribution z to be $h_z^{(t)} : [0,1] \to [0,1]$,

$$h_z^{(t)}(x) = \max_{w \in \mathcal{W}_x} \left\{\sum_{u \in V} \left(q_z^{(t)}(u) - \pi(u)\right) w(u)\right\}.$$

The maximum in the preceding display is over the set of weight functions $\mathcal{W}_x$

$$\mathcal{W}_x = \left\{w : V \to [0,1] \middle| 0 \le w(u) \le 1 \;\; \forall\, u, \;\; \text{and} \;\; \sum_{u \in V} w(u)\pi(u) = x\right\}.$$

[3]We say z is a starting distribution over a graph G when $\text{supp}(z) \subseteq V$ and $\sum_{u \in V} z(u) = 1$. Then, $q_z^{(t)} = zW^t$.

The utility of the Lovasz-Simonovits curve is demonstrated by the following observations. First, note that

$$\left\| q_v^{(t)} - \pi \right\|_{\mathrm{TV}} = \sup_{S \subseteq V} \left| q_v^{(t)}(S) - \pi(S) \right|$$

where we use the standard notation $\pi(S) := \sum_{u \in S} \pi(u)$, and likewise for $q_v^{(t)}(S)$. Moreover, observe that for any $S \subseteq V$ and any integer $t \geq 0$,

$$\left| q_v^{(t)}(S) - \pi(S) \right| \leq \max\left\{ h_v^{(t)}(\pi(S)), h_v^{(t)}(1 - \pi(S)) \right\} \tag{38}$$

(To see this, consider the weight functions $w(u) = \mathbb{1}(u \in S)$ and $w'(u) = 1 - w(u)$.) Taking the maximum on both sides of (38) over all $S \subseteq V$, we have

$$\left\| q_v^{(t)} - \pi \right\|_{\mathrm{TV}} \leq \max_{0 \leq x \leq 1} h_v^{(t)}(x).$$

Now, take $z = e_v W^3$, and trivially observe that $h_v^{(t+3)}(x) = h_z^{(t)}(x)$. To prove Lemma 10, it is therefore sufficient to show the desired upper bound (37) holds with respect to $h_z^{(t)}(x)$ for all $x \in [0,1]$, and all starting distributions $z = e_v W^3$.

Let $\mu = as(G)$, and note under the condition $a < 1/4$, $\mu < 1 - \mu$. To show the desired upper bound, we split the interval $[0,1]$ into the subinterval $[\mu, 1-\mu]$ and the remainder $[0,\mu) \cup (1-\mu, 1]$. Let $\ell_\mu(x)$ be the linear interpolator between $h_z^{(0)}(\mu)$ and $h_z^{(0)}(1-\mu)$,

$$\ell_\mu(x) = \frac{1 - \mu - x}{1 - 2\mu} h_z^{(0)}(\mu) + \frac{x - \mu}{1 - 2\mu} h_z^{(0)}(1 - \mu).$$

The following technical Lemma collects the upper bounds which hold over $[\mu, 1-\mu]$ and $[0,\mu) \cup (1-\mu, 1]$, respectively:

Lemma 12. *For any $\mu \leq x \leq 1 - \mu$ and any starting distribution $z = e_v W^3$,*

$$h_z^{(t)}(x) \leq \ell_\mu(x) + \left(\frac{1}{1 - 2\mu} + \frac{1}{\mu} \right) \left(1 - \frac{\Phi^2(G)}{2} \right)^t. \tag{39}$$

For any $0 \leq x < \mu$ or $1 - \mu < x \leq 1$

$$h_z^{(t)}(x) \leq \max\left\{ as(G), \frac{1}{2^{t+3}} + \frac{9a}{20} + \frac{1}{2 \min_{u \in V} \deg(u;G)} \right\} \tag{40}$$

Taking Lemma 12 as given, we have nearly completed our proof of Lemma 10. Note that for any $\mu \leq x \leq 1 - \mu$,

$$\begin{aligned} \ell_\mu(x) &\leq \max\left\{ h_z^{(t)}(\mu), h_z^{(t)}(1-\mu) \right\} \\ &\leq \max\left\{ as(G), \frac{1}{2^{t+3}} + \frac{9a}{20} + \frac{1}{2 \min_{u \in V} \deg(u;G)} \right\} \end{aligned}$$

with the latter inequality following from (40). Therefore, for any $x \in [0,1]$,

$$h_z^{(t)}(x) \leq \max\left\{ as(G), \frac{1}{8} + \frac{9a}{20} + \frac{1}{2 \min_{u \in V} \deg(u;G)} \right\} + \left(\frac{1}{1 - 2as(G)} + \frac{1}{as(G)} \right) \left(1 - \frac{\Phi^2(G)}{2} \right)^t$$

which is exactly the claimed result of Lemma 10.

B.3 Proof of Lemma 12

We first prove (39), and then (40).

Proof of (39): The desired result is a consequence of Theorem 1.2 of [27]. To state this theorem, we introduce the notation

$$C_\mu = \max\left\{ \frac{h_z^{(0)}(x) - \ell_\mu(x)}{\sqrt{x-\mu}}, \frac{h_z^{(0)}(x) - \ell_\mu(x)}{\sqrt{1-x-\mu}} : \mu < x < 1-\mu \right\},$$

and then the theorem itself.

Theorem 6 (Theorem 1.2 of [27]). *For any $\mu \le x \le 1-\mu$, z an arbitrary starting distribution, and an integer $t \geq 0$,*

$$h_z^{(t)}(x) \le \ell_\mu(x) + C_\mu \min\left\{\sqrt{x-\mu}, \sqrt{1-x-\mu}\right\}\left(1 - \frac{\Phi^2(G)}{2}\right)^t \tag{41}$$

In order to show (39), we must therefore provide an appropriate bound on the quantity C_μ. Precisely, we will show that for any $\mu < x < 1-\mu$,

$$h_z^{(0)}(x) - \ell_\mu(x) \le \max\left\{ \frac{h_z^{(0)}(\mu)}{1-2\mu} + \frac{h_z^{(0)}(\mu)}{\mu}, \frac{h_z^{(0)}(1-\mu)}{1-2\mu} + 1 \right\} \min\left\{\sqrt{x-\mu}, \sqrt{1-x-\mu}\right\} \tag{42}$$

which will in turn imply

$$\begin{aligned} C_\mu &\le \max\left\{ \frac{h_z^{(0)}(\mu)}{1-2\mu} + \frac{h_z^{(0)}(\mu)}{\mu}, \frac{h_z^{(0)}(1-\mu)}{1-2\mu} + 1 \right\} \\ &\le \frac{1}{\mu} + \frac{1}{1-2\mu}. \end{aligned}$$

Plugging this upper bound into (41), we obtain

$$\begin{aligned} h_z^{(t)}(x) &\le \ell_\mu(x) + \left(\frac{1}{\mu} + \frac{1}{1-2\mu}\right) \min\left\{\sqrt{x-\mu}, \sqrt{1-x-\mu}\right\}\left(1 - \frac{\Phi^2(G)}{2}\right)^t \\ &\le \ell_\mu(x) + \left(\frac{1}{\mu} + \frac{1}{1-2\mu}\right)\left(1 - \frac{\Phi^2(G)}{2}\right)^t \end{aligned}$$

which is the desired result.

It remains to show (42). To do so, we make use of an equivalent representation of the Lovasz-Simonovits curve $h_z^{(t)}$. Order the elements of $V = \{u_1, \ldots, u_N\}$, where $N = |V|$, such that

$$\frac{q_z^{(t)}(u_1)}{\pi(u_1)} \ge \frac{q_z^{(t)}(u_2)}{\pi(u_2)} \ge \ldots \ge \frac{q_z^{(t)}(u_N)}{\pi(u_N)}, \tag{43}$$

and for each $k = 1, \ldots, N$, let $U_k = \{u_1, \ldots, u_k\}$. Then for any $x \in [0,1]$, letting k satisfy $\pi(U_{k-1}) < x < \pi(U_k)$, it can be shown that[4],

$$h_z^{(t)}(x) = \sum_{j=1}^{k-1}(q_z^{(t)}(u_j) - \pi(u_j)) + \frac{x - \pi(U_{k-1})}{\pi(u_k)}\left(q_z^{(t)}(u_k) - \pi(u_k)\right). \tag{44}$$

[4] See [27] for a formal justification.

The representation of the Lovasz-Simonovits curve given by (44) makes it clear that $h_z^{(t)}(x)$ is a piecewise linear curve, with knots at the points $x_k = \pi(U_k)$ for $k = 1, \ldots, N$, where the kth linear segment has slope $v_k = q_z^{(t)}(u_k)/\pi(u_k) - 1$. By the ordering of (43), we have that $v_1(x) > v_2(x) > \ldots > v_{N-1}(x)$, and the curve is therefore concave.

In fact, letting $v(x) = \sum_{k=1}^N v_k \mathbb{1}(x \in [U_k, U_{k+1}))$ be the slope of the Lovasz-Simonovits curve, for any $x \geq \mu$ it can be shown that

$$\begin{aligned} v(x) &\leq \min_{k: u_k \leq x} v_k \\ &\leq \frac{h_z^{(0)}(\mu)}{\mu}. \end{aligned} \tag{45}$$

From the upper bound in (45) along with the concavity of $h_z^{(0)}(x)$, we have that for any $x \geq \mu$,

$$h_z^{(0)}(x) \leq h_z^{(0)}(\mu) + (x - \mu)\frac{h_z^{(0)}(\mu)}{\mu}.$$

Some algebra then yields that for any $x \geq \mu$,

$$\begin{aligned} h_z^{(0)}(x) - \ell_\mu(x) &\leq h_z^{(0)}(\mu) - \left(\frac{1-\mu-x}{1-2\mu} h_z^{(0)}(\mu) + \frac{x-\mu}{1-2\mu} h_z^{(0)}(1-\mu)\right) + \frac{h_z^{(0)}(\mu)}{\mu}(x-\mu) \\ &= \frac{x-\mu}{1-2\mu} h_z^{(0)}(\mu) + \frac{h_z^{(0)}(\mu)}{\mu}(x-\mu) - \frac{x-\mu}{1-2\mu} h_z^{(0)}(1-\mu) \\ &\leq \sqrt{x-\mu}\left(\frac{h_z^{(0)}(\mu)}{1-2\mu} + \frac{h_z^{(0)}(\mu)}{\mu}\right) \end{aligned}$$

On the other hand, $\ell_\mu(1-\mu) = h_z^{(0)}(1-\mu)$, and by the concavity of $h_z^{(0)}$ and (45), for $x \leq 1-\mu$

$$h_z^{(0)}(x) \leq h_z^{(0)}(1-\mu) + (1-x-\mu).$$

Similar manipulations to above give the upper bound

$$h_z^{(0)}(x) - \ell_\mu(x) \leq \sqrt{1-\mu-x}\left(\frac{h_z^{(0)}(1-\mu)}{1-2\mu} + 1\right)$$

and (42) follows.

Proof of (40): We let $z = e_v W^3$ for an arbitrary $v \in V$. First, we deal with the case $x \leq as(G)$. Letting $\pi(U_{k-1}) \leq x \leq \pi(U_k)$, we have

$$h_z^{(t)}(x) \leq q_z^{(t)}(U_{k-1}) + q_z^{(t)}(u_k) \tag{46}$$

We observe a few facts about the random walk defined by $q_z^{(t)}$. By the laziness of the random walk, for $u \neq v, t \geq 1$

$$q_z^{(t)}(u) \leq \frac{1}{2\min_{u \in V} \deg(u; G)} \tag{47}$$

On the other hand if $u = v$,

$$q_z^{(t)}(u) \leq \frac{1}{2^{t+3}} + \frac{1}{2\min_{u \in V} \deg(u; G)}. \tag{48}$$

Therefore by (46), (47), and (48)

$$h_z^{(t)}(x) \leq \frac{1}{2^{t+3}} + \frac{|U_{k-1}|}{2\min_{u\in V}\deg(u;G)} + \frac{1}{2\min_{u\in V}\deg(u;G)}.$$

Finally, since $x \leq as(G)$ and $x \geq \pi(U_{k-1})$,

$$\begin{aligned}
|U_{k-1}| &\leq \frac{\pi(U_{k-1})}{\min_{u\in V}\pi(v)} \\
&\leq \frac{as(G)}{\min_{u\in V}\pi(v)} \\
&\leq \frac{9a\min_{u\in V}\deg(u;G)}{10}.
\end{aligned}$$

and the desired result is shown.

Now, we turn to the case where $x \geq 1 - as(G)$. Noting that the slope $v(x)$ of the Lovasz-Simonovits curve $h_z^{(t)}$ is bounded below by -1 for all $x \in [0,1]$, by the concavity of $h_z^{(t)}$ we have

$$\begin{aligned}
h_z^{(t)}(x) &\leq h_z^{(t)}(1) + (1-x) \\
&= 1 - x \leq as(G).
\end{aligned}$$

and the proof of Lemma 10 is complete.

B.4 Proof of Lemma 11

The proof of Lemma 11 follows from straightforward algebraic manipulation. Fix $u \in V$ and let $m \geq t+1$ be arbitrary. The stationarity of π gives (see (16) of [31])

$$\begin{aligned}
\frac{\pi(u) - q_v^m(u)}{\pi(u)} &= \sum_{y\in V}\Big(\pi(y) - q^{(m-1)}(v,y)\Big)\left(\frac{q^{(1)}(y,u) - \pi(u)}{\pi(u)}\right) \\
&\overset{\text{(i)}}{=} \sum_{y\neq u}\Big(\pi(y) - q^{(m-1)}(v,y)\Big)\left(\frac{q^{(1)}(y,u) - \pi(u)}{\pi(u)}\right) + \frac{\pi(u) - q^{(m-1)}(v,u)}{\pi(u)}\left(\frac{1}{2} - \pi(u)\right) \\
&\leq \sum_{y\neq u}\Big(\pi(y) - q^{(m-1)}(v,y)\Big)\left(\frac{q^{(1)}(y,u) - \pi(u)}{\pi(u)}\right) + \frac{\pi(u) - q^{(m-1)}(v,u)}{2\pi(u)} \qquad (49)
\end{aligned}$$

where (i) follows from $q^{(1)}(u,u) = \frac{1}{2}$. Then,

$$\begin{aligned}
\sum_{y\neq u}\Big(\pi(y) - q^{(m-1)}(v,y)\Big)\left(\frac{q^{(1)}(y,u) - \pi(u)}{\pi(u)}\right) &\leq \left\|q_v^{(m-1)} - \pi\right\|_{\mathrm{TV}} \max_{y\neq u}\left|\frac{q^{(1)}(y,u) - \pi(u)}{\pi(u)}\right| \\
&\leq \left\|q_v^{(m-1)} - \pi\right\|_{\mathrm{TV}} \max\left\{1, \max_{y\neq u}\left\{\frac{q^{(1)}(y,u)}{\pi(u)}\right\}\right\} \\
&\overset{(ii)}{\leq} \left\|q_v^{(m-1)} - \pi\right\|_{\mathrm{TV}} \max\left\{1, \frac{1}{2\pi(u)\min_{u'\in V}\deg(u';G)}\right\} \\
&\leq \left\|q_v^{(m-1)} - \pi\right\|_{\mathrm{TV}} \frac{1}{s(G)} \qquad (50)
\end{aligned}$$

where (ii) follows from the fact that for $y \neq u$, $q^{(1)}(y,u) \leq 1/\left(2\min_{u \in V} \deg(u;G)\right)$. As $m-1 \geq t$, it is well-known (see e.g. [27]) that the laziness of the random walk guarantees $\left\|q_v^{(m-1)} - \pi\right\|_{\mathrm{TV}} \leq \left\|q_v^{(t)} - \pi\right\|_{\mathrm{TV}}$, and therefore by (50) and the hypothesis of Lemma 11,

$$\sum_{y \neq u} \left(\pi(y) - q^{(m-1)}(v,y)\right)\left(\frac{q^{(1)}(y,u) - \pi(u)}{\pi(u)}\right) \leq \frac{1}{14}.$$

Plugging this in to (49) and taking maximum on both sides, we obtain

$$d_{\mathrm{unif}}(q_v^{(m)}, \pi) \leq \frac{1}{14} + \frac{d_{\mathrm{unif}}(q_v^{(m-1)}, \pi)}{2} \tag{51}$$

The recurrence relation of (51) along with the initial condition $d_{\mathrm{unif}}(q_v^{(t)}, \pi) \leq 1$ yields

$$d_{\mathrm{unif}}(q_v^{(t+1)}, \pi) \leq \frac{8}{14} \Rightarrow d_{\mathrm{unif}}(q_v^{(t+2)}, \pi) \leq \frac{5}{14} \Rightarrow d_{\mathrm{unif}}(q_v^{(t+3)}, \pi) \leq \frac{1}{4}$$

and the claim is shown. We have proved Lemmas 10 and 11, and therefore Proposition 1.

B.5 Proof of Proposition 2

We prove (34) and (35) immediately before turning our attention to (36), which will require the bulk of our attention. First, however, we introduce some notation. For $S \subseteq \mathcal{C}_\sigma[X]$ and $u \in \mathcal{C}_\sigma[X]$, we will write

$$\widetilde{\mathrm{cut}}_{n,r}(S) := \mathrm{cut}(S; \widetilde{G}_{n,r}), \quad \widetilde{\deg}_{n,r}(u) := \deg(u; \widetilde{G}_{n,r}), \quad \widetilde{\mathrm{vol}}_{n,r}(S) := \mathrm{vol}(S; \widetilde{G}_{n,r})$$

and let $\widetilde{\pi}_{n,r}(u) = \widetilde{\deg}_{n,r}(u)/\widetilde{\mathrm{vol}}_{n,r}(\mathcal{C}_\sigma[X])$ be the stationary distribution of the lazy random walk over $\widetilde{G}_{n,r}$. We also let $\widetilde{\Phi}_{n,r}(S) := \Phi(S; \widetilde{G}_{n,r})$ denote the normalized cut functional over $\widetilde{G}_{n,r}$, and $\widetilde{d}_{\min} := \min_{u \in \mathcal{C}_\sigma[X]} \widetilde{\deg}_{n,r}(u)$,

Proof of (34): Applying Lemma 36 with $\delta = 1/5$, we have that

$$\widetilde{d}_{\min} \geq \frac{24}{125} \lambda_\sigma \nu_d r^d n$$

with probability at least $1 - 2n\exp\left\{-\frac{2\lambda_\sigma \nu_d r^d n}{1875}\right\}$, and therefore for any

$$n \geq \frac{53}{\lambda_\sigma \nu_d r^d} =: b_4$$

the minimum degree $\widetilde{d}_{\min} \geq 10$.

Proof of (35): We rewrite $s(\widetilde{G}_{n,r}) = \frac{9\widetilde{d}_{\min}^2}{10\widetilde{\mathrm{vol}}_{n,r}(\mathcal{C}_\sigma[X])}$. Then applying Lemma 36 with $\delta = 1/5$, we have

$$\begin{aligned} s(\widetilde{G}_{n,r}) &\geq \frac{9}{10} \cdot \frac{\left(\frac{6}{25}\right)^2 (1-\delta)^2 \lambda_\sigma^2 \nu_d r^{2d}}{(1+\delta)\widetilde{\mathrm{vol}}_{\mathbb{P},r}(\mathcal{C}_\sigma) n^2} \\ &\geq \frac{1}{40} \frac{\lambda_{\sigma^2} \nu_d r^d}{\Lambda_\sigma} \end{aligned}$$

with probability at least $1 - 2n\exp\left\{-\frac{2\lambda_\sigma \nu_d r^d n}{1875}\right\} - 2\exp\left\{-\frac{2}{25} n (\widetilde{\mathrm{vol}}_{\mathbb{P},r}(\mathcal{C}_\sigma))^2\right\}$.

Proof of (36)**:** Roughly, our goal is to show that for sufficiently large n, with probability at least $1 - \frac{b_5}{n}$

$$\min_{S \subseteq \mathcal{C}_\sigma[X]} \widetilde{\Phi}_{n,r}(S) \geq \frac{\lambda_\sigma^2 r}{\Lambda_\sigma^2 2^{14} \rho L \sqrt{d}}.$$

In order to show this bound holds uniformly over all sets $S \subseteq \mathcal{C}_\sigma[X]$, we will split the analysis into two cases based on the size of $S \subseteq \mathcal{C}_\sigma[X]$. To do so, we introduce $\mathcal{L}(G) := \{S \subseteq V : \pi(S), \pi(S^c) \geq s(G)\}$ (where as usual π denotes the stationary distribution of a lazy random walk over G.)

Lemma 13 shows that for any subset $S \subseteq \mathcal{C}_\sigma[X]$ not in $\mathcal{L}(\widetilde{G}_{n,r})$, the graph normalized cut of S is at least $1/10$. In fact, this statement holds for any graph G.

Lemma 13. *Let $G = (V, E)$ be an arbitrary undirected graph. Then, for non-empty subsets $S \subseteq V$,*

$$\min_{S \notin \mathcal{L}(G)} \Phi(S; G) \geq \frac{1}{10}.$$

Proof. The claim follows by simple manipulations:

$$\begin{aligned}
\Phi(S;G) &\geq \frac{\mathrm{cut}(S;G)}{\mathrm{vol}(S;G)} \\
&\geq \sum_{u \in S} \frac{\deg(u;G) - |S|}{\mathrm{vol}(S;G)} \\
&\geq \sum_{u \in S} \frac{\deg(u;G) - \pi(S)/(\min_{u \in V} \pi(u))}{\mathrm{vol}(S;G)} \\
&\geq \sum_{u \in S} \frac{\deg(u;G) - \frac{9}{10}\min_{u \in V} \deg(u;G)}{\mathrm{vol}(S;G)} \\
&\geq \frac{1}{10} \sum_{u \in S} \frac{\deg(u;G)}{\mathrm{vol}(S;G)} = \frac{1}{10}.
\end{aligned}$$

□

In Lemma 14, we establish a uniform lower bound on the normalized cut for the remaining sets $S \in \mathcal{L}(\widetilde{G}_{n,r})$.

