Statistical Guarantees for Local Spectral Clustering on Random Neighborhood Graphs

Alden Green AJGREEN@STAT.CMU.EDU
Department of Statistics and Data Science
Carnegie Mellon University
Pittsburgh, PA 15213

Sivaraman Balakrishnan SIVA@STAT.CMU.EDU
Department of Statistics and Data Science
Machine Learning Department
Carnegie Mellon University
Pittsburgh, PA 15213

Ryan J. Tibshirani RYANTIBS@STAT.CMU.EDU
Department of Statistics and Data Science
Machine Learning Department
Carnegie Mellon University
Pittsburgh, PA 15213

Abstract

We study the Personalized PageRank (PPR) algorithm, a local spectral method for clustering, which extracts clusters using locally-biased random walks around a given seed node. In contrast to previous work, we adopt a classical statistical learning setup, where we obtain samples from an unknown nonparametric distribution, and aim to identify sufficiently salient clusters. We introduce a trio of population-level functionals—the *normalized cut*, *conductance*, and *local spread*, analogous to graph-based functionals of the same name—and prove that PPR, run on a neighborhood graph, recovers clusters with small population normalized cut and large conductance and local spread. We apply our general theory to establish that PPR identifies connected regions of high density (density clusters) that satisfy a set of natural geometric conditions. We also show a converse result, that PPR can fail to recover geometrically poorly-conditioned density clusters, even asymptotically. Finally, we provide empirical support for our theory.

Keywords: graphs, spectral clustering, Personalized PageRank, density clustering, unsupervised learning

1. Introduction

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

When applied to geometric graphs built from a large number of samples, global spectral clustering methods can be computationally cumbersome and insensitive to the local geometry of the underlying distribution (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 the Personalized PageRank (PPR) algorithm, first introduced by Haveliwala (2003), and then further developed by several others (Spielman and Teng, 2011, 2014; Andersen et al., 2006; Mahoney et al., 2012; Allen-Zhu et al., 2013).

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), which has led many authors to develop supporting theory (Spielman and Teng, 2013; Andersen and Peres, 2009; Gharan and Trevisan, 2012; Allen-Zhu et al., 2013) that gives worst-case guarantees on traditional graph-theoretic notions of cluster quality (such as normalized cut and conductance). In contrast, in this paper we adopt a classical statistical viewpoint, and examine what the output of local clustering on a data set reveals about the underlying density f of the samples. We establish conditions on f under which PPR, when appropriately tuned and initialized inside a candidate cluster $\mathcal{C} \subseteq \mathbb{R}^d$, will approximately recover this candidate cluster. We pay special attention to the case where $\mathcal{C}$ is a *density cluster* of f—defined as a connected component of the upper level set $\{x \in \mathbb{R}^d : f(x) \geq \lambda\}$ for some $\lambda > 0$—and show precisely how PPR accounts for both geometry and density in estimating a cluster.

Before giving a more detailed overview of our main results, we formally define PPR on a neighborhood graph, review some of the aforementioned worst-case guarantees, and introduce the population-level functionals that govern the behavior of local clustering in our statistical context.

1.1 PPR clustering

We start by reviewing the PPR clustering algorithm. Let $G = (V, E)$ be an undirected, unweighted, and connected graph. We denote by $A \in \mathbb{R}^{n \times n}$ the adjacency matrix of G, with entries $A_{uv} = 1$ if $(u, v) \in E$ and 0 otherwise. We also denote by D the diagonal degree matrix, with entries $D_{uu} := \sum_{v \in V} A_{uv}$, and by I the $n \times n$ identity matrix. The PPR vector $p_v = p(v, \alpha; G)$ is defined with respect to a given seed node $v \in V$ and a teleportation parameter $\alpha \in [0, 1]$, as the solution of the following linear system:

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

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

In practice, exactly solving the system of equations (1) to compute the PPR vector may be too computationally expensive. To address this limitation, Andersen et al. (2006) introduced the ε-*approximate* PPR vector (aPPR), which we will denote by $p_v^{(\varepsilon)}$. 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 $\mathcal{O}(1/(\varepsilon\alpha))$ time, while satisfying the following uniform error bound:

$$p_v(u) - \varepsilon D_{uu} \leq p_v^{(\varepsilon)}(u) \leq p_v(u), \quad \text{for all } u \in V. \tag{2}$$

Once p_v or $p_v^{(\varepsilon)}$ is computed, the cluster estimate $\widehat{C}$ is chosen by taking a particular sweep cut. For a given level $\beta > 0$, the *β-sweep cut* of $p_v = (p_v(u))_{u \in V}$ is

$$S_{\beta,v} := \left\{ u \in V : \frac{p_v(u)}{D_{uu}} > \beta \right\}. \tag{3}$$

To define $\widehat{C}$, one computes $S_{\beta,v}$ over all $\beta \in (L, U)$ (where the range (L, U) is user-specified), and then outputs the cluster estimate $\widehat{C} = S_{\beta^*,v}$ with minimum normalized cut. For a set $C \subseteq V$ with complement $C^c = V \backslash C$, the *cut* and *volume* are respectively,

$$\mathrm{cut}(C; G) := \sum_{u \in C} \sum_{v \in C^c} \mathbf{1}\{(u, v) \in E\}, \quad \mathrm{vol}(C; G) := \sum_{u \in C} \sum_{v \in V} \mathbf{1}\{(u, v) \in E\}, \tag{4}$$

and the *normalized cut* of C is

$$\Phi(C; G) := \frac{\mathrm{cut}(C; G)}{\min\left\{\mathrm{vol}(C; G), \mathrm{vol}(C^c; G)\right\}}. \tag{5}$$

1.2 Worst-case guarantees for PPR clustering

As mentioned, most analyses of local clustering have focused on worst-case guarantees, defined with respect to functionals of an a priori fixed graph $G = (V, E)$. For instance, Andersen et al. (2006) analyze the normalized cut of the cluster estimate $\widehat{C}$ output by PPR, showing that when PPR is appropriately seeded within a candidate cluster $C \subseteq V$, the normalized cut $\Phi(\widehat{C}; G)$ is upper bounded by (a constant times) $\sqrt{\Phi(C; G)}$. Allen-Zhu et al. (2013) build on this: they introduce a second functional, the *conductance* $\Psi(G)$, defined as

$$\Psi(G) := \min_{S \subseteq V} \Phi(S; G), \tag{6}$$

and show that if $\Phi(C; G)$ is much smaller than $\Psi(G[C])^2$—where $G[C] = (C, E \cap (C \times C))$ is the subgraph of G induced by C— then (in addition to having a small normalized cut) the cluster estimate $\widehat{C}$ approximately recovers C. Our own analysis builds on that of Allen-Zhu et al. (2013), and we present a more detailed summary of their results in Section 2. For now, we merely reiterate that the conclusions of Andersen et al. (2006); Allen-Zhu et al. (2013) cannot be straightforwardly applied to our setting, where the input data are random samples $\{x_1, \ldots, x_n\}$ drawn from a distribution $\mathbb{P}$, the graph G is a random neighborhood graph formed from the samples, and the candidate cluster is a set $\mathcal{C} \subseteq \mathbb{R}^d$.[1]

1.3 PPR on a neighborhood graph

We now formally describe the statistical setting in which we operate, as well as the method we will study: PPR on a neighborhood graph. Let $X = \{x_1, \ldots, x_n\}$ be samples drawn i.i.d. from a distribution $\mathbb{P}$ on $\mathbb{R}^d$. We will assume throughout that $\mathbb{P}$ has a density f with respect to the Lebesgue measure ν on $\mathbb{R}^d$. 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 $i \neq j$ and $\|x_i - x_j\| \leq r$, where $\|\cdot\|$ is the Euclidean norm. Once the

---

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 $v \in X$, sweep cut range (L, U).

Output: cluster estimate $\widehat{C} \subseteq V$.

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

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

---

neighborhood graph $G_{n,r}$ is formed, the PPR vector p_v is then computed over $G_{n,r}$, with a resulting cluster estimate $\widehat{C} \subseteq X$. The precise algorithm is summarized in Algorithm 1.

1.4 Cluster accuracy

We need a metric to assess the accuracy with which $\widehat{C}$ estimates the candidate cluster $\mathcal{C}$. One commonly used metric is the misclassification error, i.e., the size of the symmetric set difference between $\widehat{C}$ and the empirical cluster $\mathcal{C}[X] = \mathcal{C} \cap X$ (Korostelev and Tsybakov, 1993; Polonik, 1995; Rigollet and Vert, 2009). We will consider a related metric, the volume of the symmetric set difference, which weights misclassified points according to their degree in $G_{n,r}$. To keep things simple, for a given set $S \subseteq X$ we write $\mathrm{vol}_{n,r}(S) := \mathrm{vol}(S; G_{n,r})$.

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

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

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

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

1.5 Population normalized cut, conductance, and local spread

Next we define three population-level functionals of $\mathcal{C}$—the normalized cut $\Phi_{\mathbb{P},r}(\mathcal{C})$, conductance $\Psi_{\mathbb{P},r}(\mathcal{C})$, and the local spread $s_{\mathbb{P},r}(\mathcal{C})$—which we will use to upper bound the volume of the symmetric set difference $\Delta(\widehat{C}, \mathcal{C}[X])$ with high probability.

Let the population-level *cut* of $\mathcal{C}$ be the expectation (up to a rescaling) of $\mathrm{cut}_{n,r}(\mathcal{C}[X]) := \mathrm{cut}(\mathcal{C}[X]; G_{n,r})$, and likewise let the population-level *volume* of $\mathcal{C}$ be the expectation (up to a rescaling) of $\mathrm{vol}_{n,r}(\mathcal{C}[X]) := \mathrm{vol}(\mathcal{C}[X]; G_{n,r})$; i.e., let

$$\mathrm{cut}_{\mathbb{P},r}(\mathcal{C}) := \int_{\mathcal{C}} \int_{\mathcal{C}^c} \mathbf{1}\{\|x - y\| \le r\} \, d\mathbb{P}(y) \, d\mathbb{P}(x), \;\; \mathrm{vol}_{\mathbb{P},r}(\mathcal{C}) := \int_{\mathcal{C}} \int_{\mathbb{R}^d} \mathbf{1}\{\|x - y\| \le r\} \, d\mathbb{P}(y) \, d\mathbb{P}(x),$$

1. Throughout, we use calligraphic notation to refer to subsets of $\mathbb{R}^d$.
2. Technically speaking, for each $\beta \in (L, U) \cap \{p_v(u)/D_{uu} : u \in V\}$.

where $\mathcal{C}^c := \mathbb{R}^d \backslash \mathcal{C}$. Also let $\deg_{\mathbb{P},r}(x) := \int_{\mathbb{R}^d} \mathbf{1}\{\|y - x\| \leq r\}\, d\mathbb{P}(y)$ be the expected degree of x in $G_{n,r}$.

Definition 2 (Population normalized cut) *For a set $\mathcal{C} \subseteq \mathbb{R}^d$, distribution $\mathbb{P}$, and radius $r > 0$, the* population normalized cut *is*

$$\Phi_{\mathbb{P},r}(\mathcal{C}) := \frac{\mathrm{cut}_{\mathbb{P},r}(\mathcal{C})}{\min\{\mathrm{vol}_{\mathbb{P},r}(\mathcal{C}), \mathrm{vol}_{\mathbb{P},r}(\mathcal{C}^c)\}}. \tag{7}$$

Let $\widetilde{\mathbb{P}}(\cdot) = \mathbb{P}(\cdot | x \in \mathcal{C})$ be the conditional distribution of x, i.e., let $\widetilde{\mathbb{P}}(\mathcal{S}) = \widetilde{\mathbb{P}}(\mathcal{S} \cap \mathcal{C})/\widetilde{\mathbb{P}}(\mathcal{C})$ for measurable sets $\mathcal{S} \subseteq \mathbb{R}^d$.

Definition 3 (Population conductance) *For a set $\mathcal{C} \subseteq \mathbb{R}^d$, distribution $\mathbb{P}$ and radius $r > 0$, the* population conductance *is*

$$\Psi_{\mathbb{P},r}(\mathcal{C}) = \inf_{\mathcal{S} \subseteq \mathcal{C}} \Phi_{\widetilde{\mathbb{P}},r}(\mathcal{S}). \tag{8}$$

Definition 4 (Population local spread) *For a set $\mathcal{C} \subseteq \mathbb{R}^d$, distribution $\mathbb{P}$ and radius $r > 0$, the* population local spread *is*

$$s_{\mathbb{P},r}(\mathcal{C}) := \min_{x \in \mathcal{C}} \left\{ \frac{\left(\deg_{\widetilde{\mathbb{P}},r}(x)\right)^2}{\mathrm{vol}_{\widetilde{\mathbb{P}},r}(\mathcal{C})} \right\}, \tag{9}$$

It is quite natural that $\Phi_{\mathbb{P},r}(\mathcal{C})$ and $\Psi_{\mathbb{P},r}(\mathcal{C})$ should help quantify the role geometry plays in local spectral clustering. Indeed, these functionals are the population-level analogues of the empirical quantities $\Phi_{n,r}(\mathcal{C}[X]) := \Phi(\mathcal{C}[X]; G_{n,r})$ and $\Psi_{n,r}(\mathcal{C}[X]) := \Psi(G_{n,r}\big[\mathcal{C}[X]\big])$, and as we have already mentioned, these empirical quantities can be used to upper bound the volume of the symmetric set difference. For this reason, similar population-level functionals are used by Shi et al. (2009); Schiebinger et al. (2015); García Trillos et al. (2019) in the analysis of *global* spectral clustering in a statistical context. We will comment more on the relationship between these works and our own results in Section 1.7.

The role played by $s_{\mathbb{P},r}(\mathcal{C})$ is somewhat less obvious. For now, we mention only that it plays an essential part in obtaining tight bounds on the mixing time of a particular random walk that is closely related to the PPR vector, and defer further discussion until later in Section 2.

1.6 Main results

We now informally state our two main upper bounds, regarding the recovery of a generic cluster $\mathcal{C}$, and a density cluster $\mathcal{C}_\lambda$. Theorem 5 informally summarizes the first of our main results (formally stated in Theorem 12) regarding the recovery of a generic cluster $\mathcal{C}$.

Theorem 5 (Informal) *If $\mathcal{C} \subseteq \mathbb{R}^d$ and $\mathbb{P}$ satisfy appropriate regularity conditions, and Algorithm 1 is initialized properly with respect to $\mathcal{C}$, then for all sufficiently large n, with high probability it holds that*

$$\frac{\Delta(\widehat{C}, \mathcal{C}[X])}{\mathrm{vol}_{n,r}(\mathcal{C}[X])} \leq c \cdot \Phi_{\mathbb{P},r}(\mathcal{C}) \cdot \left(\frac{\log(1/s_{\mathbb{P},r}(\mathcal{C}))}{\Psi_{\mathbb{P},r}(\mathcal{C})}\right)^2.$$

(Above, and throughout, c stands for a universal constant that may change from line to line.) Put more succinctly, we find that $\Delta(\widehat{C}, \mathcal{C}[X])$ is small when $\Phi_{\mathbb{P},r}(\mathcal{C})$ is small relative to $(\Psi_{\mathbb{P},r}(\mathcal{C})/\log(1/s_{\mathbb{P},r}(\mathcal{C})))^2$. To the best of our knowledge, this gives the first population-level guarantees for local clustering in the nonparametric statistical context.

Next, Theorem 6 informally summarizes the second of our main results (formally stated in Theorem 16) regarding the recovery of a λ-density cluster $\mathcal{C}_\lambda$ by PPR. For reasons that we explain later in Section 3, our cluster recovery statement will actually be with respect to the σ-thickened set $\mathcal{C}_{\lambda,\sigma} := \{x \in \mathbb{R}^d : \mathrm{dist}(x, \mathcal{C}_\lambda) < \sigma\}$, for a given $\sigma > 0$. The upper bound we establish is a function of various parameters that measure the conditioning of both the density cluster $\mathcal{C}_{\lambda,\sigma}$ and density f for recovery by PPR. We assume that $\mathcal{C}_{\lambda,\sigma}$ is the image of a convex set $\mathcal{K}$ of finite diameter $\mathrm{diam}(\mathcal{K}) \leq \rho < \infty$ under a Lipschitz, measure-preserving mapping g, with Lipschitz constant M. We also assume that f is bounded away from 0 and ∞ on $\mathcal{C}_{\lambda,\sigma}$:

$$0 < \lambda_\sigma \leq f(x) \leq \Lambda_\sigma < \infty \ \text{ for all } x \in \mathcal{C}_{\lambda,\sigma},$$

and additionally satisfies the following low-noise condition:

$$\inf_{y \in \mathcal{C}_{\lambda,\sigma}} f(y) - f(x) \geq \theta \cdot \mathrm{dist}(x, \mathcal{C}_{\lambda,\sigma})^\gamma \ \text{ for all } x \text{ such that } 0 < \mathrm{dist}(x, \mathcal{C}_{\lambda,\sigma}) \leq \sigma.$$

(Here $\mathrm{dist}(x, \mathcal{C}) := \inf_{y \in \mathcal{C}} \|y - x\|$.)

Theorem 6 (Informal) *If $\mathcal{C}_\lambda \subseteq \mathbb{R}^d$ is a λ-density cluster of a distribution $\mathbb{P}$, which satisfies appropriate regularity conditions, and Algorithm 1 is initialized properly with respect to $\mathcal{C}_{\lambda,\sigma}$, then for all sufficiently large n, with high probability it holds that*

$$\frac{\Delta(\widehat{C}, \mathcal{C}_{\lambda,\sigma}[X])}{\mathrm{vol}_{n,r}(\mathcal{C}_{\lambda,\sigma})} \leq c \cdot d^4 \cdot \frac{M^2\rho^2}{\sigma r} \cdot \frac{\Lambda_\sigma^2 \lambda(\lambda - \theta\frac{r^\gamma}{\gamma+1})}{\lambda_\sigma^4} \cdot \log^2\left(\frac{\Lambda_\sigma^{2/d} M\rho}{\lambda_\sigma^{2/d} 2r}\right).$$

The above result reveals the separate roles played by geometry and density in the ability of PPR to recover a density cluster. Here M, ρ, and σ capture whether $\mathcal{C}_{\lambda,\sigma}$ is geometrically well-conditioned (short and fat) or poorly-conditioned (long and thin) for recovery by PPR. Likewise, the parameters $\lambda_\sigma, \Lambda_\sigma, \gamma$, and θ measure whether the density f is well-conditioned (approximately uniform over the density cluster, and having thin tails outside of it) or poorly conditioned (vice versa). Theorem 6 says that if the thickened density cluster $\mathcal{C}_{\lambda,\sigma}$ is geometrically well-conditioned—meaning, $M^2\rho^2/(\sigma r) \approx 1$—and the density f is well-conditioned near $\mathcal{C}_{\lambda,\sigma}$—meaning, $\Lambda_\sigma \approx \lambda \approx \lambda_\sigma$ and $\lambda - \theta r^\gamma/(\gamma+1)$ is much less than λ_σ—then PPR will approximately recover $\mathcal{C}_{\lambda,\sigma}$.

1.7 Related work

We now summarize some related work (in addition to the background material already given above), regarding the theory of spectral clustering, and of density cluster recovery.

Spectral clustering. In the stochastic block model (SBM), arguably one of the simplest models of network formation, edges between nodes independently occur with probability based on a latent community membership. In the SBM, the ability of spectral algorithms to perform clustering—or community detection—is well-understood, dating back to McSherry (2001) who gives conditions under which the entire community structure can be recovered. In more recent work, Rohe et al. (2011) upper bound the fraction of nodes misclassified by a spectral algorithm for the high-dimensional (large number of blocks) SBM, and Lei and Rinaldo (2015) extend these results to the sparse (low average degree) regime. Relatedly, Clauset et al. (2008); Balakrishnan et al. (2011); Li et al. (2018) analyze the misclassification rate when the block model exhibits some hierarchical structure. The framework we consider, in which nodes correspond to data points sampled from an underlying density, and edges between nodes are formed based on geometric proximity, is quite different than the SBM, and therefore so is our analysis.

In general, the study of spectral algorithms on neighborhood graphs has been focused on establishing asymptotic convergence of eigenvalues and eigenvectors of certain sample objects to the eigenvalues and eigenfunctions of corresponding limiting operators. Koltchinskii and Gine (2000) establish convergence of spectral projections of the adjacency matrix to a limiting integral operator, with similar results obtained using simplified proofs in Rosasco et al. (2010). von Luxburg et al. (2008) studies convergence of eigenvectors of the Laplacian matrix for a neighborhood graph of fixed radius. Belkin and Niyogi (2007) and García Trillos and Slepčev (2018) extend these results to the regime where the radius $r \to 0$ as $n \to \infty$.

These results are of fundamental importance. However, they remain silent on the following natural question: do the spectra of these continuum operators induce a partition of the sample space which is "good" in some sense? Shi et al. (2009); Schiebinger et al. (2015); García Trillos et al. (2019); Hoffmann et al. (2019) address this question, showing that spectral algorithms will recover the latent labels in certain well-conditioned nonparametric mixture models. These works are probably the most similar to our own: the conditioning of these mixture models depends on population-level functionals resembling the population normalized cut and conductance introduced above, and the resulting bounds on the error of spectral clustering are comparable to those we establish in Theorem 12. However, these results focus on global rather than local methods, and impose global rather than local conditions on $\mathbb{P}$. Moreover, they do not explicitly consider recovery of density clusters, which is an important concern of our work. We comment further on the relationship between our results and these works after Theorem 12.

Density clustering. For a given threshold $\lambda \in (0, \infty)$, we denote by $\mathbb{C}_f(\lambda)$ the connected components of the density upper level set $\{x \in \mathbb{R}^d : f(x) \geq \lambda\}$. In the density clustering problem, initiated by Hartigan (1975), the goal is to recover $\mathbb{C}_f(\lambda)$. By now, density clustering, and the related problem of level-set estimation, have been thoroughly studied. For instance, Polonik (1995); Rigollet and Vert (2009); Rinaldo and Wasserman (2010); Steinwart (2015) study density clustering under the symmetric set difference metric, Tsybakov (1997); Singh et al. (2009); Jiang (2017) describe minimax optimal level-set and cluster estimators under Hausdorff loss, and Hartigan (1981); Chaudhuri and Dasgupta (2010); Kpotufe and von Luxburg (2011); Balakrishnan et al. (2013); Steinwart et al. (2017); Wang et al. (2019) consider estimation of the cluster tree $\{\mathbb{C}_f(\lambda) : \lambda \in (0, \infty)\}$.

We emphasize that our goal is not to improve on these results, nor is it to offer a better algorithm for density clustering. Indeed, seen as a density clustering algorithm, PPR has none of the optimality guarantees found in the aforementioned works. Rather, we hope to better understand the implications of our general theory by applying it within an already well-studied framework. We should also note that since we study a local algorithm, our interest will be in a local version of the density clustering problem, where the goal is to recover a single density cluster $\mathcal{C}_\lambda \in \mathbb{C}_f(\lambda)$.

1.8 Organization

We now outline the rest of the paper.

- In Section 2, we derive bounds on the error of PPR as a function of sample normalized cut, conductance, and local spread. We then show that under certain conditions the sample normalized cut, conductance, and local spread are close to their population-level counterparts, with high probability for sufficient number of samples. As a result, we obtain an upper bound on $\Delta(\widehat{C}, \mathcal{C}[X])/\mathrm{vol}_{n,r}(\mathcal{C}[X])$ purely in terms of these population-level functionals (Theorem 16).

- In Section 3, we focus on the special case where the candidate cluster $\mathcal{C} = \mathcal{C}_\lambda$ is a λ-*density cluster*—that is, a connected component of the upper level set $\{x : f(x) \geq \lambda\}$. We derive bounds on the population normalized cut, conductance, and local spread of the density cluster, which depend on λ as well as some other natural parameters. This leads to a bound on the symmetric set difference between $\widehat{C}$ and the λ-density cluster (Theorem 16).

- In Section 4, we prove a negative result: we give a hard distribution $\mathbb{P}$ with corresponding density cluster $\mathcal{C}_\lambda$ for which the symmetric set difference between $\widehat{C}$ and the λ-density cluster is provably large.

- In Section 5 we empirically investigate some of our conclusions, before ending with some discussion in Section 6.

2. Recovery of a generic cluster with PPR

In the main result (Theorem 12) of this section, we give a high probability upper bound on the volume of the symmetric set difference $\Delta(\widehat{C}, \mathcal{C}[X])$, in terms of the population normalized cut $\Phi_{\mathbb{P},r}(\mathcal{C})$, conductance $\Psi_{\mathbb{P},r}(\mathcal{C})$, and local spread $s_{\mathbb{P},r}(\mathcal{C})$. We build to this theorem slowly, giving new structural results in two distinct directions. First, we build on some previous work (mentioned in the introduction) to relate $\Delta(\widehat{C}, \mathcal{C}[X])$ to the sample normalized cut $\Phi_{n,r}(\mathcal{C}[X])$, conductance $\Psi_{n,r}(\mathcal{C}[X])$, and local spread $s_{n,r}(\mathcal{C}[X]) := s(G_{n,r}[\mathcal{C}[X]])$. Second, we argue that when n is large, each of these graph functionals can be bounded by their population-level analogues with high probability.

2.1 The fixed graph case

When PPR is run on a fixed graph $G = (V, E)$ with the goal of recovering a candidate cluster $C \subseteq V$, Allen-Zhu et al. (2013) provide the sharpest known bounds on the volume of the symmetric set difference between the cluster estimate $\widehat{C}$ and candidate cluster C. Since these results will play a major part in our analysis, in Lemma 7 we restate them for the convenience of the reader.[3]

In their most general form, the results of Allen-Zhu et al. (2013) depend on the mixing time of a lazy random walk over the induced subgraph $G[C]$. The *mixing time* of a lazy random walk over a graph G is

$$\tau_{\infty}(G) := \min\left\{t : \frac{\pi(u) - q_v^{(t)}(u)}{\pi(u)} \leq \frac{1}{4}, \text{ for all } u, v \in V\right\}; \tag{10}$$

here $q_v^{(t)} := e_v W^t$ is the distribution of a lazy random walk over G initialized at node v and run for t steps, and $\pi := \lim_{t\to\infty} q_v^{(t)}$ is the limiting distribution of $q_v^{(t)}$.

Lemma 7 (Lemma 3.4 of Allen-Zhu et al. (2013)) *For a set $C \subseteq V$, suppose that*

$$\alpha \leq \min\Big\{\frac{1}{45}, \frac{1}{2\tau_{\infty}(G[C])}\Big\}, \quad \beta \leq \frac{1}{5\mathrm{vol}(C;G)}. \tag{11}$$

Then there exists a set $C^g \subseteq C$ with $\mathrm{vol}(C^g;G) \geq \frac{1}{2}\mathrm{vol}(C;G)$ such that for any $v \in C^g$, the sweep cut $S_{\beta,v}$ satisfies

$$\mathrm{vol}(S_{\beta,v} \vartriangle C;G) \leq 6\frac{\Phi(C;G)}{\alpha\beta}. \tag{12}$$

The upper bound in (12) does not obviously depend on the conductance $\Psi(G[C])$. However, as Allen-Zhu et al. (2013) point out, letting $\pi_{\min}(G) := \min_{u\in V}\{\pi(u)\}$, it follows from Cheeger's inequality (Chung, 1997) that

$$\tau_{\infty}(G) \leq \frac{\log(1/\pi_{\min}(G))}{\Psi(G)^2}. \tag{13}$$

Therefore, setting (for instance) $\alpha = \frac{\Psi(G[C])^2}{2\log(1/\pi_{\min}(G))}$ and $\widehat{C} = S_{\beta_0,v}$ for $\beta_0 = \frac{1}{5\mathrm{vol}(C;G)}$, we obtain from (12) that

$$\frac{\mathrm{vol}(C \vartriangle \widehat{C};G)}{\mathrm{vol}(C;G)} \leq 60\frac{\Phi(C;G)\log\big(1/\pi_{\min}(G[C])\big)}{\Psi(G[C])^2}. \tag{14}$$

2.2 Improved bounds on mixing time

Having reviewed the conclusions of Allen-Zhu et al. (2013), we return now to our own setting, where the data is not a fixed graph G but instead random samples $\{x_1, \ldots, x_n\}$, and our goal is to recover a candidate cluster $\mathcal{C} \subseteq \mathbb{R}^d$. Ideally, we would like to apply (14) with $C = \mathcal{C}[X]$ and $G = G_{n,r}$, replace $\Phi_{n,r}(\mathcal{C}[X])$ and $\Psi_{n,r}(\mathcal{C}[X])$ by $\Phi_{\mathbb{P},r}(\mathcal{C})$ and $\Psi_{\mathbb{P},r}(\mathcal{C})$ inside (14), and thereby obtain an upper bound on $\Delta(\widehat{C};\mathcal{C}[X])$ that depends only on $\mathbb{P}$ and $\mathcal{C}$. Unfortunately, however, there is a catch: when the graph is $G = G_{n,r}$ and the candidate cluster is $C = \mathcal{C}[X]$, as $n \to \infty$ the sample normalized cut $\Phi_{n,r}(\mathcal{C}[X])$ and conductance $\Psi_{n,r}(\mathcal{C}[X])$ each converge

3. Lemma 7 improves on Lemma 3.4 of Allen-Zhu et al. (2013) by some constant factors, and for completeness we prove Lemma 7 in the Appendix. Nevertheless, to be clear the essential idea of Lemma 7 is no different than that of Allen-Zhu et al. (2013), and we do not claim any novelty.

to their population-level analogues, but $\pi_{\min}(G_{n,r}[\mathcal{C}[X]]) \asymp 1/n$.[4] Therefore the right hand side of (14) diverges at a $\log n$ rate, which turns (14) into a vacuous upper bound whenever the number of samples is sufficiently large.

To address this, in Proposition 8 we improve the upper bound on the mixing time in (13). Specifically, in (15) the "start penalty" of $\log(1/\pi_{\min}(G))$ is replaced by $\log(1/s(G))$, where $s(G)$ is the graph local spread, defined as

$$s(G) := d_{\min}(G) \cdot \pi_{\min}(G),$$

for $d_{\min}(G) = \min_{u \in V}\{\deg(u;G)\}$, and likewise $d_{\max}(G) = \max_{u \in V}\{\deg(u;G)\}$. Notice that $s(G) \geq \pi_{\min}(G)$.

Proposition 8 *Assume* $d_{\max}(G)/d_{\min}(G)^2 \leq 1/16$. *Then,*

$$\tau_{\infty}(G) \leq \frac{17}{\ln(2)}\left(\frac{\ln\bigl(32/s(G)\bigr)}{\Psi(G)}\right)^2. \tag{15}$$

While Proposition 8 can be applied to *any* graph G (provided that the ratio of maximum degree to squared minimum degree is at most $1/16$), it is particularly useful for geometric graphs: when $G = G_{n,r}[\mathcal{C}[X]]$ for a fixed radius $r > 0$, we have $d_{\min}(G_{n,r}[\mathcal{C}[X]]) \asymp n$, and thus $s(G_{n,r}[\mathcal{C}[X]]) \asymp 1$. We give a precise upper bound on $s(G_{n,r}[\mathcal{C}[X]])$ in Proposition 9, which does not grow with n, and in combination with Proposition 8 this allows us to remove the unwanted $\log n$ factor from the upper bound in (14).

The local spread $s(G)$ plays an intuitive role in the analysis of mixing time. Indeed, in any graph G sufficiently small sets are expanders—that is, if a set $R \subseteq V$ has cardinality less than the minimum degree, the normalized cut $\Phi(R;G)$ will be much larger than the conductance $\Psi(G)$. As a consequence, a random walk over G will rapidly mix over all small sets R, and in our analysis of the mixing time we may therefore "pretend" that the random walk was given a warm start over a larger set S. The local spread $s(G)$ simply delineates small sets R from larger sets S. Of course, the proof of Proposition 8 requires a much more intricate analysis, and—as with the proofs of all results in this paper—it is deferred to the appendix.

2.3 Sample-to-population results

In Propositions 9 and 10, we establish high probability bounds on the sample normalized cut, conductance, and local spread in terms of their population-level analogues. To establish these bounds, we impose the following regularity conditions on $\widetilde{\mathbb{P}}$ and $\mathcal{C}$.

(A1) The distribution $\widetilde{\mathbb{P}}$ has a density $\widetilde{f} : \mathcal{C} \to (0,\infty)$ with respect to Lebesgue measure. There exist $0 < f_{\min} \leq f_{\max} < \infty$ for which

$$(\forall x \in \mathcal{C}) \quad f_{\min} \leq \widetilde{f}(x) \leq f_{\max}.$$

For convenience, we will assume $f_{\min} \leq 1$ and $f_{\max} \geq 1$.

