
Local Spectral Clustering of Density Upper Level Sets

Alden Green

Sivaraman Balakrishnan

Ryan J. Tibshirani

Abstract

We analyze Personalized PageRank (PPR), a local spectral method for clustering, which extract clusters using locally-biased random walks around a user-specified seed node. In contrast to previous work, we adopt a traditional 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. We also provide empirical support for our theory.

1 Introduction

In this paper, we study the problem of clustering: splitting a given data set into groups that satisfy some notion of within-group similarity and between-group difference. We focus on spectral clustering, a family of powerful nonparametric clustering algorithms. Generally speaking, a spectral algorithm first constructs a geometric graph G , where vertices correspond to samples, and edges correspond to proximities between samples. It then learns a feature embedding based on the Laplacian 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 [Leskovec et al., 2010, Mahoney et al., 2012]. 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 Personalized PageRank (PPR), first introduced by Haveliwala [2003], then further developed by [Spielman and Teng, 2011, 2014, Andersen et al., 2006, Mahoney et al., 2012, Zhu et al., 2013], among others.

Local spectral clustering techniques have been practically very successful [Leskovec et al., 2010, Andersen et al., 2012, Gleich and Seshadhri, 2012, Mahoney et al., 2012, Wu et al., 2012], leading many authors to develop supporting theory [Spielman and Teng, 2013, Andersen and Peres, 2009, Ghahramani and Trevisan, 2012, Zhu et al., 2013] that gives worst-case guarantees on traditional graph-theoretic notions of cluster quality (such as conductance). In this paper, we adopt a more traditional statistical viewpoint, and examine what the output of local clustering on a data set reveals about the underlying density f . 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 $\lambda > 0$ (a central object of interest in the statistical clustering literature, dating back to Hartigan [1981]).

PPR on a neighborhood graph. We now describe the clustering algorithm that will be our focus for the rest of the paper. Let $X = \{x_1, \dots, x_n\}$ 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 ℓ_2 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 $D_{uu} = \sum_{v \in V} A_{uv}$, and by I the $n \times n$ identity matrix.

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

$$p = \alpha e_v + (1 - \alpha)pW, \quad (1)$$

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 the v th position and 0 elsewhere). For a level $\beta > 0$ and a target volume $\text{vol}_0 > 0$, we define a β -sweep cut of $p = (p_u)_{u \in V}$ as

$$S_\beta := \left\{ u \in V : \frac{p_u}{D_{uu}} > \frac{\beta}{\text{vol}_0} \right\}. \quad (2)$$

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 $\text{cut}(S; G_{n,r}) := \sum_{u \in S, v \in S^c} A_{uv}$, and $\text{vol}(S; G_{n,r}) := \sum_{u \in S} D_{uu}$. We define the *normalized cut* of S as

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

Having computed sweep cuts S_β over a range $\beta \in (\frac{1}{40}, \frac{1}{11})^1$, we output the cluster estimate $\hat{C} = S_{\beta^*}$ that has minimum normalized cut. For concreteness, this is summarized in Algorithm 1.

Algorithm 1 PPR on a neighborhood graph

Input: data $X = \{x_1, \dots, x_n\}$, radius $r > 0$, teleportation parameter $\alpha \in [0, 1]$, seed $v \in X$, target stationary volume $\text{vol}_0 > 0$.

Output: cluster $\hat{C} \subseteq V$.

- 1: Form the neighborhood graph $G_{n,r}$.
- 2: Compute the PPR vector $p = p(v, \alpha; G_{n,r})$ as in (1).
- 3: For $\beta \in (\frac{1}{40}, \frac{1}{11})$ compute sweep cuts S_β as in (2).
- 4: Return as a cluster $\hat{C} = S_{\beta^*}$, where

$$\beta^* = \arg \min_{\beta \in (\frac{1}{40}, \frac{1}{11})} \Phi(S_\beta; G_{n,r}).$$

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 clusters is a commonly used metric to quantify cluster estimation error [Korostelev and Tsybakov, 1993, Polonik, 1995, Rigollet and Vert, 2009].

Definition 1 (Symmetric set difference). *For an estimator $\hat{C} \subseteq X$ and set $\mathcal{S} \subseteq \mathbb{R}^d$, we define*

$$\Delta(\hat{C}, \mathcal{S}) := |\hat{C} \setminus \mathcal{S}[X] \cup \mathcal{S}[X] \setminus \hat{C}|, \quad (4)$$

the cardinality of the symmetric set difference between \hat{C} and $\mathcal{S} \cap X = \mathcal{S}[X]$.

However, the symmetric set difference does not measure whether \hat{C} can 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 [1981], and defined asymptotically.

Definition 2 (Consistent density cluster estimation). *For an estimator $\hat{C} \subseteq X$ and cluster $\mathcal{C} \in \mathbb{C}_f(\lambda)$, we say \hat{C} is a consistent estimator of \mathcal{C} if for all $\mathcal{C}' \in \mathbb{C}_f(\lambda)$ with $\mathcal{C} \neq \mathcal{C}'$, the following holds as $n \rightarrow \infty$:*

$$\mathcal{C}[X] \subseteq \hat{C} \quad \text{and} \quad \hat{C} \cap \mathcal{C}'[X] = \emptyset, \quad (5)$$

with probability tending to 1.

¹The choice of a specific range such as $(\frac{1}{40}, \frac{1}{11})$ is standard in the analysis of PPR algorithms, see, e.g., [Zhu et al., 2013].

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

1. In Section 2, we introduce a set of natural geometric conditions on the density cluster \mathcal{C} and show that when Algorithm 1 is properly initialized, the size of the symmetric set difference of \mathcal{C} and a thickened version of the density cluster \mathcal{C}_σ can be bounded in a meaningful way.
2. We further show in Section 2 that if the density cluster \mathcal{C} is particularly well-conditioned, Algorithm 1 will consistently estimate a density cluster in the sense of (5).
3. In Section 3, we detail some of the analysis required to prove our main results, and expose the parts various geometric quantities play in the difficulty of the clustering problem.
4. In Section 4, we empirically investigate the tightness of our analysis, and provide examples showing how violations of our geometric conditions impact density cluster recovery by PPR.

Our main takeaway: PPR, run on a neighborhood graph, recovers geometrically compact high-density clusters.

Related work. In addition to the background on local spectral clustering given previously, a few related lines of work are worth highlighting. [Shi et al., 2009, Schiebinger et al., 2015] examine the consistency of spectral algorithms in recovering the latent labels in certain 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.

More broadly, density clustering and level set estimation is a well-studied problem. Polonik [1995], Rigollet and Vert [2009] study density clustering under the symmetric set difference metric, Tsybakov [1997], Singh et al. [2009] describe minimax optimal level set estimators under Hausdorff loss and Hartigan [1981], Chaudhuri and Dasgupta [2010] consider consistent estimation of the cluster tree, to note but a few works. 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 Estimation of well-conditioned density clusters

We formalize some geometric conditions, and use these to define a condition number $\kappa(\mathcal{C})$, which measures the difficulty PPR will have in estimating \mathcal{C} . (Our theoretical guarantees for PPR will be framed in terms of $\kappa(\mathcal{C})$.)

Geometric conditions on density clusters. At a high level, for PPR to be successful, the underlying density cluster must be geometrically well-conditioned. At a minimum, we want to avoid sets that contain arbitrarily thin bridges or spikes. Hence, as in Chaudhuri and Dasgupta [2010], we consider a thickened version of $\mathcal{C} \in \mathbb{C}_f(\lambda)$ defined as $\mathcal{C}_\sigma := \{x \in \mathbb{R}^d : \text{dist}(x, \mathcal{C}) \leq \sigma\}$, which we call the σ -expansion of \mathcal{C} . Here $\text{dist}(x, \mathcal{C}) := \inf_{y \in \mathcal{C}} \|y - x\|$. We now list our conditions on \mathcal{C}_σ .

- (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) *Cluster separation:* For all clusters $\mathcal{C}' \in \mathbb{C}_f(\lambda)$ with $\mathcal{C}' \neq \mathcal{C}$, $\text{dist}(\mathcal{C}_\sigma, \mathcal{C}'_\sigma) > \sigma$, where $\text{dist}(\mathcal{C}_\sigma, \mathcal{C}'_\sigma) := \inf_{x \in \mathcal{C}_\sigma} \text{dist}(x, \mathcal{C}'_\sigma)$.
- (A3) *Low noise density:* There exist $\gamma, c_0 > 0$ such that for $x \in \mathbb{R}^d$ with $0 < \text{dist}(x, \mathcal{C}_\sigma) \leq \sigma$, $\inf_{x' \in \mathcal{C}_\sigma} f(x') - f(x) \geq c_0 \text{dist}(x, \mathcal{C}_\sigma)^\gamma$.
- (A4) *Lipschitz embedding:* There exists $g : \mathbb{R}^d \rightarrow \mathbb{R}^d$ with the following properties: i) we have $\mathcal{C}_\sigma = g(\mathcal{K})$, for a convex set $\mathcal{K} \subseteq \mathbb{R}^d$ with $\text{diam}(\mathcal{K}) = \sup_{x, y \in \mathcal{K}} \|x - y\| =: \rho < \infty$; ii)

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

$$\frac{1}{L} \|x - y\| \leq \|g(x) - g(y)\| \leq L \|x - y\| \text{ for all } x, y \in \mathcal{K}.$$

Succintly, \mathcal{C}_σ is the image of a convex set with finite diameter under a measure preserving, bi-Lipschitz transformation.

(A5) *Bounded volume*: Let the neighborhood graph radius $0 < r \leq \sigma/2d$ be such that

$$2 \int_{\mathcal{C}_\sigma} \mathbb{P}(B(x, r)) f(x) dx \leq \int_{\mathbb{R}^d} \mathbb{P}(B(x, r)) f(x) dx,$$

where $B(x, r)$ is the closed ball of radius r at x .

To motivate these conditions, Zhu et al. [2013] show for arbitrary graph $G = (V, E)$ and subset of vertices $S \subseteq V$, the PPR estimate \hat{C} of subset S satisfies, for a constant $c > 0$,

$$\text{vol}(\hat{C} \setminus S; G) + \text{vol}(S \setminus \hat{C}; G) \leq c(\Phi(S, G) \cdot \tau_\infty(G[S])) \text{vol}(S; G), \quad (6)$$

where $\Phi(S; G)$ is the normalized cut of S (as defined in (3)), and $\tau_\infty(G[S])$ is called the *mixing time* of a random walk over the induced subgraph $G[S]$ (to be defined precisely later, in (16)). The left-hand side in (6) resembles a (degree-weighted) form of the symmetric set difference metric in (4). As we will show in Section 3, the conditions (A1)–(A5) allow us to upper bound the normalized cut $\Phi(\mathcal{C}_\sigma[X]; G_{n,r})$, and the mixing time $\tau_\infty(G_{n,r}[\mathcal{C}_\sigma[X]])$.²

Condition number. Motivated by (6), we will define $\kappa(\mathcal{C})$ to be an upper bound on the product $\Phi(\mathcal{C}_\sigma[X]; G_{n,r}) \cdot \tau_\infty(G_{n,r}[\mathcal{C}_\sigma[X]])$. The smaller $\kappa(\mathcal{C})$ is, the more success PPR will have in recovering \mathcal{C} . Let $\theta := (r, \sigma, \lambda, \lambda_\sigma, \Lambda_\sigma, \gamma, \rho, L)$ contain the geometric parameters from (A1)–(A5).

Definition 3 (Well-conditioned density clusters). *For $\lambda > 0$ and $\mathcal{C} \in \mathbb{C}_f(\lambda)$, let \mathcal{C} satisfy (A1)–(A5) for some θ . Then, for universal constants $c_1, c_2, c_3 > 0$ to be specified later, we set*

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

and letting $\kappa(\mathcal{C}) := \Phi_u(\theta) \cdot \tau_u(\theta)$, we call \mathcal{C} a κ -well-conditioned density cluster.

We note that $\Phi_u(\theta)$ and $\tau_u(\theta)$ 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 3.

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} (with respect to some θ), we require

$$0 < r \leq \frac{\sigma}{2d}, \quad \alpha \in [1/10, 1/9] \cdot \frac{1}{\tau_u(\theta)}, \quad (8)$$

$$v \in \mathcal{C}_\sigma[X]^g, \quad \text{vol}_0 \in [3/4, 5/4] \cdot n(n-1) \int_{\mathcal{C}_\sigma} \mathbb{P}(B(x, r)) f(x) dx,$$

where $\mathcal{C}_\sigma[X]^g \subseteq \mathcal{C}_\sigma[X]$ will be some large (“good”) subset of $\mathcal{C}_\sigma[X]$. In particular, abbreviating $\text{vol}_{n,r}(S) := \text{vol}(S; G_{n,r})$ for $S \subseteq X$, we will have $\text{vol}_{n,r}(\mathcal{C}_\sigma[X]^g) \geq \text{vol}_{n,r}(\mathcal{C}_\sigma[X])/2$.

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

In practice it is clearly not feasible to set hyperparameters based on the underlying (unknown) density f . Typically, one tunes PPR over a range of hyperparameters and optimizes for some criterion such as minimum normalized cut; it is not obvious how this scheme would affect the performance of PPR in the density clustering context.

²Informally, assumptions (A2), (A3) yield an upper bound on $\text{cut}(\mathcal{C}_\sigma[X]; G_{n,r})$, and (A1) yields a lower bound on $\text{vol}(\mathcal{C}_\sigma[X]; G_{n,r})$; together with (A5), this gives an upper bound on the normalized cut. On the other hand, (A1), (A4) preclude bottlenecks in the induced subgraph $G_{n,r}[\mathcal{C}_\sigma[X]]$, and combined with the upper bound on diameter in (A4), this leads to an upper bound on the mixing time over this subgraph.

Main theorems. The results of Section 3, combined with (6), give an upper bound on the volume of $\widehat{C} \setminus \mathcal{C}_\sigma[X]$ and $\mathcal{C}_\sigma[X] \setminus \widehat{C}$,

$$\text{vol}_{n,r}(\widehat{C} \setminus \mathcal{C}_\sigma[X]) + \text{vol}_{n,r}(\mathcal{C}_\sigma[X] \setminus \widehat{C}) \leq c\kappa(\mathcal{C})\text{vol}_{n,r}(\mathcal{C}_\sigma[X]). \quad (9)$$

To translate (9) into meaningful bounds on the symmetric set difference metric $\Delta(\mathcal{C}_\sigma[X], \widehat{C})$, we want to preclude vertices $x \in X$ from having arbitrarily small degree, and so we make some regularity assumptions on $\mathcal{X} := \text{supp}(f)$. Let ν denote the Lebesgue measure on \mathbb{R}^d , and $\nu_d := \nu(B)$ be the measure of the unit ball $B = B(0, 1)$.

(A6) *Regular support:* There exists some constant $\lambda_{\min} > 0$ such that $\lambda_{\min} < f(x)$ for all $x \in \mathcal{X}$. Additionally, there exists some $c > 0$ such that for each $x \in \partial\mathcal{X}$, $\nu(B(x, r) \cap \mathcal{X}) \geq c\nu_d r^d$.

Note that the latter condition in (A6) will hold if the boundary $\partial\mathcal{X}$ is sufficiently regular. Now we present our main bound on the symmetric set difference metric.

Theorem 1. Fix $\lambda > 0$, let $\mathcal{C} \in \mathbb{C}_f(\lambda)$ be a κ -well-conditioned density cluster (with respect to some θ), and additionally assume f satisfies (A6). If Algorithm 1 is well-initialized, there exists a universal constant $c > 0$ such that with probability tending to 1 as $n \rightarrow \infty$,

$$\Delta(\mathcal{C}_\sigma[X], \widehat{C}) \leq c\kappa(\mathcal{C}) \frac{\Lambda_\sigma}{\lambda_{\min}}. \quad (10)$$

The proof of Theorem 1, along with all other proofs in this paper, is deferred to the appendix. Note that this result says the symmetric set difference metric $\Delta(\mathcal{C}_\sigma[X], \widehat{C})$ is proportional to the difficulty of the clustering problem, as measured by the condition number $\kappa(\mathcal{C})$.

Neither (9) nor (10) imply consistent density cluster estimation in the sense of (5). This notion of consistency requires a uniform bound over p : for all $\mathcal{C}' \in \mathbb{C}_f(\lambda)$, $\mathcal{C}' \neq \mathcal{C}$, and each $u \in \mathcal{C}$, $w \in \mathcal{C}'$,

$$\frac{p_w}{D_{ww}} \leq \frac{1}{40\text{vol}_0} < \frac{1}{11\text{vol}_0} \leq \frac{p_u}{D_{uu}}, \quad (11)$$

so that any sweep cut S_β for $\beta\text{vol}_0 \in [1/40, 1/11]$ (i.e., any sweep cut considered by Algorithm 1) will fulfill both conditions laid out in (5). In Theorem 2, we show that a sufficiently small upper bound on $\kappa(\mathcal{C})$ ensures such a gap exists with probability 1 as $n \rightarrow \infty$, and hence guarantees \widehat{C} will be a consistent estimator. As was the case before, we wish to preclude arbitrarily low degree vertices, this time for points $x \in \mathcal{C}'[X]$.

