

A Proofs

In this supplement, we present proofs for “Local Clustering of Density Upper Level Sets”. Sections A.1 - A.3 detail the proof for Theorem 1. A.4 develops a bound of the form of (??), which we recall links the conductance function to mixing time; this will be necessary for both Theorems ?? and ??. ?? and ?? give the proof of Theorem ??, while ??- ?? give the proof of Theorem ??. A.10 gives some general concentration results used throughout, before we finish with the proof of Theorem 4 in A.12.

A.1 Volume estimates

Let $\mathcal{A} \subseteq \mathbb{R}^d$, and for $\sigma \geq 0$, write $\sigma B := B(0, \sigma) = \{x \in \mathbb{R}^d : \|x\| \leq \sigma\}$ for the closed ball of radius σ centered at the origin (and let $B^\circ(0, \sigma)$ denote the corresponding open ball). Let $\mathcal{A}_\sigma = \mathcal{A} + \sigma B$ be the direct sum of \mathcal{A} and σB , $\mathcal{A}_\sigma = \{z = x + y : x \in \mathcal{A}, y \in \sigma B\}$.

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

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

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

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

which is sufficient to prove the claim.

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

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

Observe that since \mathcal{A} is closed and bounded, it is compact. As $B(x, \sigma)$ is compact, and the direct sum of two compact sets is itself compact, \mathcal{A}_σ is compact. Moreover,

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

so by compactness there exists $x_1, \dots, x_N \in \mathcal{A}$ such that

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

By the triangle inequality, $\mathcal{A}_\sigma + \delta B \subset \bigcup_{i=1}^N B^\circ(x_i, \sigma + \epsilon + \delta)$. Of course, for each $x_i \in \mathcal{A}$, $B(x_i, \sigma) \in \mathcal{A}_\sigma$. Summarizing our findings, we have

$$\bigcup_{i=1}^N B(x_i, \sigma) \subseteq \mathcal{A}_\sigma, \quad \mathcal{A}_\sigma + \delta B \subset \bigcup_{i=1}^N B^\circ(x_i, \sigma + \delta + \epsilon) \quad (\text{A.2})$$

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

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

Observe that $\bigcup_{i=1}^N \mathcal{A}_\sigma^{(i)} = \bigcup_{i=1}^N B(x_i, \sigma)$, so by (A.2) $\mathcal{A}_\sigma \supseteq \bigcup_{i=1}^N \mathcal{A}_\sigma^{(i)}$. As $\mathcal{A}_\sigma^{(1)}, \dots, \mathcal{A}_\sigma^{(N)}$ are non-overlapping,

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

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

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

As $\bigcup_{i=1}^N \mathcal{A}_{\sigma+\delta+\epsilon}^{(i)} = \bigcup_{i=1}^N B(x_i, \sigma + \delta + \epsilon)$, by (A.2)

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

and therefore

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

where the last inequality follows from Lemma 2. We have shown (A.1), and thus the claim. \square

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

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

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

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

since then

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

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

$$\|x\| \leq \sigma + \delta', \quad \|x - x_j\| > \sigma + \delta' \text{ for } j = 1, \dots, i-1. \quad (\text{A.3})$$

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

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

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

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

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

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

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

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

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

Let $\mathcal{C}_{\sigma, \sigma+r} := \{x : 0 < \text{dist}(x, \mathcal{C}_\sigma) < r\}$, where \mathcal{C}_σ is as in Theorem 1. Lemma 4 involves the bulk of the technical effort required to prove Theorem 1; it will be necessary to bound the expected cut size of $\mathcal{C}_\sigma[\mathbf{X}]$ in $G_{n,r}$.

Lemma 4. *Under the conditions of Theorem 1, and for any $0 < r \leq \sigma/2d$,*

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

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

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

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

$$\mathbb{P}(\mathcal{C}_{\sigma, \sigma+r}) = \int_{\mathcal{C}_{\sigma, \sigma+r}} f(x) dx = \sum_{i=0}^{k-1} \int_{\mathcal{T}_{i,k}} f(x) dx \leq \sum_{i=0}^{k-1} \nu(\mathcal{T}_{i,k}) \max_{x \in \mathcal{T}_{i,k}} f(x). \quad (\text{A.4})$$

(A1) and (A2) imply the upper bound

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

and writing

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

we have

$$\begin{aligned}
\sum_{i=0}^{k-1} \nu(\mathcal{T}_{i,k}) \max_{x \in \mathcal{T}_{i,k}} f(x) &\leq \sum_{i=0}^{k-1} \left\{ \nu_{i+1,k} - \nu_{i,k} \right\} \left(\lambda_\sigma - c_0 (rt_{i,k})^\gamma \right) \\
&= \sum_{i=1}^k \underbrace{\nu_{i,k} \left([\lambda_\sigma - c_0 (rt_{i-1,k})^\gamma] - [\lambda_\sigma - c_0 (rt_{i,k})^\gamma] \right)}_{:=\Sigma_k} + \underbrace{\left(\nu_{k,k} [\lambda_\sigma - c_0 r^\gamma] - \nu_{1,k} \lambda_\sigma \right)}_{:=\xi}
\end{aligned} \tag{A.5}$$

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

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

$$\nu_{i,k} = \nu(\mathcal{C}_\sigma + rt_{i,k}B) \leq \nu(\mathcal{C}_\sigma) \left(1 + \frac{rt_{i,k}}{\sigma} \right)^d \tag{A.6}$$

which in turn gives

$$\begin{aligned}
\Sigma_k &\leq c_0 \nu(\mathcal{C}_\sigma) r^\gamma \sum_{i=1}^k \left(1 + \frac{rt_{i,k}}{\sigma} \right)^d \left((t_{i,k})^\gamma - (t_{i-1,k})^\gamma \right) \\
&= c_0 \nu(\mathcal{C}_\sigma) r^\gamma \sum_{i=1}^k \left(1 + \frac{ru_{i,k}^{1/\gamma}}{\sigma} \right)^d (u_{i,k} - u_{i,k-1}). \quad (u_{i,k} = t_{i,k}^\gamma)
\end{aligned} \tag{A.7}$$

(A.7) is a Riemann sum, and taking the limit as $k \rightarrow \infty$ we obtain

$$\begin{aligned}
\lim_{k \rightarrow \infty} c_0 \nu(\mathcal{C}_\sigma) r^\gamma \sum_{i=1}^k \left(1 + \frac{ru_{i,k}^{1/\gamma}}{\sigma} \right)^d (u_{i,k} - u_{i,k-1}) &= c_0 \nu(\mathcal{C}_\sigma) r^\gamma \int_0^1 \left(1 + \frac{ru^{1/\gamma}}{\sigma} \right)^d du \\
&\stackrel{(i)}{\leq} c_0 \nu(\mathcal{C}_\sigma) r^\gamma \int_0^1 \left(1 + \frac{2dr u^{1/\gamma}}{\sigma} \right) du \\
&= c_0 \nu(\mathcal{C}_\sigma) r^\gamma \left(1 + \gamma \frac{2dr}{(\gamma+1)\sigma} \right).
\end{aligned} \tag{A.8}$$

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

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

$$\begin{aligned}
\xi &\stackrel{(ii)}{\leq} \nu(\mathcal{C}_\sigma) \left\{ \left(1 + \frac{r}{\sigma} \right)^d (\lambda_\sigma - c_0 r^\gamma) - \lambda_\sigma \right\} \\
&\stackrel{(iii)}{\leq} \nu(\mathcal{C}_\sigma) \left\{ \left(1 + \frac{2dr}{\sigma} \right) (\lambda_\sigma - c_0 r^\gamma) - \lambda_\sigma \right\} = \nu(\mathcal{C}_\sigma) \left\{ \frac{2dr}{\sigma} (\lambda_\sigma - c_0 r^\gamma) - c_0 r^\gamma \right\}.
\end{aligned} \tag{A.9}$$

where (ii) follows from (A.6), and (iii) from Lemma 3. As the bounds in (A.4) and (A.5) hold for all k , these along with (A.8) and (A.9) imply the desired result. \square

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

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

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

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

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

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

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

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

which by Lemma 6 has volume

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

with $\alpha = 1 - \frac{2rh - h^2}{r^2} = \frac{r^2}{4\sigma^2} \leq \frac{1}{8d}$.

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

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

□

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

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

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

where

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

and

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

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

Lemma 7. *For any $0 \leq t \leq 1$ and $\alpha \leq \frac{t^2}{8d}$,*

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

Proof. We can write

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

The first integral is simply the beta function, with

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

To upper bound the second integral, we apply the Taylor theorem with remainder to $(1-u)^{-1/2}$, obtaining

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

As a result,

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

where (iv) follows from Lemma 3, and the condition $\alpha \leq \frac{t^2}{8d}$. The result follows from the condition $\alpha \leq \frac{t^2}{8d}$. \square

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

Lemma 8.

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

A.2 Density-weighted cut and volume estimates

For notational ease, we write

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

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

Lemma 9. *Under the setup and conditions of Theorem 1, and for any $0 < r \leq \sigma/2d$,*

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

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

$$\text{cut}_{n,r} = \sum_{i=1}^n \sum_{j \neq i} \mathbf{1}(x_i \notin \mathcal{C}_\sigma) \mathbf{1}(x_j \in \mathcal{C}_\sigma) \mathbf{1}(\|x_i - x_j\| \leq r) \quad (\text{A.10})$$

and by linearity of expectation, we obtain

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

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

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

where the inequality follows from (A3), which implies $f(x) \leq \lambda$ for $x \in \mathcal{C}_\sigma \setminus \mathcal{C}$. Then, upper bounding the integral using Lemma 9 gives the final result. \square

Lemma 10. *Under the setup and conditions of Theorem 1, and for any $0 < r \leq \sigma/2d$,*

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

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

$$\text{vol}_{n,r} = \sum_{i=1}^n \sum_{j \neq i} \mathbf{1}(x_i \in \mathcal{C}_\sigma) \mathbf{1}(x_j \in B(x_i, r)) \quad (\text{A.11})$$

and by linearity of expectation we obtain

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

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

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

whence the claim then follows by Lemma 5. \square

To employ Lemmas 9 and 10 in the proof of Theorem 1, we must relate the random variable

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

to p_K and p_V .

In Lemma 11, we give probabilistic bounds on the $\text{cut}_{n,r}$, $\text{vol}_{n,r}$ and $\text{vol}_{n,r}^c$ in terms of p_K and p_V . These bounds are a straightforward consequence of Lemma 30, Hoeffding's inequality for U-statistics.

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

$$\frac{\text{cut}_{n,r}}{\binom{n}{2}} \leq p_K + \sqrt{\frac{\log(1/\delta)}{n}}, \text{ and } \frac{\text{vol}_{n,r}}{\binom{n}{2}}, \frac{\text{vol}_{n,r}^c}{\binom{n}{2}} \geq p_V - \sqrt{\frac{\log(1/\delta)}{n}}. \quad (\text{A.12})$$

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

Proof of Lemma 11. From (A.10) and (A.11), we see that $\text{cut}_{n,r}$ and $\text{vol}_{n,r}$, properly scaled, can be expressed as order-2 U -statistics,

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

with kernels

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

Similarly,

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

with kernel,

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

From Lemma 30 we therefore have

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

each with probability at least $1 - \delta$. The claim follows in light of (7), which implies $p_V^c \geq p_V$. \square

A.3 Proof of Theorem 1

The proof of Theorem 1 is more or less given by Lemmas 9, 10, and 11. All that remains is some algebra, which we take care of below.