4. For sequences (a_n) and (b_n), we say $a_n \asymp b_n$ if there exists a constant $c \geq 1$ such that $a_n/c \leq b_n \leq c a_n$ for all $n \in \mathbb{N}$.

(A2) The candidate cluster $\mathcal{C} \subseteq \mathbb{R}^d$ is a bounded, connected, open set, and for $d \geq 2$, it has a Lipschitz boundary $\partial\mathcal{C}$, meaning it is locally the graph of a Lipschitz function (e.g., see Definition 9.57 of Leoni (2017)).

In what follows, we use $b_1, b_2, \ldots$ and $B_1, B_2, \ldots$ to refer to positive constants that may depend on $\mathbb{P}$, $\mathcal{C}$, and r, but do not depend on n or δ. We explicitly keep track of all constants in our proofs.

Proposition 9 *Fix $\delta \in (0, 1/3)$. Suppose $\widetilde{\mathbb{P}}$ and $\mathcal{C}$ satisfy (A1) and (A2). Then each of the following statements hold.*

- *With probability at least $1 - 3\exp\{-b_1\delta^2 n\}$,*

$$\Phi_{n,r}(\mathcal{C}[X]) \leq (1 + 3\delta)\Phi_{\mathbb{P},r}(\mathcal{C}). \tag{16}$$

- *For any $n \in \mathbb{N}$ for which*

$$\frac{1}{n} \leq \delta \cdot \frac{2\mathbb{P}(\mathcal{C})}{3}, \tag{17}$$

the following inequality holds with probability at least $1 - (n+2)\exp\{-b_2\delta^2 n\}$:

$$s_{n,r}(\mathcal{C}[X]) \geq (1 - 4\delta)s_{\mathbb{P},r}(\mathcal{C}). \tag{18}$$

Let $p_d := 1/2$ if $d = 1$, $p_d := 3/4$ if $d = 2$, and otherwise $p_d := 1/d$ for $d \geq 3$.

Proposition 10 *Fix $\delta \in (0, 1/2)$. Suppose $\widetilde{\mathbb{P}}$ and $\mathcal{C}$ satisfy (A1) and (A2). For any $n \in \mathbb{N}$ satisfying*

$$B_1 \frac{(\log n)^{p_d}}{\min\{n^{1/2}, n^{1/d}\}} \leq \delta, \tag{19}$$

the following inequality holds with probability at least $1 - B_2/n - (n+1)\exp\{-b_3 n\}$:

$$\Psi_{n,r}(\mathcal{C}[X]) \geq (1 - 2\delta)\Psi_{\mathbb{P},r}(\mathcal{C}). \tag{20}$$

A note on the proof techniques: the upper bound in (16) follows by applying Bernstein's inequality to control the deviations of $\mathrm{cut}_{n,r}(\mathcal{C}[X])$, $\mathrm{vol}_{n,r}(\mathcal{C}[X])$, and $\mathrm{vol}_{n,r}(\mathcal{C}^c[X])$ around their expectations (noting that each of these is an order-2 U-statistic). To prove the lower bound (18), we require a union bound to control the minimum degree $d_{\min}(G_{n,r}[\mathcal{C}[X]])$, but otherwise the proof is similarly straightforward.

On the other hand, the proof of (20) is considerably more complicated. Our proof relies on the recent results of García Trillos and Slepcev (2015), who upper bound the L^∞-optimal transport distance between the empirical measure $\mathbb{P}_n$ and $\mathbb{P}$. For further details, we refer to Appendix B.4, where we prove Proposition 10, as well as García Trillos et al. (2016), who establish the asymptotic convergence of the sample conductance as $n \to \infty$ and $r \to 0$.

2.4 Cluster recovery

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 $\delta \in (0, 1/4)$ and a candidate cluster $\mathcal{C} \in \mathbb{R}^d$, we require that some of the tuning parameters of Algorithm 1 be chosen within specific ranges,

$$\begin{aligned} &\alpha \in \Big[(1-4\delta)^2, (1-2\delta)^2\Big) \cdot \frac{\alpha_{\mathbb{P},r}(\mathcal{C},\delta)}{2} \\ &(L, U) \subseteq \Big(\frac{1}{5(1+2\delta)}, \frac{1}{5(1+\delta)}\Big) \cdot \frac{1}{n(n-1)\mathrm{vol}_{\mathbb{P},r}(\mathcal{C})}, \end{aligned} \tag{21}$$

where

$$\alpha_{\mathbb{P},r}(\mathcal{C},\delta) := \frac{\ln(2)}{17} \cdot \frac{\Psi^2_{\mathbb{P},r}(\mathcal{C})}{\ln^2\Big(\frac{32}{(1-4\delta)s_{\mathbb{P},r}(\mathcal{C})}\Big)}. \tag{22}$$

Definition 11 *When the input parameters to Algorithm 1 satisfy (21) for some $\mathcal{C} \subseteq \mathbb{R}^d$ and $\delta \in (0, 1/4)$, we say the algorithm is* δ-well-initialized *with respect to $\mathcal{C}$.*

Of course, in practice it is not feasible to set tuning parameters based on the underlying (unknown) distribution $\mathbb{P}$ and candidate cluster $\mathcal{C}$. Typically, one runs PPR over some range of tuning parameter values and selects the cluster which has the smallest normalized cut.

By combining Lemma 7 and Propositions 8-10, we obtain an upper bound on $\Delta(\widehat{C}, \mathcal{C}[X])$ that depends solely on the distribution $\mathbb{P}$ and candidate cluster $\mathcal{C}$. To ease presentation, we introduce a *condition number*, defined for a given $\mathcal{C} \subseteq \mathbb{R}^d$ and $\delta \in (0, 1/4)$ as

$$\kappa_{\mathbb{P},r}(\mathcal{C},\delta) := \frac{(1+3\delta)(1+2\delta)}{(1-4\delta)^2(1-\delta)} \cdot \frac{\Phi_{\mathbb{P},r}(\mathcal{C})}{\alpha_{\mathbb{P},r}(\mathcal{C},\delta)}. \tag{23}$$

Theorem 12 *Fix $\delta \in (0, 1/4)$. Suppose $\widetilde{\mathbb{P}}$ and $\mathcal{C}$ satisfy (A1) and (A2). Then for any $n \in \mathbb{N}$ which satisfies (17), (19), and*

$$\frac{(1+\delta)}{(1-\delta)^4} \cdot B_3 \le n, \tag{24}$$

the following statement holds with probability at least $1 - B_2/n - 4\exp\{-b_1\delta^2 n\} - (2n+2)\exp\{-b_2\delta^2 n\} - (n+1)\exp\{-b_3 n\}$: there exists a set $\mathcal{C}[X]^g \subseteq \mathcal{C}[X]$ of large volume,

$$\mathrm{vol}_{n,r}(\mathcal{C}[X]^g) \ge \frac{1}{2}\mathrm{vol}_{n,r}(\mathcal{C}[X]),$$

such that if Algorithm 1 is δ-well-initialized with respect to $\mathcal{C}[X]$, and run with any seed node $v \in \mathcal{C}[X]^g$, then the PPR estimated cluster $\widehat{C}$ satisfies

$$\frac{\Delta(\widehat{C}; \mathcal{C}[X])}{\mathrm{vol}_{n,r}(\mathcal{C}[X])} \le 60 \cdot \kappa_{\mathbb{P},r}(\mathcal{C},\delta). \tag{25}$$

We now make some remarks.

- It is useful to compare Theorem 12 with what is already known regarding *global* spectral clustering in the context of nonparametric statistics. Schiebinger et al. (2015) consider the following variant of spectral clustering: first embed the data X into $\mathbb{R}^k$ using the bottom k eigenvectors of the degree-normalized Laplacian $I - D^{-1/2}AD^{-1/2}$, and then partition the embedded data into estimated clusters $\widehat{C}_1, \ldots, \widehat{C}_k$ using k-means clustering. They derive error bounds on the misclassification error that depend on a difficulty function $\varphi(\mathbb{P})$. In our context, where the goal is to successfully distinguish $\mathcal{C}$ and $\mathcal{C}^c$, thus where $k = 2$, this difficulty function is roughly

$$\varphi(\mathbb{P}) \approx \sqrt{\Phi_{\mathbb{P},r}(\mathcal{C})} \cdot \max\left\{\frac{1}{\Psi_{\mathbb{P},r}(\mathcal{C})^2}; \frac{1}{\Psi_{\mathbb{P},r}(\mathcal{C}^c)^2}\right\}. \tag{26}$$

  We point out two ways in which (25) is a tighter bound than (26). First, (26) depends on $\Psi_{\mathbb{P},r}(\mathcal{C}^c)$ in addition to $\Psi_{\mathbb{P},r}(\mathcal{C})$, and is thus a useful bound only if $\mathcal{C}^c$ and $\mathcal{C}$ are both internally well-connected. In contrast (25) depends only on $\Psi_{\mathbb{P},r}(\mathcal{C})$, and is thus a useful bound if $\mathcal{C}$ has small conductance, regardless of the conductance of $\mathcal{C}^c$. This is intuitive: PPR is a local rather than global algorithm, and as such the analysis requires only local rather than global conditions. Second, (26) depends on $\sqrt{\Phi_{\mathbb{P},r}(\mathcal{C})}$ rather than $\Phi_{\mathbb{P},r}(\mathcal{C})$, and since $\Phi_{\mathbb{P},r}(\mathcal{C}) \leq 1$ this results in a weaker bound. Schiebinger et al. (2015) provide experiments suggesting that the linear, rather than square-root, dependence is correct, and we theoretically confirm this in the local clustering setup. Of course, on the other hand (25) depends on $\log^2(1/s_{\mathbb{P},r}(\mathcal{C}))$, which is due to the locally-biased nature of the PPR algorithm, and does not appear in (26).

- Although Theorem 12 is stated with respect to the exact PPR vector p_v, for a sufficiently small choice of ε the application of (2) within the proof of Theorem 12 leads to an analogous result which holds for the aPPR vector $p_v^{(\varepsilon)}$. We formally state and prove this fact in Appendix E.

3. Recovery of a density cluster with PPR

We now apply the general theory established in the last section to the special case where $\mathcal{C} = \mathcal{C}_\lambda$ is a λ-density cluster—that is, a connected component of the upper level set $\{x \in \mathbb{R}^d : f(x) \geq \lambda\}$. In Section 4, we also derive a lower bound, giving a "hard problem" for which PPR will provably fail to recover a density cluster. Together, these results can be summarized as follows: PPR recovers a density cluster $\mathcal{C}_\lambda$ if and only if both $\mathcal{C}_\lambda$ and f are well-conditioned, meaning that $\mathcal{C}_\lambda$ is not too long and thin, and that f is approximately uniform inside $\mathcal{C}_\lambda$ while satisfying a low-noise condition near its boundary.

3.1 Recovery of well-conditioned density clusters

All results on density clustering assume the density f satisfies some regularity conditions. A basic requirement is the need to avoid clusters which contain arbitrarily thin bridges or spikes, or more generally clusters which can be disconnected by removing a subset of (Lebesgue) measure 0, and thus may not be resolved by any finite number of samples. To rule out such problematic clusters, we follow the approach of Chaudhuri and Dasgupta (2010), who assume the density is lower bounded on a thickened version of $\mathcal{C}_\lambda$, defined as $\mathcal{C}_{\lambda,\sigma} := \{x \in \mathbb{R}^d : \mathrm{dist}(x, \mathcal{C}) < \sigma\}$ for a given $\sigma > 0$. Regardless of the dimension of $\mathcal{C}_\lambda$, the set $\mathcal{C}_{\lambda,\sigma}$ is full dimensional. Under typical uniform continuity conditions, the requirement that the density be lower bounded over $\mathcal{C}_{\lambda,\sigma}$ will be satisfied. Such continuity conditions can be weakened (see for instance Rinaldo and Wasserman (2010); Steinwart (2015)) but we do not pursue the matter further.

In summary, our goal is to obtain upper bounds on $\Delta(\widehat{C}, \mathcal{C}_{\lambda,\sigma}[X])$, for some fixed λ and $\sigma > 0$. We have already derived upper bounds on the symmetric set difference of $\widehat{C}$ and a generic cluster $\mathcal{C}$ that depend on some population-level functionals of $\mathcal{C}$. What remains is to analyze these population-level functionals in the specific case where the candidate cluster is $\mathcal{C}_{\lambda,\sigma}$. To carry out this analysis, we will need to impose some conditions, and for the rest of this section we will assume the following.

(A3) *Bounded density within cluster:* There exist constants $0 < \lambda_\sigma < \Lambda_\sigma < \infty$ such that

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

(A4) *Low-noise density:* There exist $\theta \in (0, \infty)$ and $\gamma \in [0, 1]$ such that for any $x \in \mathbb{R}^d$ with $0 < \mathrm{dist}(x, \mathcal{C}_{\lambda,\sigma}) \le \sigma$,

$$\inf_{y \in \mathcal{C}_{\lambda,\sigma}} f(y) - f(x) \ge \theta \cdot \mathrm{dist}(x, \mathcal{C}_{\lambda,\sigma})^\gamma.$$

Roughly, this assumption ensures that the density decays sufficiently quickly as we move away from the target cluster $\mathcal{C}_{\lambda,\sigma}$, and is a standard assumption in the level-set estimation literature (see for instance Singh et al. (2009)).

(A5) *Lipschitz embedding:* There exists a differentiable function $g : \mathbb{R}^d \to \mathbb{R}^d$, $\rho \in (0, \infty)$ and $M \in [1, \infty)$ such that

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

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

(c) for some $M \ge 1$,

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

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

For convenience only, we will also make the following assumption.

(A6) *Bounded volume:* The volume of $\mathcal{C}_{\lambda,\sigma}$ is no more than half the total volume of $\mathbb{R}^d$:

$$\mathrm{vol}_{\mathbb{P},r}(\mathcal{C}_{\lambda,\sigma}) \le \mathrm{vol}_{\mathbb{P},r}(\mathcal{C}_{\lambda,\sigma}^c).$$

This assumption implies that the normalized cut of $\mathcal{C}_{\lambda,\sigma}$ will be equal to the ratio of $\mathrm{cut}_{\mathbb{P},r}(\mathcal{C}_{\lambda,\sigma})$ to $\mathrm{vol}_{\mathbb{P},r}(\mathcal{C}_{\lambda,\sigma})$.

Normalized cut, conductance, and local spread of a density cluster. In Lemma 13, Proposition 14, and Proposition 15, we give bounds on the population local spread, normalized cut, and conductance of $\mathcal{C}_{\lambda,\sigma}$. These bounds depend on the various geometric parameters just introduced.

Lemma 13 *Assume $\mathcal{C}_{\lambda,\sigma}$ satisfies Assumptions (A3) and (A5) for some $\lambda_\sigma, \Lambda_\sigma, \rho$ and M. Then,*

$$s_{\mathbb{P},r}(\mathcal{C}_{\lambda,\sigma}) \geq \frac{1}{4} \cdot \frac{\lambda_\sigma^2}{\Lambda_\sigma^2} \cdot \left(\frac{2r}{\rho}\right)^d \cdot \left(1 - \frac{r}{\sigma}\sqrt{\frac{d+2}{2\pi}}\right). \tag{27}$$

Proposition 14 *Assume $\mathcal{C}_{\lambda,\sigma}$ satisfies Assumptions (A3), (A4) and (A6) for some $\lambda_\sigma, \Lambda_\sigma, \theta$, and γ, and additionally that $0 < r \leq \frac{\sigma}{4d}$. Then,*

$$\Phi_{\mathbb{P},r}(\mathcal{C}_{\lambda,\sigma}) \leq \frac{16}{9} \cdot \frac{dr}{\sigma} \cdot \frac{\lambda\Big(\lambda_\sigma - \theta\frac{r^\gamma}{\gamma+1}\Big)}{\lambda_\sigma^2}. \tag{28}$$

Proposition 15 *Assume $\mathcal{C}_{\lambda,\sigma}$ satisfies Assumptions (A3) and (A5) for some $\lambda_\sigma, \Lambda_\sigma, \rho$ and M. Then,*

$$\Psi_{\mathbb{P},r}(\mathcal{C}_{\lambda,\sigma}) \geq \left(1 - \frac{r}{4\rho M}\right) \cdot \left(1 - \frac{r}{\sigma}\sqrt{\frac{d+2}{2\pi}}\right)^2 \cdot \frac{\sqrt{2\pi}}{36} \cdot \frac{r}{\rho M\sqrt{d+2}} \cdot \frac{\lambda_\sigma^2}{\Lambda_\sigma^2}. \tag{29}$$

Some remarks are in order.

- We prove Proposition 14 by separately upper bounding $\mathrm{cut}_{\mathbb{P},r}(\mathcal{C}_{\lambda,\sigma})$ and lower bounding the volume $\mathrm{vol}_{\mathbb{P},r}(\mathcal{C}_{\lambda,\sigma})$. Of these two bounds, the trickier to prove is the upper bound on the cut, which involves carefully estimating the probability mass of thin tubes around the boundary of $\mathcal{C}_{\lambda,\sigma}$.

- Proposition 15 is proved in a completely different way. The proof relies heavily on bounds on the isoperimetric ratio of convex sets (as derived by e.g., Lovász and Simonovits (1990) or Dyer and Frieze (1991)), and thus the embedding assumption (A5) and Lipschitz parameter M play an important role in proving the upper bound in Proposition 15.

- There is some interdependence between M and σ, ρ, which might lead one to hope that (A5) is non-essential. However, it is not possible to eliminate condition (A5) without incurring an additional factor of at least $(\rho/\sigma)^d$ in (29), achieved, for instance, when $\mathcal{C}_{\lambda,\sigma}$ is a dumbbell-like set consisting of two balls of diameter ρ linked by a cylinder of radius σ. In contrast, (29) depends polynomially on d, and many reasonably shaped sets—such as star-shaped sets as well as half-moon shapes of the type we consider in Section 5—satisfy (A5) for reasonably small values of M (Abbasi-Yadkori, 2016; Abbasi-Yadkori et al., 2017).

Applying these results along with Theorem 12, we obtain an upper bound on $\Delta(\widehat{C}, \mathcal{C}_{\lambda,\sigma}[X])$. In what follows, $C_{1,\delta}, C_{2,\delta}, \ldots$ are constants which may depend on δ, but not on n, $\mathbb{P}$ or $\mathcal{C}_{\lambda,\sigma}$, and which we keep track of in our proofs.

Theorem 16 *Let $C_\lambda \subseteq \mathbb{R}^d$ and $\delta \in (0,1/4)$. Suppose that $C_{\lambda,\sigma}$ satisfies (A2)-(A6) for some $\lambda_\sigma, \Lambda_\sigma, \theta, \gamma, \rho$ and M, that $0 < r \leq \sigma/4d$, and that the sample size n satisfies the same conditions as in Theorem 12. Then with probability at least $1 - B_2/n - 4\exp\{-b_1\delta^2 n\} - (2n+2)\exp\{-b_2\delta^2 n\} - (n+1)\exp\{-b_3 n\}$, the following statement holds: there exists a set $\mathcal{C}_{\lambda,\sigma}[X]^g \subseteq \mathcal{C}_{\lambda,\sigma}[X]$ of large volume,*

$$\mathrm{vol}_{n,r}(\mathcal{C}_{\lambda,\sigma}[X]^g) \geq \frac{1}{2}\mathrm{vol}_{n,r}(\mathcal{C}_{\lambda,\sigma}[X]).$$

such that if Algorithm 1 is δ-well-initialized with respect to $\mathcal{C}_{\lambda,\sigma}$, and run with any seed node $v \in \mathcal{C}_{\lambda,\sigma}[X]^g$, then the PPR estimated cluster $\widehat{C}$ satisfies

$$\frac{\Delta(\widehat{C}; \mathcal{C}_{\lambda,\sigma}[X])}{\mathrm{vol}_{n,r}(\mathcal{C}[X])} \leq C_{1,\delta} \cdot d^3(d+2) \cdot \frac{M^2\rho^2}{\sigma r} \cdot \frac{\Lambda_\sigma^2 \lambda(\lambda - \theta\frac{r^\gamma}{\gamma+1})}{\lambda_\sigma^4} \cdot \log^2\left(C_{2,\delta}^{1/d}\frac{\Lambda_\sigma^{2/d} M\rho}{\lambda_\sigma^{2/d} 2r}\right) \tag{30}$$

Several further remarks are as follows.

- Observe that while the diameter ρ is absent from our upper bound on normalized cut in Proposition 14, it enters the ultimate bound in Theorem 16 through the conductance. This reflects (what may be regarded as) established wisdom regarding spectral partitioning algorithms more generally (Guattery and Miller, 1995; Hein and Bühler, 2010), but newly applied to the density clustering setting: if the diameter ρ is large, then PPR may fail to recover $\mathcal{C}_{\lambda,\sigma}[X]$ even when $\mathcal{C}_\lambda$ is sufficiently well-conditioned to ensure that $\mathcal{C}_{\lambda,\sigma}[X]$ has a small normalized cut in $G_{n,r}$. This will be supported by simulations in Section 5.2.

- Several modifications of global spectral clustering have been proposed with the intent of making such procedures essentially independent of the shape of the density cluster $\mathcal{C}_\lambda$. For instance, Arias-Castro (2009); Pelletier and Pudlo (2011) introduce a cleaning step to remove low-degree vertices, whereas Little et al. (2020) use a weighted geometric graph, where the weights are computed with respect to a density-dependent distance. The resulting procedures come with stronger density cluster recovery guarantees. However, the key ingredient in such procedures is the explicitly density-dependent part of the algorithm, and spectral clustering functions as more of a post-processing step. These methods are as such very different in spirit to PPR, which is a bona fide (local) spectral clustering algorithm.

- As mentioned in the discussion after Theorem 12, the population normalized cut and conductance also play a leading role in the analysis of global spectral clustering algorithms. It therefore seems likely that similar bounds to (30) would apply to the output of global spectral clustering methods as well, but formalizing this is outside the scope of our work.

- The symmetric set difference does not measure whether $\widehat{C}$ can (perfectly) distinguish any two distinct clusters $\mathcal{C}_\lambda, \mathcal{C}'_\lambda \in \mathbb{C}_f(\lambda)$. In Appendix E, we show that the PPR estimate $\widehat{C}$ can in fact distinguish two distinct clusters $\mathcal{C}_\lambda$ and $\mathcal{C}'_\lambda$, but the result holds only under relatively restrictive conditions.

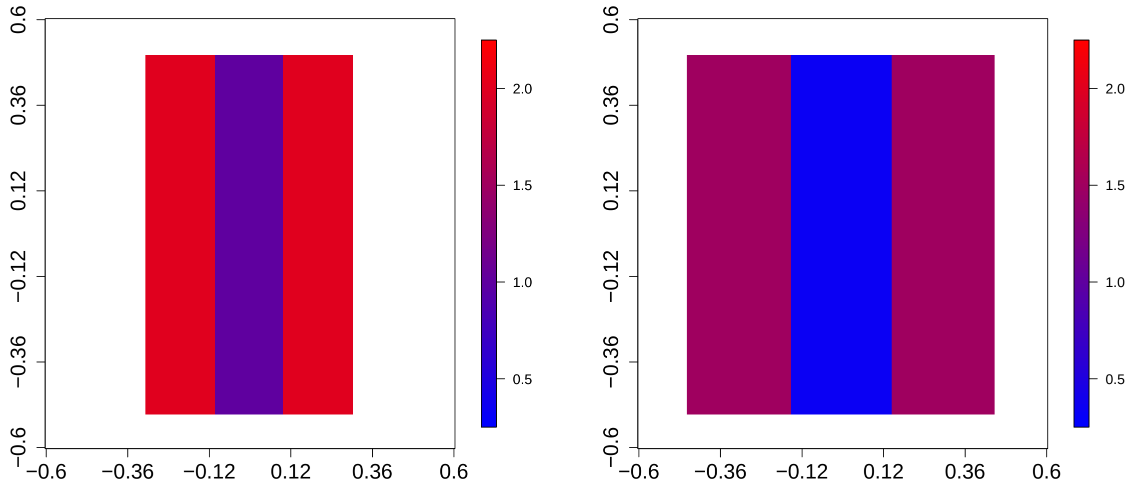


Figure 1: *The density f in (31), for $\rho = 1$, and two different choices of ϵ and σ. Left: $\epsilon = 0.3$ and $\sigma = 0.1$; right: $\epsilon = 0.2$ and $\sigma = 0.2$.*

4. Negative result

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

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

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

$$\mathbb{P} = \frac{1-\epsilon}{2}\mathbb{P}_1 + \frac{1-\epsilon}{2}\mathbb{P}_2 + \frac{\epsilon}{2}\mathbb{P}_0,$$

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

$$f(x) = \frac{1}{\rho\sigma}\left(\frac{1-\epsilon}{2}\mathbf{1}(x \in \mathcal{C}^{(1)}) + \frac{1-\epsilon}{2}\mathbf{1}(x \in \mathcal{C}^{(2)}) + \epsilon\mathbf{1}(x \in \mathcal{C}^{(0)})\right), \tag{31}$$

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

4.1 Lower bound on symmetric set difference

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

$$\mathcal{L} = \{(x_1, x_2) \in \mathcal{X} : x_2 < 0\}. \tag{32}$$

In the following Theorem, $B_{1,\delta}$ and $B_{2,\delta}$ are constants which may depend on $\delta, \mathbb{P}, \mathcal{C}_{\lambda,\sigma}$ and r, but not on n.

Theorem 17 *Fix $\delta \in (0, 1/7)$. Assume the neighborhood graph radius $r < \sigma/4$, that*

$$\max\left\{B_{1,\delta} \cdot \frac{r}{\rho}, B_{2,\delta} \cdot \frac{1}{n}\right\} < \frac{1}{18} \quad \textit{and} \quad n \geq 8\frac{(1+\delta)}{(1-\delta)}, \tag{33}$$

and that Algorithm 1 is initialized using inputs $\alpha = 36 \cdot \Phi_{n,r}(\mathcal{L}[X])$, and $(L, U) = (0, 1)$. Then the following statement holds with probability at least $1 - (B_4 + 2n + 10)\exp\{-n\delta^2 b_4\}$: there exists a set $\mathcal{C}[X]^g$ of large volume,

$$\mathrm{vol}_{n,r}(\mathcal{C}[X]^g \cap \mathcal{C}^{(1)}[X]) \geq \frac{1}{8}\mathrm{vol}_{n,r}(\mathcal{C}^{(1)}[X]; G_{n,r}),$$

suchthat for any seed node $v \in \mathcal{C}[X]^g$, the PPR estimated cluster $\widehat{C}$ satisfies

$$\frac{\sigma\rho}{r^2 n^2} \cdot \mathrm{vol}_{n,r}(\widehat{C} \vartriangle \mathcal{C}^{(1)}[X]) \geq \frac{1-\delta}{2} - C_{3,\delta} \cdot]frac\sqrt{\sigma/\rho}\epsilon^2 \cdot \sqrt{\log\left(C_{4,\delta} \cdot \frac{\rho\sigma}{\epsilon^2 r^2}\right)\frac{\sigma}{r}}, \tag{34}$$

We make a couple of remarks.

- Theorem 17 is stated with respect to a particular hard case, where the density clusters are rectangular subsets of $\mathbb{R}^2$. We chose this setting to make the theorem simple to state, and our results are generalizable to $\mathbb{R}^d$ and to non-rectangular clusters. Technically, the rectangles $\mathcal{C}^{(0)}, \mathcal{C}^{(1)}, \mathcal{C}^{(2)}$ are not σ-expansions due to their sharp corners. To fix this, one can simply modify these sets to have appropriately rounded corners, and our lower bound arguments do not need to change significantly, subject to some additional bookkeeping. Thus we ignore this technicality in our subsequent discussion.

- Although we state our lower bound with respect to PPR run on a neighborhood graph, the conclusion is likely to hold for a much broader class of spectral clustering algorithms. In the proof of Theorem 17, we rely heavily on the fact that when ϵ^2 is sufficiently greater than σ/ρ, the normalized cut of $\mathcal{C}^{(1)}$ will be much larger than that of $\mathcal{L}$. In this case, not merely PPR but any algorithm that approximates the minimum normalized cut is unlikely to recover $\mathcal{C}^{(1)}$. In particular, local spectral clustering methods that are based on truncated random walks (Spielman and Teng, 2013), global spectral clustering algorithms (Shi and Malik, 2000), and p-Laplacian based spectral embeddings (Hein and Bühler, 2010) all have provable upper bounds on the normalized cut of cluster they output, and thus we expect that they would all fail to estimate $\mathcal{C}^{(1)}$.

4.2 Comparison between upper and lower bounds

To better digest the implications of Theorem 17, we translate the results of our upper bound in Theorem 16 to the density f given in (31). Observe that $\mathcal{C}^{(1)}$ satisfies each of the Assumptions (A3)–(A6):

(A3) The density $f(x) = \frac{1-\epsilon}{2\rho\sigma}$ for all $x \in \mathcal{C}^{(1)}$.

(A4) The density $f(x) \leq \frac{\epsilon}{\rho\sigma}$ for all $x \in \mathbb{R}^2$ such that $0 < \mathrm{dist}(x, \mathcal{C}^{(1)}) \leq \sigma$. Therefore for all such x,

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

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

(A5) The set $\mathcal{C}^{(1)}$ is itself convex, and has diameter $\sqrt{\rho^2 + \sigma^2}$.

(A6) By symmetry, $\mathrm{vol}_{\mathbb{P},r}(\mathcal{C}^{(1)}) = \mathrm{vol}_{\mathbb{P},r}(\mathcal{C}^{(2)})$, and therefore $\mathrm{vol}_{\mathbb{P},r}(\mathcal{C}^{(1)}) \leq \frac{1}{2}\mathrm{vol}_{\mathbb{P},r}(\mathbb{R}^d)$.

If the user-specified parameters are initialized according to (21), we may apply Theorem 16. This implies that there exists a set $\mathcal{C}^{(1)}[X]^g \subseteq \mathcal{C}^{(1)}[X]$ with $\mathrm{vol}_{n,r}(\mathcal{C}^{(1)}[X]^g) \geq \frac{1}{2}\mathrm{vol}_{n,r}(\mathcal{C}^{(1)}[X])$ such that for any seed node $v \in \mathcal{C}^{(1)}[X]^g$, and for large enough n, the PPR estimated cluster $\widehat{C}$ satisfies with high probability

$$\frac{\mathrm{vol}_{n,r}(\widehat{C} \vartriangle \mathcal{C}^{(1)}[X])}{\mathrm{vol}_{n,r}(\mathcal{C}^{(1)}[X])} \leq 64 C_{1,\delta} \cdot \frac{\rho^2 + \sigma^2}{\sigma r} \cdot \frac{\epsilon}{1-\epsilon} \cdot \log^2\left(\sqrt{C_{2,\delta}} \frac{\rho}{2r}\right)$$

To facilitate comparisons between our upper and lower bounds set $r = \sigma/8$. Then the following statements each hold with high probability.

- If the user-specified parameters satisfy (21), and for some $a \geq 0$,

$$\frac{\epsilon}{1-\epsilon} \leq \frac{a}{512 C_{1,\delta}} \frac{\sigma^2}{(\rho^2 + \sigma^2)\log^2(\rho/\sigma\sqrt{C_{2,\delta}})},$$

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

- The population-level volume $\mathrm{vol}_{\mathbb{P},r}(\mathcal{C}^{(1)}) \leq (1-\epsilon)/2 \cdot \pi r^2/(\rho\sigma)$, and

$$\mathrm{vol}_{n,r}(\mathcal{C}^{(1)}[X]) \leq (1+\delta) \cdot n(n-1)\mathrm{vol}_{\mathbb{P},r}(\mathcal{C}^{(1)}).$$

  Therefore, if the user-specified parameters are as in Theorem 17, and

$$\epsilon \geq \sqrt{8 C_{3,\delta}} \left(\frac{\sigma}{\rho} \log\left(64 C_{4,\delta} \cdot \frac{\rho}{\epsilon^2 \sigma}\right)\right)^{1/4},$$

  then $\Delta(\widehat{C}, \mathcal{C}^{(1)}[X]) \geq \frac{1}{20}\mathrm{vol}_{n,r}(\mathcal{C}^{(1)}[X])$.