(A7) *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 $x \in \mathcal{C}'_\sigma$, $\lambda_\sigma \leq f(x)$.

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

Theorem 2. Fix $\lambda > 0$, let $\mathcal{C} \in \mathbb{C}_f(\lambda)$ be a κ -well-conditioned density cluster (with respect to some θ), and additionally assume f satisfies (A7). If Algorithm 1 is well-initialized, there exists a universal constant $c > 0$ such that if

$$\kappa(\mathcal{C}) \leq c \frac{\lambda_\sigma^2 r^d \nu_d}{\Lambda_\sigma \mathbb{P}(\mathcal{C}_\sigma)}, \quad (12)$$

then the output set $\widehat{C} \subseteq X$ is a consistent estimator for \mathcal{C} , in the sense of Definition 2.

Remark 1. We note that the restriction on $\kappa(\mathcal{C})$ imposed by (12) results in a symmetric set difference $\Delta(\mathcal{C}_\sigma[X], \widehat{C})$ on the order of r^d . In plain terms, we are able to recover a density cluster \mathcal{C} in the sense of (5) 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 of (11).

Remark 2. While taking the radius of the neighborhood graph $r \rightarrow 0$ as $n \rightarrow \infty$ (thereby ensuring $G_{n,r}$ is sparse) is computationally attractive, the presence of a factor of $\log^2(1/r)/r$ in $\kappa(\mathcal{C})$ unfortunately prevents us from making claims about the behavior of PPR in this regime. Although the restriction to a kernel function fixed in n is standard for spectral clustering theory [Schiebinger et al., 2015, von Luxburg et al., 2008], it is an interesting question whether PPR exhibits some degeneracy over r -neighborhood graphs as $r \rightarrow 0$, or if this is merely looseness in our upper bounds.

Approximate PPR vector. In practice, exactly solving (1) may be too computationally expensive. To address this limitation, Andersen et al. [2006] introduced the ϵ -approximate PPR vector (aPPR), which we will denote by $p^{(\epsilon)}$. We refer the curious reader to Andersen et al. [2006] 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 $\mathcal{O}(1/(\epsilon\alpha))$ time, while satisfying the following uniform error bound:

$$\text{for all } u \in V, \quad p(u) - \epsilon D_{uu} \leq p^{(\epsilon)}(u) \leq p(u). \quad (13)$$

Application of (13) within the proofs of Theorems 1 and 2 leads to analogous results which hold for $p^{(\epsilon)}$. We formally state and prove this in the appendix.

3 Analysis

The primary technical contribution of our work is showing that the geometric conditions (A1)–(A5) translate to meaningful bounds on the normalized cut and mixing time of $\mathcal{C}_\sigma[X]$ in $G_{n,r}$. In doing so, we elaborate on how some of the geometric conditions introduced in Section 2 contribute to the difficulty of the clustering problem.

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 $\lambda > 0$, and assume $\mathcal{C} \in \mathbb{C}_f(\lambda)$ satisfies Assumptions (A1)–(A3), (A5) for some $r, \sigma, \lambda_\sigma, c_0, \gamma > 0$ (no bound on maximum density is needed). Then for any $0 < \delta < 1$, $\epsilon > 0$, if

$$n \geq \frac{(2 + \epsilon)^2 \log(3/\delta)}{\epsilon^2} \left(\frac{25}{6\lambda_\sigma^2 \nu(\mathcal{C}_\sigma) \nu_d r^d} \right)^2, \quad (14)$$

then

$$\frac{\Phi_{n,r}(\mathcal{C}_\sigma[X])}{r} \leq c_1 \frac{d}{\sigma} \frac{\lambda}{\lambda_\sigma} \frac{(\lambda_\sigma - c_0 \frac{r^\gamma}{\gamma+1})}{\lambda_\sigma} + \epsilon, \quad (15)$$

with probability at least $1 - \delta$, where $c_1 > 0$ is a universal constant.

Remark 3. Observe that the diameter ρ is absent from Theorem 3, in contrast to the condition number $\kappa(\mathcal{C})$, which worsens (increases) as ρ increases. This reflects established wisdom regarding spectral partitioning algorithms more generally [Guattery and Miller, 1995, Hein and Bühler, 2010], albeit newly applied to the density clustering setting. It suggests that if the diameter ρ is large, PPR may fail to recover $\mathcal{C}_\sigma[X]$ even when \mathcal{C} is sufficiently well-conditioned to ensure $\mathcal{C}_\sigma[X]$ has a small normalized cut in $G_{n,r}$. This intuition will be supported by simulations in Section 4.

Mixing time. For $S \subseteq V$, denote by $G[S] = (S, E_S)$ the subgraph induced by S (where the edges are $E_S = E \cap (S \times S)$). 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 lazy random walk over $G[S]$ originating at $v \in V$. Also write $\pi = (\pi(u))_{u \in S}$ for 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 $\pi(u) = (D_S)_{uu} / \text{vol}(S; G[S])$, where we write D_S for the degree matrix of $G[S]$.) We define the *mixing time* of $G[S]$ as

$$\tau_\infty(G[S]) = \min \left\{ t : \frac{\pi(u) - q_v^{(t)}(u)}{\pi(u)} \leq \frac{1}{4}, \text{ for } u, v \in V \right\}. \quad (16)$$

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

Theorem 4. Fix $\lambda > 0$, and assume that $\mathcal{C} \in \mathbb{C}_f(\lambda)$ satisfies Assumptions (A1) and (A4) for some $\sigma, \lambda_\sigma, \Lambda_\sigma, \rho, L > 0$. Then, for any $0 < r < \sigma/2\sqrt{d}$, with probability 1,

$$\limsup_{n \rightarrow \infty} \tau_\infty(G_{n,r}[\mathcal{C}_\sigma[X]]) \leq c_2 \frac{\Lambda_\sigma^4 d^3 \rho^2 L^2}{\lambda_\sigma^4 r^2} \log^2 \left(\frac{1}{r} \right) + c_3, \quad (17)$$

for $c_2, c_3 > 0$ universal constants.

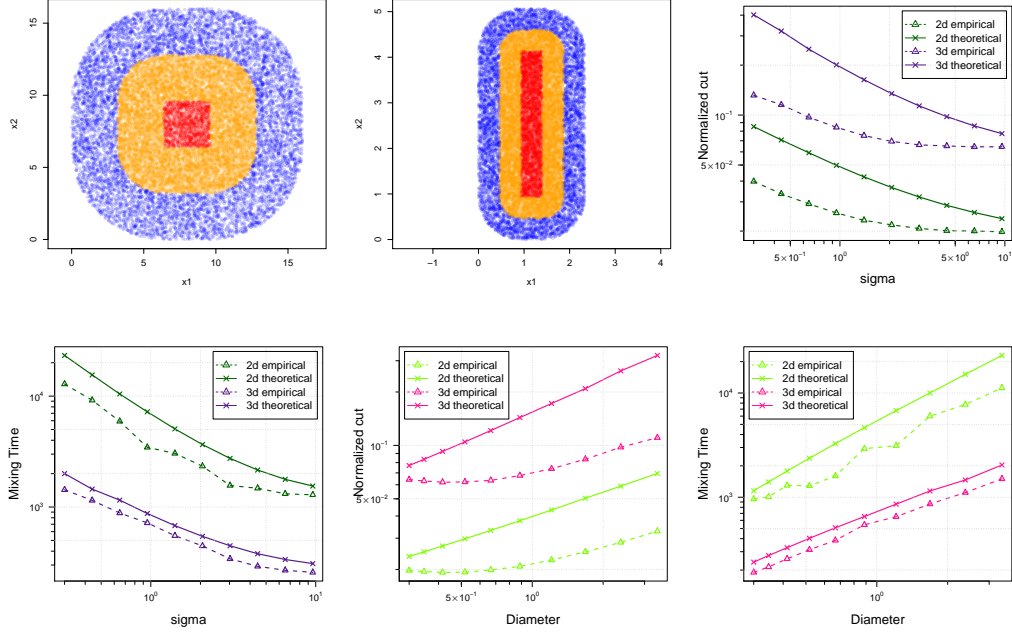


Figure 1: *Top left and top middle: samples from a geometrically well- and poor-conditioned cluster. The points in C are colored in red, points in $C_\sigma \setminus C$ are colored in yellow, and the remaining points in blue. Other panels: empirical normalized cut and mixing time, as a function of σ or ρ , versus their theoretical upper bounds.*

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

Remark 4. The embedding assumption (A4) and Lipschitz parameter L play an important role in proving the upper bound of Theorem 4. There is some interdependence between L and σ, ρ , which might lead one to hope that (A4) is non-essential. However, it is not possible to eliminate condition (A4) without incurring an additional factor of at least $(\rho/\sigma)^d$ in (17), achieved, for instance, when C_σ is a dumbbell-like set consisting of two balls of diameter ρ linked by a cylinder of radius σ . Abbasi-Yadkori et al. [2017], Abbasi-Yadkori [2016] develop theory regarding bi-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 4 both satisfy (A4) for reasonably small values of L .

4 Experiments

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

Validating theoretical bounds. We investigate the tightness of Theorems 3 and 4 via simulation. Figure 1 compares our upper bounds with the actual empirically-computed quantities (3) and (16), as we vary the diameter ρ and thickness σ of a cluster 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.³ This provides empirical evidence that Theorem 4 has the right dependency on both expansion parameter σ and diameter ρ . The story for the normalized

³We 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.

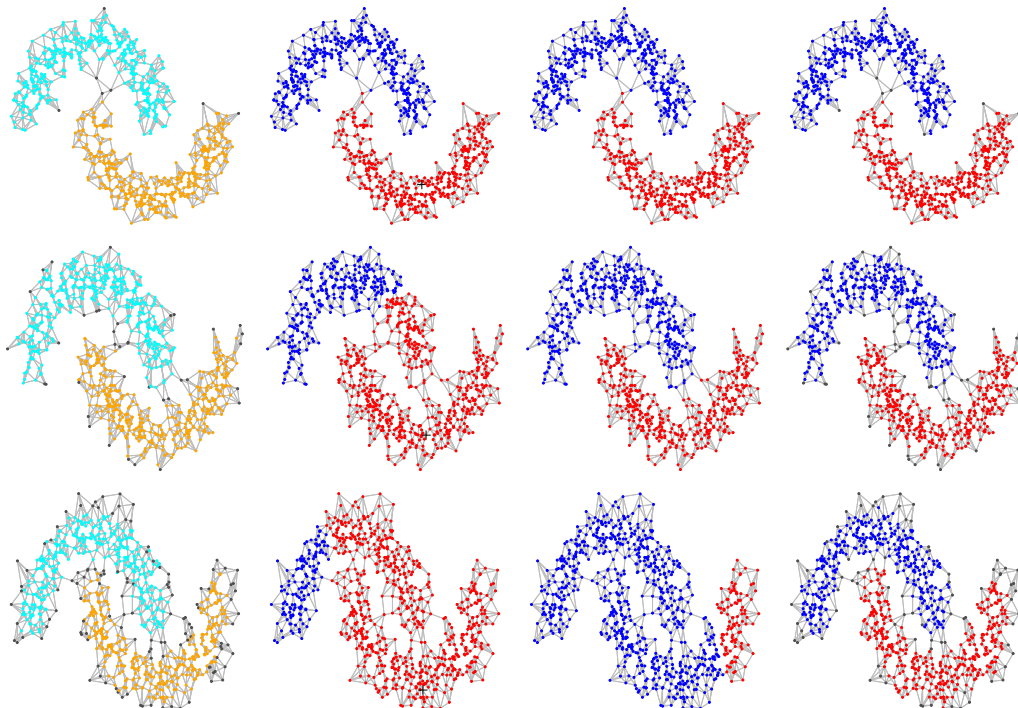


Figure 2: True density (column 1), PPR (column 2), normalized cut (column 3) and estimated density (column 4) clusters for 3 different simulated data sets. Seed node for PPR denoted by a black cross.

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 2, to drive home the main implications of Theorems 1 and 2, we show the behavior of PPR, normalized cut, and the density clustering algorithm of Chaudhuri and Dasgupta [2010] 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 and Bresson [2010]; and the last column shows a cut of the density cluster tree estimator of Chaudhuri and Dasgupta [2010]. 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, supporting our claim from Remark 3. Additionally, we note that the Chaudhuri-Dasgupta algorithm succeeds even when both PPR and normalized cut fail. While our main message was that PPR recovers geometrically well-conditioned density clusters, it would be interesting to establish that it *only* recovers such clusters, a direction for future work.

5 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 do not, however, provide theory to show that our geometric conditions are required for successful recovery of a density level set. This seems to be supported by our empirical results, and it remains a direction for future work.

In this supplement, we present proofs and experimental details for the article “Local Spectral Clustering of Density Upper Level Sets”.

A Proofs

Subsections A.1 - A.3 detail the proof for Theorem 3. In subsections A.4 and A.5, we establish some results regarding mixing time over general graphs. The proof of Theorem 4 is then carried out in subsections A.6 - A.10. A.11 gives some general concentration results we use throughout, before we finish with the proofs of Theorems 1 and 2, and the statement and proof of our results regarding the aPPR vector (Corollary 1), in subsections A.12 - A.14.

A.1 Volume estimates

We begin with some notation. Let $\mathcal{A} \subseteq \mathbb{R}^d$, and for $\sigma \geq 0$, write $\sigma B := B(0, \sigma) = \{x \in \mathbb{R}^d : \|x\| \leq \sigma\}$ 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 1 allows us to reason about the ratio $\nu(\mathcal{C}_\sigma + rB)/\nu(\mathcal{C}_\sigma)$, an important quantity in bounding the ratio $\text{cut}(\mathcal{C}_\sigma[X]; G_{n,r})/\text{vol}(\mathcal{C}_\sigma[X]; G_{n,r})$.

Lemma 1. *If \mathcal{A} is closed and bounded, then for any $\delta > 0$,*

$$\nu(\mathcal{A}_\sigma + \delta B) \leq \left(1 + \frac{\delta}{\sigma}\right)^d \nu(\mathcal{A}_\sigma). \quad (18)$$

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

$$\frac{\nu(\mathcal{A}_\sigma + \delta B)}{\nu(\mathcal{A}_\sigma)} \leq \frac{(\sigma + \delta + \epsilon)^d}{\sigma^d} \quad (19)$$

Taking the limit as $\epsilon \rightarrow 0$ results in (18).

Fix $\epsilon > 0$. Our first goal is to find a finite collection $x_1, \dots, x_N \in \mathbb{R}^d$ (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}_σ 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 $x_1, \dots, x_N \in \mathcal{A}$ such that

$$\mathcal{A}_\sigma \subset \bigcup_{i=1}^N B^\circ(x_i, \sigma + \epsilon). \quad (20)$$

As a direct consequence of (20), $\mathcal{A}_\sigma + \delta B \subset \bigcup_{i=1}^N B^\circ(x_i, \sigma + \epsilon + \delta)$, and by definition for every $x_i \in \mathcal{A}$, $B(x_i, \sigma) \in \mathcal{A}_\sigma$. Summarizing our findings, we have

$$\bigcup_{i=1}^N B(x_i, \sigma) \subseteq \mathcal{A}_\sigma, \mathcal{A}_\sigma + \delta B \subset \bigcup_{i=1}^N B^\circ(x_i, \sigma + \delta + \epsilon) \quad (21)$$

We next show a lower bound on $\nu(\mathcal{A}_\sigma)$. Partition \mathcal{A}_σ 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

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

Observe that $\bigcup_{i=1}^N \mathcal{A}_\sigma^{(i)} = \bigcup_{i=1}^N B(x_i, \sigma)$, so by (20) $\mathcal{A}_\sigma \supseteq \bigcup_{i=1}^N \mathcal{A}_\sigma^{(i)}$. As $\mathcal{A}_\sigma^{(1)}, \dots, \mathcal{A}_\sigma^{(N)}$ are non-overlapping,

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

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

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

As $\bigcup_{i=1}^N \mathcal{A}_{\sigma+\delta+\epsilon}^{(i)} = \bigcup_{i=1}^N B(x_i, \sigma + \delta + \epsilon)$, by (20)

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

and therefore

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

where the last inequality follows from Lemma 2. We have shown (19), and thus the claim. \square

We require Lemma 2 to prove Lemma 1.

Lemma 2. For $i = 1, \dots, N$ and $\mathcal{A}_\sigma^{(i)}, \mathcal{A}_{\sigma+\delta+\epsilon}^{(i)}$ as in Lemma 1,

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

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

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

since then

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

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

$$\|x\| \leq \sigma + \delta', \quad \|x - x_j\| > \sigma + \delta' \text{ for } j = 1, \dots, i-1. \quad (22)$$

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

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

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

We will need to carefully control the volume of expansion sets using the estimate in Lemma 1; Lemma 3 serves this purpose.

Lemma 3. *For any $0 \leq x \leq 1/2d$,*

$$(1+x)^d \leq 1+2dx$$

$$(1-x)^d \geq 1-2dx.$$

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

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

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

Let $\mathcal{C}_{\sigma, \sigma+r} := \{x : 0 < \text{dist}(x, \mathcal{C}_\sigma) < r\}$, where \mathcal{C}_σ is as in Theorem 3. Lemma 4 involves the bulk of the technical effort required to prove Theorem 3; it will be necessary to bound the expected cut size of $\mathcal{C}_\sigma[X]$ in $G_{n,r}$. We write $\mathbb{P}(\mathcal{A}) = \int_{\mathcal{A}} f(x) dx$ for measurable $\mathcal{A} \subseteq \mathbb{R}^d$.