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

$$\Phi_{n,r}(\mathcal{C}_\sigma[\mathbf{X}]) = \frac{p_K + \left(\frac{\text{cut}_{n,r}}{\binom{n}{2}} - p_K \right)}{p_V + \left(\frac{\min\{\text{vol}_{n,r}, \text{vol}_{n,r}^c\}}{\binom{n}{2}} - p_V \right)}. \quad (\text{A.13})$$

Assume (A.12) holds with respect to δ' , keeping in mind that this will happen with probability at least $1 - \delta$. Along with (A.13) this means

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

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

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

By Lemmas 9 and 10, we have

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

Then, by the choice of sample size in (8),

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

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

A.4 Mixing time on graphs

For $N \in \mathbb{N}$ and a collection $V = \{v_1, \dots, v_N\}$ of N vertices, take $G = (V, E)$ to be an undirected and unweighted graph, with associated adjacency matrix \mathbf{A} , degree matrix \mathbf{D} , and random walk matrix \mathbf{W} . We consider the stationary distribution π to be a mapping from V to $[0, 1]$, with $\pi(u) := \frac{\mathbf{D}_{uu}}{\text{vol}(V; G)}$ for $u \in V$. For $v \in V$, let $q : V \times V \rightarrow [0, 1]$ represent the m -step transition probability of the random walk over G originating from v , $q^{(m)}(v, u) = e_v \mathbf{W}^m e_u$. The relative pointwise mixing time over G is then

$$\tau_\infty(G) = \min \left\{ m : \forall u, v \in V, \frac{|q^{(m)}(v, u) - \pi(u)|}{\pi(u)} \leq 1/4 \right\}.$$

Two key quantities relate the mixing time to the expansion of subsets S of V . The *local spread* is defined as

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

For $f : V \rightarrow \mathbb{R}$, and $S \subseteq V$, write $f(S) := \sum_{u \in S} f(u)$. We collect some necessary facts about the local spread in Lemma 12.

Lemma 12.

- If $\pi(S) \leq s(G)$, then for every $u \in S$, $q^{(1)}(u, S^c) \geq \frac{1}{10}$.
- For any $v, u \in V$, and $m \in \mathbb{N}, m \geq 0$,

$$q^{(m)}(v, u) \leq \frac{\min_{v \in V} \{\pi(v)\}}{s(G)}.$$

Proof. If $\pi(S) \leq s(G)$, divide both sides by $\min_{v \in V} \{\pi(v)\}$ to obtain $|S| \leq 9 \min_{v \in V} \{\deg(v; G)\} / 10$, which implies $q^{(1)}(v, S^c) \geq 1/10$ for all $v \in S$. This implies the first statement.

The second statement follows from the fact $q^{(m)}(v, u) \leq 1 / \min_{v \in V} \{\deg(v; G)\}$ for any m . \square

The local spread facilitates conversion between $\tau_\infty(G)$ and the more easily manageable *total variation* mixing time, given by

$$\tau_1(G) = \min \left\{ m : \forall v \in V, \left\| \rho^{(m)}(v, \cdot) - \pi \right\|_{TV} \leq 1/4 \right\}$$

where $\rho^{(m)}(u, v) = \sum_{l=1}^m p^{(l)}(u, v)$ and $\|\rho(v, \cdot) - \pi\|_{TV} = \sum_{u \in V} |\rho(u, v) - \pi(v)|$ is the total variation norm between distributions $\rho(v, \cdot)$ and π .

Lemma 13.

$$\tau_\infty(G) \leq 2752 \cdot \tau_1(G) \log \left(4 \max \left\{ 1, \frac{1}{s(G)} \right\} \right)$$

Proof. Masking dependence on the starting vertex v for the moment, let

$$\Delta^{(m)}(u) = q^{(m)}(v, u) - \pi(u), \quad \delta^{(m)}(u) = \frac{\Delta^{(m)}(u)}{\pi(u)}$$

and denote $\mathbf{\Delta}^{(m)} = (\Delta^{(m)}(u))_{u \in V}$, $\mathbf{\delta}^{(m)} = (\delta^{(m)}(u))_{u \in V}$. For a vector $\mathbf{\Delta} = (\Delta_u)_{u \in V}$, the $L^p(\pi)$ norm is given by

$$\|\mathbf{\Delta}\|_{L^p(\pi)} = \left(\sum_{u \in V} (\Delta_u)^p \pi(u) \right)^{1/p}$$

To go between the $L^\infty(\pi)$ and $L^1(\pi)$ norms, we capitalize on a clever argument developed in [5]. For any $v, u \in V$, by the time-reversibility of the graph random walk,

$$\begin{aligned} \left| \frac{q^{(2m)}(v, u) - \pi(u)}{\pi(u)} \right| &= \left| \sum_{y \in V} \pi(y) \left(\frac{q^{(m)}(v, y) - \pi(y)}{\pi(y)} \right) \left(\frac{q^{(m)}(u, y) - \pi(y)}{\pi(y)} \right) \right| \\ &\leq \left\| (\mathbf{\delta}^{(m)}) \right\|_{L^1(\pi)} \left\| (\mathbf{\delta}^{(m)}) \right\|_{L^\infty(\pi)}, \end{aligned} \quad (\text{A.14})$$

with the second line following by Holder's inequality.

Now, we upper bound the second factor on the right hand side by observing

$$\begin{aligned} \left\| (\mathbf{\delta}^{(m)}) \right\|_{L^\infty(\pi)} &\leq \max \left\{ 1, \max_{u \in V} \frac{q^{(m)}(v, u)}{\pi(u)} \right\} \\ &\stackrel{(i)}{\leq} \max \left\{ 1, \frac{1}{s(G)} \right\} \end{aligned} \quad (\text{A.15})$$

where (i) follows from Lemma 12.

Set $m_0 = 1376\tau_1(G)$, which we will show implies $\|q^{(m)}(v, \cdot) - \pi\|_{TV} \leq \frac{1}{4}$. Now, we leverage the following well-known fact (see pg. 1 of [4]): for any $\epsilon > 0$, if $m \geq m_0 \cdot \log(1/\epsilon)$ then for any $v \in V$,

$$\|q^{(m)}(v, \cdot) - \pi\|_{TV} \leq \epsilon. \quad (\text{A.16})$$

But $\|q^{(m)}(v, \cdot) - \pi\|_{TV}$ is exactly $\|\mathbf{\delta}^{(m)}\|_{L^1(\pi)}$. Letting

$$m_1 = m_0 \cdot \log \left(4 \max \left\{ 1, \frac{1}{s(G)} \right\} \right)$$

by (A.14), (A.15), and (A.16), $\|\mathbf{\delta}^{(m)}\|_{L^\infty(\pi)} \leq 1/4$ for all $m \geq 2m_1$. Recall that

$$\left\| (\mathbf{\delta}^{(m)}) \right\|_{L^\infty(\pi)} = \sup_{u \in V} \left\{ \frac{|q^{(m)}(v, u) - \pi(u)|}{\pi(u)} \right\}.$$

and since none of the above depended on a specific choice of v , the supremum can be taken over all starting vertices v as well. Thus for all $m \geq 2m_1$, for any $v, u \in V$, $|q^m(v, u) - \pi(u)| \leq \frac{\pi(u)}{4}$.

Finally, it is known (pg.15 of [4]) that

$$\|\rho^{(m)}(v, \cdot) - \pi\|_{TV} \leq \frac{1}{4} \implies \|q^{(1376m)}(v, \cdot) - \pi\|_{TV} \leq \frac{1}{4}$$

and so $m_0 = 1376\tau_1(G)$ implies $\|q^{(m)}(v, \cdot) - \pi\|_{TV} \leq \frac{1}{4}$, completing the proof. \square

The second key quantity is the *conductance function*. For $\min_{u \in V} \{\pi(u)\} \leq t \leq 1/2$, the conductance function is defined as

$$\Phi(t; G) := \min_{\substack{S \subseteq V, \\ 0 < \pi(S) \leq t}} \Phi(S; G) \quad (\text{A.17})$$

where $\Phi(S; G)$ is the normalized cut of S in G given by (1), and in the minimization S may not be the empty set.

Lemma 14 leverages the conductance function and local spread to produce an upper bound on the total variation distance between $\rho_v^{(m)}$ and π .

Lemma 14. *If $D_{\min} > 10$, for any $v \in V$:*

$$\left\| \rho_v^{(m)} - \pi \right\|_{TV} \leq \max \left\{ \frac{1}{4}, \frac{1}{10} + \frac{70}{m} \left(\frac{20}{9} + \int_{t=s'(G)}^{1/2} \frac{4}{t\Phi^2(t; G)} dt \right) \right\}$$

where $s'(G) = s(G)/9$.

To prove Lemma 14 we first introduce a generalization of $\Phi(t; G) \cdot \Phi(t; G)$ known as a blocking conductance function.¹

Definition 1 (Blocking Conductance Function of [4]). *For $t_0 \geq \pi_{\min}$, a function $\phi(t; G) : [t_0, 1/2] \rightarrow [0, 1]$ is a blocking conductance function if for all $S \subset V$ with $\pi(S) = t \in [t_0, 1/2]$, either of the following hold:*

1. Exterior inequality. *For all $y \in [\frac{1}{2}t, t] : \phi_{\text{int}}(S) \geq \phi(\max\{t_0, y\})$*
2. Interior inequality. *For all $y \in [t, \frac{3}{2}t] : \phi_{\text{ext}}(S) \geq \phi(\max\{y, 1 - y\})$.*

where ϕ_{int} and ϕ_{ext} are defined respectively as

$$\begin{aligned} \phi_{\text{int}}(S) &= \sup_{\lambda \leq \pi(S)} \min_{\substack{B \subseteq S \\ \pi(B) \leq \lambda}} \frac{\lambda \text{cut}(S \setminus B, S^c; G)}{\text{vol}(V; G) [\pi(S)\pi(S^c)]^2} \\ \phi_{\text{ext}}(S) &= \sup_{\lambda \leq \pi(S)} \min_{\substack{B \subseteq S^c \\ \pi(B) \leq \lambda}} \frac{\lambda \text{cut}(S \setminus B, S^c; G)}{\text{vol}(V; G) [\pi(S)\pi(S^c)]^2} \end{aligned}$$

Theorem 1 (Theorem 3.2 of [4]). *Consider $\phi(t; G) : [t_0, 1/2] \rightarrow [0, 1]$ a blocking conductance function. Then, letting*

$$h^m(t_0) = \sup_{S: \pi(S) < t_0} (\rho_v^{(m)}(S) - \pi(S))$$

the following statement holds: if ϕ is a blocking conductance function,

$$\left\| \rho_v^{(m)} - \pi \right\|_{TV} \leq \max \left\{ \frac{1}{4}, h^1(t_0) + \frac{70}{m} \left(\frac{1}{\phi(t_0; G)} + \int_{t=t_0}^{1/2} \frac{4}{t\phi(t; G)} dt \right) \right\}$$

Note that in [4] this theorem is stated with respect to h^0 . However, in the subsequent proof it holds with respect to h^m , and it is observed that h^m is decreasing in m . For our purposes it is more useful to state it with respect to h^1 , as we have done.