Lemma 14. *Under the setup and conditions of Theorem 4, there exist constants b_3 and b_5 independent of n such that the following statement holds true: for any n such that*

$$\frac{n}{(\log n)^{\max\{\frac{3}{d}, 1\}}} \geq \left(\frac{1}{b_3}\right)^d$$

the following upper bound holds

$$\min_{S \in \mathcal{L}(\widetilde{G}_{n,r})} \widetilde{\Phi}_{n,r}(S) \geq \frac{\lambda_\sigma^2 r}{\Lambda_\sigma^2 2^{14} \rho L \sqrt{d}}, \tag{52}$$

with probability at least $1 - \frac{b_5}{n}$.

The desired upper bound on graph conductance (36) follows from Lemma 14, along with Lemma 13, in light of the fact $\frac{\lambda_\sigma^2 r}{\Lambda_\sigma^2 2^{13} \rho L \sqrt{d}} < \frac{1}{10}$. We turn now to the proof of Lemma 14.

B.6 Proof of Lemma 14.

The proof of Lemma 14 will essentially follow from a pair of technical results. The first of these will demonstrate that the functional $\widetilde{\Phi}_{n,r}(S)$ can be lower bounded by a population analogue $\widetilde{\Phi}_{\mathbb{P},r}(\mathcal{S})$, for an appropriately chosen $\mathcal{S} \subseteq \mathcal{C}_\sigma$; we term this latter functional the continuous normalized cut. This lower bound will hold uniformly over sets $S \subseteq \mathcal{C}_\sigma[X]$. The second technical Lemma will build on known continuous space isoperimetric inequalities to lower bound the continuous normalized cut $\widetilde{\Phi}_{\mathbb{P},r}(\mathcal{S})$ uniformly over sets $\mathcal{S} \subseteq \mathcal{C}_\sigma$.

We will now build slowly toward a formal definition of the continuous normalized cut, before establishing a relation between it and it's discrete counterpart. Let $\mathcal{S} \subseteq \mathcal{C}_\sigma$ be measurable. We introduce the *r-ball walk*, a Markov chain over $\mathcal{C}_\sigma$ with transition probability at $x \in \mathcal{C}_\sigma$ given by

$$\widetilde{P}_{\mathbb{P},r}(x;\mathcal{S}) := \frac{\mathbb{P}(\mathcal{S} \cap B(x,r))}{\mathbb{P}(\mathcal{C}_\sigma \cap B(x,r))}.$$

Denote the stationary distribution for this Markov chain by $\widetilde{\pi}_{\mathbb{P},r}$, which is defined by the relation

$$\int_{\mathcal{C}_\sigma} \widetilde{P}_{\mathbb{P},r}(x;\mathcal{S})\, d\widetilde{\pi}_{\mathbb{P},r}(x) = \widetilde{\pi}_{\mathbb{P},r}(\mathcal{S}).$$

Letting the $\mathbb{P}$-*local conductance* be given by

$$\ell_{\mathbb{P},r}(x) := \mathbb{P}\bigl(\mathcal{C}_\sigma \cap B(x,r)\bigr)$$

a bit of algebra verifies that

$$\widetilde{\pi}_{\mathbb{P},r}(\mathcal{S}) = \frac{\int_{\mathcal{S}} \ell_{\mathbb{P},r}(x)\, d\mathbb{P}(x)}{\int_{\mathcal{C}_\sigma} \ell_{\mathbb{P},r}(x)\, d\mathbb{P}(x)}.$$

We next introduce the *ergodic flow*, $\widetilde{Q}_{\mathbb{P},r}$. Let $\mathcal{S} \cap \mathcal{S}^c = \mathcal{C}_\sigma$ be a partition of $\mathcal{C}_\sigma$. Then the ergodic flow between $\mathcal{S}$ and $\mathcal{S}$ is given by

$$\widetilde{Q}_{\mathbb{P},r}(\mathcal{S},\mathcal{S}^c) := \int_{\mathcal{S}} \widetilde{P}_{\mathbb{P},r}(x;\mathcal{S}^c)\, d\widetilde{\pi}_{\mathbb{P},r}(x),$$

and the *($\mathbb{P}$) continuous normalized cut* by

$$\widetilde{\Phi}_{\mathbb{P},r}(\mathcal{S}) := \frac{\widetilde{Q}_{\mathbb{P},r}(\mathcal{S},\mathcal{S}^c)}{\min\left\{\widetilde{\pi}_{\mathbb{P},r}(\mathcal{S}), \widetilde{\pi}_{\mathbb{P},r}(\mathcal{S}^c)\right\}},$$

As the functionals $\widetilde{\Phi}_{n,r}$ and $\widetilde{\Phi}_{\mathbb{P},r}$ act in the different spaces $\mathcal{C}_\sigma[X]$ and $\mathcal{C}_\sigma$, respectively, it is not obvious how to relate them. To do so, following the lead of [44], we introduce transportation maps between the space $\mathcal{C}_\sigma$ and the sample points $\mathcal{C}_\sigma[X]$. We note that by assumption (A1), $\mathbb{P}(\mathcal{C}_\sigma) > 0$, and therefore with probability one as $n \to \infty$, the number of sample points $|\mathcal{C}_\sigma[X]|$ will be non-zero as well. We may therefore define the conditional probability measures

$$\widetilde{\mathbb{P}}(\mathcal{S}) = \frac{\mathbb{P}(\mathcal{S})}{\mathbb{P}(\mathcal{C}_\sigma)},\ \widetilde{\mathbb{P}}_n(\mathcal{S}) := \frac{1}{|\mathcal{C}_\sigma[X]|}\sum_{x_i \in \mathcal{C}_\sigma[X]} \mathbb{1}(x_i \in \mathcal{S}).$$

We then define a *transportation map* between $\widetilde{\mathbb{P}}$ and $\widetilde{\mathbb{P}}_n$ to be any measurable map $T : \mathcal{C}_\sigma \to \mathcal{C}_\sigma[X]$ such that for every $S \subseteq \mathcal{C}_\sigma[X]$,

$$\widetilde{\mathbb{P}}(T^{-1}(S)) = \widetilde{\mathbb{P}}_n(S),$$

where $T^{-1}(S) = \{x \in \mathcal{C}_\sigma : T(x) \in S\}$ is the preimage of T. Observe that by the definition of the transportation map T, for any $g \in L^1\bigl(\widetilde{\mathbb{P}}_n\bigr)$ the following change of variables formula holds

$$\int_{\mathcal{C}_\sigma} g(x)\, d\widetilde{\mathbb{P}}_n(x) = \int_{\mathcal{C}_\sigma} g\bigl(T(x)\bigr)\, d\widetilde{\mathbb{P}}(x)$$

Using the change of variables formula with an appropriate choice of g, after suitable rescaling we can relate $\widetilde{\mathrm{cut}}_{n,r}(S)$ to $\widetilde{Q}_{\mathbb{P},r}(T^{-1}(S), T^{-1}(S)^c)$. Similarly, we can relate $\widetilde{\mathrm{vol}}_{n,r}(S^c)$ to $\widetilde{\pi}_{\mathbb{P},r}(T^{-1}(S))$. Working along these lines, we obtain the following lower bound on $\widetilde{\Phi}_{n,r}(S)$, stated in terms of the transportation distance $\|\mathrm{Id} - T\|_{L^\infty(\mathbb{P})}$, where $\mathrm{Id}(x) = x$ is the identity mapping over $\mathcal{C}_\sigma$.

Lemma 15. *Let $T : \mathcal{C}_\sigma \to \mathcal{C}_\sigma[X]$ be a transportation map between $\widetilde{\mathbb{P}}$ and $\widetilde{\mathbb{P}}_n$. Suppose $\|\mathrm{Id} - T\|_{L^\infty(\widetilde{\mathbb{P}})} < \min\Bigl\{s(\widetilde{G}_{n,r}), r, \lambda_\sigma/(2^{d+1} d \Lambda_\sigma r)\Bigr\}$. Then there exists a constant $b_6 > 0$ which does not depend on the sample size n, such that for all $S \in \mathcal{L}(\widetilde{G}_{n,r})$,*

$$\widetilde{\Phi}_{n,r}(S) \geq \widetilde{\Phi}_{\mathbb{P},r}(T^{-1}(S)) - \frac{b_6\, \|\mathrm{Id} - T\|_{L^\infty(\widetilde{\mathbb{P}})}}{s(\widetilde{G}_{n,r}) - b_6\, \|\mathrm{Id} - T\|_{L^\infty(\widetilde{\mathbb{P}})}} \tag{53}$$

Clearly, Lemma 15 is useful only when combined with an upper bound on the transportation distance $\|\mathrm{Id} - T_n\|_{L^\infty(\widetilde{\mathbb{P}})}$. Proposition 5 of [44] establishes such an upper bound, with respect to transportation maps on measures supported on open, connected and bounded domains with Lipschitz boundaries. The following result is a restatement of this Proposition with respect to the domain $\mathcal{C}_\sigma$, and the measure $\widetilde{\mathbb{P}}$. Although $\mathcal{C}_\sigma$ is closed rather than open, as $\nu(\partial \mathcal{C}_\sigma) = 0$ we may apply Proposition 5 of [44] to the interior $\mathcal{C}_\sigma^o$ of $\mathcal{C}_\sigma$, and the desired result will hold for any arbitrary extension of T_n to $\mathcal{C}_\sigma$. Let $\widetilde{n} = |\mathcal{C}_\sigma[X]|$.

Theorem 7 (Restatement of Proposition 5 of [44])**.** *There exists constants b_3 and b_5 which do not depend on the sample size such that for any $\delta > 0$, the following statement holds: for any n such that*

$$\frac{n}{(\log n)^{\max\{\frac{3}{d}, 1\}}} \geq \left(\frac{1}{b_3 \delta}\right)^d$$

then with probability at least $1 - \frac{b_5}{n}$ there exists a transportation map T_n between $\widetilde{\mathbb{P}}$ and $\widetilde{\mathbb{P}}_n$ which satisfies

$$\|\mathrm{Id} - T_n\|_{L^\infty(\widetilde{\mathbb{P}})} \leq \delta.$$

Lemma 15 and Theorem 7 show that with high probability, the discrete normalized cut $\widetilde{\Phi}_{n,r}(S)$ is lower bounded by the continuous normalized cut $\widetilde{\Phi}_{\mathbb{P},r}(T_n^{-1}(S))$ over all sufficiently large sets $S \subseteq X$. The following result then supplies the last step, a uniform lower bound on the continuous normalized cut $\widetilde{\Phi}_{\mathbb{P},r}(\mathcal{S})$ for all sets $\mathcal{S} \subseteq \mathcal{C}_\sigma$. Let the $\mathbb{P}$-*continuous conductance* be given by

$$\widetilde{\Phi}_{\mathbb{P},r} := \min_{\mathcal{S} \subseteq \mathcal{C}_\sigma} \widetilde{\Phi}_{\mathbb{P},r}(\mathcal{S}).$$

Lemma 16. *Under the setup and conditions of Theorem 4, the $\mathbb{P}$-continuous conductance of the r-ball walk satisfies*

$$\widetilde{\Phi}_{\mathbb{P},r} > \frac{\lambda_\sigma^2 r}{2^{13} \Lambda_\sigma^2 \rho L \sqrt{d}}.$$

With Lemmas 15 and 16, as well as Theorem 7, in hand, we proceed to the proof of Lemma 14. Fix

$$\delta = \min\left\{\frac{\lambda_\sigma^2 \nu_d r^d}{40\Lambda_\sigma}, r, \lambda_\sigma/(2^{d+1} d\Lambda_\sigma r), \frac{\widetilde{\Phi}_{\mathbb{P},r}\lambda_\sigma^2\nu_d r^d}{40(2b_6 + \widetilde{\Phi}_{\mathbb{P},r})\Lambda_\sigma}\right\};$$

and suppose n is large enough so that $\|\mathrm{Id} - T_n\|_{L^\infty(\mathbb{P})} \leq \delta$ for the optimal transportation map T_n. By (35), $\delta \leq s(\widetilde{G}_{n,r})$. Therefore, we may apply Lemma 15 and obtain

$$\begin{aligned}
\min_{S \in \mathcal{L}(\widetilde{G}_{n,r})} \widetilde{\Phi}_{n,r}(S) &\geq \min_{S \in \mathcal{L}(\widetilde{G}_{n,r})} \widetilde{\Phi}_{\mathbb{P},r}(T_n^{-1}(S)) - \frac{b_6\,\|\mathrm{Id} - T\|_{L^\infty(\widetilde{\mathbb{P}})}}{s(\widetilde{G}_{n,r}) - b_6\,\|\mathrm{Id} - T\|_{L^\infty(\widetilde{\mathbb{P}})}} \\
&\geq \widetilde{\Phi}_{\mathbb{P},r} - \frac{b_6\,\|\mathrm{Id} - T\|_{L^\infty(\widetilde{\mathbb{P}})}}{s(\widetilde{G}_{n,r}) - b_6\,\|\mathrm{Id} - T\|_{L^\infty(\widetilde{\mathbb{P}})}} \\
&\geq \frac{\widetilde{\Phi}_{\mathbb{P},r}}{2},
\end{aligned}$$

where the last inequality follows since $\delta < \frac{\widetilde{\Phi}_{\mathbb{P},r}\cdot s(\widetilde{G}_{n,r})}{2b_6 + \widetilde{\Phi}_{\mathbb{P},r}}$. Then, Lemma 16 yields the desired result. It remains to prove Lemmas 15 and 16, which we do in the following two subsections.

B.7 Proof of Lemma 15

Let $S \subseteq \mathcal{C}_\sigma[X]$ belong to $\mathcal{L}(\widetilde{G}_{n,r})$, and denote $\mathcal{S} := T^{-1}(S)$. Further let $\Delta := \|\mathrm{Id} - T\|_{L^\infty(\widetilde{\mathbb{P}})}$, and $r^+ := r + \Delta$ and $r^- := r - \Delta$. Our goal is to lower bound

$$\frac{\widetilde{\mathrm{cut}}_{n,r}(S, S^c)}{\widetilde{\mathrm{vol}}_{n,r}(S)} \geq \widetilde{\Phi}_{\mathbb{P},r}(\mathcal{S}) - \frac{b_6\Delta}{s(\widetilde{G}_{n,r}) - b_6\Delta}$$

for some constant $b_6 > 0$. We first state the relationships between the discrete functionals $\widetilde{\mathrm{cut}}_{n,r}$ and $\widetilde{\mathrm{vol}}_{n,r}$, and the continuous functionals $\widetilde{Q}_{\mathbb{P},r}$ and $\widetilde{\pi}_{\mathbb{P},r}$, alluded to in the previous section.

Lemma 17. *For any set $S \subseteq \mathcal{C}_\sigma[X]$ and $\mathcal{S} = T^{-1}(S)$,*

$$\frac{1}{\widetilde{n}^2}\widetilde{\mathrm{cut}}_{n,r}(S) \geq \frac{\int_{\mathcal{C}_\sigma} \ell_{\mathbb{P},r^-}(x)\,d\mathbb{P}(x)}{\mathbb{P}(\mathcal{C}_\sigma)^2}\widetilde{Q}_{\mathbb{P},r^-}(\mathcal{S}, \mathcal{S}^c) \tag{54}$$

and

$$\frac{1}{\widetilde{n}^2}\widetilde{\mathrm{vol}}_{n,r}(S) \leq \frac{\int_{\mathcal{C}_\sigma} \ell_{\mathbb{P},r^+}(x)\,d\mathbb{P}(x)}{\mathbb{P}(\mathcal{C}_\sigma)^2}\widetilde{\pi}_{\mathbb{P},r^+}(\mathcal{S}) \tag{55}$$

To make use of Lemma 17, we provide deviation inequalities on $\int_{\mathcal{C}_\sigma} \ell_{\mathbb{P},r^+}(x)\,d\mathbb{P}(x)$, $\int_{\mathcal{C}_\sigma} \ell_{\mathbb{P},r^-}(x)\,d\mathbb{P}(x)$, $\widetilde{Q}_{\mathbb{P},r^-}(\mathcal{S}, \mathcal{S}^c)$, and $\widetilde{\pi}_{\mathbb{P},r^+}(\mathcal{S})$ in terms of the transportation distance Δ.

Lemma 18. *Suppose $\Delta \leq r$. Then there exist constants $b_7, b_8 \geq 0$ which do not depend on the sample size n, such that for every $\mathcal{S} \subseteq \mathcal{C}_\sigma$,*

$$\begin{aligned}
\int_{\mathcal{C}_\sigma} \ell_{\mathbb{P},r^+}(x)\,d\mathbb{P}(x) &\leq \int_{\mathcal{C}_\sigma} \ell_{\mathbb{P},r}(x)\,d\mathbb{P}(x) + b_7\Delta \\
\int_{\mathcal{C}_\sigma} \ell_{\mathbb{P},r^-}(x)\,d\mathbb{P}(x) &\geq \int_{\mathcal{C}_\sigma} \ell_{\mathbb{P},r}(x)\,d\mathbb{P}(x) - b_7\Delta \\
\widetilde{Q}_{\mathbb{P},r^-}(\mathcal{S}, \mathcal{S}^c) &\geq \widetilde{Q}_{\mathbb{P},r}(\mathcal{S}, \mathcal{S}^c) - b_8\Delta \\
\widetilde{\pi}_{\mathbb{P},r^+}(\mathcal{S}) &\leq \widetilde{\pi}_{\mathbb{P},r}(\mathcal{S}) + b_8\Delta.
\end{aligned}$$

The combination of Lemmas 17 and 18 brings us close to our goal, as demonstrated by the following manipulations:

$$\frac{\widetilde{\mathrm{cut}}_{n,r}(S,S^c)}{\widetilde{\mathrm{vol}}_{n,r}(S)} \geq \frac{\widetilde{Q}_{\mathbb{P},r^-}(\mathcal{S},\mathcal{S}^c)}{\widetilde{\pi}_{\mathbb{P},r^+}(\mathcal{S})}\frac{\int_{\mathcal{C}_\sigma} \ell_{\mathbb{P},r^-}(x)\,d\mathbb{P}(x)}{\int_{\mathcal{C}_\sigma} \ell_{\mathbb{P},r^+}(x)\,d\mathbb{P}(x)} \qquad \text{(Lemma 17)}$$

$$\geq \frac{\widetilde{Q}_{\mathbb{P},r}(\mathcal{S},\mathcal{S}^c) - b_8\varDelta}{\widetilde{\pi}_{\mathbb{P},r^+}(\mathcal{S}) + b_8\varDelta}\left(1 - \frac{2b_7\varDelta}{\int_{\mathcal{C}_\sigma} \ell_{\mathbb{P},r}(x)\,d\mathbb{P}(x)}\right) \qquad \text{(Lemma 18)}$$

$$\geq \frac{\widetilde{Q}_{\mathbb{P},r}(\mathcal{S},\mathcal{S}^c)}{\widetilde{\pi}_{\mathbb{P},r}(\mathcal{S})} - \left(\frac{2b_8}{\widetilde{\pi}_{\mathbb{P},r}(\mathcal{S})} + \frac{2b_7}{\int_{\mathcal{C}_\sigma} \ell_{\mathbb{P},r}(x)\,d\mathbb{P}(x)}\right)\varDelta \qquad (56)$$

where the last line follows from some basic algebra. The following result relates the unknown stationary distribution $\widetilde{\pi}_{\mathbb{P},r}(\mathcal{S})$ to $\widetilde{\pi}_{n,r}(S)$.

Lemma 19. *Suppose $\varDelta \leq \lambda_\sigma/(2^{d+1}d\Lambda_\sigma r)$. Then there exists a constant $b_9 > 0$ which does not depend on the sample size n such that for every $S \subseteq \mathcal{C}_\sigma[X]$,*

$$\widetilde{\pi}_{\mathbb{P},r}(\mathcal{S}) \geq \widetilde{\pi}_{n,r}(S) - b_9\varDelta$$

Combining this result with (56), we obtain

$$\frac{\widetilde{\mathrm{cut}}_{n,r}(S,S^c)}{\widetilde{\mathrm{vol}}_{n,r}(S)} \geq \widetilde{\Phi}_{\mathbb{P},r}(\mathcal{S}) - \left(\frac{2b_8}{\widetilde{\pi}_{n,r}(S) - b_9\varDelta} + \frac{2b_7}{\int_{\mathcal{C}_\sigma} \ell_{\mathbb{P},r}(x)\,d\mathbb{P}(x)}\right)\varDelta$$

$$\geq \widetilde{\Phi}_{\mathbb{P},r}(\mathcal{S}) - \left(\frac{2b_8}{s(\widetilde{G}_{n,r}) - b_9\varDelta} + \frac{2b_7}{\int_{\mathcal{C}_\sigma} \ell_{\mathbb{P},r}(x)\,d\mathbb{P}(x)}\right)\varDelta$$

where the latter inequality follows as we assumed $S \in \mathcal{L}(\widetilde{G}_{n,r})$. Choosing the constant b_6 in Lemma 15 to be $b_6 = \max\left\{2b_7/\left(\int_{\mathcal{C}_\sigma} \ell_{\mathbb{P},r}(x)\,d\mathbb{P}(x)\right) + 2b_8, b_9\right\}$, we achieve our desired result. It remains to show Lemmas 17, 18, and 19, which we now turn to.