Lemma 4. *Under the conditions of Theorem 3, and for any $0 < r \leq \sigma/2d$,*

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

Proof. We partition $\mathcal{C}_{\sigma, \sigma+r}$ into slices based on distance from \mathcal{C}_σ as follows: for $k \in \mathbb{N}$,

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

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

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

(A1) and (A3) 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

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

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

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

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

which in turn gives

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

(26) is a Riemann sum, and taking the limit as $k \rightarrow \infty$ we obtain

$$\begin{aligned} \lim_{k \rightarrow \infty} c_0 \nu(\mathcal{C}_\sigma) r^\gamma \sum_{i=1}^k \left(1 + \frac{ru_{i,k}^{1/\gamma}}{\sigma}\right)^d (u_{i,k} - u_{i,k-1}) &= c_0 \nu(\mathcal{C}_\sigma) r^\gamma \int_0^1 \left(1 + \frac{ru^{1/\gamma}}{\sigma}\right)^d du \\ &\stackrel{(i)}{\leq} c_0 \nu(\mathcal{C}_\sigma) r^\gamma \int_0^1 \left(1 + \frac{2dr u^{1/\gamma}}{\sigma}\right) du \\ &= c_0 \nu(\mathcal{C}_\sigma) r^\gamma \left(1 + \gamma \frac{2dr}{(\gamma+1)\sigma}\right). \end{aligned} \quad (27)$$

where (i) follows from Lemma 3 in light of the fact $r \leq \sigma/2d$.

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

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

where (ii) follows from (25), and (iii) from Lemma 3. As the bounds in (23) and (24) hold for all k , these along with (27) and (28) imply the desired result. \square

Lemma 5 will be necessary to lower bound the expected volume of $\mathcal{C}_\sigma[X]$ in $G_{n,r}$. Define the *uniform local conductance* $\ell_{\nu,r}(u)$ to be

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

Lemma 5. *Let $u \in \mathcal{C}_\sigma$. Then, for any $0 < r \leq \frac{\sigma}{2\sqrt{d}}$,*

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

Proof. Since $u \in \mathcal{C}_\sigma$ there exists $x \in \mathcal{C}$ such that $u \in B(x, \sigma)$, and as $B(x, \sigma) \subseteq \mathcal{C}_\sigma$,

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

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 6 has volume

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

with $\alpha = 1 - \frac{2rh-h^2}{r^2} = \frac{r^2}{4\sigma^2} \leq \frac{1}{16d}$.

Then by Lemmas 7 (applied with $t = 1$) and 8,

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

□

The following formula for the volume of the spherical cap, stated in terms of the incomplete beta function, is well known. We include it without proof.

Lemma 6. Let $\text{Cap}_r(h)$ denote a spherical cap of radius r and height h . Then,

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

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 Beta(z, w) distribution, evaluated at $1 - \alpha$.

Lemma 7. For any $0 \leq t \leq 1$ and $\alpha \leq \frac{t^2}{16d}$,

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

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

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

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

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

As a result,

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

where (iv) follows from Lemma 3, and the fact $\alpha \leq \frac{1}{16d}$. The result follows from the condition $\alpha \leq \frac{t^2}{16d}$. □

Lemma 8 follows from $\Gamma(1/2) = \sqrt{\pi}$ and the upper bound $\Gamma(x+1)/\Gamma(x+s) \leq (x+1)^{1-s}$ for $s \in [0, 1]$.

Lemma 8.

$$\frac{\Gamma(\frac{d}{2} + 1)}{\Gamma(\frac{d+1}{2})\Gamma(\frac{1}{2})} \leq \sqrt{\frac{d+2}{2\pi}}$$

A.2 Density-weighted cut and volume estimates

For notational ease, we write

$$\begin{aligned} \text{cut}_{n,r} &= \text{cut}(\mathcal{C}_\sigma[X]; G_{n,r}), \quad \mu_K = \mathbb{E}(\text{cut}_{n,r}), \quad p_K = \frac{\mu_K}{\binom{n}{2}} \\ \text{vol}_{n,r} &= \text{vol}(\mathcal{C}_\sigma[X]; G_{n,r}), \quad \mu_V = \mathbb{E}(\text{vol}_{n,r}), \quad p_V = \frac{\mu_V}{\binom{n}{2}} \\ \text{vol}_{n,r}^c &= \text{vol}(X \setminus \mathcal{C}_\sigma[X]; G_{n,r}), \quad \mu_V^c = \mathbb{E}(\text{vol}_{n,r}^c), \quad p_V^c = \frac{\mu_V^c}{\binom{n}{2}} \end{aligned}$$

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

Lemma 9. *Under the setup and conditions of Theorem 3, and for any $0 < r \leq \sigma/2d$,*

$$p_K \leq \frac{4d\nu_d r^{d+1}\lambda}{\sigma} \left(\lambda_\sigma - c_0 \frac{r^\gamma}{\gamma+1} \right) \nu(\mathcal{C}_\sigma)$$

Proof. We can write $\text{cut}_{n,r}$ as a double sum,

$$\text{cut}_{n,r} = \sum_{i=1}^n \sum_{j \neq i} \mathbf{1}(x_i \notin \mathcal{C}_\sigma) \mathbf{1}(x_j \in \mathcal{C}_\sigma) \mathbf{1}(\|x_i - x_j\| \leq r) \quad (29)$$

and by linearity of expectation, we obtain

$$p_K = \frac{\mu_K}{\binom{n}{2}} = 2 \cdot \mathbb{P}(x_i \notin \mathcal{C}_\sigma, x_j \in \mathcal{C}_\sigma, \|x_i - x_j\| \leq r). \quad (\text{for each } i, j, i \neq j)$$

Writing this with respect to the density function f , we have

$$\begin{aligned} p_K &= 2 \int_{\mathbb{R}^d \setminus \mathcal{C}_\sigma} f(x) \mathbb{P}(B(x, r) \cap \mathcal{C}_\sigma) dx \\ &= 2 \int_{\mathcal{C}_{\sigma, \sigma+r}} f(x) \mathbb{P}(B(x, r) \cap \mathcal{C}_\sigma) dx \\ &\leq 2\nu_d r^d \lambda \int_{\mathcal{C}_{\sigma, \sigma+r}} f(x) dx = 2\nu_d r^d \lambda \mathbb{P}(\mathcal{C}_{\sigma, \sigma+r}). \end{aligned}$$

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 9 gives the final result. \square

Lemma 10. *Under the setup and conditions of Theorem 3, and for any $0 < r \leq \sigma/2d$,*

$$p_V \geq \frac{12}{25} \lambda_\sigma^2 \nu_d r^d \nu(\mathcal{C}_\sigma)$$

Proof. The proof will proceed similarly to Lemma 9. We begin by writing $\text{vol}_{n,r}$ as the sum of indicator functions,

$$\text{vol}_{n,r} = \sum_{i=1}^n \sum_{j \neq i} \mathbf{1}(x_i \in \mathcal{C}_\sigma) \mathbf{1}(x_j \in B(x_i, r)) \quad (30)$$

and by linearity of expectation we obtain

$$p_V = \frac{\mu_V}{\binom{n}{2}} = 2 \cdot \mathbb{P}(x_i \in \mathcal{C}_\sigma, x_j \in B(x_i, r)). \quad (\text{for any } i, j, i \neq j.)$$

Writing this with respect to the density function f , we have

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

whence the claim then follows by Lemma 5. \square

To employ Lemmas 9 and 10 in the proof of Theorem 3, we must relate the random variable

$$\Phi_{n,r}(\mathcal{C}_\sigma[X]) = \frac{\text{cut}_{n,r}}{\min\{\text{vol}_{n,r}, \text{vol}_{n,r}^c\}}$$

to p_K and p_V .

In Lemma 11, we give probabilistic bounds on the $\text{cut}_{n,r}$, $\text{vol}_{n,r}$ and $\text{vol}_{n,r}^c$ in terms of p_K and p_V . These bounds are a straightforward consequence of Lemma 29, Hoeffding's inequality for U-statistics.

Lemma 11. *For any $\delta \in (0, 1]$,*

$$\frac{\text{cut}_{n,r}}{\binom{n}{2}} \leq p_K + \sqrt{\frac{\log(1/\delta)}{n}}, \text{ and } \frac{\text{vol}_{n,r}}{\binom{n}{2}}, \frac{\text{vol}_{n,r}^c}{\binom{n}{2}} \geq p_V - \sqrt{\frac{\log(1/\delta)}{n}}. \quad (31)$$

each with probability at least $1 - \delta$.

Proof of Lemma 11. From (29) and (30), we see that $\text{cut}_{n,r}$ and $\text{vol}_{n,r}$, properly scaled, can be expressed as order-2 U-statistics,

$$\frac{\text{cut}_{n,r}}{\binom{n}{2}} = \frac{1}{\binom{n}{2}} \sum_{1 \leq i < j \leq n} \phi_K(x_i, x_j), \quad \frac{\text{vol}_{n,r}}{\binom{n}{2}} = \frac{1}{\binom{n}{2}} \sum_{1 \leq i < j \leq n} \phi_V(x_i, x_j)$$

with kernels

$$\begin{aligned} \phi_K(x_i, x_j) &= \mathbf{1}(x_i \in \mathcal{C}_\sigma, x_j \notin \mathcal{C}_\sigma, \|x_i - x_j\| \leq r) + \mathbf{1}(x_j \in \mathcal{C}_\sigma, x_i \notin \mathcal{C}_\sigma, \|x_i - x_j\| \leq r) \\ \phi_V(x_i, x_j) &= \mathbf{1}(x_i \in \mathcal{C}_\sigma, \|x_i - x_j\| \leq r) + \mathbf{1}(x_j \in \mathcal{C}_\sigma, \|x_i - x_j\| \leq r). \end{aligned}$$

Similarly,

$$\frac{\text{vol}_{n,r}^c}{\binom{n}{2}} = \frac{1}{\binom{n}{2}} \sum_{1 \leq i < j \leq n} \phi_{V^c}(x_i, x_j)$$

with kernel,

$$\phi_{V^c}(x_i, x_j) = \mathbf{1}(x_i \notin \mathcal{C}_\sigma, \|x_i - x_j\| \leq r) + \mathbf{1}(x_j \notin \mathcal{C}_\sigma, \|x_i - x_j\| \leq r).$$

From Lemma 29 we therefore have

$$\frac{\text{cut}_{n,r}}{\binom{n}{2}} \leq p_K + \sqrt{\frac{\log(1/\delta)}{n}}, \quad \frac{\text{vol}_{n,r}}{\binom{n}{2}} \geq p_V - \sqrt{\frac{\log(1/\delta)}{n}}, \quad \frac{\text{vol}_{n,r}^c}{\binom{n}{2}} \geq p_V^c - \sqrt{\frac{\log(1/\delta)}{n}}$$

each with probability at least $1 - \delta$. The claim follows in light of (A5), which implies $p_V^c \geq p_V$. \square

A.3 Proof of Theorem 3

The proof of Theorem 3 is more or less given by Lemmas 9, 10, and 11. All that remains is some algebra, which we take care of below.

Fix $\delta \in (0, 1]$ and let $\delta' = \delta/3$. We rewrite

$$\Phi_{n,r}(\mathcal{C}_\sigma[X]) = \frac{p_K + \left(\frac{\text{cut}_{n,r}}{\binom{n}{2}} - p_K \right)}{p_V + \left(\frac{\min\{\text{vol}_{n,r}, \text{vol}_{n,r}^c\}}{\binom{n}{2}} - p_V \right)}. \quad (32)$$

Assume (31) holds with respect to δ' , keeping in mind that this will happen with probability at least $1 - \delta$. Along with (32) this means

$$\Phi_{n,r}(\mathcal{C}_\sigma[\mathbf{X}]) \leq \frac{p_K + \text{Err}_n}{p_V - \text{Err}_n}$$

for $\text{Err}_n = \sqrt{\frac{\log(1/\delta')}{n}}$. Now, some straightforward algebraic manipulations yield

$$\begin{aligned} \frac{p_K + \text{Err}_n}{p_V - \text{Err}_n} &= \frac{p_K}{p_V} \left(\frac{p_V}{p_V - \text{Err}_n} \right) + \frac{\text{Err}_n}{p_V - \text{Err}_n} \\ &= \frac{p_K}{p_V} + \left(\frac{p_K}{p_V} + 1 \right) \frac{\text{Err}_n}{p_V - \text{Err}_n} \\ &\leq \frac{p_K}{p_V} + 2 \frac{\text{Err}_n}{p_V - \text{Err}_n}. \end{aligned}$$

By Lemmas 9 and 10, we have

$$\frac{p_K}{p_V} \leq \frac{100rd}{12\sigma} \frac{\lambda}{\lambda_\sigma} \frac{\left(\lambda_\sigma - c_0 \frac{r^\gamma}{\gamma+1} \right)}{\lambda_\sigma}.$$

Then, by the choice of sample size in (14),

$$n \geq \frac{(2 + \epsilon)^2 \log\left(\frac{3}{\delta}\right)}{\epsilon^2 p_V^2}$$

which implies $2 \frac{\text{Err}_n}{p_V - \text{Err}_n} \leq \epsilon$.

A.4 Mixing time on graphs

We generalize some notation from the main text. For an undirected graph $G = (V, E)$, 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 . Denote the m -step probability distribution of this random walk originating from a particular $v \in V$ as $q^{(m)} : V \times V \rightarrow [0, 1]$, $q^{(m)}(v, u) = e_v W^m e_u$. An important intermediate quantity used to bound the relative pointwise mixing time $\tau_\infty(G)$ is the total variation distance between the distributions $q_v^{(m)} := q^{(m)}(v, \cdot)$ and π ,

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

We recall the *cut* and *volume* functionals over a graph, and introduce the *degree* functional as well. For $u \in V$, $S \subseteq V$,

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

The *local spread* is defined as

$$s(G) := \frac{9}{10} \cdot \min_{u \in V} \{\deg(u; G)\} \cdot \min_{u \in V} \{\pi(u)\}$$

Recalling the *normalized cut* $\Phi(S; G)$ is defined (as in (3)) as

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

we refer to the minimum normalized cut over V as the *conductance*,

$$\Phi(G) := \min_{S \subseteq V} \Phi(S; G).$$

Proposition 1 relates these geometric quantities to the TV distance between $q_v^{(t)}$ and stationary distribution π .

Proposition 1. For any $v \in V$, and any $0 < a \leq 1$,

$$\left\| q_v^{(t+3)} - \pi \right\|_{TV} \leq \left\{ as(G), \frac{1}{8} + \frac{1}{20} + \frac{1}{2 \min_{u \in V} \deg(u; G)} \right\} + \left(\frac{1}{1 - 2as(G)} + \frac{1}{2as(G)} \right) \left(1 - \frac{\Phi^2(G)}{2} \right)^t$$

As we will see, Proposition 1 is an essential step to proving an upper bound on the uniform mixing time. We justify this statement next, before moving on to proving Proposition 1 in the following subsection.

Uniform mixing time Consider the uniform distance⁴ between $q_v^{(t)}$ and π , given by

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

Lemma 12. Let $\left\| q_v^{(t)} - \pi \right\|_{TV} \leq \frac{1}{14} \max \left\{ 1, \frac{1}{s(G)} \right\}$. Then,

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

Proof. Fix $u \in V$ and let $m \geq t + 1$ be arbitrary. The stationarity of π gives

$$\begin{aligned} \frac{\pi(u) - q_v^m(u)}{\pi(u)} &= \sum_{y \in V} \left(\pi(y) - q^{(m-1)}(v, y) \right) \left(\frac{q^{(1)}(y, u) - \pi(u)}{\pi(u)} \right) \\ &\stackrel{(i)}{=} \sum_{y \neq u} \left(\pi(y) - q^{(m-1)}(v, y) \right) \left(\frac{q^{(1)}(y, u) - \pi(u)}{\pi(u)} \right) + \frac{\pi(u) - q^{(m-1)}(v, u)}{\pi(u)} \left(\frac{1}{2} - \pi(u) \right) \\ &\leq \sum_{y \neq u} \left(\pi(y) - q^{(m-1)}(v, y) \right) \left(\frac{q^{(1)}(y, u) - \pi(u)}{\pi(u)} \right) + \frac{\pi(u) - q^{(m-1)}(v, u)}{2\pi(u)} \end{aligned} \quad (33)$$

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

Then

$$\begin{aligned} \sum_{y \neq u} \left(\pi(y) - q^{(m-1)}(v, y) \right) \left(\frac{q^{(1)}(y, u) - \pi(u)}{\pi(u)} \right) &\leq \left\| q_v^{(m-1)} - \pi \right\|_{TV} \max_{y \neq u} \left| \frac{q^{(1)}(y, u) - \pi(u)}{\pi(u)} \right| \\ &\leq \left\| q_v^{(m-1)} - \pi \right\|_{TV} \max \left\{ 1, \max_{y \neq u} \left\{ \frac{q^{(1)}(y, u)}{\pi(u)} \right\} \right\} \\ &\leq \left\| q_v^{(m-1)} - \pi \right\|_{TV} \max \left\{ 1, \frac{1}{s(G)} \right\} \end{aligned} \quad (34)$$

since for $y \neq u$, $q^{(1)}(y, u) \leq 1/(2 \min_{u \in V} \deg(u; G))$. As $m - 1 \geq t$, it is well known Lovász and Simonovits [1990] that the laziness of the random walk guarantees $\left\| q_v^{(m-1)} - \pi \right\|_{TV} \leq \left\| q_v^{(t)} - \pi \right\|_{TV}$, and therefore by (34) and hypothesis,

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

Plugging this in to (33) and taking maximum on both sides, we obtain

$$d_{\text{unif}}(q_v^{(m)}, \pi) \leq \frac{2}{7} + \frac{d_{\text{unif}}(q_v^{(m-1)}, \pi)}{2} \quad (35)$$

The recurrence relation of (35) along with the initial condition $d_{\text{unif}}(q_v^{(t)}, \pi) \leq 1$ yields

$$d_{\text{unif}}(q_v^{(t+1)}, \pi) \leq \frac{8}{14} \Rightarrow d_{\text{unif}}(q_v^{(t+2)}, \pi) \leq \frac{10}{28} \Rightarrow d_{\text{unif}}(q_v^{(t+3)}, \pi) \leq \frac{1}{4}$$

and the claim is shown. \square

⁴Note d_{unif} is not formally a distance as it is not symmetric.