Proof of Lemma 14. Consider the function $\phi_0(t, G) : [s(G), 1/2] \rightarrow [0, 1]$ defined by

$$\phi_0(t; G) = \begin{cases} \frac{1}{5}, & t = s'(G) \\ \frac{1}{4}\Phi^2(t; G), & t \in (s'(G), 1/2] \end{cases} \quad (\text{A.18})$$

¹For more details, see [4]

Lemma 15. *If $D_{\min} > 10$, ϕ_0 is a blocking conductance function.*

We take Lemma 15 as given, and defer the proof until after the proof of Lemma 14.

Lemma 15 and Theorem 1 together yield:

$$\|\rho^t - \pi\|_{TV} \leq \max \left\{ \frac{1}{4}, h^1(s'(G)) + \frac{70}{m} \left(5 + \int_{t=s'(G)}^{1/2} \frac{4}{t\Phi^2(t; G)} \right) \right\}$$

Then, $h^1(s'(G)) \leq 1/10$ follows exactly from the proof of Lemma 12, except now $\pi(S) \leq s'(G)$ results in the sharper bound of $\mathbf{q}_u^{(1)}(S^c) \geq 9/10$ for every $u \in S$. \square

Lemma 15. The condition $D_{\min} > 10$ ensures that $s(G) \geq \pi_{\min}$.

It is known that $\frac{1}{4}\Phi^2(x; G)$ satisfies the exterior inequality for all $t \in (\pi_{\min}, 1/2]$.

For $t = s'(G)$ we will instead use the interior inequality. For any S such that $\pi(S) \leq s'(G)$, the following statement holds: for every $u \in S$, $\text{cut}(u, S^c; G) \geq 9/10 \cdot \deg(u; G)$. Fixing $\lambda = \pi(S)/2$, we have

$$\begin{aligned} \phi_{\text{int}}(S) &\geq \min_{\substack{B \subset S \\ \pi(B) \leq \lambda}} \frac{\lambda \text{cut}(S \setminus B, S^c; G)}{\text{vol}(V; G) [\lambda(1 - \lambda)]^2} \\ &\geq \min_{\substack{B \subset S \\ \pi(B) \leq \lambda}} \frac{9\lambda \sum_{u \in S \setminus B} \deg(u; G)}{10 \text{vol}(V; G) [\lambda(1 - \lambda)]^2} \\ &\geq \frac{9\lambda^2}{20[\lambda^2(1 - \lambda)^2]} \geq \frac{9}{20}. \end{aligned}$$

\square

A.5 Population-level conductance profile.

We introduce the r -ball walk, a Markov chain over \mathcal{C}_σ with transition probability given by

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

where $\mathfrak{B}(\mathcal{C}_\sigma)$ is the Borel σ -algebra of \mathcal{C}_σ .

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

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

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

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

a bit of algebra verifies that

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

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

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

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

$$\tilde{\Phi}_{\mathbb{P},r}(\mathcal{S}) := \frac{\tilde{Q}_{\mathbb{P},r}(\mathcal{S}, \mathcal{S}^c)}{\min \{\tilde{\pi}_{\mathbb{P},r}(\mathcal{S}), \tilde{\pi}_{\mathbb{P},r}(\mathcal{S}^c)\}}, \quad (\mathcal{S} \in \mathfrak{B}(\mathcal{C}_\sigma))$$

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

$$\tilde{\Phi}_{\mathbb{P},r}(t) := \min_{\substack{\mathcal{S} \in \mathfrak{B}(\mathcal{C}_\sigma) \\ 0 < \tilde{\pi}_{\mathbb{P},r}(\mathcal{S}) \leq t}} \tilde{\Phi}_{\mathbb{P},r}(\mathcal{S}) \quad (0 < t \leq 1/2)$$

where $\mathcal{S}^c = \mathcal{C}_\sigma \setminus \mathcal{S}$.

Proposition 1. *Let \mathcal{C} satisfy Assumption (A1) for some $\lambda_\sigma \leq \Lambda_\sigma$ and Assumption (A4) for some convex set \mathcal{K} with diameter D , and measure-preserving mapping $g : \mathcal{K} \rightarrow \mathcal{C}_\sigma$ with biLipschitz constant L . Then, for any $0 < r < \frac{\sigma}{2\sqrt{d}}$, the continuous conductance function of the r -ball walk satisfies*

$$\tilde{\Phi}_{\mathbb{P},r}(t) > \frac{\lambda_\sigma^2 r}{2^{12} \Lambda_\sigma^2 D L \sqrt{d}}.$$

for any $0 < t \leq \frac{1}{2}$.

Similar results are already known (see e.g. [2]) when the density f is uniform (or log-concave) and \mathcal{C}_σ is itself a convex set – indeed, in this case stronger versions of it exist, though we will not require them. We first prove an analogous result for the special case of $f \equiv 1$ everywhere on \mathcal{C}_σ . For $x \in \mathcal{C}_\sigma, \mathcal{S} \in \mathfrak{B}(\mathcal{C}_\sigma)$, let

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

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

$$\tilde{\Phi}_{\nu,r}(\mathcal{S}) := \frac{\tilde{Q}_{\nu,r}(\mathcal{S}, \mathcal{S}^c)}{\min \{\pi_{\nu,r}(\mathcal{S}), \pi_{\nu,r}(\mathcal{S}^c)\}}, \quad \tilde{\Phi}_{\nu,r}(t) := \min_{\substack{\mathcal{S} \in \mathfrak{B}(\mathcal{C}_\sigma) \\ 0 < \pi_{\nu,r}(\mathcal{S}) \leq t}} \tilde{\Phi}_{\nu,r}(\mathcal{S}).$$

Lemma 16. *Let \mathcal{C} satisfy Assumption (A4) for some convex set \mathcal{K} with diameter D , and measure-preserving mapping $g : \mathcal{K} \rightarrow \mathcal{C}_\sigma$ with biLipschitz constant L . Then, for any $0 < r < \frac{\sigma}{2\sqrt{d}}$, the uniform conductance function of the r -ball walk satisfies*

$$\tilde{\Phi}_{\nu,r}(t) > \frac{r}{2^{12} D L \sqrt{d}}.$$

for any $0 < t \leq 1/2$.

Most of the technical work needed to show Proposition 1 involves proving Lemma 16. We defer this work to the next subsection, and first show that Proposition 1 is a simple consequence of Lemma 16 along with (A1).

Proof of Proposition 1. Assume Lemma 16 holds. Since $t \leq 1/2$, for any Borel $\mathcal{S} \subseteq \mathcal{C}_\sigma$ with $\tilde{\pi}_{\mathbb{P},r}(\mathcal{S}) \leq t$,

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

By (A1) we obtain

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

Noting that $\tilde{\Phi}_{\nu,r}(\mathcal{S}) = \tilde{\Phi}_{\nu,r}(\mathcal{S}^c)$, and as $0 < \tilde{\pi}_{\mathbb{P},r}(\mathcal{S}) < 1/2$, one of $0 < \pi_{\nu,r}(\mathcal{S}) \leq 1/2$ or $0 < \pi_{\nu,r}(\mathcal{S}^c) \leq 1/2$ must hold, the statement follows by Lemma 16.

A.6 Proof of Lemma 16.

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

Lemma 17 (Isoperimetry of Lipschitz embeddings of convex sets.). *Let \mathcal{C} satisfy Assumption (A4) for some convex set \mathcal{K} with diameter D , and measure-preserving mapping $g : \mathcal{K} \rightarrow \mathcal{C}_\sigma$ with biLipschitz constant L . Then, for any partition $(\Omega_1, \Omega_2, \Omega_3)$ of \mathcal{C}_σ ,*

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

The proof of Lemma 17 from first principles is non-trivial, even in the convex setting, and is a primary technical contribution of the seminal work [3], extended by [1] among others. Once the result is shown in the case where Ω is convex, however, generalizing to the setting implied Assumption (A4) is not difficult.

Proof of Lemma 17. For $\Omega_i, i = 1, 2, 3$, denote the preimage

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

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

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

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

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

By Theorem 2.2 of [3],

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

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

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

□

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

Lemma 18 (One-step distributions). *Let $u, v \in \mathcal{C}_\sigma$ be such that*

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

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

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

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

The key result needed to show Lemma 18 is regarding the volume of the overlap $B(u, r) \cap B(v, r)$.

Lemma 19. *Let $u, v \in \mathbb{R}^d$ be points such that $\|u - v\| \leq t \frac{r}{\sqrt{d}}$ for some $0 < t < 1/8$. Then,*

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

Proof. We will treat only the case where $\|u - v\| = t \frac{r}{\sqrt{d}}$; if they are closer together the overlap of the volume will only increase. Then, it is not hard to see that $I = B(u, r) \cap B(v, r)$ is comprised of the union of two disjoint spherical caps, and thus

$$\nu(I) = 2\nu(\text{Cap}_r(r(1 - \frac{t}{2\sqrt{d}}))).$$

From Lemma 6 we therefore obtain

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

where

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

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

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

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

Proof of Lemma 18. Let $S_1 \cup S_2 = \mathcal{C}_\sigma$ be an arbitrary partition of \mathcal{C}_σ . We will show that

$$\tilde{P}_{\nu,r}(u; S_1) - \tilde{P}_{\nu,r}(v; S_1) \leq 1 + \frac{3\sqrt{3}t}{4\sqrt{2\pi}} - \ell.$$

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

Now, note that

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

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

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

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

$$1 - \tilde{P}_{\nu,r}(u; S_2) - \tilde{P}_{\nu,r}(v; S_1) \leq 1 - \frac{1}{\nu_d r^d} \nu(\mathcal{C}_\sigma \cap I) \quad (\text{A.19})$$

As (A.19) demonstrates, the overlap of the one-step distributions is related to the volume of the intersection between $B(u, r)$ and $B(v, r)$ within \mathcal{C}_σ .

From here, some simple manipulations yield

$$\begin{aligned} \nu(\mathcal{C}_\sigma \cap I) &= \nu(I) - \nu(I \setminus \mathcal{C}_\sigma) \\ &\geq \nu(I) - \max \{ \nu(B(u, r) \setminus \mathcal{C}_\sigma), \nu(B(v, r) \setminus \mathcal{C}_\sigma) \} \\ &\geq \nu_d r^d \left(1 - \frac{3\sqrt{3}t}{4\sqrt{2\pi}} - (1 - \ell) \right) = \nu_d r^d \left(\ell - \frac{3\sqrt{3}t}{4\sqrt{2\pi}} \right) \end{aligned} \quad (\text{A.20})$$

where the last inequality follows from Lemma 19. (A.20) along with (A.19) then give the desired result. \square

Proof of Lemma 16. Let $S_1 \cup S_2 = \mathcal{C}_\sigma$, and let $\ell \geq 0$ satisfy $\ell \nu_d r^d \leq \ell_{\nu,r}(x)$. We will show that

$$\int_{S_1} \tilde{P}_{\nu,r}(x; S_2) d\pi_{\nu,r}(x) \geq \frac{\sqrt{2\pi} r \ell^4}{24DL\sqrt{d}} \min \{ \pi_{\nu,r}(S_1), \pi_{\nu,r}(S_2) \}$$

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

Now, consider the sets

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

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

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

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

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

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

a symmetric result holds.