Proof of Lemma 17: We begin with $\widetilde{\mathrm{cut}}_{n,r}(S)$.

$$\frac{1}{\widetilde{n}^2}\widetilde{\mathrm{cut}}_{n,r}(S) = \frac{1}{\widetilde{n}^2}\sum_{x_i,x_j \in \mathcal{C}_\sigma[X]} \mathbb{1}(\|x_i - x_j\| \leq r)\mathbb{1}(x_i \in S)\mathbb{1}(x_j \notin S)$$

$$= \iint_{\mathcal{C}_\sigma \times \mathcal{C}_\sigma} \mathbb{1}(\|x-y\| \leq r)\mathbb{1}(x \in S)\mathbb{1}(y \notin S)\,d\widetilde{\mathbb{P}}_n(x)\,d\widetilde{\mathbb{P}}_n(y)$$

$$= \iint_{\mathcal{C}_\sigma \times \mathcal{C}_\sigma} \mathbb{1}(\|T(x)-T(y)\| \leq r)\mathbb{1}(T(x) \in S)\mathbb{1}(T(y) \notin S)\,d\widetilde{\mathbb{P}}(x)\,d\widetilde{\mathbb{P}}(y)$$

(change of variables)

$$\geq \iint_{\mathcal{C}_\sigma \times \mathcal{C}_\sigma} \mathbb{1}(\|x-y\| \leq r^-)\mathbb{1}(T(x) \in S)\mathbb{1}(T(y) \notin S)\,d\widetilde{\mathbb{P}}(x)\,d\widetilde{\mathbb{P}}(y)$$

$$= \int_{\mathcal{S}}\int_{\mathcal{S}^c \cap B(x,r^-)} 1\,d\widetilde{\mathbb{P}}(y)\,d\widetilde{\mathbb{P}}(x)$$

By definition we have $\frac{d\mathbb{P}(x)}{d\widetilde{\mathbb{P}}(x)} = \mathbb{P}(\mathcal{C}_\sigma)$. Therefore,

$$\begin{aligned}
\int_{\mathcal{S}}\int_{\mathcal{S}^c\cap B(x,r^-)} d\widetilde{\mathbb{P}}(y)d\widetilde{\mathbb{P}}(x) &= \frac{1}{\mathbb{P}(\mathcal{C}_\sigma)^2}\int_{\mathcal{S}}\int_{\mathcal{S}^c\cap B(x,r^-)} d\mathbb{P}(y)d\mathbb{P}(x) \\
&= \frac{\int_{\mathcal{C}_\sigma}\ell_{\mathbb{P},r^-}(x)d\mathbb{P}(x)}{\mathbb{P}(\mathcal{C}_\sigma)^2}\int_{\mathcal{S}}\frac{\mathbb{P}\big(\mathcal{S}^c\cap B(x,r^-)\big)}{\ell_{\mathbb{P},r^-}(x)}d\widetilde{\pi}_{\mathbb{P},r^-}(x) \\
&= \frac{\int_{\mathcal{C}_\sigma}\ell_{\mathbb{P},r^-}\,d\mathbb{P}(x)}{\mathbb{P}(\mathcal{C}_\sigma)^2}\widetilde{Q}_{\mathbb{P},r^-}(\mathcal{S},\mathcal{S}^c). \qquad (57)
\end{aligned}$$

We obtain an upper bound on $\widetilde{\mathrm{vol}}_{n,r}(S)$ by similar manipulations, as follows:

$$\begin{aligned}
\frac{1}{\widetilde{n}^2}\widetilde{\mathrm{vol}}_{n,r}(S) &\le \frac{1}{\widetilde{n}^2}\sum_{x_i,x_j\in\mathcal{C}_\sigma[X]}\mathbb{1}(\|x_i-x_j\|\le r)\mathbb{1}(x_i\in S) \\
&= \iint_{\mathcal{C}_\sigma\times\mathcal{C}_\sigma}\mathbb{1}(\|x-y\|\le r)\mathbb{1}(x\in S)\,d\widetilde{\mathbb{P}}_n(x)\,d\widetilde{\mathbb{P}}_n(y) \\
&= \iint_{\mathcal{C}_\sigma\times\mathcal{C}_\sigma}\mathbb{1}(\|T(x)-T(y)\|\le r)\mathbb{1}(T(x)\in S)\,d\widetilde{\mathbb{P}}(x)\,d\widetilde{\mathbb{P}}(y) \\
&\le \iint_{\mathcal{C}_\sigma\times\mathcal{C}_\sigma}\mathbb{1}(\|x-y\|\le r^+)\mathbb{1}(x\in S)\,d\widetilde{\mathbb{P}}(x)\,d\widetilde{\mathbb{P}}(y) \\
&= \int_{\mathcal{S}}\int_{\mathcal{C}_\sigma\cap B(x,r^+)} 1\,d\widetilde{\mathbb{P}}(y)\,d\widetilde{\mathbb{P}}(x) \\
&= \frac{1}{\mathbb{P}(\mathcal{C}_\sigma)^2}\int_{\mathcal{S}}\int_{\mathcal{C}_\sigma\cap B(x,r^+)} 1\,d\mathbb{P}(y)\,d\mathbb{P}(x) \\
&= \frac{1}{\mathbb{P}(\mathcal{C}_\sigma)^2}\int_{\mathcal{S}}\ell_{\mathbb{P},r^+}(x)\,d\mathbb{P}(x) \\
&= \frac{\int_{\mathcal{C}_\sigma}\ell_{\mathbb{P},r}(x)\,d\mathbb{P}(x)}{\mathbb{P}(\mathcal{C}_\sigma)^2}\widetilde{\pi}_{\mathbb{P},r^+}(\mathcal{S}). \qquad (58)
\end{aligned}$$

Proof of Lemma 18: Consider the set

$$\mathcal{R}(x) := \big\{y\in\mathcal{C}_\sigma : y\in B(x,r^+), y\notin B(x,r^-)\big\}$$

for $x\in\mathcal{C}_\sigma$, and observe that

$$\int_{\mathcal{R}(x)} d\mathbb{P}(y) \le \Lambda_\sigma\nu_d\left((r+\delta)^d-(r-\Delta)^d\right)$$

and therefore

$$\int_{\mathcal{C}_\sigma}\int_{\mathcal{R}(x)} d\mathbb{P}(y)\,d\mathbb{P}(x) \le \mathbb{P}(\mathcal{C}_\sigma)\Lambda_\sigma\nu_d\left((r+\Delta)^d-(r-\Delta)^d\right). \qquad (59)$$

A first-order Taylor expansion of $(r+\Delta)^d$ results in the upper bound $(r+\Delta)^d \le r^d + d\Delta(r+\Delta)^{d-1}$, and similarly $(r-\Delta)^d \ge r^d - d\Delta(r+\Delta)^{d-1}$. Plugging these bounds into (59) yields

$$\begin{aligned}
\int_{\mathcal{C}_\sigma}\int_{\mathcal{R}(x)} d\mathbb{P}(y)\,d\mathbb{P}(x) &\le \mathbb{P}(\mathcal{C}_\sigma)\Lambda_\sigma\nu_d\big(2d(r+\Delta)^{d-1}\Delta\big) \\
&\qquad\text{(1st-order Taylor expansion of } (r+\Delta)^d) \\
&\le \mathbb{P}(\mathcal{C}_\sigma)\Lambda_\sigma\nu_d 2^d d r^{d-1}\Delta =: b_7\Delta. \qquad (60)
\end{aligned}$$

where the second line follows from the condition $\Delta \leq r$. Using (60), we proceed to obtain each of the four bounds stated in Lemma 18. First, as $\ell_{\mathbb{P},r^+}(x) \leq 1$ for all $x \in \mathcal{C}_\sigma$, we have

$$
\begin{aligned}
\int_{\mathcal{C}_\sigma} \ell_{\mathbb{P},r^+}(x)\, d\mathbb{P}(x) &\leq \int_{\mathcal{C}_\sigma} \ell_{\mathbb{P},r}(x)\, d\mathbb{P}(x) + \int_{\mathcal{C}_\sigma} \int_{\mathcal{R}(x)} d\mathbb{P}(y)\, d\mathbb{P}(x) \\
&\leq \int_{\mathcal{C}_\sigma} \ell_{\mathbb{P},r}(x)\, d\mathbb{P}(x) + b_7 \mathit{\Delta}.
\end{aligned}
$$

An equivalent bound for $\int_{\mathcal{C}_\sigma} \ell_{\mathbb{P},r^-}(x)\, d\mathbb{P}(x)$ is obtained by similar reasoning. We now lower bound $\widetilde{Q}_{\mathbb{P},r^-}(\mathcal{S}, \mathcal{S}^c)$,

$$
\begin{aligned}
\widetilde{Q}_{\mathbb{P},r^-}(\mathcal{S}, \mathcal{S}^c) &= \int_{\mathcal{S}} \widetilde{P}_{\mathbb{P},r^-}(x; \mathcal{S})\, d\widetilde{\pi}_{\mathbb{P},r}(x) \\
&= \frac{\int_{\mathcal{S}} \mathbb{P}(\mathcal{S}^c \cap B(x, r^-)) d\mathbb{P}(x)}{\int_{\mathcal{C}_\sigma} \mathbb{P}(\mathcal{C}_\sigma \cap B(x, r^-)) d\mathbb{P}(x)} \\
&\geq \frac{\int_{\mathcal{S}} \mathbb{P}(\mathcal{S}^c \cap B(x, r)) d\mathbb{P}(x) - \int_{\mathcal{C}_\sigma} \int_{\mathcal{R}(x)} d\mathbb{P}(y)\, d\mathbb{P}(x)}{\int_{\mathcal{C}_\sigma} \mathbb{P}(\mathcal{C}_\sigma \cap B(x, r)) d\mathbb{P}(x)} \\
&\geq \widetilde{Q}_{\mathbb{P},r}(\mathcal{S}, \mathcal{S}^c) - \frac{b_7 \mathit{\Delta}}{\int_{\mathcal{C}_\sigma} \ell_{\mathbb{P},r}(x)\, d\mathbb{P}(x)} =: \widetilde{Q}_{\mathbb{P},r}(\mathcal{S}, \mathcal{S}^c) - b_8 \mathit{\Delta}.
\end{aligned}
$$

Finally, we upper bound $\widetilde{\pi}_{\mathbb{P},r^+}(\mathcal{S})$,

$$
\begin{aligned}
\widetilde{\pi}_{\mathbb{P},r^+}(\mathcal{S}) &= \frac{\int_{\mathcal{S}} \ell_{\mathbb{P},r^+}(x)\, d\mathbb{P}(x)}{\int_{\mathcal{C}_\sigma} \ell_{\mathbb{P},r^+}(x)\, d\mathbb{P}(x)} \\
&\leq \frac{\int_{\mathcal{S}} \ell_{\mathbb{P},r}(x)\, d\mathbb{P}(x) + \int_{\mathcal{S}} \int_{\mathcal{R}(x)} d\mathbb{P}(y)\, d\mathbb{P}(x)}{\int_{\mathcal{C}_\sigma} \ell_{\mathbb{P},r}(x)\, d\mathbb{P}(x)} \\
&\leq \widetilde{\pi}_{\mathbb{P},r}(\mathcal{S}) + b_8 \mathit{\Delta}.
\end{aligned}
$$

Proof of Lemma 19: The proof of Lemma 19 will not be too different from the proof of Lemma 18. From the change of variables formula, we have

$$
\begin{aligned}
\widetilde{\pi}_{n,r}(S) &= \frac{\int_{\mathcal{S}} \int_{\mathcal{C}_\sigma} \mathbb{1}(\|x - y\| \leq r)\, d\widetilde{\mathbb{P}}_n(y)\, d\widetilde{\mathbb{P}}_n(x)}{\int_{\mathcal{C}_\sigma} \int_{\mathcal{C}_\sigma} \mathbb{1}(\|x - y\| \leq r)\, d\widetilde{\mathbb{P}}_n(y)\, d\widetilde{\mathbb{P}}_n(x)} \\
&\leq \frac{\int_{\mathcal{S}} \ell_{\mathbb{P},r^+}(x)\, d\mathbb{P}(x)}{\int_{\mathcal{C}_\sigma} \ell_{\mathbb{P},r^-}(x)\, d\mathbb{P}(x)} \\
&\leq \frac{\int_{\mathcal{S}} \ell_{\mathbb{P},r}(x)\, d\mathbb{P}(x) + b_7 \mathit{\Delta}}{\int_{\mathcal{C}_\sigma} \ell_{\mathbb{P},r}(x)\, d\mathbb{P}(x) - b_7 \mathit{\Delta}} \\
&\leq \widetilde{\pi}_{\mathbb{P},r}(\mathcal{S}) + \frac{2 b_7 \mathit{\Delta}}{\int_{\mathcal{C}_\sigma} \ell_{\mathbb{P},r}(x)\, d\mathbb{P}(x) - b_7 \mathit{\Delta}} \\
&\leq \widetilde{\pi}_{\mathbb{P},r}(\mathcal{S}) + \frac{4 b_7 \mathit{\Delta}}{\int_{\mathcal{C}_\sigma} \ell_{\mathbb{P},r}(x)\, d\mathbb{P}(x)} =: \widetilde{\pi}_{\mathbb{P},r}(\mathcal{S}) + b_9 \mathit{\Delta}.
\end{aligned}
$$

where the last inequality follows from the assumption $\mathit{\Delta} \leq \lambda_\sigma / (2^{d+1} d r \Lambda_\sigma)$, which implies $b_7 \mathit{\Delta} \leq \int_{\mathcal{C}_\sigma} \ell_{\mathbb{P},r}(x)\, d\mathbb{P}(x)/2$.

B.8 Proof of Lemma 16.

We recall that our goal is to lower bound the continuous conductance $\widetilde{\Phi}_{\mathbb{P},r}$. Results along these lines are already known (see e.g. [19]) when the density f is uniform (or log-concave) and $\mathcal{C}_\sigma$ is itself a convex set – indeed, in this case stronger versions of it exist, though we will not require them. In [1], a statement of this sort is made with respect to uniform density f and $\mathcal{C}_\sigma$ a Lipschitz deformation of a convex set, but for completeness we produce all proofs here.

Population-level conductance with uniform density. We first state an analogous result for the special case where the density f is uniform everywhere on $\mathcal{C}_\sigma$. For $x \in \mathcal{C}_\sigma$ and measurable $\mathcal{S} \subseteq \mathcal{C}_\sigma$, let

$$\widetilde{P}_{\nu,r}(x;\mathcal{S}) := \frac{\nu(\mathcal{S} \cap B(x,r))}{\nu(\mathcal{C}_\sigma \cap B(x,r))}, \quad \widetilde{\pi}_{\nu,r}(\mathcal{S}) = \frac{\int_{\mathcal{S}} \ell_{\nu,r}(x)\,dx}{\int_{\mathcal{C}_\sigma} \ell_{\nu,r}(x)\,dx}, \quad \widetilde{Q}_{\nu,r}(\mathcal{S}_1,\mathcal{S}_2) := \int_{\mathcal{S}_1} \widetilde{P}_{\nu,r}(x;\mathcal{S}_2)\,d\widetilde{\pi}_{\nu,r}(x).$$

The uniform continuous normalized cut and conductance are then defined analogously to the $\mathbb{P}$-weighted case,

$$\widetilde{\Phi}_{\nu,r}(\mathcal{S}) := \frac{\widetilde{Q}_{\nu,r}(\mathcal{S},\mathcal{S}^c)}{\min\{\widetilde{\pi}_{\nu,r}(\mathcal{S}), \widetilde{\pi}_{\nu,r}(\mathcal{S}^c)\}}, \quad \widetilde{\Phi}_{\nu,r} := \min_{\mathcal{S} \subseteq \mathcal{C}_\sigma} \widetilde{\Phi}_{\nu,r}(\mathcal{S}).$$

where the minimum is over measurable sets $\mathcal{S} \subseteq \mathcal{C}_\sigma$.

Lemma 20. *Let $\mathcal{C}_\sigma$ satisfy Assumption (A3) for some convex set $\mathcal{K}$ with diameter ρ, and measure-preserving mapping $g : \mathcal{K} \to \mathcal{C}_\sigma$ with Lipschitz constant L. Then, for any $0 < r \le \frac{\sigma}{2\sqrt{d}}$, the uniform conductance $\widetilde{\Phi}_{\nu,r}$ satisfies*

$$\widetilde{\Phi}_{\nu,r} > \frac{r}{2^{13}\rho L\sqrt{d}}.$$

Most of the technical work needed to show Lemma 16 involves proving Lemma 20. We first show that Lemma 16 is a simple consequence of Lemma 20 – and (A1) – before turning to prove Lemma 20.

To relate the uniform- and $\mathbb{P}$-continuous conudctances, observe that by (A1) we obtain

$$\begin{aligned}
\frac{\widetilde{Q}_{\mathbb{P},r}(\mathcal{S},\mathcal{S}^c)}{\widetilde{\pi}_{\mathbb{P},r}(\mathcal{S})} &= \frac{\int_{\mathcal{S}} \mathbb{P}(\mathcal{S}^c \cap B(x,r))\,d\mathbb{P}(x)}{\int_{\mathcal{S}} \mathbb{P}(\mathcal{C}_\sigma \cap B(x,r))\,d\mathbb{P}(x)} \\
&\ge \frac{\lambda_\sigma^2 \int_{\mathcal{S}} \nu(\mathcal{S}^c \cap B(x,r))\,dx}{\Lambda_\sigma^2 \int_{\mathcal{S}} \nu(\mathcal{C}_\sigma \cap B(x,r))\,dx} \\
&\ge \frac{\lambda_\sigma^2}{\Lambda_\sigma^2} \frac{\widetilde{Q}_{\nu,r}(\mathcal{S},\mathcal{S}^c)}{\pi_{\nu,r}(\mathcal{S})}
\end{aligned}$$

with equivalent reasoning leading to bound $\widetilde{Q}_{\mathbb{P},r}(\mathcal{S},\mathcal{S}^c)/\widetilde{\pi}_{\mathbb{P},r}(\mathcal{S}) \le (\lambda_\sigma^2/\Lambda_\sigma^2)\widetilde{Q}_{\nu,r}(\mathcal{S},\mathcal{S}^c)/\widetilde{\pi}_{\nu,r}(\mathcal{S})$. Lemma 16 therefore follows from Lemma 20, and it remains to show the latter Lemma.

Proof of Lemma 20: Lower bounds on the conductance of geometric random walks follow a typical pattern. The first step is to establish that an isoperimetric inequality holds uniformly over the domain of interest, in this case $\mathcal{C}_\sigma$. The following result contains the isoperimetric inequality we will use. It is stated and proved in [1]; for completeness, we reproduce the proof here.

Lemma 21 (Isoperimetry of Lipschitz embeddings of convex sets.)**.** *Let $\mathcal{C}_\sigma$ satisfy Assumption (A3) for some convex set $\mathcal{K}$ with diameter ρ, and measure-preserving mapping $g : \mathcal{K} \to \mathcal{C}_\sigma$ with Lipschitz constant L. Then, for any partition $(\Omega_1, \Omega_2, \Omega_3)$ of $\mathcal{C}_\sigma$,*

$$\nu(\Omega_3) \geq 2\frac{\mathrm{dist}(\Omega_1, \Omega_2)}{\rho L} \min(\nu(\Omega_1), \nu(\Omega_2))$$

The proof of Lemma 21 from first principles is non-trivial, even when the domain $\mathcal{C}_\sigma$ is itself convex, and is a primary technical contribution of the seminal work [27], extended by [11] among (many) others. Once the result is shown in the convex setting, however, it is not hard to show that it applies to Lipschitz transformations of convex sets as well.

Proof of Lemma 21. For $\Omega_i, i = 1, 2, 3$, denote the preimage

$$R_i = \{x \in \mathcal{K} : g(x) \in \Omega_i\}$$

For any $x \in R_1, y \in R_2$,

$$\|x - y\| \geq \frac{1}{L} \|g(x) - g(y)\| \geq \frac{1}{L}\mathrm{dist}(\Omega_1, \Omega_2).$$

Since $x \in \Omega_1$ and $y \in \Omega_2$ were arbitrary, we have

$$\mathrm{dist}(R_1, R_2) \geq \frac{1}{L}\mathrm{dist}(\Omega_1, \Omega_2).$$

By Theorem 2.2 of [27],

$$\begin{aligned}\nu(R_3) &\geq 2\frac{\mathrm{dist}(R_1, R_2)}{\rho} \min\{\nu(R_1), \nu(R_2)\} \\ &\geq \frac{2}{\rho L}\mathrm{dist}(\Omega_1, \Omega_2) \min\{\nu(R_1), \mathrm{vol}(R_2)\}\end{aligned}$$

and by the measure-preserving property of g, this implies

$$\nu(\Omega_3) \geq \frac{2}{\rho L}\mathrm{dist}(\Omega_1, \Omega_2) \min\{\nu(\Omega_1), \nu(\Omega_2)\}.$$

□

With an isoperimetric inequality in hand, the next step towards showing a lower bound on conductance is to prove that for any pair of sufficiently close points $u, v \in \mathcal{C}_\sigma$, the total variation distance between the distributions $\widetilde{P}_{\nu,r}(u; \cdot)$ and $\widetilde{P}_{\nu,r}(v; \cdot)$ is small.

Lemma 22 (One-step distributions)**.** *Let $u, v \in \mathcal{C}_\sigma$ be such that*

$$\|u - v\| \leq \frac{rt}{2\sqrt{d}}$$

for some $0 < t < 1$, and further assume there exists $\ell > 0$ such that $\ell_{\nu,r}(u), \ell_{\nu,r}(v) \geq \ell \nu_d r^d$. Then,

$$\left\|\widetilde{P}_{\nu,r}(u; \cdot) - \widetilde{P}_{\nu,r}(v; \cdot)\right\|_{\mathrm{TV}} \leq 1 + \frac{3\sqrt{3}t}{4\sqrt{2\pi}} - \ell$$

where $\|P - Q\|_{\mathrm{TV}}$ is the total variation distance between probabilities P and Q.

For the moment, we will take Lemma 22 as given and prove Lemma 20, before returning to prove Lemma 22. Deriving a lower bound on conductance from a isoperimetric inequality (i.e., Lemma 21) along with a bound on the one-step distributions is standard. Let $\mathcal{S}_1 \cup \mathcal{S}_2 = \mathcal{C}_\sigma$ be an arbitrary partition of $\mathcal{C}_\sigma$, and let $\ell = 6/25$. Note by Lemma 3, as $r \leq \sigma/(2\sqrt{d})$, $\ell \nu_d r^d \leq \ell_{\nu,r}(x)$ for all $x \in \mathcal{C}_\sigma$. We will show that

$$\int_{\mathcal{S}_1} \widetilde{P}_{\nu,r}(x;\mathcal{S}_2) d\widetilde{\pi}_{\nu,r}(x) \geq \frac{\sqrt{2}r\ell^4}{48\rho L\sqrt{d}} \min\left\{\widetilde{\pi}_{\nu,r}(\mathcal{S}_1), \widetilde{\pi}_{\nu,r}(\mathcal{S}_2)\right\}. \tag{61}$$

Then, dividing both sides by $\min\{\widetilde{\pi}_{\nu,r}(\mathcal{S}_1), \widetilde{\pi}_{\nu,r}(\mathcal{S}_2)\}$ yields the desired result (61), as $\mathcal{S}_1, \mathcal{S}_2$ was an arbitrary partition.

