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 (11), which we recall links the conductance function to mixing time; this will be necessary for both Theorems 2 and 3. A.5 and A.6 give the proof of Theorem 2, while A.7- A.10 give the proof of Theorem 3. A.11 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 $\epsilon > 0$,

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

which is sufficient to prove the claim.

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

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

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, $\mathcal{A}_\sigma$ is compact. Moreover,

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

so by compactness there exists $x_1, \ldots, x_N \in \mathcal{A}$ 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

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

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

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

Observe that $\bigcup_{i=1}^N \mathcal{A}_\sigma^{(i)} = \bigcup_{i=1}^N B(x_i, \sigma)$, 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 $\mathcal{A}_{\sigma+\epsilon+\delta}^{(1)} := B(x_1, \sigma + \delta + \epsilon)$ and

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

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. ☐

Lemma 2. *For $i = 1, \dots, N$ and $A_\sigma^{(i)}, A_{\sigma+\delta+\epsilon}^{(i)}$ 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

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

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 $x' \in \mathcal{A}_\sigma^{(i)}$. ☐

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

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

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

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

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

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

Let $\mathcal{C}_{\sigma,\sigma+r} := \{x : 0 < \mathrm{dist}(x, \mathcal{C}_\sigma) < r\}$, where $\mathcal{C}_\sigma$ 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 $\mathcal{C}_\sigma$ as follows: for $k \in \mathbb{N}$,

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

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

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

(A1) and (A2) imply the upper bound

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

and writing

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

we have

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

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

We first consider the term Σ_k. $\mathcal{C}$ has finite diameter by (A4). Letting $\overline{\mathcal{C}}$ be the closure of $\mathcal{C}$, we observe that $\overline{\mathcal{C}_\sigma} = \overline{\mathcal{C}} + \sigma B$, and moreover for any $\delta > 0$, $\nu(\overline{\mathcal{C}_\sigma} + \delta B) = \nu(\mathcal{C}_\sigma + \delta B)$ (as $\partial(\mathcal{C}_\sigma + \delta B)$ is Lipschitz and therefore has measure zero). As a result, for each $t_{i,k}, i = 1, \ldots, k$ we may apply Lemma 1 to $\overline{\mathcal{C}}$ and obtain

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

which in turn gives

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

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

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

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:

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

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. ☐

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

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

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

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

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

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

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

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

which by Lemma 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,

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

□

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

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

where

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

and

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

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

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

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

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

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

As a result,

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

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

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

Lemma 8.

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

A.2 Density-weighted cut and volume estimates

For notational ease, we write

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

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* $0 < r \le \sigma/2d$,

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

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

$$\mathrm{cut}_{n,r} = \sum_{i=1}^{n}\sum_{j \ne i}\mathbf{1}(x_i \notin \mathcal{C}_\sigma)\mathbf{1}(x_j \in \mathcal{C}_\sigma)\mathbf{1}(\|x_i - x_j\| \le r) \tag{A.10}$$

and by linearity of expectation, we obtain

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

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

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

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 $\mathrm{vol}_{n,r}$ as the sum of indicator functions,

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

and by linearity of expectation we obtain

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

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

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

to p_K and p_V.

In Lemma 11, we give probabilistic bounds on the $\mathrm{cut}_{n,r}$, $\mathrm{vol}_{n,r}$ and $\mathrm{vol}_{n,r}^c$ in terms of p_K and p_V. These bounds are a straightforward consequence of Lemma 26, Hoeffding's inequality for U-statistics.

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

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

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

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

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

with kernels

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

Similarly,

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

with kernel,

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

From Lemma 26 we therefore have

$$\frac{\mathrm{cut}_{n,r}}{\binom{n}{2}} \le p_K + \sqrt{\frac{\log(1/\delta)}{n}}, \quad \frac{\mathrm{vol}_{n,r}}{\binom{n}{2}} \ge p_V - \sqrt{\frac{\log(1/\delta)}{n}}, \quad \frac{\mathrm{vol}^c_{n,r}}{\binom{n}{2}} \ge p_V^c - \sqrt{\frac{\log(1/\delta)}{n}}$$

each with probability at least $1 - \delta$. The claim follows in light of (6), which implies $p_V^c \ge p_V$. □

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

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

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

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

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

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

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 (7),

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

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

A.4 Mixing time on graphs

For $N \in \mathbb{N}$ and a set V of N vertices, take $G = (V, E)$ to be an undirected and unweighted graph, with associated adjacency matrix $\mathbf{A}$, random walk matrix $\mathbf{W}$, and stationary distribution $\boldsymbol{\pi} = (\pi_u)_{u\in V}$ where $\pi_v = \frac{\mathbf{D}_{vv}}{\mathrm{vol}(V;G)}$. For $v \in V$,

$$q_{vu}^{(m)} = e_v \mathbf{W}^m e_u, \quad \mathbf{q}_v^{(m)} = \left(q_{vu}^{(m)}\right)_{u\in V}, \quad \mathbf{q}_v = (\mathbf{q}_{v\cdot}^{(1)}, \mathbf{q}_{v\cdot}^{(2)}, \ldots), \tag{A.14}$$

denote respectively the m-step transition probability, distribution, and sequence distributions of the random walk over G originating at v. Letting $\mathbf{q} = (\mathbf{q}_v)_{v\in V}$, the relative pointwise mixing time is thus

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

where we include the extra dependency on $\mathbf{q}$ as we will need to consider mixing of various random walks.

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

$$s(G) := \frac{9D_{\min}}{10}\pi_{\min}$$

for $D_{\min} := \min_{v\in V}\mathbf{D}_{vv}$ and $\pi_{\min} := D_{\min}/\mathrm{vol}(V;G)$.

where $\beta(S) := \inf_{v\in S}\mathbf{q}_v^{(1)}(S^c)$, and by convention we let $\mathbf{p}(S) = \sum_{u\in S} p_u$ for any distribution vector $\mathbf{p} = (p_u)_{u\in V}$ over V. We collect some necessary facts about the local spread in Lemma 12.

Lemma 12.
- *If $\boldsymbol{\pi}(S) \leq s(G)$, then for every $u \in S$, $\mathbf{q}_u^{(1)}(S^c) \geq 1/10$.*
- *For any $v, u \in V$, and $m \in N$ greater than 0, $q_{vu}^{(m)}/\pi_{\min} \leq 1/s(G)$.*

Proof. If $t = \boldsymbol{\pi}(S) \leq \frac{9D_{\min}}{10}\pi_{\min}$, divide both sides by $\pi_{\min}$ to obtain

$$|S| \leq \frac{9D_{\min}}{10}$$

which implies $\mathbf{q}_v^{(1)}(S^c) \geq 1/10$ for all $v \in S$. This implies the first statement.

The second statement follows from the fact $q_{vu}^{(m)} \leq 1/D_{\min}$ for any m. $\square$

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

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

where

$$\boldsymbol{\rho}_v^{(m)} = \frac{1}{m}\sum_{k=1}^{m+1} \mathbf{q}_v^m, \quad \boldsymbol{\rho}_v = \left(\boldsymbol{\rho}_v^{(1)}, \boldsymbol{\rho}_v^{(2)}, \boldsymbol{\rho}_v^{(3)} \ldots\right), \quad \boldsymbol{\rho} = (\boldsymbol{\rho}_v)_{v\in V} \tag{A.15}$$

and $\|\mathbf{p}-\boldsymbol{\pi}\|_{TV} = \sum_{v\in V}|p_v - \pi_v|$ is the total variation norm between distributions $\mathbf{p}$ and $\boldsymbol{\pi}$.

Lemma 13. *For* $\mathbf{q}$ *as in* (A.14) *and* $\boldsymbol{\rho}$ *as in* (A.15),

$$\tau_\infty(\mathbf{q}; G) \le 2752\tau_1(\boldsymbol{\rho}; G)\log\left(4\max\left\{1, \frac{1}{s(G)}\right\}\right)$$

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

$$\Delta_u^{(m)} = q_{vu}^{(m)} - \pi_u, \quad \delta_u^{(m)} = \frac{\Delta_u^{(m)}}{\pi_u}$$

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

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

To go between the $L^\infty(\boldsymbol{\pi})$ and $L^1(\boldsymbol{\pi})$ norms, we have

$$\begin{aligned}\left\|\boldsymbol{\delta}^{(2m)}\right\|_{L^\infty(\pi)} &\overset{(i)}{\le} \left\|\boldsymbol{\delta}^{(m)}\right\|^2_{L^2(\pi)} \\ &= \left\|(\boldsymbol{\delta}^{(m)})^2\right\|_{L^1(\pi)} \\ &\overset{(ii)}{\le} \left\|(\boldsymbol{\delta}^{(m)})\right\|_{L^1(\pi)}\left\|(\boldsymbol{\delta}^{(m)})\right\|_{L^\infty(\pi)}\end{aligned}$$

where (i) is a result of [2] and (ii) follows from Holder's inequality. Now, we upper bound the second factor on the right hand side by observing

$$\begin{aligned}\left\|(\boldsymbol{\delta}^{(m)})\right\|_{L^\infty(\pi)} &\le \max\left\{1, \max_{u\in V}\frac{q_{vu}^{(m)}}{\pi_u}\right\} \\ &\overset{(iii)}{\le} \max\left\{1, \frac{1}{s(G)}\right\}\end{aligned}$$

where (iii) follows from Lemma 12.

Now, we leverage the following well-known fact [4]: for any $\epsilon > 0$, if $m \ge \tau_1(\mathbf{q}_v^{(m)}; G)\cdot\log(1/\epsilon)$ then

$$\left\|\mathbf{q}_v^{(m)} - \boldsymbol{\pi}\right\|_{TV} \le \epsilon.$$

But $\left\|\mathbf{q}_v^{(m)} - \boldsymbol{\pi}\right\|_{TV}$ is exactly $\|(\boldsymbol{\delta}^{(m)})\|_{L^1(\pi)}$. Therefore, picking

$$m_0 = \tau_1(\mathbf{q}_v^{(m)}; G)\cdot\log\left(4\max\left\{1, \frac{1}{s(G)}\right\}\right)$$

implies $\|(\boldsymbol{\delta}^{(m)})\|_{L^\infty(\pi)} \le 1/4$ for all $m \ge 2m_0$. Then,

$$\left\|(\boldsymbol{\delta}^{(m)})\right\|_{L^\infty(\pi)} = \sup_u\left\{\frac{\left|q_{vu}^{(m)} - \boldsymbol{\pi}_u\right|}{\boldsymbol{\pi}_u}\right\}.$$

and since none of the above depended on a specific choice for v, the supremum can be taken over all starting vertices v as well. Thus $\tau_\infty(\mathbf{q}^{(m)}; G) \leq 2m_0$.