Together, Proposition 1 and Lemma 12 imply the main result of this section.

Proposition 2. Assume $\min_{u \in V} \deg(u; G) \geq 10$. Then,

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

Proof. Fix $a = \frac{1}{18}$, and let $\tau_0 = \frac{2}{\Phi^2(G)} \log \left(\frac{80}{as(G)} \right)$. Then,

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

and so by Proposition 1,

$$\left\| q_v^{(\tau_0+3)} - \pi \right\|_{TV} \leq \max \left\{ \frac{1}{20}, \frac{1}{8} + \frac{1}{20} + \frac{1}{20} \right\} + \left(\frac{2}{as(G)} \right) \frac{as(G)}{80} \leq \frac{1}{4}.$$

It is a well known fact Montenegro [2002] that if $\left\| q_v^{(t)} - \pi \right\|_{TV} \leq \frac{1}{4}$, then for any $0 < \epsilon < 1$, $\left\| q_v^{(t \log_2(1/\epsilon))} - \pi \right\|_{TV} \leq \epsilon$. Therefore, letting $\tau_1 = (\tau_0 + 3) \log(\frac{14}{s(G)})$,

$$\left\| q_v^{(\tau_1)} - \pi \right\|_{TV} \leq \frac{1}{14s(G)}$$

and so by Lemma 12, $d_{\text{unif}}(q_v^{(\tau_1+3)}, \pi) \leq \frac{1}{4}$. \square

A.5 Proof of Proposition 1.

For arbitrary starting distribution q (meaning $\text{supp}(q) \subseteq V$ and $\sum_{u \in V} q(u) = 1$), and for $t \geq 0$ an integer, let $q^{(t)}$ be the t -step probability distribution of the lazy random walk with starting distribution q . Consider the distance function $h_q^{(t)}, t \geq 0$,

$$h_q^{(t)}(x) = \max \left\{ \sum_{u \in V} (q^{(m)}(u) - \pi(u)) w(u) \right\}$$

where the maximum is over all $w : V \rightarrow [0, 1]$ such that $0 \leq w(u) \leq 1$ for all u , and $\sum_{u \in V} w(u) \pi(u) = x$. Writing $h_v^{(t)}(x) := h_{e_v}^{(t)}(x)$ in a small abuse of notation, in Lemmas 13 and 14, we give an upper bound on $h_v^{(t)}(x)$ for all $0 \leq x \leq 1$.

Remark 5. $h_q^{(t)}$ permits an equivalent definition. Order the elements of $V = \{u_1, \dots, u_N\}$ ($N = |V|$), such that

$$\frac{q^{(m)}(u_1)}{\pi(u_1)} \geq \frac{q^{(m)}(u_2)}{\pi(u_2)} \geq \dots \geq \frac{q^{(m)}(u_N)}{\pi(u_N)}$$

and let $U_k = \{u_1, \dots, u_k\}$. Then for any x , letting k satisfy $\pi(U_{k-1}) < x < \pi(U_k)$, it can be shown that,

$$h_q^{(t)}(x) = \sum_{j=1}^{k-1} (q^{(m)}(u_j) - \pi(u_j)) + \frac{x - \pi(U_{k-1})}{\pi(u_k)} (q^{(m)}(u_k) - \pi(u_k)). \quad (36)$$

where we have used the standard notation $\pi(S) = \sum_{u \in S} \pi(u)$. The formulation on the right hand side of (36) has come to be known as the *Lovasz-Simonovits curve*, a concave, piecewise linear curve.

Mixing over large sets For $0 \leq \mu \leq 1$ and $\mu \leq x \leq 1 - \mu$ let

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

be the linear interpolator between $h_q^{(0)}(\mu)$ and $h_q^{(0)}(1 - \mu)$.

Lemma 13. For any $0 \leq \mu \leq 1/2$ and $\mu \leq x \leq 1 - \mu$ and $t \geq 0$,

$$h_q^{(t)}(x) \leq \ell_\mu(x) + \max \left\{ \frac{h_q^{(0)}(\mu)}{1-2\mu} + \frac{h_q^{(0)}(\mu)}{\mu}, \frac{h_q^{(0)}(1-\mu)}{1-2\mu} + 1 \right\} \left(1 - \frac{\Phi^2(G)}{2} \right)^t$$

Theorem 13 is a direct consequence of Theorem 1.2 of Lovász and Simonovits [1990]. To state the latter, we introduce

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

Theorem 5 (Theorem 1.2 of Lovász and Simonovits [1990]). For any $0 \leq \mu \leq \frac{1}{2}$, $\mu \leq x \leq 1 - \mu$ and an integer $t \geq 0$,

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

Proof of Lemma 13. Fix $0 \leq \mu \leq \frac{1}{2}$. We will show that for all $\mu \leq x \leq 1 - \mu$

$$h_q^{(0)}(x) - \ell_\mu(x) \leq \max \left\{ \frac{h_q^{(0)}(\mu)}{1-2\mu} + \frac{h_q^{(0)}(\mu)}{\mu}, \frac{h_q^{(0)}(1-\mu)}{1-2\mu} + 1 \right\} \min \left\{ \sqrt{x-\mu}, \sqrt{1-x-\mu} \right\} \quad (37)$$

whence the claim follows by Theorem 5.

We observe that the slope $v(x)$ of each linear segment of the Lovasz-Simonovits curve satisfies

$$-1 \leq v(x) \leq \frac{h_q^{(0)}(\mu)}{\mu}. \quad (38)$$

Note that $\ell_\mu(\mu) = h_q^{(0)}(\mu)$, and for $x \geq \mu$,

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

by the concavity of h_0 along with (38). Some algebra then yields

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

On the other hand, $\ell_\mu(1-\mu) = h_q^{(0)}(1-\mu)$, and by the concavity of $h_q^{(0)}$ and (38), for $x \leq 1 - \mu$

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

Similar manipulations to above give the upper bound

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

and (37) follows. \square

Mixing over small sets.

Lemma 14. *Let $0 \leq a \leq 1$, and $t \geq 1$ an integer. Then for any $x \leq as(G)$ or $x \geq 1 - as(G)$,*

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

Proof. First, we deal with the case $x \leq as(G)$. Letting U_k be as in (36), we have

$$h_v^{(t)}(x) \leq q_v^{(m)}(U_{k-1}) + q_v^{(m)}(u_k) \quad (39)$$

We will rely on the key fact that for any $u \neq v, t \geq 1$,

$$q_v^{(t)}(u) \leq \frac{1}{2 \min_{u \in V} \deg(u; G)} \quad (40)$$

On the other hand if $u = v$,

$$q_v^{(m)}(u) \leq \frac{1}{2^t} + \frac{1}{2 \min_{u \in V} \deg(u; G)}. \quad (41)$$

Therefore by (39), (40), and (41)

$$h_v^{(t)}(x) \leq \frac{1}{2^t} + \frac{|U_k|}{2 \min_{u \in V} \deg(u; G)} + \frac{1}{2 \min_{u \in V} \deg(u; G)}.$$

Since $x \leq as(G)$,

$$|U_k| \leq \frac{x}{10 \min_{u \in V} (\pi(v))} \leq \frac{9a \min_{u \in V} \deg(u; G)}{10}.$$

For any $0 \leq b \leq 1$, $x \geq 1 - b$ implies $h_v^{(t)}(x) \leq b$. Taking $b = as(G)$, the claim is shown. \square

Proof of Proposition 1. For any $S \subseteq V$ and any integer $t \geq 0$,

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

and taking the maximum over all $S \subseteq V$ on both sides, we have

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

Letting $q = e_v W^3$, observe that $h_v^{(t+3)}(x) = h_q^{(t)}(x)$. Fix $\mu = as(G)$. Then, for $\mu \leq x \leq 1 - \mu$, by Lemma 13,

$$\begin{aligned} h_v^{(t+3)}(x) &= h_q^{(t)}(x) \\ &\leq \max \left\{ h_q^{(0)}(\mu), h_q^{(0)}(1 - \mu) \right\} + \left(\frac{1}{1 - 2\mu} + \frac{1}{2\mu} \right) \left(1 - \frac{\Phi^2(G)}{2} \right)^t \\ &\leq \max \left\{ as(G), \frac{1}{8} + \frac{1}{20} + \frac{1}{2 \min_{u \in V} \deg(u; G)} \right\} + \left(\frac{1}{1 - 2as(G)} + \frac{1}{2as(G)} \right) \left(1 - \frac{\Phi^2(G)}{2} \right)^t \end{aligned}$$

where the last inequality comes from application of Lemma 14 to $h_q^{(t)} = h_v^{(t+3)}$.

For $x \leq \mu$ or $x \geq 1 - \mu$, again by Lemma 14,

$$h_v^{(t+3)}(x) \leq \max \left\{ as(G), \frac{1}{8} + \frac{1}{20} + \frac{1}{2 \min_{u \in V} \deg(u; G)} \right\}$$

and the proof is complete.

A.6 Population-level conductance.

We turn now to lower bounding the conductance over $\tilde{G}_{n,r} := G_{n,r}[\mathcal{C}_\sigma[X]]$ (in general, we will use tilde notation to refer to quantities computed over \mathcal{C}_σ or over the induced subgraph $\tilde{G}_{n,r}$). Our first step will be to introduce a population analogue $\tilde{\Phi}_{\mathbb{P},r}$ of the conductance, and prove a lower bound on this quantity. Throughout this section, let $\mathcal{S}, \mathcal{S}_1, \mathcal{S}_2 \subseteq \mathcal{C}_\sigma$ be measurable.

We introduce the r -ball walk, a Markov chain over \mathcal{C}_σ with transition probability at $x \in \mathcal{C}_\sigma$ given by

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

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

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

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

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

a bit of algebra verifies that

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

We next introduce the *ergodic flow*, $\tilde{Q}_{\mathbb{P},r}$. Let $\mathcal{S}_1 \cap \mathcal{S}_2 = \mathcal{C}_\sigma$ be a partition of \mathcal{C}_σ . Then the ergodic flow between \mathcal{S}_1 and \mathcal{S}_2 is given by

$$\tilde{Q}_{\mathbb{P},r}(\mathcal{S}_1, \mathcal{S}_2) := \int_{\mathcal{S}_1} \tilde{P}_{\mathbb{P},r}(x; \mathcal{S}_2) d\tilde{\pi}_{\mathbb{P},r}(x),$$

the \mathbb{P} -(continuous) normalized cut by

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

and the \mathbb{P} -(continuous) conductance by

$$\tilde{\Phi}_{\mathbb{P},r} := \min_{\mathcal{S} \in \mathfrak{B}(\mathcal{C}_\sigma)} \tilde{\Phi}_{\mathbb{P},r}(\mathcal{S})$$

where $\mathcal{S}^c = \mathcal{C}_\sigma \setminus \mathcal{S}$ and $\mathfrak{B}(\mathcal{C}_\sigma)$ is the Borel σ -algebra of \mathcal{C}_σ .

Proposition 3. *Let \mathcal{C} satisfy Assumption (A1) for some $\lambda_\sigma \leq \Lambda_\sigma$ and Assumption (A4) for some convex set \mathcal{K} with diameter ρ , and measure-preserving mapping $g : \mathcal{K} \rightarrow \mathcal{C}_\sigma$ with biLipschitz constant L . Then, for any $0 < r < \frac{\sigma}{2\sqrt{d}}$, the \mathbb{P} -continuous conductance of the r -ball walk satisfies*

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

Similar results are already known (see e.g. Kannan et al. [2004]) when the density f is uniform (or log-concave) and \mathcal{C}_σ is itself a convex set – indeed, in this case stronger versions of it exist, though we will not require them. In Abbasi-Yadkori [2016], a statement of this sort is made with respect to uniform density f and \mathcal{C}_σ a biLipschitz deformation of a convex set, but for completeness we produce all proofs here.

Population-level conductance with uniform density We first prove an analogous result for the special case of $f \propto 1$ everywhere on \mathcal{C}_σ . For $x \in \mathcal{C}_\sigma$, $\mathcal{S} \in \mathfrak{B}(\mathcal{C}_\sigma)$, let

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

The uniform continuous normalized cut and conductance are then defined analogously to the weighted case,

$$\tilde{\Phi}_{\nu,r}(\mathcal{S}) := \frac{\tilde{Q}_{\nu,r}(\mathcal{S}, \mathcal{S}^c)}{\min\{\pi_{\nu,r}(\mathcal{S}), \pi_{\nu,r}(\mathcal{S}^c)\}}, \quad \tilde{\Phi}_{\nu,r} := \min_{\mathcal{S} \in \mathfrak{B}(\mathcal{C}_\sigma)} \tilde{\Phi}_{\nu,r}(\mathcal{S}).$$

Lemma 15. *Let \mathcal{C} satisfy Assumption (A4) for some convex set \mathcal{K} with diameter ρ , and measure-preserving mapping $g : \mathcal{K} \rightarrow \mathcal{C}_\sigma$ with biLipschitz constant L . Then, for any $0 < r < \frac{\sigma}{2\sqrt{d}}$, the uniform conductance of the r -ball walk satisfies*

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

Most of the technical work needed to show Proposition 3 involves proving Lemma 15. We defer this work to the next subsection, and first show that Proposition 3 is a simple consequence of Lemma 15 along with (A1).

Proof of Proposition 3. Assume Lemma 15 holds, and note that for any $\mathcal{S} \subseteq \mathcal{C}_\sigma$ with $\tilde{\pi}_{\mathbb{P},r}(\mathcal{S}) \leq 1/2$,

$$\tilde{\Phi}_{\mathbb{P},r}(\mathcal{S}) = \frac{\tilde{Q}_{\mathbb{P},r}(\mathcal{S}, \mathcal{S}^c)}{\tilde{\pi}_{\mathbb{P},r}(\mathcal{S})}$$

By (A1) we obtain

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

and the statement follows by Lemma 15.

A.7 Proof of Lemma 15.

As is standard, the proof of a lower bound on the conductance relies on an isoperimetric inequality.

Lemma 16 (Isoperimetry of Lipschitz embeddings of convex sets.). *Let \mathcal{C} satisfy Assumption (A4) for some convex set \mathcal{K} with diameter ρ , and measure-preserving mapping $g : \mathcal{K} \rightarrow \mathcal{C}_\sigma$ with biLipschitz constant L . Then, for any partition $(\Omega_1, \Omega_2, \Omega_3)$ of \mathcal{C}_σ ,*

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

The proof of Lemma 16 from first principles is non-trivial, even in the convex setting, and is a primary technical contribution of the seminal work Lovász and Simonovits [1990], extended by Dyer et al. [1991] among others. Once the result is shown in the case where Ω is convex, however, generalizing to the setting given by Assumption (A4) is not difficult.

Proof of Lemma 16. For $\Omega_i, i = 1, 2, 3$, denote the preimage

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

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

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

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

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

By Theorem 2.2 of Lovász and Simonovits [1990],

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

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

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

□

We will also need an upper bound on the distance between transition probability densities $\tilde{P}_{\nu,r}(u, \cdot)$ and $\tilde{P}_{\nu,r}(v, \cdot)$ for $u, v \in \mathcal{C}_\sigma$ close together.

Lemma 17 (One-step distributions). *Let $u, v \in \mathcal{C}_\sigma$ be such that*

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

for some $0 < t < 1/8$, and further assume there exists $\ell > 0$ such that $\ell_{\nu,r}(u), \ell_{\nu,r}(v) \geq \ell \nu_d r^d$. Then,

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

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

The key result needed to show Lemma 17 deals with volume of the overlap $B(u, r) \cap B(v, r)$.