Ignoring constants and log factors, we can summarize the above conclusions as follows: if ϵ is much less than $(\sigma/\rho)^2$, then PPR will approximately recover the density cluster $\mathcal{C}^{(1)}$, whereas if ϵ is much greater than $(\sigma/\rho)^{1/4}$ then PPR will fail to recover $\mathcal{C}^{(1)}$, even if it is reasonably initialized with a seed node $v \in \mathcal{C}^{(1)}$. Jointly, these upper and lower bounds give a relatively precise characterization of what it means for a density cluster to be well- or poorly-conditioned for recovery using PPR.[5]

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

Of course, it is not hard to show that in the example under consideration, classical plug-in density cluster estimators can consistently recover $\mathcal{C}^{(1)}$, even if ϵ is large compared to σ/ρ. That PPR has trouble recovering density clusters here (where standard plug-in approaches do not) is not meant to be a knock on PPR. Rather, it simply reflects that while classical density clustering approaches are specifically designed to identify high-density regions regardless of their geometry, PPR relies on geometry as well as density when forming the output cluster.

5. Experiments

We provide numerical experiments to investigate the tightness of our theoretical results in Section 3, and compare the performance of PPR with a density clustering algorithm on the "two moons" dataset. We defer details of the experimental settings to Appendix F.

5.1 Validating theoretical bounds

We investigate the tightness of Lemma 13 and Propositions 14 and 15— i.e., the bounds on population functionals required for the eventual density cluster recovery result in Theorem 16— via simulation. Figure 2 compares our bounds on normalized cut, conductance, and local spread of a density cluster with the actual empirically-computed quantities, when samples are drawn from a mixture of uniform distributions over rectangular clusters. In the first row we vary the diameter ρ of the candidate cluster, in the second row we vary the width σ, and in the third row we vary the ratio $(\lambda - \theta)/\lambda$ of the density within and outside the cluster. In almost all cases, it is encouraging to see that our bounds track closely with their empirical counterparts, and are loose by roughly an order of magnitude at most. The one exception to this is the dependence of local spread on the width σ; this theoretical deficiency stems from a loose bound on the volume of sets with large aspect ratio (meaning ρ/σ is much greater than 1), but in any case the local spread contributes only log factors to the ultimate bound on cluster recovery. On the other hand, the looseness in each of these bounds will propagate to our eventual upper bound on $\Delta(\widehat{C}, \mathcal{C}_{\lambda,\sigma}[X])/\mathrm{vol}_{n,r}(\mathcal{C}_{\lambda,\sigma}[X])$, which as a result is loose by several orders of magnitude.

5.2 Empirical behavior of PPR

In Figure 3, to drive home the implications of Sections 3 and 4, we compare the behavior of PPR 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. We also examine the cluster which minimizes the normalized cut; as we have discussed previously, this can be seen as a middle ground between the geometric sensitivity of PPR, and the geometric insensitivity of density clustering. The first column shows the empirical density clusters $\mathcal{C}_\lambda[X]$ and $\mathcal{C}'_\lambda[X]$ 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 Bresson et al. (2012); and the last column shows a cut of the density cluster tree estimator of Chaudhuri and Dasgupta (2010). Each row corresponds to a different separation between the two moons. In the second row, we see that as the two moons become less well-separated, PPR becomes

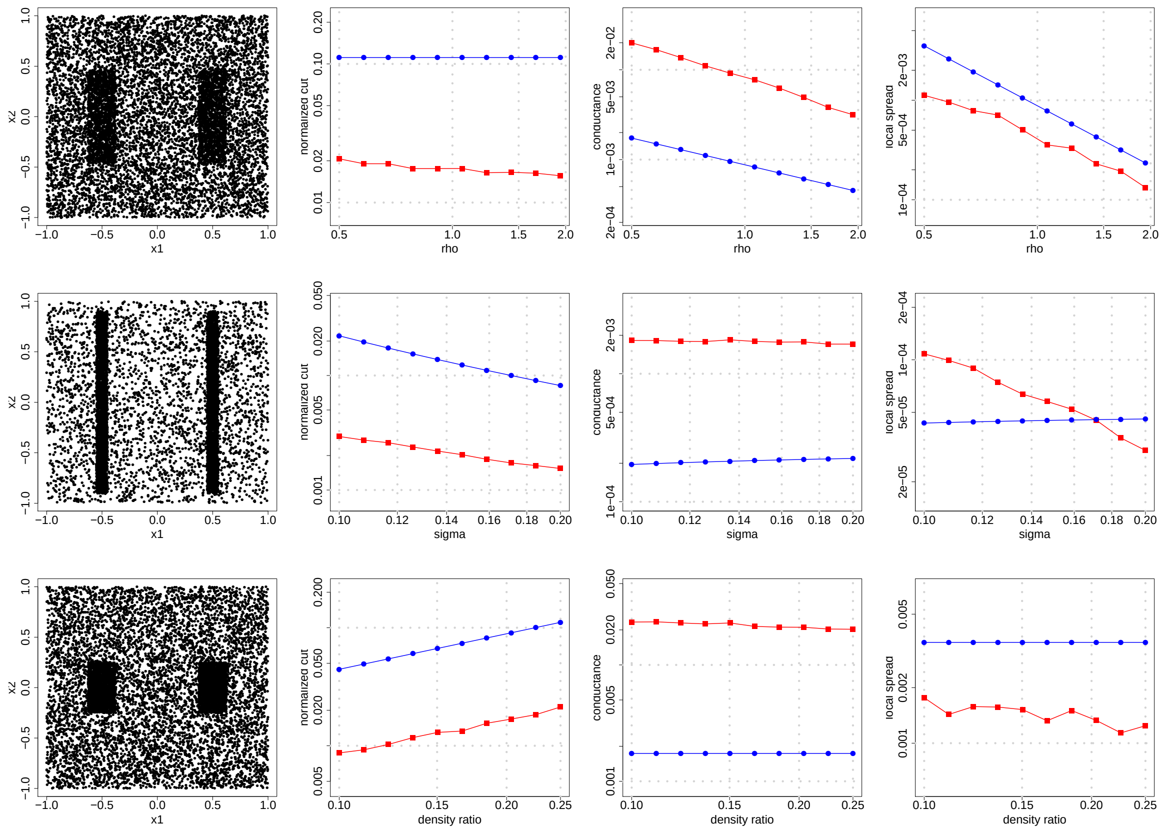


Figure 2: *Empirical normalized cut, conductance, and local spread (in red), versus their theoretical bounds (in blue). In the first row we vary the diameter ρ, in the second row we vary the thickness σ, and in the third row we vary the density ratio $(\lambda - \theta)/\lambda$. The first column shows $n = 8000$ samples for three different parameter values.*

unable to recover the density clusters, but normalized cut still succeeds in doing so. In the third row, we see that the Chaudhuri-Dasgupta algorithm succeeds even when both PPR and normalized cut fail. This supports one of our main messages, which is that PPR recovers only geometrically well-conditioned density clusters.

6. Discussion

In this work, we have analyzed the behavior of PPR in the classical setup of nonparametric statistics. We have shown how PPR depends on the distribution $\mathbb{P}$ through the population normalized cut, conductance, and local spread, and established upper bounds on the error with which PPR recovers an arbitrary candidate cluster $\mathcal{C} \subseteq \mathbb{R}^d$. In the particularly important case where $\mathcal{C} = \mathcal{C}_\lambda$ is a λ-density cluster, we have shown that PPR recovers $\mathcal{C}_\lambda$ if and only if both the density cluster and density are well-conditioned. We now conclude by summarizing a couple of interesting directions for future work.

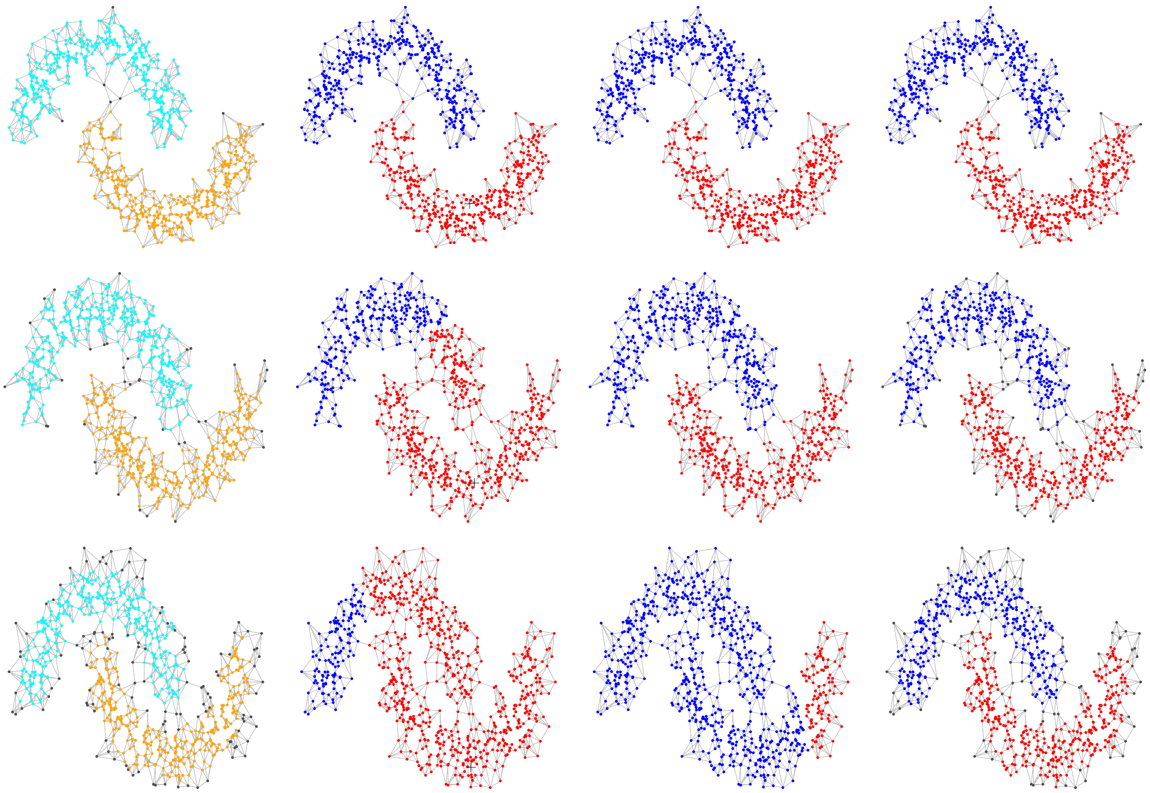


Figure 3: *True density (column 1), PPR (column 2), minimum normalized cut (column 3) and estimated density (column 4) clusters for 3 different simulated data sets. Seed node for PPR denoted by a black cross.*

Letting the radius of the neighborhood graph shrink, $r \to 0$ as $n \to \infty$, would be computationally attractive, as it would ensure that the graph $G_{n,r}$ is sparse. However, the bounds (25) and (30) will blow up as the radius r goes to 0, preventing us from making claims about the behavior of PPR in this regime. Although the restriction to a kernel function fixed in n is common in spectral clustering theory (von Luxburg et al., 2008; Schiebinger et al., 2015; Singer and Wu, 2017), recent works (Shi, 2015; Calder and García Trillos, 2019; García Trillos and Slepčev, 2018; García Trillos et al., 2020; Yuan et al., 2020) have demonstrated that spectral methods have meaningful continuum limits when $r \to 0$ as $n \to \infty$, and given precise rates of convergence. García Trillos et al. (2019) have applied these results to analyze global spectral clustering in the nonparametric mixture model, obtaining asymptotic upper bounds that do not depend on r; it seems plausible that similar bounds could be obtained for local spectral clustering with PPR, although the arguments would necessarily be quite different.

In another direction, it would be very useful to find reasonable conditions under which the ratio $\Delta(\widehat{C}, \mathcal{C}[X])/\mathrm{vol}_{n,r}(\mathcal{C}[X])$ would tend to 0 as $n \to \infty$. It seems likely that such a strong result would entail bounds on the L^∞-error of PPR. Though most results thus far derive bounds only on the L^1- or L^2-error of spectral clustering methods, some recent works (Dunson et al., 2020; Calder et al., 2020) have established L^∞-bounds on the error with which the eigenvectors of a graph Laplacian matrix approximate the eigenvectors of a weighted Laplace-Beltrami operator. It is not clear whether the techniques used in these works can be applied to PPR.

Acknowledgments

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

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.

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

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.

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

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.

Ery Arias-Castro. Clustering based on pairwise distances when the data is of mixed dimensions. *arXiv preprint arXiv:0909.2353*, 2009.

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

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

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

Xavier Bresson, Thomas Laurent, David Uminsky, and James Brecht. Convergence and energy landscape for cheeger cut clustering. In *Advances in Neural Information Processing Systems 25*, pages 1385–1393, 2012.

Jeff Calder and Nicolas García Trillos. Improved spectral convergence rates for graph laplacians on epsilon-graphs and k-nn graphs. *arXiv preprint arXiv:1910.13476*, 2019.

Jeff Calder, Nicolas Garcia Trillos, and Marta Lewicka. Lipschitz regularity of graph laplacians on random data clouds. *arXiv preprint arXiv:2007.06679*, 2020.

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

Fan RK Chung. *Spectral graph theory*. American Mathematical Soc., 1997.

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

R. M. Dudley. Distances of probability measures and random variables. *Ann. Math. Statist.*, 39(5):1563–1572, 10 1968.

David B Dunson, Hau-Tieng Wu, and Nan Wu. Spectral convergence of graph laplacian and heat kernel reconstruction in l^∞ from random samples, 2020.

Martin Dyer and Alan Frieze. Computing the volume of convex bodies: a case where randomness provably helps. Technical Report 91-104, Carnegie Mellon University, 1991.

Nicolas García Trillos and Dejan Slepcev. On the rate of convergence of empirical measures in infinity-transportation distance. *Canadian Journal of Mathematics*, 67(6):1358–1383, 2015.

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

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

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

Nicolás García Trillos, Moritz Gerlach, Matthias Hein, and Dejan Slepčev. Error estimates for spectral convergence of the graph laplacian on random geometric graphs toward the laplace–beltrami operator. *Foundations of Computational Mathematics*, 20(4):827–887, 2020.

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

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.

Peter M Gruber. *Convex and discrete geometry*, volume 336. Springer Science & Business Media, 2007.

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. *Clustering algorithms*. John Wiley & Sons, Inc., 1975.

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 23*, pages 847–855, 2010.

Franca Hoffmann, Bamdad Hosseini, Assad A Oberai, and Andrew M Stuart. Spectral analysis of weighted laplacians arising in data clustering. *arXiv preprint arXiv:1909.06389*, 2019.

Heinrich Jiang. Density level set estimation on manifolds with DBSCAN. In *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *ICML'17*, pages 1684–1693, 2017.

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

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

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

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

Giovanni Leoni. *A first Course in Sobolev Spaces*. American Mathematical Society, 2017.

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.

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

Anna V Little, Mauro Maggioni, and James M Murphy. Path-based spectral clustering: Guarantees, robustness to outliers, and fast algorithms. *Journal of Machine Learning Research*, 21(6):1–66, 2020.

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

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.

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

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

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

Bruno Pelletier and Pierre Pudlo. Operator norm convergence of spectral clustering on level sets. *Journal of Machine Learning Research*, 12(12):385–416, 2011.

Wolfgang Polonik. Measuring mass concentrations and estimating density contour clusters-an excess mass approach. *Ann. Statist.*, 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.

Alessandro Rinaldo and Larry Wasserman. Generalized density clustering. *Ann. Statist.*, 38(5): 2678–2722, 10 2010.

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

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

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

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

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

Zuoqiang Shi. Convergence of laplacian spectra from random samples. *arXiv preprint arXiv:1507.00151*, 2015.

Amit Singer and Hau-Tieng Wu. Spectral convergence of the connection laplacian from random samples. *Information and Inference: A Journal of the IMA*, 6(1):58–123, 2017.

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.

Ingo Steinwart. Fully adaptive density-based clustering. *Ann. Statist.*, 43(5):2132–2167, 10 2015.

Ingo Steinwart, Bharath K Sriperumbudur, and Philipp Thomann. Adaptive clustering using kernel density estimators. *arXiv preprint arXiv:1708.05254*, 2017.

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

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

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

Daren Wang, Xinyang Lu, and Alessandro Rinaldo. Dbscan: Optimal rates for density-based cluster estimation. *Journal of machine learning research*, 2019.

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

Amber Yuan, Jeff Calder, and Braxton Osting. A continuum limit for the pagerank algorithm. *arXiv preprint arXiv:2001.08973*, 2020.

The proofs of our major theorems largely consist of (at most) three modular parts.

1. **Fixed graph results.** Results which hold with respect to an arbitrary graph G, and are stated with respect to functionals (i.e. normalized cut, conductance, and local spread) of G;

2. **Sample-to-population results.** For the specific choice $G = G_{n,r}$, results relating the aforementioned functionals to their population analogues.

3. **Bounds on population functionals.** (In the case of density clustering only.) When the candidate cluster is a λ-density cluster, bounds on population functionals as a function of λ, as well as the other relevant parameters introduced in Section 3.

Appendices A-C will correspond to each of these three parts. In Appendix D, we will combine these parts to prove the major theorems of our main text, Theorems 12 and 16, as well as our negative result, Theorem 17. In Appendix E we derive upper bounds for the aPPR vector, and show that under certain conditions the PPR vector can perfectly separate two density clusters. Finally, in Appendix F we give relevant details regarding our experiments.

Appendix A. Fixed graph results

In this section, we give all results that hold with respect to an arbitrary graph G. For the convenience of the reader, we begin by reviewing some notation from the main text, and also introduce some new notation.

Notation. The graph $G = (V, E)$ is an undirected and connected but otherwise arbitrary graph, defined over vertices $V = \{1, \ldots, n\}$ with $m = |E|$ total edges. The adjacency matrix of G is A, the degree matrix is D, and the lazy random walk matrix over G is $W = (I + D^{-1}A)/2$. The lazy random walk originating at node $v \in V$ has distribution $q(v, t; G) = e_v W^t$ after t steps; we use the notational shorthand $q_v^{(t)} := q(v, t; G)$. The stationary distribution of the lazy random walk is $\pi := \pi(G) := \lim_{t\to\infty} q_v^{(t)}$ is given by $\pi(u) = \deg(u; G)/\mathrm{vol}(u; G)$.

For a starting distribution s (by distribution we mean a vector with non-negative entries), the PPR vector $p_s = p(s, \alpha; G)$ is the solution to

$$p_s = \alpha s + (1 - \alpha) p_s W. \tag{35}$$

When $s = e_v$, we write $p_v := p_{e_v}$. It is easy to check that $p_s = \alpha \sum_{t=0}^{\infty} (1 - \alpha)^t q_s^{(t)}$. Note that s need not be a probability distribution (i.e. its entries need not sum to 1) to make sense of (35).

Given a distribution q (for instance, $q = q_v^{(t)}$ for $t \in \mathbb{N}$, $q = p_v$, or $q = \pi$) and $\beta \in (0, 1)$, the β-sweep cut of q is

$$S_\beta(q) = \left\{ u : \frac{q(u)}{\deg(u; G)} > \beta \right\};$$

in the special case where $q = p_v$ we write $S_{\beta,v}$ for $S_\beta(p_v)$. The argument of $S_\beta(\cdot)$ will usually be clear from context, in which case we will drop it and simply write S_β. For $j = 1, \ldots, n$, let β_j be the smallest value of $\beta \in (0, 1)$ such that the sweep cut S_{β_j} contains at least j vertices. For notational ease, we will write $S_j := S_{\beta_j}$, and $S_0 = \emptyset$.

We now introduce the *Lovasz-Simonovits curve* $h_q(\cdot) : [0, 2m] \to [0, 1]$ to measure the extent to which a distribution q is mixed. To do so, we first define a piecewise linear function $q[\cdot] : [0, 2m] \to [0, 1]$. Letting $q(S) := \sum_{u \in S} q(u)$, we take $q[\mathrm{vol}(S_j)] = q(S_j)$ for each sweep cut S_j, and then extend $q[\cdot]$ by piecewise linear interpolation to be defined everywhere on its domain. Then the mixedness of q is measured by

$$h_q(k) := q[k] - \frac{k}{2m}.$$

The Lovasz-Simonovits curve is a non-negative function, with $h_q(0) = h_q(2m) = 0$. The stationary distribution π is mixed, i.e. $h_\pi(k) = 0$ for all $k \in [0, 2m]$. Finally, both $q[\cdot]$ and $h_q(\cdot)$ are concave functions, which will be an important fact later on.

The conductance of V is abbreviated as $\Psi(G) := \Psi(V; G)$, and likewise for the local spread $s(G) := s(V; G)$. Finally, for convenience we introduce the following functionals:

$$\begin{aligned} d_{\max}(C; G) &:= \max_{u \in C} \deg(u; G), & d_{\min}(C; G) &:= \min_{u \in C} \deg(u; G) \\ d_{\max}(G) &:= d_{\max}(V; G), & d_{\min}(G) &:= d_{\min}(V; G) \end{aligned}$$

We note that $d_{\min}(G)^2 \leq d_{\min}(G) \cdot n \leq \mathrm{vol}(G) \leq d_{\max}(G) \cdot n$, and that for any $S \subseteq V$, $|S| \cdot d_{\min}(G) \leq \mathrm{vol}(S; G)$ (where $|S|$ is the cardinality of S.)

Organization. In the following sections we establish: (Section A.1) an upper bound on the misclassification error of PPR in terms of α and $\Phi(C; G)$ (Lemma 7), and an analogous result for aPPR (Corollary 18); (Section A.2) a uniform bound on the perturbations of the PPR vector, to be used later in the proof of Theorem 44 (consistency of PPR); (Section A.3) upper bounds on the mixedness of $q_v^{(t)}$ (as a function of t) and p_v (as a function of α), which will be helpful in the proofs of Proposition 8 and Theorem 17; (Section A.4) an upper bound on $\tau_\infty(G)$ in terms of $\Psi(G)$ and $s(G)$ (Proposition 8); and (Section A.5) an upper bound on the normalized cut $\Phi(\widehat{C}; G)$ in terms of $\Phi(C; G)$, to be used later in the proof of Theorem 17 (negative example).

A.1 Misclassification error of clustering with PPR and aPPR

For a candidate cluster $C \subseteq V$, we use the tilde-notation $\widetilde{G} = G[C]$ to refer to the subgraph of G induced by C. Similarly we write $\widetilde{q}_v^{(t)} := q(v, t; \widetilde{G})$ for the t-step distribution of the lazy random walk over $\widetilde{G}$, $\widetilde{\pi} = \pi(G[C])$ for the stationary distribution of $\widetilde{q}_v^{(t)}$ (we will always assume $G[C]$ is connected), and $\widetilde{p}_v := p(v, \alpha; \widetilde{G})$ for the PPR vector over $\widetilde{G}$.

Proof of Lemma 7. As mentioned in the main text, Lemma 7 is equivalent, up to constants, to Lemma 3.4 in Allen-Zhu et al. (2013), and the proof of Lemma 7 proceeds along very similar lines to the proof of that lemma. In fact, we directly use the following three inequalities, derived in that work:

- **(c.f. Lemma 3.2 of Allen-Zhu et al. (2013))** For any seed node $v \in C$, the PPR vector is lower bounded,

$$\widetilde{p}_v(u) \geq \frac{3}{4}\big(1 - \alpha \cdot \tau_\infty(\widetilde{G})\big) \cdot \widetilde{\pi}(u), \quad \text{for every } u \in C. \tag{36}$$

- **(c.f. Corollary 3.3 of Allen-Zhu et al. (2013))** For any seed node $v \in C$, there exists a so-called leakage distribution $\ell = \ell(v)$ such that $\mathrm{supp}(\ell) \subseteq C$, $\|\ell\|_1 \leq 2\Phi(C;G)/\alpha$, and

$$p_v(u) \geq \widetilde{p}_v(u) - \widetilde{p}_\ell(u), \quad \text{for every } u \in C. \tag{37}$$

- **(c.f. Lemma 3.1 of Allen-Zhu et al. (2013))** There exists a set $C^g \subset C$ with $\mathrm{vol}(C^g;G) \geq \frac{1}{2}\mathrm{vol}(C;G)$ such that for any seed node $v \in C^g$, the following inequality holds

$$p_v(C^c) \leq 2\frac{\Phi(C;G)}{\alpha}. \tag{38}$$

We use (36)-(38) to separately upper bound $\mathrm{vol}(S_{\beta,v} \setminus C;G)$, $\mathrm{vol}(C^{\mathrm{int}} \setminus S_{\beta,v};G)$ and $\mathrm{vol}(C^{\mathrm{bdry}} \setminus S_{\beta,v};G)$; here $C^{\mathrm{int}} \cup C^{\mathrm{bdry}} = C$ is a partition of C, with

$$C^{\mathrm{int}} := \Big\{u \in C : \deg(u;\widetilde{G}) > \big(1 - \alpha \cdot \beta \cdot \mathrm{vol}(C;G)\big)\deg(u;G)\Big\},$$

consisting of those vertices $u \in C$ with sufficient large degree in $\widetilde{G}$.

First we upper bound $\mathrm{vol}(S_{\beta,v} \setminus C;G)$. Observe that for any $u \in S_{\beta,v} \setminus C$, $p_v(u) > \beta \cdot \deg(u;G)$. Summing up over all such vertices, from (38) we conclude that

$$\mathrm{vol}(S_{\beta,v} \setminus C;G) \leq \frac{p_v(C^c)}{\beta} \leq 2\frac{\Phi(C;G)}{\beta \cdot \alpha}. \tag{39}$$

Next we upper bound $\mathrm{vol}(C^{\mathrm{int}} \setminus S_{\beta,v};G)$. From (36) and (37) we see that

$$p_v(u) \geq \frac{3}{4}\big(1 - \alpha \cdot \tau_\infty(\widetilde{G})\big) \cdot \widetilde{\pi}(u) - \widetilde{p}_\ell(u) \;\; \text{for all } u \in C.$$

If additionally $u \notin S_{\beta,v}$ then $p_v(u) \leq \beta \deg(u;G)$, and for all such $u \in C \setminus S_{\beta,v}$,

$$\frac{3}{4}\big(1 - \alpha \cdot \tau_\infty(\widetilde{G})\big) \cdot \widetilde{\pi}(u) - \beta \deg(u;G) \leq \widetilde{p}_\ell(u). \tag{40}$$

On the other hand, for any $u \in C^{\mathrm{int}}$ it holds that

$$\widetilde{\pi}(u) = \frac{\deg(u;\widetilde{G})}{\mathrm{vol}(\widetilde{G})} \geq \frac{\deg(u;\widetilde{G})}{\mathrm{vol}(G)} \geq \frac{(1 - \alpha\beta\mathrm{vol}(C;G))\deg(u;G)}{\mathrm{vol}(C;G)};$$

by plugging this in to (40) we obtain

$$\left(\frac{3(1 - \alpha\beta\mathrm{vol}(C;G)) \cdot \big(1 - \alpha\tau_\infty(\widetilde{G})\big)}{4\mathrm{vol}(C;G)} - \beta\right) \cdot \deg(u;G) \leq \widetilde{p}_\ell(u), \quad \text{for all } u \in C^{\mathrm{int}} \setminus S_{\beta,v};$$

and summing over all such u gives

$$\left(\frac{3(1 - \alpha\beta\mathrm{vol}(C;G)) \cdot \big(1 - \alpha\tau_\infty(\widetilde{G})\big)}{4\mathrm{vol}(C;G)} - \beta\right) \cdot \mathrm{vol}\big(C^{\mathrm{int}} \setminus S_{\beta,v};G\big) \leq \widetilde{p}_\ell\big(C^{\mathrm{int}} \setminus S_{\beta,v}\big) \leq 2\frac{\Phi(C;G)}{\alpha}.$$

The upper bounds on α and β in (11) imply

$$\left(\frac{3(1-\alpha\beta\mathrm{vol}(C;G))\cdot\big(1-\alpha\tau_\infty(\widetilde{G})\big)}{4\mathrm{vol}(C;G)}-\beta\right)\geq\frac{2}{3}\beta,$$

and we conclude that

$$\mathrm{vol}(C^{\mathrm{int}}\setminus S_{\beta,v};G)\leq\frac{3\Phi(C;G)}{\alpha\beta}. \tag{41}$$

Finally, we upper bound $\mathrm{vol}(C^{\mathrm{bdry}}\setminus S_{\beta,v};G)$. Indeed, for any $u\in C^{\mathrm{bdry}}$,

$$\frac{1}{\mathrm{vol}(C;G)}\sum_{w\notin C}\mathbf{1}((u,w)\in E)\geq\alpha\cdot\beta\cdot\deg(u;G)$$

and summing over all such vertices yields

$$\mathrm{vol}(C^{\mathrm{bdry}};G)\leq\frac{1}{\alpha\beta\mathrm{vol}(C;G)}\sum_{\substack{u\in C^{\mathrm{bdry}}\\ w\notin C}}\mathbf{1}((u,w)\in E)\leq\frac{\Phi(C;G)}{\alpha\cdot\beta}. \tag{42}$$

The claim follows upon summing the upper bounds in (39), (41) and (42).

If the cluster estimate $\widehat{C}$ is instead obtained by sweep cutting the aPPR vector $p_v^{(\varepsilon)}$, a similar upper bound on $\mathrm{vol}(\widehat{C}\bigtriangleup C)$ holds, provided that ε is sufficiently small.

Corollary 18 *For a set $C\subseteq V$, suppose that α,β satisfy (11), and additionally that*

$$\varepsilon\leq\frac{1}{25\mathrm{vol}(C;G)}. \tag{43}$$

Then there exists a set $C^g\subset C$ with $\mathrm{vol}(C^g;G)\geq\frac{1}{2}\mathrm{vol}(C^g;G)$ such that for any $v\in C^g$, the sweep cut $S_{\beta,v}$ of the aPPR vector $p_v^{(\varepsilon)}$ satisfies

$$\mathrm{vol}(S_{\beta,v}\bigtriangleup C;G)\leq 6\frac{\Phi(C;G)}{\alpha\beta}. \tag{44}$$

Proof of Corollary 18. Recall that the upper bound (12) on $\mathrm{vol}(\widehat{C}\bigtriangleup C;G)$ comes from combining the upper bounds on $\mathrm{vol}(\widehat{C}\setminus C;G)$, $\mathrm{vol}(C^{\mathrm{int}}\setminus\widehat{C};G)$ and $\mathrm{vol}(C^{\mathrm{bdry}}\setminus\widehat{C};G)$ in (39), (41) and (42). From the upper bound $p_v^{(\varepsilon)}(u)\leq p_v(u)$ for all $u\in V$, it is clear that both (39) and (42) continue to hold when the aPPR vector is used instead of the PPR vector.

It remains only to establish an upper bound on $\mathrm{vol}(C^{\mathrm{int}}\setminus\widehat{C};G)$. For any $u\in C\setminus S_{\beta,v}$, from inequality (37) and the lower bound $p_v^{(\varepsilon)}(u)\geq p_v(u)-\varepsilon\deg(u;G)$ in (2) we deduce that

$$\frac{3}{4}\big(1-\alpha\cdot\tau_\infty(\widetilde{G})\big)\cdot\widetilde{\pi}(u)-(\beta+\varepsilon)\deg(u;G)\leq\widetilde{p}_\ell(u). \tag{45}$$

Following the same steps as used in the proof of Lemma 7 yields the following inequality:

$$\left(\frac{3(1-\alpha\beta\mathrm{vol}(C;G))\cdot\big(1-\alpha\tau_\infty(\widetilde{G})\big)}{4\mathrm{vol}(C;G)}-\beta\right)\cdot\mathrm{vol}\big(C^{\mathrm{int}}\setminus S_{\beta,v};G\big)\leq\widetilde{p}_\ell\big(C^{\mathrm{int}}\setminus S_{\beta,v}\big)\leq 2\frac{\Phi(C;G)}{\alpha}.$$

The upper bounds on α, β in (11), and on ε in (43), imply that

$$\left(\frac{3(1-\alpha\beta\mathrm{vol}(C;G))\cdot\big(1-\alpha\tau_\infty(\widetilde{G})\big)}{4\mathrm{vol}(C;G)} - \beta\right) \geq \frac{2}{3}\beta,$$

and we conclude that

$$\mathrm{vol}(C^{\mathrm{int}} \setminus S_{\beta,v}; G) \leq \frac{3\Phi(C;G)}{\alpha\beta}. \tag{46}$$

Summing the right hand sides of (37), (42), and (46) yields the claim.