Finally, it is known [4] that

$$\tau_1(\mathbf{q}^{(m)}; G) \leq 1376\tau_1(\boldsymbol{\rho}^{(m)}; G)$$

and so the desired result holds. □

The second key quantity is the *conductance function*

$$\Phi(t; G) := \min_{\substack{S \subseteq V, \\ \boldsymbol{\pi}(S) \leq t}} \Phi(S; G) \qquad (\pi_{\min} \leq t < 1) \tag{A.16}$$

where $\Phi(S; G)$ is the normalized cut of S in G given by (1).

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

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

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

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 $\phi(t; G) : [t_0, 1/2] \to [0, 1]$ is a* blocking conductance function *if for all $S \subset V$ with $\boldsymbol{\pi}(S) = t \in [t_0, 1/2]$, either of the following hold:*

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

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

$$\phi_{int}(S) = \sup_{\lambda \leq \boldsymbol{\pi}(S)} \min_{\substack{B \subset S \\ \boldsymbol{\pi}(B) \leq \lambda}} \frac{\lambda \mathrm{cut}(S \setminus B, S^c; G)}{\mathrm{vol}(V; G)\left[\boldsymbol{\pi}(S)\boldsymbol{\pi}(S^c)\right]^2}$$

$$\phi_{ext}(S) = \sup_{\lambda \leq \boldsymbol{\pi}(S)} \min_{\substack{B \subset S^c \\ \boldsymbol{\pi}(B) \leq \lambda}} \frac{\lambda \mathrm{cut}(S \setminus B, S^c; G)}{\mathrm{vol}(V; G)\left[\boldsymbol{\pi}(S)\boldsymbol{\pi}(S^c)\right]^2}$$

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

$$h^m(t_0) = \sup_{S : \boldsymbol{\pi}(S) < t_0} \left(\boldsymbol{\rho}_v^{(m)}(S) - \boldsymbol{\pi}(S) \right)$$

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

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

[1] For more details, see [4]

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 $\phi_0(t, G) : [s(G), 1/2] \to [0, 1]$ defined by

$$\phi_0(t; G) = \begin{cases} \frac{1}{5}, & t = s'(G) \\ \frac{1}{4}\Phi^2(t; G), & t \in (s'(G), 1/2] \end{cases} \tag{A.17}$$

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:

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

Then, $h^1(s'(G)) \leq 1/10$ follows exactly from the proof of Lemma 12, except now $\boldsymbol{\pi}(S) \leq s'(G)$ 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 $\boldsymbol{\pi}(S) \leq s'(G)$, the following statement holds: for every $u \in S$, $\mathrm{cut}(u, S^c; G) \geq 9/10 \cdot \deg(u; G)$. Fixing $\lambda = \pi(S)/2$, we have

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

$\square$

A.5 Conductance function and local spread: non-convex case.

We begin with some notation. Write $\mathcal{C}_\sigma[\mathbf{X}] = \widetilde{\mathbf{X}}$, and $G_{n,r}[\mathcal{C}_\sigma[\mathbf{X}]]$ as $\widetilde{G}_{n,r}$. For $S \subset \widetilde{\mathbf{X}}$, let $\widetilde{\mathrm{cut}}_{n,r}(S) = \mathrm{cut}(S; \widetilde{G}_{n,r})$ and similarly $\widetilde{\mathrm{vol}}_{n,r}(S) = \mathrm{vol}(S; \widetilde{G}_{n,r})$.

Consider $\mathbf{z} \subset \mathcal{C}_\sigma$ such that $\mathcal{N}_{\mathbf{z}} = \{B(z, r/3) : z \in \mathbf{z}\}$ is an internal covering of $\mathcal{C}_\sigma$, meaning $\mathcal{N}_{\mathbf{z}} \supseteq \mathcal{C}_\sigma$. Then, we write

$$\begin{aligned}
\widetilde{B}_{\min} &= \min_{z \in \mathbf{z}} \left|B(z, r/3) \cap \widetilde{\mathbf{X}}\right|, \quad \widetilde{D}_{\min} = \min_{\widetilde{x} \in \widetilde{\mathbf{X}}} \left|\widetilde{\mathrm{cut}}_{n,r}(x)\right| \\
\widetilde{B}_{\max} &= \min_{z \in \mathbf{z}} \left|B(z, r/3) \cap \widetilde{\mathbf{X}}\right|, \quad \widetilde{D}_{\max} = \min_{\widetilde{x} \in \widetilde{\mathbf{X}}} \left|\widetilde{\mathrm{cut}}_{n,r}(x)\right|
\end{aligned}$$

Both the conductance function and local spread will depend heavily on these quantities. Lemma 16 collects the bounds we will need.

Lemma 16. *Let $\mathcal{C}_\sigma$ satisfy the conditions of Theorem 2. For sufficiently large n, and $r \le \sigma/4d$, each of the following bounds hold with probability $1-\delta$:*

$$\widetilde{B}_{\max} \le \left(1 + \sqrt{3^d \frac{3(\log|\mathcal{N}_{\mathbf{z}}| + \log(1/\delta))}{n\nu_d r^d \Lambda_\sigma}}\right) n\nu_d \left(\frac{r}{3}\right)^d \Lambda_\sigma$$
$$\widetilde{B}_{\min} \ge \left(1 - \sqrt{3^d \frac{2(\log|\mathcal{N}_{\mathbf{z}}| + \log(1/\delta))}{n\nu_d r^d \lambda_\sigma \beta_d}}\right) n\nu_d \left(\frac{r}{3}\right)^d \lambda_\sigma \beta_d$$
$$\widetilde{D}_{\max} \le \left(1 + \sqrt{\frac{3(\log n + \log(1/\delta))}{n\nu_d r^d \Lambda_\sigma}}\right) n\nu_d r^d \Lambda_\sigma$$
$$\widetilde{D}_{\min} \ge \left(1 - \sqrt{\frac{2(\log n + \log(1/\delta))}{n\nu_d r^d \lambda_\sigma \beta_d}}\right) n\nu_d \left(\frac{r}{3}\right)^d \lambda_\sigma \beta_d$$
$$\left(1 - \sqrt{\frac{2\log(1/\delta)}{n\lambda_d \nu_d \sigma^d}}\right) n\lambda_\sigma \nu_d \sigma^d \le \widetilde{n} \le \left(1 + \sqrt{\frac{3\log(1/\delta)}{n\Lambda_\sigma \nu_d D^d}}\right) n\Lambda_d \nu_d D^d \tag{A.18}$$

where $\widetilde{n} = \left|\widetilde{\mathbf{X}}\right|$ and $\beta_d \ge 1/2^d$.

In particular, fix $\epsilon > 0$. Then, for n as specified in Theorem 2, the following event:

$$\widetilde{B}_{\max} \le (1+\epsilon)\, n\nu_d \left(\frac{r}{3}\right)^d \Lambda_\sigma, \quad \widetilde{D}_{\max} \le (1+\epsilon)\, n\nu_d r^d \Lambda_\sigma$$
$$\widetilde{B}_{\min} \ge (1-\epsilon)\, n\nu_d \left(\frac{r}{3}\right)^d \lambda_\sigma \beta_d, \quad \widetilde{D}_{\min} \ge (1-\epsilon)\, n\nu_d r^d \lambda_\sigma \beta_d$$
$$(1-\epsilon) n\lambda_\sigma \nu_d \sigma^d \le \widetilde{n} \le (1+\epsilon) n\Lambda_\sigma \nu_d D^d \tag{A.19}$$

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

Proof. We observe that for any $s \le \sigma/4d$ and any $x \in \mathcal{C}_\sigma$,

$$\nu(B(x,s) \cap \mathcal{C}_\sigma) \ge \left(\frac{s}{2}\right)^d = s^d \beta_d.$$

(In fact, tighter bounds can be shown to hold, but we will not need them). Therefore, by (A1):

$$\lambda_\sigma \nu_d s^d \beta_d \le \mathbb{P}\left(B(z,s) \cap \mathcal{C}_\sigma\right) \le \Lambda_\sigma \nu_d s^d$$

In particular, this holds for $s = r$ and $s = r/3$, and for every $z \in \mathbf{z}$ as well as every $z \in \widetilde{\mathbf{X}}$. Now, by (A4) and (A1) we also have

$$\Lambda_\sigma \nu_d \sigma^d \le \mathbb{P}(\mathcal{C}_\sigma) \le \Lambda_\sigma \nu_d D^d$$

The proof of each statement in (A.18) then follows from application of Lemma 27.

To show (A.19), we note that $|\mathcal{N}_z|$ is less than the covering number of the D-ball in d dimensions. Therefore

$$|\mathcal{N}_z| \le \left(\frac{6D}{r} + 1\right)^d.$$

It is then immediately apparent that n chosen as in Theorem 2 yields (A.19). $\square$

Now, we consider the conductance function and local spread computed over $\widetilde{G}_{n,r}$, which we refer to by

$$\widetilde{\Phi}_{n,r}(t) = \Phi(t; \widetilde{G}_{n,r}), \quad \widetilde{s}_{n,r} = s(\widetilde{G}_{n,r}).$$

where the restriction in the minimization problem of (A.16) is with respect to $\widetilde{\pi}_{n,r}$ the stationary distribution over $\widetilde{G}_{n,r}$.

We will bound $\widetilde{\Phi}_{n,r}(1/2)$ and $\widetilde{s}_{n,r}$ under the event that (A.19) holds, noting that this occurs with probability at least $1 - \delta$.