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

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

By Lemma 18, we therefore have

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

and since $u \in S'_1, v \in S'_2$ were arbitrary, the same inequality holds for $\text{dist}(S'_1, S'_2)$. Therefore by Lemma 17

$$\text{vol}(S'_3) \geq \frac{2\sqrt{2\pi r\ell}}{3DL\sqrt{3d}} \min \{ \text{vol}(S'_1), \text{vol}(S'_2) \}$$

We now prove the desired result:

$$\begin{aligned} \int_{S_1} \tilde{P}_{\nu,r}(x; S_2) &= \frac{1}{2} \left(\int_{S_2} \tilde{P}_{\nu,r}(x; S_2) d\pi_{\nu,r}(x) \right) \\ &\geq \frac{\ell}{8} \pi_{\nu,r}(S'_3) \\ &\geq \frac{\ell^2}{8\nu(\mathcal{C}_\sigma)} \nu(S'_3) \\ &\geq \frac{\sqrt{2}r\ell^3}{12DL\sqrt{d}\nu(\mathcal{C}_\sigma)} \min \{ \nu(S'_1), \nu(S'_2) \} \\ &\geq \frac{\sqrt{2}r\ell^4}{12DL\sqrt{d}} \min \{ \pi_{\nu,r}(S'_1), \pi_{\nu,r}(S'_2) \} \\ &\geq \frac{\sqrt{2}r\ell^4}{24DL\sqrt{d}} \min \{ \pi_{\nu,r}(S_1), \pi_{\nu,r}(S_2) \}. \end{aligned}$$

A.7 Graph conductance profile and local spread.

Let $\tilde{G}_{n,r} := G_{n,r}[\mathcal{C}_\sigma[\mathbf{X}]]$ denote the subgraph induced by $\mathcal{C}_\sigma[\mathbf{X}]$. $\tilde{\Phi}_{n,r}(t)$, the graph conductance profile over $\tilde{G}_{n,r}$ is defined in (??). Write $\tilde{\pi}_{\min} := \min_{u \in \mathcal{C}_\sigma[\mathbf{X}]} \tilde{\pi}_{n,r}(u)$.

Proposition 2 (Lower bound on graph conductance profile). *Let \mathcal{C} satisfy Assumption (A1) for some $\lambda_\sigma \leq \Lambda_\sigma$ and Assumption (A4) for some convex set \mathcal{K} with diameter D , and measure-preserving mapping $g : \mathcal{K} \rightarrow \mathcal{C}_\sigma$ with biLipschitz constant L . Then, with probability one the following lower bound holds on the graph conductance function:*

$$\liminf_{n \rightarrow \infty} \tilde{\Phi}_{n,r}(t) > \frac{\lambda_\sigma^2 r}{\Lambda_\sigma^2 2^{12} DL \sqrt{d}} \quad (0 < t \leq \frac{1}{2})$$

In order to prove Proposition 2, we will need to split the analysis into two cases based on the size of $S \subseteq \mathcal{C}_\sigma[\mathbf{X}]$. For any graph $G = (V, E)$, let $\mathcal{L}(G) = \{S \subseteq V : \pi(S) \geq s(G)\}$ (where as usual π denotes the stationary distribution of a random walk over G .)

In Lemma 20, we establish a uniform lower bound on the normalized cut of all sets $S \in \mathcal{L}(\tilde{G}_{n,r})$.

Lemma 20. *Let \mathcal{C} satisfy Assumption (A1) for some $\lambda_\sigma \leq \Lambda_\sigma$ and Assumption (A4) for some convex set \mathcal{K} with diameter D , and measure-preserving mapping $g : \mathcal{K} \rightarrow \mathcal{C}_\sigma$ with biLipschitz constant L . Then with probability one,*

$$\liminf_{n \rightarrow \infty} \left\{ \min_{S \in \mathcal{L}(\tilde{G}_{n,r})} \tilde{\Phi}_{n,r}(S) \right\} \geq \frac{\lambda_\sigma^2 r}{\Lambda_\sigma^2 2^{12} DL \sqrt{d}}$$

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

The proof of Lemma 20 is nontrivial, and we devote the following subsection to it.

Lemma 21 shows that for the remaining small sets, the graph normalized cut is of constant order.

Lemma 21. *Let $G = (V, E)$ be an arbitrary undirected graph. Then, for non-empty subsets $S \subseteq V$,*

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

Proof. Observing that

$$\Phi(S; G) \geq \sum_{u \in S} q^{(1)}(u, S^c) \frac{\deg(u; G)}{\text{vol}(S)},$$

the claim follows by Lemma 12. □

Proposition 2 follows in light of the fact $\frac{\lambda_\sigma^2 r}{\Lambda_\sigma^2 2^{12} D L \sqrt{d}} < \frac{1}{10}$.

Lemma 22. *With probability one, the following bound holds:*

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

Proof. We rewrite $s(\tilde{G}_{n,r}) = \frac{9 \widetilde{\deg_{\min}^2}}{10 \text{vol}_{n,r}(\mathcal{C}_\sigma[\mathbf{X}])}$. Therefore by Lemma 32, with probability one

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

□

A.8 Proof of Lemma 20

Let $\mathfrak{B}(\mathcal{C}_\sigma)$ be the Borel σ -algebra of \mathcal{C}_σ , and define the conditional probability measures

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

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

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

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

If a sequence of transportation maps $\{T_n\}_{n \in \mathbb{N}}$ satisfies $\lim_{n \rightarrow \infty} \|\text{Id} - T_n\|_{L^1(\tilde{\mathbb{P}})} = 0$ (where Id is the identity mapping), we refer to it as a sequence of *stagnating transportation maps*. Proposition 5 of [6] establishes that, for open connected domains with Lipschitz boundaries, such stagnating transportation maps exist. In particular, as $\mathcal{C}_\sigma[\mathbf{X}]$ consists of $\tilde{n} = |\mathcal{C}_\sigma[\mathbf{X}]|$ points sampled independently from $\tilde{\mathbb{P}}$, and \mathcal{C}_σ is a connected domain with Lipschitz boundary, the following implication is immediate.

Lemma 23. *With probability one, there exists a sequence of transportation maps $\{T_n\}_{n \in \mathbb{N}}$, $T_n : \mathcal{C}_\sigma \rightarrow \mathcal{C}_\sigma[\mathbf{X}]$ such that the following statement holds:*

$$\limsup_{n \rightarrow \infty} \frac{\tilde{n}^{1/d} \|\text{Id} - T_n\|_{L^\infty(\tilde{\mathbb{P}})}}{(\log \tilde{n})^{p_d}} \leq C \quad (\text{A.21})$$

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

Note that although \mathcal{C}_σ is closed not open, as $\nu(\partial\mathcal{C}_\sigma) = 0$ we may apply Proposition 5 of [6] to \mathcal{C}_σ° , and (A.21) will hold for arbitrary extension of T_n to \mathcal{C}_σ .

Graph cuts to continuous cuts. We use stagnating transportation maps to relate the discrete graph normalized cuts to the continuous normalized cuts discussed in Section A.5. We fix some notation, letting $r_n^\pm = r \pm \|\text{Id} - T_n\|_{L^\infty(\tilde{\mathbb{P}})}$ for a transportation map T_n .

Lemmas 24 and 25 provide the necessary bounds for the cut and vol functionals in terms of continuous analogues.

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

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

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

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

Now, we proceed

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

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

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

which is the desired upper bound. \square

Lemma 25. Let $S \subseteq \mathcal{C}_\sigma[\mathbf{X}]$, and let T_n be a transportation map between $\tilde{\mathbb{P}}$ and $\tilde{\mathbb{P}}_n$. Then, letting $\mathcal{S} = T_n^{-1}(S)$,

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

Proof. As in the proof of Lemma 24, let $\varphi : \mathcal{C}_\sigma[\mathbf{X}] \rightarrow \{0, 1\}$ be the characteristic function for S .

We proceed according to a very similar set of steps as Lemma 24:

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

We conclude similarly to the proof of Lemma 24,

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

□

Lemma 26. Let $\{T_n\}_{n \in \mathbb{N}}$ be a sequence of transportation maps from $\tilde{\mathbb{P}}$ to $\tilde{\mathbb{P}}_n$, and fix $S \subseteq \mathcal{C}_\sigma[\mathbf{X}]$. Then, letting $\mathcal{S} = T_n^{-1}(S)$,

$$\tilde{\Phi}_{n,r}(S) \geq \frac{\int_{\mathcal{C}_\sigma} \ell_{\mathbb{P},r_n^-}(x) f(x) dx}{\int_{\mathcal{C}_\sigma} \ell_{\mathbb{P},r_n^+}(x) f(x) dx} \frac{\min \left\{ \tilde{\pi}_{\mathbb{P},r_n^-}(\mathcal{S}), \tilde{\pi}_{\mathbb{P},r_n^-}(\mathcal{S}^c) \right\}}{\min \left\{ \tilde{\pi}_{\mathbb{P},r_n^+}(\mathcal{S}), \tilde{\pi}_{\mathbb{P},r_n^+}(\mathcal{S}^c) \right\}} \tilde{\Phi}_{\mathbb{P},r_n^-}(S) \quad (\text{A.22})$$

Proof. By Lemmas 24 and 25,

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

Then, noting that $\mathcal{S}^c = T_n^{-1}(\mathcal{S}^c)$, Lemmas 24 and 25 also imply

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

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

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

and the proof is complete. □

Perturbation analysis. If $r_n^- = r = r_n^+$, (A.22) would simplify to $\tilde{\Phi}_{n,r}(S) \geq \tilde{\Phi}_{\mathbb{P},r}(S)$. We show that this conclusion is robust to asymptotically negligible perturbations of r .

Lemma 27 (Continuity of local conductance). *For $\{T_n\}_{n \in \mathbb{N}}$ a sequence of stagnating transportation maps,*

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

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

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

and therefore

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

We upper bound the remainder term

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

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

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

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

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

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

Lemma 28 (Continuity of stationary distribution). *Let $c > 0$ be a fixed constant, and $\{T_n\}_{n \in \mathbb{N}}$ be a sequence of stagnating transportation maps. Then,*

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

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

Proof. It will be sufficient to show that

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

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

The proof proceeds similarly to Lemma 27. Letting

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

we may rewrite

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

By the stagnating property of T_n ,

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

where the convergence occurs as $n \rightarrow \infty$. On the other hand, by hypothesis

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

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

Lemma 29 (Stationary distribution lower bound). *Let $\{T_n\}_{n \in \mathbb{N}}$ be a sequence of stagnating transportation maps from \mathbb{P} to \mathbb{P}_n . Then with probability one,*

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

Proof. Fix $\epsilon > 0$, and let $S \in \mathcal{L}(\tilde{G}_{n,r})$ be arbitrary, writing $\mathcal{S} := T_n^{-1}(S)$. Letting $\mathcal{R}_n(x)$ be as in the proof of Lemma 28, we have

$$\tilde{\pi}_{n,r}(S) \leq \tilde{\pi}_{\mathbb{P}, r_n^-}(\mathcal{S}) + \frac{\int_{\mathcal{S}} \int_{\mathcal{R}_n(x)} f(y) f(x) dy dx}{\int_{\mathcal{C}_{\sigma}} \ell_{\mathbb{P}, r_n^-}(x) f(x) dx}. \quad (\text{A.23})$$

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

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

which along with (A.23) implies

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

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

By Lemma 22, with probability one as $n \rightarrow \infty$,

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

and the claim is shown. \square

Proof of Lemma 20 By Lemma 23, with probability one there exists a sequence of stagnating transportation maps from $\tilde{\mathbb{P}}$ to $\tilde{\mathbb{P}}_n$, which we will denote $\{T_n\}_{n \in \mathbb{N}}$.