A.2 Uniform bounds on PPR

As mentioned in our main text, in order to prove Theorem 44, we require a uniform bound on the PPR vector. Actually, we require two such bounds: for a candidate cluster $C \subseteq V$ and an alternative cluster $C' \subseteq V$, we require a lower bound on $p_v(u)$ for all $u \in C$, and an upper bound on $p_v(u')$ for all $u' \in C'$. In Lemma 19 we establish an upper bound that holds for all vertices u in the interior C_o of C, and a lower bound holds for all vertices u' in the interior of C'_o of C'; here

$$C_o = \Big\{u \in C : \deg(u, \widetilde{G}) = \deg(u; G)\Big\}, \quad \text{and} \quad C'_o = \Big\{u \in C' : \deg(u, G[C']) = \deg(u; G)\Big\},$$

and we remind the reader that $\widetilde{G} = G[C]$.

Lemma 19 *Let C and C' be disjoint subsets of V, and suppose that*

$$\alpha \leq \frac{1}{2\tau_\infty(\widetilde{G})}.$$

Then there exists a set $C^g \subseteq C$ with $\mathrm{vol}(C^g; G) \geq \mathrm{vol}(C;G)/2$ such that for any $v \in C^g$,

$$p_v(u) \geq \frac{3}{8}\widetilde{\pi}(u) - \frac{2\Phi(C;G)}{d_{\min}(\widetilde{G})\cdot\alpha} \quad \textit{for all } u \in C_o \tag{47}$$

and

$$p_v(u') \leq \frac{2\Phi(C;G)}{d_{\min}(C';G)\cdot\alpha} \quad \textit{for all } u \in C'_o. \tag{48}$$

"Leakage" and "soakage" vectors. To prove Lemma 19, we will make use of the following explicit representation of the *leakage* distribution ℓ from (38), as well as an analogously defined *soakage* distribution s:

$$\begin{aligned} \ell^{(t)} &:= e_v(W\widetilde{I})^t(I - D^{-1}\widetilde{D}), \qquad & \ell = \sum_{t=0}^{\infty}(1-\alpha)^t\ell^{(t)} \\ s^{(t)} &:= e_v(W\widetilde{I})^tW(I-\widetilde{I}), \qquad & s = \sum_{t=0}^{\infty}(1-\alpha)^t s^{(t)}. \end{aligned} \tag{49}$$

In the above, $\widetilde{I} \in \mathbb{R}^{n \times n}$ is a diagonal matrix with $I_{uu} = 1$ if $u \in C$ and 0 otherwise, and $\widetilde{D}$ is the diagonal matrix with $\widetilde{D}_{uu} = \deg(u; \widetilde{G})$ if $u \in C$, and 0 otherwise.

These quantities admit a natural interpretation in terms of random walks. For $u \in C$, $\ell^{(t)}(u)$ is the probability that a lazy random walk over G originating at v stays within C for t steps, arriving at u on the tth step, and then "leaks out" of C on the $(t+1)$st step. On the other hand, for $u \notin C$, $s^{(t)}(u)$ is the probability that a lazy random walk over G originating at v stays within C for t steps and is then "soaked up" into u on the $(t+1)$st step. The vectors ℓ and s then give the total mass leaked and soaked, respectively, by the PPR vector.

Three properties of ℓ and s are worth pointing out. First, $\mathrm{supp}(\ell) \subseteq C \setminus C_o$ and $\mathrm{supp}(s) \subseteq V \setminus C$. Second, $\|\ell^{(t)}\|_1 = \|s^{(t)}\|_1$ for all $t \in \mathbb{N}$, and so $\|\ell\|_1 = \|s\|_1$. Third, for any $u \in V \setminus C$, $p_v(u) = p_s(u)$. The first two properties are immediate. The third property follows by the law of total probability, which implies that

$$q_v^{(\tau)}(u) = \sum_{t=0}^{\tau} q_{s^{(t)}}^{(\tau-t)}(u), \quad \text{for all } u \in V \setminus C.$$

or in terms of the PPR vector,

$$p_v(u) = \alpha \sum_{\tau=0}^{\infty} (1-\alpha)^{\tau} q_v^{(\tau)}(u) = \alpha \sum_{\tau=0}^{\infty} \sum_{t=0}^{\tau} (1-\alpha)^{\tau} q_{s^{(t)}}^{(\tau-t)}(u).$$

Substituting $\Delta = \tau + t$ and rearranging gives the claimed property, as

$$p_v(u) = \alpha \sum_{\tau=0}^{\infty} \sum_{t=0}^{\tau} (1-\alpha)^{\tau} q_{s^{(t)}}^{(\tau-t)}(u) = \sum_{\Delta=0}^{\infty} \sum_{t=0}^{\infty} (1-\alpha)^{\Delta+t} q_{s^{(t)}}^{(\Delta)}(u) = \alpha \sum_{\Delta=0}^{\infty} (1-\alpha)^{\Delta} q_s^{(\Delta)}(u) = p_s(u).$$

Proof of Lemma 19. We first show (47). From (38) and (37), we have that

$$p_v(u) \geq \frac{3}{4}\bigl(1 - \alpha \cdot \tau_{\infty}(\widetilde{G})\bigr) \cdot \widetilde{\pi}(u) - \widetilde{p}_{\ell}(u) \quad \text{for all } u \in C,$$

where ℓ has support $\mathrm{supp}(\ell) \subseteq C$ with $\|\ell\|_1 \leq 2\Phi(C;G)/\alpha$. Recalling that $u \in C_o$ implies that $u \notin \mathrm{supp}(\ell)$, as a consequence of (66),

$$\widetilde{p}_{\ell}(u) \leq \frac{\|\ell\|_1}{d_{\min}(\widetilde{G})} \quad \text{for all } u \in C_o,$$

establishing (47). The proof of (48) follows similarly:

$$p_v(u) = p_s(u) \overset{(i)}{\leq} \frac{\|s\|_1}{d_{\min}(C';G)} = \frac{\|\ell\|_1}{d_{\min}(C';G)}, \quad \text{for all } u \in C_o',$$

where the presence of $d_{\min}(C';G)$ on the right hand side of (i) can be verified by inspecting (67).

A.3 Mixedness of lazy random walk and PPR vectors

In this subsection, we give upper bounds on $h^{(t)} := h_{q_v^{(t)}}$ and $h^{(\alpha)} := h_{p_v}$. Although similar bounds exist in the literature (see in particular Theorem 1.1 of (Lovász and Simonovits, 1990) and Theorem 3 of (Andersen et al., 2006)), we could not find precisely the results we needed, and so for completeness we state and prove these results ourselves.

Theorem 20 *For any $k \in [0, 2m]$, $t_0 \in \mathbb{N}$ and $t \geq t_0$,*

$$h^{(t)}(k) \leq \frac{1}{2^{t_0}} + \frac{d_{\max}(G)}{d_{\min}(G)^2} + \frac{m}{d_{\min}(G)^2}\left(1 - \frac{\Psi(G)^2}{8}\right)^{t-t_0}. \tag{50}$$

Theorem 21 *Let ϕ be any constant in $[0,1]$. Either the following bound holds for any $t \in \mathbb{N}$ and any $k \in [d_{\max}(G), 2m - d_{\max}(G)]$:*

$$h^{(\alpha)}(k) \leq \alpha t + \frac{2\alpha}{1+\alpha} + \frac{d_{\max}(G)}{d_{\min}(G)^2} + \frac{m}{d_{\min}(G)^2}\left(1 - \frac{\phi^2}{8}\right)^t,$$

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

The proofs of these upper bounds will be similar to each other (in places word-for-word alike), and will follow a similar approach and use similar notation to that of (Lovász and Simonovits, 1990; Andersen et al., 2006). For $h : [0, 2m] \to [0, 1]$, $0 \leq K_0 \leq m$ and $k \in [K_0, 2m - K_0]$, define

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

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

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

where we use the notation $\bar{k} := \min\{k, 2m - k\}$, and treat $0/0$ as equal to 1. Our first pair of Lemmas upper bound $h^{(t)}$ and $h^{(\alpha)}$ as a function of L_{K_0} and $C(K_0)$. Lemma 22 implies that if t is large relative to $\Psi(G)$, then $h^{(t)}(\cdot)$ must be small.

Lemma 22 (c.f. Theorem 1.2 of (Lovász and Simonovits, 1990)) *For any $K_0 \in [0, m]$, $k \in [K_0, 2m - K_0]$, $t_0 \in \mathbb{N}$ and $t \geq t_0$,*

$$h^{(t)}(k) \leq L_{K_0}(k; h^{(t_0)}) + C(K_0; h^{(t_0)})\sqrt{\bar{k}} \cdot \left(1 - \frac{\Psi(G)^2}{8}\right)^{t-t_0} \tag{51}$$

Lemma 23 implies that if the PPR random walk is not well mixed, then some sweep cut of p_v must have small normalized cut.

Lemma 23 (c.f Theorem 3 of (Andersen et al., 2006)) *Let $\phi \in [0,1]$. Either the following bound holds for any $t \in \mathbb{N}$, any $K_0 \in [0, m]$, and any $k \in [K_0, 2m - K_0]$:*

$$h^{(\alpha)}(k) \leq \alpha t + L_{K_0}(k; h^{(\alpha)}) + C(K_0; h^{(\alpha)})\sqrt{\bar{k}}\left(1 - \frac{\phi^2}{8}\right)^t \tag{52}$$

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

In order to make use of these Lemmas, we require upper bounds on $L_{K_0}(\cdot, h)$ and $C(K_0; h)$, for each of $h = h^{(t_0)}$ and $h = h^{(\alpha)}$. Of course, trivially $L_{K_0}(k; h) \leq \max\{h(K_0); h(2m - K_0)\}$ for any $k \in [K_0, 2m - K_0]$. As it happens, this observation will lead to sufficient upper bounds on $L_{K_0}(k, h)$ for both $h = h^{(t_0)}$ (Lemma 24) and $h = h^{(\alpha)}$ (Lemma 25).

Lemma 24 *For any $t_0 \in \mathbb{N}$ and $K_0 \in [0, m]$, the following inequalities hold:*

$$h^{(t_0)}\big(2m - K_0\big) \leq \frac{K_0}{2m} \quad \textit{and} \quad h^{(t_0)}\big(K_0\big) \leq \frac{K_0}{d_{\min}(G)^2} + \frac{1}{2^{t_0}}. \tag{53}$$

As a result, for any $k \in [K_0, 2m - K_0]$,

$$L_{K_0}(k; h^{(t_0)}) \leq \max\Big\{\frac{K_0}{2m}, \frac{K_0}{d_{\min}(G)^2} + \frac{1}{2^{t_0}}\Big\} = \frac{K_0}{d_{\min}(G)^2} + \frac{1}{2^{t_0}}. \tag{54}$$

Lemma 25 *For any $\alpha \in [0, 1]$ and $K_0 \in [0, m]$, the following inequalities hold:*

$$h^{(\alpha)}\big(2m - K_0\big) \leq \frac{K_0}{2m} \quad \textit{and} \quad h^{(\alpha)}\big(K_0\big) \leq \frac{K_0}{d_{\min}(G)^2} + \frac{2\alpha}{1 + \alpha}. \tag{55}$$

As a result, for any $k \in [K_0, 2m - K_0]$,

$$L_{K_0}(k; h^{(\alpha)}) \leq \max\Big\{\frac{K_0}{2m}, \frac{K_0}{d_{\min}(G)^2} + \frac{2\alpha}{1 + \alpha}\Big\} = \frac{K_0}{d_{\min}(G)^2} + \frac{2\alpha}{1 + \alpha}. \tag{56}$$

We next establish an upper bound on $C_{K_0}(k; h)$, which rests on the following key observation: since $h(k)$ is concave and $L_{K_0}(K_0; h) = h(K_0)$, it holds that

$$\frac{h(k) - L_{K_0}(k)}{\sqrt{k}} \leq \begin{cases} h'(K_0)\sqrt{k}, & k \leq m \\ -h'(2m - K_0)\sqrt{2m - k}, & k > m. \end{cases} \tag{57}$$

(Since h is not differentiable at $k = k_j$, here h' refers to the right derivative of h.)

Lemma 26 gives good estimates for $h'(K_0)$ and $h'(2m - K_0)$, which hold for both $h = h^{(t_0)}$ and $h = h^{(\alpha)}$, and result in an upper bound on $C(K_0; h)$. Both the statement and proof of this Lemma rely on the following explicit representation of the Lovasz-Simonovits curve $h_q(\cdot)$. Order the vertices $q(u_{(1)})/\deg(u_{(1)}; G) \geq q(u_{(2)})/\deg(u_{(2)}; G) \geq \cdots \geq q(u_{(n)})/\deg(u_{(n)}; G)$. Then for each $j = 0, \ldots, n - 1$, and for all $k \in [\mathrm{vol}(S_j), \mathrm{vol}(S_{j+1}))$, the function $h_q(k)$ satisfies

$$h_q(k) = \sum_{i=0}^{j} \big(q(u_{(i)}) - \pi(u_{(i)})\big) + \frac{\big(k - \mathrm{vol}(S_j; G)\big)}{\deg(u_{(j+1)}; G)} \big(q(u_{(j+1)}) - \pi(u_{(j+1)})\big). \tag{58}$$

Lemma 26 *The following statements hold for both $h = h^{(\alpha)}$ and $h = h^{(t_0)}$.*

- *Let $K_0 = k_1 = \deg(v; G)$ if $u_{(1)} = v$, and otherwise $K = 0$. Then*

$$h'\big(K_0\big) \leq \frac{1}{d_{\min}(G)^2}. \tag{59}$$

- *For all $K_0 \in [0, m]$,*

$$h'(2m - K_0) \geq -\frac{d_{\max}(G)}{d_{\min}(G) \cdot \mathrm{vol}(G)}. \tag{60}$$

As a result, letting $K_0 = \deg(v; G)$ if $u_{(1)} = v$, and otherwise letting $K_0 = 0$, we have

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

A.3.1 Proof of Theorems 20 and 21

Proof of Theorem 20. Take $K_0 = 0$ if $u_{(1)} \neq v$, and otherwise take $K_0 = \deg(v; G)$. Combining Lemmas 22, 24 and 26, we obtain that for any $k \in [K_0, 2m - K_0]$,

$$\begin{aligned} h^{(t)}(k) &\leq \frac{1}{2^{t_0}} + \frac{K_0}{d_{\min}(G)^2} + \frac{\sqrt{m}}{d_{\min}(G)^2}\sqrt{\bar{k}}\Big(1 - \frac{\Psi^2(G)}{8}\Big)^{t-t_0} \\ &\leq \frac{1}{2^{t_0}} + \frac{d_{\max}(G)}{d_{\min}(G)^2} + \frac{m}{d_{\min}(G)^2}\Big(1 - \frac{\Psi^2(G)}{8}\Big)^{t-t_0}, \end{aligned}$$

where the second inequality follows since we have chosen $K_0 \leq d_{\max}(G)$, and since $\bar{k} \leq m$. If $K_0 = 0$, we are done.

Otherwise, we must still establish that (51) is a valid upper bound when $k \in [0, \deg(v; G)) \cup (2m - \deg(v; G), 2m]$. If $k \in [0, \deg(v; G))$ then

$$h^{(t)}(k) \overset{(63)}{\leq} h^{(t_0)}(k) \overset{\text{(i)}}{\leq} h^{(t_0)}(K_0) \overset{(53)}{\leq} \frac{K_0}{d_{\min}(G)^2} + \frac{1}{2^{t_0}}, \tag{61}$$

where (i) follows since $k \in [0, K_0]$, and $h^{(t_0)}$ is linear over $[0, K_0)$ with $h^{(t_0)}(0) = 0$ and $h^{(t_0)}(K_0) \geq 0$. For similar reasons,

$$h^{(t)}(k) \leq h^{(t_0)}(k) \leq h^{(t_0)}(2m - K_0) \leq \frac{\deg(v; G)}{2m}. \tag{62}$$

Since the ultimate upper bounds in (61) and (62) are each no greater than that of (51), the claim follows.

Proof of Theorem 21. The proof of Theorem 21 follows immediately from Lemmas 23, 25 and 26, taking $K_0 = 0$ if $u_{(1)} \neq v$ and otherwise $K_0 = \deg(v; G)$.

A.3.2 Proofs of lemmas

In what follows, for a distribution q and vertices $u, w \in V$, we write $q(u, w) := q(u)/d(u) \cdot 1\{(u, w) \in E\}$, and similarly for a collection of dyads $\widetilde{E} \subseteq V \times V$ we write $q(\widetilde{E}) := \sum_{(u,w) \in \widetilde{E}} q(u, w)$.

Proof of Lemma 22. We will prove Lemma 22 by induction on t. In the base case $t = t_0$, observe that $C(K_0; h^{(t_0)}) \cdot \sqrt{\bar{k}} \geq h^{(t_0)}(k) - L_{K_0}(k; h^{(t_0)})$ for all $k \in [K_0, 2m - K_0]$, which implies

$$L_{K_0}(k; h^{(t_0)}) + C(K_0; h^{(t_0)}) \cdot \sqrt{\bar{k}} \geq h^{(t_0)}(k).$$

Now, we proceed with the inductive step, assuming that the inequality holds for $t_0, t_0 + 1, \ldots, t-1$, and proving that it thus also holds for t. By the definition of L_{K_0}, the inequality (51) holds when $k = K_0$ or $k = 2m - K_0$. We will additionally show that (51) holds for every $k_j = \mathrm{vol}(S_j), j = 1, 2, \ldots, n$ such that $k_j \in [K_0, 2m - K_0]$. This suffices to show that the inequality (51) holds for all $k \in [K_0, 2m - K_0]$, since the right hand side of (51) is a concave function of k.

Now, we claim that for each k_j, it holds that

$$q_v^{(t)}[k_j] \leq \frac{1}{2}\Big(q_v^{(t-1)}[k_j - \bar{k}_j \Psi(G)] + q_v^{(t-1)}[k_j + \bar{k}_j \Psi(G)]\Big). \tag{63}$$

To establish this claim, we note that for any $u \in V$

$$q_v^{(t)}(u) = \frac{1}{2} q_v^{(t-1)}(u) + \frac{1}{2} \sum_{w \in V} q_v^{(t-1)}(w, u) = \frac{1}{2} \sum_{w \in V} \big(q_v^{(t-1)}(u, w) + q_v^{(t-1)}(w, u)\big),$$

and consequently for any $S \subset V$,

$$\begin{aligned} q_v^{(t)}(S) &= \frac{1}{2}\big\{q_v^{(t-1)}(\mathrm{in}(S)) + q_v^{(t-1)}(\mathrm{out}(S))\big\} \\ &= \frac{1}{2}\big\{q_v^{(t-1)}\big(\mathrm{in}(S) \cup \mathrm{out}(S)\big) + q_v^{(t-1)}\big(\mathrm{in}(S) \cap \mathrm{out}(S)\big)\big\} \end{aligned}$$

where $\mathrm{in}(S) = \{(u, w) \in E : u \in S\}$ and $\mathrm{out}(S) = \{(w, u) \in E : w \in S\}$. We deduce that

$$\begin{aligned} q_v^{(t)}[k_j] = q_v^{(t)}(S_j) &= \frac{1}{2}\big\{q_v^{(t-1)}\big(\mathrm{in}(S_j) \cup \mathrm{out}(S_j)\big) + q_v^{(t-1)}\big(\mathrm{in}(S_j) \cap \mathrm{out}(S_j)\big)\big\} \\ &\leq \frac{1}{2}\big\{q_v^{(t-1)}\big[|\mathrm{in}(S_j) \cup \mathrm{out}(S_j)|\big] + q_v^{(t-1)}\big[|\mathrm{in}(S_j) \cap \mathrm{out}(S_j)|\big]\big\} \\ &= \frac{1}{2}\big\{q_v^{(t-1)}\big[k_j + \mathrm{cut}(S_j; G)\big] + q_v^{(t-1)}\big[k_j - \mathrm{cut}(S_j; G)\big]\big\} \\ &\leq \frac{1}{2}\big\{q_v^{(t-1)}\big[k_j + \bar{k}_j \Phi(S_j; G)\big] + q_v^{(t-1)}\big[k_j - \bar{k}_j \Phi(S_j; G)\big]\big\} \\ &\leq \frac{1}{2}\big\{q_v^{(t-1)}\big[k_j + \bar{k}_j \Psi(G)\big] + q_v^{(t-1)}\big[k_j - \bar{k}_j \Psi(G)\big]\big\}, \end{aligned}$$

establishing (63). The final two inequalities both follow from the concavity of $q_v^{(t)}[\cdot]$.

Subtracting $k_j/2m$ from both sides, we get

$$h^{(t)}(k_j) \leq \frac{1}{2}\big\{h^{(t-1)}\big(k_j + \bar{k}_j \Psi(G)\big) + h^{(t-1)}\big(k_j - \bar{k}_j \Psi(G)\big)\big\}. \tag{64}$$

At this point, we divide our analysis into cases.

Case 1. Assume $k_j - \Psi(G)\bar{k}_j$ and $k_j + 2\Psi(G)\bar{k}_j$ are both in $[K_0, 2m - K_0]$. We are therefore in a position to apply our inductive hypothesis to both terms on the right hand side of (64), and obtain the following:

$$\begin{aligned}
h^{(t)}(k_j) \leq \frac{1}{2}\Big(&L_{K_0}\big(k_j - \Psi(G)\bar{k}_j; h^{(t_0)}\big) + L_{K_0}\big(k_j + \Psi(G)\bar{k}_j; h^{(t_0)}\big)\Big) \ + \\
&\frac{1}{2}C\big(K_0; h^{(t_0)}\big) \cdot \Big(\sqrt{k_j - \Psi(G)\bar{k}_j} + \sqrt{k_j + \Psi(G)\bar{k}_j}\Big)\left(1 - \frac{\Psi(G)^2}{8}\right)^{t-t_0-1} \\
&= L_{K_0}(k; h^{(t_0)}) + \frac{1}{2}\left(C(K_0; h^{t_0})\big(\sqrt{k_j - \Psi(G)\bar{k}_j} + \sqrt{k_j + \Psi(G)\bar{k}_j}\big)\left(1 - \frac{\Psi(G)^2}{8}\right)^{t-t_0-1}\right) \\
&\leq L_{K_0}(k; h^{(t_0)}) + \frac{1}{2}\left(C(K_0; h^{(t_0)})\big(\sqrt{\bar{k}_j - \Psi(G)\bar{k}_j} + \sqrt{\bar{k}_j + \Psi(G)\bar{k}_j}\big)\left(1 - \frac{\Psi(G)^2}{8}\right)^{t-t_0-1}\right).
\end{aligned}$$

A Taylor expansion of $\sqrt{1 + \Psi(G)}$ around $\Psi(G) = 0$ yields the following bound:

$$\sqrt{1 + \Psi(G)} + \sqrt{1 - \Psi(G)} \leq 2 - \frac{\Psi(G)^2}{4},$$

and therefore

$$\begin{aligned}
h^{(t)}(k_j) &\leq L_{K_0}(k; h^{(t_0)}) + \frac{C(K_0; h^{(t_0)})}{2} \cdot \sqrt{\bar{k}_j} \cdot \left(2 - \frac{\Psi(G)^2}{4}\right)\left(1 - \frac{\Psi(G)^2}{8}\right)^{t-1} \\
&= L_{K_0}(k_j; h^{(t_0)}) + C(K_0; h^{(t_0)})\sqrt{\bar{k}_j}\left(1 - \frac{\Psi(G)^2}{8}\right)^{t-t_0}.
\end{aligned}$$

Case 2. Otherwise one of $k_j - 2\Psi(G)\bar{k}_j$ or $k_j + 2\Psi(G)\bar{k}_j$ is not in $[K_0, 2m - K_0]$. Without loss of generality assume $k_j < m$, so that (i) we have $k_j - 2\Psi(G)\bar{k}_j < K_0$ and (ii) $k_j + (k_j - K_0) \leq 2m - K_0$. We deduce the following:

$$\begin{aligned}
h^{(t)}(k_j) &\overset{\text{(i)}}{\leq} \frac{1}{2}\Big(h^{(t-1)}(K_0) + h^{(t-1)}\big(k_j + (k_j - K_0)\big)\Big) \\
&\overset{\text{(ii)}}{\leq} \frac{1}{2}\Big(h^{(t_0)}(K_0) + h^{(t)}\big(k_j + (k_j - K_0)\big)\Big) \\
&\overset{\text{(iii)}}{\leq} \frac{1}{2}\left(L_{K_0}(K_0; h^{(t_0)}) + L_{K_0}(2k_j - K_0; h^{(t_0)}) + C(K_0; h^{(t_0)})\sqrt{2k_j - K_0}\left(1 - \frac{\Psi(G)^2}{8}\right)^{t-t_0-1}\right) \\
&\leq L_{K_0}(k_j; h^{(t_0)}) + C(K_0; h^{(t_0)})\frac{\sqrt{2\bar{k}_j}}{2}\left(1 - \frac{\Psi(G)^2}{8}\right)^{t-t_0-1} \\
&\leq L_{K_0}(k_j; h^{(t_0)}) + C(K_0; h^{(t_0)})\sqrt{\bar{k}_j} \cdot \left(1 - \frac{\Psi(G)^2}{8}\right)^{t-t_0}
\end{aligned}$$

where ((i)) follows from (64) and the concavity of $h^{(t-1)}$, we deduce (ii) from (64), which implies that $h^{(t)}(k) \leq h^{(t_0)}(k)$, and (iii) follows from applying the inductive hypothesis to $h^{(t-1)}(2k_j - K_0)$.

Proof (of Lemma 23). We will show that if $\Phi(S_j; g) \geq \phi$ for each $j = 1, \ldots, n$, then (52) holds for all t and any $k \in [K_0, 2m - K_0]$.

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

$$L_{K_0}(k; h^{(\alpha)}) + C(K_0; h^{(\alpha)}) \cdot \sqrt{\bar{k}} \geq h^{(\alpha)}(k).$$

Now, we proceed with the inductive step. By the definition of L_{K_0}, the inequality (52) holds when $k = K_0$ or $k = 2m - K_0$. We will additionally show that (52) holds for every $k_j = \mathrm{vol}(S_j), j = 1, 2, \ldots, n$ such that $k_j \in [K_0, 2m - K_0]$. This suffices to show that the inequality (52) holds for all $k \in [K_0, 2m - K_0]$, since the right hand side of (52) is a concave function of k.

By Lemma 5 of Andersen et al. (2006), we have that

$$\begin{aligned}
p_v[k_j] &\leq \alpha + \frac{1}{2}\bigl(p_v[k_j - \mathrm{cut}(S_j; G)] + p_v[k_j + \mathrm{cut}(S_j; G)]\bigr) \\
&\leq \alpha + \frac{1}{2}\bigl(p_v[k_j - \Phi(S_j; G)\bar{k}_j] + p_v[k_j + \Phi(S_j; G)\bar{k}_j]\bigr) \\
&\leq \alpha + \frac{1}{2}\bigl(p_v[k_j - \phi\bar{k}_j] + p_v[k_j + \phi\bar{k}_j]\bigr)
\end{aligned}$$

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

$$h^{(\alpha)}(k_j) \leq \alpha + \frac{1}{2}\bigl(h^{(\alpha)}(k_j - \phi\bar{k}_j) + h^{(\alpha)}(k_j + \phi\bar{k}_j)\bigr) \tag{65}$$

From this point, we divide our analysis into cases.

Case 1. Assume $k_j - 2\phi\bar{k}_j$ and $k_j + 2\phi\bar{k}_j$ are both in $[K_0, 2m - K_0]$. We are therefore in a position to apply our inductive hypothesis to (65), yielding

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

and therefore

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

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

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

Proof of Lemma 24. We will prove that the inequalities of (53) hold at the knot points of $h^{(t_0)}$, whence they follow for all $K_0 \in [0, m]$.

We first prove the upper bound on $h^{(t_0)}(2m - K_0)$, when $2m - K_0 = k_j$ for some $j = 0, \ldots, n-1$. Indeed, the following manipulations show the upper bound holds for $h_q(\cdot)$ regardless of the distribution q. Noting that $h_q(2m) = 0$, we have that,

$$
h_q(k_j) = h_q(k_j) - h_q(2m) = \sum_{i=j+1}^{n} q(u_{(i)}) - \pi(u_{(i)}) \leq \sum_{i=j+1}^{n} \pi(u_{(i)}) = 1 - \frac{k_j}{2m} = \frac{K_0}{2m}.
$$

In contrast, when $K_0 = k_j$ the upper bound on $h^{(t_0)}(\cdot)$ depends on the properties of $q = q_v^{(t_0)}$. In particular, we claim that for any $t \in \mathbb{N}$,

$$
q_v^{(t)}(u) \leq \begin{cases} \dfrac{1}{d_{\min}(G)}, & \text{if } u \neq v \\ \dfrac{1}{d_{\min}(G)} + \dfrac{1}{2^t}, & \text{if } u = v. \end{cases} \tag{66}
$$

This claim follows straightforwardly by induction. In the base case $t = 0$, the claim is obvious. If the claim holds true for a given $t \in \mathbb{N}$, then for $u \neq v$,

$$
\begin{aligned}
q_v^{(t+1)}(u) &= \frac{1}{2}\sum_{w \neq u} q_v^{(t)}(w, u) + \frac{1}{2}q_v^{(t)}(u) \\
&\leq \frac{1}{2d_{\min}(G)}\sum_{w \neq u} q_v^{(t)}(w) + \frac{1}{2d_{\min}(G)} \qquad (67) \\
&\leq \frac{1}{d_{\min}(G)},
\end{aligned}
$$

where the last inequality holds because $q_v^{(t)}$ is a probability distribution (i.e. the sum of its entries is equal to 1). Similarly, if $u = v$, then

$$\begin{aligned}
q_v^{(t+1)}(v) &= \frac{1}{2}\sum_{w \neq v} q_v^{(t)}(w,v) + \frac{1}{2}q_v^{(t)}(v) \\
&\leq \frac{1}{2d_{\min}(G)}\sum_{w \neq u} q_v^{(t)}(w) + \frac{1}{2d_{\min(G)}} + \frac{1}{2^{t+1}} \\
&\leq \frac{1}{d_{\min}(G)} + \frac{1}{2^{t+1}},
\end{aligned}$$

and the claim (66) is shown. The upper bound on $h^{(t_0)}(K_0)$ for $K_0 = k_j$ follows straightforwardly:

$$h^{(t_0)}(K_0) \leq \sum_{i=0}^{j} q_v^{(t_0)}(u_{(j)}) \leq \frac{j}{d_{\min}(G)} + \frac{1}{2^{t_0}} \leq \frac{K_0}{d_{\min}(G)^2} + \frac{1}{2^{t_0}},$$

where the last inequality follows since $\mathrm{vol}(S) \geq |S| \cdot d_{\min}(G)$ for any set $S \subseteq V$.

Proof of Lemma 25. We have already established the first upper bound in (55), in the proof of Lemma 24. Then, noting that from (66),

$$p_v(u) = \alpha\sum_{t=0}^{\infty}(1-\alpha)^t q_v^{(t)}(u) \leq \begin{cases} \alpha\displaystyle\sum_{t=0}^{\infty}(1-\alpha)^t\Big(\frac{1}{d_{\min}(G)} + \frac{1}{2^t}\Big) = \frac{1}{d_{\min}(G)} + \frac{2\alpha}{1-\alpha} & \text{if } u = v \\ \alpha\displaystyle\sum_{t=0}^{\infty}(1-\alpha)^t\frac{1}{d_{\min}(G)} = \frac{1}{d_{\min}(G)} & \text{if } u \neq v, \end{cases} \tag{68}$$

the second upper bound in (55) follows similarly to the proof of the equivalent upper bound in Lemma 24.

Proof of Lemma 26. The result of the Lemma follows obviously from (57), once we show (59)-(60). We begin by showing (59). Inspecting the representation (58), we see that for any distribution q and knot point k_j, the right derivative of h_q can always be upper bounded,

$$h_q'(k_j) \leq \frac{q(u_{(j+1)})}{\deg(u_{(j+1)};G)}.$$

We have chosen $K_0 = k_j$ so that $v \neq u_{(j+1)}$, and so (66) implies that $h_q'(k_j) \leq 1/(d_{\min}(G)^2)$, for either $q = q_v^{(t)}$ or $q = p_v$.

