A Proofs for Section 2

Proof of Proposition 2. By Cirelson's theorem, we know that

$$\mathbb{P}\{\max_{v \in C} \epsilon_v \geq \mathbb{E} \max_{v \in C} \epsilon_v + u\} \leq e^{-u^2/2\sigma^2}$$

and we have that $\mathbb{E} \min_{v \in C} \epsilon_v$ is within a constant factor of $-\sigma\sqrt{2 \log |C|}$ by the Majorizing-Measure theorem. Hence, under H_1^C we have that $\max_{v \in C} y_v \geq \mu/\sqrt{|C|} + L^{-1}\sigma\sqrt{2\log|C|} - \sigma\sqrt{2\log(1/\delta)}$ with probability at least $1-\delta$ for some universal constant L. By similar reasoning, under H_0 we have that $\max_{v \in V} |y_v| \leq L\sigma\sqrt{2\log n} + \sigma\sqrt{2\log(2/\delta)}$. We have analogous opposite bounds that prove the necessary condition. Hence, a threshold, τ, distinguishes H_0 and H_1 if

$$\begin{aligned}\mu/\sqrt{|C|} + L^{-1}\sigma\sqrt{2\log|C|} - \sigma\sqrt{2\log 1/\delta} > \tau \\ > L\sigma\sqrt{2\log n} + \sigma\sqrt{2\log 1/\delta}\end{aligned}$$

which occurs if $\mu/\sigma - \sqrt{2|C|\log n} = \omega(1)$. Choosing C such that $|C| \geq n/2$ we have the sufficient condition. □

Proof of Proposition 3. Under H_0, $\frac{1}{n}\sum_{v \in V} y_v$ is normally distributed with mean 0 and variance σ^2/n. Meanwhile under H_1^C, $\frac{1}{n}\sum_{v \in V} y_v$ is normally distributed with mean $\mu\sqrt{|C|}/n$ and variance σ^2/n. Hence, the test statistic asymptotically distinguished H_0 from H_1 if and only if

$$\Phi\left(-\frac{\mu}{\sigma}\sqrt{\frac{|C|}{n}}\right) = o(1)$$

where Φ is the CDF of the standard normal. □

B Proofs for Section 3

B.1 Properties of the Spanning Tree Wavelet Basis

Lemma 14. *The output* $\mathbf{B}$ *of the Spanning Tree Wavelet Construction has the following properties:*

1. $\mathbf{B}$ *is an orthonormal basis of* $\mathbb{R}^n$*, in particular there are* n *vectors.*
2. $\mathbf{B}$ *can be computed in* $O(n \log n \log d_{\mathcal{T}})$ *time.*

Proof. Before we dive into the Spanning Tree Wavelet Construction, we must analyze the FormWavelets routine. FormWavelets operates on the subtrees c_i as if they form a chain structure, and it constructs the Haar Wavelet basis over this chain structure. A fairly straightforward inductive argument shows that, coupled with the all-ones vector, the output of FormWavelets is an orthonormal basis over the subtrees c_i. In particular, a linear combination of these vectors can be used to obtain a vector that is constant and non-zero on a single subtree and zero elsewhere.

Analyzing the Spanning Tree Wavelet Construction also proceeds by induction. On the two-node graph, running the Spanning Tree Wavelet Construction gives a basis with two elements $(1/\sqrt{2}, 1/\sqrt{2})^T$ and $(1/\sqrt{2}, -1/\sqrt{2})^T$ which is clearly an orthobasis. For any tree $\mathcal{T}$, FormWavelets gives an orthonormal basis over the subtrees, and the inductive hypothesis gives the basis for each of these individual subtrees. By induction, recursing on each subtree $\mathcal{T}_i$ gives a set of orthonormal vectors that are also orthogonal to $\mathbf{1}_{\mathcal{T}_i}$. Consequently $\mathbf{B}$ is an orthobasis.