Now, consider the sets

$$\mathcal{S}_1' = \left\{x \in \mathcal{S}_1 : \widetilde{P}_{\nu,r}(x;\mathcal{S}_2) < \frac{\ell}{4}\right\}$$
$$\mathcal{S}_2' = \left\{x \in \mathcal{S}_2 : \widetilde{P}_{\nu,r}(x;\mathcal{S}_1) < \frac{\ell}{4}\right\}$$

and $\mathcal{S}_3' = \mathcal{C}_\sigma \setminus (\mathcal{S}_1' \cup \mathcal{S}_2')$. Suppose $\widetilde{\pi}_{\nu,r}(\mathcal{S}_1') < \widetilde{\pi}_{\nu,r}(\mathcal{S}_1)/2$. Then,

$$\int_{\mathcal{S}_1} \widetilde{P}_{\nu,r}(x;\mathcal{S}_2)\, d\widetilde{\pi}_{\nu,r}(x) \geq \frac{\ell \widetilde{\pi}_{\nu,r}(\mathcal{S}_1)}{8}$$

Similarly, if $\widetilde{\pi}_{\nu,r}(\mathcal{S}_1') < \widetilde{\pi}_{\nu,r}(\mathcal{S}_1)/2$, then since

$$\int_{\mathcal{S}_1} \widetilde{P}_{\nu,r}(x;\mathcal{S}_2)\, d\widetilde{\pi}_{\nu,r}(x) = \int_{\mathcal{S}_2} \widetilde{P}_{\nu,r}(x;\mathcal{S}_1)\, d\widetilde{\pi}_{\nu,r}(x)$$

a symmetric result holds. In either case, (61) follows.

Now, suppose $\pi_{\nu,r}(\mathcal{S}_1') \geq \pi_{\nu,r}(\mathcal{S}_1)/2$, and likewise for $\mathcal{S}_2'$ and $\mathcal{S}_2$. For every $u \in \mathcal{S}_1'$ and $v \in \mathcal{S}_2'$ we have that

$$\left\|\widetilde{P}_{\nu,r}(u;\cdot) - \widetilde{P}_{\nu,r}(v;\cdot)\right\|_{\mathrm{TV}} \geq 1 - \widetilde{P}_{\nu,r}(u;\mathcal{S}_1) - \widetilde{P}_{\nu,r}(v;\mathcal{S}_2) > 1 - \frac{\ell}{2}.$$

and thus by Lemma 22,

$$\|u - v\| \geq \frac{\sqrt{2\pi}r\ell}{3\sqrt{3d}} \geq \frac{\sqrt{2}r\ell}{3\sqrt{d}}.$$

Since $u \in \mathcal{S}_1', v \in \mathcal{S}_2'$ were arbitrary, the same inequality holds for $\mathrm{dist}(\mathcal{S}_1', \mathcal{S}_2')$. Therefore by Lemma 21

$$\nu(\mathcal{S}_3') \geq \frac{\sqrt{2}r\ell}{3\rho L\sqrt{d}} \min\left\{\nu(\mathcal{S}_1'), \nu(\mathcal{S}_2')\right\}. \tag{62}$$

Let $\mathcal{S} \subseteq \mathcal{C}_\sigma$ be arbitrary, and note that an upper (lower) bound on $\nu(\mathcal{S})$ implies an upper (lower) bound on $\widetilde{\pi}_{\nu,r}(\mathcal{S})$, as follows:

$$\ell \nu_d r^d \nu(\mathcal{S}) \leq \widetilde{\pi}_{\nu,r}(\mathcal{S}) \leq \nu_d r^d \nu(\mathcal{S}). \tag{63}$$

We now prove (61), the desired result:

$$\begin{aligned}
\int_{\mathcal{S}_1} \widetilde{P}_{\nu,r}(x;\mathcal{S}_2)\,dx &= \frac{1}{2}\Big(\int_{\mathcal{S}_1} \widetilde{P}_{\nu,r}(x;\mathcal{S}_2)\,d\widetilde{\pi}_{\nu,r}(x) + \int_{\mathcal{S}_2} \widetilde{P}_{\nu,r}(x;\mathcal{S}_1)\,d\widetilde{\pi}_{\nu,r}(x)\Big) \\
&\geq \frac{\ell}{8}\Big(\widetilde{\pi}_{\nu,r}(\mathcal{S}_1 \setminus \mathcal{S}_1') + \widetilde{\pi}_{\nu,r}(\mathcal{S}_2 \setminus \mathcal{S}_2')\Big) \\
&\geq \frac{\ell}{8}\widetilde{\pi}_{\nu,r}(\mathcal{S}_3') \\
&\geq \frac{\ell^2}{8\nu(\mathcal{C}_\sigma)}\nu(\mathcal{S}_3') && \text{(by (63))} \\
&\geq \frac{\sqrt{2}r\ell^3}{24\rho L\sqrt{d}\nu(\mathcal{C}_\sigma)} \min\left\{\nu(\mathcal{S}_1'), \nu(\mathcal{S}_2')\right\} && \text{(by (62))} \\
&\geq \frac{\sqrt{2}r\ell^4}{24\rho L\sqrt{d}} \min\left\{\widetilde{\pi}_{\nu,r}(\mathcal{S}_1'), \widetilde{\pi}_{\nu,r}(\mathcal{S}_2')\right\} && \text{(by (63))} \\
&\geq \frac{\sqrt{2}r\ell^4}{48\rho L\sqrt{d}} \min\left\{\widetilde{\pi}_{\nu,r}(\mathcal{S}_1), \widetilde{\pi}_{\nu,r}(\mathcal{S}_2)\right\}
\end{aligned}$$

where the last line follows since $\pi_{\nu,r}(\mathcal{S}_1') \geq \pi_{\nu,r}(\mathcal{S}_1)/2$, and likewise for $\mathcal{S}_2'$ and $\mathcal{S}_2$. To prove Lemma 20, it remains to prove Lemma 22, and we do this next.

Proof of Lemma 22: The key result needed to show Lemma 22 deals with volume of the overlap $B(u,r) \cap B(v,r)$. We state this result and immediately prove it.

Lemma 23. *Let $u, v \in \mathbb{R}^d$ be points such that $\|u - v\| \leq t\frac{r}{2\sqrt{d}}$ for some $0 < t < 1$. Then,*

$$\nu(B(u,r) \cap B(v,r)) \geq \nu_d r^d \left(1 - \frac{3\sqrt{3}t}{4\sqrt{2\pi}}\right)$$

Proof. We will treat only the case where $\|u - v\| = t\frac{r}{2\sqrt{d}}$; if they are closer together the overlap of the volume will only increase. Then, it is not hard to see that $\mathcal{I} := B(u,r) \cap B(v,r)$ consists of two symmetric spherical caps, each with height

$$h = r\left(1 - \frac{t}{4\sqrt{d}}\right)$$

As a result, by Lemma 7 we have

$$\nu(\mathcal{I}) = \nu_d r^d I_{1-\alpha}(\frac{d+1}{2}; \frac{1}{2})$$

where

$$\alpha = 1 - \frac{2r^2(1 - \frac{t}{4\sqrt{d}}) - r^2(1 - \frac{t}{4\sqrt{d}})^2}{r^2} = \frac{t^2}{16d}.$$

Expanding the incomplete beta function in integral form, we therefore have

$$\begin{aligned}
\nu(\mathcal{I}) &= \nu_d r^d \frac{\Gamma\big(\frac{d}{2}+1\big)}{\Gamma\big(\frac{d+1}{2}\big)\Gamma\big(\frac{1}{2}\big)} \int_0^{1-\alpha} u^{(d-1)/2}(1-u)^{-1/2}du \\
&\overset{\text{(i)}}{\geq} \nu_d r^d \left(1 - \frac{\Gamma\big(\frac{d}{2}+1\big)}{\Gamma\big(\frac{d+1}{2}\big)\Gamma\big(\frac{1}{2}\big)} \frac{3t}{4\sqrt{d}}\right) \\
&\overset{\text{(ii)}}{\geq} \nu_d r^d \left(1 - \frac{3\sqrt{3}t}{4\sqrt{2\pi}}\right)
\end{aligned}$$

where (i) follows from Lemma 8 (which we may validly apply since $\alpha \leq \frac{t^2}{16d}$), and (ii) from Lemma 9. $\square$

We now complete the proof of Lemma 22. Let $\mathcal{S}_1 \cup \mathcal{S}_2 = \mathcal{C}_\sigma$ be an arbitrary partition of $\mathcal{C}_\sigma$. We will show that

$$\widetilde{P}_{\nu,r}(u;\mathcal{S}_1) - \widetilde{P}_{\nu,r}(v;\mathcal{S}_1) \leq 1 + \frac{3\sqrt{3}t}{4\sqrt{2\pi}} - \ell.$$

Since this will hold for arbitrary $\mathcal{S}_1 \subset \mathcal{C}_\sigma$, it will hold for the supremum over all such $\mathcal{S}_1$ as well, and therefore the same upper bound will hold for $\left\|\widetilde{P}_{\nu,r}(u;\cdot) - \widetilde{P}_{\nu,r}(v;\cdot)\right\|_{\mathrm{TV}}$. Now, note that

$$\widetilde{P}_{\nu,r}(u;\mathcal{S}_1) - \widetilde{P}_{\nu,r}(v;\mathcal{S}_1) = 1 - \widetilde{P}_{\nu,r}(u;\mathcal{S}_2) - \widetilde{P}_{\nu,r}(v;\mathcal{S}_1)$$

As before, let $\mathcal{I} = B(u,r) \cap B(v,r)$. Then we have

$$\widetilde{P}_{\nu,r}(u;\mathcal{S}_2) \geq \frac{1}{\nu(B(u,r))}\nu(\mathcal{S}_2 \cap (B(u,r)) \geq \frac{1}{\nu(B(u,r))}\nu(\mathcal{S}_2 \cap \mathcal{I})$$

with a symmetric inequality holding for $\widetilde{P}_{\nu,r}(v;\mathcal{S}_1)$. As a result,

$$1 - \widetilde{P}_{\nu,r}(u;\mathcal{S}_2) - \widetilde{P}_{\nu,r}(v;\mathcal{S}_1) \leq 1 - \frac{1}{\nu_d r^d}\nu(\mathcal{C}_\sigma \cap \mathcal{I}) \tag{64}$$

From here, some simple manipulations yield

$$\begin{aligned}
\nu(\mathcal{C}_\sigma \cap \mathcal{I}) &= \nu(\mathcal{I}) - \nu(\mathcal{I} \setminus \mathcal{C}_\sigma) \\
&\geq \nu(\mathcal{I}) - \max\left\{\nu\big(B(u,r) \setminus \mathcal{C}_\sigma\big), \nu\big(B(v,r) \setminus \mathcal{C}_\sigma\big)\right\} \\
&\geq \nu_d r^d \left(1 - \frac{3\sqrt{3}t}{4\sqrt{2\pi}}\right) - \max\left\{\nu\big(B(u,r) \setminus \mathcal{C}_\sigma\big), \nu\big(B(v,r) \setminus \mathcal{C}_\sigma\big)\right\} \qquad \text{(Lemma 23)} \\
&\geq \nu_d r^d \left(1 - \frac{3\sqrt{3}t}{4\sqrt{2\pi}} - (1-\ell)\right) = \nu_d r^d \left(\ell - \frac{3\sqrt{3}t}{4\sqrt{2\pi}}\right) \qquad (65)
\end{aligned}$$

where the last inequality follows by the hypothesis $\ell_{\nu,r}(u), \ell_{\nu,r}(v) \geq \ell \nu_d r^d$. Then (65) along with (64) give the desired result

$$\begin{aligned}
\widetilde{P}_{\nu,r}(u;\mathcal{S}_1) - \widetilde{P}_{\nu,r}(v;\mathcal{S}_1) &\leq 1 - \frac{1}{\nu_d r^d}\nu(\mathcal{C}_\sigma \cap \mathcal{I}) \\
&\leq 1 - \ell + \frac{3\sqrt{3}t}{4\sqrt{2\pi}}.
\end{aligned}$$

We have completed our proof of Lemma 22, and therefore (in turn), Lemmas 20, 16, and 14. As a result, Proposition 2 is proved, and the proof of Theorem 4 is complete.

C Proof of Theorem 1

In this section, we prove Theorem 1. The theorem is a consequence of Lemma 3.4 in Zhu et al. [51], which relates the quality of the PPR output to normalized cut and mixing time.

Lemma 24 (Lemma 3.4 of Zhu et al. [51].). *Let $G = (V, E)$ be a undirected, unweighted graph, let $A \subseteq V$, and let $p_v^{(\varepsilon)}$ be an ε-approximation to the PPR vector $p_v := p(v, \alpha; G)$. For $\beta \in (0, 1)$, the sweep cut $S_{\beta,v}$ is*

$$S_{\beta,v} = \left\{ u \in V : \frac{p_v^{(\varepsilon)}(u)}{\deg(u;G)} \geq \beta \right\}$$

Suppose $\alpha \leq \frac{1}{9\tau_\infty(G[S])}$, $\varepsilon \leq \frac{1}{20}\mathrm{vol}(A;G)$. Then there exists a set $A^g \subset A$ with $\mathrm{vol}(A^g;G) \geq \frac{1}{2}\mathrm{vol}(A^g;G)$ such that for any $v \in S^g$ and $\beta \in \frac{1}{\mathrm{vol}(A;G)}(\frac{1}{100}, \frac{2}{5})$ the sweep cut $S_{\beta,v}$ satisfies

$$\mathrm{vol}(A \vartriangle S_{\beta,v}; G) \leq \left(218 \frac{\Phi(A;G)}{\alpha} \right) \mathrm{vol}(A;G)$$

Proof of Theorem 1. First, we verify that the conditions of Lemma 24 are met when $S = \mathcal{C}_\sigma[X]$, $G = G_{n,r}$, and the user-specified parameters satisfy the initialization conditions (8). By Theorem 4 and (8), the teleportation parameter α is upper bounded

$$\alpha < \frac{1}{9\tau_u(\theta)} \leq \frac{1}{9\tau_\infty(G_{n,r}(\mathcal{C}_\sigma[X])}$$

with probability at least $1 - \frac{b_2}{n}$. Since we use an exact PPR vector $p_v = p_v^{(0)}$ to construct the sweep cut sets in Algorithm 1, we may apply Lemma 24. By this Lemma, there exists a set $\mathcal{C}_\sigma[X]^g \subset \mathcal{C}_\sigma[X]$ with $\mathrm{vol}_{n,r}(\mathcal{C}_\sigma[X]^g) \geq \frac{1}{2}\mathrm{vol}_{n,r}(\mathcal{C}_\sigma[X])$ such that for any $\beta \in \frac{1}{\mathrm{vol}_{n,r}(\mathcal{C}_\sigma[X])}(\frac{1}{100}, \frac{2}{5})$, the sweep cut $S_{\beta,v}$ of $p(\alpha, v; G_{n,r})$ satisfies

$$\begin{aligned} \mathrm{vol}_{n,r}(\mathcal{C}_\sigma[X] \vartriangle S_{\beta,v}) &\leq \left(218 \frac{\Phi_{n,r}(\mathcal{C}_\sigma[X])}{\alpha} \right) \mathrm{vol}_{n,r}(\mathcal{C}_\sigma[X]) \\ &\leq 2180 \cdot \kappa(\mathcal{C}) \cdot \mathrm{vol}_{n,r}(\mathcal{C}_\sigma[X]) \end{aligned} \tag{66}$$

where the latter inequality holds with probability at least $1 - 3\exp\{-bn\}$, and follows from Theorem 3 and the lower bound $\alpha \geq \frac{1}{10\tau_u(\theta)}$ given in (8).

Now we show that we can apply (66) to the cluster estimate $\widehat{C}$ output by Algorithm 1. Observe that $\widehat{C}$ is itself a sweep cut as $\widehat{C} = S_{\beta,v}$ for some $\beta \in \left(\frac{1}{50}, \frac{1}{5}\right) \cdot \frac{1}{2\binom{n}{2}\mathrm{vol}_{\mathbb{P},r}(\mathcal{C}_\sigma)}$. Moreover, applying Lemma 36 with $\delta = 1/2$,

$$\frac{1}{2}\mathrm{vol}_{\mathbb{P},r}(\mathcal{C}_\sigma) \leq \frac{\mathrm{vol}_{n,r}(\mathcal{C}_\sigma[X])}{2\binom{n}{2}} \leq \frac{3}{2}\mathrm{vol}_{\mathbb{P},r}(\mathcal{C}_\sigma) \tag{67}$$

with probability at least $1 - 2\exp\{-\frac{1}{4}n\mathrm{vol}_{\mathbb{P},r}(\mathcal{C}_\sigma)^2\}$. Therefore, conditional on (67) the cluster estimate $\widehat{C} = S_{\beta,v}$ for some $\beta \in \frac{1}{\mathrm{vol}_{n,r}(\mathcal{C}_\sigma[X])}\left(\frac{1}{100}, \frac{2}{5}\right)$, and therefore the upper bound (66) holds with respect to $\mathrm{vol}_{n,r}(\widehat{C} \vartriangle \mathcal{C}_\sigma[X])$. To summarize, we have shown that when

$$n \geq b_1(\log(n))^{\max\{\frac{3}{d},1\}}$$

then

$$\mathrm{vol}_{n,r}(\mathcal{C}_\sigma[X] \vartriangle \widehat{C}) \leq 2180 \cdot \kappa(\mathcal{C}) \cdot \mathrm{vol}_{n,r}(\mathcal{C}_\sigma[X])$$

with probability at least $1 - \frac{b_2}{n} - 3\exp\left\{-(b + \frac{1}{4}\mathrm{vol}_{\mathbb{P},r}(\mathcal{C}_\sigma)^2)n\right\}$. We have proved Theorem 1 (upon appropriate choice of constants b_2 and c in the theorem statement).

D Proof of Theorem 2

As a reminder, to prove Theorem 2 it will be sufficient to show that

$$\max_{u' \in \mathcal{C}'[X]} \frac{p_v(u')}{\deg_{n,r}(u')} \leq \frac{1}{100\binom{n}{2}\mathrm{vol}_{\mathbb{P},r}(\mathcal{C}_\sigma)} < \frac{1}{10\binom{n}{2}\mathrm{vol}_{\mathbb{P},r}(\mathcal{C}_\sigma)} \leq \min_{u \in \mathcal{C}[X]} \frac{p_v(u)}{\deg_{n,r}(u)} \tag{68}$$

with probability at least $1 - \frac{b_2}{n}$.

To show this, we first establish in Lemma 25 that when Algorithm 1 is properly initialized, the resulting PPR vector is large for every vertex $u \in \mathcal{C}$, and small for every vertex $u' \in \mathcal{C}'$. Let $\overline{\pi}_{n,r} : \mathcal{C}_\sigma[X] \to [0,1]$ be the vector given by[5]

$$\overline{\pi}_{n,r}(u) := \frac{\widetilde{\deg}_{n,r}(u)}{\mathrm{vol}_{n,r}(\mathcal{C}_\sigma[X])}.$$

Additionally, let $d'_{\min} = \min_{u \in \mathcal{C}'[X]} \deg_{n,r}(u')$.

Lemma 25. *Let $0 < r < \sigma$ and $\alpha \leq \frac{1}{9\tau_\infty(\widetilde{G}_{n,r})}$. Then the following statement holds: there exists a good set $\mathcal{C}_\sigma[X]^g \subseteq \mathcal{C}_\sigma[X]$ with $\mathrm{vol}_{n,r}(\mathcal{C}_\sigma[X]^g) \geq \mathrm{vol}_{n,r}(\mathcal{C}_\sigma[X])/2$ such that the following bounds hold with respect to $p_v := p(v, \alpha; G_{n,r})$ for any $v \in \mathcal{C}_\sigma[X]^g$:*

- *For each $u \in \mathcal{C}[X]$,*
$$p_v(u) \geq \frac{4}{5}\overline{\pi}_{n,r}(u) - \frac{20\Phi_{n,r}(\mathcal{C}_\sigma[X])/\alpha}{\widetilde{d}_{\min}} \tag{69}$$

- *Let $\mathcal{C}' \neq \mathcal{C} \in \mathbb{C}_f(\lambda)$ be another λ-density cluster. Then for each $u' \in \mathcal{C}'[X]$,*
$$p_v(u') \leq \frac{20\Phi_{n,r}(\mathcal{C}_\sigma[X])/\alpha}{d'_{\min}}. \tag{70}$$

We immediately note that by assumption, Algorithm 1 is well-initialized, and therefore the seed node v is chosen in $\mathcal{C}_\sigma[X]^g$. Since we additionally assume $\mathcal{C}$ is a κ-well-conditioned density cluster, we have that $r < \sigma/(2d) < \sigma$, and the upper bound

$$\alpha < \frac{1}{9\tau_u(\theta)} \leq \frac{1}{(9\tau_\infty(G_{n,r}[\mathcal{C}_\sigma[X]]))}$$

holds with probability at least $1 - \frac{b_2}{n}$. Therefore, all the conditions of Lemma 25 are met.

We now collect the estimates on graph functionals we will use to complete the proof of Theorem 2.

- By Theorem 3,
$$\Phi(\mathcal{C}_\sigma[X]; G_{n,r}) \leq \Phi_u(\theta)$$
with probability at least $1 - 3\exp\{-nb\}$. By (8), $\alpha \geq 1/(10\tau_u(\theta))$ and therefore $2\Phi(\mathcal{C}_\sigma[X]; G_{n,r})/\alpha \leq 20\kappa(\mathcal{C})$.