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

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

where $r_n^\pm := r \pm \|\text{Id} - T_n\|_{L^\infty(\tilde{\mathbb{P}})}$. By Proposition 1 and Lemma 26 we have that

$$\begin{aligned}\tilde{\Phi}_{n,r}(S) &\geq \xi_n \gamma_n(S) \tilde{\Phi}_{\mathbb{P},r_n^-}(T_n^{-1}(S)) \\ &\geq \xi_n \gamma_n(S) \frac{\lambda_\sigma^2 r_n^-}{2^{12} \Lambda_\sigma^2 D L \sqrt{d}}.\end{aligned}\tag{A.24}$$

By Lemma 27, with probability one

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

By Lemma 29, letting c be any constant satisfying $0 < c < \frac{\lambda_\sigma^2 \nu_d r^d}{20 \Lambda_\sigma}$, with probability one there exists some $m \in \mathbb{N}$ such that for all $n \geq m$,

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

and therefore by Lemma 28

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

As Lemma 23 implies $r_n^- \rightarrow r$ with probability one, an application of Slutsky's Theorem to (A.24) completes the proof.

A.9 Proof of Theorem 2

Lemma 14 implies

$$\tau_1(\tilde{G}_{n,r}) \leq 490 \cdot \left(\frac{20}{9} + \int_{s'(\tilde{G}_{n,r})}^{1/2} \frac{4}{t \tilde{\Phi}_{n,r}^2(t)} dt \right)\tag{A.25}$$

where we recall $s'(G) = \frac{s(G)}{9}$.

By Lemmas 20 and 22 for sufficiently large n , and for any $0 < t \leq \frac{1}{2}$

$$s(\tilde{G}_{n,r}) \geq \frac{\lambda_\sigma^2 \nu_d r^d}{20 \Lambda_\sigma}, \quad \tilde{\Phi}_{n,r}(t) > \frac{\lambda_\sigma^2 r}{\Lambda_\sigma^2 2^{12} D L \sqrt{d}}$$

which along with (A.25) and Lemma 13 implies the desired result.

A.10 Concentration inequalities

Given a symmetric kernel function $k : \mathcal{X}^m \rightarrow \mathbb{R}$, and data $\{x_1, \dots, x_n\}$, we define the *order- m U statistic* to be

$$U := \frac{1}{\binom{n}{m}} \sum_{1 \leq i_1 < \dots < i_m \leq n} k(x_{i_1}, \dots, x_{i_m})$$

For Lemmas 30, let $X_1, \dots, X_n \in \mathcal{X}$ be independent and identically distributed. We will additionally assume the order- m kernel function k satisfies the boundedness property $\sup_{x_1, \dots, x_m} |k(x_1, \dots, x_m)| \leq 1$.

Lemma 30 (Hoeffding's inequality for U -statistics.). *For any $t > 0$,*

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

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

$$\begin{aligned} U &\leq \mathbb{E}U + \sqrt{\frac{m \log(1/\delta)}{2n}}, \\ U &\geq \mathbb{E}U - \sqrt{\frac{m \log(1/\delta)}{2n}} \end{aligned}$$

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

We will employ a sharper concentration inequality for $\sum_{i=1}^n X_i$.

Lemma 31. *Let $X_i \in \{0, 1\}$ for $i = 1, \dots, n$ and let $\mu = \mathbb{E}(\sum_{i=1}^n X_i)$. Then,*

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

We will require asymptotic bounds on several different degree and volume functionals, which follow from Lemma 31. Let

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

Lemma 32. *Under the conditions of Theorem 4, for any $\epsilon > 0$, each of the following bounds hold with probability tending to one as $n \rightarrow \infty$:*

$$\frac{\widetilde{\deg}_{\min}}{n}, \frac{\deg'_{\min}}{n} \geq (1 - \epsilon) \frac{6}{25} \lambda_\sigma r^d \nu_d \quad (\text{A.26})$$

$$\frac{\deg_{\min}}{n} \geq (1 - \epsilon) \lambda_{\min} r^d \nu_d \quad (\text{A.27})$$

$$\frac{\deg_{\max}}{n} \leq (1 + \epsilon) \Lambda_\sigma r^d \nu_d \quad (\text{A.28})$$

$$\frac{|\mathcal{C}_\sigma[\mathbf{X}]|}{n} \leq (1 + \epsilon) \mathbb{P}(\mathcal{C}_\sigma) \quad (\text{A.29})$$

$$\frac{1}{2} \text{vol}_0 \leq \text{vol}_{n,r}(\mathcal{C}_\sigma[\mathbf{X}]) \leq \frac{3}{2} \text{vol}_0 \quad (\text{A.30})$$

Additionally, with probability one,

$$\begin{aligned} \liminf_{n \rightarrow \infty} \frac{\widetilde{\deg}_{\min}}{n} &\geq \frac{6}{25} \lambda_\sigma r^d \nu_d \\ \limsup_{n \rightarrow \infty} \frac{\widetilde{\text{vol}}_{n,r}(\mathcal{C}_\sigma[\mathbf{X}])}{n(n-1)} &\leq \mathbb{P}(\mathcal{C}_\sigma) \Lambda_\sigma \nu_d r^d \end{aligned}$$

Proof. We note that for any $x \in \mathcal{C}_\sigma \cup \mathcal{C}'_\sigma$, by (A1) and (A5) along with Lemma 5,

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

Then, by Lemma 31 along with a union bound

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

As the same bound holds for \deg'_{\min} , (A.26) is shown. (A.27) and (A.28) follow from similar reasoning. (A.29) follows directly from the law of large numbers.

Finally, note that

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

can be rewritten as a U-statistic, and furthermore

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

Therefore by Lemma 30

$$\left| \text{vol}_{n,r}(\mathcal{C}_\sigma[\mathbf{X}]) - n(n-1) \int_{\mathcal{C}_\sigma} \mathbb{P}(B(x, r)) dx \right| \leq \frac{3}{20} n(n-1) \int_{\mathcal{C}_\sigma} \mathbb{P}(B(x, r)) dx \quad (\text{A.31})$$

with probability at least $1 - 2 \exp(-\frac{3}{20} n \int_{\mathcal{C}_\sigma} \mathbb{P}(B(x, r)) dx) \rightarrow 0$ as $n \rightarrow \infty$. Then, application of the triangle inequality to (A.31), along with the range $\text{vol}_0 \in [3/4, 5/4] \cdot n(n-1) \int_{\mathcal{C}_\sigma} \mathbb{P}(B(x, r)) dx$ given by (10), yields

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

By similar reasoning to (A.31), for any $\epsilon > 0$, letting $\mu = \int_{\mathcal{C}_\sigma} \mathbb{P}(B(x, r)) dx \geq \mathbb{E}\left(\frac{\widetilde{\text{vol}}_{n,r}(\mathcal{C}_\sigma[\mathbf{X}])}{n(n-1)}\right)$ we have

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

The final two statements therefore both follow from the Borel-Cantelli Lemma. \square

A.11 Proof of Theorem 3

In this section, we prove both Theorem 3 and the relevant portions of Corollary 1.

Lemma 33 is a restatement of Lemma 3.4 of [7], translated into notation consistent with the rest of this paper. It is stated with respect to an arbitrary $A \subseteq G = (V, E)$, and a good set $A^g \subseteq A$ with $\text{vol}(A^g; G) \geq \text{vol}(A; G)/2$.

Lemma 33. *Let $\epsilon \geq 0$, $v \in A^g$, $\alpha \leq \frac{\Psi(A; G)}{9}$, and $\text{vol}_0 \in [1-c, 1+c]\text{vol}(A; G)$ for $0 < c < 1$. Consider taking sweep cuts of the **approximate PPR vector** $\mathbf{p}^{(\epsilon)} := \mathbf{p}^{(\epsilon)}(v, \alpha; G)$ of the form*

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

²The sweep cuts of Lemma 3.4 of [7] are stated directly with respect to $\text{vol}(A; G)$, rather than vol_0 ; this discrepancy leads to the extra factors of $1-c$ and $1+c$ which appear in our bounds as opposed to theirs.

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

$$\text{vol}(S_\beta \setminus A; G) \leq \frac{2\Phi(A; G)}{\alpha\beta(1-c)}\text{vol}(A; G), \quad (\text{A.32})$$

$$\text{vol}(A \setminus S_\beta; G) \leq \left(\frac{2\Phi(A; G)}{\alpha(\frac{3}{5} - \beta(1+c))} + 8\Phi(A; G) \right) \text{vol}(A; G), \quad (\text{A.33})$$

Proof of Theorem 3. Observe that by Theorem ?? and the upper bounds $\alpha \leq \frac{\Psi(\theta)}{9}$, $r < \sigma/2d$ given in (10), with probability one as $n \rightarrow \infty$, $\alpha \leq \frac{\Psi_{n,r}(\mathcal{C}_\sigma[\mathbf{X}])}{9}$. Moreover, i) we restrict $v \in \mathcal{C}_\sigma[\mathbf{X}]^g$; ii) the output of Algorithm 1 $\hat{C} = S_\beta$ for some $1/40 < \beta < 1/11$; iii) by Lemma 32, $\text{vol}_0 \in [1/2, 3/2]\text{vol}_{n,r}(\mathcal{C}_\sigma[\mathbf{X}])$, and iv) the PPR vector $\mathbf{p}(v, \alpha; G_{n,r})$ is simply an approximate PPR vector with $\epsilon = 0$. All conditions of Lemma 33 are therefore satisfied, and consequently by (A.32), (A.33), Theorem 1, and the lower bound $\alpha \geq \Psi(\theta)/10$, the following bounds hold with probability tending to 1 as $n \rightarrow \infty$:

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

Noting that

$$\begin{aligned} \text{vol}_{n,r}(\hat{C} \setminus \mathcal{C}_\sigma[\mathbf{X}]) &\geq |\hat{C} \setminus \mathcal{C}_\sigma[\mathbf{X}]| \min_{x \in \hat{C} \setminus \mathcal{C}_\sigma[\mathbf{X}]} \{\deg_{n,r}(x)\} \\ \text{vol}_{n,r}(\mathcal{C}_\sigma[\mathbf{X}] \setminus \hat{C}) &\geq |\mathcal{C}_\sigma[\mathbf{X}] \setminus \hat{C}| \min_{x \in \mathcal{C}_\sigma[\mathbf{X}] \setminus \hat{C}} \{\deg_{n,r}(x)\}, \text{ and} \\ \text{vol}_{n,r}(\mathcal{C}_\sigma[\mathbf{X}]) &\leq |\mathcal{C}_\sigma[\mathbf{X}]| \max_{x \in \mathcal{C}_\sigma[\mathbf{X}]} \{\deg_{n,r}(x)\} \end{aligned}$$

the claim follows in light of Lemma 32.

Proof of Corollary 1 misclassification rates. The choice $\epsilon = \frac{1}{20\text{vol}_0}$, along with Lemma 32 and the upper bound on vol_0 given by (10), together imply that $\epsilon < \frac{1}{10}\text{vol}_{n,r}(\mathcal{C}_\sigma[\mathbf{X}])$ with probability tending to one as $n \rightarrow \infty$. We may therefore apply Lemma 33 with respect to $\mathbf{p}^{(\epsilon)}$. All other aspects of the proof of Theorem 3 may be applied exactly as stated.

