Metric Embedding Theory and its Algorithms (67720) – Exercise 4

Mike Greenbaum 211747639

January 2021

1. (a) Define T to be the MST of C_n, therefore we can denote $V_T = \{v_1, \ldots, v_n\}$ s.t. $E_T = \{\{v_i, v_{i+1}\} \mid 1 \le i \le n-1\}$.
Define $e_i = \{v_i, v_{i+1}\}$ and $e_n = \{v_1, v_n\}$.
Due to T being the MST, it follows that $\forall 1 \le i \le n-1, w(e_i) \le w(e_n)$, else we could have removed e_i and added e_n and get a better tree in contradiction to minimallity.
Define $A = \{(v_i, v_j) \mid d_T(v_i, v_j) \ne d_G(v_i, v_j) \wedge i < j\}$.
If $|A| = 0$, then the tree T embeds isometrically and we have $(1,1)$-spanning tree cover which is also $(1,2)$-spanning tree cover.
Else $|A| \ne 0$, therefore define $a = \min_{(v_i,v_j)\in A}(j-i)$.
Define $B = \{(v_i, v_j) \mid j - i = a\} \cap A$. Notice that B is not empty, denote $(v_i, v_j) \in B$. Let $(v_{i'}, v_{j'}) \in A$. We will prove a couple of cases:

i. If $v_i < v_{i'} < v_j$, due to (v_i, v_j) having the minimal distance it follows that $v_i < v_{i'} < v_j < v_{j'}$. Due to $(v_i, v_j) \in A$, it follows that the minimal path between v_i, v_j is $v_i \to v_1 \to v_n \to v_{j'} \to v_j$, therefore

$$d_T(v_i, v_{i'}) + d_T(v_{i'}, v_j) = d_T(v_i, v_j) > d_G(v_i, v_j) = d_T(v_i, v_1) + w(e_n) + d_T(v_n, v_{j'}) + d_T(v_{j'}, v_j)$$

Due to $(v_{i'}, v_{j'}) \in A$, it follows that the minimal path between $v_{i'}, v_{j'}$ is $v_{i'} \to v_i \to v_1 \to v_n \to v_{j'}$, therefore

$$d_T(v_{i'}, v_j) + d_T(v_j, v_{j'}) = d_T(v_{i'}, v_{j'}) > d_G(v_{i'}, v_{j'}) = d_T(v_{i'}, v_i) + d_T(v_i, v_1) + w(e_n) + d_T(v_n, v_{j'})$$

Subtracting the equations we get that

$$d_T(v_i, v_{i'}) - d_T(v_j, v_{j'}) > d_T(v_{j'}, v_j) - d_T(v_i, v_{i'}) \implies \boxed{d_T(v_i, v_{i'}) > d_T(v_{j'}, v_j)}$$

Subtracting the results in the opposite way we get that

$$d_T(v_j, v_{j'}) - d_T(v_i, v_{i'}) > d_T(v_i, v_{i'}) - d_T(v_{j'}, v_j) \implies \boxed{d_T(v_{j'}, v_j) > d_T(v_i, v_{i'})}$$

and we got a contradiction.

ii. If $v_i < v_{j'} < v_j$, due to (v_i, v_j) having the minimal distance it follows that $v_{i'} < v_i < v_{j'} < v_j$. Due to $(v_i, v_j) \in A$, it follows that the minimal path between v_i, v_j is $v_i \to v_{i'} \to v_1 \to v_n \to v_j$, therefore

$$d_T(v_i, v_{j'}) + d_T(v_{j'}, v_j) = d_T(v_i, v_j) > d_G(v_i, v_j) = d_T(v_i, v_{i'}) + d_T(v_{i'}, 1) + w(e_n) + d_T(v_n, v_{j'})$$

Due to $(v_{i'}, v_{j'}) \in A$, it follows that the minimal path between $v_{i'}, v_{j'}$ is $v_{i'} \to v_1 \to v_n \to v_j \to v_{j'}$, therefore

$$d_T(v_{i'}, v_i) + d_T(v_i, v_{j'}) = d_T(v_{i'}, v_{j'}) > d_G(v_{i'}, v_{j'}) = d_T(v_{i'}, v_1) + w(e_n) + d_T(v_n, v_j) + d_T(v_{j'}, v_j)$$