- By Lemma 36, for any $\delta \in (0,1)$,
$$\widetilde{d}_{\min} \geq \frac{6}{25}(1-\delta)\lambda_\sigma r^d \nu_d n,$$
with probability at least $1 - n\exp\left\{-\frac{2\delta^2\lambda_\sigma\nu_d r^d n}{75(1+\frac{\delta}{3})}\right\}$.

[5] Note that $\overline{\pi}_{n,r}$ is distinct from $\widetilde{\pi}_{n,r}$, as we normalize by $\mathrm{vol}_{n,r}(\mathcal{C}_\sigma[X])$ rather than $\widetilde{\mathrm{vol}}_{n,r}(\mathcal{C}_\sigma[X])$.

- By Lemma 36, for any $\delta \in (0,1)$,

$$d'_{\min} \geq (1-\delta)\lambda_\sigma r^d \nu_d n,$$

with probability at least $1 - n\exp\left\{-\frac{\delta^2 \lambda_\sigma \nu_d r^d n}{3(1+\frac{\delta}{3})}\right\}$.

- By Lemma 36, for any $\delta \in (0,1)$,

$$d_{\max} \leq (1+\delta)\Lambda_\sigma \nu_d r^d$$

with probability at least $1 - n\exp\left\{-\frac{2\delta^2 \lambda_\sigma \nu_d r^d n}{75(1+\frac{\delta}{3})}\right\}$.

-

$$\widetilde{\mathrm{vol}}_{n,r}(\mathcal{C}_\sigma[X]) \leq 2(1+\delta)\mathrm{vol}_{\mathbb{P},r}(\mathcal{C}_\sigma)\binom{n}{2}$$

with probability at least $1 - \exp\left\{-\delta^2(\mathrm{vol}_{\mathbb{P},r}(\mathcal{C}_\sigma))^2\right\}$.

In the rest of this proof we assume the above bounds hold. For all $u \in \mathcal{C}[X]$, $\widetilde{\deg}_{n,r}(u) = \deg_{n,r}(u)$, so by (69),

$$\begin{aligned}
\frac{p_v(u)}{\deg_{n,r}(u)} &\geq \frac{4}{5\mathrm{vol}_{n,r}(\mathcal{C}_\sigma[X])} - \frac{20\kappa(\mathcal{C})}{\widetilde{d}^2_{\min}} \\
&\geq \frac{4}{5(1+\delta)\mathrm{vol}_{\mathbb{P},r}(\mathcal{C}_\sigma)n^2} - \frac{20\kappa(\mathcal{C})}{\left(\frac{6}{25}(1-\delta)\lambda_\sigma \nu_d r^d n\right)^2} \\
&\geq \frac{1}{10\binom{n}{2}\mathrm{vol}_{\mathbb{P},r}(\mathcal{C}_\sigma)},
\end{aligned}$$

where the last inequality follows by choosing $\delta = 1/2$ and (11). For all $u' \in \mathcal{C}'[X]$,

$$\begin{aligned}
\frac{p_v(u')}{\deg_{n,r}(u')} &\leq \frac{20\kappa(\mathcal{C})}{(d'_{\min})^2} \\
&\leq \frac{20\kappa(\mathcal{C})}{\left((1-\delta)\lambda_\sigma \nu_d r^d n\right)^2} \\
&\leq \frac{1}{100\binom{n}{2}\mathrm{vol}_{\mathbb{P},r}(\mathcal{C}_\sigma)}
\end{aligned} \tag{71}$$

where the last inequality again follows by choosing $\delta = 1/2$ and (11). We have shown (68) and therefore proved Theorem 2 (upon an appropriate choice of constants b_2 and c in the statement of the theorem).

We defer the proof of Lemma 25, and first extend Theorems 1 and 2 to apply with respect to the aPPR vector.

D.1 Proof of Corollary 1.

To prove the first claim of Corollary 1, it will be sufficient to show that the conditions of Lemma 24 are still met when we use the approximate PPR vector $p_v^{(\varepsilon)}$ rather than the exact PPR vector p_v. In particular, we must show that $\varepsilon \leq \frac{1}{20}\mathrm{vol}_{n,r}(\mathcal{C}_\sigma)$. However by Lemma 36,

$$\mathrm{vol}_{n,r}(\mathcal{C}_\sigma[X]) \geq (1-\delta)\mathrm{vol}_{\mathbb{P},r}(\mathcal{C}_\sigma) \tag{72}$$

with probability at least $1 - \exp\left\{-n\delta^2(\mathrm{vol}_{\mathbb{P},r}(\mathcal{C}_\sigma))^2\right\}$, and by setting $\delta = 1/5$ the desired claim follows.

We now prove the second claim, by showing that both bounds in (68) hold with respect to the aPPR vector. The upper bound follows immediately. Since $p_v^{(\varepsilon)}(x) \leq p_v(x)$ for all $x \in X$, the upper bound on $\max_{u' \in \mathcal{C}'[X]} p_v(u')$ also applies to $\max_{u' \in \mathcal{C}'[X]} p_v^{(\varepsilon)}(u')$.

To show the desired lower bound on $\min_{u \in \mathcal{C}[X]} p_v^{(\varepsilon)}(u)$, we use the uniform approximation guarantee in (12). As previously observed, since $r < \sigma$, for any $u \in \mathcal{C}_\sigma$ the ball $B(u, r) \subseteq \mathcal{C}_\sigma$, and so $\widetilde{\deg}_{n,r}(u) = \deg_{n,r}(u)$. Along with (13) and (72), this means

$$
\varepsilon \deg_{n,r}(u) \leq \frac{\overline{\pi}_{n,r}(u)}{25(1-\delta)}
$$

and therefore by (69), for every $u \in \mathcal{C}[X]$,

$$
p_v^{(\varepsilon)}(u) \geq \left\{\frac{4}{5} - \frac{1}{25(1-\delta)}\right\} \overline{\pi}_{n,r}(u) - \frac{20\Phi_{n,r}(\mathcal{C}_\sigma[X])/\alpha}{\widetilde{d}_{\min}}.
$$

Using the arguments given in the proof of Theorem 2, and choosing the universal constant in (11) to be sufficiently small, the upper bound in (71) follows.

D.2 Proof of Lemma 25

In this subsection, we will let D and W be the degree and lazy random walk matrices over $G_{n,r}$. Additionally we let $\widetilde{D}$ and $\widetilde{W}$ be the degree and lazy random walk matrices over the induced subgraph $\widetilde{G}_{n,r}$. Given a starting distribution q with $\mathrm{supp}(q) \subseteq \mathcal{C}_\sigma[X]$, we let

$$
\widetilde{p}_q = \alpha q + (1-\alpha)\widetilde{p}_q \widetilde{W} \tag{73}
$$

be the PPR vector originating from q over $\widetilde{G}_{n,r}$. (When the starting distribution $q = e_v$ is a point mass at a seed node $v \in \mathcal{C}_\sigma[X]$, we write $\widetilde{p}_v := \widetilde{p}_{e_v}$ in a slight abuse of notation).

Our analysis will involve *leakage* and *soakage* vectors, defined by

$$
\begin{aligned}
\ell_t &:= e_v(W\widetilde{I})^t(I - D^{-1}\widetilde{D}),\ \ell := \sum_{t=0}^{\infty}(1-\alpha)^t \ell_t, \\
s_t &:= e_v(W\widetilde{I})^t(W\widetilde{I}^c),\ s := \sum_{t=0}^{\infty}(1-\alpha)^t s_t. \qquad (74)
\end{aligned}
$$

where I is the $n \times n$ identity matrix, $\widetilde{I}$ is an $n \times n$ diagonal matrix with $\widetilde{I}_{uu} = 1$ if $u \in \mathcal{C}_\sigma[X]$ and 0 otherwise, and $\widetilde{I}^c = I - \widetilde{I}$.

In words, for $u \in \mathcal{C}_\sigma[X]$, $\ell_t(u)$ is the probability that a random walk over $G_{n,r}$ originating from $v \in \mathcal{C}_\sigma[X]$ stays within $\widetilde{G}_{n,r}$ for t steps, arriving at u on the tth step, and then "leaks out" of $\mathcal{C}_\sigma[X]$ on the $t+1$th step. For $w \in X \setminus \mathcal{C}_\sigma[X]$, $\ell_t(w) = 0$. By contrast, for w again in $X \setminus \mathcal{C}_\sigma[X]$, $s_t(w)$ is the probability that a random walk originating from v stays within $\mathcal{C}_\sigma[X]$ for t steps, and then is "soaked up" into w on the $t+1$ step, while $s_t(u) = 0$ for all $u \in \mathcal{C}_\sigma[X]$. The vectors ℓ and s then give the total mass leaked and soaked, respectively, by the PPR vector.

We first prove (69), and begin by restating some results of [51], adapted to our notation. By Lemma 3.1 of [51], there exists a good set $\mathcal{C}_\sigma[X]^g \subseteq \mathcal{C}_\sigma[X]$ with $\mathrm{vol}(\mathcal{C}_\sigma[X]^g; G_{n,r}) \geq \mathrm{vol}(\mathcal{C}_\sigma[X]; G_{n,r})/2$ such that for every $v \in \mathcal{C}_\sigma[X]^g$

$$p_v(u) \geq \widetilde{p}_v(u) - \widetilde{p}_\ell(u), \quad \text{and} \quad \|\ell\|_1 \leq \frac{2\Phi_{n,r}(\mathcal{C}_\sigma[X])}{\alpha}. \tag{75}$$

(The result $\|\ell\|_1 \leq \frac{2\Phi_{n,r}(\mathcal{C}_\sigma[X])}{\alpha}$ is the only result in the proof of Theorem 2 which relies on the restriction $v \in \mathcal{C}_\sigma[X]^g$.)

If additionally $\alpha \leq \frac{1}{9\tau_\infty(\widetilde{G}_{n,r})}$, then by Corollary 3.3 of [51], for every $u \in \mathcal{C}_\sigma[X]$

$$\widetilde{p}_v(u) \geq \frac{4}{5}\overline{\pi}_{n,r}(u)$$

and along with (75), we obtain

$$p_v(u) \geq \frac{4}{5}\overline{\pi}_{n,r}(u) - \widetilde{p}_\ell(u).$$

We proceed to show the upper bound $\widetilde{p}_\ell(u) \leq \|\ell\|_1 / \widetilde{d}_{\min}$, whence (69) follows by (75). We note two facts regarding $\widetilde{p}_\ell(u)$, which hold for all $u \in \mathcal{C}[X]$.

1. Since $r < \sigma$, $(u, w) \notin G_{n,r}$ for any $w \notin \mathcal{C}_\sigma$. As a result, for all $t \geq 1$, $\ell_t(u) = 0$ and by extension, $\ell(u) = 0$ as well.
2. For any q such that $\sum_{w \in \mathcal{C}_\sigma[X]} q(w) \leq 1$ and $u \notin \mathrm{supp}(q)$, and any $t \geq 1$,

$$\begin{aligned} q\widetilde{W}^t(u) &\leq \|q\|_1 \max_{v \neq u} W_{vu} \\ &\leq \frac{1}{2\widetilde{d}_{\min}} \end{aligned} \tag{76}$$

   where last inequality follows from the fact $(u, w) \in \widetilde{G}_{n,r}$ implies $w \in \mathcal{C}_\sigma$, and therefore $\deg(w; \widetilde{G}_{n,r}) \geq \widetilde{d}_{\min}$.

These facts, along with some basic algebra, lead to the desired lower bound on $\widetilde{p}_\ell(u)$ for every $u \in \mathcal{C}[X]$:

$$\begin{aligned} \widetilde{p}_\ell(u) &= \alpha \sum_{t=0}^{\infty} (1-\alpha)^t \left(\ell \widetilde{W}^t\right)(u) \\ &= \|\ell\|_1 \alpha \sum_{t=0}^{\infty} (1-\alpha)^t \left(\frac{\ell}{\|\ell\|_1} \widetilde{W}^t\right)(u) \\ &= \|\ell\|_1 \alpha \sum_{t=1}^{\infty} (1-\alpha)^t \left(\frac{\ell}{\|\ell\|_1} \widetilde{W}^t\right)(u) \\ &\leq \frac{\|\ell\|_1}{\widetilde{d}_{\min}}. && \text{(since } u \notin \mathrm{supp}(\ell)\text{)} \end{aligned}$$

and (69) is proved.

We turn to showing (70). By Lemma 26, for all $u' \notin \mathcal{C}_\sigma[X]$,

$$p_v(u') \leq p_s(u').$$

Note that by (A2), $\mathrm{dist}(\mathcal{C}_\sigma, \mathcal{C}') > r$. Therefore for every $u \in \mathcal{C}_\sigma[X]$ and $u' \in \mathcal{C}'[X]$ $(u', u) \notin E$ and so $s(u') = 0$. Some manipulations, similar to those in the preceding part of the proof, yield a lower bound on $p_v(u')$ in terms of $\|s\|_1$:

$$\begin{aligned}
p_s(u') &= \alpha \sum_{t=0}^{\infty} (1-\alpha)^t \left(sW^t\right)(u') \\
&= \|s\|_1 \, \alpha \sum_{t=0}^{\infty} (1-\alpha)^t \left(\frac{s}{\|s\|_1} W^t\right)(u') \\
&= \|s\|_1 \, \alpha \sum_{t=1}^{\infty} (1-\alpha)^t \left(\frac{s}{\|s\|_1} W^t\right)(u') \\
&\leq \frac{\|s\|_1}{2d'_{\min}}
\end{aligned}$$

where the last inequality follows from precisely the same reasoning as (76). The claim follows in light of Lemma 28, along with (75).

D.3 Linear Algebra Facts

We state here a number of basic facts which follow from matrix manipulations, which are used in the proof of Theorem 2.

Lemma 26. *For any $v \in \mathcal{C}_\sigma[X]$ and $u \notin \mathcal{C}_\sigma[X]$,*

$$p_v(u) \leq p_s(u)$$

where s is defined as in (74) and depends implicitly upon v.

Proof. The statement follows from Lemma 27 along with a series of algebraic manipulations,

$$\begin{aligned}
p_v(u) &= \alpha \sum_{T=0}^{\infty} (1-\alpha)^t \left(e_v W^T\right)(u) \\
&= \alpha \sum_{T=1}^{\infty} (1-\alpha)^t \left(e_v W^T\right)(u) \\
&\leq \alpha \sum_{T=1}^{\infty} (1-\alpha)^T \left(\sum_{t=0}^{T-1} s_t W^{T-t-1}\right)(u) && \text{(Lemma 27)} \\
&= \alpha \sum_{t=0}^{\infty} \sum_{T=t+1}^{\infty} (1-\alpha)^T \left(s_t W^{T-t-1}\right)(u) \\
&= \alpha \sum_{t=0}^{\infty} \sum_{\Delta=0}^{\infty} (1-\alpha)^{\Delta+t+1} \left(s_t W^{\Delta}\right)(u) \\
&\leq \alpha \sum_{t=0}^{\infty} \sum_{\Delta=0}^{\infty} (1-\alpha)^{\Delta+t} \left(s_t W^{\Delta}\right)(u) \\
&= \alpha \sum_{\Delta=0}^{\infty} (1-\alpha)^{\Delta} \left(s W^{\Delta}\right)(u) \\
&= p_s(u)
\end{aligned}$$

$\square$

Let $s_t := q(WI_S)^t(W(I_{S^c}))$ be the soakage vector out of $S \subseteq V$, where I_S is a $|V| \times |V|$ diagonal matrix with $(I_S)_{uu} = 1$ if $u \in S$ and 0 otherwise, and $I_{S^c} := I - I_S$.

Lemma 27. *Let $G = (V, E)$ be an unweighted, undirected graph with associated random walk matrix W. For any $T \in \mathbb{N}, T \geq 1$, $q \in \mathbb{R}^{|V|}$, and $S \subseteq V$*

$$qW^T = \sum_{t=0}^{T-1} s_t W^{T-t-1} + q(WI_S)^T \tag{77}$$

In particular, if $u \in V \setminus S$, then

$$qW^T(u) = \sum_{t=0}^{T-1} \left(s_t W^{T-t-1}\right)(u) \tag{78}$$

Proof. We show (77), from which (78) is an immediate consequence. To show (77), we proceed by induction on T. When $T = 1$,

$$qW = qWI_S + qWI_{S^c} = qWI_S + s_0.$$

Then, for $T \in \mathbb{N},\ T \geq 2$,

$$\begin{aligned}
qW^T &= qW^{T-1}W \\
&= \left\{\sum_{t=0}^{T-2} s_t W^{T-2-t} + q(WI_S)^{T-1}\right\} W && \text{(by the inductive hypothesis)} \\
&= \sum_{t=0}^{T-2} s_t W^{T-1-t} + q(WI_S)^{T-1}(WI_S + WI_{S^c}) \\
&= \sum_{t=0}^{T-1} s_t W^{T-1-t} + q(WI_S)^{T-1}(WI_S)
\end{aligned}$$

and the proof is complete. $\square$

Lemma 28. *Letting s_t, ℓ_t s and ℓ be as in (74),*

$$\|s_t\|_1 = \|\ell_t\|_1\,, \text{ for each } t \geq 0$$

and therefore $\|s\|_1 = \|\ell\|_1$.

Proof. By the definition of s_t and ℓ_t, we have

$$\begin{aligned}
\|s_t\|_1 &= \left\|q_t(W\widetilde{I}^c)\right\|_1 \\
&= \sum_{u \in X} \sum_{u' \in X} q_t(u)(W\widetilde{I}^c)(u, u') \\
&= \sum_{u \in \mathcal{C}_\sigma[X]} \sum_{u' \in \mathcal{C}_\sigma[X]^c} \frac{q_t(u)}{(D)_{uu}} \mathbb{1}((u, u') \in G_{n,r}) \\
&= \sum_{u \in \mathcal{C}_\sigma[X]} \frac{q_t(u)\left((D)_{uu} - (\widetilde{D})_{uu}\right)}{(D)_{uu}} \\
&= \left\|q_t(I - D^{-1}\widetilde{D})\right\|_1 = \|\ell_t\|_1\,.
\end{aligned}$$

$\square$

E Proof of Lower Bound.

To prove Theorem 5, we will proceeding according to the following steps:

1. We study the spectral partitioning properties of PPR on an arbitrary graph G, and show that when suitably initialized inside a subset $S \subset V$, the normalized cut of the PPR sweep cut is upper bounded by (a function of) $\Phi(S; G)$.

2. We specialize to the graph $G = G_{n,r}$ and the subset $\mathcal{L}[X] \subset X$, and show that the normalized cut $\Phi_{n,r}(\mathcal{L}[X])$ is small (with high probability) when the diameter ρ is large.

3. We reason that for the input parameters given in Theorem 5, the output of Algorithm 1 $\widehat{C}$ must therefore also have small normalized cut.

4. On other hand, we show that when the noise parameter ϵ is not too small, the empirical density cluster $\mathcal{C}^{(1)}[X]$ will have large normalized cut $\Phi_{n,r}(\mathcal{C}^{(1)}[X])$. In fact, we generalize this to hold for any set $A \subset X$ for which the symmetric set distance metric $\Delta(A, \mathcal{C}_1[X])$ is small.

5. We conclude that the symmetric set distance metric $\Delta(\widehat{C}, \mathcal{C}^{(1)}[X])$ must not be small.

We devote the subsequent sections to proving each of the aforementioned steps.

E.1 Spectral partitioning properties of PPR.

Let $G = (V, E)$ be an undirected, unweighted graph with $m = |E|$ total edges, defined on vertices $V = \{v_1, \ldots, v_n\}$. Let C be a subset of the vertices V, Recall that for a given $\beta \in (0, 1)$ the sweep cut

$$S_{\beta,v} = \left\{u \in V : \frac{p_v(u)}{\deg(u; G)} > \beta\right\}$$

The following theorem relates the normalized cut of the sweep sets $\Phi(S_\beta; G)$ to the normalized cut of C; it is stated with respect to the graph functionals

$$d_{\max} := \max_{u \in V} \deg(u; G), \quad \text{and} \quad d_{\min} := \min_{u \in V} \deg(u; G).$$

Theorem 8. *Let $C \subseteq V$ satisfy the following conditions:*

- $\mathrm{vol}(C; G) \leq \frac{2}{3}\mathrm{vol}(G)$,
- $|C| \geq \frac{d_{\max}}{d_{\min}}$, *and*
- $\frac{20\Phi(C;G)}{1+10\Phi(C;G)} + \frac{d_{\max}}{2d_{\min}^2} \leq \frac{1}{10}$.

Suppose $60\Phi(C; G) \leq \alpha \leq 70\Phi(C; G)$, and let $(L, U) = (0, 1)$. Then, there exists a subset $C^g \subset C$ with $\mathrm{vol}(C^g; G) \geq \frac{5}{6}\mathrm{vol}(C; G)$ such that for any $v \in C^g$ the following statement holds: For the PPR vector $p_v := p(v, \alpha; G)$, the minimum conductance sweep cut set satisfies

$$\min_{\beta \in (0,1)} \Phi(S_{\beta,v}; G) \leq \sqrt{11200\left\{\log\left(\frac{m}{d_{\min}^2}\right) + \log 20\right\}\Phi(C; G)}$$

Although this theorem appears quite similar to standard results in the PPR literature – for instance, Theorem 6 of Andersen et al. [4] – crucially the above bound depends on $\log\left(\frac{m}{d_{\min}^2}\right)$ rather than $\log m$. In the case where $d_{\min} \asymp n$, this amounts to replacing a factor of $O(\log m)$ by a factor of $O(1)$, and therefore allows us to obtain meaningful results in the limit as $m \to \infty$.

Notwithstanding these improvements, the proof of Theorem 8 follows the same general outline as the proof of Theorem 6 of Andersen et al. [4]. We now walk through this outline step by step, modifying the results of Andersen et al. [4] as needed. As with their work, we begin by proving a mixing time bound on the PPR vector p_v.