Computing the basis involves finding a balancing vertex and then recursively forming the basis elements. Finding a balancing vertex can be done in linear time by precomputing all of the subtree sizes using a depth-first tree traversal. Computing the basis just involves another pre-order traversal of the tree, where at each level in the tree we construct $O(\lceil \log d_{\mathcal{T}} \rceil)$ basis elements, with a total of $O(n\lceil \log d_{\mathcal{T}} \rceil)$ non-zero coefficients. Repeating this across the $\lceil \log n \rceil$ levels gives the running time bound. □

B.2 Proof of Lemma 4

Before we proceed with the proof, we state and prove two results on the performance of the algorithm:

Lemma 15. *Let* $\mathcal{T}$ *be a tree.* FindBalance *returns a vertex* v *such that the largest connected component of* $\mathcal{T}\backslash v$ *is of size at most* $\lceil |\mathcal{T}|/2 \rceil$ *in* $O(|\mathcal{T}|)$ *time.*

Proof. Let the objective be the size of the largest connected components of $\mathcal{T}\backslash v$. Every move in *FindBalance* reduces the objective by at least 1 and the objective can be at most $|\mathcal{T}|-1$ so it must terminate in less than $|\mathcal{T}|$ moves. Now at any step of *FindBalance*, if the objective is greater than $\lceil |\mathcal{T}|/2 \rceil$, the cumulative size of the remaining connected components is less than $\lfloor |\mathcal{T}|/2 \rfloor$. Hence, in the next step the connected component formed by these is less than $\lceil |\mathcal{T}|/2 \rceil$. Thus, the program cannot terminate at a move directly after the objective is greater than $\lceil |\mathcal{T}|/2 \rceil$. □

We will also require the following claim. Indeed, controlling the depth of the recursion in the wavelet construction is the sine qua non for controlling the sparsity, $\|\mathbf{B}^T\mathbf{x}\|_0$.

Claim 16. *The wavelet construction has recursion depth at most* $\lceil \log d_{\mathcal{T}} \rceil \lceil \log n \rceil$.

Proof. Whenever *FormWavelet* is applied it increases the height of the dendrogram by at most $\lceil \log d_{\mathcal{T}} \rceil$. By lemma 15 the size of the remaining components is halved, so the algorithm terminates in at most $\lceil \log n \rceil$ steps. □

Proof of Lemma 4. We will show that any edge $e \in \mathcal{T}$ supports at most $\lceil \log d_{\mathcal{T}} \rceil \lceil \log n \rceil$ basis elements in $\mathbf{B}$, and this will imply the result. We will say that an edge e supports a basis element $\mathbf{b}$ if $e \subseteq supp(\nabla_{\mathcal{T}} \mathbf{b})$. It follows that for a basis element $\mathbf{b}$, if $\mathbf{b}^T \mathbf{x} \neq 0$ then $\exists e$ that supports b. Let no_basis(e) be the number of basis elements that are supported by e (no_basis(e) $= 0$ if $e \notin supp(\nabla_{\mathcal{T}} \mathbf{x})$). We then have

$$||\mathbf{B}^T \mathbf{x}||_0 \leq \sum_{e \in supp(\nabla_{\mathcal{T}} \mathbf{x})} \text{no\_basis}(e)$$

Consider some edge e. If e supports some subtree $\mathcal{T}_{sub}$ (we use this interchangeably with supporting a basis element formed by partitioning $\mathcal{T}_{sub}$ into two groups), then e supports at most one of $\mathcal{T}_{sub}$'s subtrees. This implies that no_basis(e) is upper bounded by the depth of the recursion. By the claim, we find that,

$$\begin{aligned} ||\mathbf{B}^T \mathbf{x}||_0 &\leq \sum_{e \in supp(\nabla_{\mathcal{T}} \mathbf{x})} \lceil \log d_{\mathcal{T}} \rceil \lceil \log n \rceil \\ &\leq ||\nabla_{\mathcal{T}} \mathbf{x}||_0 \lceil \log d_{\mathcal{T}} \rceil \lceil \log n \rceil \end{aligned}$$

proving the first claim. The second claim is obvious from the fact that $\mathcal{T}$ contains a subset of the edges in $\mathcal{G}$, so every cut has larger cut size in $\mathcal{G}$ than it does in $\mathcal{T}$. □