A.12 Proof of Theorem 4

We begin with some notation. Let $\tilde{\mathbf{D}}$ and $\tilde{\mathbf{W}}$ be the degree and random walk matrices for the induced subgraph $G_{n,r}(\mathcal{C}_\sigma[\mathbf{X}]) = \tilde{G}_{n,r}$, with vertex set $\tilde{V} = \mathcal{C}_\sigma[\mathbf{X}]$ and edge set $\tilde{E} = \{(u, v) \in G_{n,r} : u, v \in \mathcal{C}_\sigma[\mathbf{X}]\}$. Consider the function $\bar{\pi}_{n,r} : \mathcal{C}_\sigma[\mathbf{X}] \rightarrow [0, 1]$

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

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

Lemma 34 provides uniform error bounds on $\mathbf{p}(u) - \bar{\pi}_{n,r}(u)$ over all $x \in \mathbf{X}$ (where in an abuse of notation we consider $\bar{\pi}_{n,r}(x) = 0$ for $x \in \mathbf{X} \setminus \mathcal{C}_\sigma[\mathbf{X}]$).

Lemma 34. Let $0 < r < \sigma$ and $\alpha \leq \frac{\Psi_{n,r}(\mathcal{C}_\sigma[\mathbf{X}])}{9}$. Then the following statement holds: there exists a good set $\mathcal{C}_\sigma[\mathbf{X}]^g \subseteq \mathcal{C}_\sigma[\mathbf{X}]$ with $\text{vol}(\mathcal{C}_\sigma[\mathbf{X}]^g; G_{n,r}) \geq \text{vol}(\mathcal{C}_\sigma[\mathbf{X}]; G_{n,r})/2$ such that the following bounds hold with respect to $\mathbf{p} := \mathbf{p}(v, \alpha; G_{n,r})$ for any $v \in \mathcal{C}_\sigma[\mathbf{X}]^g$:

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

$$\mathbf{p}(u) \geq \frac{4}{5} \bar{\pi}_{n,r}(u) - \frac{20\Phi_{n,r}(\mathcal{C}_\sigma[\mathbf{X}])}{\Psi_{n,r}(\mathcal{C}_\sigma[\mathbf{X}]) \deg_{\min}} \quad (\text{A.34})$$

- Let $\mathcal{C}' \neq \mathcal{C} \in \mathbb{C}_f(\lambda)$ satisfy (A3) with respect to \mathcal{C} . Then for each $u' \in \mathcal{C}'[\mathbf{X}]$,

$$\mathbf{p}(u') \leq \frac{20\Phi_{n,r}(\mathcal{C}_\sigma[\mathbf{X}])}{\Psi_{n,r}(\mathcal{C}_\sigma[\mathbf{X}]) \deg'_{\min}}. \quad (\text{A.35})$$

Proof. The proof of Lemma 34 is lengthy, but not difficult. Given starting distribution q with $\text{supp}(q) \subseteq \mathcal{C}_\sigma[\mathbf{X}]$, we let

$$\tilde{\mathbf{p}}_q = \alpha q + (1 - \alpha) \tilde{\mathbf{p}}_q \tilde{\mathbf{W}} \quad (\text{A.36})$$

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

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

$$\begin{aligned} \ell_t &:= e_v(\mathbf{W}\tilde{\mathbf{I}})^t(\mathbf{I} - \mathbf{D}^{-1}\tilde{\mathbf{D}}), \quad \ell := \sum_{t=0}^{\infty} (1 - \alpha)^t \ell_t, \\ s_t &:= e_v(\mathbf{W}\tilde{\mathbf{I}})^t(\mathbf{W}\tilde{\mathbf{I}}^c), \quad s := \sum_{t=0}^{\infty} (1 - \alpha)^t s_t. \end{aligned} \quad (\text{A.37})$$

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

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

We first prove (A.34), and begin by restating some results of [7], adapted to our notation. By Lemma 3.1 of [7], there exists a good set $\mathcal{C}_\sigma[\mathbf{X}]^g \subseteq \mathcal{C}_\sigma[\mathbf{X}]$ with $\text{vol}(\mathcal{C}_\sigma[\mathbf{X}]^g; G_{n,r}) \geq \text{vol}(\mathcal{C}_\sigma[\mathbf{X}]; G_{n,r})/2$ such that for every $v \in \mathcal{C}_\sigma[\mathbf{X}]^g$

$$\mathbf{p}(u) \geq \tilde{\mathbf{p}}_v(u) - \tilde{\mathbf{p}}_\ell(u), \quad \text{and} \quad \|\ell\|_1 \leq \frac{2\Phi_{n,r}(\mathcal{C}_\sigma[\mathbf{X}])}{\alpha}. \quad (\text{A.38})$$

(The result $\|\ell\|_1 \leq \frac{2\Phi_{n,r}(\mathcal{C}_\sigma[\mathbf{X}])}{\alpha}$ is the only result in the proof of Theorem 4 which relies on the restriction $v \in \mathcal{C}_\sigma[\mathbf{X}]^g$.)

If additionally $\alpha \leq \frac{\Psi_{n,r}(\mathcal{C}_\sigma[\mathbf{X}])}{9}$, then by Corollary 3.3 of [7], for every $u \in \mathcal{C}_\sigma[\mathbf{X}]$

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

and along with (A.38), we obtain

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

We proceed to show the upper bound $\tilde{\mathbf{p}}_\ell(u) \leq \|\ell\|_1 / \deg_{\min}$, whence (A.34) follows by (A.38). We note two facts regarding $\tilde{\mathbf{p}}_\ell(u)$, which hold for all $u \in \mathcal{C}[\mathbf{X}]$. First, since $r < \sigma$, $(u, w) \notin G_{n,r}$ for any $w \notin \mathcal{C}_\sigma$. As a result,

for all $t \geq 1$, $\ell_t(u) = 0$ and by extension, $\ell(u) = 0$ as well. Second, for any q such that $\sum_{w \in \mathcal{C}_\sigma[\mathbf{X}]} q(w) \leq 1$, and any $t \geq 1$,

$$\begin{aligned} q \widetilde{\mathbf{W}}^t(u) &\leq \|q\|_1 \|\widetilde{\mathbf{W}}_{\cdot u}\|_\infty \\ &\leq \frac{1}{\deg_{\min}} \end{aligned} \tag{A.39}$$

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

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

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

and (A.34) is proved.

We turn to showing (A.35). By Lemma 35, for all $u' \notin \mathcal{C}_\sigma[\mathbf{X}]$,

$$\mathbf{p}_v(u') \leq \mathbf{p}_s(u').$$

Note that by (A3), for every $u \in \mathcal{C}_\sigma[\mathbf{X}]$, $(u', u) \notin E$ and therefore $s(u') = 0$. Some manipulations, similar to those in the preceding part of the proof, yield a lower bound on $\mathbf{p}_v(u')$ in terms of $\|s\|_1$:

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

where the last inequality follows from precisely the same reasoning as (A.39). The claim follows in light of Lemma 37, along with (A.38). \square

We turn now to the proof of Theorem 4. Let $\widetilde{\pi}_{\min} := \min_{u \in \mathcal{C}[\mathbf{X}]} \widetilde{\pi}_{n,r}(u)$.

Proof of Theorem 4.

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

Observe that by Theorem ?? and the upper bounds $\alpha \leq \frac{\Psi}{9}$ and $r < \sigma/2d$ given in (10), with probability one as $n \rightarrow \infty$, $\alpha \leq \frac{\Psi_{n,r}(\mathcal{C}_\sigma[\mathbf{X}])}{9}$. As a result, with probability one as $n \rightarrow \infty$, the conditions of Theorem 4 subsume the requirements of Lemma 34, and by that lemma, the following inequalities hold for each $u \in \mathcal{C}[\mathbf{X}]$, $u' \in \mathcal{C}'[\mathbf{X}]$:

$$\begin{aligned} \mathbf{p}(u) &\geq \frac{4}{5} \bar{\pi}_{n,r}(u) - \frac{2\Phi_{n,r}(\mathcal{C}_\sigma[\mathbf{X}])}{\alpha \widetilde{\deg}_{\min}} \\ \mathbf{p}(u') &\leq \frac{2\Phi_{n,r}(\mathcal{C}_\sigma[\mathbf{X}])}{\alpha \deg'_{\min}} \end{aligned} \quad (\text{A.40})$$

We will show that under the conditions of Theorem 4, the following bounds hold with probability tending to one as $n \rightarrow \infty$:

$$\begin{aligned} \frac{2\widetilde{\Phi}_{n,r}(\mathcal{C}_\sigma[\mathbf{X}])}{\alpha \widetilde{\deg}_{\min}} &\leq \frac{\widetilde{\deg}_{\min}}{125 \text{vol}_{n,r}(\mathcal{C}_\sigma[\mathbf{X}])} \\ \frac{2\widetilde{\Phi}_{n,r}(\mathcal{C}_\sigma[\mathbf{X}])}{\alpha \deg'_{\min}} &\leq \frac{\deg'_{\min}}{125 \text{vol}_{n,r}(\mathcal{C}_\sigma[\mathbf{X}])} \end{aligned} \quad (\text{A.41})$$

The bounds in (A.41) along with (A.40) imply that for each $u \in \mathcal{C}[\mathbf{X}]$,

$$\begin{aligned} \frac{\mathbf{p}(u)}{\deg_{n,r}(u)} &\geq \frac{4}{5} \frac{\widetilde{\deg}_{n,r}(u)}{\deg_{n,r}(u) \text{vol}_{n,r}(\mathcal{C}_\sigma[\mathbf{X}])} - \frac{\widetilde{\deg}_{\min}}{125 \deg_{n,r}(u) \text{vol}_{n,r}(\mathcal{C}_\sigma[\mathbf{X}])} \\ &\geq \frac{22}{125 \text{vol}_{n,r}(\mathcal{C}_\sigma[\mathbf{X}])} \quad (\text{by (A.27) and (A.28), applied with } \epsilon = \frac{1}{47}) \\ &\geq \frac{44}{375 \text{vol}_0}. \quad (\text{by (A.30)}) \end{aligned}$$

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

$$\begin{aligned} \frac{\mathbf{p}(u')}{\deg_{n,r}(u')} &\leq \frac{\deg'_{\min}}{\deg_{n,r}(u') 125 \text{vol}_{n,r}(\mathcal{C}_\sigma[\mathbf{X}])} \\ &\leq \frac{1}{125 \text{vol}_{n,r}(\mathcal{C}_\sigma[\mathbf{X}])} \\ &\leq \frac{8}{375 \text{vol}_0}. \quad (\text{by (A.30)}) \end{aligned}$$

and as every sweep cut S_β under consideration in Algorithm 1 satisfies $8/375 < \frac{\beta}{\text{vol}_0} < 44/375$, (6) must hold for the output $\widehat{C} = S_{\beta^*}$.

It remains to prove (A.41). Apply Lemma 32 with $\epsilon = 1/11$ to obtain

$$\frac{\deg_{\min}'^2}{\text{vol}_{n,r}(\mathcal{C}_\sigma[\mathbf{X}])}, \frac{\widetilde{\deg}_{\min}^2}{\text{vol}_{n,r}(\mathcal{C}_\sigma[\mathbf{X}])} \geq \left(\frac{6}{25}\right)^2 \frac{\lambda_\sigma^2 \nu_d r^d}{\Lambda_\sigma \mathbb{P}(\mathcal{C}_\sigma)} \quad (\text{A.42})$$

where the bound holds with probability tending to one as $n \rightarrow \infty$.

By Theorem 1 and the lower bound $\alpha \geq \frac{1}{10\Psi}$ given by (10), with probability tending to one as $n \rightarrow \infty$,

$$\frac{2\Phi_{n,r}(\mathcal{C}_\sigma[\mathbf{X}])}{\alpha} \leq 20\kappa_1(\mathcal{C}). \quad (\text{A.43})$$

(A.43), (A.42) and (16) together show (A.41).