E.1.1 Mixing time of PPR.

To quantify the mixing of a PPR vector p_v, we introduce the function $p[\cdot] : [0, 2m] \to [0, 1]$. For $j = 1, \dots, n$, let β_j be the smallest value of $\beta \in (0, 1)$ such that S_{β_j} contains at least j vertices. (For notational ease, we will write $S_i := S_{\beta_i}$, so that $S_1, S_2, \dots, S_n$ comprise the n unique sweep cuts of p_v.) For each $j = 1, \dots, n$, we let $p[\mathrm{vol}(S_j)] = \sum_{u \in S} p_v(u)$. Additionally, we let $p[0] = 0$ and $p[2m] = 1$. Finally, we extend $p[\cdot]$ by piecewise interpolation to be defined everywhere on its domain. The mixedness of the PPR vector is then measured by the function $h : [0, 2m] \to [0, 1]$, defined as

$$h(k) = p[k] - \frac{k}{2m}.$$

Next, for a given $0 \le K_0 \le m$, let

$$L_{K_0}(k) = \frac{2m - K_0 - k}{2m - 2K_0} h(K_0) + \frac{k - K_0}{2m - 2K_0} h(2m - K_0)$$

be the linear interpolator of $h(K_0)$ and $h(2m - K_0)$, and additionally let

$$C(K_0) = \max\left\{\frac{h(k) - L_{K_0}(k)}{\sqrt{\overline{k}}} : K_0 < k < 2m - K_0\right\}.$$

where we use the notation $\overline{k} := \min\{k, 2m - k\}$.

Theorem 9 implies that if the PPR random walk is not well mixed, then some sweep cut of p_v must have small normalized cut.

Theorem 9. *Let $p_v = p(v, \alpha; G)$ be a PPR vector, and let ϕ be any constant in $[0, 1]$. Then, either the following bound holds for any integer t, any $0 < K_0 < m$, and any $k \in [K_0, 2m - K_0]$:*

$$h(k) \le \alpha t + L_{K_0}(k) + C(K_0)\sqrt{\overline{k}}\left(1 - \frac{\phi^2}{8}\right)^t \tag{79}$$

or else there exists some sweep cut S_j of p_v such that $\Phi(S_j; G) < \phi$.

Proof (of Theorem 9). The proof of Theorem 9 is essentially a combination of the proofs of Theorem 3 in Andersen et al. [4] and Theorem 1.2 in Lovász and Simonovits [27]. We will show that if $\Phi(S_j) > \phi$ for each $j = 1, \dots, n$, then (79) holds for all t and any $k \in (K_0, 2m - K_0)$.

We proceed by induction on t. Our base case will be $t = 0$. Observe that $C(K_0) \cdot \sqrt{\overline{k}} \ge h(k) - L_{K_0}(k)$ for all $k \in [K_0, 2m - K_0]$, which implies

$$L_{K_0}(k) + C(K_0) \cdot \sqrt{\overline{k}} \ge h(k).$$

Now, we proceed with the inductive step. By the definition of L_{K_0}, the inequality (79) holds when $k = K_0$ or $k = 2m - K_0$. We will additionally show that (79) holds for every $k_j = \mathrm{vol}(S_j), j = 1, 2, \ldots, n$ such that $k_j \in [K_0, 2m - K_0]$. Once this is shown, the concavity of the expression on the right-hand side of (79) implies that the inequality holds for all $k \in [K_0, 2m - K_0]$.

By Lemma 5 of Andersen et al. [4], we have that

$$\begin{aligned}
p[k_j] &\leq \alpha + \frac{1}{2}\left(p[k_j - |\partial(S_j)|] + p[k_j + |\partial S_j|]\right) \\
&\leq \alpha + \frac{1}{2}\left(p[k_j - \Phi(S_j)\bar{k}_j] + p[k_j + \Phi(S_j)\bar{k}_j]\right) \\
&\leq \alpha + \frac{1}{2}\left(p[k_j - \phi\bar{k}_j] + p[k_j + \phi\bar{k}_j]\right)
\end{aligned}$$

and subtracting $k_j/2m$ from both sides, we get

$$h(k_j) \leq \alpha + \frac{1}{2}\big(h(k_j - \phi\bar{k}_j) + h(k_j + \phi\bar{k}_j)\big) \tag{80}$$

From this point, we divide our analysis into cases.

Case 1. Assume $k_j - 2\phi\bar{k}_j$ and $k_j + 2\phi\bar{k}_j$ are both in $[K_0, 2m - K_0]$. We are therefore in a position to apply our inductive hypothesis to (80), yielding

$$\begin{aligned}
h(k_j) &\leq \alpha + \alpha(t-1)\frac{1}{2}\left(L_{K_0}(k_j - \phi\bar{k}_j) + L_{K_0}(k_j + \phi\bar{k}_j) + C(K_0)\big(\sqrt{k_j - \phi\bar{k}_j} + \sqrt{k_j + \phi\bar{k}_j}\big)\left(1 - \frac{\phi^2}{8}\right)^{t-1}\right) \\
&\leq \alpha t + L_{K_0}(k) + \frac{1}{2}\left(C(K_0)\big(\sqrt{k_j - \phi\bar{k}_j} + \sqrt{k_j + \phi\bar{k}_j}\big)\left(1 - \frac{\phi^2}{8}\right)^{t-1}\right) \\
&\leq \alpha t + L_{K_0}(k) + \frac{1}{2}\left(C(K_0)\big(\sqrt{\bar{k}_j - \phi\bar{k}_j} + \sqrt{\bar{k}_j + \phi\bar{k}_j}\big)\left(1 - \frac{\phi^2}{8}\right)^{t-1}\right).
\end{aligned}$$

A Taylor expansion of $\sqrt{1+\phi}$ around $\phi = 0$ yields the following bound:

$$\sqrt{1+\phi} + \sqrt{1-\phi} \leq 2 - \frac{\phi^2}{4},$$

and therefore

$$h(k_j) \leq \alpha t + L_{K_0}(k) + \frac{C(K_0)}{2} \cdot \sqrt{\bar{k}_j} \cdot \left(2 - \frac{\phi^2}{4}\right)\left(1 - \frac{\phi^2}{8}\right)^{t-1} = \alpha t + L_{K_0}(k) + C(K_0)\sqrt{\bar{k}_j}\left(1 - \frac{\phi^2}{8}\right)^{t}.$$

Case 2.

Now, assume one of $k_j - 2\phi\bar{k}_j$ or $k_j + 2\phi\bar{k}_j$ is not in $[K_0, 2m - K_0]$. Without loss of generality assume $k_j < m$, so that (i) we have $k_j - 2\phi\bar{k}_j < K_0$ and (ii) $k_j + (k_j - K_0) \leq 2m - K_0$. By the concavity of h, and applying the inductive hypothesis to $h(2k_j - K_0)$, we have

$$\begin{aligned}
h(k_j) &\leq \alpha + \frac{1}{2}\Big(h(K_0) + h\big(k_j + (k_j - K_0)\big)\Big) \\
&\leq \alpha + \frac{\alpha(t-1)}{2} + \frac{1}{2}\left(L_{K_0}(K_0) + L_{K_0}(2k_j - K_0) + C(K_0)\sqrt{2k_j - K_0}\left(1 - \frac{\phi^2}{8}\right)^{t-1}\right) \\
&\leq \alpha t + L_{K_0}(k_j) + C(K_0)\frac{\sqrt{2\bar{k}_j}}{2}\left(1 - \frac{\phi^2}{8}\right)^{t-1} \\
&\leq \alpha t + L_{K_0}(k_j) + C(K_0)\sqrt{\bar{k}_j} \cdot \left(1 - \frac{\phi^2}{8}\right)^{t}
\end{aligned}$$

□

As a sanity check, we confirm that Theorem 9 is no weaker than Theorem 3 of Andersen et al. [4]. It is not hard to show that $h(k) \leq \min\{1, \sqrt{k}\}$, and therefore that $C(K_0) \leq 1$ for any K_0. Setting $K_0 = 0$ in Theorem 9, we therefore recover Theorem 3 of Andersen et al. [4].

We now proceed to identify when Theorem 9 may offer some improvement on Theorem 3 of Andersen et al. [4], by showing when we can upper bound $C(K_0) << 1$. The critical point is that since $h(k)$ is concave and $L_{K_0}(K_0) = h(K_0)$ the upper bound

$$\frac{h(k) - L_{K_0}(k)}{\sqrt{k}} \leq h'(K_0)\sqrt{k}$$

holds whenever $k < m$. For similar reasons, when $k > m$,

$$\frac{h(k) - L_{K_0}(k)}{\sqrt{k}} \leq -h'(2m - K_0)\sqrt{2m - k}.$$

(Since h is not differentiable at points $k = \mathrm{vol}(S_j)$, here we use h' to denote the left derivative of h whenever $k < m$, and the right derivative of h whenever $k \geq m$)

The following Lemma gives good estimates for $h'(K_0)$ and $h'(2m - K_0)$, and a resulting upper bound on $C(K_0)$.

Lemma 29. *There exists $K_0 \in \{0, \deg(v; G)\}$ such that*

$$h'(K_0) \leq \frac{1}{2d_{\min}^2}. \tag{81}$$

Additionally, for all $K_0 \in [0, 2m]$,

$$h'(2m - K_0) \geq -\frac{d_{\max}}{d_{\min}\mathrm{vol}(G)}. \tag{82}$$

As a result,

$$C(K_0) \leq \frac{\sqrt{m}}{d_{\min}^2}.$$

Proof (of Lemma 29). The result of the Lemma is obvious once we show (81) and (82). To show either inequality, it will be useful to work with an alternative representation of h. In particular, whenever $\mathrm{vol}(S_j) \leq k < \mathrm{vol}(S_{j+1})$ (where we let $S_0 = \emptyset$), the function $h(k)$ may be written as

$$h(k) = \sum_{i=0}^{j} \big(p_v(u_{(i)}) - \pi(u_{(i)}; G)\big) + \frac{\big(k - \mathrm{vol}(S_j; G)\big)}{\deg(u_{(j+1)}; G)} \big(p_v(u_{(j+1)}) - \pi(u_{(j+1)}; G)\big) \tag{83}$$

where the vertices are ordered $\frac{p_v(u_{(1)})}{\deg(u_{(1)};G)} \geq \frac{p_v(u_{(2)})}{\deg(u_{(2)};G)} \geq \cdots \geq \frac{p_v(u_{(n)})}{\deg(u_{(n)};G)}$, and as usual $\pi(u; G) = \frac{\deg(u;G)}{\mathrm{vol}(G)}$.

From this representation, it is not hard to verify that the left derivative $h'(k)$ can be upper bounded

$$h'(k) \leq \frac{p(v_{(j+1)})}{\deg(v_{(j+1)}; G)} \tag{84}$$

We now upper bound $p(u)$ uniformly over all u except the seed node v. For any $u \in V$ besides the seed node v, we can show by induction that

$$e_v W^t(u) \leq \frac{1}{2d_{\min}}$$

for any $t \geq 0$, and therefore

$$p(\alpha, \chi_v)(u) = \alpha \sum_{t=0}^{\infty} (1-\alpha)^t \chi_v W^t(u) \leq \frac{1}{2d_{\min}}. \tag{85}$$

As a result, by (84), for either $K_0 = \deg(v; G)$ (in the case where $v_{(1)} = v$) or otherwise for $K_0 = 0$, the inequality $h'(K_0) \leq \frac{1}{2d_{\min}^2}$ holds, proving (81). The inequality (82) follows immediately from the representation (83), since

$$h'(k) \geq -\frac{\pi(v_{(j+1)})}{d(v_{(j+1)})} \geq -\frac{\pi_{\max}}{d_{\min}},$$

and the proof of the Lemma is therefore complete. □

To apply Theorem 9, we must also upper bound the linear interpolator $L_{K_0}(k)$. Of course, trivially $L_{K_0}(k) \leq \max\{h(K_0), h(2m - K_0)\}$ for all k. As it happens, this observation will lead to a sufficient upper bound on L_{K_0}.

Lemma 30. *Assume $s = \chi_v$ for some $v \in V$. Let $K_0 = \mathrm{vol}(S_j)$ for some $j = 0, \ldots, n$. Then,*

$$h(2m - K_0) \leq \frac{K_0}{2m} \text{ and } h(K_0) \leq \frac{K_0}{2d_{\min}^2} + \frac{2\alpha}{1+\alpha}.$$

and as a result for any $k \in \mathbb{R}$,

$$L_{K_0}(k) \leq \frac{2\alpha}{1+\alpha} + \frac{K_0}{2d_{\min}^2}.$$

Proof (of Lemma 30). We make use of the representation (83) to prove the desired upper bounds on $h(2m - K_0)$ and $h(K_0)$. We first upper bound $h(2m - K_0)$,

$$\begin{aligned} h(2m - K_0) &= \sum_{i=1}^{j} p(v_{(i)}) - \pi(v_{(i)}) \\ &\leq 1 - \sum_{i=1}^{j} \pi(v_{(i)}) \\ &= 1 - \sum_{i=1}^{j} \frac{d(v_i)}{2m} = \frac{K_0}{2m}. \end{aligned}$$

We will upper bound $h(K_0)$ by $p[\mathrm{vol}(S_j)] \leq p_v(v) + \sum_{u \in S_j \setminus \{v\}} p_v(u)$. In the proof of Lemma 29 we have already given an upper bound on $p_v(u)$ when $u \neq v$. Now, we additionally observe that for all t,

$$e_v W^t(v) \leq \frac{1}{2d_{\min}} + \left(\frac{1}{2}\right)^t$$

and therefore $p_v(v) \leq \frac{1}{2d_{\min}} + \frac{2\alpha}{1+\alpha}$. As a result,

$$h(K_0) \leq \frac{2\alpha}{1+\alpha} + \frac{|S_j|}{2d_{\min}} \leq \frac{2\alpha}{1+\alpha} + \frac{K_0}{2d_{\min}^2}, \tag{86}$$

where the latter inequality follows since $K_0 = \mathrm{vol}(S_j) \geq |S_j| \cdot d_{\min}$. □

Combining Theorem 9, Lemma 29 and Lemma 30, we have the following result.

Corollary 2. *Let $p_v = p(v, \alpha; G)$ be a PPR vector with seed node $v \in V$, and let ϕ be any constant in $[0, 1]$. Then, either the following bound holds for any integer t and any $k \in [d_{\max}, 2m - d_{\min}]$:*

$$h(k) \leq \alpha t + \frac{2\alpha}{1+\alpha} + \frac{d(v)}{2d_{\min}^2} + \frac{\sqrt{m}}{d_{\min}^2} \cdot \sqrt{k}\left(1 - \frac{\phi^2}{8}\right)^t$$

or there exists some sweep cut S_j of p_v such that $\Phi(S_j; G) < \phi$.

We arrive now at the main result of this section. It is similar in form to Theorem 2 of Andersen et al. [4] but reflects the improvements due to using Corollary 2. To simplify notation, we will write the total mass placed by p_v on a subset $S \subset V$ as $p_v(S) := \sum_{u \in S} p_v(u)$.

Theorem 10. *Let $p_v = p(v, \alpha; G)$ be a PPR vector with seed node $v \in V$. Suppose there exists some $\delta > \frac{2\alpha}{1+\alpha} + \frac{d_{\max}}{2d_{\min}^2}$, such that*

$$p_v(S) - \frac{\mathrm{vol}(S; G)}{\mathrm{vol}(G)} > \delta \tag{87}$$

for a set S with cardinality $|S| \geq \frac{d_{\max}}{d_{\min}}$. Then there exists a sweep cut S_j of p, such that

$$\Phi(S_j) < \sqrt{\frac{16\alpha\left\{\log\left(\frac{m}{d_{\min}^2}\right) + \log\left(\frac{2}{\delta'}\right)\right\}}{\delta'}}$$

where $\delta' = \delta - \frac{2\alpha}{1+\alpha} + \frac{d(v)}{2d_{\min}^2}$.

Proof. Suppose the assumption of the theorem is satisified, that is there exists a set $S \subset V$ with cardinality $|S| \geq \frac{d_{\max}}{d_{\min}}$ which satisfies (87). Then for $j = |S|$ the sweep cut S_j has volume at least $d_{\max}$, and by hypothesis $h(\mathrm{vol}(S_j)) > \delta$.

Now, letting

$$t = \frac{8}{\phi^2}\left\{\log\left(\frac{m}{d_{\min}^2}\right) + \log\left(\frac{2}{\delta'}\right)\right\}, \quad \phi^2 = \frac{16\alpha\left\{\log\left(\frac{m}{d_{\min}^2}\right) + \log(\frac{2}{\delta'})\right\}}{\delta'}$$

we have that

$$\alpha t + \frac{2\alpha}{1+\alpha} + \frac{d(v)}{2d_{\min}^2} + \frac{\sqrt{m}}{d_{\min}^2} \cdot \sqrt{k}\left(1 - \frac{\phi^2}{8}\right)^t \leq \frac{\delta'}{2} + \frac{2\alpha}{1+\alpha} + \frac{d(v)}{2d_{\min}^2} + \frac{\delta'}{2} < \delta,$$

and the Theorem follows by Corollary 2. □

E.1.2 Improved Local Partitioning with PPR.

As in Andersen et al. [4], the mixing time results of the previous section lead to an upper bound on the normalized cut $\Phi(\widehat{C}; G)$. First, we restate a theorem of Andersen et al. [4] which lower bounds the probability mass $p(v, \alpha; G)(C)$ as a function of the normalized cut $\Phi(C)$.

Theorem 11. *For any set C and any constant α, there exists a subset $C^g \subset C$ with $\mathrm{vol}(C^g; G) \geq \frac{5}{6}\mathrm{vol}(C; G)$, such that for any vertex $v \in C^g$, the PPR vector $p(v, \alpha; G)$ satisfies*

$$p(v, \alpha; G) \geq 1 - 6\frac{\Phi(C; G)}{\alpha}.$$

We are now in a position to prove Theorem 8 by combining Corollary 2 and Theorem 11.

Proof (of Theorem 8). Since $\alpha \geq 60\Phi(C)$ and $v \in C^g$, by Theorem 11,

$$p_v(C) \geq \frac{9}{10}.$$

This inequality along with the assumption $\mathrm{vol}(C) \leq \frac{2}{3}\mathrm{vol}(G)$ implies that $p_v(C) - \frac{\mathrm{vol}(C)}{\mathrm{vol}(G)} \geq \frac{1}{5}$. Since we assume $|C| \geq \frac{d_{\max}}{d_{\min}}$, the hypothesis of Theorem 10 is satisfied with $\delta = 1/5$. Therefore, the minimum conductance sweep cut satisfies

$$\min_{j=1,\ldots,n} \Phi(S_j; G) \leq \sqrt{\frac{1120 \cdot \Phi(C; G)\left\{\log\left(\frac{m}{d_{\min}^2}\right) + \log\left(\frac{2}{\delta'}\right)\right\}}{\delta'}}$$

Finally, we assume $\frac{20\Phi(C)}{1+10\Phi(C)} + \frac{d_{\max}}{2d_{\min}^2} \leq \frac{1}{10}$ which implies that

$$\delta' = \delta - \frac{20\alpha}{1+10\alpha} + \frac{d_{\max}}{2d_{\min}^2} \geq \frac{1}{10}$$

completing the proof of the theorem. □

E.2 Normalized cut of $\mathcal{L}[X]$.

Recall that for any set $\mathcal{A} \subset \mathcal{X}$, the $\mathbb{P}$-weighted *cut* and *volume* functionals can be written as

$$\mathrm{cut}_{\mathbb{P},r}(\mathcal{A}) = \int_{\mathcal{A}}\int_{\mathcal{X}\setminus\mathcal{A}} \mathbb{1}(\|x-y\| \leq r)\, d\mathbb{P}(x)\, d\mathbb{P}(y), \mathrm{vol}_{\mathbb{P},r}(\mathcal{A}) := \int_{\mathcal{A}}\int_{\mathcal{X}} \mathbb{1}(\|x-y\| \leq r)\, d\mathbb{P}(x)\, d\mathbb{P}(y),$$

and the continuous *normalized cut* is

$$\Phi_{\mathbb{P},r}(\mathcal{A}) := \frac{\mathrm{cut}_{\mathbb{P},r}(\mathcal{A})}{\min\{\mathrm{vol}_{\mathbb{P},r}(\mathcal{A}), \mathrm{vol}_{\mathbb{P},r}(\mathcal{X}\setminus\mathcal{A})\}}.$$

We now upper bound the normalized cut $\Phi_{\mathbb{P},r}(\mathcal{L})$ as a function of the diameter ρ, and the neighborhood graph radius r. Our bounds will be simple and not tight, but will display the right dependence on these parameters, and so will be sufficient for our purposes.

To upper bound $\mathrm{cut}_{\mathbb{P},r}(\mathcal{L})$, note that for any $x = (x_1, x_2) \in \mathcal{L}$, if $x_2 \leq -r$ the ball $B(x, r)$ and the set $\mathcal{X} \setminus \mathcal{L}$ are disjoint. This implies

$$\begin{aligned}
\mathrm{cut}_{\mathbb{P},r}(\mathcal{L}) &\leq \mathbb{P}(\{x \in \mathcal{X} : -r < x_2 < 0\}) \cdot \max_{x\in\mathcal{X}} \mathbb{P}(B(x,r)) \\
&\leq \frac{r}{2\rho} \cdot \frac{\pi r^2}{2\sigma\rho}.
\end{aligned}$$

By symmetry, $\mathrm{vol}_{\mathbb{P},r}(\mathcal{L}) = \mathrm{vol}_{\mathbb{P},r}(\mathcal{X} \setminus \mathcal{L})$, and therefore to upper bound $\Phi_{\mathbb{P},r}(\mathcal{L})$, it is sufficient to lower bound $\mathrm{vol}_{\mathbb{P},r}(\mathcal{L})$. We have

$$\begin{aligned}
\mathrm{vol}_{\mathbb{P},r}(\mathcal{L}) &\geq \mathbb{P}(\{x \in \mathcal{C}_1 \cap \mathcal{L} : \mathrm{dist}(x, \partial \mathcal{C}_1) > r\}) \cdot \frac{\pi r^2}{2\sigma\rho} \\
&= \frac{(\sigma - 2r)(\rho - r)}{2\sigma\rho} \cdot \frac{\nu_d r^d}{2\sigma\rho} \\
&\geq \frac{3}{16} \cdot \frac{\pi r^2}{2\sigma\rho}
\end{aligned}$$

where the last inequality follows since $r \leq \frac{1}{4}\sigma < \frac{1}{4}\rho$. Therefore, $\Phi_{\mathbb{P},r}(\mathcal{L}) \leq \frac{8r}{3\rho}$.