B.3 Proof of Theorem 5

Proof. Under the null $\mathbf{x} = 0$, and we have that

$$||\mathbf{B}^T \mathbf{y}||_\infty = ||\mathbf{B}^T \boldsymbol{\epsilon}||_\infty < \sigma \sqrt{2 \log(n/\delta)}$$

with probability at least $1 - \delta$. So, as long as $\tau = \sigma \sqrt{2 \log(n/\delta)}$ then we control the probability of false alarm (type 1 error). For a element $\mathbf{x}$ of the alternative, let the index, i^*, achieve the maximum of $\mathbf{B}^T \mathbf{x}$ (i.e. $||\mathbf{B}^T \mathbf{x}||_\infty = |\mathbf{B}^T \mathbf{x}|_{i^*}$). Then $|\mathbf{B}^T \mathbf{y}|_{i^*} \geq |\mathbf{B}^T \mathbf{x}|_{i^*} - \sigma \sqrt{2 \log(1/\delta)}$ with probability at least $1 - \delta$ and

$$|\mathbf{B}^T \mathbf{x}|_{i^*}^2 = ||\mathbf{B}^T \mathbf{x}||_\infty^2 \geq \frac{\sum_{i:(\mathbf{B}^T \mathbf{x})_i \neq 0} (\mathbf{B}^T \mathbf{x})_i^2}{||\mathbf{B}^T \mathbf{x}||_0} = \frac{||\mathbf{x}||_2^2}{||\mathbf{B}^T \mathbf{x}||_0}$$

Taking square roots and combining this with Lemma 4,

$$||\mathbf{B}^T \mathbf{x}||_\infty \geq \frac{||\mathbf{x}||_2}{\sqrt{||\nabla_{\mathcal{T}} \mathbf{x}||_0 \lceil \log d_{\mathcal{T}} \rceil \lceil \log n \rceil}}$$

from which we have the result that under H_1,

$$||\mathbf{B}^T \mathbf{y}||_\infty \geq \frac{||\mathbf{x}||_2}{\sqrt{||\nabla_{\mathcal{T}} \mathbf{x}||_0 \lceil \log d_{\mathcal{T}} \rceil \lceil \log n \rceil}} - \sigma \sqrt{2 \log(1/\delta)}$$

Furthermore, because the first wavelet coefficient is just the vertex average, $\mathbf{B}_0^T \mathbf{y}$, we know that under H_1 with probability at least $1 - \delta$,

$$|\mathbf{B}_0^T \mathbf{y}| \geq \sqrt{\frac{|C|}{n}} \mu - \sigma \sqrt{2 \log(1/\delta)}$$

Forcing the maximum of these lower bounds to be greater than τ gives us our result. □

C Proofs For Section 5

C.1 Proof of Corollary 12

First we restate Corollary 9 from [32]:

Corollary 17. *Consider an unweighted symmetric or mutual k-NN graph built from a sequence $X_1, \ldots, X_n$ drawn i.i.d. from a density p. Then there exists constants c_1, c_2, c_3 such that with probability at least $1 - c_1 n \exp(-k c_2)$ we have uniformly for all $i \neq j$ that:*

$$\left| \frac{k}{2m} H_{ij} - \frac{k}{d_j} \right| \leq c_3 \frac{n^{2/d}}{k^{1+2/d}}$$