Subtracting the equations we get that

$$d_T(v_j, v_{j'}) - d_T(v_i, v_{i'}) > d_T(v_i, v_{i'}) - d_T(v_{j'}, v_j) \implies \boxed{d_T(v_{j'}, v_j) > d_T(v_i, v_{i'})}$$

Subtracting the results in the opposite way we get that

$$d_T(v_i, v_{i'}) - d_T(v_j, v_{j'}) > d_T(v_{j'}, v_j) - d_T(v_i, v_{i'}) \implies \boxed{d_T(v_i, v_{i'}) > d_T(v_{j'}, v_j)}$$

and we got a contradiction.

So for every $(v_{i'}, v_{j'}) \in A$ it follows that $v_{i'} \le v_i < v_j \le v_{j'}$. Therefore, the minimal path between $v_{i'} \to v_{j'}$ doesn't pass through e_i.
Define $C = \{(v_i, v_j) \mid (v_i, v_j) \notin A \wedge i < j\})$. Notice that $A \cup C = \binom{V}{2}$.
Therefore, if we define $T_A = \langle V, E \setminus \{e_i\}\rangle$ we will get a tree s.t. $\forall (x,y) \in A$, $d_{T_A}(x,y) = d_G(x,y)$ i.e. $dist(f_{T_A}) \mid_A = 1$. And $\forall (x,y) \in C$, it follows from definition of A that $d_T(v_i, v_j) = d_G(v_i, v_j)$ i.e. $dist(f_T) \mid_C = 1$
Therefore, $(T_A, A), (T, C)$ are 2 spanning trees that cover all of V with distortion 1, therefore, we showed that there is $(1,2)$-spanning tree cover for G, as required.

(b) Let $f_1, \dots, f_k$ be the embeddings of G. Let D be the distribution of $f_1, \dots, f_k$.
Notice that $1 \leq \exists i \leq l$ s.t. $\mathbb{P}(f_i \sim D) \geq \frac{1}{k}$ (because there are k elements that sum to 1).
Notice that $\exists x \neq y \in X$ s.t. $dist_{f_i}(x, y) = \Omega(n)$ (proved in class, Theorem 8.11).
Therefore, $\mathbb{E}[dist_f(x, y)] \geq dist_{f_i}(x, y) \cdot \mathbb{P}(f_i \sim D) \geq \frac{dist_{f_i}(x,y)}{k} = \Omega\left(\frac{n}{k}\right)$.
Therefore, we showed that for every k embeddings and for any distribution, there is a pair that its expected distortion $\Omega\left(\frac{n}{k}\right)$, and therefore the expected distortion of the graph is $\Omega\left(\frac{n}{k}\right)$, as required.

(c) Let $E = \{e_1, \dots, e_n\}$ s.t. $\forall i \in [n], w(e_i) \geq w(e_{i+1})$.
Define $f_i = \langle V, E \setminus \{e_i\} \rangle$. Notice that embeddings are non-contractive. Define $D_i = \frac{w(e_i)}{\sum_{j=1}^{k} w(e_j)}$ and $D = (D_1, \dots, D_k)$ will be the distribution over $(f_1, \dots, f_k)$.
Notice that $w(G) = \sum_{j=1}^{n} w(e_j)$.
Let $v \neq u \in X$, let $v \to u$ be the minimal path between v to u, let $e'_1, \dots, e'_l$ be the edges of the path. Define $A = \{e_1, \dots, e_k\} \cap \{e'_1, \dots, e'_l\}$.
Note that the probability that the path still exists in the chosen tree is $1 - \frac{\sum_{e_j \in A} w(e_j)}{\sum_{j=1}^{k} w(e_j)}$
Notice that if the minimal path is not in the tree, then we removed an edge from A which happens with prob. $\frac{\sum_{e_j \in A} w(e_j)}{\sum_{j=1}^{k} w(e_j)}$ and we know that $\sum_{e_j \in A} w(e_j) \leq d_G(u, v)$ because $e_j \in A$ are all in the minimal path and the weights are non negative.
Denote the random variable $X =$ "the minimal path between u and v exists in the tree"