Lemma 17. *If the bounds given by* (A.19) *hold, then*

$$\widetilde{s}_{n,r} \geq \frac{9}{10} \frac{(1+\epsilon)^2}{(1-\epsilon)^2} \frac{\lambda_\sigma^2}{\Lambda_\sigma^2} \frac{r^d}{D^d} \beta_d^2$$

Proof. The local spread can be written as

$$\widetilde{s}_{n,r} = \frac{9}{10} \frac{\widetilde{D}_{\min}^2}{\widetilde{\mathrm{vol}}_{n,r}(\widetilde{G}_{n,r})} \geq \frac{9}{10} \frac{\widetilde{D}_{\min}^2}{\widetilde{D}_{\max} \widetilde{n}}.$$

Then apply the relevant results of Lemma 16. $\square$

Lemma 18. *If the bounds given by* (A.19) *hold, then:*

$$\widetilde{\Phi}_{n,r}(1/2) \geq \frac{\lambda_\sigma (1-\epsilon) \beta_d}{4 \Lambda_\sigma (1+\epsilon) 3^d} \left(1 + \frac{(1-\epsilon) r^d \lambda_\sigma}{(1+\epsilon) D^d \Lambda_\sigma} \right)$$

Proof. Fix $S \subset \widetilde{\mathbf{X}}$ with $\widetilde{\pi}_{n,r}(S) \leq 1/2$. Partition $\mathcal{N}_{\mathbf{z}} = \mathcal{N}_{\mathbf{z}}^+ \cup \mathcal{N}_{\mathbf{z}}^-$, where

$$\begin{aligned}
\mathcal{N}_{\mathbf{z}}^- &= \left\{ B(z, r/3) : 2 \left| B(z, r/3) \cap S \right| \leq \left| B(z, r/3) \cap \widetilde{\mathbf{X}} \right| \right\} \\
\mathcal{N}_{\mathbf{z}}^+ &= \mathcal{N}_{\mathbf{z}} \setminus \mathcal{N}_{\mathbf{z}}^-
\end{aligned}$$

and correspondingly $S^- = \mathcal{N}_{\mathbf{z}}^- \cap S$, $S^+ = \mathcal{N}_{\mathbf{z}}^+ \cap S$, so

$$\frac{\widetilde{\mathrm{cut}}_{n,r}(S)}{\widetilde{\mathrm{vol}}_{n,r}(S)} = \frac{\widetilde{\mathrm{cut}}_{n,r}(S^-; \widetilde{G}_{n,r} \setminus S) + \widetilde{\mathrm{cut}}_{n,r}(S^+; \widetilde{G}_{n,r} \setminus S)}{\widetilde{\mathrm{vol}}_{n,r}(S^-) + \widetilde{\mathrm{vol}}_{n,r}(S^+)}.$$

It is immediately apparent that the following bounds hold for all $S \subset \widetilde{\mathbf{X}}$:

$$\begin{aligned}
\widetilde{\mathrm{cut}}_{n,r}(S^-; \widetilde{G}_{n,r} \setminus S) &\geq \frac{|S^-| \, \widetilde{B}_{\min}}{2} \\
\widetilde{\mathrm{vol}}_{n,r}(S^-) &\leq \left| S^- \right| \widetilde{D}_{\max} \\
\widetilde{\mathrm{vol}}_{n,r}(S^+) &\leq \widetilde{\mathrm{vol}}_{n,r}(\widetilde{G}_{n,r}) \mathbf{1}(\left| N_{\mathbf{z}}^+ \right| > 0)
\end{aligned}$$

If moreover $|N_{\mathbf{z}}^+| < |N_{\mathbf{z}}|$, then

$$\widetilde{\mathrm{cut}}_{n,r}(S^+; \widetilde{G}_{n,r} \setminus S) \geq \frac{\widetilde{B}_{\min}^2}{4} \mathbf{1}(\left| N_{\mathbf{z}}^+ \right| > 0)$$

follows from the fact that the graph $H_{n,r} = (\mathbf{z}, E_H)$, with $(z_i, z_j) \in E_H$ if $\|z_i - z_j\| \leq r/3$, is connected. As a result, if $|N_{\mathbf{z}}^+| < |N_{\mathbf{z}}|$ we have

$$\frac{\widetilde{\mathrm{cut}}_{n,r}(S)}{\widetilde{\mathrm{vol}}_{n,r}(S)} \geq \frac{\widetilde{B}_{\min}}{4 \widetilde{D}_{\max}} + \frac{\widetilde{B}_{\min}^2}{8 \widetilde{\mathrm{vol}}_{n,r}(\widetilde{G}_{n,r})} \tag{A.20}$$

using the inequality $2(A+B)/(C+D) \geq A/C + B/D$ for A, B, C, D non-negative.

If, on the other hand , $|\mathcal{N}_{\mathbf{z}}^{+}| = \mathcal{N}_{\mathbf{z}}$, then (A.20) holds with respect to S^c. Then, because $\widetilde{\pi}_{n,r}(S) \leq 1/2$,

$$\frac{\widetilde{\mathrm{cut}}_{n,r}(S^c)}{\widetilde{\mathrm{vol}}_{n,r}(S^c)} \leq \frac{\widetilde{\mathrm{cut}}_{n,r}(S)}{\widetilde{\mathrm{vol}}_{n,r}(S)}$$

and so we get the exact statement of (A.20). Noting, as in the proof of Lemma 17, that $\widetilde{\mathrm{vol}}_{n,r}(\widetilde{G}_{n,r}) \leq \widetilde{n} \cdot \widetilde{D}_{\max}$, the relevant results of Lemma 16 yield the desired inequality. $\square$

A.6 Proof of Theorem 2

Throughout this proof, we will condition on the events of Lemmas 17 and 18, namely

$$\widetilde{s}_{n,r} \geq \frac{9}{10}\frac{(1+\epsilon)^2}{(1-\epsilon)^2}\frac{\lambda_\sigma^2}{\Lambda_\sigma^2}\frac{r^d}{D^d}\beta_d^2$$

$$\widetilde{\Phi}_{n,r}(1/2) \geq \frac{\lambda_\sigma(1-\epsilon)\beta_d}{4\Lambda_\sigma(1+\epsilon)3^d}\left(1 + \frac{(1-\epsilon)r^d\lambda_\sigma}{(1+\epsilon)D^d\Lambda_\sigma}\right)$$

noting that for n as chosen in Theorem 3, this will occur with probability at least $1-\delta$ (by Lemma 16).

As a reminder, we write $\mathcal{C}_\sigma[\mathbf{X}] = \widetilde{\mathbf{X}}$, and $G_{n,r}[\mathcal{C}_\sigma[\mathbf{X}]]$ as $\widetilde{G}_{n,r}$. Fix arbitrary $v \in \widetilde{\mathbf{X}}$, and let

$$\widetilde{\mathbf{q}}_n^{(m)} = e_v \mathbf{W}_{\widetilde{\mathbf{X}}}^t, \quad \widetilde{\mathbf{q}}_n = (\widetilde{q}_n^{(1)}, \widetilde{\mathbf{q}}_n^{(2)}, \ldots)$$

Our goal is to upper bound $\tau_\infty(\widetilde{\mathbf{q}}_n; \widetilde{G}_{n,r})$. (As we will see, this bound will hold over all such starting vertices $v \in \widetilde{\mathbf{X}}$.)

By Lemma 12,

$$\begin{aligned}\tau_\infty(\widetilde{\mathbf{q}}_n; \widetilde{G}_{n,r}) &\leq 2752\tau_1(\widetilde{\mathbf{q}}_n; \widetilde{G}_{n,r}) \max\left\{2, \log\left(\frac{4}{\widetilde{s}_{n,r}}\right)\right\} \\ &\leq 2752\tau_1(\widetilde{\mathbf{q}}_n; \widetilde{G}_{n,r}) \max\left\{2, d\log\left(\frac{8D(1+\epsilon)^2\Lambda_\sigma^2}{r(1-\epsilon)^2\lambda_\sigma^2}\right)\right\}\end{aligned}$$

We now upper bound $\tau_1(\widetilde{\mathbf{q}}_n; \widetilde{G}_{n,r})$. From Lemma 14, we have that

$$\begin{aligned}\tau_1(\widetilde{\mathbf{q}}_n; \widetilde{G}_{n,r}) &\leq \frac{1400}{3}\left(5 + \int_{\widetilde{s}_{n,r}}^{1/2} \frac{4}{t\widetilde{\Phi}_{n,r}^2(t)} dt\right) \\ &\leq \frac{1400}{3}\left(5 + \int_{s_{\mathbb{P},r}}^{1/2} \frac{4}{t\widetilde{\Phi}_{n,r}^2(t)} dt\right)\end{aligned} \tag{A.21}$$

where $s_{\mathbb{P},r} = \frac{9}{10}\frac{(1+\epsilon)^2}{(1-\epsilon)^2}\frac{\lambda_\sigma^2}{\Lambda_\sigma^2}\frac{r^d}{D^d}\beta_d^2$.(Since r remains constant, for sufficiently large n the lower bound on $\widetilde{D}_{\min}$ of Lemma 16 will be at least 10, and therefore Lemma 15 holds.)

Now, we can upper bound the average conductance integral:

$$\begin{aligned}\int_{s_{\mathbb{P},r}}^{1/2} \frac{4}{t\widetilde{\Phi}_{n,r}^2(t)} dt &\leq \int_{s_{\mathbb{P},r}}^{1/2} \frac{4}{t\widetilde{\Phi}_{n,r}^2(1/2)} dt \\ &\leq 64\frac{9^d\Lambda_\sigma^2(1+\epsilon)^2\beta_d^2}{\lambda_\sigma^2(1-\epsilon)^2}\left(1 + \frac{(1-\epsilon)r^d\lambda_\sigma}{(1+\epsilon)D^d\Lambda_\sigma}\right)^{-2}\log s_{\mathbb{P},r}.\end{aligned}$$

Plugging this in to (A.21) gives the desired upper bound on $\tau_\infty(\widetilde{q}_n, \widetilde{G}_{n,r})$, which translates to the lower bound of (10).

A.7 Population-level conductance function: convex case.