On the other hand, the inequality (60) follows immediately from the representation (58), since for any $K_0 \in [0,m]$, taking j so that $2m - K_0 \in [k_j, k_{j+1})$,

$$h'(2m-K_0) \geq -\frac{\pi(u_{(j+1)})}{\deg(u_{(j+1)};G)} \geq -\frac{d_{\max}(G)}{d_{\min}(G)\cdot \mathrm{vol}(G)}.$$

A.4 Proof of Proposition 8

To prove Proposition 8, we first give an upper bound on the total variation distance between $q_v^{(t)}$ and its limiting distribution π, then upgrade to the desired uniform upper bound (15). The *total variation distance* between distributions q and p is

$$\mathrm{TV}(q,p) := \frac{1}{2}\sum_{u \in v}\bigl|q(u) - p(u)\bigr|$$

It follows from the representation (58) that

$$\mathrm{TV}(q,\pi) = \max_{S \subseteq V}\Bigl\{q(S) - \pi(S)\Bigr\} = \max_{j=1,\ldots,n}\Bigl\{q(S_j) - \pi(S_j)\Bigr\} = \max_{k \in [0,2m]} h_q(k),$$

so that Theorem 20 gives an upper bound on $\mathrm{TV}(q_v^{(t)},\pi)$. We can then use the following result to upgrade to a uniform upper bound.

Lemma 27 *For any $t_* \in \mathbb{N}$,*

$$\max_{u \in V}\Biggl\{\frac{\pi(u) - q_v^{(t_*+1)}(u)}{\pi(u)}\Biggr\} \leq \frac{1}{s(G)}\Biggl(\sum_{t=0}^{t_*}\frac{\mathrm{TV}(q_v^{(t)},\pi)}{2^{t_*-t}} + \frac{\mathrm{TV}(q_v^{(0)},\pi)}{2^{t_*}}\Biggr).$$

The proof of Proposition 8 is then straightforward.

Proof of Proposition 8. Put $T = 8/(\Psi(G)^2)\ln(4/s(G)) + 4$. We will use Theorem 20 to show that $\mathrm{TV}(q_v^{(T)},\pi) \leq 1/4$. This will in turn imply (Montenegro (2002) pg. 13) that for all $t \geq t_* := T\log_2(32/s(G))$,

$$\mathrm{TV}(q_v^{(t)},\pi) \leq \frac{1}{32}s(G). \tag{69}$$

Finally, let $\tau_* = t_* + 4\log_2(1/s(G))$. Applying Lemma 27 gives

$$\begin{aligned}
\max_{u \in V}\Biggl\{\frac{\pi(u) - q_v^{(\tau_*+1)}(u)}{\pi(u)}\Biggr\} &\leq \frac{1}{s(G)}\Biggl(\sum_{t=0}^{\tau_*}\frac{\mathrm{TV}(q_v^{(t)},\pi)}{2^{\tau_*-t}} + \frac{\mathrm{TV}(q_v^{(0)},\pi)}{2^{\tau_*}}\Biggr) \\
&= \frac{1}{s(G)}\Biggl(\sum_{t=t_*+1}^{\tau_*}\frac{\mathrm{TV}(q_v^{(t)},\pi)}{2^{\tau_*-t}} + \sum_{t=0}^{t_*}\frac{\mathrm{TV}(q_v^{(t)},\pi)}{2^{\tau_*-t}} + \frac{\mathrm{TV}(q_v^{(0)},\pi)}{2^{\tau_*}}\Biggr) \\
&\leq \frac{1}{4},
\end{aligned}$$

where the final inequality follows from (69) and the crude upper bound $\mathrm{TV}(q,\pi) \leq 1$, which holds for any distribution q. Taking maximum over all $v \in V$, we conclude that $\tau_\infty(G) \leq \tau_* + 1$, which implies the claim of Proposition 8.

It remains to show that $\mathrm{TV}(q_v^{(T)}, \pi) \leq 1/4$. Choosing $t_0 = 4$ in the statement of Theorem 20, we have that

$$\begin{aligned}
\mathrm{TV}(q_v^{(T)}, \pi) &\leq \frac{1}{16} + \frac{d_{\max}(G)}{d_{\min}(G)^2} + \frac{1}{2s(G)}\Bigl(1 - \frac{\Psi(G)^2}{8}\Bigr)^{T-4} \\
&\leq \frac{1}{8} + \frac{1}{2s(G)}\Bigl(1 - \frac{\Psi(G)^2}{8}\Bigr)^{T-4} \\
&\leq \frac{1}{8} + \frac{1}{2s(G)}\exp\Bigl(-\frac{\Psi(G)^2}{8}(T-4)\Bigr) = \frac{1}{4},
\end{aligned}$$

where the middle inequality follows by assumption.

Proof of Lemma 27. Our goal will be to establish the recurrence relation (71). To derive (71), the key observation is the following equivalence (see equation (16) of (Morris and Peres, 2005)):

$$\begin{aligned}
\frac{\pi(u) - q_v^{(t+1)}(u)}{\pi(u)} &= \sum_{w \in V}\bigl(\pi(w) - q_v^{(t)}(w)\bigr)\cdot\Bigl(\frac{q_w^{(1)}(u) - \pi(u)}{\pi(u)}\Bigr) \\
&= \sum_{w \neq u}\bigl(\pi(w) - q_v^{(t)}(w)\bigr)\cdot\Bigl(\frac{q_w^{(1)}(u) - \pi(u)}{\pi(u)}\Bigr) + \bigl(\pi(u) - q_v^{(t)}(u)\bigr)\cdot\Bigl(\frac{q_u^{(1)}(u) - \pi(u)}{\pi(u)}\Bigr)
\end{aligned} \tag{70}$$

We separately upper bound each term on the right hand side of (70). The sum over all $w \neq u$ can be related to the TV distance between $q_v^{(t)}$ and π using Hölder's inequality,

$$\begin{aligned}
\sum_{w \neq u}\bigl(\pi(w) - q_v^{(t)}(w)\bigr)\cdot\Bigl(\frac{q_w^{(1)}(u) - \pi(u)}{\pi(u)}\Bigr) &\leq 2\mathrm{TV}(q_v^{(t)}, \pi)\cdot\max_{w \neq u}\Bigl|\frac{q_w^{(1)}(u) - \pi(u)}{\pi(u)}\Bigr| \\
&\leq 2\mathrm{TV}(q_v^{(t)}, \pi)\cdot\max\Bigl\{1, \max_{w \neq u}\frac{q_w^{(1)}(u)}{\pi(u)}\Bigr\} \\
&\leq 2\mathrm{TV}(q_v^{(t)}, \pi)\cdot\frac{m}{d_{\min}(G)^2} = \frac{\mathrm{TV}(q_v^{(t)}, \pi)}{s(G)}.
\end{aligned}$$

On the other hand, the second term on the right hand side of (70) satisfies

$$\bigl(\pi(u) - q_v^{(t)}(u)\bigr)\cdot\Bigl(\frac{q_u^{(1)}(u) - \pi(u)}{\pi(u)}\Bigr) \leq \bigl(\pi(u) - q_v^{(t)}(u)\bigr)\cdot\Bigl(\frac{1/2 - \pi(u)}{\pi(u)}\Bigr) \leq \frac{\pi(u) - q_v^{(t)}(u)}{2\pi(u)},$$

so that we obtain the recurrence relation

$$\frac{\pi(u) - q_v^{(t+1)}(u)}{\pi(u)} \leq \frac{\mathrm{TV}(q_v^{(t)}, \pi)}{s(G)} + \frac{\pi(u) - q_v^{(t)}(u)}{2\pi(u)}. \tag{71}$$

From (71) alogn with the initial condition

$$\Bigl\{\frac{\pi(u) - q_v^{(1)}(u)}{\pi(u)}\Bigr\} \leq 1 \leq 2(1 - \pi(v)) \leq 2\frac{\mathrm{TV}(q_v^{(0)}, \pi)}{s(G)},$$

—where the second inequality follows because $\pi(v) \leq d_{\max}(G)/(2m) \leq d_{\max}(G)/d_{\min}(G)^2 \leq 1/16$—we obtain the upper bound

$$\frac{\pi(u) - q_v^{(t+1)}(u)}{\pi(u)} \leq \frac{1}{s(G)}\left(\sum_{t=0}^{t_*} \frac{\mathrm{TV}(q_v^{(t)}, \pi)}{2^{t_* - t}} + \frac{\mathrm{TV}(q_v^{(0)}, \pi)}{2^{t_*}}\right).$$

This inequality holds for each $u \in V$, and taking the maximum over u completes the proof of Lemma 27.

A.5 Spectral partitioning properties of PPR

The following theorem is the main result of Section A.5. It relates the normalized cut of the sweep sets $\Phi(S_\beta; G)$ to the normalized cut of a candidate cluster $C \subseteq V$, when p_v is properly initialized within C.

Theorem 28 (c.f. Theorem 6 of Andersen et al. (2006)) *Suppose that*

$$d_{\max}(G) \leq \mathrm{vol}(C; G) \leq \max\left\{\frac{2}{3}\mathrm{vol}(G); \mathrm{vol}(G) - d_{\max}(G)\right\} \tag{72}$$

and

$$\max\left\{288\Phi(C; G) \cdot \ln\left(\frac{36}{s(G)}\right), 72\Phi(C; G) + \frac{d_{\max}(G)}{d_{\min}(G)^2}\right\} < \frac{1}{18}. \tag{73}$$

Set $\alpha = 36 \cdot \Phi(C; G)$. The following statement holds: there exists a set $C^g \subseteq C$ of large volume, $\mathrm{vol}(C^g; G) \geq 5/6 \cdot \mathrm{vol}(C; G)$, such that for any $v \in C^g$, the minimum normalized cut of the sweep sets of p_v satisfies

$$\min_{\beta \in (0,1)} \Phi(S_{\beta,v}; G) < 72\sqrt{\Phi(C; G) \cdot \ln\left(\frac{36}{s(G)}\right)}. \tag{74}$$

A few remarks:

- Theorem 28 is similar to Theorem 6 of Andersen et al. (2006), but crucially the above bound depends on $\log\big(1/s(G)\big)$ rather than $\log m$. In the case where $d_{\min}(G)^2 \asymp \mathrm{vol}(G)$ and thus $s(G) \asymp 1$, this amounts to replacing a factor of $O(\log m)$ by a factor of $O(1)$, and therefore allows us to obtain meaningful results in the limit as $m \to \infty$.
- For simplicity, we have chosen to state Theorem 28 with respect to a specific choice of $\alpha = 36 \cdot \Phi(C; G)$, but if $\alpha \approx 36 \cdot \Phi(C; G)$ then the Theorem will still hold up to constant factors.

It follows from Markov's inequality (see Theorem 4 of Andersen et al. (2006)) that there exists a set $C^g \subseteq C$ of volume $\mathrm{vol}(C^g; G) \geq 5/6 \cdot \mathrm{vol}(C; G)$ such that for any $v \in C^g$,

$$p_v(C) \geq 1 - \frac{6\Phi(C; G)}{\alpha}. \tag{75}$$

The claim of Theorem 28 is a consequence of (75) along with Theorem 21, as we now demonstrate.

Proof of Theorem 28. From (75), the upper bound in (72), and the choice of $\alpha = 36 \cdot \Phi(C;G)$,

$$p_v(C) - \pi(C) \geq \frac{1}{3} - \frac{6\Phi(C;G)}{\alpha} = \frac{1}{6}. \tag{76}$$

Now, put

$$t_* = \frac{1}{648\Phi(C;G)}, \quad \phi_*^2 = \frac{8}{t_*} \cdot \ln\Bigl(\frac{36}{s(G)}\Bigr),$$

and note that by (73) $\phi_*^2 \in [0,1]$. It therefore follows from (76) and Theorem 21 that either

$$\frac{1}{6} \leq p_v(C) - \pi(C) \leq \frac{1}{18} + 72\Phi(C;G) + \frac{d_{\mathrm{max}}(G)}{d_{\mathrm{min}}(G)^2} + \frac{1}{2s(G)} \cdot \left(1 - \frac{\phi_*^2}{8}\right)^{t_*}, \tag{77}$$

or $\min_{\beta \in (0,1)} \Phi(S_{\beta,v};G) \leq \phi_*^2$. But by (73)

$$72\Phi(C;G) + \frac{d_{\mathrm{max}}(G)}{d_{\mathrm{min}}(G)^2} < \frac{1}{18},$$

and we have chosen ϕ_* precisely so that

$$\frac{1}{2s(G)} \cdot \left(1 - \frac{\phi_*^2}{8}\right)^{t_*} \leq \frac{1}{2s(G)} \exp\Bigl(-\frac{\phi_*^2 t_*}{8}\Bigr) \leq \frac{1}{18}.$$

Thus the inequality (77) cannot hold, and so it must be that $\min_{\beta \in (0,1)} \Phi(S_{\beta,v};G) \leq \phi_*^2$. This is exactly the claim of the theorem.

Appendix B. Sample-to-population bounds

In this appendix, we prove Propositions 9 and 10, by establishing high-probability finite-sample bounds on various functionals of the random graph $G_{n,r}$: cut, volume, and normalized cut (B.2), minimum and maximum degree, and local spread (B.3), and conductance (B.4). To establish these results, we will use several different concentration inequalities, and we begin by reviewing these in (B.1). Throughout, we denote the empirical probability of a set $\mathcal{S} \subseteq \mathbb{R}^d$ as $\mathbb{P}_n(\mathcal{S}) = \sum_{i=1}^n \mathbf{1}\{x_i \in \mathcal{S}\}/n$, and the conditional (on being in $\mathcal{C}$) empirical probability as $\widetilde{\mathbb{P}}_n = \sum_{i=1}^n \mathbf{1}\{x_i \in (\mathcal{S} \cap \mathcal{C})\}/\widetilde{n}$, where $\widetilde{n} = |\mathcal{C}[X]|$ is the number of sample points that are in $\mathcal{C}$. For a probability measure $\mathbb{Q}$, we also write

$$d_{\mathrm{min}}(\mathbb{Q}) := \inf_{x \in \mathrm{supp}(\mathbb{Q})} \deg_{\mathbb{P},r}(x), \quad \text{and} \quad d_{\mathrm{max}}(\mathbb{Q}) := \sup_{x \in \mathrm{supp}(\mathbb{Q})} \deg_{\mathbb{P},r}(x). \tag{78}$$

B.1 Review: concentration inequalities

We use Bernstein's inequality to control the deviations of the empirical probability of $\mathcal{S}$.

Lemma 29 (Bernstein's Inequality.) *Fix $\delta \in (0,1)$. For any measurable $\mathcal{S} \subseteq \mathbb{R}^d$, each of the inequalities,*

$$(1-\delta)\mathbb{P}(\mathcal{S}) \leq \mathbb{P}_n(\mathcal{S}) \quad \textit{and} \quad \mathbb{P}_n(\mathcal{S}) \leq (1+\delta)\mathbb{P}(\mathcal{S}),$$

hold with probability at least $1 - \exp\bigl\{-n\delta^2\mathbb{P}(\mathcal{S})/(2+2\delta)\bigr\} \geq 1 - \exp\bigl\{-n\delta^2\mathbb{P}(\mathcal{S})/4\bigr\}$.

Many graph functionals are order-2 U-statistics, and we use Bernstein's inequality to control the deviations of these functionals from their expectations. Recall that U_n is an order-2 U-statistic with kernel $\varphi : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ if

$$U_n = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i} \varphi(x_i, x_j).$$

We write $\|\varphi\|_\infty = \sup_{x,y} |\varphi(x,y)|$.

Lemma 30 (Bernstein's Inequality for Order-2 U-statistics.) *Fix $\delta \in (0,1)$. Assume $\|\varphi\|_\infty \leq 1$. Then each of the inequalities,*

$$(1-\delta)\mathbb{E}U_n \leq U_n \quad \textit{and} \quad U_n \leq (1+\delta)\mathbb{E}U_n,$$

hold with probability at least $1 - \exp\{-n\delta^2 \mathbb{E}U_n/(4 + 4\delta/3)\} \geq 1 - \exp\{-n\delta^2\mathbb{E}U_n/6\}$.

Finally, we use Lemma 31—a combination of Bernstein's inequality and a union bound—to upper and lower bound $d_{\max}(G_{n,r})$ and $d_{\min}(G_{n,r})$. For measurable sets $\mathcal{S}_1, \ldots, \mathcal{S}_M$, we denote $p_{\min} := \min_{m=1,\ldots,M} \mathbb{P}(\mathcal{A}_m)$, and likewise let $p_{\max} := \max_{m=1,\ldots,M} \mathbb{P}(\mathcal{A}_m)$

Lemma 31 (Bernstein's inequality + union bound.) *Fix $\delta \in (0,1)$. For any measurable $\mathcal{S}_1, \ldots, \mathcal{S}_M \subseteq \mathbb{R}^d$, each of the inequalities*

$$(1-\delta)p_{\min} \leq \min_{m=1,\ldots,M} \mathbb{P}_n(\mathcal{A}_m), \quad \textit{and} \quad \max_{m=1,\ldots,M} \mathbb{P}_n(\mathcal{A}_m) \leq (1+\delta)p_{\max}$$

hold with probability at least $1 - M\exp\{-n\delta^2 p_{\min}/(2+2\delta)\} \geq 1 - M\exp\{-n\delta^2 p_{\min}/4\}$.

B.2 Sample-to-population: normalized cut

In this subsection we establish (16). For a set $\mathcal{S} \subseteq \mathbb{R}^d$, both $\mathrm{cut}_{n,r}(\mathcal{S}[X])$ and $\mathrm{vol}_{n,r}(\mathcal{S}[X])$ are order-2 U-statistics:

$$\mathrm{cut}_{n,r}(\mathcal{S}[X]) = \sum_{i=1}^{n} \sum_{j \neq i} \mathbf{1}\{\|x_i - x_j\| \leq r\} \cdot \mathbf{1}\{x_i \in \mathcal{S}\} \cdot \mathbf{1}\{x_j \notin \mathcal{S}\},$$

and

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

Therefore with probability at least $1 - \exp\{-n\delta^2 \mathrm{cut}_{\mathbb{P},r}(\mathcal{S})/4\}$,

$$\frac{1}{n(n-1)}\mathrm{cut}_{n,r}(\mathcal{S}[X]) \leq (1+\delta)\mathrm{cut}_{\mathbb{P},r}(\mathcal{S}),$$

and likewise with probability at least $1 - \exp\{-n\delta^2\mathrm{vol}_{\mathbb{P},r}(\mathcal{S})/4\} - \exp\{-n\delta^2\mathrm{vol}_{\mathbb{P},r}(\mathcal{S}^c)/4\}$,

$$(1-\delta)\mathrm{vol}_{\mathbb{P},r}(\mathcal{S}) \leq \frac{1}{n(n-1)}\mathrm{vol}_{n,r}(\mathcal{S}[X]), \quad \text{and} \quad (1-\delta)\mathrm{vol}_{\mathbb{P},r}(\mathcal{S}^c) \leq \frac{1}{n(n-1)}\mathrm{vol}_{n,r}(\mathcal{S}^c[X])$$

Consequently, for any $\delta \in (0, 1/3)$,

$$\Phi_{n,r}(\mathcal{C}[X]) \leq \frac{1+\delta}{1-\delta} \cdot \frac{\mathrm{cut}_{\mathbb{P},r}(\mathcal{C})}{\min\{\mathrm{vol}_{\mathbb{P},r}(\mathcal{C}), \mathrm{vol}_{\mathbb{P},r}(\mathcal{C}^c)\}} = \frac{1+\delta}{1-\delta} \cdot \Phi_{\mathbb{P},r}(\mathcal{C}) \leq (1+3\delta) \cdot \Phi_{\mathbb{P},r}(\mathcal{C})$$

with probability at least $1 - 3\exp\{-n\delta^2 \mathrm{cut}_{\mathbb{P},r}(\mathcal{C})/4\}$. This establishes (16) upon taking $b_1 := 3\mathrm{cut}_{\mathbb{P},r}(\mathcal{C})/4$.

B.3 Sample-to-population: local spread

In this subsection we establish (18). To ease the notational burden, let $\widetilde{G}_{n,r} := G_{n,r}\big[\mathcal{C}[X]\big]$. Conditional on $\widetilde{n}$, it follows from Lemma 31 that with probability at least $1 - \widetilde{n}\exp\{-(\widetilde{n} - 1)\delta^2 d_{\min}(\widetilde{\mathbb{P}})/4\}$,

$$(1-\delta) \cdot d_{\min}(\widetilde{\mathbb{P}}) \leq \frac{1}{\widetilde{n}-1} d_{\min}(\widetilde{G}_{n,r}), \tag{79}$$

Likewise it follows from Lemma 30 that with probability at least $1 - \exp\{-\widetilde{n}\delta^2 \mathrm{vol}_{\widetilde{\mathbb{P}},r}(\mathcal{C})/6\}$,

$$(1-\delta) \cdot \mathrm{vol}_{\widetilde{\mathbb{P}},r}(\mathcal{C}) \leq \frac{1}{\widetilde{n}(\widetilde{n}-1)} \mathrm{vol}(\widetilde{G}_{n,r}).$$

Finally, it follows from Lemma 29 that with probability at least $1 - \exp\{-n\delta^2 \mathbb{P}(\mathcal{C})/4\}$

$$\widetilde{n} \geq (1-\delta) \cdot n \cdot \mathbb{P}(\mathcal{C}), \tag{80}$$

and therefore by (17), $(\widetilde{n}-1)/\widetilde{n} \geq 1 - \delta$. Consequently for any $\delta \in (0, 1/3)$,

$$s_{n,r}(\mathcal{C}[X]) = \frac{d_{\min}(\widetilde{G}_{n,r})^2}{\mathrm{vol}(\widetilde{G}_{n,r})} = \frac{\widetilde{n}-1}{\widetilde{n}} \cdot \frac{\frac{1}{(\widetilde{n}-1)^2} d_{\min}(\widetilde{G}_{n,r})^2}{\frac{1}{\widetilde{n}(\widetilde{n}-1)} \mathrm{vol}(\widetilde{G}_{n,r})} \geq \frac{(1-\delta)^3}{(1+\delta)} \cdot \frac{d_{\min}(\widetilde{\mathbb{P}})^2}{\mathrm{vol}_{\widetilde{\mathbb{P}},r}(\mathcal{C})} \geq (1-4\delta) \cdot s_{\mathbb{P},r}(\mathcal{C}).$$

with probability at least $1 - n\exp\{-n\mathbb{P}(\mathcal{C}) \cdot \delta^2 d_{\min}(\widetilde{\mathbb{P}})/9\} - \exp\{-n\mathbb{P}(\mathcal{C})\delta^2 \cdot \mathrm{vol}_{\widetilde{\mathbb{P}},r}(\mathcal{C})/14\} - \exp\{-n\delta^2\mathbb{P}(\mathcal{C})/4\}$. This establishes (18) upon taking $b_2 := \mathbb{P}(\mathcal{C}) \cdot d_{\min}(\widetilde{\mathbb{P}})/14$.

B.4 Sample-to-population: conductance

In this section we establish (20). As mentioned in our main text, the proof of (20) relies on a high-probability upper bound of the ∞-transportation distance between $\mathbb{P}$ and $\mathbb{P}_n$, from (García Trillos and Slepcev, 2015). We begin by reviewing this upper bound, which we restate in Theorem 32. Subsequently in Proposition 33, we relate the ∞-transportation distance between two measures $\mathbb{Q}_1$ and $\mathbb{Q}_2$ to the difference of their conductances. Together these results will imply (20).

Review: ∞-transportation distance and transportation maps. We give a brief review of some of the main ideas regarding ∞-transportation distance, and transportation maps. This discussion is largely taken from (García Trillos and Slepcev, 2015; García Trillos et al., 2016), and the reader should consult these works for more detail.

For two measures $\mathbb{Q}_1$ and $\mathbb{Q}_2$ on a domain D, the ∞-*transportation distance* $\Delta_\infty(\mathbb{Q}_1, \mathbb{Q}_2)$ is

$$\Delta_\infty(\mathbb{Q}_1, \mathbb{Q}_2) := \inf_{\gamma} \Big\{ \mathrm{esssup}_\gamma \big\{ |x-y| : (x,y) \in D \times D \big\} : \gamma \in \Gamma(\mathbb{Q}_1, \mathbb{Q}_2) \Big\}$$

where $\Gamma(\mathbb{Q}_1, \mathbb{Q}_2)$ is the set of all couplings of $\mathbb{Q}_1$ and $\mathbb{Q}_2$, that is the set of all probability measures on $D \times D$ for which the marginal distribution in the first variable is $\mathbb{Q}_1$, and the marginal distribution in the second variable is $\mathbb{Q}_2$.

Suppose $\mathbb{Q}_1$ is absolutely continuous with respect to the Lebesgue measure. Then $\Delta_\infty(\mathbb{Q}_1, \mathbb{Q}_2)$ can be more simply defined in terms of push-forward measures and transportation maps. For a Borel map $T : D \to D$, the *push-forward* of $\mathbb{Q}_1$ by T is $T_\sharp \mathbb{Q}_1$, defined for Borel sets U as

$$T_\sharp \mathbb{Q}_1(U) = \mathbb{Q}_1(T^{-1}(U)).$$

A *transportation map* from $\mathbb{Q}_1$ to $\mathbb{Q}_2$ is a Borel map T for which $T_\sharp \mathbb{Q}_1 = \mathbb{Q}_2$. Transportation maps satisfy two important properties. First, the transportation distance can be formulated in terms of transportation maps:

$$\Delta_\infty(\mathbb{Q}_1, \mathbb{Q}_2) = \inf_T \|\mathrm{Id} - T\|_{L^\infty(\mathbb{Q}_1)}$$

where $\mathrm{Id} : D \to D$ is the identity mapping, and the infimum is over transportation maps T from $\mathbb{Q}_1$ to $\mathbb{Q}_2$. Second, they result in the following change of variables formula; if $T_\sharp \mathbb{Q}_1 = \mathbb{Q}_2$, then for any $g \in L^1(\mathbb{Q}_2)$,

$$\int g(y)\, d\mathbb{Q}_2(y) = \int g(T(x))\, d\mathbb{Q}_1(x). \tag{81}$$

∞-transportation distance between empirical and population measures. We now review the relevant upper bound on $\Delta_\infty(\mathbb{P}, \mathbb{P}_n)$, which holds under the following mild regularity conditions.

(A1) The distribution $\mathbb{P}$ has density $g : D \to (0, \infty)$ such that there exist $g_{\min} \leq 1 \leq g_{\max}$ for which

$$(\forall x \in D) \quad g_{\min} \leq g(x) \leq g_{\max}.$$

(A2) The distribution $\mathbb{P}$ is defined on a bounded, connected, open domain $D \subseteq \mathbb{R}^d$. If $d \geq 2$ then additionally D has Lipschitz boundary.

When $d = 1$, it follows from Proposition 6.2 of Dudley (1968) that $\Delta_\infty(\mathbb{P}, \mathbb{P}_n) \leq B_5 \|F - F_n\|_\infty$ for some positive constant B_5, and in turn from the DKW inequality that

$$\Delta_\infty(\mathbb{P}, \mathbb{P}_n) \leq B_5 \sqrt{\frac{\ln(2n/B_2)}{n}} \tag{82}$$

with probability at least $1 - B_2/n$.

When $d \geq 2$, García Trillos and Slepcev (2015) derive an upper bound on the transportation distance $\Delta_\infty(\mathbb{P}, \mathbb{P}_n)$.

Theorem 32 (Theorem 1.1 of García Trillos and Slepcev (2015)) *Suppose $\mathbb{P}$ satisfies (A1) and (A2). Then, there exists positive constants B_2 and B_5 that do not depend on n, such that with probability at least $1 - B_2/n$:*

$$\Delta_\infty(\mathbb{P}, \mathbb{P}_n) \leq B_5 \cdot \begin{cases} \dfrac{\ln(n)^{3/4}}{n^{1/2}}, & \text{if } d = 2, \\ \dfrac{\ln(n)^{1/d}}{n^{1/d}}, & \text{if } d \geq 3. \end{cases}$$

Assuming the candidate cluster $\mathcal{C}$ and conditional distribution $\widetilde{\mathbb{P}}$ satisfy (A1) and (A2), then (82) ($d = 1$) or Theorem 32 ($d \geq 2$) apply to $\Delta_\infty(\widetilde{\mathbb{P}}, \widetilde{\mathbb{P}}_n)$; we will use these upper bounds on $\Delta_\infty(\widetilde{\mathbb{P}}, \widetilde{\mathbb{P}}_n)$ to show (20).

Lower bound on conductance using transportation maps. Let $\mathbb{Q}_1$ and $\mathbb{Q}_2$ be probability measures, with $\mathbb{Q}_1$ absolutely continuous with respect to Lebesgue measure, and let T be a transportation map from $\mathbb{Q}_1$ to $\mathbb{Q}_2$. We write $\Delta_T(\mathbb{Q}_1, \mathbb{Q}_2) := \|\mathrm{Id} - T\|_{L^\infty(\mathbb{Q}_1)}$. To facilitate easy comparison between the conductances of two arbitrary distributions, let $\Psi_r(\mathbb{Q}) := \Psi_{\mathbb{Q},r}(\mathrm{supp}(\mathbb{Q}))$ for a distribution $\mathbb{Q}$. In the following Proposition, we lower bound $\Psi_r(\mathbb{Q}_2)$ by $\Psi_r(\mathbb{Q}_1)$, plus an error term that depends on $\Delta(\mathbb{Q}_1, \mathbb{Q}_2)$.

Proposition 33 *Let $\mathbb{Q}_1$ be a probability measure that admits a density g with respect to $\nu(\cdot)$, let $\mathbb{Q}_2$ be an arbitrary probability measure, and let T be a transportation map from $\mathbb{Q}_1$ to $\mathbb{Q}_2$. Suppose $\Delta_T(\mathbb{Q}_1, \mathbb{Q}_2) \leq r/(4(d-1))$. It follows that*

$$\Psi_r(\mathbb{Q}_2) \geq \Psi_r(\mathbb{Q}_1) \cdot \left(1 - \frac{B_6 \Delta_T(\mathbb{Q}_1, \mathbb{Q}_2)}{\big(1 - \Psi_r(\mathbb{Q}_1)\big) \cdot \big(d_{\min}(\mathbb{Q}_2)\big)^2}\right) - \frac{B_6 \Delta_T(\mathbb{Q}_1, \mathbb{Q}_2)}{\big(1 - \Psi_r(\mathbb{Q}_1)\big) \cdot \big(d_{\min}(\mathbb{Q}_2)\big)^2}, \tag{83}$$

where $B_6 := 4d\nu_d r^{d-1} \cdot \max_{x \in \mathbb{R}^d}\{g(x)\}$ is a positive constant that does not depend on $\mathbb{Q}_2$.

We note that the lower bound can also be stated with respect to the ∞-optimal transport distance $\Delta_\infty(\mathbb{Q}_1, \mathbb{Q}_2)$.