Proof of Corollary 1: consistent cluster estimation. Using (17) with $\epsilon = \frac{1}{20}\text{vol}_0$ along with (A.34) we obtain

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

and similar reasoning as used in the proof of Theorem 4 yields

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

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

A.13 Linear Algebra Facts

We state here a number of basic facts which follow from matrix manipulations, which are used in the proof of Theorem 4.

Lemma 35. *For any $v \in \mathcal{C}_\sigma[\mathbf{X}]$ and $u \notin \mathcal{C}_\sigma[\mathbf{X}]$,*

$$\mathbf{p}_v(u) \leq \mathbf{p}_s(u)$$

where s is defined as in (A.37) and depends implicitly upon v .

Proof. The statement follows from Lemma 36 along with a series of algebraic manipulations,

$$\begin{aligned} \mathbf{p}_v(u) &= \alpha \sum_{T=0}^{\infty} (1-\alpha)^T (e_v \mathbf{W}_n^T)(u) \\ &= \alpha \sum_{T=1}^{\infty} (1-\alpha)^T (e_v \mathbf{W}_n^T)(u) \\ &\leq \alpha \sum_{T=1}^{\infty} (1-\alpha)^T \left(\sum_{t=0}^{T-1} s_t \mathbf{W}_n^{T-t-1} \right)(u) && \text{(Lemma 36)} \\ &= \alpha \sum_{t=0}^{\infty} \sum_{T=t+1}^{\infty} (1-\alpha)^T (s_t \mathbf{W}_n^{T-t-1})(u) \\ &= \alpha \sum_{t=0}^{\infty} \sum_{\Delta=0}^{\infty} (1-\alpha)^{\Delta+t+1} (s_t \mathbf{W}_n^{\Delta})(u) \\ &\leq \alpha \sum_{t=0}^{\infty} \sum_{\Delta=0}^{\infty} (1-\alpha)^{\Delta+t} (s_t \mathbf{W}_n^{\Delta})(u) \\ &= \alpha \sum_{\Delta=0}^{\infty} (1-\alpha)^{\Delta} (s \mathbf{W}_n^{\Delta})(u) \\ &= \mathbf{p}_s(u) \end{aligned}$$

□

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

Lemma 36. Let $G = (V, E)$ be an unweighted, undirected graph with associated random walk matrix \mathbf{W} . For any $T \in \mathbb{N}, T \geq 1$, $q \in \mathbb{R}^{|V|}$, and $S \subseteq V$

$$q\mathbf{W}^T = \sum_{t=0}^{T-1} s_t \mathbf{W}^{T-t-1} + q(\mathbf{W}\mathbf{I}_S)^T \quad (\text{A.44})$$

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

$$q\mathbf{W}^T(u) = \sum_{t=0}^{T-1} (s_t \mathbf{W}^{T-t-1})(u) \quad (\text{A.45})$$

Proof. We show (A.44), from which (A.45) is an immediate consequence.

To show (A.44), we proceed by induction on T . When $T = 1$,

$$q\mathbf{W} = q\mathbf{W}\mathbf{I}_S + q\mathbf{W}\mathbf{I}_{S^c} = q\mathbf{W}\mathbf{I}_S + s_0.$$

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

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

and the proof is complete. \square

Lemma 37. Letting s_t, ℓ_t and ℓ be as in (A.37),

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

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

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

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

\square

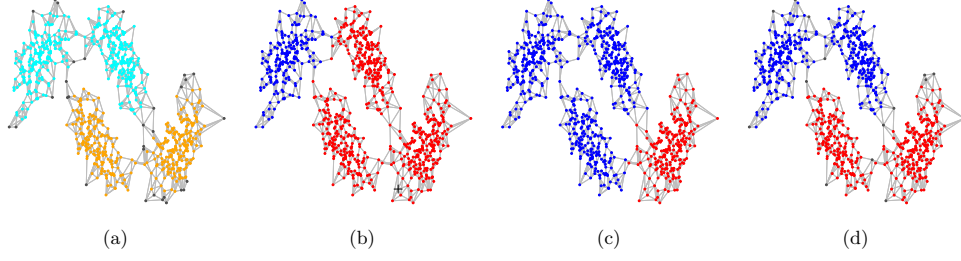


Figure 1

B Additional Experiments

C OLD STUFF

C.1 Volume estimates

We will fix $\mathcal{A} \subset \mathbb{R}^d$ to be an arbitrary set. To simplify expressions, for the σ -expansion \mathcal{A}_σ , we will write the set difference between \mathcal{A}_σ and the $(\sigma + r)$ -expansion $\mathcal{A}_{\sigma+r}$ as

$$\mathcal{A}_{\sigma, \sigma+r} := \{x : 0 < \text{dist}(x, \mathcal{A}_\sigma) \leq r\},$$

where as a reminder $\text{dist}(x, \mathcal{A}) = \min_{x' \in \mathcal{A}} \|x - x'\|$.

Lemma 38. We will repeatedly employ Lemma ?? and Lemma 3 in tandem. As a first example, in Lemma 38, we bound the ratio of $\nu(\mathcal{A})$ to $\nu(\mathcal{A}_{-\delta})$, where we write $\partial\mathcal{A}$ for the boundary of \mathcal{A} , and for $\delta > 0$ we let

$$\mathcal{A}_{-\delta} := \{x \in \mathcal{A} : \text{dist}(x, \partial\mathcal{A}) > \delta\}.$$

This will be useful when we bound $\text{vol}(\mathcal{C}_\sigma)$.

Lemma 38. For $\sigma, \mathcal{A}_\sigma$ as in Lemma ??, let $r > 0$ satisfy $r \leq \sigma/4d$. Then,

$$\frac{\nu(\mathcal{A}_\sigma)}{\nu(\mathcal{A}_{\sigma-r})} \leq 2.$$

Proof. Fix $q = \sigma - r$. Then,

$$\begin{aligned} \nu(\mathcal{A}_\sigma) &= \nu(\mathcal{A}_{q+\sigma-q}) = \nu(\mathcal{A}_q + (\sigma - q)B) \\ &\leq \nu(\mathcal{A}_q + \frac{(\sigma - q)}{q} \mathcal{A}_q) = \left(1 + \frac{\sigma - q}{q}\right)^d \nu(\mathcal{A}_q) \end{aligned}$$

where the inequality follows from Lemma ?. Of course, $\sigma - q = r$, and $\frac{r}{q} \leq \frac{1}{2d}$ for $r \leq \frac{1}{4d}$. The claim then follows from Lemma 3. \square

C.2 Proof of Lemma 34

Proof. We will write $\mathbf{W}_n = \mathbf{D}_n \mathbf{A}_n^{-1}$ for the transition probability matrix over $G_{n,r}$, and let $\widetilde{\mathbf{D}}_n$ and $\widetilde{\mathbf{W}}_n$ be the degree and random walk matrices for the subgraph $\widetilde{G}_{n,r}$.

We introduce *leakage* and *soakage* vectors, defined by

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

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

Roughly, the proof of Lemma 34 will unfold in four steps. The first two will result in the lower bound of (A.34), while the latter two will imply the upper bound in (A.35). We briefly summarize the approach before diving into the formal proof:

1. For $u \in \mathcal{C}[\mathbf{X}]$, we use the results of [7] to produce the lower bound

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

where

$$\tilde{\mathbf{p}}_\ell = \alpha \ell + (1 - \alpha) \tilde{\mathbf{p}}_\ell \tilde{\mathbf{W}}_n$$

is the PPR random walk over $\tilde{G}_{n,r}$, and ℓ has bounded norm $\|\ell\|_1 \leq 2 \frac{\Phi_{n,r}(\mathcal{C}_\sigma[\mathbf{X}])}{\alpha}$.

2. Since $r < \sigma$, for any $u \in \mathcal{C}[\mathbf{X}]$ there are no edges between u and $\mathbf{X} \setminus \mathcal{C}_\sigma[\mathbf{X}]$. Therefore, the page-rank vector $\tilde{\mathbf{p}}_\ell$ will not assign more than $\|\ell\|_1 / d_{\min}(\mathcal{C}_\sigma[\mathbf{X}])$ probability mass to any vertex in $\mathcal{C}'[\mathbf{X}]$. This observation will conclude our proof of (A.34).
3. For vertices $u' \in G_{n,r} / \mathcal{C}_\sigma[\mathbf{X}]$, we can upper bound $p_v(u) \leq p_s(u')$. In particular, this hold for all $u' \in \mathcal{C}'[\mathbf{X}]$.
4. Since $r < \sigma$, there are no edges between u' and $G / \mathcal{C}'[\mathbf{X}]$. Therefore, the page-rank vector p_s will assign no more than $\|s\|_1 / d_{\min}(\mathcal{C}_\sigma[\mathbf{X}])$ probability mass to any vertex in $\mathcal{C}'[\mathbf{X}]$. Additionally, s has bounded norm $\|s\|_1 \leq \|\ell\|_1$. This will conclude our proof of (A.35), and hence Lemma 34.

Step 1 We will begin by restating some results of [7].

For seed node v , we write

$$\tilde{\mathbf{p}}_v = \alpha e_v + (1 - \alpha) \tilde{\mathbf{p}}_v \tilde{\mathbf{W}}_n \tag{A.46}$$

$$= \alpha \sum_{t=0}^{\infty} (1 - \alpha)^t \left(e_v \tilde{\mathbf{W}}_n^t \right) \tag{A.47}$$

From Lemma 3.1 of [7] we have that there exists a good set $\mathcal{C}_\sigma[\mathbf{X}]^g \subseteq \mathcal{C}_\sigma[\mathbf{X}]$ with $\text{vol}(\mathcal{C}_\sigma[\mathbf{X}]^g) \geq \text{vol}(\mathcal{C}_\sigma[\mathbf{X}]) / 2$ for all $v \in \mathcal{C}_\sigma[\mathbf{X}]^g$, $u \in \mathcal{C}_\sigma[\mathbf{X}]$

$$\begin{aligned}p_u &\geq \tilde{\mathbf{p}}_v(u) - \tilde{\mathbf{p}}_\ell(u) \\ \|\ell\|_1 &\leq \frac{2\tilde{\Phi}_{n,r}}{\alpha}\end{aligned} \tag{A.48}$$

where $\tilde{\mathbf{p}}_v = (\tilde{\mathbf{p}}_v(u))$ and likewise for $\tilde{\mathbf{p}}_\ell = (\tilde{\mathbf{p}}_\ell(u))$. (This will be the only time we need to restrict ourselves to this 'good set'.)

Moreover if, as we have specified, $\alpha \leq \tilde{\Psi}_{n,r}/9$, Lemma 3.2 of [7] yields a lower bound on \tilde{p}

$$\tilde{\mathbf{p}}_v(u) \geq \frac{4}{5} \tilde{\pi}_{n,r}(u). \quad (\text{A.49})$$

Step 2 We turn to upper bounding $\tilde{\mathbf{p}}_\ell(u)$. For any $u \in \mathcal{C}[\mathbf{X}]$, we have

$$\begin{aligned} \tilde{\mathbf{p}}_\ell(u) &= \alpha \sum_{t=0}^{\infty} (1-\alpha)^t \left(\ell \tilde{\mathbf{W}}_n^t \right) (u) \\ &= \|\ell\|_1 \alpha \sum_{t=0}^{\infty} (1-\alpha)^t \left(\frac{\ell}{\|\ell\|_1} \tilde{\mathbf{W}}_n^t \right) (u) \\ &\stackrel{(i)}{=} \|\ell\|_1 \alpha \sum_{t=1}^{\infty} (1-\alpha)^t \left(\frac{\ell}{\|\ell\|_1} \tilde{\mathbf{W}}_n^t \right) (u) \\ &\stackrel{(ii)}{\leq} \|\ell\|_1 \frac{1}{\tilde{D}_{\min}} \end{aligned} \quad (\text{A.50})$$

where we use $\left(\frac{\ell}{\|\ell\|_1} \tilde{\mathbf{W}}_n^t \right) (u)$ to denote $\ell \tilde{\mathbf{W}}_n^t e_u$.