When $\mathcal{C}$ is convex, we will make use of the theory developed in A.4 with respect to the conductance function $\Phi(t; G_{n,r}[\mathcal{C}_\sigma(\mathbf{X})])$. First, however, we introduce a population-level analogue to $\Phi(t; G_{n,r}[\mathcal{C}_\sigma(\mathbf{X})])$ over the set $\mathcal{C}_\sigma$, which we denote $\widetilde{\Phi}_{\mathbb{P},r}$. (In general, we will adopt the convention of using $\widetilde{f}$ to denote functionals computed with respect to $\mathcal{C}_\sigma$.)

For $\mathcal{S} \subset \mathbb{R}^d$

$$\nu_{\mathbb{P}}(\mathcal{S}) := \int_{\mathcal{S}} f(x) dx$$

is the weighted volume.

The r-ball walk over $\mathcal{C}_\sigma$ is a Markov chain. For $x \in \mathcal{C}_\sigma$ and $\mathcal{S}, \mathcal{S}' \subset \mathcal{C}_\sigma$, the transition probability is given by

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

The stationary distribution $\pi_{\mathbb{P},r}$ thus satisfies

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

for all $\mathcal{S} \in \mathcal{C}_\sigma$. A simple calculation yields

$$\ell_{\mathbb{P},r}(x) := \nu_{\mathbb{P}}(\mathcal{C}_\sigma \cap B(x,r)) \quad \pi_{\mathbb{P},r}(\mathcal{S}) := \frac{1}{\int_{\mathcal{C}_\sigma} f(x) \ell_{\mathbb{P},r}(x) dx} \int_{\mathcal{S}} f(x) \ell_{\mathbb{P},r}(x) dx,$$

and therefore the ergodic flow is

$$\begin{aligned}
\widetilde{Q}_{\mathbb{P},r}(\mathcal{S}, \mathcal{S}') &:= \int_{\mathcal{S}} d\pi_{\mathbb{P},r}(x) P_{\mathbb{P},r}(x; \mathcal{S}') dx \\
&= \frac{1}{\int_{\mathcal{C}_\sigma} f(x) \ell_{\mathbb{P},r}(x) dx} \int_{\mathcal{S}} f(x) \left(\int_{\mathcal{S}' \cap B(x,r)} f(x') dx' \right) dx
\end{aligned}$$

The continuous conductance function is then

$$\begin{aligned}
\widetilde{\Phi}_{\mathbb{P},r}(t) &:= \min_{\substack{\mathcal{S} \subset \mathcal{C}_\sigma, \\ \pi_{\mathbb{P},r}(\mathcal{S}) \leq t}} \frac{\widetilde{Q}_{\mathbb{P},r}(\mathcal{S}, \mathcal{C}_\sigma \setminus \mathcal{S})}{\pi_{\mathbb{P},r}(\mathcal{S})} \\
&= \min_{\substack{\mathcal{S} \subset \mathcal{C}_\sigma, \\ \pi_{\mathbb{P},r}(\mathcal{S}) \leq t}} \frac{\int_{\mathcal{S}} f(x) \left(\int_{(\mathcal{C}_\sigma \setminus \mathcal{S}) \cap B(x,r)} f(x') dx' \right) dx}{\int_{\mathcal{S}} f(x) \left(\int_{\mathcal{C}_\sigma \cap B(x,r)} f(x') dx' \right) dx}.
\end{aligned}$$

For $m > 0$ and $0 < t_0 < t_1 < \ldots < t_m < 1$, denote the *stepwise approximation to g* by $\bar{g}$, defined as

$$\bar{g}(t) = g(t_i), \quad \text{for } t \in [t_{i-1}, t_i] \tag{A.22}$$

The stepwise approximation will be important to showing the consistency results of Section A.9 can be translated to a uniform bound. Lemma 19 shows that the approximation will not overly degrade our estimates of the population-level conductance function.

Lemma 19. • *For any function f monotone decreasing in t on the interval $[t_0, t_m]$, $\bar{f}(t) \leq f(t)$ for all $t \in [t_0, t_m]$.*

- *Fix*

$$g(t) = \log\left(\frac{1}{t}\right) \text{ for } x \in [t_0, 1/2]$$

If for all i in $1, \ldots, m$, $(t_i - t_{i-1}) \leq t_0/2$, then $\bar{g}(t) \geq g(t)/2$.

Proof. The first statement is immediately obvious, and we turn to proving the second.

The upper bound $g(t) \geq \bar{g}(t)$ follows immediately from the fact that $g(t)$ is a decreasing function along with the first statement.

By the concavity of the log function,

$$\bar{g}(t) = \log\left(\frac{1}{t_i}\right) \geq \log\left(\frac{1}{t}\right) - \frac{(t_i - t)}{t}.$$

As a result,

$$\bar{g}(t) - \frac{g(t)}{2} \geq \frac{\log\left(\frac{1}{t}\right)}{2} - \frac{(t_i - t)}{t} \geq 1/2 - 1/2 = 0.$$

□

The following theorem is found in [3]. It gives a bound population-level conductance function over convex bodies, when the density is uniform.

Theorem 2 (Restatement of [3] Theorem 4.6)**.** *Let $K \subset \mathbb{R}^d$ be a convex body of diameter D. Then for any $\mathcal{S} \subset K$ with $\pi_{\nu,r}(\mathcal{S}) \leq 1/2$,*

$$\frac{Q_{\nu,r}(\mathcal{S}, \mathcal{C}_\sigma \setminus \mathcal{S})}{\pi_{\nu,r}(\mathcal{S})} > \min\left\{\frac{1}{288\sqrt{d}}, \frac{r}{81\sqrt{d}D}\log\left(1 + \frac{1}{\pi_{\nu,r}(\mathcal{S})}\right)\right\}. \tag{A.23}$$

Lemma 20. *Under the conditions on $\mathcal{C}_\sigma$ given by Theorem 3, the following bounds hold:*

- *for $0 < t < 1/2$,*

$$\widetilde{\Phi}_{\mathbb{P},r}(t) > \min\left\{\frac{1}{288\sqrt{d}}, \frac{r}{81\sqrt{d}D}\log\left(1 + \frac{\lambda_\sigma^2}{\Lambda_\sigma^2 t}\right)\right\} \cdot \frac{\lambda_\sigma^4}{\Lambda_\sigma^4}$$

- *Let*

$$M = \frac{2^{d+1} D^d \Lambda_\sigma^2}{r^d \lambda_\sigma^2}$$

and $t_i = (i+1)/M$ for $i = 0, \ldots, m-1$. Then, for $1/M < t < 1/2$

$$\overline{\Phi}_{\mathbb{P},r}(t) > \min\left\{\frac{1}{288\sqrt{d}}, \frac{r}{162\sqrt{d}D}\mathrm{Log}\left(\frac{\Lambda_\sigma^2}{\lambda_\sigma^2 t}\right)\right\} \cdot \frac{\lambda_\sigma^4}{\Lambda_\sigma^4}$$

where $\overline{\Phi}_{\mathbb{P},r}(t)$ is defined as in (A.22) with respect to $t_0, \ldots t_{M-1}$, and $\mathrm{Log}(A/t) = \max\{\log(1 + 2A), \log(A/t)\}$.

Before we prove Lemma 20, note that the choice of M is made to ensure t_0 is greater than the local spread of $G_{n,r}\left[\mathcal{C}_\sigma[\mathbf{X}]\right]$, as we will see in Section A.8.

Proof of Lemma 20. We note that

$$\pi_{\mathbb{P},r}(S) \leq \pi_{\nu,r}(S) \cdot \frac{\Lambda_\sigma^2}{\lambda_\sigma^2}, \quad Q_{\mathbb{P},r}(\mathcal{S}, \mathcal{C}_\sigma \setminus \mathcal{S}) \geq Q_{\nu,r}(\mathcal{S}, \mathcal{C}_\sigma \setminus \mathcal{S}) \cdot \frac{\lambda_\sigma^2}{\Lambda_\sigma^2}$$

Plugging these estimates in to (A.23) gives

$$\frac{Q_{\mathbb{P},r}(\mathcal{S}, \mathcal{C}_\sigma \setminus \mathcal{S})}{\pi_{\mathbb{P},r}(\mathcal{S})} > \min\left\{\frac{1}{288\sqrt{d}}, \frac{r}{81\sqrt{d}D} \log\left(1 + \frac{\lambda_\sigma^2}{\Lambda_\sigma^2 \pi_{\mathbb{P},r}(\mathcal{S})}\right)\right\} \cdot \frac{\lambda_\sigma^4}{\Lambda_\sigma^4}$$

and since the right hand side is decreasing in $\pi_{\mathbb{P},r}(\mathcal{S})$, the desired lower bound holds on $\widetilde{\Phi}_{\mathbb{P},r}(t)$. The bound on $\overline{\Phi}_{\mathbb{P},r}(t)$ then follows from $\operatorname{Log}(A/t) \leq \log(1 + 1/t)$ for all $0 < t < 1/2$ and application of Lemma 19. ☐

A.8 Consistency of local spread and conductance function: convex case.

The introduction of the stepwise approximation allows us to make use of Lemma 21, which gives us (pointwise) consistency of the discrete graph functionals $\widetilde{\Phi}_{n,r}(t)$ to the continuous functionals $\widetilde{\Phi}_{\mathbb{P},r}(t)$.

We use $\omega_r(1)$ to denote a term which goes to infinity as $r \to 0$, and likewise $o_r(1)$ to denote a term which goes to 0 as $r \to 0$.

Lemma 21. *Fix $0 < t < 1/2$. Under the conditions on $\mathcal{C}_\sigma$ given by Theorem 3, the following statement holds: with probability one, as $n \to \infty$,*

$$\liminf_{n\to\infty} \widetilde{\Phi}_{n,r}(t) \geq \min\left\{\widetilde{\Phi}_{\mathbb{P},r}(t), c_d r \omega_r(1)\right\} \tag{A.24}$$

where c_d is a constant which may depend on the dimension d (as well as the distribution $\mathbb{P}$), but not r.

As a consequence, for M and $(t_i)_{i=0}^{M-1}$ defined as in Lemma 20, we have that

$$\liminf_{n\to\infty} \overline{\Phi}_{n,r}(t) \geq \min\left\{\overline{\Phi}_{\mathbb{P},r}(t), c_d r \omega_r(1)\right\} \tag{A.25}$$

We defer the proof of pointwise consistency to Section A.9. For now, we show that (A.25) is immediately implied by (A.24).