Then, Lemma 36 implies that the graph functionals $\mathrm{cut}_{n,r}(\mathcal{L}[X])$ and $\mathrm{vol}_{n,r}(\mathcal{L}[X])$–and in turn $\Phi_{n,r}(\mathcal{L}[X])$– concentrate around their expectations. Precisely, we have that

$$\begin{aligned}
\Phi_{n,r}(\mathcal{L}[X]) &= \frac{\mathrm{cut}_{n,r}(\mathcal{L}[X])}{\min\{\mathrm{vol}_{n,r}(\mathcal{L}[X]), \mathrm{vol}_{n,r}((\mathcal{X} \setminus \mathcal{L})[X])\}} \\
&\leq \frac{3}{2}\Phi_{\mathbb{P},r}(\mathcal{A}) \leq \frac{4r}{\rho}
\end{aligned} \tag{88}$$

with probability at least $1 - 3\exp\{-\frac{1}{25}n(\mathrm{cut}_{\mathbb{P},r}(\mathcal{L}))^2\}$.

E.3 Normalized cut of $\widehat{C}$.

To upper bound $\Phi_{n,r}(\widehat{C})$, we need to show that the conditions of Theorem 8 are met with respect to the graph $G = G_{n,r}$ and subset $C = \mathcal{L}[X]$. To do, we require the following inequalities, which are satisfied with probability at least $1 - 2n\exp\left\{-\frac{\pi\epsilon r^2\delta^2 n}{8\rho\sigma(1+\frac{\delta}{3})}\right\} - 6\exp\left\{-n\delta^2(\mathrm{cut}_{\mathbb{P},r}(\mathcal{L}[X]))^2\right\}$:

- For any $r \in (0, \frac{1}{4}\sigma)$ and any $x \in \mathcal{X}$,

  $$\frac{\epsilon}{4\rho\sigma}\pi r^2 \leq \mathbb{P}((B(x,r))) \leq \frac{1}{2\rho\sigma}\pi r^2.$$

  Therefore by Lemma 34, $\frac{(1-\delta)\epsilon\pi r^2}{4\rho\sigma}n \leq d_{\min} \leq d_{\max} \leq \frac{(1+\delta)\pi r^2}{2\rho\sigma}n$.

- $\frac{(1-\delta)}{2}n \leq |\mathcal{L}[X]| \leq \frac{(1+\delta)}{2}n$,

- $\mathrm{vol}_{n,r}(\mathcal{L}[X]) \leq (1+\delta)\mathrm{vol}_{\mathbb{P},r}(\mathcal{L}) = \frac{(1+\delta)}{2}\mathrm{vol}_{\mathbb{P},r}(\mathcal{X}) \leq \frac{(1+\delta)}{2}\mathrm{vol}(G_{n,r})$, and

- $(1-\delta)\mathrm{cut}_{\mathbb{P},r}(\mathcal{L}[X]) \leq \mathrm{cut}_{n,r}(\mathcal{L}[X]) \leq (1+\delta)\mathrm{cut}_{\mathbb{P},r}(\mathcal{L}[X])$.

We now condition on these inequalities, and letting $\delta = \frac{2}{67}$ we verify that under the setup of Theorem 5, each of the conditions of Theorem 8 are met:

- $\mathrm{vol}(\mathcal{L}[X]) \leq \frac{(1+\delta)}{2(1-\delta)}\mathrm{vol}(G_{n,r}) \leq \frac{2}{3}\mathrm{vol}(G_{n,r})$ since $\delta < 1/7$,

- $|\mathcal{L}[X]| \geq \frac{n(1-\delta)}{2} \geq \frac{2(1+\delta)}{(1-\delta)\epsilon} \geq \frac{d_{\max}}{d_{\min}}$ and $\frac{d_{\max}}{2d_{\min}^2} \leq \frac{8(1+\delta)}{(1-\delta)^2\epsilon^2\rho\sigma\pi r^2} \cdot \frac{1}{n} \leq \frac{1}{10}$ by the assumed lower bound on the sample size,

- $\Phi_{n,r}(\mathcal{L}[X]) \leq \frac{4r}{\rho} \leq \frac{1}{10}$, by assumption on r and ρ, and

- $60\Phi_{n,r}(\mathcal{L}[X]) \leq \frac{60(1+\delta)}{1-\delta}\Phi_{\mathbb{P},r}(\mathcal{L}) \leq \alpha \leq \frac{65(1+\delta)}{1-\delta}\Phi_{n,r}(\mathcal{L}[X]) \leq 70\Phi_{n,r}(\mathcal{L}[X])$ since $\delta < 2/67$.

We may therefore apply Theorem 8, which allow us to upper bound the minimum conductance sweep cut $\min_{\beta \in (0,1)} \Phi(S_{\beta,v}; G)$ or equivalently the output of Algorithm 1.

To be precise, we have that there exists a set $\mathcal{L}[X]^g \subset \mathcal{L}[X]$ with $\mathrm{vol}_{n,r}(\mathcal{L}[X]^g) \geq \frac{5}{6}\mathrm{vol}_{n,r}(\mathcal{L}[X])$, such that the following statement holds for any $v \in \mathcal{L}[X]^g$: when Algorithm 1 is run with inputs $X, r < \frac{1}{4}\sigma, \alpha = 65\Phi_{\mathbb{P},r}(\mathcal{L}[X]), v \in \mathcal{L}[X]^g$ and $(L,U) = (0,1)$, the resulting PPR cluster estimate $\widehat{C}$ satisfies

$$\begin{aligned}
\Phi_{n,r}(\widehat{C}) &\leq \sqrt{11200\left\{\log\left(\frac{m}{d_{\min}^2}\right) + \log 20\right\}\Phi_{n,r}(\mathcal{L}[\mathcal{X}])} \\
&\leq \sqrt{89600\left\{\log\left(\frac{\rho\sigma}{\epsilon^2\pi r^2}\right) + \log 20\right\}\frac{r}{\rho}} \qquad (89)
\end{aligned}$$

with probability at least $1 - 2n\exp\left\{-\frac{\pi\epsilon r^2 n}{8978\rho\sigma}\right\} - 6\exp\left\{-\frac{1}{1123}(\mathrm{cut}_{\mathbb{P},r}(\mathcal{L}))^2 n\right\}$ (where the latter inequality follows from (88) and Lemma 36.)

E.4 Lower bound on normalized cut.

The precise statement we will prove is contained in the following Lemma.

Lemma 31. *The normalized cut $\Phi_{n,r}(A)$ is upper bounded*

$$\Phi_{n,r}(A) \geq \frac{1}{12\pi}\left(1 - 4\frac{\sigma\rho}{r^2n^2}\mathrm{vol}_{n,r}(A \vartriangle \mathcal{C}^{(1)}[X])\right)\frac{\epsilon^2 r}{\sigma}$$

uniformly over all $A \subset X$ with probability at least $1 - \exp\left\{-2n\delta^2(\mathrm{vol}_{\mathbb{P},r}(\mathcal{X}))^2\right\} - \frac{12\sigma\rho}{r^2}\exp\left\{-\frac{\delta^2\epsilon r^2 n}{\rho\sigma(3+\delta)}\right\} - \frac{2\rho}{r}\exp\left\{-\frac{\delta^2\pi r^3 n}{2\sigma\rho^2(3+\delta)}\right\}$.

Proof. To lower bound the normalized cut $\Phi_{n,r}(A)$, we must lower bound $\mathrm{cut}_{n,r}(A)$ and upper bound $\mathrm{vol}_{n,r}(A)$. A naive upper bound on the volume is simply

$$\mathrm{vol}_{n,r}(A) \leq \mathrm{vol}_{n,r}(G_{n,r}) \overset{(i)}{\leq} (1+\delta)\mathrm{vol}_{\mathbb{P},r}(\mathcal{X})n^2 \leq (1+\delta)\frac{\pi r^2}{\rho\sigma}n^2 \qquad (90)$$

where (i) holds with probability at least $1 - \exp\left\{-2n\delta^2(\mathrm{vol}_{\mathbb{P},r}(\mathcal{X}))^2\right\}$, and it turns out this will suffice for our purposes. (Here and in the rest of this proof we take $\delta = 1/2$.)

We turn to lower bounding $\mathrm{cut}_{n,r}(A)$. We will approximate the cut of A by discretizing the space $\mathcal{X}$ into bins, relate the cut of A to the boundary of the binned set $\overline{A}$, and then lower bound the size of the boundary of $\overline{A}$.

Let (k_1, k_2) for $k_1 \in \left[\frac{6\sigma}{r}\right], k_2 \in \left[\frac{2\rho}{r}\right]$ be the upper right corner of the cube

$$Q_{(k_1,k_2)} = \left[-\frac{3\sigma}{2} + \frac{(k_1-1)}{2}r, -\frac{3\sigma}{2} + \frac{k_1}{2}r\right] \times \left[-\frac{\rho}{2} + \frac{(k_2-1)}{2}r, -\frac{\rho}{2} + \frac{k_2}{2}r\right]$$

and let $\overline{Q} = \left\{Q_{(k_1,k_2)} : k_1 \in \left[\frac{6\sigma}{r}\right], k_2 \in \left[\frac{2\rho}{r}\right]\right\}$ be the collection of such cubes. For a set $A \subset X$ we define the binned set $\overline{A} \subset \overline{Q}$ as follows

$$\overline{A} := \left\{Q \in \overline{Q} : \mathbb{P}_n(A \cap Q) \geq \frac{1}{2}\mathbb{P}_n(Q)\right\},$$

and we let

$$\partial\overline{A} := \left\{ Q_{(k_1,k_2)} \in \overline{A} : \exists(\ell_1,\ell_2) \in \left[\frac{3\sigma}{r}\right] \times \left[\frac{\rho}{r}\right] \text{ such that } Q_{(\ell_1,\ell_2)} \notin \overline{A}, \|k-\ell\|_1 = 1 \right\}.$$

be the boundary set of $\overline{A}$ in $\overline{Q}$. Intuitively, every point $x_i \in A$ in the boundary set of $\overline{A}$ will have many edges to $X \setminus A$. Formally, letting $Q_{\min} := \min_{Q \in \overline{Q}} \mathbb{P}_n(Q)$, we have

$$\mathrm{cut}_{n,r}(A) \geq \mathrm{cut}_{n,r}(A \cap \{x_i \in \overline{A}\}) \geq \frac{1}{4} \left|\partial\overline{A}\right| Q_{\min}^2, \tag{91}$$

where the last inequality follows since for every cube $Q_k \in \partial\overline{A}$, there exists a cube $Q_\ell \notin \overline{A}$ such that $\|i-j\|_1 \leq 1$, and since each cube has side length $r/2$, this implies that for every $x_i \in Q_k$ and $x_j \in Q_\ell$ the edge (x_i,x_j) belongs to $G_{n,r}$.

Now we move on lower bounding the size of the boundary $\left|\partial\overline{A}\right|$. To do so, we divide $\mathcal{X}$ into slices horizontally. Let $R_k = \left\{(x_1,x_2) \in \mathcal{X} : x_2 \in \left[-\frac{\rho}{2} + \frac{(k-1)}{2}r, -\frac{\rho}{2} + \frac{k}{2}r\right]\right\}$ be the kth horizontal slice, and $\overline{R}_k = \left\{Q_{(k_1,k)} \in \overline{Q} : k_1 \in [\frac{6\sigma}{r}]\right\}$ be the binned version of R_k. For each k, either

1. $\overline{R}_k \cap \overline{A} = \emptyset$, in which case
$$\mathrm{vol}_{n,r}\Big(\big(A \vartriangle C_1[X]\big) \cap R_k\Big) \geq \frac{1}{2}\mathrm{vol}_{n,r}(C_1[X] \cap R_k), \quad \text{or}$$
2. $\overline{R}_k \cap \overline{A} = \overline{R}_k$, in which case
$$\mathrm{vol}_{n,r}\Big(\big(A \vartriangle C_1[X]\big) \cap R_k\Big) \geq \frac{1}{2}\mathrm{vol}_{n,r}(C_2[X] \cap R_k), \quad \text{or}$$
3. $\overline{R}_k \cap \partial\overline{A} \neq \emptyset$.

Let $N(R)$ be the number of slices for which $\overline{R}_k \cap \partial\overline{A} \neq \emptyset$. By the cases elucidated above, letting

$$R_{\min} := \min_k \Big\{\mathrm{vol}_{n,r}(C_1[X] \cap R_k) \wedge \mathrm{vol}_{n,r}(C_2[X] \cap R_k)\Big\}$$

we obtain the following lower bound on the volume of the symmetric set difference,

$$\mathrm{vol}_{n,r}(A \vartriangle C_1[X]) \geq \frac{1}{2} R_{\min} \left[\frac{2\rho}{r} - N(R)\right]. \tag{92}$$

Finally note that $\left|\partial\overline{A}\right| \geq N(R)$. Therefore combining (91) and (92), we have that

$$\begin{aligned}\mathrm{cut}_{n,r}(A) &\geq \frac{1}{4} N(R) Q_{\min}^2 \\ &\geq \frac{1}{2}\left(\frac{\rho}{r} - \frac{\mathrm{vol}_{n,r}(A \vartriangle C_1[X])}{R_{\min}}\right) Q_{\min}^2 \end{aligned} \tag{93}$$

for all $A \subset X$.

It remains to lower bound the random quantities $R_{\min}$ and $Q_{\min}$. To do so, we first lower bound the expected probability of any cell Q,

$$\min_{Q \in \overline{Q}} \mathbb{P}(Q) \geq \frac{\epsilon r^2}{\rho\sigma}.$$

and the expected volume of $\mathcal{C}^{(1)}[X] \cap R_k$ and $\mathcal{C}^{(2)}[X] \cap R_k$,

$$\mathrm{vol}_{\mathbb{P},r}(\mathcal{C}^{(1)} \cap R_k) = \mathrm{vol}_{\mathbb{P},r}(\mathcal{C}^{(2)} \cap R_k) \geq \frac{\pi r^3}{2\sigma\rho^2} \tag{94}$$

Since $Q_{\min}$ and $R_{\min}$ are obtained by taking the minimum of functionals over a fixed number of sets in n, they concentrate tightly around their means. Specifically, note that the total number of cubes is $|\overline{Q}| = \frac{12\sigma\rho}{r^2}$, and the total number of horizontal slices is $\frac{2\rho}{r}$. Along with (93) and (94), by Lemma 34

$$Q_{\min} \geq (1-\delta)\frac{\epsilon r^2}{\rho\sigma} \quad \text{and} \quad R_{\min} \geq \frac{(1-\delta)}{2}\frac{\pi r^3}{\sigma\rho^2},$$

with probability at least $1 - \frac{12\sigma\rho}{r^2}\exp\left\{-\frac{\delta^2\epsilon r^2 n}{\rho\sigma(3+\delta)}\right\} - \frac{2\rho}{r}\exp\left\{-\frac{\delta^2\pi^2 r^6 n}{4\sigma^2\rho^4}\right\}$. Combining these lower bounds with (90) and (93), we obtain

$$\Phi_{n,r}(A) \geq \frac{(1-\delta)^2}{2(1+\delta)\pi}\left(1 - 2\frac{\sigma\rho}{(1-\delta)r^2n^2}\mathrm{vol}_{n,r}(A \vartriangle \mathcal{C}^{(1)}[X])\right)\frac{\epsilon^2 r}{\sigma},$$

and plugging in $\delta = 1/2$ yields the claim of Lemma 31. □

Conclusion. Combining (89) and Lemma 31, we have that there exists a set $\mathcal{L}[X]^g \subset \mathcal{L}[X]$ with $\mathrm{vol}_{n,r}(\mathcal{L}[X]^g) \geq \frac{5}{6}\mathrm{vol}_{n,r}(\mathcal{L}[X])$ such that for any seed node $v \in \mathcal{L}[X]^g$, the following bounds hold:

$$\frac{1}{12\pi}\left(1 - 4\frac{\sigma\rho}{r^2n^2}\mathrm{vol}_{n,r}(\widehat{C} \vartriangle \mathcal{C}^{(1)}[X])\right)\frac{\epsilon^2 r}{\sigma} \leq \Phi_{n,r}(\widehat{C}) \leq \sqrt{89600\left\{\log\left(\frac{\rho\sigma}{\epsilon^2\pi r^2}\right) + \log 20\right\}\frac{r}{\rho}},$$

with probability at least $1 - b_2/n$ for an appropriate choice of constant b_2. Solving for $\mathrm{vol}_{n,r}(\widehat{C} \vartriangle \mathcal{C}^{(1)}[X])$ in the previous equation, we obtain (19) (for an appropriate choice of constant c). Finally, we show that the volume of $\mathcal{L}[X]^g$ is sufficiently large to ensure that it includes many points in $\mathcal{C}^{(1)}[X]$:

$$\begin{aligned}
\mathrm{vol}_{n,r}(\mathcal{L}[X]^g \cap \mathcal{C}^{(1)}[X]) &\geq \mathrm{vol}_{n,r}(\mathcal{L}[X]^g) - \mathrm{vol}_{n,r}((\mathcal{L}[X]^g \cap (\mathcal{C}^{(0)} \cup \mathcal{C}^{(2)})[X])[X]) \\
&\geq \frac{5}{6}\mathrm{vol}_{n,r}(\mathcal{L}[X]) - \mathrm{vol}_{n,r}((\mathcal{L} \cap (\mathcal{C}^{(0)} \cup \mathcal{C}^{(1)})[X])[X]) \\
&\geq \frac{5}{6}\mathrm{vol}_{n,r}((\mathcal{L} \cap \mathcal{C}^{(1)})[X]) - \frac{1}{6}\mathrm{vol}_{n,r}(\mathcal{L}[X]) \\
&\geq \left(\frac{5}{6}(1-\delta) - \frac{1}{2}(1+\delta)\right)\mathrm{vol}_{\mathbb{P},r}(\mathcal{L} \cap \mathcal{C}^{(1)}) \\
&\geq \frac{(1-\delta)}{2}\left(\frac{5}{6}(1-\delta) - \frac{1}{2}(1+\delta)\right)\mathrm{vol}_{n,r}(\mathcal{C}^{(1)}[X])
\end{aligned}$$

where the final two inequalities follow from Lemma 36 and hold with probability at least $1 - 3\exp\left\{-n\delta^2(\mathrm{vol}_{\mathbb{P},r}(\mathcal{L} \cap \mathcal{C}^{(1)}))^2\right\}$. Setting $\delta = 1/13$, we have that $\mathrm{vol}_{n,r}(\mathcal{L}[X]^g \cap \mathcal{C}^{(1)}[X]) \geq \frac{1}{10}\mathrm{vol}_{n,r}(\mathcal{C}^{(1)}[X])$.

F Bounding the misclassification rate.

A common loss function for clustering is the misclassification rate.

Definition 5 (Misclassification rate.)**.** For an estimator $\widehat{C} \subseteq X$ and set $\mathcal{S} \subseteq \mathbb{R}^d$, we define

$$\Delta_{\mathrm{mc}}(\widehat{C}, \mathcal{S}) := |\mathcal{C} \vartriangle \mathcal{S}[X]|\,, \tag{95}$$

the cardinality of the symmetric set difference between $\widehat{C}$ and $\mathcal{S} \cap X = \mathcal{S}[X]$.

In order to upper bound the misclassification rate of a PPR cluster estimate, we will need to slightly modify our approach to computing sweep cuts, and no longer normalize by degree; we formally define this modified algorithm in Algorithm 2. Intuitively, this change helps us avoid including many low-degree vertices $w \notin \mathcal{C}_\sigma[X]$ in our estimated cluster.

Algorithm 2 Unnormalized PPR on a neighborhood graph

Input: data $X = \{x_1, \ldots, x_n\}$, radius $r > 0$, teleportation parameter $\alpha \in [0, 1]$, seed $v \in X$, target stationary probability $\pi_0 > 0$, range (L, U).
Output: cluster $\widehat{C} \subseteq V$.

1: Form the neighborhood graph $G_{n,r}$.
2: Compute the PPR vector $p_v = p(v, \alpha; G_{n,r})$ as in (1).
3: For $\beta \in (\frac{1}{40}, \frac{1}{11})$ compute sweep cuts S_β

$$S_\beta := \{u \in V : p_v(u) > \beta\pi_0\}\,. \tag{96}$$

4: Return as a cluster $\widehat{C}_{\mathrm{un}} = S_{\beta^*}$, where

$$\beta^* = \operatorname*{arg\,min}_{\beta \in (L,U)} \Phi(S_\beta; G_{n,r}).$$

We say Algorithm 2 is well-initialized if r, α and v satisfy (8), and additionally

$$\pi_0 = \frac{\lambda_\sigma}{\Lambda_\sigma \mathbb{P}(\mathcal{C}_\sigma) n}, \quad \text{and} \quad (L, U) \in \left(\frac{1}{100}, \frac{1}{50}\right). \tag{97}$$

Theorem 12. *Fix $\lambda > 0$ let $\mathcal{C} \in \mathbb{C}_f(\lambda)$ be a κ-well-conditioned density cluster, and assume If Algorithm 2 is well-initialized with respect to $\mathcal{C}$. Then for any*

$$n \geq b_1 (\log n)^{\max\{\frac{3}{d}, 1\}}$$

there exists a set $\mathcal{C}_\sigma[X]^g \subseteq \mathcal{C}_\sigma[X]$ of large volume, $\mathrm{vol}_{n,r}(\mathcal{C}_\sigma[X]^g) \geq \mathrm{vol}_{n,r}(\mathcal{C}_\sigma[X])/2$, such that the following holds: if Algorithm 2 is run with any seed node $v \in \mathcal{C}_\sigma[X]^g$, then the PPR estimated cluster $\widehat{C}$ satisfies

$$\Delta(\mathcal{C}_\sigma[X], \widehat{C}) \leq c\kappa(\mathcal{C}) \frac{\Lambda_\sigma}{\lambda_\sigma}, \tag{98}$$

with probability at least $1 - \frac{b_2}{n}$.