Proof of Proposition 33. Throughout this proof, we will write $\Delta_{12} = \Delta_T(\mathbb{Q}_1, \mathbb{Q}_2)$, and $\overline{\mathrm{vol}}_{\mathbb{Q},r}(\mathcal{R}) = \min\big\{\mathrm{vol}_{\mathbb{Q},r}(\mathcal{R}), \mathrm{vol}_{\mathbb{Q},r}(\mathcal{R}^c)\big\}$ for conciseness. Naturally, the proof of Proposition 33 involves using the transportation map T to relate $\mathrm{cut}_{\mathbb{Q}_2,r}(\cdot)$ to $\mathrm{cut}_{\mathbb{Q}_1,r}(\cdot)$, and likewise $\mathrm{vol}_{\mathbb{Q}_2,r}(\cdot)$ to $\mathrm{vol}_{\mathbb{Q}_1,r}(\cdot)$. Define the remainder term $R^{(\Delta)}_{\epsilon,\mathbb{Q}_1}(x) = \int \mathbf{1}\{\epsilon \leq \|x - y\| \leq \epsilon + \Delta\}\, d\mathbb{Q}_1(y)$ for any $\epsilon, \Delta > 0$. Then for any set $\mathcal{S} \subseteq \mathrm{supp}(\mathbb{Q}_2)$, we have that

$$\begin{aligned}
\mathrm{cut}_{\mathbb{Q}_2,r}(\mathcal{S}) &= \iint \mathbf{1}\{\|x - y\| \leq r\} \cdot \mathbf{1}\{x \in \mathcal{S}\} \cdot \mathbf{1}\{y \in \mathcal{S}^c\}\, d\mathbb{Q}_2(y)\, d\mathbb{Q}_2(x) \\
&\overset{\text{(i)}}{=} \iint \mathbf{1}\{\|T(x) - T(y)\| \leq r\} \cdot \mathbf{1}\{x \in T^{-1}(\mathcal{S})\} \cdot \mathbf{1}\{y \in T^{-1}(\mathcal{S})^c\}\, d\mathbb{Q}_1(y)\, d\mathbb{Q}_1(x) \\
&\overset{\text{(ii)}}{\geq} \iint \mathbf{1}\{\|x - y\| \leq r - 2\Delta_{12}\} \cdot \mathbf{1}\{x \in T^{-1}(\mathcal{S})\} \cdot \mathbf{1}\{y \in T^{-1}(\mathcal{S}^c)\}\, d\mathbb{Q}_1(y)\, d\mathbb{Q}_1(x) \\
&= \mathrm{cut}_{\mathbb{Q}_1,r}\big(T^{-1}(\mathcal{S})\big) - \int R^{(2\Delta_{12})}_{r-2\Delta_{12},\mathbb{Q}_1}(x)\, d\mathbb{Q}_1(x) \qquad (84)
\end{aligned}$$

where (i) follows from the change of variables formula (81), and (ii) follows from the triangle inequality. Similar reasoning implies that

$$\mathrm{vol}_{\mathbb{Q}_2,r}(\mathcal{S}) \leq \mathrm{vol}_{\mathbb{Q}_1,r}\big(T^{-1}(\mathcal{S})\big) + \int R_{r,\mathbb{Q}_1}^{(2\Delta_{12})}(x)\, d\mathbb{Q}_1(x). \tag{85}$$

For any $x \in \mathbb{R}^d$, since $0 \leq \Delta_{12} \leq r/(4(d-1))$, the remainder terms can be upper bounded as follows:

$$R_{r-2\Delta_{12},\mathbb{Q}_1}^{(2\Delta_{12})}(x) \leq \nu_d r^d \Big\{1 - \Big(1 - \frac{2\Delta_{12}}{r}\Big)^d\Big\} \cdot \max_{x \in \mathbb{R}^d}\{g(x)\} \leq \underbrace{4d\nu_d r^{d-1} \cdot \max_{x\in\mathbb{R}^d}\{g(x)\}}_{=B_6} \cdot \Delta_{12},$$

and

$$R_{r,\mathbb{Q}_1}^{(2\Delta_{12})}(x) \leq \nu_d r^d \Big\{\Big(1 + \frac{2\Delta_{12}}{r}\Big)^d - 1\Big\} \cdot \max_{x\in\mathbb{R}^d}\{g(x)\} \leq B_6 \cdot \Delta_{12}.$$

Plugging these bounds on the remainder terms back into (84) and (85) respectively, we see that

$$\begin{aligned}
\Phi_{\mathbb{Q}_2,r}(\mathcal{S}) &\geq \frac{\mathrm{cut}_{\mathbb{Q}_1,r}\big(T^{-1}(\mathcal{S})\big) - B_6\Delta_{12}}{\overline{\mathrm{vol}}_{\mathbb{Q}_1,r}(T^{-1}(\mathcal{S})) + B_6\Delta_{12}} \\
&= \Phi_{\mathbb{Q}_1,r}(T^{-1}(\mathcal{S})) \cdot \left(\frac{\overline{\mathrm{vol}}_{\mathbb{Q}_1,r}(T^{-1}(\mathcal{S}))}{\overline{\mathrm{vol}}_{\mathbb{Q}_1,r}(T^{-1}(\mathcal{S})) + B_6\Delta_{12}}\right) - \frac{B_6\Delta_{12}}{\overline{\mathrm{vol}}_{\mathbb{Q}_1,r}(T^{-1}(\mathcal{S})) + B_6\Delta_{12}} \\
&\overset{(85)}{\geq} \Phi_{\mathbb{Q}_1,r}(T^{-1}(\mathcal{S})) \cdot \left(\frac{\overline{\mathrm{vol}}_{\mathbb{Q}_2,r}(\mathcal{S}) - B_6\Delta_{12}}{\overline{\mathrm{vol}}_{\mathbb{Q}_2,r}(\mathcal{S})}\right) - \frac{B_6\Delta_{12}}{\overline{\mathrm{vol}}_{\mathbb{Q}_2,r}(\mathcal{S})}.
\end{aligned}$$

We would like to conclude by taking an infimum over $\mathcal{S}$ on both sides, but in order to ensure that the remainder term is small we must specially handle the case where $\overline{\mathrm{vol}}_{\mathbb{Q}_2,r}(\mathcal{S})$ is small. Let

$$\mathfrak{L}_r(\mathbb{Q}_1,\mathbb{Q}_2) = \big\{\mathcal{S} \subseteq \mathrm{supp}(\mathbb{Q}_2) : \overline{\mathrm{vol}}_{\mathbb{Q}_2,r}(\mathcal{S}) \geq (1 - \Psi_r(\mathbb{Q}_1)) \cdot d_{\min}(\mathbb{Q}_2)^2\big\}.$$

On the one hand, taking an infimum over all sets $\mathcal{S} \in \mathfrak{L}_r(\mathbb{Q}_1,\mathbb{Q}_2)$, we have that

$$\inf_{\mathcal{S}:\mathcal{S}\in\mathfrak{L}_r(\mathbb{Q}_1,\mathbb{Q}_2)} \Phi_{\mathbb{Q}_2,r}(\mathcal{S}) \geq \Psi_r(\mathbb{Q}_1) \cdot \left(1 - \frac{B_6\Delta_{12}}{(1-\Psi_r(\mathbb{Q}_1)) \cdot d_{\min}(\mathbb{Q}_2)^2}\right) - \frac{B_6\Delta_{12}}{(1-\Psi_r(\mathbb{Q}_1)) \cdot d_{\min}(\mathbb{Q}_2)^2}$$

On the other hand, we claim that

$$\Phi_{r,\mathbb{Q}_2}(\mathcal{R}) \geq \Psi_r(\mathbb{Q}_1), \quad \text{for any } \mathcal{R} \notin \mathfrak{L}(\mathbb{Q}_1,\mathbb{Q}_2). \tag{86}$$

To derive (86), suppose that $\mathcal{R} \subseteq \mathrm{supp}(\mathbb{Q}_2)$ and $\mathcal{R} \notin \mathfrak{L}(\mathbb{Q}_1,\mathbb{Q}_2)$. Without loss of generality, we shall assume that $\mathrm{vol}_{\mathbb{Q}_2,r}(\mathcal{R}) \leq (1-\Psi_r(\mathbb{Q}_1)) \cdot d_{\min}(\mathbb{Q}_2)^2$ (otherwise we can work with respect to $\mathcal{R}^c$.) Then, for all $x \in \mathcal{R}$,

$$\int \mathbf{1}\{\|x-y\| \leq r\} \cdot \mathbf{1}\{y \in \mathcal{R}^c\}\, d\mathbb{Q}_2(y) \geq \deg_{\mathbb{Q}_2,r}(x) - \mathbb{Q}_2(\mathcal{R}) \geq \deg_{\mathbb{Q}_2,r}(x) - \frac{\mathrm{vol}_{\mathbb{Q}_2,r}(\mathcal{R})}{d_{\min}(\mathbb{Q}_2)} \geq d_{\min}(\mathbb{Q}_2) \cdot \Psi_r(\mathbb{Q}_2),$$

whence integrating over all $x \in \mathcal{R}$ and dividing by $\mathrm{vol}_{\mathbb{Q}_2,r}(\mathcal{R})$ yields (86). This completes the proof of Proposition 33.

Putting the pieces together. First, we note that

$$\Psi_r(\widetilde{\mathbb{P}}) = \Psi_{\widetilde{\mathbb{P}},r}(\mathrm{supp}(\widetilde{\mathbb{P}})) = \Psi_{\mathbb{P},r}(\mathcal{C}), \quad \text{and} \quad \Psi_r(\widetilde{\mathbb{P}}_n) = \Psi_{\widetilde{\mathbb{P}}_n,r}(\mathrm{supp}(\widetilde{\mathbb{P}}_n)) = \Psi_{n,r}(\mathcal{C}[X]),$$

so that we may apply Proposition 33 to get a lower bound on $\Psi_{n,r}(\mathcal{C}[X])$ in terms of $\Psi_{\mathbb{P},r}(\mathcal{C})$, $\Delta_\infty(\widetilde{\mathbb{P}},\widetilde{\mathbb{P}}_n)$, and $d_{\min}(\widetilde{\mathbb{P}}_n)$. To begin, we recall from Section B.3 that the lower bound on minimum degree,

$$d_{\min}(\widetilde{\mathbb{P}}_n) = \frac{1}{\widetilde{n}}\big(d_{\min}(\widetilde{G}_{n,r}) + 1\big) \geq \frac{1}{\sqrt{2}} d_{\min}(\widetilde{\mathbb{P}}),$$

is satisfied with probability at least $1 - (n+1)\exp\{-nb_2/16\}$. On the other hand, taking

$$b_6 := \frac{1}{B_6}\Psi_r(\widetilde{\mathbb{P}}) \cdot (1 - \Psi_r(\widetilde{\mathbb{P}})) \cdot d_{\min}(\widetilde{\mathbb{P}})^2, \quad \text{and} \quad B_1 := B_5\Big(\min\Big\{b_6, \frac{r}{4(d-1)}\Big\}\Big)^{-1},$$

by (82) (if $d = 1$) or Theorem 32 (if $d \geq 2$) along with (19), we have that

$$\Delta_\infty(\widetilde{\mathbb{P}},\widetilde{\mathbb{P}}_n) \leq B_5 \frac{(\log n)^{p_d}}{\min\{n^{1/2}, n^{1/d}\}} \leq \min\Big\{b_6, \frac{r}{4(d-1)}\Big\} \cdot \delta$$

with probability at least $1 - B_2/n$. Finally, appealing first to Proposition 33 and then to the bounds we have just established on $d_{\min}(\widetilde{\mathbb{P}}_n)$ and $\Delta_\infty(\widetilde{\mathbb{P}},\widetilde{\mathbb{P}}_n)$, we conclude that the sample conductance is lower bounded,

$$\Psi_r(\widetilde{\mathbb{P}}_n) \geq \Psi_r(\widetilde{\mathbb{P}}) \cdot \left(1 - \frac{B_6\Delta_\infty(\widetilde{\mathbb{P}}_n,\widetilde{\mathbb{P}})}{\big(1 - \Psi_r(\widetilde{\mathbb{P}})\big)\cdot\big(d_{\min}(\widetilde{\mathbb{P}}_n)\big)^2}\right) - \frac{B_6\Delta_\infty(\widetilde{\mathbb{P}}_n,\widetilde{\mathbb{P}})}{\big(1 - \Psi_r(\widetilde{\mathbb{P}})\big)\cdot\big(d_{\min}(\widetilde{\mathbb{P}}_n)\big)^2} \geq \Psi_r(\widetilde{\mathbb{P}})(1 - 2\delta)$$

with probability at least $1 - B_2/n - (n+1)\exp\{-nb_2/16\}$, establishing (20) upon taking $b_3 := b_2/16$.

Appendix C. Population functionals for density clusters

In this appendix, we prove Lemma 13 (in Section C.4), Proposition 14 (in Section C.5), and Proposition 15 (in Section C.6), by establishing bounds on the population-level local spread, normalized cut, and conductance of a thickened density cluster $\mathcal{C}_{\lambda,\sigma}$. In these proofs, we make use of some estimates on the volume of spherical caps (given in Section C.1); some isoperimetric inequalities (Section C.2), and some reverse isoperimetric inequalities (Section C.3). Finally, in Section C.7, for the hard case distribution $\mathbb{P}$ defined in (31) and $\mathcal{L}$ defined in (32), we establish bounds on the population-level normalized cut $\Phi_{\mathbb{P},r}(\mathcal{L})$ and local spread $s_{\mathbb{P},r}(\mathcal{X})$; these will be useful in the proof of Theorem 17. Throughout, we write $\nu_d := \nu(B(0,1))$ for the Lebesgue measure of a d-dimensional unit ball.

C.1 Balls, spherical caps, and associated estimates

In this section, we derive lower bounds on the volume of the intersection between two balls in $\mathbb{R}^d$, and the volume of a spherical cap. Results of this type are well-known, but since we could not find exactly the statements we desire, for completeness we also supply proofs. We use the notation $B(x,r)$ for a ball of radius r centered at $x \in \mathbb{R}^d$, and $\mathrm{cap}_r(h)$ for a spherical cap of height h and radius r. Recall that the Lebesgue measure of a spherical cap is

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

where $a = (r-h)^2/r^2$, and

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

is the cumulative distribution function of a Beta(z,w) distribution, evaluated at $1-a$. (Here $\Gamma(\cdot)$ is the gamma function).

Lemma 34 *For any $x, y \in \mathbb{R}^d$ and $r > 0$, it holds that*

$$\nu\big(B(x,r) \cap B(y,r)\big) \geq \nu_d r^d \left(1 - \frac{\|x-y\|}{r}\sqrt{\frac{d+2}{2\pi}}\right). \tag{87}$$

For any $x, y \in \mathbb{R}^d$ and $r, \sigma > 0$ such that $\|x-y\| \leq \sigma$, it holds that,

$$\nu\big(B(x,r) \cap B(y,\sigma)\big) \geq \frac{1}{2}\nu_d r^d \left(1 - \frac{r}{\sigma}\sqrt{\frac{d+2}{2\pi}}\right). \tag{88}$$

Lemma 35 *For any $0 < h \leq r$, and $a = 1 - (2rh - h^2)/r^2$,*

$$\nu\big(\mathrm{cap}_r(h)\big) \geq \frac{1}{2}\nu_d r^d \left(1 - 2\sqrt{a} \cdot \sqrt{\frac{d+2}{2\pi}}\right).$$

An immediate implication of (88) is that for any $x \in \mathcal{C}_{\lambda,\sigma}$,

$$\nu\big(B(x,r) \cap \mathcal{C}_{\lambda,\sigma}\big) \geq \frac{1}{2}\nu_d r^d \left(1 - \frac{r}{\sigma}\sqrt{\frac{d+2}{2\pi}}\right). \tag{89}$$

Proof of Lemma 34. First, we prove (87). The intersection $B(x,r) \cap B(y,r)$ consists of two symmetric spherical caps, each of height $h = r - \frac{\|x-y\|}{2}$. As a result, by Lemma 35 we have

$$\nu\big(B(x,r) \cap B(y,r)\big) \geq \nu_d r^d \big(1 - 2\sqrt{a} \cdot \sqrt{\frac{d+2}{2\pi}}\big)$$

where $a = \|x-y\|^2/(4r^2)$, and the claim follows.

Next we prove (88). Assume that $\|x-y\| = \sigma$, as otherwise if $0 \leq \|x-y\| < \sigma$ the volume of the overlap will only be larger. Then $B(x,r) \cap B(y,\sigma)$ contains a spherical cap of radius r and height $h = r - \frac{r^2}{2\sigma}$, from Lemma 35 we deduce

$$\nu\big(B(x,r) \cap B(y,\sigma)\big) \geq \frac{1}{2}\nu_d r^d \left(1 - 2\sqrt{a} \cdot \sqrt{\frac{d+2}{2\pi}}\right)$$

for $a = (r-h)^2/r^2 = r^2/(4\sigma^2)$, and the claim follows.

Proof of Lemma 35. For any $0 \leq a \leq 1$, we have that

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

The first integral is simply

$$\int_0^1 u^{(d-1)/2}(1-u)^{-1/2}du = \frac{\Gamma\big(\frac{d+1}{2}\big)\Gamma\big(\frac{1}{2}\big)}{\Gamma\big(\frac{d}{2}+1\big)},$$

whereas for all $u \in [0,1]$ and $d \geq 1$, the second integral can be upper bounded as follows:

$$\int_{1-a}^1 u^{(d-1)/2}(1-u)^{-1/2}du \leq \int_{1-a}^1 (1-u)^{-1/2}du = \int_0^a u^{-1/2}du = 2\sqrt{a}.$$

As a result,

$$\nu\big(\mathrm{cap}_r(h)\big) \geq \frac{1}{2}\nu_d r^d\left(1 - 2\sqrt{a}\frac{\Gamma(\frac{d}{2}+1)}{\Gamma(\frac{d+1}{2})\Gamma(\frac{1}{2})}\right) \geq \frac{1}{2}\nu_d r^d\left(1 - 2\sqrt{a}\cdot\sqrt{\frac{d+2}{2\pi}}\right).$$

C.2 Isoperimetric inequalities

Dyer and Frieze (1991) establish the following isoperimetric inequality for convex sets.

Lemma 36 (Isoperimetry of a convex set.) *For any partition $(\mathcal{R}_1, \mathcal{R}_2, \mathcal{R}_3)$ of a convex set $\mathcal{K} \subseteq \mathbb{R}^d$, it holds that*

$$\nu(\mathcal{R}_3) \geq 2\frac{\mathrm{dist}(\mathcal{R}_1, \mathcal{R}_2)}{\mathrm{diam}(\mathcal{K})}\min(\nu(\mathcal{R}_1), \nu(\mathcal{R}_2)).$$

Abbasi-Yadkori (2016) points out that if $\mathcal{S}$ is the image of a convex set under a Lipschitz measure-preserving mapping $g : \mathbb{R}^d \to \mathbb{R}^d$, a similar inequality can be obtained.

Corollary 37 (Isoperimetry of Lipschitz embeddings of convex sets.) *Suppose $\mathcal{S}$ is the image of a convex set $\mathcal{K}$ under a mapping $g : \mathbb{R}^d \to \mathbb{R}^d$ such that*

$$\|g(x) - g(y)\| \leq M \cdot \|x - y\|, \quad \textit{for all } x, y \in \mathcal{K}, \textit{ and } \quad \det(\nabla g(x)) = 1 \quad \textit{for all } x \in \mathcal{K}.$$

Then for any partition $(\Omega_1, \Omega_2, \Omega_3)$ of $\mathcal{S}$,

$$\nu(\Omega_3) \geq 2\frac{\mathrm{dist}(\Omega_1, \Omega_2)}{\mathrm{diam}(\mathcal{K})M}\min(\nu(\Omega_1), \nu(\Omega_2)).$$

C.3 Reverse isoperimetric inequalities

For any set $\mathcal{C} \subseteq \mathbb{R}^d$ and $\sigma > 0$, let $\mathcal{C}_\sigma := \{x : \mathrm{dist}(x, \mathcal{C}) \leq \sigma\}$. We begin with an upper bound on the volume of $\mathcal{C}_{\sigma+\delta}$ as compared to $\mathcal{C}_\sigma$.

Lemma 38 *For any bounded set $\mathcal{C} \subseteq \mathbb{R}^d$ and $\sigma, \delta > 0$, it holds that*

$$\nu(\mathcal{C}_{\sigma+\delta}) \leq \nu(\mathcal{C}_\sigma) \cdot \Big(1 + \frac{\delta}{\sigma}\Big)^d. \tag{90}$$

Lemma 38 is a reverse isoperimetric inequality. To see this, note that if $\delta \leq \sigma/d$ then $\big(1+\delta/\sigma\big)^d \leq 1 + d \cdot \delta/(\sigma - d\delta)$, and we deduce from (90) that

$$\nu(\mathcal{C}_{\sigma+\delta} \setminus \mathcal{C}_\sigma) = \nu(\mathcal{C}_{\sigma+\delta}) - \nu(\mathcal{C}_\delta) \leq \frac{d\delta}{\sigma - d\delta} \cdot \nu(\mathcal{C}_\sigma). \tag{91}$$

We use (91) along with Assumption (A4) to derive a density-weighted reverse isoperimetric inequality.

Lemma 39 *Let $\mathcal{C}_{\lambda,\sigma}$ satisfy Assumption (A1) and (A4) for some θ, γ and λ_σ. Then for any $0 < r \leq \sigma/d$, it holds that*

$$\mathbb{P}\big(\mathcal{C}_{\lambda,\sigma+r} \setminus \mathcal{C}_{\lambda,\sigma}\big) \leq \Big(1 + \frac{dr}{\sigma - dr}\Big) \cdot \frac{dr}{\sigma} \cdot \Big(\lambda_\sigma - \theta\frac{r^\gamma}{\gamma+1}\Big) \cdot \nu(\mathcal{C}_{\lambda,\sigma}). \tag{92}$$

Proof of Lemma 38. Fix $\delta' > 0$, and take $\epsilon = \delta + \delta'$. We will show that

$$\nu(\mathcal{C}_{\sigma+\epsilon}) \leq \nu(\mathcal{C}_\sigma) \cdot \Big(1 + \frac{\epsilon}{\sigma}\Big)^d, \tag{93}$$

whence taking a limit as $\delta' \to 0$ yields the claim.

To show (93), we need to construct a particular disjoint covering $\mathcal{A}_1(\sigma+\epsilon), \ldots, \mathcal{A}_N(\sigma+\epsilon)$ of $\mathcal{C}_{\sigma+\delta}$. To do so, we first take a finite set of points $x_1, \ldots, x_N$ such that the net $B(x_1, \sigma+\epsilon), \ldots, B(x_N, \sigma+\epsilon)$ covers $\mathcal{C}_{\sigma+\delta}$. Note that such a covering exists for some finite $N = N(\epsilon)$ because $\mathcal{C}_{\sigma+\delta}$ is bounded, and the closure of $\mathcal{C}_{\sigma+\delta}$ is thus a compact subset of $\cup_{x \in \mathcal{C}} B(x, \sigma+\epsilon)$. Defining $\mathcal{A}_1(s), \ldots, \mathcal{A}_N(s)$ for a given $s > 0$ to be

$$\mathcal{A}_1(s) := B(x_1, s), \quad \text{and} \quad \mathcal{A}_{j+1}(s) := B(x_{j+1}, s) \setminus \bigcup_{i=1}^{j} B(x_i, s) \quad \text{for } j = 1, \ldots, N-1,$$

we have that $\mathcal{A}_1(\sigma+\epsilon), \ldots, \mathcal{A}_N(\sigma+\epsilon)$ is a disjoint covering of $\mathcal{C}_{\sigma+\delta}$, and so $\nu(\mathcal{C}_{\sigma+\delta}) \leq \sum_{j=1}^N \nu\big(\mathcal{A}_j(\sigma+\epsilon)\big)$.

We claim that for all $j = 1, \ldots, N$, the function $s \mapsto \nu\big(\mathcal{A}_j(s)\big)/\nu\big(B(x_j, s)\big)$ is monotonically non-increasing in s. Once this claim is verified, it follows that

$$\nu(\mathcal{A}_j(\sigma+\epsilon)) = \nu\big(B(x_j, \sigma+\epsilon)\big) \cdot \frac{\nu\big(\mathcal{A}_j(\sigma+\epsilon)\big)}{\nu\big(B(x_j, \sigma+\epsilon)\big)} \leq \Big(1+\frac{\epsilon}{\sigma}\Big)^d \cdot \nu\big(B(x_j, \sigma)\big) \cdot \frac{\nu\big(\mathcal{A}_j(\sigma)\big)}{\nu\big(B(x_j, \sigma)\big)} = \Big(1+\frac{\epsilon}{\sigma}\Big)^d \cdot \nu\big(\mathcal{A}_j(\sigma)$$

and summing over j, we see that

$$\nu(\mathcal{C}_{\sigma+\delta}) \leq \sum_{j=1}^N \nu\big(\mathcal{A}_j(\sigma+\epsilon)\big) \leq \Big(1+\frac{\epsilon}{\sigma}\Big)^d \cdot \sum_{j=1}^N \nu\big(\mathcal{A}_j(\sigma)\big) \leq \Big(1+\frac{\epsilon}{\sigma}\Big)^d \nu(\mathcal{C}_\sigma).$$

The last inequality follows since $\mathcal{A}_1(\sigma), \ldots, \mathcal{A}_N(\sigma)$ are disjoint subsets of the closure of $\mathcal{C}_\sigma$.

It remains to verify that $s \mapsto \nu\bigl(\mathcal{A}_j(s)\bigr)/\nu\bigl(B(x_j, s)\bigr)$ is monotonically non-increasing. For any $0 < s < t$ and $j = 1, \ldots, N$, suppose $x \in \mathcal{A}_j(T) - x_j$, meaning $x \in B(0, t)$ and $x \notin B(x_i - x_j, t)$ for any $i = 1, \ldots, j-1$. Thus $(s/t)x \in B(0, s)$, and

$$\|(s/t)x - (x_i - x_j)\| \geq \|x - (x_i - x_j)\| - \|x - (s/t)x\| > t - (1 - s/t)\|x\| \geq t - (1 - s/t)t = s,$$

or in other words $(s/t)x \notin B(x_i - x_j, s)$ for any $i = 1, \ldots, j-1$. Consequently,

$$\Bigl(\mathcal{A}_j(t) - x_j\Bigr) \subset \frac{t}{s} \cdot \Bigl(\mathcal{A}_j(s) - x_j\Bigr),$$

and applying $\nu(\cdot)$ to both sides yields the claim.

Proof of Lemma 39. Fix $k \in \mathbb{N}$. To establish (92), we partition $\mathcal{C}_{\lambda,\sigma+r} \setminus \mathcal{C}_{\lambda,\sigma}$ into thin tubes $\mathcal{T}_1, \ldots, \mathcal{T}_k$, with the jth tube $\mathcal{T}_j$ defined as $\mathcal{T}_j := \mathcal{C}_{\lambda,\sigma+jr/k} \setminus \mathcal{C}_{\lambda,\sigma+(j-1)r/k}$. We upper bound the Lebesgue measure of each tube $\mathcal{T}_j$ using (91):[6]

$$\nu(\mathcal{T}_j) \leq \frac{dr/k}{\sigma - dr/k}\nu(\mathcal{C}_{\lambda,\sigma+(j-1)r/k}) \leq \frac{dr/k}{\sigma - dr/k}\nu(\mathcal{C}_{\lambda,\sigma+r}) \leq \Bigl(1 + \frac{dr}{\sigma - dr}\Bigr) \cdot \frac{dr/k}{\sigma - dr/k} \cdot \nu(\mathcal{C}_{\lambda,\sigma}),$$

and the maximum density within each tube using (A4):

$$\max_{x \in \mathcal{T}_j} f(x) \leq \lambda_\sigma - \theta\Bigl(\frac{j-1}{k}r\Bigr)^\gamma;$$

combining these upper bounds, we see that

$$\mathbb{P}\bigl(\mathcal{C}_{\lambda,\sigma+r} \setminus \mathcal{C}_{\lambda,\sigma}\bigr) = \sum_{j=1}^{k} \mathbb{P}(\mathcal{T}_j) \leq \Bigl(1 + \frac{dr}{\sigma - dr}\Bigr) \cdot \frac{dr/k}{\sigma - dr/k} \cdot \nu(\mathcal{C}_{\lambda,\sigma}) \cdot \Biggl(\sum_{j=0}^{k-1} \lambda_\sigma - \theta r^\gamma \Bigl(\frac{j}{k}\Bigr)^\gamma\Biggr). \tag{94}$$

Treating the sum in the previous expression as a Riemann sum of a non-increasing function evaluated at $0, \ldots, k-1$ gives the upper bound

$$\sum_{j=0}^{k-1} \lambda_\sigma - \theta r^\gamma \Bigl(\frac{j}{k}\Bigr)^\gamma \leq \lambda_\sigma + \int_0^{k-1} \Bigl(\lambda_\sigma - \theta r^\gamma \Bigl(\frac{x}{k}\Bigr)^\gamma\Bigr)\,dx \leq k\lambda_\sigma + (k-1)\frac{\theta r^\gamma}{\gamma+1}\Bigl(\frac{k-1}{k}\Bigr)^\gamma,$$

and plugging back in to (94), we obtain

$$\mathbb{P}\bigl(\mathcal{C}_{\lambda,\sigma+r} \setminus \mathcal{C}_{\lambda,\sigma}\bigr) \leq \Bigl(1 + \frac{dr}{\sigma - dr}\Bigr) \cdot \frac{dr}{\sigma - dr/k}\nu(\mathcal{C}_{\lambda,\sigma}) \cdot \Biggl(\lambda - \frac{\theta r^\gamma}{\gamma+1} \cdot \Bigl(\frac{k-1}{k}\Bigr)^{\gamma+1}\Biggr).$$

The above inequality holds for any $k \in \mathbb{N}$, and taking the limit of the right hand side as $k \to \infty$ yields the claim.

6. Note that $\mathcal{C}$ must be bounded, since the density $f(x) \geq \lambda_\sigma > 0$ for all $x \in \mathcal{C}$.

C.4 Proof of Lemma 13

The population-level local spread of $\mathcal{C}_{\lambda,\sigma}$ is

$$s_{\mathbb{P},r}(\mathcal{C}_{\lambda,\sigma}) = \frac{\big(d_{\min}(\widetilde{\mathbb{P}})\big)^2}{\mathrm{vol}_{\widetilde{\mathbb{P}},r}(\mathcal{C}_{\lambda},\sigma)},$$

where we recall that $\widetilde{\mathbb{P}}(\mathcal{S}) = \frac{\mathbb{P}(\mathcal{S}\cap\mathcal{C}_{\lambda,\sigma})}{\mathbb{P}(\mathcal{C}_{\lambda,\sigma})}$ for Borel sets $\mathcal{S}$, and $d_{\min}(\widetilde{\mathbb{P}}) := \min_{x\in\mathcal{C}_{\lambda,\sigma}}\{\deg_{\widetilde{\mathbb{P}},r}(x)\}^2$. To lower bound $s_{\mathbb{P},r}(\mathcal{C}_{\lambda,\sigma})$, we first lower bound $d_{\min}(\widetilde{\mathbb{P}})$, and then upper bound $\mathrm{vol}_{\widetilde{\mathbb{P}},r}(\mathcal{C}_{\lambda},\sigma)$. Using the lower bound $f(x) \geq \lambda_\sigma$ for all $x \in \mathcal{C}_{\lambda,\sigma}$ stipulated in (A3), we deduce that

$$\begin{aligned}
d_{\min}(\widetilde{\mathbb{P}}) &= \min_{x\in\mathcal{C}_{\lambda,\sigma}}\Big\{\int \mathbf{1}\{\|x-y\|\leq r\}\,d\widetilde{\mathbb{P}}(y)\Big\} \\
&\geq \frac{\lambda_\sigma}{\mathbb{P}(\mathcal{C}_{\lambda,\sigma})}\cdot\min_{x\in\mathcal{C}_{\lambda,\sigma}}\Big\{\int_{\mathcal{C}_{\lambda,\sigma}} \mathbf{1}\{\|x-y\|\leq r\}\,dy\Big\} \\
&\geq \frac{\lambda_\sigma}{\mathbb{P}(\mathcal{C}_{\lambda,\sigma})}\cdot\frac{1}{2}\nu_d r^d\cdot\Big(1-\frac{r}{\sigma}\sqrt{\frac{d+2}{2\pi}}\Big),
\end{aligned}$$

where the final inequality follows from Lemma 34.

On the other hand, using the upper bound $f(x) \leq \Lambda_\sigma$ for all $x \in \mathcal{C}_{\lambda,\sigma}$, we deduce that

$$\begin{aligned}
\mathrm{vol}_{\widetilde{\mathbb{P}},r}(\mathcal{C}_{\lambda,\sigma}) &= \iint \mathbf{1}\{\|x-y\|\leq r\}\,d\widetilde{\mathbb{P}}(y)\,d\widetilde{\mathbb{P}}(x) \\
&\leq \frac{\Lambda_\sigma^2}{\mathbb{P}(\mathcal{C}_{\lambda,\sigma})^2}\cdot\int_{\mathcal{C}_{\lambda,\sigma}}\int_{\mathcal{C}_{\lambda,\sigma}} \mathbf{1}\{\|x-y\|\leq r\}\,dy\,dx \\
&\leq \frac{\Lambda_\sigma^2}{\mathbb{P}(\mathcal{C}_{\lambda,\sigma})^2}\cdot\nu_d r^d\cdot\nu(\mathcal{C}_{\lambda,\sigma}) \\
&\leq \frac{\Lambda_\sigma^2}{\mathbb{P}(\mathcal{C}_{\lambda,\sigma})^2}\cdot\nu_d^2 r^d\cdot\Big(\frac{\rho}{2}\Big)^d;
\end{aligned}$$

the final inequality follows from (A5), which along with the isodiametric inequality for convex sets (Gruber, 2007) implies $\nu(\mathcal{C}_{\lambda,\sigma}) = \nu(\mathcal{K}) \leq \nu_d(\rho/2)^d$. The claim of Lemma 13 follows.

C.5 Proof of Proposition 14

By Assumption (A6), we have that $\Phi_{\mathbb{P},r}(\mathcal{C}_{\lambda,\sigma}) = \mathrm{cut}_{\mathbb{P},r}(\mathcal{C}_{\lambda,\sigma})/\mathrm{vol}_{\mathbb{P},r}(\mathcal{C}_{\lambda,\sigma})$, and to prove Proposition 14 we must therefore upper bound $\mathrm{cut}_{\mathbb{P},r}(\mathcal{C}_{\lambda,\sigma})$ and lower bound $\mathrm{vol}_{\mathbb{P},r}(\mathcal{C}_{\lambda,\sigma})$.