Proof of (A.25). We take as given that for any $0 < t < 1/2$,

$$\liminf_{n\to\infty} \widetilde{\Phi}_{n,r}(t) \geq \widetilde{\Phi}_{\mathbb{P},r}(t).$$

In particular, for sufficiently large n this will occur for each of $t_0, t_1, \ldots, t_m$ and therefore

$$\liminf_{n\to\infty} \overline{\Phi}_{n,r} \geq \overline{\Phi}_{\mathbb{P},r}$$

uniformly over $[1/m, 1/2]$. ☐

A.9 Pointwise consistency of conductance function: convex case.

We will rely heavily on results of [5], which prove the same result but consider only a pointwise result on $\widetilde{\Phi}_{n,r}(1/2)$ rather than over the entire conductance function.

Let $\widetilde{\mathbf{X}} = \mathcal{C}_\sigma[\mathbf{X}] = \{\widetilde{x}_1, \ldots, \widetilde{x}_{\widetilde{n}}\}$, and $\widetilde{n} = \left|\widetilde{\mathbf{X}}\right|$. Then

$$\widetilde{\mathbb{P}}_n := \frac{1}{\widetilde{n}} \sum_{\widetilde{x}_i \in \widetilde{\mathbf{X}}} \delta_{\widetilde{x}_i}$$

is the empirical distribution of $\widetilde{\mathbf{X}}$. Likewise, for $\mathcal{S} \subset \mathcal{C}_\sigma$ let $\widetilde{\mathbb{P}}$ be the conditional distribution $\mathbb{P}(x \in S | x \in \mathcal{C}_\sigma)$, given by

$$\widetilde{\mathbb{P}}(\mathcal{S}) := \frac{\nu_{\mathbb{P}}(\mathcal{S})}{\nu_{\mathbb{P}}(\mathcal{C}_\sigma)}.$$

A Borel map $T : \mathcal{C}_\sigma \to \widetilde{\mathbf{X}}$ is a *transportation map* between $\widetilde{\mathbb{P}}$ and $\widetilde{\mathbb{P}}_n$ if

$$\widetilde{\mathbb{P}}(\mathcal{S}) = \widetilde{\mathbb{P}}_n(T(\mathcal{S}))$$

for all $\mathcal{S} \in \mathcal{C}_\sigma$.

Lemma 22 (Proposition 5 of [5])**.** *There exists a sequence of transportation maps $(T_{\widetilde{n}})$ from $\widetilde{\mathbb{P}}$ and $\widetilde{\mathbb{P}}_n$ such that*

$$\limsup_{\widetilde{n} \to \infty} \frac{\widetilde{n}^{1/d} \left\| \mathrm{Id} - T_{\widetilde{n}} \right\|_{L^\infty(\widetilde{\mathbb{P}})}}{(\log \widetilde{n})^{p_d}} \leq C$$

where $p_d = 1/d$ for $d \geq 3$ and $3/4$ if $d = 2$.

These are referred to stagnating transportation maps. We refer the curious reader to [5] for more details.

For $S \subset \widetilde{\mathbf{X}}$, we will denote $\mathrm{vol}(S; \widetilde{G}_{n,r})$ by $\widetilde{\mathrm{vol}}_{n,r}(S)$, and likewise $\mathrm{cut}(S; \widetilde{G}_{n,r})$ by $\widetilde{\mathrm{cut}}_{n,r}(S)$.

Consider a sequence of sets $(S_{\widetilde{n}})_{\widetilde{n} \in \mathbb{N}}$, with $u_{\widetilde{n}} = \mathbf{1}_{S_{\widetilde{n}}}$ the characteristic function of $S_{\widetilde{n}}$. Similarly, for $\mathcal{S} \subset \mathcal{C}_\sigma$ let $u = \mathbf{1}_{\mathcal{S}}$.

Definition 2. *For a sequence $(u_{\widetilde{n}}) \in L^1(\widetilde{\mathbb{P}}_{\widetilde{n}})$ and $u \in L^1(\widetilde{\mathbb{P}})$, we say $(u_{\widetilde{n}})$ converges TL^1 to u if there exists a sequence of stagnating transportation maps $(T_{\widetilde{n}})$ such that*

$$\int_{\mathcal{C}_\sigma} |u(x) - (u_{\widetilde{n}}) \circ T_{\widetilde{n}}(x)| \, d\widetilde{\mathbb{P}}(x) \to 0$$

and denote it $u_{\widetilde{n}} \overset{TL^1}{\to} u$.

Lemma 23. *If $(u_{\widetilde{n}}) \overset{TL^1}{\to} u$, with probability one*

$$\lim_{n \to \infty} \frac{\widetilde{\mathrm{cut}}_{n,r}(S_{\widetilde{n}})}{\widetilde{n}^2} = \int_{\mathcal{S}} \widetilde{f}(x) \left(\int_{(\mathcal{C}_\sigma \setminus \mathcal{S}) \cap B(x,r)} \widetilde{f}(x') dx' \right) dx$$

where $\widetilde{f}$ is the density function of $\widetilde{\mathbb{P}}$ over $\mathcal{C}_\sigma$.

Proof. We note immediately that $n \to \infty$ implies $\widetilde{n} \to \infty$ with probability one.

Now, we can write

$$\begin{aligned}
\frac{\widetilde{\mathrm{cut}}_{n,r}(S_{\widetilde{n}})}{\widetilde{n}^2} &= \frac{1}{\widetilde{n}^2} \sum_{i=1}^{\widetilde{n}} \sum_{j=1}^{\widetilde{n}} u_{\widetilde{n}}(\widetilde{x}_i) \big(1 - u_{\widetilde{n}}(\widetilde{x}_j) \mathbf{1}(\|\widetilde{x}_i - \widetilde{x}_j\| \leq r)\big) \\
&= \int_{\mathcal{C}_\sigma} \left(\int_{\mathcal{C}_\sigma \cap B(x,r)} u_{\widetilde{n}}(x) \big(1 - u_{\widetilde{n}}(x')\big) d\widetilde{\mathbb{P}}_n(x') \right) d\widetilde{\mathbb{P}}_n(x) \\
&= \int_{\mathcal{C}_\sigma} (u_{\widetilde{n}} \circ T_{\widetilde{n}}(x)) \left(\int_{\mathcal{C}_\sigma \cap B(T_{\widetilde{n}}(x),r)} (1 - u_{\widetilde{n}} \circ T_{\widetilde{n}}(x')) \, d\widetilde{\mathbb{P}}(x') \right) d\widetilde{\mathbb{P}}(x).
\end{aligned}$$

Note that, for any $x \in \mathcal{C}_\sigma$,

$$\lim_{n\to\infty} \nu\big(B(T_{\widetilde{n}}(x), r) \setminus B(x,r)\big) = 0$$
$$\lim_{n\to\infty} \nu\big(B(x,r) \setminus B(T_{\widetilde{n}}(x), r)\big) = 0.$$

and therefore

$$\lim_{n\to\infty} \int_{\mathcal{C}_\sigma \cap B(T_{\widetilde{n}}(x),r)} (1 - u_{\widetilde{n}} \circ T_{\widetilde{n}}(x'))\, d\widetilde{\mathbb{P}}(x') = \lim_{n\to\infty} \int_{\mathcal{C}_\sigma \cap B(x,r)} (1 - u_{\widetilde{n}} \circ T_{\widetilde{n}}(x'))\, d\widetilde{\mathbb{P}}(x').$$

An application of the bounded convergence theorem yields

$$\lim_{n\to\infty} \int_{\mathcal{C}_\sigma} (u_{\widetilde{n}} \circ T_{\widetilde{n}}(x)) \left(\int_{\mathcal{C}_\sigma \cap B(T_{\widetilde{n}}(x),r)} (1 - u_{\widetilde{n}} \circ T_{\widetilde{n}}(x'))\, d\widetilde{\mathbb{P}}(x') \right) d\widetilde{\mathbb{P}}(x) =$$
$$\lim_{n\to\infty} \int_{\mathcal{C}_\sigma} (u_{\widetilde{n}} \circ T_{\widetilde{n}}(x)) \left(\int_{\mathcal{C}_\sigma \cap B(x,r)} (1 - u_{\widetilde{n}} \circ T_{\widetilde{n}}(x'))\, d\widetilde{\mathbb{P}}(x') \right) d\widetilde{\mathbb{P}}(x).$$

Letting

$$\mathcal{I}_n^1 = \int_{\mathcal{C}_\sigma} (u(x)) \left(\int_{\mathcal{C}_\sigma \cap B(x,r)} (u(x') - u_{\widetilde{n}} \circ T_{\widetilde{n}}(x'))\, d\widetilde{\mathbb{P}}(x') \right) d\widetilde{\mathbb{P}}(x)$$
$$\mathcal{I}_n^2 = \int_{\mathcal{C}_\sigma} (u_{\widetilde{n}} \circ T_{\widetilde{n}}(x) - u(x)) \left(\int_{\mathcal{C}_\sigma \cap B(x,r)} (1 - u_{\widetilde{n}} \circ T_{\widetilde{n}}(x'))\, d\widetilde{\mathbb{P}}(x') \right) d\widetilde{\mathbb{P}}(x)$$

we have

$$\int_{\mathcal{C}_\sigma} (u_{\widetilde{n}} \circ T_{\widetilde{n}}(x)) \left(\int_{\mathcal{C}_\sigma \cap B(x,r)} (1 - u_{\widetilde{n}} \circ T_{\widetilde{n}}(x'))\, d\widetilde{\mathbb{P}}(x') \right) d\widetilde{\mathbb{P}}(x) =$$
$$\int_{\mathcal{C}_\sigma} u(x) \left\{ \int_{\mathcal{C}_\sigma \cap B(x,r)} (1 - u(x')) d\widetilde{\mathbb{P}}(x') \right\} d\widetilde{\mathbb{P}}(x) + \mathcal{I}_n^1 + \mathcal{I}_n^2. \tag{A.26}$$

Recalling that $u = 1_{\mathcal{S}}$, we can see

$$\int_{\mathcal{C}_\sigma} u(x) \left\{ \int_{\mathcal{C}_\sigma \cap B(x,r)} (1 - u(x')) d\widetilde{\mathbb{P}}(x') \right\} d\widetilde{\mathbb{P}}(x) = \int_{\mathcal{S}} \widetilde{f}(x) \left(\int_{(\mathcal{C}_\sigma \setminus \mathcal{S}) \cap B(x,r)} \widetilde{f}(x') dx' \right) dx \tag{A.27}$$

Since $(u_{\widetilde{n}}) \overset{TL^1}{\to} u$, another application of the bounded convergence theorem yields $\lim_{n\to\infty} \mathcal{I}_n^1 = \lim_{n\to\infty} \mathcal{I}_n^2 = 0$. Therefore by (A.26) and (A.27) the final result holds. □

Lemma 24. *If $u_{\widetilde{n}} \overset{TL^1}{\to} u$, then*

$$\lim_{n\to\infty} \frac{\widetilde{\mathrm{vol}}_{n,r}(S_{\widetilde{n}})}{\widetilde{n}^2} = \int_{\mathcal{S}} \widetilde{f}(x) \left(\int_{\mathcal{C}_\sigma \cap B(x,r)} \widetilde{f}(x') dx' \right) dx$$

with probability one.