Lemma 18. *Let $u, v \in \mathbb{R}^d$ be points such that $\|u - v\| \leq t \frac{r}{2\sqrt{d}}$ for some $0 < t < 1/8$. 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 $\|u - v\| = t \frac{r}{\sqrt{d}}$; if they are closer together the overlap of the volume will only increase. Then, it is not hard to see that $I = B(u, r) \cap B(v, r)$ consists of two symmetric spherical caps, each with height

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

From Lemma 6 we therefore obtain

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

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

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

where (i) follows from Lemma 7 (which we may validly apply since $\alpha \leq \frac{t^2}{16d}$), and (ii) from Lemma 8. □

Proof of Lemma 17. Let $\mathcal{S}_1 \cup \mathcal{S}_2 = \mathcal{C}_\sigma$ be an arbitrary partition of \mathcal{C}_σ . We will show that

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

Since this will hold for arbitrary $S_1 \in \mathfrak{B}(\mathcal{C}_\sigma)$, it will hold for the infimum over all such S_1 as well, and therefore the same lower bound will hold for $\left\| \tilde{P}_{\nu,r}(u; \cdot) - \tilde{P}_{\nu,r}(v; \cdot) \right\|_{TV}$.

Now, note that

$$\tilde{P}_{\nu,r}(u; S_1) - \tilde{P}_{\nu,r}(v; S_1) = 1 - \tilde{P}_{\nu,r}(u; S_2) - \tilde{P}_{\nu,r}(v; S_1)$$

Let $I := B(u, r) \cap B(v, r)$. Then we have

$$\tilde{P}_{\nu,r}(u; S_2) \geq \frac{1}{\nu(B(u, r))} \nu(S_2 \cap (B(u, r))) \geq \frac{1}{\nu(B(u, r))} \nu(S_2 \cap I)$$

with a symmetric inequality holding for $\tilde{P}_{\nu,r}(v; S_1)$. As a result,

$$1 - \tilde{P}_{\nu,r}(u; S_2) - \tilde{P}_{\nu,r}(v; S_1) \leq 1 - \frac{1}{\nu_d r^d} \nu(\mathcal{C}_\sigma \cap I) \quad (42)$$

As (42) demonstrates, the overlap of the one-step distributions is related to the volume of the intersection between $B(u, r)$ and $B(v, r)$ within \mathcal{C}_σ .

From here, some simple manipulations yield

$$\begin{aligned} \nu(\mathcal{C}_\sigma \cap I) &= \nu(I) - \nu(I \setminus \mathcal{C}_\sigma) \\ &\geq \nu(I) - \max \{ \nu(B(u, r) \setminus \mathcal{C}_\sigma), \nu(B(v, r) \setminus \mathcal{C}_\sigma) \} \\ &\geq \nu_d r^d \left(1 - \frac{3\sqrt{3}t}{4\sqrt{2\pi}} - (1 - \ell) \right) = \nu_d r^d \left(\ell - \frac{3\sqrt{3}t}{4\sqrt{2\pi}} \right) \end{aligned} \quad (43)$$

where the last inequality follows from Lemma 18. (43) along with (42) then give the desired result. \square

Proof of Lemma 15. Let $S_1 \cup S_2 = \mathcal{C}_\sigma$ be arbitrary, and let $\ell \geq 0$ satisfy $\ell \nu_d r^d \leq \ell_{\nu,r}(x)$. We will show that

$$\int_{S_1} \tilde{P}_{\nu,r}(x; S_2) d\pi_{\nu,r}(x) \geq \frac{\sqrt{2\pi} r \ell^4}{48 \rho L \sqrt{d}} \min \{ \pi_{\nu,r}(S_1), \pi_{\nu,r}(S_2) \}$$

Once we have shown this, Lemma 5 gives the bound $\ell \geq \frac{6}{25}$. Then, dividing both sides by $\pi_{\nu,r}(S_1)$ yields the desired result, since S_1 was arbitrary.

Now, consider the sets

$$\begin{aligned} S'_1 &= \left\{ x \in S_1 : \tilde{P}_{\nu,r}(x; S_2) < \frac{\ell}{4} \right\} \\ S'_2 &= \left\{ x \in S_1 : \tilde{P}_{\nu,r}(x; S_2) < \frac{\ell}{4} \right\} \end{aligned}$$

and $S'_3 = \mathcal{C}_\sigma \setminus (S'_1 \cup S'_2)$.

Suppose $\pi_{\nu,r}(S'_1) < \pi_{\nu,r}(S_1)/2$. Then,

$$\int_{S_1} \tilde{P}_{\nu,r}(x; S_2) d\pi_{\nu,r}(x) \geq \frac{\ell \pi_{\nu,r}(S_1)}{8}$$

Similarly, if $\pi_{\nu,r}(S'_1) < \pi_{\nu,r}(S_1)/2$, then since

$$\int_{S_1} \tilde{P}_{\nu,r}(x; S_2) d\pi_{\nu,r}(x) = \int_{S_2} \tilde{P}_{\nu,r}(x; S_2) d\pi_{\nu,r}(x)$$

a symmetric result holds.

So we can assume $\pi_{\nu,r}(S'_1) \geq \pi_{\nu,r}(S_1)/2$, and likewise for S_2 . Now, for every $u \in S'_1, v \in S'_2$, we have that

$$\left\| \tilde{P}_{\nu,r}(u; \cdot) - \tilde{P}_{\nu,r}(v; \cdot) \right\|_{TV} \geq 1 - \tilde{P}_{\nu,r}(u; S_1) - \tilde{P}_{\nu,r}(v; S_2) > 1 - \frac{\ell}{2}.$$

By Lemma 17, we therefore have

$$\|u - v\| \geq \frac{2\sqrt{2\pi r\ell}}{6\sqrt{3d}}.$$

and since $u \in \mathcal{S}'_1, v \in \mathcal{S}'_2$ were arbitrary, the same inequality holds for $\text{dist}(\mathcal{S}'_1, \mathcal{S}'_2)$. Therefore by Lemma 16

$$\nu(\mathcal{S}'_3) \geq \frac{2\sqrt{2\pi r\ell}}{6\rho L\sqrt{3d}} \min\{\nu(\mathcal{S}'_1), \nu(\mathcal{S}'_2)\}$$

We now prove the desired result:

$$\begin{aligned} \int_{\mathcal{S}_1} \tilde{P}_{\nu,r}(x; \mathcal{S}_2) &= \frac{1}{2} \left(\int_{\mathcal{S}_2} \tilde{P}_{\nu,r}(x; \mathcal{S}_2) d\pi_{\nu,r}(x) \right) \\ &\geq \frac{\ell}{8} \pi_{\nu,r}(\mathcal{S}'_3) \\ &\geq \frac{\ell^2}{8\nu(\mathcal{C}_\sigma)} \nu(\mathcal{S}'_3) \\ &\geq \frac{\sqrt{2}r\ell^3}{24\rho L\sqrt{d}\nu(\mathcal{C}_\sigma)} \min\{\nu(\mathcal{S}'_1), \nu(\mathcal{S}'_2)\} \\ &\geq \frac{\sqrt{2}r\ell^4}{24\rho L\sqrt{d}} \min\{\pi_{\nu,r}(\mathcal{S}'_1), \pi_{\nu,r}(\mathcal{S}'_2)\} \\ &\geq \frac{\sqrt{2}r\ell^4}{48\rho L\sqrt{d}} \min\{\pi_{\nu,r}(\mathcal{S}_1), \pi_{\nu,r}(\mathcal{S}_2)\}. \end{aligned}$$

A.8 Graph conductance and local spread.

Recall that $\tilde{G}_{n,r} = G_{n,r}[\mathcal{C}_\sigma[X]]$ denotes the subgraph induced by $\mathcal{C}_\sigma[X]$. We denote the graph conductance over $\tilde{G}_{n,r}$ as $\tilde{\Phi}_{n,r} := \Phi(G_{n,r}[\mathcal{C}_\sigma[X]])$. To simplify notation, for $S \subseteq \mathcal{C}_\sigma[X]$ and $u \in \mathcal{C}_\sigma[X]$, we will write

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

and let $\tilde{\pi}_{n,r}(u) = \widetilde{\text{deg}}_{n,r}(u) / \widetilde{\text{vol}}_{n,r}(\mathcal{C}_\sigma[X])$ be the stationary distribution of the lazy random walk over $\tilde{G}_{n,r}$.

Proposition 4 (Lower bound on graph conductance). *Let \mathcal{C} satisfy Assumption (A1) for some $\lambda_\sigma \leq \Lambda_\sigma$ and Assumption (A4) for some convex set \mathcal{K} with diameter ρ , and measure-preserving mapping $g : \mathcal{K} \rightarrow \mathcal{C}_\sigma$ with biLipschitz constant L . Then, with probability 1 the following lower bound holds on the graph conductance:*

$$\liminf_{n \rightarrow \infty} \tilde{\Phi}_{n,r} \geq \frac{\lambda_\sigma^2 r}{\Lambda_\sigma^2 2^{13} \rho L \sqrt{d}}$$

In order to prove Proposition 4, we will need to split the analysis into two cases based on the size of $S \subseteq \mathcal{C}_\sigma[X]$. For any graph $G = (V, E)$, let $\mathcal{L}(G) = \{S \subseteq V : \pi(S) \geq s(G)\}$ (where as usual π denotes the stationary distribution of a random walk over G .)

In Lemma 19, we establish a uniform lower bound on the normalized cut of all sets $S \in \mathcal{L}(\tilde{G}_{n,r})$.

Lemma 19. *Let \mathcal{C} satisfy Assumption (A1) for some $\lambda_\sigma \leq \Lambda_\sigma$ and Assumption (A4) for some convex set \mathcal{K} with diameter ρ , and measure-preserving mapping $g : \mathcal{K} \rightarrow \mathcal{C}_\sigma$ with biLipschitz constant L . Then with probability 1,*

$$\liminf_{n \rightarrow \infty} \left\{ \min_{S \in \mathcal{L}(\tilde{G}_{n,r})} \tilde{\Phi}_{n,r}(S) \right\} \geq \frac{\lambda_\sigma^2 r}{\Lambda_\sigma^2 2^{13} \rho L \sqrt{d}}$$

where $\mathcal{L}(\tilde{G}_{n,r}) = \{S \subseteq \mathcal{C}_\sigma[X] : \tilde{\pi}_{n,r}(S) \geq s(\tilde{G}_{n,r})\}$.

The proof of Lemma 19 is nontrivial, and we devote the following subsection to it.

Lemma 20 shows that for the remaining small sets, the graph normalized cut is of constant order.

Lemma 20. *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:

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

□

Proposition 4 follows in light of the fact $\frac{\lambda_\sigma^2 r}{\Lambda_\sigma^2 2^{13} \rho L \sqrt{d}} < \frac{1}{10}$.

In Lemma 21, we lower bound the local spread $s(\tilde{G}_{n,r})$.

Lemma 21. *With probability one, the following bound holds:*

$$\liminf_{n \rightarrow \infty} s(\tilde{G}_{n,r}) \geq \frac{\lambda_\sigma^2 \nu_d r^d}{20 \Lambda_\sigma}$$

Proof. Letting $\widetilde{\deg}_{\min} := \min_{u \in \mathcal{C}_\sigma[X]} \widetilde{\deg}_{n,r}(u)$, we rewrite $s(\tilde{G}_{n,r}) = \frac{9 \widetilde{\deg}_{\min}^2}{10 \text{vol}_{n,r}(\mathcal{C}_\sigma[X])}$. Then by Lemma 31, with probability one

$$\liminf_{n \rightarrow \infty} s(\tilde{G}_{n,r}) \geq \frac{36 \lambda_\sigma r^d \nu_d}{625 \mathbb{P}(\mathcal{C}_\sigma) \Lambda_\sigma} \geq \frac{\lambda_\sigma^2 \nu_d r^d}{20 \Lambda_\sigma}.$$

□

A.9 Proof of Lemma 19

For $S \subseteq \mathcal{C}_\sigma$ measurable, define the conditional probability measures

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

A Borel map $T : \mathcal{C}_\sigma \rightarrow \mathcal{C}_\sigma[X]$ is said to be a *transportation map* between $\tilde{\mathbb{P}}$ and $\tilde{\mathbb{P}}_n$ if for any $S \subseteq \mathcal{C}_\sigma[X]$

$$\tilde{\mathbb{P}}(T^{-1}(S)) = \tilde{\mathbb{P}}_n(S).$$

where $T^{-1}(S) = \{x \in \mathcal{C}_\sigma : T(x) \in S\}$ is the preimage of T .

If a sequence of transportation maps $\{T_n\}_{n \in \mathbb{N}}$ satisfies $\lim_{n \rightarrow \infty} \|\text{Id} - T_n\|_{L^\infty(\tilde{\mathbb{P}})} = 0$ (where Id is the identity mapping), we refer to it as a sequence of *stagnating transportation maps*. Proposition 5 of Trillos et al. [2016] establishes that, for open connected domains with Lipschitz boundaries, such stagnating transportation maps exist. In particular, as $\mathcal{C}_\sigma[X]$ consists of $\tilde{n} := |\mathcal{C}_\sigma[X]|$ points sampled independently from $\tilde{\mathbb{P}}$, and \mathcal{C}_σ is a connected domain with Lipschitz boundary, the following implication is immediate.

Lemma 22. *With probability one, there exists a sequence of transportation maps $\{T_n\}_{n \in \mathbb{N}}$, $T_n : \mathcal{C}_\sigma \rightarrow \mathcal{C}_\sigma[X]$ such that the following statement holds:*

$$\limsup_{n \rightarrow \infty} \frac{\tilde{n}^{1/d} \|\text{Id} - T_n\|_{L^\infty(\tilde{\mathbb{P}})}}{(\log \tilde{n})^{p_d}} \leq C \quad (44)$$

where $\text{Id}(x) = x$ is the identity mapping over \mathcal{C}_σ , C is a universal constant and $p_d = 3/4$ for $d = 2$ and $1/d$ for $d \geq 3$.

Note that although \mathcal{C}_σ is closed not open, as $\nu(\partial\mathcal{C}_\sigma) = 0$ we may apply Proposition 5 of Trillos et al. [2016] to the interior \mathcal{C}_σ° of \mathcal{C}_σ , and (44) will hold for any arbitrary extension of T_n to \mathcal{C}_σ .

Graph cuts to continuous cuts. We use stagnating transportation maps to relate discrete graph normalized cuts to the continuous normalized cuts discussed in Section A.6. We fix some notation, letting $r_n^\pm = r \pm \|\text{Id} - T_n\|_{L^\infty(\tilde{\mathbb{P}})}$ for a transportation map T_n .

Lemmas 23 and 24 provide the necessary bounds for the cut and vol functionals in terms of continuous analogues.

Lemma 23. *Let $S \subseteq \mathcal{C}_\sigma[X]$ and T_n be a transportation map between $\tilde{\mathbb{P}}$ and $\tilde{\mathbb{P}}_n$. Then, letting $\mathcal{S} := T_n^{-1}(S)$,*

$$\frac{1}{\tilde{n}^2} \widetilde{\text{vol}}_{n,r}(S) \leq \frac{\int_{\mathcal{C}_\sigma} \ell_{\mathbb{P},r_n^+}(x) f(x) dx}{2\mathbb{P}(\mathcal{C}_\sigma)^2} \tilde{\pi}_{\mathbb{P},r_n^+}(\mathcal{S}).$$

Proof. Let $\varphi : \mathcal{C}_\sigma[X] \rightarrow \{0, 1\}$ be the characteristic function for S , meaning

$$\varphi(x) = \begin{cases} 1, & x \in S \\ 0, & \text{otherwise} \end{cases}$$

Now, we proceed

$$\begin{aligned} \frac{1}{\tilde{n}^2} \widetilde{\text{vol}}_{n,r}(S) &\leq \frac{1}{2\tilde{n}^2} \sum_{x_i, x_j \in \mathcal{C}_\sigma[X]} \mathbf{1}(\|x_i - x_j\| \leq r) |\varphi(x_i)| \\ &= \frac{1}{2} \iint_{\mathcal{C}_\sigma \times \mathcal{C}_\sigma} \mathbf{1}(\|x - y\| \leq r) |\varphi(x)| d\tilde{\mathbb{P}}_n(x) d\tilde{\mathbb{P}}_n(y) \\ &= \frac{1}{2} \iint_{\mathcal{C}_\sigma \times \mathcal{C}_\sigma} \mathbf{1}(\|T_n(x) - T_n(y)\| \leq r) |\varphi \circ T_n(x)| d\tilde{\mathbb{P}}(x) d\tilde{\mathbb{P}}(y) \\ &\quad \text{(change of variables)} \\ &\leq \frac{1}{2} \iint_{\mathcal{C}_\sigma \times \mathcal{C}_\sigma} \mathbf{1}(\|x - y\| \leq r_n^+) |\varphi \circ T_n(x)| d\tilde{\mathbb{P}}(x) d\tilde{\mathbb{P}}(y) \\ &= \frac{1}{2} \int_{\mathcal{S}} \int_{\mathcal{C}_\sigma \cap B(x, r_n^+)} 1 d\tilde{\mathbb{P}}(y) d\tilde{\mathbb{P}}(x) \end{aligned}$$

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

$$\begin{aligned} \int_{\mathcal{S}} \int_{\mathcal{C}_\sigma \cap B(x, r_n^+)} 1 d\tilde{\mathbb{P}}(y) d\tilde{\mathbb{P}}(x) &= \frac{1}{\mathbb{P}(\mathcal{C}_\sigma)^2} \int_{\mathcal{S}} \int_{\mathcal{C}_\sigma \cap B(x, r_n^+)} 1 d\mathbb{P}(y) d\mathbb{P}(x) \\ &= \frac{1}{\mathbb{P}(\mathcal{C}_\sigma)^2} \int_{\mathcal{S}} \ell_{\mathbb{P},r_n^+}(x) f(x) dx \\ &= \frac{\int_{\mathcal{C}_\sigma} \ell_{\mathbb{P},r}(x) f(x) dx}{\mathbb{P}(\mathcal{C}_\sigma)^2} \tilde{\pi}_{\mathbb{P},r_n^+}(\mathcal{S}) \end{aligned}$$

and we have the desired upper bound. \square

Lemma 24. Let $S \subseteq \mathcal{C}_\sigma[X]$, and let T_n be a transportation map between $\tilde{\mathbb{P}}$ and $\tilde{\mathbb{P}}_n$. Then, letting $\mathcal{S} = T_n^{-1}(S)$,

$$\frac{1}{\tilde{n}^2} \widetilde{\text{cut}}_{n,r}(S) \geq \frac{\int_{\mathcal{C}_\sigma} \ell_{\mathbb{P},r_n^-} f(x) dx}{2\mathbb{P}(\mathcal{C}_\sigma)^2} \tilde{Q}_{\mathbb{P},r_n^-}(\mathcal{S}, \mathcal{S}^c).$$

Proof. As in the proof of Lemma 23, let $\varphi : \mathcal{C}_\sigma[X] \rightarrow \{0, 1\}$ be the characteristic function for S .