Proof of Corollary 12. We focus on the symmetric k-NN graph in which we connect v_i to v_j if v_i is in the k-nearest neighbors of v_j or vice versa. In this graph, every node has degree $\geq k$ which will be crucial in our analysis. Our goal is to bound the effective resistance of every edge, so that we can subsequently bound r_{max} and apply Corollary 9. From the definition of r_e we have:

$$\begin{aligned} r_{ij} &= \frac{1}{2} \left(\frac{H_{ij}}{m} + \frac{H_{ji}}{m} \right) \\ &\leq 2c_3 \frac{n^{2/d}}{k^{2+2/d}} + \frac{1}{d_i} + \frac{1}{d_j} \\ &\leq 2c_3 \frac{n^{2/d}}{k^{2+2/d}} + \frac{2}{k} \end{aligned}$$

Where the first line is the definition of r_{ij}, the second line follows from Corollary 17 and the last line follows from the fact that $d_i \geq k$ for each vertex. Since $k(k/n)^{2/d} \to \infty$, we see that $r_{ij} = O(\frac{1}{k})$. Moreover, with this scaling of k, that the probability in Corollary 17 is going to 1. We can therefore bound r_{max} as:

$$r_{max} \leq \rho \left(2c_3 \frac{n^{2/d}}{k^{2+2/d}} + \frac{2}{k} \right) = O\left(\frac{\rho}{k} \right)$$

Since the first term is going to zero with n. Plugging in this bound on r_{max} into Theorem 9 gives the result (and substituting $d = k$). □

C.2 Proof of Corollary 13

As before, we first state Corollary 8 from [32]:

Corollary 18. *Consider an unweighted ϵ-graph built from the sequence $X_1, \ldots, X_n$ drawn i.i.d. from the density p .Then there exists constants $c_1, \ldots c_5 > 0$ such that with probability at least $1 - c_1 n \exp(-c_2 n\epsilon^D) - c_3 \exp(-c_4 n\epsilon^D)/\epsilon^D$, we have uniformly for all $i \neq j$ that:*

$$\left| \frac{n\epsilon^D}{2m} H_{ij} - \frac{n\epsilon^D}{d_j} \right| \leq \frac{c_5}{n\epsilon^{D+1}}$$

Proof of Corollary 13. Some manipulation of the result in Corollary 18 reveals that:

$$H_{ij} \leq \frac{2m}{d_j} + \frac{2c_5 m}{n^2 \epsilon^{2D+2}}$$

Under our scaling, the second term goes to zero and the probability in Corollary 18 goes to one, so $H_{ij} = O(m/d_j)$. We will now give a lower bound on d_j. If X_i is in the ball of radius ϵ centered at X_j, then we connect X_i and X_j. Thus d_j is exactly the number of vertices in the $B(X_j; \epsilon)$. The regularity condition on p in [32] requires that there exists constants α and ϵ_0 such that for all $\epsilon < \epsilon_0$ and for all $x \in \text{supp}(p)$, $\text{vol}(B(x;\epsilon) \cap \text{supp}(p)) \geq \alpha \text{vol}(B(x;\epsilon))$. By this fact, the fact that the density is lower bounded by $p_{\min}$, and by the fact that $\epsilon \to 0$, we know that for sufficiently large n, $p(B(X_j;\epsilon)) \geq p_{\min} \alpha c_D \epsilon^D$ where $c_D \epsilon^D$ is the volume of a D-dimensional ball of radius ϵ. The probability that $X_i \in B(X_j;\epsilon)$ is distributed as a Bernoulli random variable with mean $\geq \alpha p_{min} c_D \epsilon^D$. By Hoeffding's inequality and a union bound we get that:

$$d_j \geq n\alpha p_{min} c_D \epsilon^D + \sqrt{n \log(n)} = \Omega(n\epsilon^D)$$

for all vertices j with probability $\geq 1 - 1/n$. Using the definition of $r_{i,j}$ along with the bound on H_{ij} and d_j we have that uniformly for all pairs i, j:

$$r_{i,j} = O(\frac{1}{n\epsilon^D})$$

Plugging in this bound into Theorem 9 gives us the result. $\square$