Proof. We note that

$$\frac{\widetilde{\mathrm{vol}}_{n,r}(S_{\widetilde{n}})}{\widetilde{n}^2} = \frac{1}{\widetilde{n}^2}\sum_{i=1}^{\widetilde{n}}\sum_{j=1}^{\widetilde{n}} u_{\widetilde{n}}(\widetilde{x}_i)\mathbf{1}(\|\widetilde{x}_i - \widetilde{x}_j\| \leq r) - \frac{1}{\widetilde{n}}$$
$$= \int_{\mathcal{C}_\sigma}\int_{\mathcal{C}_\sigma \cap B(x,r)} \left(u_{\widetilde{n}}(x)d\widetilde{\mathbb{P}}_n(x')\right) d\widetilde{\mathbb{P}}(x) - \frac{1}{\widetilde{n}}$$

Of course, $\lim_{n\to\infty}\frac{1}{\widetilde{n}} = 0$, and so

$$\lim_{n\to\infty}\frac{\widetilde{\mathrm{vol}}_{n,r}(S_{\widetilde{n}})}{\widetilde{n}^2} = \lim_{n\to\infty}\int_{\mathcal{C}_\sigma}\left(\mathcal{C}_\sigma \cap B(x,r)u_{\widetilde{n}}(x)d\widetilde{\mathbb{P}}_n(x')\right)d_{\widetilde{\mathbb{P}}}(x)$$

The proof then proceeds analogously to Lemma 23. ☐

Lemma 25 can be found in [1] (Theorem 3.1) or [5] (Lemma 23).

Lemma 25. *If* $u_{\widetilde{n}} \overset{TL^1}{\not\to} u$ *for some* $u \in L^1\nu$*, with probability one:*

$$\liminf_{n\to\infty}\frac{\widetilde{\mathrm{cut}}_{n,r}(\mathcal{S}_{\widetilde{n}})}{\widetilde{n}^2} \geq c_d r^{d+1}\omega_r(1)$$

where c_d *is a constant which does not depend on* r *but may depend on* $\mathcal{C}_\sigma$ *and* f*, and* $\omega_r(1) \to \infty$ *as* $r \to 0$*.*

Proof of (A.24). Let $(S^\star_{\widetilde{n}})$

$$\frac{\widetilde{\mathrm{cut}}_{n,r}(S^\star_{\widetilde{n}})}{\widetilde{\mathrm{vol}}_{n,r}(S^\star_{\widetilde{n}})} = \widetilde{\Phi}_{n,r}(t), \quad \widetilde{\pi}_{n,r}(S_{\widetilde{n}}) \leq t$$

be the sequence of minimizers of the normalized cut with stationary distribution at most t in the graph $\widetilde{G}_{n,r}$. Denote $u^\star_{\widetilde{n}} = 1_{S^\star_{\widetilde{n}}}$, and assume that $u^\star_{\widetilde{n}} \overset{TL^1}{\to} u$, for $u = 1_{\mathcal{S}}$, $\mathcal{S} \subset \mathcal{C}_\sigma$. Then, by Lemmas 23 and 24,

$$\lim_{n\to\infty}\frac{\widetilde{\mathrm{cut}}_{n,r}(S_{\widetilde{n}})}{\widetilde{\mathrm{vol}}_{n,r}(S_{\widetilde{n}})} = \frac{\int_{\mathcal{S}}\widetilde{f}(x)\left(\int_{(\mathcal{C}_\sigma\setminus\mathcal{S})\cap B(x,r)}\widetilde{f}(x')dx'\right)dx}{\int_{\mathcal{S}}\widetilde{f}(x)\left(\int_{\mathcal{C}_\sigma\cap B(x,r)}\widetilde{f}(x')dx'\right)dx}$$
$$\overset{(i)}{=} \frac{\int_{\mathcal{S}}f(x)\left(\int_{(\mathcal{C}_\sigma\setminus\mathcal{S})\cap B(x,r)}f(x')dx'\right)dx}{\int_{\mathcal{S}}f(x)\left(\int_{\mathcal{C}_\sigma\cap B(x,r)}f(x')dx'\right)dx} = \frac{\widetilde{Q}_{\mathbb{P},r}(\mathcal{S},\mathcal{C}_\sigma\setminus\mathcal{S})}{\pi_{\mathbb{P},r}(\mathcal{S})}.$$

where (i) holds because the normalization factors present in $\widetilde{f}$ cancel. From Lemma 24, we also have that $\lim_{n\to\infty}\widetilde{\pi}_{n,r}(S_{\widetilde{n}}) = \pi_{\mathbb{P},r}(\mathcal{S})$, and therefore $\pi_{\mathbb{P},r}(\mathcal{S}) \leq t$. As a result,

$$\liminf_{n\to\infty}\widetilde{\Phi}_{n,r}(t) = \frac{\widetilde{Q}_{\mathbb{P},r}(\mathcal{S},\mathcal{C}_\sigma\setminus\mathcal{S})}{\pi_{\mathbb{P},r}(\mathcal{S})} \geq \widetilde{\Phi}_{\mathbb{P},r}(t).$$

On the other hand, if $u^\star_{\widetilde{n}}$ does not converge TL^1, then

$$\frac{\widetilde{\mathrm{cut}}_{n,r}(S_{\widetilde{n}})}{\widetilde{n}^2} \geq c_d r^{d+1}\omega_r(1)$$

Additionally,

$$\frac{\widetilde{\mathrm{vol}}_{n,r}(S_{\widetilde{n}})}{\widetilde{n}^2} \leq \frac{\widetilde{\mathrm{vol}}_{n,r}(\widetilde{G}_{n,r})}{\widetilde{n}^2}$$

and

$$\limsup_{n\to\infty} \frac{\widetilde{\mathrm{vol}}_{n,r}(\widetilde{G}_{n,r})}{\widetilde{n}^2} \leq \nu_d r^d \Lambda_\sigma$$

As a result,

$$\liminf_{n\to\infty} \frac{\widetilde{\mathrm{cut}}_{n,r}(S_{\widetilde{n}})}{\widetilde{\mathrm{vol}}_{n,r}(S_{\widetilde{n}})} \geq c_d r \omega_r(1).$$

□

A.10 Proof of Theorem 3

Throughout this proof, we will refer to the subgraph $G_{n,r}\left[\mathcal{C}_\sigma[\mathbf{X}]\right]$ as $\widetilde{G}_{n,r}$.

Fix arbitrary $v = x_i \in \mathcal{C}_\sigma[\mathbf{X}]$, and let

$$\widetilde{\mathbf{q}}_n^{(m)} = e_v \mathbf{W}^t_{\mathcal{C}_\sigma[\mathbf{X}]}, \quad \widetilde{\mathbf{q}}_n = (\widetilde{q}_n^{(1)}, \widetilde{q}_n^{(2)}, \ldots)$$

Our goal is to upper bound $\tau_\infty(\widetilde{\mathbf{q}}_n; \widetilde{G}_{n,r})$.

By Lemmas 12 and 17,

$$\begin{aligned}
\tau_\infty(\widetilde{\mathbf{q}}_n; \widetilde{G}_{n,r}) &\leq 2752\tau_1(\widetilde{\mathbf{q}}_n; \widetilde{G}_{n,r}) \max\left\{2, \log\left(\frac{4}{\widetilde{s}_{n,r}}\right)\right\} \\
&\leq 2752\tau_1(\widetilde{\mathbf{q}}_n; \widetilde{G}_{n,r}) \max\left\{2, 4d\log\left(\frac{2D\Lambda_\sigma^2}{\lambda_\sigma^2}\right)\right\}
\end{aligned}$$

We now upper bound $\tau_1(\widetilde{\mathbf{q}}_n; \widetilde{G}_{n,r})$. From Lemma 14, we have that

$$\limsup_{n\to\infty} \tau_1(\widetilde{\mathbf{q}}_n; \widetilde{G}_{n,r}) \leq \limsup_{n\to\infty} \frac{1400}{3}\left(5 + \int_{\widetilde{s}_{n,r}}^{1/2} \frac{4}{t\widetilde{\Phi}^2_{n,r}(t)} dt\right) \tag{A.28}$$