We proceed according to a very similar set of steps as Lemma 23:

$$\begin{aligned} \frac{1}{\tilde{n}^2} \widetilde{\text{cut}}_{n,r}(S) &= \frac{1}{2\tilde{n}^2} \sum_{x_i, x_j \in \mathcal{C}_\sigma[X]} \mathbf{1}(\|x_i - x_j\| \leq r) |\varphi(x_i) - \varphi(x_j)| \\ &= \frac{1}{2} \iint_{\mathcal{C}_\sigma \times \mathcal{C}_\sigma} \mathbf{1}(\|x - y\| \leq r) |\varphi(x) - \varphi(y)| d\tilde{\mathbb{P}}_n(x) d\tilde{\mathbb{P}}_n(y) \\ &= \frac{1}{2} \iint_{\mathcal{C}_\sigma \times \mathcal{C}_\sigma} \mathbf{1}(\|T_n(x) - T_n(y)\| \leq r) |\varphi \circ T_n(x) - \varphi \circ T_n(y)| d\tilde{\mathbb{P}}(x) d\tilde{\mathbb{P}}(y) \\ &\geq \frac{1}{2} \iint_{\mathcal{C}_\sigma \times \mathcal{C}_\sigma} \mathbf{1}(\|x - y\| \leq r_n^-) |\varphi \circ T_n(x) - \varphi \circ T_n(y)| d\tilde{\mathbb{P}}(x) d\tilde{\mathbb{P}}(y) \\ &= \frac{1}{2} \int_{\mathcal{S}} \int_{\mathcal{S}^c \cap B(x, r_n^-)} 1 d\tilde{\mathbb{P}}(y) d\tilde{\mathbb{P}}(x) \end{aligned}$$

We conclude similarly to the proof of Lemma 23,

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

□

Putting Lemmas 23 and 24 together yields the following result, a lower bound on graph normalized cut.

Lemma 25. Let $\{T_n\}_{n \in \mathbb{N}}$ be a sequence of transportation maps from $\tilde{\mathbb{P}}$ to $\tilde{\mathbb{P}}_n$, and fix $S \subseteq \mathcal{C}_\sigma[X]$. Then, letting $\mathcal{S} = T_n^{-1}(S)$,

$$\tilde{\Phi}_{n,r}(S) \geq \frac{\int_{\mathcal{C}_\sigma} \ell_{\mathbb{P},r_n^-}(x) f(x) dx}{\int_{\mathcal{C}_\sigma} \ell_{\mathbb{P},r_n^+}(x) f(x) dx} \frac{\min \left\{ \tilde{\pi}_{\mathbb{P},r_n^-}(\mathcal{S}), \tilde{\pi}_{\mathbb{P},r_n^-}(\mathcal{S}^c) \right\}}{\min \left\{ \tilde{\pi}_{\mathbb{P},r_n^+}(\mathcal{S}), \tilde{\pi}_{\mathbb{P},r_n^+}(\mathcal{S}^c) \right\}} \tilde{\Phi}_{\mathbb{P},r_n^-}(\mathcal{S}) \quad (45)$$

Proof. By Lemmas 23 and 24,

$$\frac{\widetilde{\text{cut}}_{n,r}(S)}{\widetilde{\text{vol}}_{n,r}(S)} \geq \frac{\int_{\mathcal{C}_\sigma} \ell_{\mathbb{P},r_n^-}(x) f(x) dx}{\int_{\mathcal{C}_\sigma} \ell_{\mathbb{P},r_n^+}(x) f(x) dx} \frac{\tilde{Q}_{\mathbb{P},r_n^-}(\mathcal{S}, \mathcal{S}^c)}{\tilde{\pi}_{\mathbb{P},r_n^+}(\mathcal{S})}.$$

Then, noting that $\mathcal{S}^c = T_n^{-1}(\mathcal{S}^c)$, Lemmas 23 and 24 also imply

$$\frac{\widetilde{\text{cut}}_{n,r}(\mathcal{S}^c)}{\widetilde{\text{vol}}_{n,r}(\mathcal{S}^c)} \geq \frac{\int_{\mathcal{C}_\sigma} \ell_{\mathbb{P},r_n^-}(x) f(x) dx}{\int_{\mathcal{C}_\sigma} \ell_{\mathbb{P},r_n^+}(x) f(x) dx} \frac{\tilde{Q}_{\mathbb{P},r_n^-}(\mathcal{S}^c, \mathcal{S})}{\tilde{\pi}_{\mathbb{P},r_n^+}(\mathcal{S}^c)}$$

and as $\tilde{Q}_{\mathbb{P},r_n^-}(\cdot, \cdot)$ is symmetric in its arguments we obtain

$$\frac{\widetilde{\text{cut}}(S)}{\min \left\{ \widetilde{\text{vol}}(S), \widetilde{\text{vol}}(\mathcal{S}^c) \right\}} \geq \frac{\int_{\mathcal{C}_\sigma} \ell_{\mathbb{P},r_n^-}(x) f(x) dx}{\int_{\mathcal{C}_\sigma} \ell_{\mathbb{P},r_n^+}(x) f(x) dx} \frac{\tilde{Q}_{\mathbb{P},r_n^-}(\mathcal{S}, \mathcal{S}^c)}{\min \left\{ \tilde{\pi}_{\mathbb{P},r_n^+}(\mathcal{S}), \tilde{\pi}_{\mathbb{P},r_n^+}(\mathcal{S}^c) \right\}},$$

and the proof is complete. □

Perturbation analysis. If $r_n^- = r = r_n^+$, (45) would simplify to $\tilde{\Phi}_{n,r}(S) \geq \tilde{\Phi}_{\mathbb{P},r}(S)$. We show that this conclusion is robust to asymptotically negligible perturbations of r .

Lemma 26 (Continuity of local conductance). *For $\{T_n\}_{n \in \mathbb{N}}$ a sequence of stagnating transportation maps,*

$$\limsup_{n \rightarrow \infty} \frac{\int_{\mathcal{C}_\sigma} \ell_{\mathbb{P},r_n^+}(x) f(x) dx}{\int_{\mathcal{C}_\sigma} \ell_{\mathbb{P},r_n^-}(x) f(x) dx} = 1$$

Proof. Letting $\mathcal{R}_n(x) := \{x' \in \mathcal{C}_\sigma : x' \in B(x, r_n^+), x' \notin B(x, r_n^-)\}$, we have

$$\int_{\mathcal{C}_\sigma} \ell_{\mathbb{P},r_n^+}(x) f(x) dx = \int_{\mathcal{C}_\sigma} \ell_{\mathbb{P},r_n^-}(x) f(x) dx + \int_{\mathcal{C}_\sigma} \int_{\mathcal{R}_n(x)} f(y) f(x) dy dx.$$

and therefore

$$\frac{\int_{\mathcal{C}_\sigma} \ell_{\mathbb{P},r_n^+}(x) f(x) dx}{\int_{\mathcal{C}_\sigma} \ell_{\mathbb{P},r_n^-}(x) f(x) dx} = 1 + \frac{\int_{\mathcal{C}_\sigma} \int_{\mathcal{R}_n(x)} f(y) f(x) dy dx}{\int_{\mathcal{C}_\sigma} \ell_{\mathbb{P},r_n^-}(x) f(x) dx}.$$

We upper bound the remainder term

$$\int_{\mathcal{C}_\sigma} \int_{\mathcal{R}_n(x)} f(y) f(x) dy dx \leq \mathbb{P}(\mathcal{C}_\sigma) \Lambda_\sigma \nu_d ((r_n^+)^d - r^d) \rightarrow 0$$

where the convergence happens as $n \rightarrow \infty$ by the stagnating property of $\{T_n\}_{n \in \mathbb{N}}$.

We apply a similar analysis to the denominator. By Lemma 5

$$\int_{\mathcal{C}_\sigma} \ell_{\mathbb{P},r_n^-}(x) f(x) dx = \int_{\mathcal{C}_\sigma} \int_{B(x, r_n^-)} f(y) f(x) dy dx \geq \frac{6}{25} \mathbb{P}(\mathcal{C}_\sigma) \lambda_\sigma (r_n^-)^d \nu_d \rightarrow \mathbb{P}(\mathcal{C}_\sigma) \lambda_\sigma^2 r^d \nu_d > 0$$

where the convergence happens as $n \rightarrow \infty$ by the stagnating property of $\{T_n\}_{n \in \mathbb{N}}$.

Then the desired result follows from an application of Slutsky's Theorem. \square

Lemma 27 (Continuity of stationary distribution). *Let $c > 0$ be any fixed constant, and $\{T_n\}_{n \in \mathbb{N}}$ be a sequence of stagnating transportation maps. Then,*

$$\liminf_{n \rightarrow \infty} \frac{\min \left\{ \tilde{\pi}_{\mathbb{P},r_n^-}(\mathcal{S}), \tilde{\pi}_{\mathbb{P},r_n^-}(\mathcal{S}^c) \right\}}{\min \left\{ \tilde{\pi}_{\mathbb{P},r_n^+}(\mathcal{S}), \tilde{\pi}_{\mathbb{P},r_n^+}(\mathcal{S}^c) \right\}} = 1$$

uniformly over all measurable sets $\mathcal{S} \subseteq \mathcal{C}_\sigma$ satisfying $\min \{ \tilde{\pi}_{\mathbb{P},r}(\mathcal{S}), \tilde{\pi}_{\mathbb{P},r}(\mathcal{S}^c) \} > c$.

Proof. It will be sufficient to show that

$$\liminf_{n \rightarrow \infty} \frac{\tilde{\pi}_{\mathbb{P},r_n^-}(\mathcal{S})}{\tilde{\pi}_{\mathbb{P},r_n^+}(\mathcal{S})} \quad \text{and} \quad \liminf_{n \rightarrow \infty} \frac{\tilde{\pi}_{\mathbb{P},r_n^-}(\mathcal{S}^c)}{\tilde{\pi}_{\mathbb{P},r_n^+}(\mathcal{S}^c)} = 1,$$

and we will show only the former, the proof of the latter being identical.

The proof proceeds similarly to Lemma 26. Letting

$$\mathcal{R}_n(x) := \{x' \in \mathcal{S} : x' \in B(x, r_n^+), x' \notin B(x, r_n^-)\}$$

we may rewrite

$$\frac{\tilde{\pi}_{\mathbb{P},r_n^-}(\mathcal{S})}{\tilde{\pi}_{\mathbb{P},r_n^+}(\mathcal{S})} = 1 - \frac{\int_{\mathcal{S}} \int_{\mathcal{R}_n(x)} f(y) f(x) dy dx}{\tilde{\pi}_{\mathbb{P},r_n^+}(\mathcal{S})}.$$

By the stagnating property of T_n ,

$$\int_{\mathcal{S}} \int_{\mathcal{R}_n(x)} f(y) f(x) dy dx \leq \mathbb{P}(\mathcal{S}) \Lambda_\sigma ((r_n^+)^d - (r_n^-)^d) \rightarrow 0$$

where the convergence occurs uniformly over all \mathcal{S} as $n \rightarrow \infty$. On the other hand, by hypothesis

$$\liminf_{n \rightarrow \infty} \tilde{\pi}_{\mathbb{P},r_n^+}(\mathcal{S}) \geq c > 0.$$

and the result follows by Slutsky's Theorem. \square

Lemma 28 (Stationary distribution lower bound). *Let $\{T_n\}_{n \in \mathbb{N}}$ be a sequence of stagnating transportation maps from \mathbb{P} to \mathbb{P}_n . Then with probability one,*

$$\liminf_{n \rightarrow \infty} \min_{S \in \mathcal{L}(\tilde{G}_{n,r})} \tilde{\pi}_{\mathbb{P},r}(T_n^{-1}(S)) \geq \frac{\lambda_\sigma^2 \nu_d r^d}{20\Lambda_\sigma}$$

Proof. Fix $\epsilon > 0$, and let $S \in \mathcal{L}(\tilde{G}_{n,r})$ be arbitrary, writing $\mathcal{S} := T_n^{-1}(S)$. Letting $\mathcal{R}_n(x)$ be as in the proof of Lemma 27, we have

$$\tilde{\pi}_{n,r}(S) \leq \tilde{\pi}_{\mathbb{P},r_n^-}(\mathcal{S}) + \frac{\int_{\mathcal{S}} \int_{\mathcal{R}_n(x)} f(y)f(x) dy dx}{\int_{\mathcal{C}_\sigma} \ell_{\mathbb{P},r_n^-}(x)f(x) dx}. \quad (46)$$

Clearly $\tilde{\pi}_{\mathbb{P},r_n^-}(\mathcal{S}) \leq \tilde{\pi}_{\mathbb{P},r}(\mathcal{S})$. Moreover

$$\frac{\int_{\mathcal{S}} \int_{\mathcal{R}_n(x)} f(y)f(x) dy dx}{\int_{\mathcal{C}_\sigma} \ell_{\mathbb{P},r_n^-}(x)f(x) dx} \leq \frac{\Lambda_\sigma}{\lambda_\sigma} \left(\frac{r_n^+ - r_n^-}{r_n^-} \right)^d$$

which along with (46) implies

$$s(\tilde{G}_{n,r}) \leq \tilde{\pi}_{n,r}(S) \leq \tilde{\pi}_{\mathbb{P},r}(\mathcal{S}) + \frac{\Lambda_\sigma}{\lambda_\sigma} \left(\frac{r_n^+ - r_n^-}{r_n^-} \right)^d \rightarrow \tilde{\pi}_{\mathbb{P},r}(\mathcal{S}),$$

where the first lower bound comes from the fact $S \in \mathcal{L}(\tilde{G}_{n,r})$, and the convergence is uniform over S as $n \rightarrow \infty$ by the stagnating property of $\{T_n\}_{n \in \mathbb{N}}$.

By Lemma 21, with probability one as $n \rightarrow \infty$,

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

and the claim is shown. \square

Proof of Lemma 19 By Lemma 22, with probability one there exists a sequence of stagnating transportation maps from $\tilde{\mathbb{P}}$ to $\tilde{\mathbb{P}}_n$, which we will denote $\{T_n\}_{n \in \mathbb{N}}$.

Let $S \subseteq \mathcal{C}_\sigma[X]$ be arbitrary, and define

$$\xi_n := \frac{\int_{\mathcal{C}_\sigma} \ell_{\mathbb{P},r_n^-}(x)f(x) dx}{\int_{\mathcal{C}_\sigma} \ell_{\mathbb{P},r_n^+}(x)f(x) dx}, \quad \gamma_n(S) := \frac{\min \left\{ \tilde{\pi}_{\mathbb{P},r_n^-}(T_n^{-1}(S^c)), \tilde{\pi}_{\mathbb{P},r_n^-}(T_n^{-1}(S)) \right\}}{\min \left\{ \tilde{\pi}_{\mathbb{P},r_n^+}(T_n^{-1}(S^c)), \tilde{\pi}_{\mathbb{P},r_n^+}(T_n^{-1}(S)) \right\}}$$

where $r_n^\pm := r \pm \|\text{Id} - T_n\|_{L^\infty(\tilde{\mathbb{P}})}$. By Proposition 3 and Lemma 25 we have that

$$\begin{aligned} \tilde{\Phi}_{n,r}(S) &\geq \xi_n \gamma_n(S) \tilde{\Phi}_{\mathbb{P},r_n^-}(T_n^{-1}(S)) \\ &\geq \xi_n \gamma_n(S) \frac{\lambda_\sigma^2 r_n^-}{2^{13} \Lambda_\sigma^2 \rho L \sqrt{d}}. \end{aligned} \quad (47)$$

By Lemma 26, with probability one

$$\liminf_{n \rightarrow \infty} \xi_n = 1.$$

By Lemma 28, letting c be any constant satisfying $0 < c < \frac{\lambda_\sigma^2 \nu_d r^d}{20\Lambda_\sigma}$, with probability one there exists some $m \in \mathbb{N}$ such that for all $n \geq m$,

$$\min_{S \in \mathcal{L}(\tilde{G}_{n,r})} \tilde{\pi}_{\mathbb{P},r}(T_n^{-1}(S)) \geq c > 0$$

and therefore by Lemma 27

$$\liminf_{n \rightarrow \infty} \left\{ \inf_{S \in \mathcal{L}(\tilde{G}_{n,r})} \gamma_n(S) \right\} = 1.$$

As Lemma 22 implies $r_n^- \rightarrow r$ with probability one, an application of Slutsky's Theorem to (47) completes the proof.