Let $\mathcal{C}_{\lambda,\sigma+r} = \{x : \mathrm{dist}(x,\mathcal{C}_\lambda) \leq \sigma + r\}$. We upper bound $\mathrm{cut}_{\mathbb{P},r}(\mathcal{C}_{\lambda,\sigma})$ in terms of the probability mass of $\mathcal{C}_{\lambda,\sigma+r}\setminus\mathcal{C}_{\lambda,\sigma}$:

$$\begin{aligned}
\mathrm{cut}_{\mathbb{P},r}(\mathcal{C}_{\lambda,\sigma}) &= \iint \mathbf{1}\{\|x-y\|\leq r\}\cdot\mathbf{1}\{x\in\mathcal{C}_{\lambda,\sigma}\}\cdot\mathbf{1}\{y\notin\mathcal{C}_{\lambda,\sigma}\}\,d\mathbb{P}(y)\,d\mathbb{P}(x) \\
&\leq \iint \mathbf{1}\{\|x-y\|\leq r\}\cdot\mathbf{1}\{x\in\mathcal{C}_{\lambda,\sigma}\}\cdot\mathbf{1}\{y\in\mathcal{C}_{\lambda,\sigma+r}\setminus\mathcal{C}_{\lambda,\sigma}\}\,d\mathbb{P}(y)\,d\mathbb{P}(x) \\
&\leq \lambda\nu_d r^d\cdot\mathbb{P}\big(\mathcal{C}_{\lambda,\sigma+r}\setminus\mathcal{C}_{\lambda,\sigma}\big).
\end{aligned}$$

On the other hand, using the lower bound $f(x) \geq \lambda_\sigma$ for all $x \in \mathcal{C}_{\lambda,\sigma}$, we lower bound $\mathrm{vol}_{\mathbb{P},r}(\mathcal{C}_{\lambda,\sigma})$ in terms of the Lebesgue measure of $\mathcal{C}_{\lambda,\sigma}$:

$$\begin{aligned}
\mathrm{vol}_{\mathbb{P},r}(\mathcal{C}_{\lambda,\sigma}) &= \iint \mathbf{1}\{\|x-y\| \leq r\} \cdot \mathbf{1}\{x \in \mathcal{C}_{\lambda,\sigma}\}\, d\mathbb{P}(y)\, d\mathbb{P}(x) \\
&\geq \lambda_\sigma^2 \cdot \iint \mathbf{1}\{\|x-y\| \leq r\} \cdot \mathbf{1}\{x, y \in \mathcal{C}_{\lambda,\sigma}\}\, dy\, dy \\
&\geq \lambda_\sigma^2 \cdot \frac{1}{2}\nu_d r^d \cdot \Bigl(1 - \frac{r}{\sigma}\sqrt{\frac{d+2}{2\pi}}\Bigr) \cdot \nu(\mathcal{C}_{\lambda,\sigma}).
\end{aligned}$$

The claim of Proposition 14 follows upon using Lemma 39 to upper bound $\mathbb{P}\bigl(\mathcal{C}_{\lambda,\sigma+r} \setminus \mathcal{C}_{\lambda,\sigma}\bigr)$.

C.6 Proof of Proposition 15

The following Lemma lower bounds the population-level uniform conductance $\Psi_{\nu,r}(\mathcal{C}_{\lambda,\sigma})$. We note that for convex sets, results of this type are well known (see e.g. Vempala (2005) and references therein).

Lemma 40 *Suppose $\mathcal{C}_{\lambda,\sigma}$ satisfies Assumption (A5) with respect to some $\rho \in (0,\infty)$ and $M \in [1,\infty)$. For any $0 < r \leq \sigma \cdot \sqrt{2\pi/(d+2)}$, it holds that*

$$\Psi_{\nu,r}(\mathcal{C}_{\lambda,\sigma}) \geq \Bigl(1 - \frac{r}{4\rho M}\Bigr) \cdot \Bigl(1 - \frac{r}{\sigma}\sqrt{\frac{d+2}{2\pi}}\Bigr)^2 \cdot \frac{\sqrt{2\pi}}{36} \cdot \frac{r}{\rho M\sqrt{d+2}}. \tag{95}$$

Noting that $\Psi_{\mathbb{P},r}(\mathcal{C}_{\lambda,\sigma}) \geq \Psi_{\nu,r}(\mathcal{C}_{\lambda,\sigma}) \cdot \lambda_\sigma^2/\Lambda_\sigma^2$, Proposition 15 follows from (95).

Proof of Lemma 40. For ease of notation, throughout this proof we write $\widetilde{\nu}$ for the uniform probability measure over $\mathcal{C}_{\lambda,\sigma}$, put $\ell\nu_d r^d := \min_{x \in \mathcal{C}_{\lambda,\sigma}} \nu(B(x,r) \cap \mathcal{C}_\sigma)$ and $a := r/(2\rho M)$.

Let $\mathcal{S}$ be an arbitrary measurable subset of $\mathcal{C}_{\lambda,\sigma}$, and let $\mathcal{R} = \mathcal{C}_{\lambda,\sigma} \setminus \mathcal{S}$. For a given $\delta \in (0,1)$, let the δ-interior of $\mathcal{S}$ be

$$\mathcal{S}^\delta := \{x \in \mathcal{S} : \nu\bigl(B(x,r) \cap \mathcal{R}\bigr) \leq \ell\delta\nu_d r^d\};$$

define $\mathcal{R}^\delta$ likewise, and let $\mathcal{B}^\delta = \mathcal{C}_{\lambda,\sigma} \setminus (\mathcal{S}^\delta \cup \mathcal{R}^\delta)$ consist of the remaining boundary points. As is standard (see for example Dyer and Frieze (1991); Lovász and Simonovits (1990)), the proof of Lemma 40 uses several inequalities to lower bound the normalized cut $\Phi_{\widetilde{\nu},r}(\mathcal{S})$.

- **Bounds on cut and volume.** We can lower bound $\mathrm{cut}_{\widetilde{\nu},r}(\mathcal{S})$ as follows:

$$\begin{aligned}
\nu(\mathcal{C}_{\lambda,\sigma})^2 \cdot \mathrm{cut}_{\widetilde{\nu},r}(\mathcal{S}) &= \int_{\mathcal{S}}\int_{\mathcal{R}} \mathbf{1}(\|x-y\| \leq r)\, dy\, dx \\
&= \frac{1}{2}\Bigl(\int_{\mathcal{S}}\int_{\mathcal{R}} \mathbf{1}(\|x-y\| \leq r)\, dy\, dx + \int_{\mathcal{R}}\int_{\mathcal{S}} \mathbf{1}(\|x-y\| \leq r)\, dy\, dx\Bigr) \\
&\geq \frac{1}{2}\delta\ell\nu_d r^d \cdot \nu(\mathcal{B}^\delta).
\end{aligned}$$

We can upper bound $\mathrm{vol}_{\widetilde{\nu},r}(\mathcal{S})$ as follows:

$$\nu(\mathcal{C}_{\lambda,\sigma})^2 \cdot \mathrm{vol}_{\widetilde{\nu},r}(\mathcal{S}) = \int_{\mathcal{C}_{\lambda,\sigma}} \int_{\mathcal{S}} \mathbf{1}(\|x-y\| \leq r)\, dy\, dx \leq \nu_d r^d \nu(\mathcal{S})$$

and likewise for $\mathrm{vol}_{\widetilde{\nu},r}(\mathcal{R})$. Therefore,

$$\Phi_{\widetilde{\nu},r}(\mathcal{S}) \geq \frac{\delta\ell \cdot \nu(\mathcal{B}^\delta)}{2 \cdot \min\{\nu(\mathcal{S}), \nu(\mathcal{R})\}}. \tag{96}$$

- **Isoperimetric inequality.** Applying Corollary 37, we have that

$$\nu(\mathcal{B}^\delta) \geq \frac{2 \cdot \mathrm{dist}(\mathcal{S}^\delta, \mathcal{R}^\delta)}{\rho M} \cdot \min\bigl\{\nu(\mathcal{S}^\delta), \nu(\mathcal{R}^\delta)\bigr\}. \tag{97}$$

- **Lebesgue measure of δ-interiors.** Suppose $\nu(\mathcal{S}^\delta) \leq (1-a) \cdot \nu(\mathcal{S})$ or $\nu(\mathcal{R}^\delta) \leq (1-a) \cdot \nu(\mathcal{R})$. Then $\nu(\mathcal{B}^\delta) \geq a \cdot \min\{\nu(\mathcal{S}), \nu(\mathcal{R})\}$, and combined with (96) we have that $\Phi_{\widetilde{\nu},r}(\mathcal{S}) \geq \delta a \ell/2$. Otherwise,

$$\min\bigl\{\nu(\mathcal{S}^\delta), \nu(\mathcal{R}^\delta)\bigr\} \geq (1-a) \cdot \min\bigl\{\nu(\mathcal{S}), \nu(\mathcal{R})\bigr\}. \tag{98}$$

- **Distance between δ-interiors.** For any $x \in \mathcal{S}^\delta$ and $y \in \mathcal{R}^\delta$, we have that

$$\begin{aligned}
\nu\bigl(B(x,r) \cap B(y,r)\bigr) &= \nu\bigl(B(x,r) \cap B(y,r) \cap \mathcal{R}\bigr) + \nu\bigl(B(x,r) \cap B(y,r) \cap \mathcal{S}\bigr) + \nu\bigl(B(x,r) \cap B(y,r) \cap \mathcal{C}_{\lambda,\sigma}^c\bigr) \\
&\leq \nu\bigl(B(x,r) \cap \mathcal{R}\bigr) + \nu\bigl(B(y,r) \cap \mathcal{S}\bigr) + \nu\bigl(B(x,r) \cap \mathcal{C}_{\lambda,\sigma}^c\bigr) \\
&\leq \bigl(2\ell\delta + (1-\ell)\bigr) \cdot \nu_d r^d.
\end{aligned}$$

It follows from (87) that

$$\|x-y\| \geq \frac{r}{\nu_d r^d} \cdot \Bigl(\nu_d r^d - \nu\bigl(B(x,r) \cap B(y,r)\bigr)\Bigr) \cdot \sqrt{\frac{2\pi}{d+2}} \geq r \cdot \ell \cdot (1-2\delta) \cdot \sqrt{\frac{2\pi}{d+2}},$$

and taking the infimum over all $x \in \mathcal{S}^\delta$ and $y \in \mathcal{R}^\delta$, we have

$$\mathrm{dist}(\mathcal{S}^\delta, \mathcal{R}^\delta) \geq r \cdot \ell \cdot (1-2\delta) \cdot \sqrt{\frac{2\pi}{d+2}}. \tag{99}$$

Combining (96)-(99) and taking $\delta = 1/3$ implies that

$$\Phi_{\widetilde{\nu},r}(\mathcal{S}) \geq \min\Biggl\{(1-a) \cdot \frac{r}{\rho M} \cdot \frac{\ell^2}{9} \cdot \sqrt{\frac{2\pi}{d+2}}, \frac{a\ell}{6}\Biggr\}$$

and the claim follows from (89), which implies that $\ell \geq 1/2 \cdot (1 - r/\sigma)\sqrt{2\pi/(d+2)}$.

C.7 Population functionals, hard case

Let $\mathbb{P}$ be the hard case distribution over rectangular domain $\mathcal{X}$, defined as in (31), and $\mathcal{L}$ the lower half of $\mathcal{X}$. Suppose $r \in (0, \sigma/2)$. Then the population normalized cut $\Phi_{\mathbb{P},r}(\mathcal{L})$ is upper bounded,

$$\Phi_{\mathbb{P},r}(\mathcal{L}) \leq \frac{8}{3} \cdot \frac{r}{\rho}. \tag{100}$$

and the population local spread $s_{\mathbb{P},r}(\mathcal{X})$ is lower bounded,

$$s_{\mathbb{P},r}(\mathcal{X}) \geq \frac{\pi r^2 \epsilon^2}{2\rho\sigma} \tag{101}$$

Proof of (100). Noting that $\mathrm{vol}_{\mathbb{P},r}(\mathcal{L}) = \mathrm{vol}_{\mathbb{P},r}(\mathcal{X} \setminus \mathcal{L})$, it suffices to upper bound $\mathrm{cut}_{\mathbb{P},r}(\mathcal{L})$ and lower bound $\mathrm{vol}_{\mathbb{P},r}(\mathcal{L})$. Note that for any $x = (x_1, x_2) \in \mathcal{L}$, if $x_2 \leq -r$ the ball $B(x,r)$ and the set $\mathcal{X} \setminus \mathcal{L}$ are disjoint. As a result,

$$\mathrm{cut}_{\mathbb{P},r}(\mathcal{L}) \leq \mathbb{P}\Big(\big\{x \in \mathcal{X} : x_2 \in (-r, 0)\big\}\Big) \cdot d_{\max}(\mathbb{P}) \leq \frac{r}{2\rho} \cdot \frac{\pi r^2}{2\sigma\rho}.$$

On the other hand, noting that $\deg_{\mathbb{P},r}(x) \geq \frac{\pi r^2}{2\sigma\rho}$ for all $x \in \mathcal{C}^{(1)}$ such that $\mathrm{dist}(x, \partial\mathcal{C}^{(1)}) > r$, we have

$$\begin{aligned}
\mathrm{vol}_{\mathbb{P},r}(\mathcal{L}) &\geq \mathbb{P}\Big(\big\{x \in \mathcal{C}^{(1)} \cap \mathcal{L} : \mathrm{dist}(x, \partial\mathcal{C}^{(1)}) > r\big\}\Big) \cdot \frac{\pi r^2}{2\sigma\rho} \\
&= \frac{(\sigma - 2r)(\rho - r)}{2\sigma\rho} \cdot \frac{\pi r^2}{2\sigma\rho} \\
&\geq \frac{3}{16} \cdot \frac{\pi r^2}{2\sigma\rho}
\end{aligned}$$

where the last inequality follows since $r \leq \frac{1}{4}\sigma \leq \frac{1}{4}\rho$.

Proof of (101). The statement follows since

$$d_{\min}(\mathbb{P}) \geq \frac{\pi r^2}{2} \cdot \min_{x \in \mathcal{X}} f(x) = \frac{\pi r^2}{2} \cdot \frac{\epsilon}{\rho\sigma},$$

and

$$\mathrm{vol}_{\mathbb{P},r}(\mathcal{X}) \leq d_{\max}(\mathbb{P}) \leq \frac{\pi r^2}{2\sigma\rho}.$$

Appendix D. Proof of major theorems

We now prove the three major theorems of our paper: Theorem 12 (in Section D.1), Theorem 16, and Theorem 17. Throughout, we use the notation $\widetilde{n} = |\mathcal{C}[X]|$ and $\widetilde{G}_{n,r} = G_{n,r}\big[\mathcal{C}[X]\big]$ as defined above.

D.1 Proof of Theorem 12

We begin by recalling some probabilistic estimates needed for the proof of Theorem 12, along with the probability with which they hold.

Probabilistic estimates. Throughout the proof of Theorem 12, we will assume (i) that the inequalities (16)-(20) are satisfied; (ii) that the volume of $\mathcal{C}[X]$ is upper and lower bounded,

$$(1-\delta)\cdot \mathrm{vol}_{\mathbb{P},r}(\mathcal{C}) \leq \frac{1}{n(n-1)}\mathrm{vol}_{n,r}(\mathcal{C}[X]) \leq (1+\delta)\cdot \mathrm{vol}_{\mathbb{P},r}(\mathcal{C}); \tag{102}$$

(iii) that the number of sample points in $\mathcal{C}$ is lower bounded,

$$\widetilde{n} \geq (1-\delta)\cdot n\cdot \mathbb{P}(\mathcal{C}) \overset{(17)}{\Longrightarrow} \widetilde{n}-1 \geq (1-\delta)^2\cdot n\cdot \mathbb{P}(\mathcal{C}); \tag{103}$$

and finally (iv) that the minimum and maximum degree of $\widetilde{G}_{n,r}$ are lower and upper bounded respectively,

$$\frac{1}{\widetilde{n}-1}d_{\min}(\widetilde{G}_{n,r}) \geq (1-\delta)\cdot d_{\min}(\widetilde{\mathbb{P}}), \quad \text{and} \quad \frac{1}{\widetilde{n}-1}d_{\max}(\widetilde{G}_{n,r}) \leq (1+\delta)\cdot d_{\max}(\widetilde{\mathbb{P}}). \tag{104}$$

By Propositions 9 and 10, and Lemmas 29-31, these inequalities are satisfied with probability at least $1 - B_2/n - 4\exp\{-b_1\delta^2 n\} - (2n+2)\exp\{-b_2\delta^2 n\} - (n+1)\exp\{-b_3 n\}$.

Proof of Theorem 12. We use Lemma 7 to upper bound $\Delta(\widehat{C},\mathcal{C}[X])$. In order to do so, we must verify that the tuning parameters α and (L,U) satisfy the condition (11) of this lemma, i.e. that $\alpha \leq 1/\big(2\tau_\infty(\widetilde{G}_{n,r})\big)$ and $U \leq 1/\big(5\mathrm{vol}_{n,r}(\mathcal{C}[X])\big)$. In order to verify the upper bound on α, we will use Proposition 8 to upper bound $\tau_\infty(\widetilde{G}_{n,r})$, which we may validly apply because

$$\frac{d_{\max}(\widetilde{G}_{n,r})}{\big(d_{\min}(\widetilde{G}_{n,r})\big)^2} \leq \frac{(1+\delta)}{(1-\delta)^2}\cdot\frac{d_{\max}(\widetilde{\mathbb{P}})}{(\widetilde{n}-1)\cdot\big(d_{\min}(\widetilde{\mathbb{P}})\big)^2} \leq \frac{(1+\delta)}{(1-\delta)^4}\cdot\frac{d_{\max}(\widetilde{\mathbb{P}})}{n\mathbb{P}(\mathcal{C})\big(d_{\min}(\widetilde{\mathbb{P}})\big)^2} \leq \frac{1}{16}.$$

The last inequality in the above follows by taking $B_3 := 16\cdot d_{\max}(\widetilde{\mathbb{P}})/\big(d_{\min}(\widetilde{\mathbb{P}})\big)^2$ in (24).

Therefore by Proposition 8, along with inequalities (18) and (20) and the initialization conditions (21) and (22), we have that $\alpha \leq 1/45 \wedge 1/\big(2\tau_\infty(\widetilde{G}_{n,r})\big)$. On the other hand, by the upper bound on $\mathrm{vol}_{n,r}(\mathcal{C}[X])$ given in (102) and the initialization condition (21), we have that $U \leq 1/\big(5\mathrm{vol}_{n,r}(\mathcal{C}[X])\big)$. In summary, we have confirmed that the condition (11) is satisfied.

Invoking Lemma 7, we conclude that there exists a set $\mathcal{C}[X]^g \subset \mathcal{C}[X]$ of volume at least $\mathrm{vol}_{n,r}(\mathcal{C}[X]^g) \geq \mathrm{vol}_{n,r}(\mathcal{C}[X])/2$, such that for any $\beta \in (L,U)$,

$$\mathrm{vol}_{n,r}(S_{\beta,v} \vartriangle \mathcal{C}[X]) \leq 60\cdot\frac{\Phi_{n,r}(\mathcal{C}[X])}{\alpha L} \leq 60\frac{(1+2\delta)}{(1-4\delta)^2}\cdot\frac{\Phi_{n,r}(\mathcal{C}[X])}{\alpha_{\mathbb{P},r}(\mathcal{C},\delta)}\cdot n(n-1)\mathrm{vol}_{\mathbb{P},r}(\mathcal{C})$$

Noting that $\widehat{C} = S_{\beta,v}$ for some $\beta \in (L,U)$, the claimed upper bound (25) on $\Delta(\widehat{C},\mathcal{C}[X])$ then follows from the upper bound (16) on $\Phi_{n,r}(\mathcal{C}[X])$ and the upper bound on $\mathrm{vol}_{\mathbb{P},r}(\mathcal{C})$ in (102).

D.2 Proof of Theorem 16

From Theorem 12, we have that with probability $1-B_2/n-4\exp\{-b_1\delta^2 n\}-(2n+2)\exp\{-b_2\delta^2 n\}-(n+1)\exp\{-b_3 n\}$, there exists a set $\mathcal{C}_{\lambda,\sigma}[X]^g \subset \mathcal{C}_{\lambda,\sigma}[X]$ of volume at least $\mathrm{vol}_{n,r}(\mathcal{C}_{\lambda,\sigma}[X]^g) \geq \mathrm{vol}_{n,r}(\mathcal{C}_{\lambda,\sigma}[X])/2$, such that

$$\begin{aligned}\frac{\Delta(\widehat{C},\mathcal{C}_{\lambda,\sigma}[X])}{\mathrm{vol}_{n,r}(\mathcal{C}_{\lambda,\sigma}[X])} &\leq 60 \cdot \frac{(1+3\delta)(1+2\delta)}{(1-4\delta)^2(1-2\delta)} \cdot \frac{\Phi_{\mathbb{P},r}(\mathcal{C}_{\lambda,\sigma})}{\alpha_{\mathbb{P},r}(\mathcal{C}_{\lambda,\sigma},\delta)} \\ &\leq \frac{1020}{\ln(2)} \cdot \frac{(1+3\delta)(1+2\delta)}{(1-4\delta)^2(1-2\delta)} \cdot \frac{\Phi_{\mathbb{P},r}(\mathcal{C}_{\lambda,\sigma})}{\Psi_{\mathbb{P},r}(\mathcal{C}_{\lambda,\sigma})^2} \cdot \ln^2\Bigl(\frac{32}{(1-3\delta)s_{\mathbb{P},r}(\mathcal{C})}\Bigr)\end{aligned}$$

The claimed upper bound (30) on $\Delta(\widehat{C},\mathcal{C}_{\lambda,\sigma}[X])$ then follows from the bounds (27)-(29) on the population-level local spread, normalized cut, and conductance of $\mathcal{C}_{\lambda,\sigma}$, noting that the condition $r \leq \sigma/(4d)$ implies that $(1-r/(4\rho L)) \geq 1-1/16$ and $1-r/\sigma\cdot\sqrt{(d+2)/(2\pi)} \geq 1-1/\sqrt{32}$, and taking

$$C_{1,\delta} := \frac{1175040}{\pi\cdot\ln(2)\cdot(1-1/16)^2\cdot(1-1/\sqrt{32})^4}\cdot\frac{(1+3\delta)(1+2\delta)}{(1-4\delta)^2(1-2\delta)}, \text{ and } C_{2,\delta} := \frac{144}{(1-1/\sqrt{32})}\cdot\frac{1}{1-3\delta}.$$

D.3 Proof of Theorem 17

We start by defining some constants, to make our proof statements easier to digest. Put

$$\begin{aligned} C_{3,\delta} &:= \frac{288(1+\delta)}{(1-\delta)}\sqrt{8/3+8\delta}, & C_{4,\delta} &:= \frac{72}{(1-3\delta)\pi}, \\ B_{1,\delta} &:= 768\cdot(1+3\delta)\cdot\ln\Bigl(C_{4,\delta}\frac{\rho\sigma}{r^2\epsilon^2}\Bigr), & B_{2,\delta} &:= \frac{(1+\delta)^2}{(1-\delta)^2}\cdot\frac{\rho\sigma}{r^2\epsilon^2} \\ B_4 &:= 1+\frac{48\sigma\rho}{r^2}+\frac{4\rho}{r}, & b_4 &:= b_8 \wedge \mathrm{cut}_{\mathbb{P},r}(\mathcal{L}) \wedge d_{\min}(\mathbb{P})/14 \wedge \mathrm{vol}_{\mathbb{P},r}(\mathcal{L}\cap\mathcal{C}^{(1)}), \\ b_8 &:= \mathrm{vol}_{\mathbb{P},r}(\mathcal{X})/4 \wedge \frac{\epsilon r^2}{4\rho\sigma} \wedge \frac{\pi r^3}{8\sigma\rho^2}. \end{aligned}$$

To prove Theorem 17, we use Theorem 28, Proposition 9 and (100) to show that the cluster estimate $\widehat{C}$ must have a small normalized cut. On the other hand, in Lemma 41 we establish that any set $Z \subseteq X$ which is close to $\mathcal{C}^{(1)}[X]$—meaning $\mathrm{vol}_{n,r}(Z \vartriangle \mathcal{C}^{(1)}[X])$ is small—has a large normalized cut.

Lemma 41 *Fix $\delta \in (0,1)$. With probability at least $1-B_4\exp\{-n\delta^2 b_8\}$, the following statement holds:*

$$\Phi_{n,r}(Z) \geq \frac{(1-\delta)^2}{4(1+\delta)\pi}\left(1-2\frac{\sigma\rho}{(1-\delta)r^2n^2}\mathrm{vol}_{n,r}(Z \vartriangle \mathcal{C}^{(1)}[X])\right)\frac{\epsilon^2 r}{\sigma}, \quad \textit{for all } Z \subseteq X. \tag{105}$$

We therefore conclude that $\mathrm{vol}_{n,r}(\widehat{C} \vartriangle \mathcal{C}^{(1)}[X])$ must be large. In the remainder of this, we detail the probabilistic estimates used in the proof of Theorem 17, and then give a formal proof of Theorem 17 and then of Lemma 41.

Probabilistic estimates. In addition to (105), we will assume (i) that the graph normalized cut of $\mathcal{L}$ and local spread of $\mathcal{X}$ are respectively upper and lower bounded,

$$\Phi_{n,r}(\mathcal{L}[X]) \leq (1+3\delta)\cdot\Phi_{\mathbb{P},r}(\mathcal{L}), \quad \text{and} \quad s_{n,r}(X) \geq (1-3\delta)\cdot s_{\mathbb{P},r}(\mathcal{X});$$

(ii) that the graph volume of $\mathcal{L}$ is upper and lower bounded,

$$(1-\delta)\mathrm{vol}_{\mathbb{P},r}(\mathcal{L}) \leq \frac{1}{n(n-1)}\mathrm{vol}_{n,r}(\mathcal{L}[X]) \leq (1+\delta)\mathrm{vol}_{\mathbb{P},r}(\mathcal{L});$$

(iii) that the graph volumes of $\mathcal{L}\cap\mathcal{C}^{(1)}$ and $\mathcal{C}^{(1)}$ are respectively lower and upper bounded,

$$(1-\delta)\mathrm{vol}_{\mathbb{P},r}(\mathcal{L}\cap\mathcal{C}^{(1)}) \leq \frac{1}{n(n-1)}\mathrm{vol}_{n,r}(\mathcal{L}[X]\cap\mathcal{C}^{(1)}[X]) \text{ and } \frac{1}{n(n-1)}\mathrm{vol}_{n,r}(\mathcal{C}^{(1)}[X]) \leq (1+\delta)\mathrm{vol}_{\mathbb{P},r}(\mathcal{C}^{(1)}); \tag{106}$$

(iv) that the graph volume of $\mathcal{X}$ is lower bounded,

$$\frac{1}{n(n-1)}\mathrm{vol}_{n,r}(X) \geq (1-\delta)\mathrm{vol}_{\mathbb{P},r}(\mathcal{X});$$

and finally (v) that the maximum degree of $G_{n,r}$ is upper bounded,

$$\frac{1}{n-1}d_{\max}(G_{n,r}) \leq (1+\delta)d_{\max}(\mathbb{P}).$$

It follows from Lemma 41, Propositions 9 and 10, and Lemmas 29 and 31 that these estimates are together satisfied with probability at least $1 - B_4\exp\{-n\delta^2 b_8\} - 3\exp\{-n\delta^2\mathrm{cut}_{\mathbb{P},r}(\mathcal{L})\} - (2n+2)\exp\{-n\delta^2\cdot d_{\min}(\mathbb{P})/14\} - 5\exp\{-n\delta^2\mathrm{vol}_{\mathbb{P},r}(\mathcal{L}\cap\mathcal{C}^{(1)})\} \geq 1-(B_4+2n+10)\exp\{-n\delta^2 b_4\}$.

Proof of Theorem 17. As mentioned, we would like to use Theorem 28 to upper bound $\Phi_{n,r}(\widehat{C})$, and so we first verify that the conditions of Theorem 28 are met. In particular, we have each of the following.

- Recall that $n \geq 8\cdot(1+\delta)/(1-\delta)$ (33) and that $\mathrm{vol}_{\mathbb{P},r}(\mathcal{L}) \geq 3/16\cdot\pi r^2/(2\sigma\rho)$ (as shown in the proof of (100)). It is additionally clear that $d_{\max}(\mathbb{P}) \leq \pi r^2/(2\rho\sigma)$, and consequently,

$$d_{\max}(G_{n,r}) \leq (n-1)\cdot(1+\delta)d_{\max}(\mathbb{P}) \leq \frac{1}{3}n^2(1-\delta)\mathrm{vol}_{\mathbb{P},r}(\mathcal{L}) \leq \frac{1}{3}\mathrm{vol}_{n,r}(\mathcal{L}[X]). \tag{107}$$

  Therefore the lower bound in condition (72) is satisfied.

- Note that $\delta \in (0,1/7)$ implies $(1-\delta)/(1+\delta) > 3/4$, and additionally that $\mathrm{vol}_{\mathbb{P},r}(\mathcal{L}) \leq \mathrm{vol}_{\mathbb{P},r}(\mathcal{X})/2$. It follows that

$$\mathrm{vol}_{n,r}(X) \geq n(n-1)(1-\delta)\mathrm{vol}_{\mathbb{P},r}(\mathcal{X}) \geq 2n(n-1)(1-\delta)\mathrm{vol}_{\mathbb{P},r}(\mathcal{L}) \geq 2\frac{(1-\delta)}{(1+\delta)}\mathrm{vol}_{n,r}(\mathcal{L}[X]) \geq \frac{3}{2}\mathrm{vol}_{n,r}(\mathcal{L}[X]). \tag{108}$$

  Therefore the upper bound in condition (72) is satisfied.

- By (100), the normalized cut of $\mathcal{L}$ satisfies the following upper bound,

$$\Phi_{n,r}(\mathcal{L}[X]) \leq (1+3\delta) \cdot \Phi_{\mathbb{P},r}(\mathcal{L}) \leq (8/3 + 8\delta) \cdot \frac{r}{\rho}, \tag{109}$$

and by (101) the local spread of $\mathcal{X}$ satisfies the following lower bound,

$$s_{n,r}(X) \geq (1-3\delta) \cdot s_{\mathbb{P},r}(\mathcal{X}) \geq (1-3\delta) \cdot \frac{\pi r^2}{2\rho\sigma}. \tag{110}$$

The constants $B_{1,\delta}$ and $B_{2,\delta}$ in assumption (33) are chosen so that condition (73) is satisfied.

As a result, we may apply Theorem 28, and deduce the following: there exists a set $\mathcal{L}[X]^g \subset \mathcal{L}$ of large volume, $\mathrm{vol}_{n,r}(\mathcal{L}[X]^g) \geq 5/6 \cdot \mathrm{vol}_{n,r}(\mathcal{L}[X])$, such that for any seed node $v \in \mathcal{L}[X]^g$, the normalized cut of the PPR cluster estimate $\Phi_{n,r}(\widehat{C})$ satisfies the following upper bound:

$$\Phi_{n,r}(\widehat{C}) < 72\sqrt{\Phi_{n,r}(\mathcal{L}[X]) \cdot \ln\Big(\frac{36}{s_{n,r}(X)}\Big)} \leq 72\sqrt{(8/3+8\delta) \cdot \frac{r}{\rho} \cdot \ln\Big(\frac{72\rho\sigma}{(1-3\delta)\pi r^2 \epsilon^2}\Big)}.$$

Combined with Lemma 41, this implies

$$\frac{(1-\delta)^2}{4(1+\delta)\pi}\left(1 - 2\frac{\sigma\rho}{(1-\delta)r^2 n^2}\mathrm{vol}_{n,r}(\widehat{C} \vartriangle \mathcal{C}^{(1)}[X])\right)\frac{\epsilon^2 r}{\sigma} \leq 72\sqrt{(8/3+8\delta) \cdot \frac{r}{\rho} \cdot \ln\Big(\frac{72\rho\sigma}{(1-3\delta)\pi r^2\epsilon^2}\Big)}, \tag{111}$$

and solving for $\mathrm{vol}_{n,r}(\widehat{C} \vartriangle \mathcal{C}^{(1)}[X])$ yields (34).