(Since r remains constant, for sufficiently large n, $\mathbf{D}_{xx} > C$ will be fulfilled for any $x \in \mathcal{C}_\sigma[\mathbf{X}]$, and any $C < \infty$.) We set aside the constant term for the moment and turn to the integral. By Lemma 17,

$$\limsup_{n\to\infty} \int_{\widetilde{s}_{n,r}}^{1/2} \frac{4}{t\widetilde{\Phi}^2_{n,r}(t)} dt \leq \limsup_{n\to\infty} \int_{s_{\mathbb{P},r}}^{1/2} \frac{4}{t\widetilde{\Phi}^2_{n,r}(t)} dt$$

where $s_{\mathbb{P},r}$ is as in the proof of Theorem 2. We now replace the discrete conductance function $\widetilde{\Phi}_{n,r}$ by the stepwise approximation to the continuous conductance function, $\overline{\Phi}_{n,r}$:

$$\begin{aligned}
\limsup_{n\to\infty} \int_{s_{\mathbb{P},r}}^{1/2} \frac{4}{t\widetilde{\Phi}^2_{n,r}(t)} dt &\overset{(i)}{\leq} \limsup_{n\to\infty} \int_{s_{\mathbb{P},r}}^{1/2} \frac{4}{t\overline{\Phi}^2_{n,r}(t)} dt \\
&= \int_{s_{\mathbb{P},r}}^{1/2} \limsup_{n\to\infty} \frac{4}{t\overline{\Phi}^2_{n,r}(t)} dt \\
&\overset{(ii)}{\leq} \int_{s_{\mathbb{P},r}}^{1/2} \frac{4}{t\overline{\Phi}^2_{\mathbb{P},r}(t)} dt + c_d^2 \log(s_{\mathbb{P},r}) \frac{1}{r^2} o_r(1)
\end{aligned}$$

where (i) follows from Lemma 19 and (ii) from Lemma 21 (along with the continuous mapping theorem). Now, we make use of Lemma 20:

$$\int_{s_{\mathbb{P},r}}^{1/2} \frac{4}{t\overline{\Phi}_{\mathbb{P},r}^2(t)} dt \leq \frac{\Lambda_\sigma^8}{\lambda_\sigma^8} \cdot \left(331776 \int_{s_{\mathbb{P},r}}^{1/2} \frac{d}{t} dt + \int_{s_{\mathbb{P},r}}^{1/2} \frac{81 d D^2}{r^2 t \mathrm{Log}(\frac{\Lambda_\sigma^2}{t\lambda_\sigma^2})} dt \right)$$

$$\leq \frac{\Lambda_\sigma^8}{\lambda_\sigma^8} \cdot \left(331776 \underbrace{\int_{s_{\mathbb{P},r}}^{1/2} \frac{d}{t} dt}_{:=\mathcal{J}_1} + 81 \underbrace{\int_{s_{\mathbb{P},r}}^{\lambda_\sigma^2/(4\Lambda_\sigma^2)} \frac{dD^2}{r^2 t \log(\frac{\Lambda_\sigma^2}{t\lambda_\sigma^2})} dt}_{:=\mathcal{J}_2} + 81 \underbrace{\int_{\lambda_\sigma^2/(4\Lambda_\sigma^2)}^{1/2} \frac{dD^2}{r^2 t \log(1 + \frac{4\lambda_\sigma^2}{\Lambda_\sigma^2})} dt}_{:=\mathcal{J}_3} \right)$$

Computing a few simple integrals yields the following upper bounds on $\mathcal{J}_1, \mathcal{J}_2, \mathcal{J}_3$:

$$\begin{aligned}
\mathcal{J}_1 &\leq d^2 \log\left(\frac{2D\Lambda_\sigma^2}{r\lambda_\sigma^2}\right) \\
\mathcal{J}_2 &\leq \frac{dD^2}{r^2} \left[\log(2d) + \log\left(\log\left(\frac{2D}{r}\right)\right)\right] \\
\mathcal{J}_3 &\overset{(iii)}{\leq} 2\frac{dD^2}{r^2} \frac{\Lambda_\sigma^2}{\lambda_\sigma^2} \log\left(4\frac{\Lambda_\sigma^2}{\lambda_\sigma^2}\right)
\end{aligned}$$

where (iii) uses the upper bound $\frac{1}{\log(1+x)} \leq \frac{1}{x}$.

Plugging these bounds in to (A.28) gives the desired upper bound on $\tau_\infty(\widetilde{q}_n, \widetilde{G}_{n,r})$, which translates to the lower bound of (10).

A.11 Concentration inequalities

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

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

For Lemmas 26, let $X_1, \ldots, X_n \in \mathcal{X}$ be independent and identically distributed. We will additionally assume the order-m kernel function k satisfies the boundedness property $\sup_{x_1,\ldots,x_m} |k(x_1, \ldots, x_m)| \leq 1$.

Lemma 26 (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 27. *Let $X_i \in \{0,1\}$ for $i = 1,\ldots,n$ and let $\mu = \mathbb{E}(\sum_{i=1}^n X_i)$. Then,*

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

A.12 Proof of Theorem 4

Recall that for teleportation parameter $\alpha \in (0,1]$ and seed node $v \in \mathbf{X}$, we write $\mathbf{p}(v,\alpha;G_{n,r})$ for the PPR vector computed over neighborhood graph $G_{n,r}$; let $\mathbf{p}(u)$ denote the entry of $\mathbf{p}(v,\alpha;G_{n,r})$ at vertex $u \in \mathbf{X}$. Recalling that we write $\widetilde{\boldsymbol{\pi}}_{n,r}$ for the stationary distribution over $\widetilde{G}_{n,r}$, write $\widetilde{\boldsymbol{\pi}}_{n,r}(u)$ denote the stationary distribution at vertex $u \in \mathcal{C}_\sigma[\mathbf{X}]$.

Lemma 28 states the key results needed to link Theorems 1 and 3 with Theorem 4.

Lemma 28. *Let $0 < r < \sigma$ and*

$$\frac{\Psi_{n,r}(\mathcal{C}_\sigma[\mathbf{X}])}{10} \leq \alpha \leq \frac{\Psi_{n,r}(\mathcal{C}_\sigma[\mathbf{X}])}{9}. \tag{A.29}$$

There exists a good set $\mathcal{C}_\sigma[\mathbf{X}]^g \subseteq \mathcal{C}_\sigma[\mathbf{X}]$ with $\mathrm{vol}(\mathcal{C}_\sigma[\mathbf{X}]^g;G_{n,r}) \geq \mathrm{vol}(\mathcal{C}_\sigma[\mathbf{X}];G_{n,r})/2$ such that the following statements hold with respect to $\mathbf{p}(v,\alpha;G_{n,r})$ for any $v \in \mathcal{C}_\sigma[\mathbf{X}]^g$:

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

$$\mathbf{p}(u) \geq \frac{4}{5}\widetilde{\pi}_{n,r}(u) - \frac{2\Phi_{n,r}(\mathcal{C}_\sigma[\mathbf{X}])}{\Psi_{n,r}(\mathcal{C}_\sigma[\mathbf{X}])\widetilde{D}_{\min}} \tag{A.30}$$

- *For any $\mathcal{C}' \in \mathbb{C}_f(\lambda)$ such that $\mathcal{C}' \neq \mathcal{C}$, and for each $u' \in \mathcal{C}'_\sigma[\mathbf{X}]$,*

$$\mathbf{p}(u') \leq \frac{2\Phi_{n,r}(\mathcal{C}_\sigma[\mathbf{X}])}{\Psi_{n,r}(\mathcal{C}_\sigma[\mathbf{X}])\widetilde{D}_{\min}} \tag{A.31}$$

Proof. The proof of Lemma 28 is lengthy, but not difficult. We begin with some notation. Recall that $\mathbf{W}_n = \mathbf{D}_n\mathbf{A}_n^{-1}$ is the transition probability matrix over $G_{n,r}$. Let $\widetilde{\mathbf{D}}_n$ and $\widetilde{\mathbf{W}}_n$ be the degree and random walk matrices for the subgraph $\widetilde{G}_{n,r}$. Then, given $q \in \mathbb{R}^{\mathcal{C}_\sigma[\mathbf{X}]}$ we let

$$\widetilde{\mathbf{p}}_q = \alpha q + (1-\alpha)\widetilde{\mathbf{p}}_q\widetilde{\mathbf{W}}_n \tag{A.32}$$

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

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

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

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

In words, for $u \in \mathcal{C}_\sigma[\mathbf{X}]$, $\ell_t(u)$ is the probability that a random walk originating from v stays within $\mathcal{C}_\sigma[\mathbf{X}]$ for t steps, arriving at the tth step at u, and then leaks out of $\mathcal{C}_\sigma[\mathbf{X}]$ on the $t+1$th step. For $w \notin \mathcal{C}_\sigma[\mathbf{X}]$, $s_t(w) = 0$. By contrast, $s_t(w)$ is the probability that a random walk originating from v stays within $\mathcal{C}_\sigma[\mathbf{X}]$ for t steps, and then moves to 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 clearly have the same interpretation, but with respect to the total mass leaked and soaked by the PPR vector.

We first prove (A.30), and begin by restating some results of [6], adapted to our notation. By Lemma 3.1 of [6], there exists a good set $\mathcal{C}_\sigma[\mathbf{X}]^g \subseteq \mathcal{C}_\sigma[\mathbf{X}]$ with $\mathrm{vol}(\mathcal{C}_\sigma[\mathbf{X}]^g; G_{n,r}) \geq \mathrm{vol}(\mathcal{C}_\sigma[\mathbf{X}]; G_{n,r})/2$ such that for every $v \in \mathcal{C}_\sigma[\mathbf{X}]^g$

$$\mathbf{p}(u) \geq \widetilde{\mathbf{p}}_v(u) - \widetilde{\mathbf{p}}_\ell(u), \text{ and } \|\ell\|_1 \leq \frac{2\Phi_{n,r}(\mathcal{C}_\sigma[\mathbf{X}])}{\alpha}. \tag{A.33}$$

If moreover $\alpha \leq \frac{\Psi_{n,r}(\mathcal{C}_\sigma[\mathbf{X}])}{9}$, then additionally by Corollary 3.3 of [6], for every $u \in \mathcal{C}_\sigma[\mathbf{X}]$

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

and along with (A.33), we obtain

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

We proceed to upper bound $\widetilde{\mathbf{p}}_\ell(u) \leq \|\ell\|_1 / \widetilde{D}_{\min}$, whence (A.30) follows by (A.33). We observe two facts regarding $\widetilde{\mathbf{p}}_\ell(u)$. First, for any $u \in \mathcal{C}$, 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 $q \in \mathbb{R}^{\mathcal{C}_\sigma[\mathbf{X}]}$ with $\sum_{w \in \mathcal{C}_\sigma[\mathbf{X}]} q(w) = 1$, and any $t \geq 1$,

$$\begin{aligned} q\widetilde{\mathbf{W}}_n^t(u) &\leq \|q\|_1 \, ||\widetilde{\mathbf{W}}_{\cdot u}||_\infty \\ &\leq \frac{1}{\widetilde{D}_{\min}} \end{aligned}$$

where $\widetilde{\mathbf{W}}_{\cdot u}$ is the uth 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

$$\deg(w; \widetilde{G}_{n,r}) \geq \widetilde{D}_{\min}.$$

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

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

and (A.30) is proved.