F.1 Proof of Theorem 12.

The proof of Theorem 12 follows from Corollary 3.3 of Zhu et al. [51].

Corollary 3.3 of Zhu. Let $G = (V, E)$ be an undirected, unweighted graph, let $p_v := p(v, \alpha; G)$ be a PPR vector with seed node $v \in V$ and teleportation parameter $\alpha \in (0, 1)$.

Lemma 32 (Corollary 3.3 of Zhu et al. [51].)**.** *Let $A \subseteq G$, and suppose $\alpha \leq \frac{1}{9\tau_\infty(G[A])}$. Then, there exists a set $A^g \subseteq A$ with $\mathrm{vol}(A^g; G) \geq \frac{1}{2}\mathrm{vol}(A; G)$ such that the following statement holds: for any $v \in A$; the PPR vector p_v satisfies*

$$p_v(V \setminus A) \leq 2\frac{\Phi(A; G)}{\alpha} \tag{99}$$

and additionally there exists a residual vector $p_\ell \in [0, 1]^V$ with $\|p_\ell\|_1 \leq \frac{2\Phi(A;G)}{\alpha}$ such that

$$\textit{for all } u \in A, \quad p_v(u) \geq \frac{4}{5}\frac{\deg(u; G[A])}{\mathrm{vol}(A; G)} - p_\ell(u). \tag{100}$$

Upper bound on $|S_\beta \vartriangle A|$. For given π_0 and $\beta \in (0, 1)$, consider the sweep cut

$$S_{\beta,v} := \{u \in V : p_v(u) \geq \beta\pi_0\}.$$

Suppose the conditions $\alpha \leq \frac{1}{9\tau_\infty(G[A])}$ and $v \in A^g$ are met. Then by (99),

$$|S_{\beta,v} \setminus A| \leq \frac{p_v(V \setminus A)}{\min_{u \in S_\beta} p_v(u)} \leq \frac{2\Phi(A; G)}{\alpha\beta\pi_0}. \tag{101}$$

To upper bound $|A \setminus S_\beta|$, note that for every $u \in A \setminus S_\beta$, $p_v(u) \leq \beta\pi_0$. Therefore by (100),

$$p_\ell(u) \geq \frac{4\deg(u; G[A])}{5\mathrm{vol}(A; G)} - \beta\pi_0, \text{for all } u \in A \setminus S_\beta$$

Since additionally $\|p_\ell\|_1 \leq 2\Phi(S; G)/\alpha$, we have

$$\begin{aligned}
\sum_{u \in A \setminus S_{\beta,v}} \left(\frac{4\deg(u; G[A])}{5\mathrm{vol}(A; G)} - \beta\pi_0\right) &\leq \frac{2\Phi(S; G)}{\alpha} \Longrightarrow \\
|A \setminus S_{\beta,v}| \underbrace{\left(\frac{4\min_{u \in A}\deg(u; G[A])}{|A|\, 5\max_{u \in A}\deg(u; G)} - \beta\pi_0\right)}_{:=T_1(A;G)} &\leq \frac{2\Phi(S; G)}{\alpha} \Longrightarrow \\
|A \setminus S_{\beta,v}| &\leq \frac{2\Phi(S; G)}{\alpha T_1(A; G)}.
\end{aligned} \tag{102}$$

Bounds on graph functionals for $G_{n,r}$. To apply (101) and (102) when $G = G_{n,r}$ and $A = \mathcal{C}_\sigma[X]$, we need to verify the condition $\alpha \leq \frac{1}{9\tau_\infty(\widetilde{G}_{n,r})}$, and additionally upper bound $\Phi_{n,r}(\mathcal{C}_\sigma[X])$ and $T_1(G_{n,r}; \mathcal{C}_\sigma[X])$. We now state the necessary upper bounds, and the probability with which they hold:

- Since $\alpha < \frac{1}{9\tau_u(\mathcal{C}_\sigma)}$, by Theorem 4

  $$\alpha \leq \frac{1}{9\tau_\infty(\widetilde{G}_{n,r})},$$

  with probability at least $1 - \frac{b_2}{n} - 2n\exp\{-b_3 n\} - 2\exp\{-b_4 n\}$.

- By Theorem 3,

$$\Phi_{n,r}(\mathcal{C}_\sigma[X]) \leq \Phi_u(\mathcal{C}_\sigma)$$

with probability at least $1 - 3\exp\{-nb\}$.

- By Lemma 36,

$$\frac{\min_{u \in \mathcal{C}_\sigma[X]} \widetilde{\deg}_{n,r}(u)}{\max_{u \in \mathcal{C}_\sigma[X]} \deg_{n,r}(u)} = \frac{\widetilde{d}_{\min}}{d_{\max}} \geq \frac{(1-\delta)6\lambda_\sigma}{(1+\delta)25\Lambda_\sigma},$$

and

$$|\mathcal{C}_\sigma[X]| \leq (1+\delta)n\mathbb{P}(\mathcal{C}_\sigma[X])$$

all with probability at least $1 - 2n\exp\left\{-\frac{2\delta^2\lambda_\sigma\nu_d r^d n}{75(1+\frac{\delta}{3})}\right\} - 2\exp\left\{-2\delta^2\mathbb{P}(\mathcal{C}_\sigma)^2\right\}$. These bounds along with the initialization conditions (97) imply

$$\begin{aligned} T_1(G_{n,r}, \mathcal{C}_\sigma[X]) &\geq \frac{\lambda_\sigma}{\Lambda_\sigma\,|\mathcal{C}_\sigma[X]|}\left(\frac{(1-\delta)6}{(1+\delta)25} - (1+\delta)\beta\right) \\ &\geq \frac{\lambda_\sigma}{\Lambda_\sigma\,|\mathcal{C}_\sigma[X]|}\left(\frac{(1-\delta)6}{(1+\delta)25} - \frac{(1+\delta)}{50}\right). \end{aligned} \tag{103}$$

We assume these high probability bounds are satisfied with $\delta = \frac{1}{2}$ and move on to upper bounding $\Delta_{\mathrm{mc}}(\widehat{C}_{\mathrm{un}}, \mathcal{C}_\sigma[X])$.

Upper bound on misclassification rate. Since $\alpha \leq \frac{1}{9\tau_\infty(\widetilde{G}_{n,r})}$, we may apply Lemma 32 to the graph $G_{n,r}$ and subset $\mathcal{C}_\sigma[X]$. Combined with the inequalities we've already derived, this implies the following: there exists a subset $\mathcal{C}_\sigma[X]^g \subset \mathcal{C}_\sigma[X]$ with $\mathrm{vol}_{n,r}(\mathcal{C}_\sigma[X]) \leq \frac{1}{2}\mathrm{vol}_{n,r}(\mathcal{C}_\sigma[X])$ such that the following bounds hold,

$$|S_{\beta,v} \setminus \mathcal{C}_\sigma[X]| \leq \frac{2\Phi_{n,r}(\mathcal{C}_\sigma[X])}{\alpha\beta\pi_0} \leq \frac{2000\kappa(\mathcal{C})}{\pi_0} \leq \frac{4000\kappa(\mathcal{C})\Lambda_\sigma}{\lambda_\sigma}\,|\mathcal{C}_\sigma[X]|\,,$$

and

$$|\mathcal{C}_\sigma[X] \setminus S_{\beta,v}| \leq \frac{2\Phi_{n,r}(\mathcal{C}_\sigma[X])}{\alpha T_1(\mathcal{C}_\sigma[X]; G_{n,r})} \leq \frac{1000\kappa(\mathcal{C})\Lambda_\sigma}{\lambda_\sigma}\,|\mathcal{C}_\sigma[X]|\,.$$

This completes the proof of Theorem 12.

G Probabilistic bounds

In our theory, we frequently appeal to concentration of graph functionals such as degree, volume, and cut size around their means. In this section, we establish such concentration inequalities. We begin with Lemma 33, Hoeffding's inequality, which we will use to bound the empirical probability of a fixed set.

Lemma 33 (Hoeffding's Inequality.)**.** *Let $\mathcal{A}$ be a subset of $\mathbb{R}^d$. Then,*

$$(1-\delta)\mathbb{P}(\mathcal{A}) \leq \mathbb{P}_n(\mathcal{A}) \leq (1+\delta)\mathbb{P}(\mathcal{A})$$

with probability at least $1 - \exp\left\{-2\delta^2(\mathbb{P}(\mathcal{A}))^2\right\}$.

To bound minimum and maximum degrees, we will use Lemma 34, which is a combination of Bernstein's inequality and a union bound.

Lemma 34 (Bernstein's inequality + Union bound.)**.** *For $M \geq 1$, let $\mathcal{A}_1, \ldots, \mathcal{A}_M$ be subsets of $\mathbb{R}^d$. Denote the minimum probability mass among these sets as $p_{\min} := \min_{m=1,\ldots,M} \mathbb{P}(\mathcal{A}_m)$, and likewise let $p_{\max} := \max_{m=1,\ldots,M} \mathbb{P}(\mathcal{A}_m)$. Then*

$$(1-\delta)p_{\min} \leq \min_{m=1,\ldots,M} \mathbb{P}_n(\mathcal{A}_m) \leq \max_{m=1,\ldots,M} \mathbb{P}_n(\mathcal{A}_m) \leq (1+\delta)p_{\max}$$

with probability at least $1 - 2M\exp\left\{-\frac{\frac{1}{3}\delta^2 p_{\min} n}{1+\frac{\delta}{3}}\right\}$.

The above Lemma will allow us to upper and lower bound the minimum and maximum degrees within the neighborhood graphs $G_{n,r}$ and $\widetilde{G}_{n,r}$. To bound cut size and volume functionals, we will use Hoeffding's inequality for U-statistics. Recall that U_n is an order-2 U-statistic with kernel $h : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ if

$$U_n = \frac{1}{2\binom{n}{2}} \sum_{i=1}^{n} \sum_{j \neq i} h(x_i, x_j).$$

Lemma 35 (Hoeffding's Inequality for U-statistics.)**.** *Assume $\|h\|_\infty \leq 1$. Then,*

$$(1-\delta)\mathbb{E}(U_m) \leq U_n \leq (1+\delta)\mathbb{E}(U_m).$$

with probability at least $1 - 2\exp\left(-\delta^2(\mathbb{E}(U_m))^2 n\right)$.

We collect the bounds on graph functionals we need in Lemma 36. Let $S \subset X$ and $\mathcal{S} \subseteq \mathbb{R}^d$. As a reminder, for notational ease we write $\widetilde{G}_{n,r} = G_{n,r}[\mathcal{C}_\sigma[X]]$ and

$$\widetilde{d}_{\min} = \min_{u \in \mathcal{C}_\sigma[X]} \widetilde{\deg}_{n,r}(u), \quad d'_{\min} = \min_{u' \in \mathcal{C}'[X]} \deg_{n,r}(u')$$
$$d_{\max} = \max_{u \in \mathcal{C}_\sigma[X]} \deg_{n,r}(u), \quad \widetilde{\mathrm{vol}}_{n,r}(S) = \mathrm{vol}(S; \widetilde{G}_{n,r}).$$

Additionally, let $\widetilde{\mathrm{vol}}_{\mathbb{P},r}(\mathcal{S}) := \frac{1}{2\binom{n}{2}}\mathbb{E}[\widetilde{\mathrm{vol}}_{n,r}(\mathcal{S}[X])]$.

Lemma 36.

The following statements hold for any $\delta \in (0,1)$:
With probability at least $1 - 2\exp\{-n\delta^2(\widetilde{\mathrm{vol}}_{\mathbb{P},r}(\mathcal{S}))^2\}$,

$$(1-\delta)\widetilde{\mathrm{vol}}_{\mathbb{P},r}(\mathcal{S}) \leq \frac{\widetilde{\mathrm{vol}}_{n,r}(\mathcal{S}[X])}{2\binom{n}{2}} \leq (1+\delta)\widetilde{\mathrm{vol}}_{\mathbb{P},r}(\mathcal{S}). \tag{104}$$

With probability at least $1 - 2\exp\left\{-n\delta^2 \mathrm{vol}_{\mathbb{P},r}(\mathcal{S})^2\right\}$,

$$(1-\delta)\mathrm{vol}_{\mathbb{P},r}(\mathcal{S}) \leq \frac{\mathrm{vol}_{n,r}(\mathcal{S}[X])}{2\binom{n}{2}} \leq 2(1+\delta)\mathrm{vol}_{\mathbb{P},r}(\mathcal{S}). \tag{105}$$

With probability at least $1 - 2\exp\left\{-n\delta^2 \mathrm{cut}_{\mathbb{P},r}(\mathcal{S})^2\right\}$,

$$(1-\delta)\mathrm{cut}_{\mathbb{P},r}(\mathcal{S}) \leq \frac{\mathrm{cut}_{n,r}(\mathcal{S}[X])}{2\binom{n}{2}} \leq (1+\delta)\mathrm{vol}_{\mathbb{P},r}(\mathcal{S}). \tag{106}$$

With probability at least $1 - 2\exp\{-n\delta^2\mathbb{P}(\mathcal{C}_\sigma)^2\}$,

$$(1-\delta)\mathbb{P}(\mathcal{C}_\sigma) \le |\mathcal{C}_\sigma[X]| \le (1+\delta)\mathbb{P}(\mathcal{C}_\sigma). \tag{107}$$

The following statements hold for any $0 < r \le \frac{\sigma}{2\sqrt{d}}$*:*
Assuming $\mathcal{C}_\sigma$ *satisfies (A1), with probability at least* $1 - 2n\exp\left\{-\frac{2\delta^2\lambda_\sigma\nu_d r^d n}{75(1+\frac{\delta}{3})}\right\}$,

$$\frac{6}{25}(1-\delta)\lambda_\sigma\nu_d r^d n \le \widetilde{d}_{\min} \le d_{\max} \le (1+\delta)\Lambda_\sigma\nu_d r^d n. \tag{108}$$

Assuming $\mathcal{C}'$ *satisfies (A5), with probability at least* $1 - n\exp\left\{-\frac{\delta^2\lambda_\sigma\nu_d r^d n}{3(1+\frac{\delta}{3})}\right\}$,

$$d'_{\min} \ge (1-\delta)\lambda_\sigma\nu_d r^d n. \tag{109}$$

Proof. (of Lemma 36)

Proof of (108), (109): Under (A1), we have that for every $x_j \in \mathcal{C}_\sigma[X]$

$$\mathbb{E}(\widetilde{\deg}_{n,r}(x_j)) \le \Lambda_\sigma\nu_d r^d;$$

furthermore, recalling the weighted local conductance $\ell_{\mathbb{P},r}(x) = \mathbb{P}(B(x,r)\cap\mathcal{C}_\sigma)$, we have,

$$\mathbb{E}(\widetilde{\deg}_{n,r}(x_j)) = \ell_{\mathbb{P},r}(x_j) \ge \frac{6}{25}\lambda_\sigma\nu_d r^d$$

where the last inequality follows from Lemma 3. Under (A5) we have for every $x_i \in \mathcal{C}'[X]$

$$\mathbb{E}(\widetilde{\deg}_{n,r}(x_i)) \ge \lambda_\sigma\nu_d r^d.$$

Then (108) and (109) each follow from Lemma 34.

Proof of (104), (105), (106): We have that $\frac{1}{2\binom{n}{2}}\widetilde{\mathrm{vol}}_{n,r}(\mathcal{C}_\sigma[X])$, $\frac{1}{2\binom{n}{2}}\mathrm{vol}_{n,r}(\mathcal{S}[X])$, $\frac{1}{2\binom{n}{2}}\mathrm{cut}_{n,r}(\mathcal{S}[X])$ are each order-2 U statistics, with respective kernels

$$\begin{aligned}
h_{\widetilde{\mathrm{vol}}}(x_i,x_j) &= \mathbb{1}(x_i\in\mathcal{C}_\sigma)\mathbb{1}(x_j\in\mathcal{C}_\sigma)\mathbb{1}(\|x_i-x_j\|\le r),\\
h_{\mathrm{cut}}(x_i,x_j) &= \mathbb{1}(x_i\in\mathcal{C}_\sigma)\mathbb{1}(x_j\notin\mathcal{C}_\sigma)\mathbb{1}(\|x_i-x_j\|\le r)\\
h_{\mathrm{vol}}(x_i,x_j) &= \mathbb{1}(x_i\in\mathcal{C}_\sigma)\mathbb{1}(\|x_i-x_j\|\le r).
\end{aligned}$$

We may therefore apply Lemma 35 in each case to obtain the stated bounds.

Proof of (107): Apply Lemma 33 to $\mathcal{C}_\sigma$. □

H Experiments

In Section H, we detail the experimental settings of Section 6 in the main text, and include an additional figure.

H.1 Experimental settings for Figure 2

We sample points according to the density function q, where for $x \in \mathbb{R}^d$

$$q(x) := \begin{cases} \lambda, & x \in [0,\sigma] \times \rho^{d-1} =: \mathcal{C}, \\ \lambda - \mathrm{dist}(x,\mathcal{C})\eta, & x \in \mathcal{C}_\sigma \setminus \mathcal{C}, \\ (\lambda - \sigma\eta) - \mathrm{dist}(x,\mathcal{C}_\sigma)^\gamma, & x \in (\mathcal{C}_\sigma + \sigma B) \setminus \mathcal{C}_\sigma, \\ 0, & \text{otherwise}, \end{cases} \tag{110}$$

where $\lambda = \frac{150}{81}\sigma^\gamma$ and $\eta = \frac{15}{81}\sigma^{\gamma-1}$.

In the top-left and top-middle, we show draws of $n = 20000$ samples from two different density functions. In the top-left panel, $\sigma = \rho = 3.2$, while in the top-middle panel $\sigma = .1$ and $\rho = 3.2$. (For both, $d = 2$).

The remaining four panels (top-right and the bottom row) in Figure 2 show the change in normalized cut and mixing time, respectively, as the parameters σ (top-right and bottom-left) and ρ (bottom-middle and bottom-right) are varied. In the top-right and bottom-left panels $\sigma = .1 \cdot \sqrt{2}^j, j = 1, \ldots, 10$, and ρ is fixed at 3.2. In the bottom-middle and bottom-right panels, $\rho = .1 \cdot \sqrt{2}^j, j = 1, \ldots, 10$ and σ is fixed at .1. For each panel, the solid lines show, up to constants, the theoretical upper bound, given by Theorem 3 for the top-right and bottom-left panels and Theorem 4 for the bottom-middle and bottom-right panels. The dashed lines show the computed empirical value, averaged over m trials ($m = 100$ for the normalized cut, dashed lines in the top-right and bottom-left panels, and $m = 20$ for the mixing time, dashed lines in the bottom-middle and bottom-right panels). For each trial across all parameters, r, the neighborhood graph radius, is set throughout to be as small as possible such that the resulting graph is connected, for computational efficiency. Green lines correspond to dimension $d = 2$, whereas purple/pink lines correspond to $d = 3$.

H.2 Experimental settings for Figure 3

To form each of the three rows in Figure 3, 800 points are independently sampled following a 'two moons plus Gaussian noise model'. Formally, the (respective) generative models for the data are

$$Z \sim \mathrm{Bern}(1/2), \theta \sim \mathrm{Unif}(0,\pi) \tag{111}$$

$$X(Z,\theta) = \begin{cases} \mu_1 + (r\cos(\theta), r\sin(\theta)) + \sigma\epsilon, & \text{if } Z = 1 \\ \mu_2 + (r\cos(\theta), -r\sin(\theta)) + \sigma\epsilon, & \text{if } Z = 0 \end{cases} \tag{112}$$

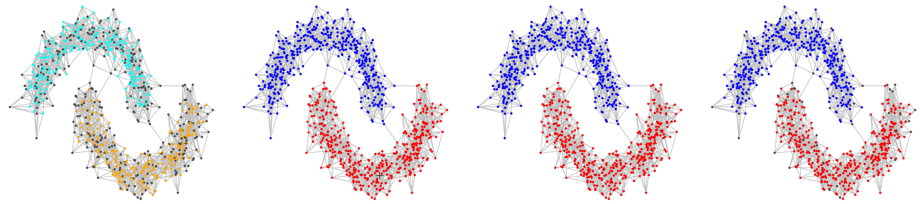
where

$$\mu_1 = (-.5, 0),\ \mu_2 = (0,0),\ \epsilon \sim N(0, I_2) \tag{row 1}$$

$$\mu_1 = (-.5, -.07),\ \mu_2 = (0,.07),\ \epsilon \sim N(0, I_2) \tag{row 2}$$

$$\mu_1 = (-.5, -.125),\ \mu_2 = (0,.125),\ \epsilon \sim N(0, I_2) \tag{row 3}$$

for I_d the $d \times d$ identity matrix. The first column consists of the empirical density clusters $C[X]$ and $C'[X]$ for a particular threshold λ of the density function; the second column shows the PPR plus minimum normalized sweep cut cluster, with hyperparameter α and all sweep cuts considered; the third column shows the global minimum normalized cut, computed according to the algorithm of [43]; and the last column shows a cut of the density cluster tree estimator of [9].



(a) (b) (c) (d)

Figure 4:

H.3 Performance of PPR with high-dimensional noise.

Figure 4 is similar to Figure 3 of the main text, with parameters

$$\mu_1 = (-.5, -.025),\ \mu_2 = (0, .025),\ \epsilon \sim N(0, I_{10}).$$

The gray dots in (a) (as in the left-hand column of Figure 3 in the main text) represent observations in low-density regions. While the PPR sweep cut (b) has relatively high symmetric set difference with the chosen density cut, it still recovers $C[X]$ in the sense of Definition 2.