We conclude by observing that the set $\mathcal{L}[X]^g$ must have significant overlap with $\mathcal{C}^{(1)}[X]$. In particular,

$$\begin{aligned}
\mathrm{vol}_{n,r}(\mathcal{L}[X]^g \cap \mathcal{C}^{(1)}[X]) &\geq \mathrm{vol}_{n,r}\big((\mathcal{L} \cap \mathcal{C}^{(1)})[X]\big) - \frac{1}{6}\mathrm{vol}_{n,r}\big(\mathcal{L}[X]\big) \\
&\overset{\text{(i)}}{\geq} n(n-1) \cdot \left((1-\delta) - \frac{1}{2}(1+\delta)\right)\mathrm{vol}_{\mathbb{P},r}(\mathcal{L} \cap \mathcal{C}^{(1)}) \\
&\overset{(ii)}{\geq} n(n-1) \cdot \frac{1}{7}\mathrm{vol}_{\mathbb{P},r}(\mathcal{L} \cap \mathcal{C}^{(1)}) \\
&\overset{(iii)}{\geq} \frac{1}{8}\mathrm{vol}_{n,r}(\mathcal{L} \cap \mathcal{C}^{(1)})
\end{aligned}$$

where in (i) we have used $\mathrm{vol}_{\mathbb{P},r}(\mathcal{L}) \leq 3\mathrm{vol}_{\mathbb{P},r}(\mathcal{C}^{(1)})$, and in (ii) and (iii) we have used $\delta \in (0, 1/7)$.

Proof of Lemma 41. To lower bound the normalized cut $\Phi_{n,r}(Z)$, it suffices to lower bound $\mathrm{cut}_{n,r}(Z)$ and upper bound $\mathrm{vol}_{n,r}(Z)$. A crude upper bound on the volume is simply

$$\mathrm{vol}_{n,r}(Z) \leq \mathrm{vol}_{n,r}(G_{n,r}) \overset{(i)}{\leq} (1+\delta)\mathrm{vol}_{\mathbb{P},r}(\mathcal{X})n(n-1) \leq (1+\delta)\frac{\pi r^2}{\rho\sigma}n^2 \tag{112}$$

where by Lemma 30, inequality (i) holds with probability at least $1 - \exp\{-n\delta^2\mathrm{vol}_{\mathbb{P},r}(\mathcal{X})/4\}$. This crude upper bound will suffice for our purposes.

We turn to lower bounding $\mathrm{cut}_{n,r}(Z)$. Establishing this lower bound is considerably more involved, and we start by giving an outline of our approach to build intuition.

- *Discretization.* Intuitively, we will approximate the cut of Z by discretizing the space $\mathcal{X}$ into a collection of rectangular bins $\bar{Q}$. We will consider a subset of bins $\partial\bar{Z} \subseteq \bar{Q}$, which we call the discretized boundary of Z, and which consists of those bins $Q \in \bar{Q}$ containing many points $x_i \in Z$ adjacent, in the graph $G_{n,r}$, to many $x_j \in X \setminus Z$. This discretization allows us to relate $\mathrm{cut}_{n,r}(Z)$ to the number of bins in the discretized boundary $\partial\bar{Z}$, as in (113).

- *Slicing.* In order to lower bound the number of bins in the discretized boundary $|\partial\bar{Z}|$, we consider a second partition of $\mathcal{X}$ into horizontal slices R. We argue that for each slice R, one of two things must be true: either R contains a bin belonging to the discretized boundary of Z, $Q \in \partial\bar{Z}$, or $(Z \vartriangle \mathcal{C}^{(1)}[X]) \cap R$ has a "substantial" volume,
$$\mathrm{vol}_{n,r}\Big((Z \vartriangle \mathcal{C}^{(1)}[X]) \cap R\Big) \geq R_{\min},$$
where $R_{\min}$ is a random quantity defined in (115). Summing over all slices R then gives a lower bound on $|\partial\bar{Z}|$, as in (115).

- *Probabilistic Guarantees.* The lower bound (115) depends on two random quantities: the minimum number of points in any bin (denoted by $Q_{\min}$), and $R_{\min}$, which was discussed in the previous bullet. We give suitable lower bounds on each quantity, which hold with high probability, and which suffice to complete the proof of Lemma 41.

Now, we fill in the details, working step-by-step.

<u>*Step 1: Discretization.*</u> We begin by discretizing $\mathcal{X}$ into a collection of bins. To that end, to each $k_1 \in \left[\frac{12\sigma}{r}\right], k_2 \in \left[\frac{4\rho}{r}\right]$ associate the bin[7]
$$Q_{(k_1,k_2)} := \left[-\frac{3\sigma}{2} + \frac{(k_1-1)}{4}r, -\frac{3\sigma}{2} + \frac{k_1}{4}r\right] \times \left[-\frac{\rho}{2} + \frac{(k_2-1)}{4}r, -\frac{\rho}{2} + \frac{k_2}{4}r\right];$$
and let $\overline{Q} = \left\{Q_{(k_1,k_2)} : k_1 \in \left[\frac{12\sigma}{r}\right], k_2 \in \left[\frac{4\rho}{r}\right]\right\}$ be the collection of such bins. Next, we define the binned set $\overline{Z} \subset \overline{Q}$ to contain all bins $Q \in \bar{Q}$ for which $Z \cap Q$ is larger than $X \setminus Z \cap Q$; in mathematical notation,
$$\overline{Z} := \left\{Q \in \overline{Q} : \mathbb{P}_n(Z \cap Q) \geq \frac{1}{2}\mathbb{P}_n(Q)\right\}.$$
Then the discretized boundary $\partial\bar{Z}$ is
$$\partial\overline{Z} := \left\{Q_{(k_1,k_2)} \in \overline{Z} : \exists(\ell_1,\ell_2) \in \left[\frac{12\sigma}{r}\right] \times \left[\frac{4\rho}{r}\right] \text{ such that } Q_{(\ell_1,\ell_2)} \notin \bar{Z}, \|(k_1,k_2)-(\ell_1,\ell_2)\|_1 = 1\right\}.$$
Intuitively, every point $x_i \in Z$ which belongs to a cube in the boundary set $\partial\overline{Z}$ will have many edges to $X \setminus Z$. Formally, letting $Q_{\min} := \min_{Q \in \overline{Q}} \mathbb{P}_n(Q)$, we have
$$\mathrm{cut}_{n,r}(Z) \geq \mathrm{cut}_{n,r}(Z \cap \{x_i \in \overline{Z}\}) \geq \frac{1}{4}\left|\partial\overline{Z}\right| Q_{\min}^2. \tag{113}$$

7. Here we are assuming, without loss of generality, that $12\sigma/r \in \mathbb{N}$ and $4\rho/r \in \mathbb{N}$, and using the notation $[x] = \{1, \ldots, x\}$ for $x \in \mathbb{N}$.

To establish the last inequality, we reason as follows: first, for every cube $Q_{(k_1,k_2)} \in \partial\overline{Z}$, there exists a cube $Q_{(\ell_1,\ell_2)} \notin \overline{Z}$ such that $\|(k_1,k_2)-(\ell_1,\ell_2)\|_1 \leq 1$; second, since each cube has side length $r/4$, this implies that for every $x_i \in Q_{(k_1,k_2)}$ and $x_j \in Q_{(\ell_1,\ell_2)}$ the edge (x_i,x_j) belongs to $G_{n,r}$.

Step 2: Slicing. Now we move on lower bounding the size of the discretized boundary $\bigl|\partial\bar{Z}\bigr|$. To do so, we divide $\mathcal{X}$ into slices horizontally. Let $R_k = \Bigl\{(x_1,x_2) \in \mathcal{X} : x_2 \in \bigl[-\frac{\rho}{2} + \frac{(k-1)}{4}r, -\frac{\rho}{2} + \frac{k}{4}r\bigr]\Bigr\}$ be the kth horizontal slice, and $\bar{R}_k = \bigl\{Q_{(k_1,k)} \in \overline{Q} : k_1 \in [\frac{12\sigma}{r}]\bigr\}$ be the binned version of R_k. For every k, we are in one of three cases:

1. $\bar{R}_k \cap \bar{Z} = \emptyset$, in which case
$$\mathrm{vol}_{n,r}\Bigl(\bigl(Z \vartriangle \mathcal{C}^{(1)}[X]\bigr) \cap R_k\Bigr) \geq \frac{1}{2}\mathrm{vol}_{n,r}(\mathcal{C}^{(1)}[X] \cap R_k), \quad \text{or}$$
2. $\bar{R}_k \cap \bar{Z} = \bar{R}_k$, in which case
$$\mathrm{vol}_{n,r}\Bigl(\bigl(Z \vartriangle \mathcal{C}^{(1)}[X]\bigr) \cap R_k\Bigr) \geq \frac{1}{2}\mathrm{vol}_{n,r}(\mathcal{C}^{(2)}[X] \cap R_k), \quad \text{or}$$
3. $\bar{R}_k \cap \partial\bar{Z} \neq \emptyset$.

In words, these cases describe, for each slice R, the situation in which the set Z has a "substantial" boundary (Case 3), or the situation in which Z does not have a substantial boundary, and instead has a substantial difference with $\mathcal{C}^{(1)}$ (Cases 1 and 2).

Now, let $N(R)$ be the number of slices which contain a boundary cell,
$$N(R) := \#\Bigl\{k \in \left[\frac{2\rho}{r}\right] : \bar{R}_k \cap \partial\bar{Z} \neq \emptyset\Bigr\}.$$
and let $R_{\min}$ be the minimum volume of the intersection of any slice R_k with the density clusters $\mathcal{C}^{(1)}, \mathcal{C}^{(2)}$; formally,
$$R_{\min} := \min_k \Bigl\{\mathrm{vol}_{n,r}(\mathcal{C}^{(1)}[X] \cap R_k) \wedge \mathrm{vol}_{n,r}(\mathcal{C}^{(2)}[X] \cap R_k)\Bigr\}. \tag{114}$$
Summing over all slices R_k, we have from our previous analysis that,
$$\mathrm{vol}_{n,r}\Bigl(Z \vartriangle \mathcal{C}^{(1)}[X]\Bigr) = \sum_{k \in [2\rho/r]} \mathrm{vol}_{n,r}\Bigl(\bigl(Z \vartriangle \mathcal{C}^{(1)}[X]\bigr) \cap R_k\Bigr) \geq R_{\min}\left[\frac{2\rho}{r} - \frac{N(R)}{2}\right].$$
Rearranging this expression and noting that $\bigl|\partial\bar{Z}\bigr| \geq N(R)$, we obtain the following lower bound on the cardinality of the discretized boundary,
$$\bigl|\partial\bar{Z}\bigr| \geq N(R) \geq 2\Bigl(\frac{2\rho}{r} - \frac{\mathrm{vol}_{n,r}(Z \vartriangle \mathcal{C}^{(1)}[X])}{R_{\min}}\Bigr). \tag{115}$$

Combining (113) and (115), we have that

$$\begin{aligned}\operatorname{cut}_{n,r}(Z) &\geq \frac{1}{4}N(R)Q_{\min}^2 \\ &\geq \left(\frac{\rho}{r} - \frac{\operatorname{vol}_{n,r}(Z \vartriangle \mathcal{C}^{(1)}[X])}{2R_{\min}}\right)Q_{\min}^2 \end{aligned} \tag{116}$$

for all $Z \subset X$.

<u>*Step 3: Probabilistic guarantees.*</u> It remains to lower bound the random quantities $R_{\min}$ and $Q_{\min}$. To do so, we first lower bound the expected probability of any cell Q,

$$\min_{Q \in \overline{Q}} \mathbb{P}(Q) \geq \frac{\epsilon r^2}{4\rho\sigma}, \tag{117}$$

and the expected volume of $\mathcal{C}^{(1)}[X] \cap R_k$ and $\mathcal{C}^{(2)}[X] \cap R_k$,

$$\operatorname{vol}_{\mathbb{P},r}(\mathcal{C}^{(1)} \cap R_k) = \operatorname{vol}_{\mathbb{P},r}(\mathcal{C}^{(2)} \cap R_k) \geq \frac{\pi r^3}{4\sigma\rho^2} \quad \text{for all } k. \tag{118}$$

Since $Q_{\min}$ and $R_{\min}$ are obtained by taking the minimum of functionals over a fixed number of sets in n, they concentrate tightly around their means. Specifically, note that the total number of cubes is $\left|\overline{Q}\right| = \frac{48\sigma\rho}{r^2}$, and the total number of horizontal slices is $\frac{4\rho}{r}$. Along with (117) and (118), by Lemma 31

$$Q_{\min} \geq (1-\delta)\frac{\epsilon r^2}{4\rho\sigma}n \quad \text{and} \quad R_{\min} \geq (1-\delta)\frac{\pi r^3}{4\sigma\rho^2}n(n-1),$$

with probability at least $1 - \frac{48\sigma\rho}{r^2}\exp\left\{-\frac{n\delta^2\epsilon r^2}{4\rho\sigma}\right\} - \frac{4\rho}{r}\exp\left\{-\frac{n(n-1)\delta^2\pi r^3}{8\sigma\rho^2}\right\}$. Combining these lower bounds with (112) and (116), we obtain

$$\Phi_{n,r}(Z) \geq \frac{(1-\delta)^2}{4(1+\delta)\pi}\left(1 - 2\frac{\sigma\rho}{(1-\delta)r^2n^2}\operatorname{vol}_{n,r}(Z \vartriangle \mathcal{C}^{(1)}[X])\right)\frac{\epsilon^2 r}{\sigma},$$

for all $Z \subseteq X$.

Appendix E. Additional results: aPPR and consistency of PPR

In this appendix, we prove two additional results regarding PPR remarked upon in our main text. In Section E.1, we show that clustering using the aPPR vector satisfies an equivalent guarantee to Theorem 12. In Section E.2, we show that the PPR vector can perfectly distinguish two distinct density clusters $\mathcal{C}_\lambda, \mathcal{C}'_\lambda$.

E.1 Generic cluster recovery with aPPR

Our formal claim regarding cluster recovery with aPPR is contained in Corollary 42.

Corollary 42 *Consider instead of Algorithm 1 using the approximate PPR vector from Andersen et al. (2006) satisfying (2), and forming the corresponding cluster estimate $\widehat{C}$ in the same manner. Then provided we take*

$$\varepsilon = \frac{1}{25(1+\delta)n(n-1)\mathrm{vol}_{\mathbb{P},r}(\mathcal{C})}, \tag{119}$$

under the assumptions of Theorem 12 the upper bound on symmetric set difference in (25) still holds.

Proof of Corollary 42. Note that the choice of ε in (119) implies $\varepsilon \leq 1/\bigl(25\mathrm{vol}_{n,r}(\mathcal{C}[X])\bigr)$ with probability at least $1-\exp\bigl\{-n\delta^2\mathrm{vol}_{\mathbb{P},r}(\mathcal{C})\bigr\}$. The proof of Corollary 42 is then identical to that of Theorem 12, except one uses Corollary 18 rather than Lemma 7 to relate the symmetric set difference to the graph normalized cut and mixing time.

E.2 Perfectly distinguishing two density clusters

As mentioned in our main text, the symmetric set difference does not measure whether $\widehat{C}$ can (perfectly) distinguish any two distinct clusters $\mathcal{C}_\lambda, \mathcal{C}'_\lambda \in \mathbb{C}_f(\lambda)$. We therefore also study a second notion of cluster estimation, first introduced by Hartigan (1981).

Definition 43 *For an estimator $\widehat{C} \subseteq X$ and distinct clusters $\mathcal{C}_\lambda, \mathcal{C}'_\lambda \in \mathbb{C}_f(\lambda)$, we say $\widehat{C}$* separates *$\mathcal{C}_\lambda$ from $\mathcal{C}'_\lambda$ if*

$$\mathcal{C}_\lambda[X] \subseteq \widehat{C} \quad \textit{and} \quad \widehat{C} \cap \mathcal{C}'_\lambda[X] = \emptyset. \tag{120}$$

The bound on symmetric set difference (30) does not imply (120), which requires a uniform bound over the PPR vector p_v. As an example, suppose that we were able to show that for all $\mathcal{C}' \in \mathbb{C}_f(\lambda), \mathcal{C}' \neq \mathcal{C}$, and each $u \in \mathcal{C}, u' \in \mathcal{C}'$,

$$\frac{p_v(u')}{\deg(u';G)} \leq \frac{1}{10} \cdot \frac{1}{n(n-1)\mathrm{vol}_{\mathbb{P},r}(\mathcal{C}_{\lambda,\sigma})} < \frac{1}{5} \cdot \frac{1}{n(n-1)\mathrm{vol}_{\mathbb{P},r}(\mathcal{C}_{\lambda,\sigma})} \leq \frac{p_v(u)}{\deg(u;G)}. \tag{121}$$

Then, any (L,U) satisfying (21) and any sweep cut S_β for $\beta \in (L,U)$ would result in a cluster estimate $\widehat{C}$ fulfilling both conditions laid out in (120). In Theorem 44, we show that a sufficiently small upper bound on $\Delta(\widehat{C}, \mathcal{C}_{\lambda,\sigma}[X])$ ensures that with high probability the uniform bound (121) is satisfied, and hence implies that $\widehat{C}$ will separate $\mathcal{C}_\lambda$ from $\mathcal{C}'_\lambda$. In what follows, put

$$c_{1,\delta} := \frac{(1-\delta)^5}{4} \cdot \min\Bigl\{\Bigl(\frac{3}{8(1+\delta)} - \frac{1}{5}\Bigr), \frac{1}{10}\Bigr\}$$

and note that if $\delta \in (0, 7/8)$ then $c_{1,\delta} > 0$. (In fact, we will have to take $\delta \in (0,1/4)$ in order to use Propositions 9 and 10). Additionally, denote $\mathbb{P}'$ for the conditional distribution of a sample point given that it falls in $\mathcal{C}'_{\lambda,\sigma}$, i.e. $\mathbb{P}'(\mathcal{S}) := \mathbb{P}(\mathcal{S} \cap \mathcal{C}'_{\lambda,\sigma})/\mathbb{P}(\mathcal{C}'_{\lambda,\sigma})$, and $G'_{n,r} := G_{n,r}[\mathcal{C}'_{\lambda,\sigma}]$ for the subgraph of $G_{n,r}$ induced by $\mathcal{C}'_{\lambda,\sigma}$.

Theorem 44 *For any $\delta \in (0,1/4)$ any $n \in \mathbb{N}$ such that*

$$\frac{1}{n} \leq \delta \cdot \frac{4\mathbb{P}(\mathcal{C}'_{\lambda,\sigma})}{3} \tag{122}$$

and otherwise under the same conditions as Theorem 12, the following statement holds with probability at least $1 - B_2/n - 4\exp\{-b_1\delta^2 n\} - (n+2)\exp\{-b_3 n\} - 3(n+3)\exp\{-b_7\delta^2 n\}$*: there exists a set* $\mathcal{C}_{\lambda,\sigma}[X]^g \subseteq \mathcal{C}_{\lambda,\sigma}[X]$ *of large volume,* $\mathrm{vol}_{n,r}(\mathcal{C}_{\lambda,\sigma}[X]^g) \geq \mathrm{vol}_{n,r}(\mathcal{C}_{\lambda,\sigma}[X])/2$*, such that if Algorithm 1 is* δ*-well-initialized and run with any seed node* $v \in \mathcal{C}_{\lambda,\sigma}[X]^g$*, and moreover*

$$\kappa_{\mathbb{P},r}(\mathcal{C}_{\lambda,\sigma},\delta) \leq c_{1,\delta} \cdot \frac{\min\bigl\{\mathbb{P}(\mathcal{C}_{\lambda,\sigma})^2 \cdot d_{\min}(\widetilde{\mathbb{P}})^2, \mathbb{P}(\mathcal{C}'_{\lambda,\sigma})^2 \cdot d_{\min}(\mathbb{P}')^2\bigr\}}{\mathrm{vol}_{\mathbb{P},r}(\mathcal{C}_{\lambda,\sigma})} \tag{123}$$

then the PPR estimated cluster $\widehat{C}$ *satisfies* (120).

Before we prove Theorem 44, we make a few brief remarks:

- In one sense, Theroem 44 is a strong result: if the density clusters $\mathcal{C}_\lambda, \mathcal{C}'_\lambda$ satisfies the requirement (123), and we are willing to ignore the behavior of the algorithm in low-density regions, Theorem 44 guarantees that PPR will *perfectly distinguish* the candidate cluster $\mathcal{C}_\lambda$ from $\mathcal{C}'_\lambda$.

- On the other hand, unfortunately the requirement (123) is rather restrictive. Suppose the density cluster $\mathcal{C}_{\lambda,\sigma}$ satisfies (A3). Then from the following chain of inequalities,

$$\frac{\Delta(\widehat{C}, \mathcal{C}_{\lambda,\sigma}[X])}{\mathrm{vol}_{n,r}(\mathcal{C}_{\lambda,\sigma}[X])} \overset{(\text{Thm. }12)}{\leq} \kappa_{\mathbb{P},r}(\mathcal{C},\delta) \overset{(123)}{\leq} c_{1,\delta} \cdot \frac{\mathbb{P}(\mathcal{C}_{\lambda,\sigma})^2 \cdot d_{\min}(\widetilde{\mathbb{P}})^2}{\mathrm{vol}_{\mathbb{P},r}(\mathcal{C}_{\lambda,\sigma})} \overset{(A3)}{\leq} c_{1,\delta} \cdot \frac{\Lambda_\sigma}{\lambda_\sigma}\nu_d r^d,$$

  we see that in order for (123) to be met, it is necessary that $\Delta(\widehat{C}, \mathcal{C}_{\lambda,\sigma}[X])/\mathrm{vol}_{n,r}(\mathcal{C}_{\lambda,\sigma}[X])$ be on the order of r^d. In plain terms, we are able to recover a density cluster $\mathcal{C}_\lambda$ in the strong sense of (120) only when we can guarantee the volume of the symmetric set difference will be very small. This strong condition is the price we pay in order to obtain the uniform bound in (121).

- The proof of Theorem 44 relies heavily on Lemma 19. This lemma—or more accurately, the equation (66) used in the proof of the lemma—can be thought of as a smoothness result for the PPR vector, showing that the mass of $p_v(\cdot)$ cannot be overly concentrated at any one vertex $u \in V$. However, (66) is a somewhat crude bound. By plugging a stronger result on the smoothness of $p_v(\cdot)$ in to the proof of Lemma 19, we could improve the uniform bounds of the lemma, and in turn show that the conclusion of Theorem 44 holds under weaker conditions than (123).

Now we recall some probabilistic estimates before proceeding to the proof of Theorem 44.

Probabilistic estimates. As in the proof of Theorem 12, we will assume that the inequalities (16)-(20) and (102)-(104) are satisfied. We will additionally assume that

$$n' \geq (1-\delta) \cdot n \cdot \mathbb{P}(\mathcal{C}'_{\lambda,\sigma}) \overset{(122)}{\Longrightarrow} n' - 1 \geq (1-\delta)^2 \cdot n \cdot \mathbb{P}(\mathcal{C}'_{\lambda,\sigma}) \tag{124}$$

and that

$$\frac{1}{n'-1} d_{\min}(G'_{n,r}) \geq (1-\delta) \cdot d_{\min}(\mathbb{P}'). \tag{125}$$

By Propositions 9-10 and Lemmas 29-31, these inequalities hold with probability at least $1 - B_2/n - 4\exp\{-b_1\delta^2 n\} - (n+1)\exp\{-b_3 n\} - (3n+3)\exp\{-b_7\delta^2 n\}$, taking $b_7 := b_2 \wedge \mathbb{P}(\mathcal{C}'_{\lambda,\sigma}) \cdot d_{\min}(\mathbb{P}')/9$.

Proof of Theorem 44. We have already verified in the proof of Theorem 12 that $\alpha \leq 1/(2\tau_\infty(\widetilde{G}_{n,r}))$, and we may therefore apply Lemma 19, which gives a lower bound on $p_v(u)$ for all $u \in \mathcal{C}_{\lambda,\sigma}[X]_o$ and an upper bound on $p_v(u')$ for all $p_v(u')$ for all $u' \in \mathcal{C}'_{\lambda,\sigma}[X]_o$. These bounds are useful because $r \leq \sigma$, which implies that $\mathcal{C}_\lambda[X] \subseteq \mathcal{C}_{\lambda,\sigma}[X]_o$ and likewise that $C'_\lambda[X] \subseteq \mathcal{C}'_{\lambda,\sigma}[X]_o$. We will show that these bounds in turn imply (121), from which the claim of the theorem follows.

We begin with the lower bound in (121). From (in order) Lemma 19, our assorted probabilistic estimates, and the assumed lower bound (123) on $\kappa_{\mathbb{P},r}(\mathcal{C},\delta)$, we have that for all $u \in \mathcal{C}_\lambda[X]$,

$$
\begin{aligned}
\frac{p_v(u)}{\deg_{n,r}(u)} &\geq \frac{3}{8\mathrm{vol}_{n,r}(\mathcal{C}[X])} - 2\frac{\Phi_{n,r}(\mathcal{C}_{\lambda,\sigma}[X])}{d_{\min}(\widetilde{G}_{n,r})^2\alpha} \\
&\geq \frac{1}{n(n-1)}\left(\frac{3}{8(1+\delta)\mathrm{vol}_{\mathbb{P},r}(\mathcal{C}_{\lambda,\sigma})} - 4\cdot\frac{n(n-1)}{(\widetilde{n}-1)^2}\cdot\frac{\kappa_{\mathbb{P},r}(\mathcal{C}_{\lambda,\sigma},\delta)}{(1-\delta)d_{\min}(\widetilde{\mathbb{P}})^2}\right) \\
&\geq \frac{1}{n(n-1)}\left(\frac{3}{8(1+\delta)\mathrm{vol}_{\mathbb{P},r}(\mathcal{C}_{\lambda,\sigma})} - 4\cdot\frac{\kappa_{\mathbb{P},r}(\mathcal{C}_{\lambda,\sigma},\delta)}{(1-\delta)^5\mathbb{P}(\mathcal{C})^2 d_{\min}(\widetilde{\mathbb{P}})^2}\right) \\
&\geq \frac{1}{5n^2\mathrm{vol}_{\mathbb{P},r}(\mathcal{C}_{\lambda,\sigma})}.
\end{aligned}
$$

An equivalent derivation implies the upper bound in (121): for all $u' \in \mathcal{C}'_\lambda[X]$,

$$
\begin{aligned}
\frac{p_v(u')}{\deg_{n,r}(u')} &\leq 2\frac{\Phi_{n,r}(\mathcal{C}_{\lambda,\sigma}[X])}{d_{\min}(G'_{n,r})^2\alpha} \\
&\leq 4\cdot\frac{1}{(n'-1)^2}\cdot\frac{\kappa(\mathcal{C}_{\lambda,\sigma},\delta)}{(1-\delta)d_{\min}(\mathbb{P}')^2} \\
&\leq 4\frac{\kappa(\mathcal{C}_{\lambda,\sigma},\delta)}{n^2(1-\delta)^5\mathbb{P}(\mathcal{C}'_{\lambda,\sigma})^2 d_{\min}(\mathbb{P}')^2} \leq \frac{1}{10n^2\mathrm{vol}_{\mathbb{P},r}(\mathcal{C}_{\lambda,\sigma})},
\end{aligned}
$$

completing the proof of Theorem 44.

Appendix F. Experimental details

Finally, we detail the settings of our experiments, and include an additional figure.

F.1 Experimental settings for Figure 2

Let $\mathcal{R}_{\sigma,\rho} = [-\sigma/2,\sigma/2]\times[-\rho/2,\rho/2]$ be the two-dimensional rectangle of width σ and height ρ, centered at the origin. We sample $n = 8000$ points according to the density function $f_{\rho,\sigma,\lambda}$, defined over domain $\mathcal{X} = [-1,1]^2$ and parameterized by ρ,σ and λ as follows:

$$
f_{\rho,\sigma,\lambda}(x) := \begin{cases} \lambda, & \text{if } x \in R_{\sigma,\rho} - (-.5,0) \text{ or } x \in R_{\sigma,\rho} + (-.5,0) \\ \dfrac{4-2\lambda\rho\sigma}{1-2\rho\sigma}, & \text{if } x \in \mathcal{X},\ x \notin R_{\sigma,\rho} - (-.5,0) \text{ and } x \notin R_{\sigma,\rho} + (-.5,0). \end{cases} \tag{126}
$$

Then $\theta := \lambda - \frac{4-2\lambda\rho\sigma}{1-2\rho\sigma}$ measures the difference in density between the density clusters and the rest of the domain. The first column displays $n = 8000$ points sampled from three different parameterizations of $f_{\rho,\sigma,\lambda}$:

$$\rho = .913, \qquad \sigma = .25, \qquad (\lambda-\theta)/\lambda = .25 \qquad \text{(top panel)}$$
$$\rho = .25, \qquad \sigma =, \qquad (\lambda-\theta)/\lambda = .05 \qquad \text{(middle panel)}$$
$$\rho = .5, \qquad \sigma = .25, \qquad (\lambda-\theta)/\lambda = .12 \qquad \text{(bottom panel.)}$$

In each of the first, second, and third rows, we fix two parameters and vary the third. In the first row, we fix $\sigma = .25$, $(\lambda-\theta)/\lambda = .25$, and vary ρ from .25 to 2. In the second row, we fix $\rho = 1.8$, $(\lambda-\theta)/\lambda = .05$, and vary σ from .1 to .2 In the third row, we fix $\rho = .5$, $\sigma = .25$ and vary $(\lambda-\theta)/\lambda$ from .1 to .25. In the first and third rows, we take $r = \sigma/8$; in the second row, where we vary σ, we take $r = .1/8$.

F.2 Experimental settings for Figure 3

To form each of the three rows in Figure 3, $n = 800$ points are independently sampled following a 'two moons plus Gaussian noise model'. Formally, the (respective) generative models for the data are

$$Z \sim \mathrm{Bern}(1/2), \theta \sim \mathrm{Unif}(0,\pi) \tag{127}$$

$$X(Z,\theta) = \begin{cases} \mu_1 + (r\cos(\theta), r\sin(\theta)) + \sigma\epsilon, & \text{if } Z = 1 \\ \mu_2 + (r\cos(\theta), -r\sin(\theta)) + \sigma\epsilon, & \text{if } Z = 0 \end{cases} \tag{128}$$

where

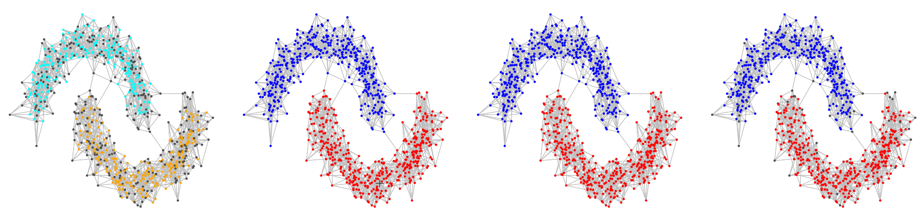
$$\mu_1 = (-.5, 0),\ \mu_2 = (0,0),\ \epsilon \sim N(0, I_2) \qquad \text{(row 1)}$$
$$\mu_1 = (-.5, -.07),\ \mu_2 = (0,.07),\ \epsilon \sim N(0, I_2) \qquad \text{(row 2)}$$
$$\mu_1 = (-.5, -.125),\ \mu_2 = (0,.125),\ \epsilon \sim N(0, I_2) \qquad \text{(row 3)}$$

for I_d the $d \times d$ identity matrix. In all cases $\sigma = .07$. In each case λ is taken as small as possible such that there exist exactly two distinct density clusters, which we call $\mathcal{C}_\lambda$ and $\mathcal{C}'_\lambda$; r is taken as small as possible so that each vertex has at least 2 neighbors. The first column consists of the empirical density clusters $\mathcal{C}_\lambda[X]$ and $\mathcal{C}'_\lambda[X]$ 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 Bresson et al. (2012); and the last column shows a cut of the density cluster tree estimator of Chaudhuri and Dasgupta (2010).

Performance of PPR with high-dimensional noise. Figure 4 is similar to Figure 3 of the main text, but with parameters

$$\mu_1 = (-.5, -.025),\ \mu_2 = (0, .025),\ \epsilon \sim N(0, I_{10}).$$

The gray dots in (a) (as in the left-hand column of Figure 3 in the main text) represent observations in low-density regions. While the PPR sweep cut (b) has relatively high symmetric set difference with the chosen density cut, it still recovers separates $C_\lambda[X]$ and $\mathcal{C}'_\lambda[X]$, in the sense of Definition 43.



(a) (b) (c) (d)

Figure 4: *True density (column 1), PPR (column 2), minimum normalized cut (column 3) and estimated density (column 4) clusters for two-moons with 10 dimensional noise. Seed node for PPR denoted by a black cross.*