A.10 Proof of Theorem 4

Since r is fixed, $\widetilde{\deg}_{\min} \rightarrow \infty$ with n . Therefore for sufficiently large n we may apply Proposition 2 and obtain

$$\tau_{\infty}(\tilde{G}_{n,r}) \leq \frac{2}{\tilde{\Phi}_{n,r}^2} \log \left(\frac{1440}{s(\tilde{G}_{n,r})} \right) \log \left(\frac{14}{s(\tilde{G}_{n,r})} \right) + 3 \log \left(\frac{14}{s(\tilde{G}_{n,r})} \right) + 3 \quad (48)$$

and the claim follows from Proposition 4 and Lemma 21. In the statement of Theorem 4 we omit the factor of $3 \log \left(\frac{14}{s(\tilde{G}_{n,r})} \right)$ as the first and third terms in (48) dominate.

A.11 Concentration inequalities

Given a symmetric kernel function $K : \mathbb{R}^d \rightarrow \mathbb{R}$ with $\sup_{x,y \in \mathbb{R}^d} |K(x,y)| \leq 1$, and independent and identically distributed data $\{x_1, \dots, x_n\}$, we define the *order-2 U statistic* to be

$$U := \frac{1}{\binom{n}{2}} \sum_{1 \leq i < j \leq n} K(x_i, x_j).$$

Lemma 29 (Hoeffding's inequality for U -statistics.). *For any $t > 0$,*

$$\mathbb{P}(|U - \mathbb{E}U| \geq t) \leq 2 \exp \{-nt^2\}$$

Further, for any $\delta > 0$, we have

$$\begin{aligned} U &\leq \mathbb{E}U + \sqrt{\frac{\log(1/\delta)}{n}}, \\ U &\geq \mathbb{E}U - \sqrt{\frac{\log(1/\delta)}{n}} \end{aligned}$$

each with probability at least $1 - \delta$.

We will employ a sharper inequality to bound the sum of independent Bernoulli random variables.

Lemma 30. *Let $X_i \in \{0, 1\}$ for $i = 1, \dots, n$ and let $\mu = \sum_{i=1}^n \mathbb{E}(X_i)$. Then,*

$$\begin{aligned} \mathbb{P}\left(\sum_{i=1}^n X_i > (1 + \epsilon)\mu\right) &\leq \exp\left(\frac{-\epsilon^2 \mu}{3}\right) \\ \mathbb{P}\left(\sum_{i=1}^n X_i < (1 - \epsilon)\mu\right) &\leq \exp\left(\frac{-\epsilon^2 \mu}{2}\right) \end{aligned}$$

We will require large-sample and asymptotic bounds on several different degree and volume functionals, which follow from Lemmas 29 and 30. Let

$$\begin{aligned} \widetilde{\deg}_{\min} &:= \min_{u \in \mathcal{C}_{\sigma}[X]} \deg(u; G_{n,r}[\mathcal{C}_{\sigma}[X]]), \quad \widetilde{\deg}_{\max} := \max_{u \in \mathcal{C}_{\sigma}[X]} \deg(u; G_{n,r}[\mathcal{C}_{\sigma}[X]]) \\ \deg'_{\min} &:= \min_{u \in \mathcal{C}'_{\sigma}[X]} \deg(u; G_{n,r}), \quad \deg_{\max} := \max_{u \in \mathcal{C}_{\sigma}[X]} \deg(u; G_{n,r}) \\ \deg_{\min} &:= \min_{u \in X} \deg(u; G_{n,r}). \end{aligned}$$

Lemma 31. *Under the geometric conditions (A1)-(A5), for any $\epsilon > 0$, each of the following bounds hold with probability tending to one as $n \rightarrow \infty$:*

$$\frac{\widetilde{\deg}_{\min}}{n} \geq (1 - \epsilon) \frac{6}{25} \lambda_{\sigma} r^d \nu_d \quad (49)$$

$$\frac{\deg_{\max}}{n} \leq (1 + \epsilon) \Lambda_{\sigma} r^d \nu_d \quad (50)$$

$$\frac{|\mathcal{C}_{\sigma}[X]|}{n} \leq (1 + \epsilon) \mathbb{P}(\mathcal{C}_{\sigma}) \quad (51)$$

$$\frac{1}{2} \text{vol}_0 \leq \text{vol}_{n,r}(\mathcal{C}_{\sigma}[X]) \leq \frac{3}{2} \text{vol}_0 \quad (52)$$

Additionally, with probability one,

$$\liminf_{n \rightarrow \infty} \frac{\widetilde{\deg}_{\min}}{n} \geq \frac{6}{25} \lambda_{\sigma} r^d \nu_d \quad (53)$$

$$\limsup_{n \rightarrow \infty} \frac{\widetilde{\text{vol}}_{n,r}(\mathcal{C}_{\sigma}[X])}{n(n-1)} \leq \mathbb{P}(\mathcal{C}_{\sigma}) \Lambda_{\sigma} \nu_d r^d \quad (54)$$

If additionally (A6) holds: for any $\epsilon > 0$, the following bound holds with probability tending to one as $n \rightarrow \infty$:

$$\frac{\deg_{\min}}{n} \geq (1 - \epsilon) \frac{6}{25} \lambda_{\min} r^d \nu_d \quad (55)$$

If additionally (A7) holds (but not necessarily (A6)): for any $\epsilon > 0$, the following bound holds with probability tending to one as $n \rightarrow \infty$:

$$\frac{\deg'_{\min}}{n} \geq (1 - \epsilon) \frac{6}{25} \lambda_{\sigma} r^d \nu_d \quad (56)$$

Proof. We note that for any $x \in \mathcal{C}_{\sigma} \cup \mathcal{C}'_{\sigma}$, by (A1) and (A7) along with Lemma 5,

$$\frac{6}{25} \lambda_{\sigma} r^d \nu_d \leq \mathbb{P}(B(x, r))$$

Then, by Lemma 30 along with a union bound

$$\mathbb{P}\left(\widetilde{\deg}_{\min} \leq n(1 - \epsilon) \frac{6}{25} \lambda_{\sigma} r^d \nu_d\right) \leq n \exp\left\{-\frac{6n\epsilon^2 \lambda_{\sigma} r^d \nu_d}{50}\right\} \xrightarrow{n \rightarrow \infty} 0.$$

and we have shown (49). Under (A7), the same bound holds for \deg'_{\min} , and (56) follows. (55) and (50) follow from similar reasoning.

(51) follows directly from the law of large numbers.

Finally, note that

$$\text{vol}_{n,r}(\mathcal{C}_{\sigma}[X]) = \sum_{i=1}^n \sum_{j \neq i}^n \mathbf{1}(x_i \in \mathcal{C}_{\sigma}) \mathbf{1}(\|x_i - x_j\| \leq r)$$

can be rewritten as a U-statistic, and furthermore

$$\int_{\mathcal{C}_{\sigma}} \mathbb{P}(B(x, r)) dx = \mathbb{E}(\mathbf{1}(x_i \in \mathcal{C}_{\sigma}) \mathbf{1}(\|x_i - x_j\| \leq r)).$$

Therefore by Lemma 29

$$|\text{vol}_{n,r}(\mathcal{C}_{\sigma}[X]) - n(n-1) \int_{\mathcal{C}_{\sigma}} \mathbb{P}(B(x, r)) dx| \leq \frac{3}{20} n(n-1) \int_{\mathcal{C}_{\sigma}} \mathbb{P}(B(x, r)) dx \quad (57)$$

with probability at least $1 - 2 \exp(-\frac{3}{20} n \int_{\mathcal{C}_{\sigma}} \mathbb{P}(B(x, r)) dx) \rightarrow 0$ as $n \rightarrow \infty$. Then, application of the triangle inequality to (57), along with the range $\text{vol}_0 \in [3/4, 5/4] \cdot n(n-1) \int_{\mathcal{C}_{\sigma}} \mathbb{P}(B(x, r)) dx$ given by (8), yields

$$|\text{vol}_{n,r}(\mathcal{C}_{\sigma}[X]) - \text{vol}_0| \leq \left(\frac{3}{20} + \frac{1}{4}\right) n(n-1) \int_{\mathcal{C}_{\sigma}} \mathbb{P}(B(x, r)) dx \leq \frac{1}{2} \text{vol}_0.$$

By similar reasoning to (57), for any $\epsilon > 0$, letting $\mu = \int_{\mathcal{C}_{\sigma}} \mathbb{P}(B(x, r)) dx \geq \mathbb{E}\left(\frac{\widetilde{\text{vol}}_{n,r}(\mathcal{C}_{\sigma}[X])}{n(n-1)}\right)$ we have

$$\mathbb{P}\left(\frac{\widetilde{\text{vol}}_{n,r}(\mathcal{C}_{\sigma}[X])}{n(n-1)} \geq (1 + \epsilon)\mu\right) \leq \exp(-n\epsilon^2 \mu).$$

The two asymptotic statements (53) and (54) therefore both follow from the Borel-Cantelli Lemma. \square

Before we proceed to the next subsection, we include a formal statement of our claims regarding the aPPR vector.

Corollary 1. Fix $\lambda > 0$, and let $\mathcal{C} \in \mathbb{C}_f(\lambda)$ be a κ -well-conditioned cluster (with respect to some θ). Choose input parameters $\alpha, r, \text{vol}_0, v$ to be well-initialized in the sense of (8), set $\epsilon = \frac{1}{20\text{vol}_0}$, and modify Algorithm 1 to compute the aPPR vector $p^{(\epsilon)}$ rather than the exact PPR vector p , with resulting output \hat{C} .

1. If (A6) holds, then (10) is still a valid upper bound for the misclassification error of \hat{C} .
2. If (A7) and (12) hold, then $\hat{C} \subseteq X$ is a consistent estimator for \mathcal{C} , in the sense of Definition 2.

A.12 Proof of Theorem 1

In this section, we prove both Theorem 1 and the relevant portions of Corollary 1.

Lemma 32 is a restatement of Lemma 3.4 of Zhu et al. [2013], translated into notation consistent with the rest of this paper. It is stated with respect to an arbitrary $S \subseteq G = (V, E)$, and a good set $S^g \subseteq S$ with $\text{vol}(S^g; G) \geq \text{vol}(S; G)/2$.⁵

Lemma 32. Let $\epsilon \geq 0$, $v \in S^g$, $\alpha \leq \frac{1}{9\tau_\infty(G[S])}$, and $\text{vol}_0 \in [1 - c, 1 + c]\text{vol}(S; G)$ for $0 < c < 1$. Consider taking sweep cuts of the approximate PPR vector $p^{(\epsilon)} := p^{(\epsilon)}(v, \alpha; G)$ of the form

$$S_\beta := \left\{ u \in V : p^{(\epsilon)} \geq \beta \frac{\deg(u; G)}{\text{vol}_0} \right\}$$

For any $\beta < \frac{3}{5(1+c)}$, and $0 \leq \epsilon \leq \frac{1}{10}\text{vol}(S; G)$ the following bounds hold:

$$\text{vol}(S_\beta \setminus S; G) \leq \frac{2\Phi(A; G)}{\alpha\beta(1-c)}\text{vol}(S; G), \quad (58)$$

$$\text{vol}(S \setminus S_\beta; G) \leq \left(\frac{2\Phi(S; G)}{\alpha(\frac{3}{5} - \beta(1+c))} + 8\Phi(S; G) \right) \text{vol}(S; G). \quad (59)$$

Proof of Theorem 1. Observe that by Theorem 4 and the upper bounds $\alpha \leq \frac{1}{9\tau_u(\theta)}$, $r < \sigma/2d$ given in (8), with probability one as $n \rightarrow \infty$, $\alpha \leq \frac{1}{9\tau_\infty(G[S])}$. Moreover,

- (i) we restrict $v \in \mathcal{C}_\sigma[X]^g$
- (ii) the output of Algorithm 1 $\hat{C} = S_\beta$ for some $1/40 < \beta < 1/11$
- (iii) by Lemma 31, $\text{vol}_0 \in [1/2, 3/2]\text{vol}_{n,r}(\mathcal{C}_\sigma[X])$
- (iv) the PPR vector $p(v, \alpha; G_{n,r})$ is simply an approximate PPR vector with $\epsilon = 0$.

All conditions of Lemma 32 are therefore satisfied, and consequently by (58), (59), Theorem 3, and the lower bound $\alpha \geq 1/10\tau_u(\theta)$, the following bounds hold with probability tending to 1 as $n \rightarrow \infty$:

$$\text{vol}_{n,r}(\hat{C} \setminus \mathcal{C}_\sigma[X]) \leq 1600\kappa(\mathcal{C})\text{vol}_{n,r}(\mathcal{C}_\sigma[X]), \quad \text{and} \quad \text{vol}_{n,r}(\mathcal{C}_\sigma[X] \setminus \hat{C}) \leq 52\kappa(\mathcal{C})\text{vol}_{n,r}(\mathcal{C}_\sigma[X])$$

Noting that

$$\begin{aligned} |\hat{C} \setminus \mathcal{C}_\sigma[X]| &\leq \frac{\text{vol}_{n,r}(\hat{C} \setminus \mathcal{C}_\sigma[X])}{\deg_{\min}} \\ |\mathcal{C}_\sigma[X] \setminus \hat{C}| &\leq \frac{\text{vol}_{n,r}(\mathcal{C}_\sigma[X] \setminus \hat{C})}{\deg_{\min}}, \quad \text{and} \\ |\mathcal{C}_\sigma[X]| &\geq \frac{\text{vol}_{n,r}(\mathcal{C}_\sigma[X])}{\deg_{\max}} \end{aligned}$$

⁵The sweep cuts of Lemma 3.4 of Zhu et al. [2013] are stated directly with respect to $\text{vol}(S; G)$, rather than vol_0 ; this discrepancy leads to the extra factors of $1 - c$ and $1 + c$ which appear in our bounds as opposed to theirs.

the claim follows in light of (55), (49) and (50) from Lemma 31.

Proof of Corollary 1 misclassification rates. The choice $\epsilon = \frac{1}{20\text{vol}_0}$, along with Lemma 31 and the upper bound on vol_0 given by (8), together imply that $\epsilon < \frac{1}{10}\text{vol}_{n,r}(\mathcal{C}_\sigma[X])$ with probability tending to one as $n \rightarrow \infty$. We may therefore apply Lemma 32 with respect to $p^{(\epsilon)}$. All other aspects of the proof of Theorem 1 may be applied exactly as stated.

A.13 Proof of Theorem 2

We begin with some notation. Let D and W be the degree and lazy random walk matrices over $G_{n,r}$, and \tilde{D} and \tilde{W} be the degree and lazy random walk matrices for the induced subgraph $\tilde{G}_{n,r}$. Consider $\bar{\pi}_{n,r} : \mathcal{C}_\sigma[X] \rightarrow [0, 1]$ given by

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

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

Lemma 33 provides uniform error bounds on $p(x) - \bar{\pi}_{n,r}(x)$.