(i) follows from the fact that since $r < \sigma$, $\text{cut}(\mathcal{C}[\mathbf{X}], G_{n,r}/\mathcal{C}_\sigma[\mathbf{X}]; G_{n,r}) = 0$. Therefore $(\mathbf{D}_n^{-1})_{uu}(\tilde{\mathbf{D}}_n)_{uu} = 1$, and as a result

$$(\ell \tilde{\mathbf{W}}_n^0)(u) = \ell(u) = 0.$$

To see (ii), let $q = \frac{\ell}{\|\ell\|_1} \tilde{\mathbf{W}}_n^{t-1}$. Then

$$\begin{aligned} \left(\frac{\ell}{\|\ell\|_1} \tilde{\mathbf{W}}_n^t \right) (u) &= \left(q \tilde{\mathbf{W}}_n \right) (u) \\ &\leq \|q\|_1 \|\tilde{\mathbf{W}}_{\cdot u}\|_\infty \\ &\stackrel{(iii)}{\leq} \frac{1}{\tilde{D}_{\min}}. \end{aligned}$$

where $\tilde{\mathbf{W}}_{\cdot u}$ is the u th column of $\tilde{\mathbf{W}}_n$. (iii) then follows from the fact that any vertex in $\mathcal{C}[\mathbf{X}]$ is connected only to vertices in $\mathcal{C}_\sigma[\mathbf{X}]$, and therefore every entry of $\tilde{\mathbf{W}}_{\cdot u}$ is either 0 or at most $1/\tilde{D}_{\min}$.

Combined, (A.50), (A.49), and (A.48) imply

$$p_v(u) \geq \frac{4}{5} \tilde{\pi}_{n,r}(u) - 18 \frac{\Phi_{n,r}(\mathcal{C}_\sigma[\mathbf{X}])}{\tilde{D}_{\min} \alpha}.$$

for any $v \in \mathcal{C}_\sigma[\mathbf{X}]$.

Step 3 To get the corresponding upper bound on $p_v(u')$, we will use the soakage vectors s and s_t . We will first argue that s is a worse starting distribution – meaning it puts uniformly more mass outside the cluster – than simply starting at v .

Lemma 39. *For all $u' \notin \mathcal{C}_\sigma[\mathbf{X}]$,*

$$\mathbf{p}_v(u') \leq \mathbf{p}_s(u'). \quad (\text{A.51})$$

Proof. We have

$$\begin{aligned}\mathbf{p}_v(u') &= \alpha \sum_{T=0}^{\infty} (1-\alpha)^T (e_v \mathbf{W}_n^T)(u) \\ &\stackrel{(i)}{=} \alpha \sum_{T=1}^{\infty} (1-\alpha)^T (e_v \mathbf{W}_n^T)(u')\end{aligned}$$

where (i) follows from $v \in \mathcal{C}_\sigma$, $u \notin \mathcal{C}_\sigma$ and therefore $e_v(u) = 0$.

Lemma 36 allows us to make the transition to sums of soakage vectors.

Lemma 40. *Let $G = (V, E)$ be a graph, with associated random walk matrix W .*

For any $T \geq 1$, q vector, $S \subset V$, and $s_t = s_t(S^c, q)$

$$qW^T = \sum_{t=0}^{T-1} s_t W^{T-t-1} + q(WI_S)^T \quad (\text{A.52})$$

We prove Lemma 36 after completing the proof of Lemma 39.

Now, along with the fact $u \notin \mathcal{C}_\sigma$, we have

$$(e_v \mathbf{W}_n^T)(u') = \sum_{t=0}^{T-1} (s_t \mathbf{W}_n^{T-t-1})(u')$$

and so

$$\begin{aligned}\mathbf{p}_v(u) &= \alpha \sum_{T=1}^{\infty} (1-\alpha)^T \left(\sum_{t=0}^{T-1} s_t \mathbf{W}_n^{T-t-1} \right)(u') \\ &= \alpha \sum_{t=0}^{\infty} \sum_{T=t+1}^{\infty} (1-\alpha)^T (s_t \mathbf{W}_n^{T-t-1})(u') \\ &= \alpha \sum_{t=0}^{\infty} \sum_{\Delta=0}^{\infty} (1-\alpha)^{\Delta+t+1} (s_t \mathbf{W}_n^{\Delta})(u') \\ &\leq \alpha \sum_{t=0}^{\infty} \sum_{\Delta=0}^{\infty} (1-\alpha)^{\Delta+t} (s_t \mathbf{W}_n^{\Delta})(u') \\ &= \alpha \sum_{\Delta=0}^{\infty} (1-\alpha)^{\Delta} (s \mathbf{W}_n^{\Delta})(u') \\ &= \mathbf{p}_s(u')\end{aligned}$$

□

Proof of Lemma 36. Proceed by induction. When $T = 1$,

$$\begin{aligned}qW &= q(WI_S) + q(WI_{S^c}) \\ &= q(WI_S)^T + s_0\end{aligned}$$

Assume true for T_0 . For $T = T_0 + 1$,

$$\begin{aligned}
qW^T &= qW^{T_0}W \\
&= \left\{ \sum_{t=0}^{T_0-1} s_t W^{T_0-1-t} + q(WI_S)^{T_0} \right\} W \\
&= \sum_{t=0}^{T_0-1} s_t W^{T-1-t} + q(WI_S)^{T_0} (WI_S + WI_{S^c}) \\
&= \sum_{t=0}^{T-1} s_t W^{T-1-t} + q(WI_S)^T
\end{aligned}$$

□

Step 4 Just as we upper bounded the probability mass $\tilde{\mathbf{p}}_\ell$ could assign to any one vertex, we can upper bound

$$\begin{aligned}
\mathbf{p}_s(u') &= \alpha \sum_{t=0}^{\infty} (1-\alpha)^t (s \mathbf{W}_n^t)(u') \\
&= \|s\|_1 \alpha \sum_{t=0}^{\infty} (1-\alpha)^t \left(\frac{s}{\|s\|_1} \mathbf{W}_n^t \right)(u') \\
&= \|s\|_1 \alpha \sum_{t=1}^{\infty} (1-\alpha)^t \left(\frac{s}{\|s\|_1} \mathbf{W}_n^t \right)(u') \\
&\leq \|s\|_1 \frac{1}{\tilde{D}_{\min}}.
\end{aligned} \tag{A.53}$$

Finally, letting $q_t = e_v(\mathbf{W}_n \tilde{\mathbf{I}}_n)^t$ for ease of notation, we have

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

and as a result $\|s\|_1 = \|\ell\|_1$. Combining with $\|\ell\|_1 \leq 2 \frac{\tilde{\Phi}_{n,r}}{\alpha}$ and (A.53) yields the desired upper bound. □

Proof of Theorem 4. We note that by Theorems 1 and ??,

$$\kappa_2(\mathcal{C}) \geq \frac{\Phi_{n,r}(\mathcal{C}_\sigma[\mathbf{X}])}{\Psi_{n,r}(\mathcal{C}_\sigma[\mathbf{X}])}.$$

As a result Lemma 34 implies

$$\begin{aligned} p_u &\geq \frac{4}{5}\tilde{\pi}_{n,r}(u) - \frac{18\kappa_2(\mathcal{C})}{\tilde{D}_{\min}} & (u \in \mathcal{C}[\mathbf{X}]) \\ p_{u'} &\leq \frac{18\kappa_2(\mathcal{C})}{\tilde{D}_{\min}} & (u' \in \mathcal{C}'[\mathbf{X}]) \end{aligned} \quad (\text{A.54})$$

We turn to bounding $\tilde{\pi}_{n,r}(u)$. Clearly,

$$\begin{aligned} \tilde{\pi}_{n,r}(u) &\geq \frac{\tilde{D}_{\min}}{\widetilde{\text{vol}}_{n,r}(\mathcal{C}_\sigma[\mathbf{X}])} \\ &\geq \frac{\tilde{D}_{\min}}{\tilde{n}\tilde{D}_{\max}}. \end{aligned}$$

Application of Lemma 32 then yields

$$\tilde{\pi}_{n,r}(u) \geq 8 \frac{\lambda_\sigma}{\nu(\mathcal{C}_\sigma)\Lambda_\sigma^2} \quad (\text{A.55})$$

as well as

$$\frac{1}{\tilde{D}_{\min}} \leq \frac{2}{\nu_d r^d \lambda_\sigma} \quad (\text{A.56})$$

with probability tending to 1 as $n \rightarrow \infty$, for all $u \in \mathcal{C}[\mathbf{X}]$ (indeed, all $u \in \mathcal{C}_\sigma[\mathbf{X}]$.)

Combining (A.54), (A.55) and (A.56), along with the requirement on $\kappa_2(\mathcal{C})$ given by (??), we have

$$\begin{aligned} p_u &\geq 3/5 \frac{\lambda_\sigma}{\nu(\mathcal{C}_\sigma)\Lambda_\sigma^2} \\ p_{u'} &\leq 1/5 \frac{\lambda_\sigma}{\nu(\mathcal{C}_\sigma)\Lambda_\sigma^2} \end{aligned}$$

for any $u \in \mathcal{C}$, $u' \in \mathcal{C}'$. As a result, if $\pi_0 \in (2/5, 3/5) \cdot \frac{\lambda_\sigma}{\nu(\mathcal{C}_\sigma)\Lambda_\sigma^2}$, as $n \rightarrow \infty$ with probability tending to one any sweep cut of the form of (4), including the output set \hat{C} , will successfully recover \mathcal{C} in the sense of (6).

References

- [1] Martin Dyer, Alan Frieze, and Ravi Kannan. A random polynomial-time algorithm for approximating the volume of convex bodies. *Journal of the ACM (JACM)*, 38(1):1–17, 1991.
- [2] Ravi Kannan, Santosh Vempala, and Adrian Vetta. On clusterings: Good, bad and spectral. *J. ACM*, 51(3):497–515, May 2004.
- [3] László Lovász and Miklós Simonovits. The mixing rate of markov chains, an isoperimetric inequality, and computing the volume. In *Proceedings [1990] 31st annual symposium on foundations of computer science*, pages 346–354. IEEE, 1990.
- [4] Ravi Montenegro. *Faster mixing by isoperimetric inequalities*. PhD thesis, Yale University, 2002.
- [5] Ben Morris and Yuval Peres. Evolving sets, mixing and heat kernel bounds. *Probability Theory and Related Fields*, 133(2):245–266, 2005.
- [6] Nicolás García Trillos, Dejan Slepčev, James Von Brecht, Thomas Laurent, and Xavier Bresson. Consistency of cheeger and ratio graph cuts. *Journal of Machine Learning Research*, 17(1):6268–6313, jan 2016.
- [7] Zeyuan Allen Zhu, Silvio Lattanzi, and Vahab S Mirrokni. A local algorithm for finding well-connected clusters. In *ICML (3)*, pages 396–404, 2013.