□

Lemma 29. *Let $\mathcal{C}_\sigma$ satisfy the conditions of Theorem 4. For $r < \sigma$, the following statements hold with probability tending to one as $n \to \infty$:*

$$\begin{aligned}
D_{\min}(\mathcal{C}_\sigma[X]; \widetilde{G}_{n,r}) &\geq \frac{1}{2}\nu_d r^d \lambda_\sigma \\
D_{\max}(\mathcal{C}_\sigma[X]; \widetilde{G}_{n,r}) &\leq 2\nu_d r^d \Lambda_\sigma \\
\widetilde{\mathrm{vol}}_{n,r}(\widetilde{G}_{n,r}) &\leq 2\nu(\mathcal{C}_\sigma)\Lambda_\sigma
\end{aligned}$$

where $D_{\min}(\mathcal{C}_\sigma[X]; \widetilde{G}_{n,r})$ is the minimum degree of any vertex $v \in \mathcal{C}_\sigma[X]$ in the subgraph $\widetilde{G}_{n,r}$, and analogously for $D_{\max}(\mathcal{C}_\sigma[X]; \widetilde{G}_{n,r})$.

The statement follows immediately from Lemma 16.

Proof of Theorem 4 We note that by Theorems 1 and 2,

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

As a result Lemma 28 implies

$$\begin{aligned}
p_u &\geq \frac{4}{5}\widetilde{\pi}_{n,r}(u) - \frac{18\kappa_2(\mathcal{C})}{\widetilde{D}_{\min}} \qquad && (u \in \mathcal{C}[\mathbf{X}]) \\
p_{u'} &\leq \frac{18\kappa_2(\mathcal{C})}{\widetilde{D}_{\min}} && (u' \in \mathcal{C}'[\mathbf{X}])
\end{aligned} \tag{A.34}$$

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

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

Application of Lemma 29 then yields

$$\widetilde{\pi}_{n,r}(u) \geq 8\frac{\lambda_\sigma}{\nu(\mathcal{C}_\sigma)\Lambda_\sigma^2} \tag{A.35}$$

as well as

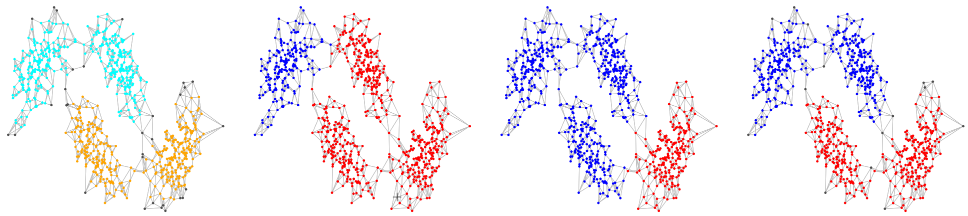
$$\frac{1}{\widetilde{D}_{\min}} \leq \frac{2}{\nu_d r^d \lambda_\sigma} \tag{A.36}$$

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

Combining (A.34), (A.35) and (A.36), along with the requirement on $\kappa_2(\mathcal{C})$ given by (14), 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 $n \to \infty$ with probability tending to one any sweep cut of the form of (4), including the output set $\widehat{C}$, will successfully recover $\mathcal{C}$ in the sense of (5).



(a) (b) (c) (d)

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 $\mathcal{A}_\sigma$, we will write the set difference between $\mathcal{A}_\sigma$ and the $(\sigma + r)$-expansion $\mathcal{A}_{\sigma+r}$ as

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

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

Lemma 30. We will repeatedly employ Lemma **??** and Lemma 3 in tandem. As a first example, in Lemma 30, 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

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

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

Lemma 30. *For σ, $\mathcal{A}_\sigma$ 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. □

C.2 Proof of Lemma 28

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

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

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

Roughly, the proof of Lemma 28 will unfold in four steps. The first two will result in the lower bound of (A.30), while the latter two will imply the upper bound in (A.31). We briefly summarize the approach before diving into the formal proof:

1. For $u\in\mathcal{C}[\mathbf{X}]$, we use the results of [6] to produce the lower bound
$$\mathbf{p}(u)\geq\frac{4}{5}\widetilde{\pi}_{n,r}(u)-\widetilde{\mathbf{p}}_\ell(u)$$
where
$$\widetilde{\mathbf{p}}_\ell=\alpha\ell+(1-\alpha)\widetilde{\mathbf{p}}_\ell\widetilde{\mathbf{W}}_n$$
is the PPR random walk over $\widetilde{G}_{n,r}$, 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 $\widetilde{\mathbf{p}}_\ell$ 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.30).
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.31), and hence Lemma 28.

Step 1 We will begin by restating some results of [6].

For seed node v, we write

$$\widetilde{\mathbf{p}}_v=\alpha e_v+(1-\alpha)\widetilde{\mathbf{p}}_v\widetilde{\mathbf{W}}_n \tag{A.37}$$
$$=\alpha\sum_{t=0}^{\infty}(1-\alpha)^t\left(e_v\widetilde{\mathbf{W}}_n^t\right) \tag{A.38}$$

From Lemma 3.1 of [6] we have that there exists a good set $\mathcal{C}_\sigma[\mathbf{X}]^g\subseteq\mathcal{C}_\sigma[\mathbf{X}]$ with $\mathrm{vol}(\mathcal{C}_\sigma[\mathbf{X}]^g)\geq\mathrm{vol}(\mathcal{C}_\sigma[\mathbf{X}])/2$ for all $v\in\mathcal{C}_\sigma[\mathbf{X}]^g$, $u\in\mathcal{C}_\sigma[\mathbf{X}]$

$$p_u\geq\widetilde{\mathbf{p}}_v(u)-\widetilde{\mathbf{p}}_\ell(u)$$
$$||\ell||_1\leq\frac{2\widetilde{\Phi}_{n,r}}{\alpha} \tag{A.39}$$

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

Moreover if, as we have specified, $\alpha \leq \widetilde{\Psi}_{n,r}/9$, Lemma 3.2 of [6] yields a lower bound on $\widetilde{p}$

$$\widetilde{\mathbf{p}}_v(u) \geq \frac{4}{5}\widetilde{\pi}_{n,r}(u). \tag{A.40}$$

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

$$\begin{aligned}
\widetilde{\mathbf{p}}_\ell(u) &= \alpha \sum_{t=0}^{\infty} (1-\alpha)^t \left(\ell \widetilde{\mathbf{W}}_n^t\right)(u) \\
&= \|\ell\|_1 \alpha \sum_{t=0}^{\infty} (1-\alpha)^t \left(\frac{\ell}{\|\ell\|_1} \widetilde{\mathbf{W}}_n^t\right)(u) \\
&\overset{\text{(i)}}{=} \|\ell\|_1 \alpha \sum_{t=1}^{\infty} (1-\alpha)^t \left(\frac{\ell}{\|\ell\|_1} \widetilde{\mathbf{W}}_n^t\right)(u) \\
&\overset{\text{(ii)}}{\leq} \|\ell\|_1 \frac{1}{\widetilde{D}_{\min}}
\end{aligned} \tag{A.41}$$

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

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

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

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

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

where $\widetilde{\mathbf{W}}_{\cdot u}$ is the uth column of $\widetilde{\mathbf{W}}_n$. (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 $\widetilde{\mathbf{W}}_{\cdot u}$ is either 0 or at most $1/\widetilde{D}_{\min}$.

Combined, (A.41), (A.40), and (A.39) imply

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

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

$$\mathbf{p}_v(u') \leq \mathbf{p}_s(u'). \tag{A.42}$$

Proof. We have

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

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

Lemma 32 allows us to make the transition to sums of soakage vectors.

Lemma 32. *Let $G = (V, E)$ be a graph, with associated random walk matrix W.*

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

$$qW^T = \sum_{t=0}^{T-1} s_t W^{T-t-1} + q(WI_S)^T \tag{A.43}$$

We prove Lemma 32 after completing the proof of Lemma 31.

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

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

and so

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

□

Proof of Lemma 32. 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 $\widetilde{\mathbf{p}}_\ell$ could assign to any one vertex, we can upper bound

$$\begin{aligned}
\mathbf{p}_s(u') &= \alpha \sum_{t=0}^{\infty} (1-\alpha)^t \left(s\mathbf{W}_n^t\right)(u') \\
&= \|s\|_1 \alpha \sum_{t=0}^{\infty} (1-\alpha)^t \left(\frac{s}{\|s\|_1}\mathbf{W}_n^t\right)(u') \\
&= \|s\|_1 \alpha \sum_{t=1}^{\infty} (1-\alpha)^t \left(\frac{s}{\|s\|_1}\mathbf{W}_n^t\right)(u') \\
&\leq \|s\|_1 \frac{1}{\widetilde{D}_{\min}}.
\end{aligned} \tag{A.44}$$

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

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

and as a result $\|s\|_1 = \|\ell\|_1$. Combining with $\|\ell\|_1 \leq 2\frac{\widetilde{\Phi}_{n,r}}{\alpha}$ and (A.44) yields the desired upper bound.

□

References

[1] Giovanni Alberti and Giovanni Bellettini. A non-local anisotropic model for phase transitions: asymptotic behaviour of rescaled energies. *European Journal of Applied Mathematics*, 9(3):261–284, 1998.

[2] Itai Benjamini and Elchanan Mossel. On the mixing time of a simple random walk on the super critical percolation cluster. *Probability Theory and Related Fields*, 125(3):408–420, Mar 2003.

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

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

[5] 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.

[6] Zeyuan Allen Zhu, Silvio Lattanzi, and Vahab S Mirrokni. A local algorithm for finding well-connected clusters. In *ICML (3)*, pages 396–404, 2013.