Lemma 33. *Let $0 < r < \sigma$ and $\alpha \leq \frac{1}{9\tau_\infty(\tilde{G}_{n,r})}$. Then the following statement holds: there exists a good set $\mathcal{C}_\sigma[X]^g \subseteq \mathcal{C}_\sigma[X]$ with $\text{vol}_{n,r}(\mathcal{C}_\sigma[X]^g) \geq \text{vol}_{n,r}(\mathcal{C}_\sigma[X])/2$ such that the following bounds hold with respect to $p := p(v, \alpha; G_{n,r})$ for any $v \in \mathcal{C}_\sigma[X]^g$:*

- For each $u \in \mathcal{C}[X]$,

$$p(u) \geq \frac{4}{5}\bar{\pi}_{n,r}(u) - \frac{20\Phi_{n,r}(\mathcal{C}_\sigma[X])\tau_\infty(\tilde{G}_{n,r})}{\widetilde{\deg}_{\min}} \quad (60)$$

- Let $\mathcal{C}' \neq \mathcal{C} \in \mathbb{C}_f(\lambda)$ satisfy the cluster separation condition (A2) with respect to \mathcal{C} . Then for each $u' \in \mathcal{C}'[X]$,

$$p(u') \leq \frac{20\Phi_{n,r}(\mathcal{C}_\sigma[X])\tau_\infty(\tilde{G}_{n,r})}{\deg'_{\min}}. \quad (61)$$

Proof. The proof of Lemma 33 is lengthy, but not difficult. Given starting distribution q with $\text{supp}(q) \subseteq \mathcal{C}_\sigma[X]$, we let

$$\tilde{p}_q = \alpha q + (1 - \alpha)\tilde{p}_q\tilde{W} \quad (62)$$

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

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

$$\begin{aligned} \ell_t &:= e_v(W\tilde{I})^t(I - D^{-1}\tilde{D}), \quad \ell := \sum_{t=0}^{\infty} (1 - \alpha)^t \ell_t, \\ s_t &:= e_v(W\tilde{I})^t(W\tilde{I}^c), \quad s := \sum_{t=0}^{\infty} (1 - \alpha)^t s_t. \end{aligned} \quad (63)$$

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

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 $\tilde{G}_{n,r}$ for t steps, arriving at u on the t th 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 (60), and begin by restating some results of [Zhu et al., 2013], adapted to our notation. By Lemma 3.1 of [Zhu et al., 2013], there exists a good set $\mathcal{C}_\sigma[X]^g \subseteq \mathcal{C}_\sigma[X]$ with $\text{vol}(\mathcal{C}_\sigma[X]^g; G_{n,r}) \geq \text{vol}(\mathcal{C}_\sigma[X]; G_{n,r})/2$ such that for every $v \in \mathcal{C}_\sigma[X]^g$

$$p(u) \geq \tilde{p}_v(u) - \tilde{p}_\ell(u), \quad \text{and} \quad \|\ell\|_1 \leq \frac{2\Phi_{n,r}(\mathcal{C}_\sigma[X])}{\alpha}. \quad (64)$$

(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 $\alpha \leq \frac{1}{9\tau_\infty(\tilde{G}_{n,r})}$, then by Corollary 3.3 of [Zhu et al., 2013], for every $u \in \mathcal{C}_\sigma[X]$

$$\tilde{p}(u) \geq \frac{4}{5}\bar{\pi}_{n,r}(u)$$

and along with (64), we obtain

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

We proceed to show the upper bound $\tilde{p}_\ell(u) \leq \|\ell\|_1 / \widetilde{\deg_{\min}}$, whence (60) follows by (64). We note two facts regarding $\tilde{p}_\ell(u)$, 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, $\tilde{\ell}(u) = 0$ as well.
2. For any q such that $\sum_{w \in \mathcal{C}_\sigma[X]} q(w) \leq 1$ and $u \notin \text{supp}(q)$, and any $t \geq 1$,

$$\begin{aligned} q\widetilde{W}^t(u) &\leq \|q\|_1 \max_{v \neq u} W_{vu} \\ &\leq \frac{1}{\widetilde{2\deg_{\min}}} \end{aligned} \quad (65)$$

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

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

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

and (60) is proved.

We turn to showing (61). By Lemma 34, for all $u' \notin \mathcal{C}_\sigma[X]$,

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

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

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

where the last inequality follows from precisely the same reasoning as (65). The claim follows in light of Lemma 36, along with (64). \square

We turn now to the proof of Theorem 2.

Proof of Theorem 2.

In the proof that follows, every inequality will hold with probability tending to one as $n \rightarrow \infty$.

Observe that by Theorem 4 and the upper bounds $\alpha \leq \frac{1}{9\tau_u(\theta)}$ and $r < \sigma/2d$ given in (8), with probability one as $n \rightarrow \infty$, $\alpha \leq \frac{1}{9\tau_u(\theta)}$. As a result, with probability one as $n \rightarrow \infty$, the conditions of Theorem 2 subsume the requirements of Lemma 33, and by that lemma, the following inequalities hold for each $u \in \mathcal{C}[X]$, $u' \in \mathcal{C}'[X]$:

$$\begin{aligned}
p(u) &\geq \frac{4}{5} \pi_{n,r}(u) - \frac{2\Phi_{n,r}(\mathcal{C}_\sigma[X])}{\alpha \widetilde{\deg}_{\min}} \\
p(u') &\leq \frac{2\Phi_{n,r}(\mathcal{C}_\sigma[X])}{\alpha \deg'_{\min}}
\end{aligned} \tag{66}$$

We will show that under the conditions of Theorem 2, the following bounds hold with probability tending to one as $n \rightarrow \infty$:

$$\begin{aligned}
\frac{2\widetilde{\Phi}_{n,r}(\mathcal{C}_\sigma[X])}{\alpha \widetilde{\deg}_{\min}} &\leq \frac{\widetilde{\deg}_{\min}}{125 \text{vol}_{n,r}(\mathcal{C}_\sigma X)} \\
\frac{2\widetilde{\Phi}_{n,r}(\mathcal{C}_\sigma[X])}{\alpha \deg'_{\min}} &\leq \frac{\deg'_{\min}}{125 \text{vol}_{n,r}(\mathcal{C}_\sigma[X])}
\end{aligned} \tag{67}$$

The bounds in (67) along with (66) imply that for each $u \in \mathcal{C}[X]$,

$$\begin{aligned}
\frac{p(u)}{\deg_{n,r}(u)} &\geq \frac{4}{5} \frac{\widetilde{\deg}_{n,r}(u)}{\deg_{n,r}(u) \text{vol}_{n,r}(\mathcal{C}_\sigma[X])} - \frac{\widetilde{\deg}_{\min}}{125 \deg_{n,r}(u) \text{vol}_{n,r}(\mathcal{C}_\sigma[X])} \\
&\geq \frac{22}{125 \text{vol}_{n,r}(\mathcal{C}_\sigma[X])} \quad (\text{by (55) and (50), applied with } \epsilon = \frac{1}{47}) \\
&\geq \frac{44}{375 \text{vol}_0}. \tag{by (52)}
\end{aligned}$$

Similarly, for each $u' \in \mathcal{C}'[X]$,

$$\begin{aligned}
\frac{p(u')}{\deg_{n,r}(u')} &\leq \frac{\deg'_{\min}}{\deg_{n,r}(u') 125 \text{vol}_{n,r}(\mathcal{C}_\sigma[X])} \\
&\leq \frac{1}{125 \text{vol}_{n,r}(\mathcal{C}_\sigma[X])} \\
&\leq \frac{8}{375 \text{vol}_0}. \tag{by (52)}
\end{aligned}$$

and as every sweep cut S_β under consideration in Algorithm 1 satisfies $8/375 < \frac{\beta}{\text{vol}_0} < 44/375$, (5) must hold for the output $\hat{C} = S_{\beta^*}$.

It remains to prove (67). First, apply Lemma 31 with $\epsilon = 1/11$ to obtain

$$\frac{\deg_{\min}'^2}{\text{vol}_{n,r}(\mathcal{C}_\sigma[X])}, \frac{\widetilde{\deg_{\min}}^2}{\text{vol}_{n,r}(\mathcal{C}_\sigma[X])} \geq \left(\frac{6}{25}\right)^2 \frac{\lambda_\sigma^2 \nu_d r^d}{\Lambda_\sigma \mathbb{P}(\mathcal{C}_\sigma)} \quad (68)$$

Then, by Theorem 3 and the lower bound $\alpha \geq \frac{1}{10\tau_u(\theta)}$ given by (8),

$$\frac{2\Phi_{n,r}(\mathcal{C}_\sigma[X])}{\alpha} \leq 20\kappa(\mathcal{C}). \quad (69)$$

(69), (68) and (12) together show (67).

Proof of Corollary 1: consistent cluster estimation. Using (13) with $\epsilon = \frac{1}{20}\text{vol}_0$ along with (60) we obtain

$$p^{(\epsilon)}(u) \geq \frac{4}{5}\bar{\pi}_{n,r}(u) - \frac{\deg_{n,r}(u)}{20\text{vol}_0} - \frac{2\Phi_{n,r}(\mathcal{C}_\sigma[X])}{\alpha \deg_{\min}}, \quad \text{for all } u \in \mathcal{C}[X]$$

and similar reasoning as used in the proof of Theorem 2 yields

$$\frac{p^{(\epsilon)}(u)}{\deg_{n,r}(u)} \geq \frac{15}{125\text{vol}_0} \geq \frac{1}{11\text{vol}_0}, \quad \text{for all } u \in \mathcal{C}[X].$$

On the other hand $p^{(\epsilon)}(u) \leq p(u)$. Therefore $\frac{p^{(\epsilon)}(u')}{\deg_{n,r}(u')} \leq \frac{8}{375\text{vol}_0}$, and the claim is proved.

A.14 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 34. 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 (63) and depends implicitly upon v .

Proof. The statement follows from Lemma 35 along with a series of algebraic manipulations,

$$\begin{aligned} p_v(u) &= \alpha \sum_{T=0}^{\infty} (1-\alpha)^T (e_v W^T)(u) \\ &= \alpha \sum_{T=1}^{\infty} (1-\alpha)^T (e_v W^T)(u) \\ &\leq \alpha \sum_{T=1}^{\infty} (1-\alpha)^T \left(\sum_{t=0}^{T-1} s_t W^{T-t-1} \right)(u) \quad (\text{Lemma 35}) \\ &= \alpha \sum_{t=0}^{\infty} \sum_{T=t+1}^{\infty} (1-\alpha)^T (s_t W^{T-t-1})(u) \\ &= \alpha \sum_{t=0}^{\infty} \sum_{\Delta=0}^{\infty} (1-\alpha)^{\Delta+t+1} (s_t W^\Delta)(u) \\ &\leq \alpha \sum_{t=0}^{\infty} \sum_{\Delta=0}^{\infty} (1-\alpha)^{\Delta+t} (s_t W^\Delta)(u) \\ &= \alpha \sum_{\Delta=0}^{\infty} (1-\alpha)^\Delta (s W^\Delta)(u) \\ &= p_s(u) \end{aligned}$$

□

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 35. *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 $S \subseteq V$*

$$qW^T = \sum_{t=0}^{T-1} s_t W^{T-t-1} + q(WI_S)^T \quad (70)$$

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

$$qW^T(u) = \sum_{t=0}^{T-1} (s_t W^{T-t-1})(u) \quad (71)$$

Proof. We show (70), from which (71) is an immediate consequence.

To show (70), we proceed by induction on T . When $T = 1$,

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

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

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

and the proof is complete. \square

Lemma 36. *Letting s_t, ℓ_t be as in (63),*

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

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

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

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

\square

B Experiments

In Section B, we detail the experimental settings of Section 4 in the main text, and include an additional figure.

B.1 Experimental settings for Figure 1

We sample points according to the density function q , where for $x \in \mathbb{R}^d$

$$q(x) := \begin{cases} \lambda, & x \in [0, \sigma] \times \rho^{d-1} =: \mathcal{C}, \\ \lambda - \text{dist}(x, \mathcal{C})\eta, & x \in \mathcal{C}_\sigma \setminus \mathcal{C}, \\ (\lambda - \sigma\eta) - \text{dist}(x, \mathcal{C}_\sigma)^\gamma, & x \in (\mathcal{C}_\sigma + \sigma B) \setminus \mathcal{C}_\sigma, \\ 0, & \text{otherwise,} \end{cases} \quad (72)$$

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 1 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 $\sigma = .1 \cdot \sqrt{2}^j, j = 1, \dots, 10$, and ρ is fixed at 3.2. In the bottom-middle and bottom-right panels, $\rho = .1 \cdot \sqrt{2}^j, j = 1, \dots, 10$ 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 $d = 3$.

B.2 Experimental settings for Figure 2

To form each of the three rows in Figure 2, 800 points are independently sampled following a 'two moons plus Gaussian noise model'. Formally, the (respective) generative models for the data are

$$Z \sim \text{Bern}(1/2), \theta \sim \text{Unif}(0, \pi) \quad (73)$$

$$X(Z, \theta) = \begin{cases} \mu_1 + (r \cos(\theta), r \sin(\theta)) + \sigma\epsilon, & \text{if } Z = 1 \\ \mu_2 + (r \cos(\theta), -r \sin(\theta)) + \sigma\epsilon, & \text{if } Z = 0 \end{cases} \quad (74)$$

where

$$\mu_1 = (-.5, 0), \mu_2 = (0, 0), \epsilon \sim N(0, I_2) \quad (\text{row 1})$$

$$\mu_1 = (-.5, -.07), \mu_2 = (0, .07), \epsilon \sim N(0, I_2) \quad (\text{row 2})$$

$$\mu_1 = (-.5, -.125), \mu_2 = (0, .125), \epsilon \sim N(0, I_2) \quad (\text{row 3})$$

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 Szlam and Bresson [2010]; and the last column shows a cut of the density cluster tree estimator of Chaudhuri and Dasgupta [2010].

B.3 Performance of PPR with high-dimensional noise.

Figure 3 is similar to Figure 2 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 2 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.

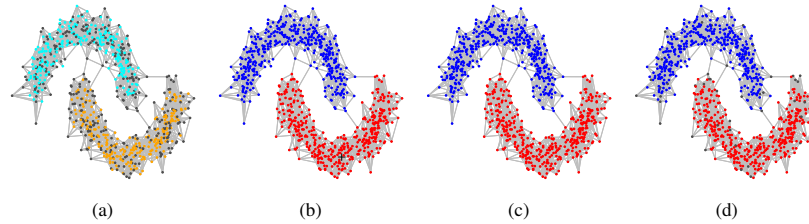


Figure 3

References

- Yasin Abbasi-Yadkori. Fast mixing random walks and regularity of incompressible vector fields. *arXiv preprint arXiv:1611.09252*, 2016.
- 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.
- 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. ISBN 978-1-60558-506-2. doi: 10.1145/1536414.1536449.
- 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*, pages 475–486, 2006.
- 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.
- Kamalika Chaudhuri and Sanjoy Dasgupta. Rates of convergence for the cluster tree. In J. D. Lafferty, C. K. I. Williams, J. Shawe-Taylor, R. S. Zemel, and A. Culotta, editors, *Advances in Neural Information Processing Systems 23*, pages 343–351. Curran Associates, Inc., 2010.
- 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.
- Shayan Oveis Gharan and Luca Trevisan. Approximating the expansion profile and almost optimal local graph clustering. In *Foundations of Computer Science (FOCS), 2012 IEEE 53rd Annual Symposium on*, pages 187–196. IEEE, 2012.
- 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.
- Stephen Guattery and Gary L Miller. On the performance of spectral graph partitioning methods. In *SODA*, volume 95, pages 233–242, 1995.
- John A. Hartigan. Consistency of single-linkage for high-density clusters. *Journal of the American Statistical Association*, 1981.
- 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.
- 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*, pages 847–855, 2010.
- Ravi Kannan, Santosh Vempala, and Adrian Vetta. On clusterings: Good, bad and spectral. *J. ACM*, 51(3):497–515, May 2004.

- Aleksandr P. Korostelev and Alexandre B. Tsybakov. *Minimax theory of image reconstruction*. Springer, 1993.
- 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.
- László Lovász and Miklós Simonovits. The mixing rate of markov chains, an isoperimetric inequality, and computing the volume. In *Proceedings [1990] 31st annual symposium on foundations of computer science*, pages 346–354. IEEE, 1990.
- 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.
- Ravi Montenegro. *Faster mixing by isoperimetric inequalities*. PhD thesis, Yale University, 2002.
- Wolfgang Polonik. Measuring mass concentrations and estimating density contour clusters-an excess mass approach. *The Annals of Statistics*, 23(3):855–881, 1995.
- Philippe Rigollet and Régis Vert. Optimal rates for plug-in estimators of density level sets. *Bernoulli*, 15(4):1154–1178, 2009.
- Geoffrey Schiebinger, Martin J. Wainwright, and Bin Yu. The geometry of kernelized spectral clustering. *Ann. Statist.*, 43(2):819–846, 04 2015.
- Tao Shi, Mikhail Belkin, and Bin Yu. Data spectroscopy: Eigenspaces of convolution operators and clustering. *Ann. Statist.*, 37(6B):3960–3984, 12 2009.
- Aarti Singh, Clayton Scott, and Robert Nowak. Adaptive hausdorff estimation of density level sets. *Ann. Statist.*, 37(5B):2760–2782, 10 2009.
- Daniel A Spielman and Shang-Hua Teng. Spectral sparsification of graphs. *SIAM Journal on Computing*, 40(4):981–1025, 2011.
- 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.
- 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.
- Arthur Szlam and Xavier Bresson. Total variation, cheeger cuts. In *ICML*, pages 1039–1046, 2010.
- 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, jan 2016.
- Alexandre B Tsybakov. On nonparametric estimation of density level sets. *The Annals of Statistics*, 25(3):948–969, 1997.
- Ulrike von Luxburg, Mikhail Belkin, and Olivier Bousquet. Consistency of spectral clustering. *Ann. Statist.*, 36(2):555–586, 04 2008.
- Xiao-Ming Wu, Zhenguo Li, Anthony M. So, John Wright, and Shih fu Chang. Learning with partially absorbing random walks. In F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 25*, pages 3077–3085. Curran Associates, Inc., 2012.
- Zeyuan Allen Zhu, Silvio Lattanzi, and Vahab S Mirrokni. A local algorithm for finding well-connected clusters. In *ICML (3)*, pages 396–404, 2013.