$$\begin{aligned}
\mathbb{E}[d_f(f(u), f(v))] &= \mathbb{P}(x) \cdot \mathbb{E}[d_f(f(u), f(v)) \mid X] + (1 - \mathbb{P}(X)) \cdot \mathbb{E}[d_f(f(u), f(v)) \mid \neg X] \\
&\leq d_G(u, v) + \left(\frac{\sum_{e_j \in A} w(e_j)}{\sum_{j=1}^{k} w(e_j)}\right) \cdot w(G) \leq d_G(u, v) + \frac{d_G(u, v)}{\sum_{j=1}^{k} w(e_j)} \cdot w(G) \\
&\leq d_G(u, v) \cdot \left(1 + \frac{\sum_{j=1}^{n} w(e_j)}{\sum_{j=1}^{k} w(e_j)}\right) = d_G(u, v) \left(1 + \frac{\sum_{j=1}^{k} w(e_j) + \sum_{j=k+1}^{n} w(e_j)}{\sum_{j=1}^{k} w(e_j)}\right) \\
&\leq d_G(u, v) \left(1 + 1 + \frac{\sum_{j=k+1}^{n} w(e_k)}{\sum_{j=1}^{k} w(e_k)}\right) \leq d_G(u, v) \left(1 + 1 + \frac{n}{k}\right) = d_G(u, v) \cdot O\left(\frac{n}{k}\right)
\end{aligned}$$

Notice that we showed k trees $(f_1, \dots, f_k)$ (non-contracting) and distribution D s.t. $\forall u \neq v, \mathbb{E}[d_f(f(u), f(v))] = d_G(u, v) \cdot O\left(\frac{n}{k}\right)$.
Therefore, the expected distortion of the graph is $O\left(\frac{n}{k}\right)$, as required.

(d) I proved the bonus and its proof is listed below (after q2). By using the bonus and the fact that $\dim(X) = O(1)$, we get immediately what we wanted to prove. I still keep here the original proof in case my solution to the bonus is faulty.
We will first prove that there is a Δ-probalistic embedding with the metric $d(x, y) = \min\{|x - y|, 1 - |x - y|\}$ for $[0, 1)$.
Let $0 < \eta < \frac{1}{16}$. Let $k \sim Uni([0, \Delta])$. Define the following partition on $[0, 1)$, $\forall 1 \leq i \leq \frac{1}{\Delta} - 1, S_i = B\left(k + (i-1) \cdot \Delta, \frac{\Delta}{2}\right) \setminus \{k + i \cdot \Delta\}$.
First $\forall x \in [k, 1)$, $x \in S_{\lfloor \frac{x-k}{\Delta} \rfloor}$.
And $\forall x \in [0, k), x \in S_{\frac{1}{\Delta}}$.
Therefore, $\bigcup_{i=1}^{\frac{1}{\Delta}} S_i = [0, 1)$.
Assume that $x \in S_i \wedge S_j$, therefore, the distance centers between S_i and S_j are at most $2 \cdot \frac{\Delta}{2} = \Delta$ apart and it is at least Δ by construction. Therefore, we get that x needs to be $\frac{\Delta}{2}$ apart from the centers. wlog $i = j + 1$, therefore $x \notin S_i$ because we removed $k + i \cdot \Delta$.
Therefore $\forall i \neq j, S_i \cap S_j = \emptyset$ and notice that $Diam(S_i) \leq \Delta$.
Therefore, $S_1, \dots, S_{\frac{1}{\Delta}}$ is a partition.
Define $S_i(x)$ the center of $x's$ cluster. Notice that $|x - S_i(x)| \sim Uni(0, \frac{\Delta}{2})$

$$\begin{aligned}
\mathbb{P}(B(x, \eta \cdot \Delta) \not\subseteq P(x)) &\leq \mathbb{P}(d_X(x, S_i(x)) \geq 1 - \eta \cdot \Delta) \leq \mathbb{P}(|x - S_i(x)| \geq 1 - \eta \cdot \Delta) + \mathbb{P}(1 - |x - S_i(x)| \geq 1 - \eta \cdot \Delta) \\
&\leq 2\eta + \eta = O(1) \cdot \eta
\end{aligned}$$

Therefore, we proved that there exists a Δ- bounded probabilistic partition P with padding parameter $O(1)$ for $X = [0, 1)$.

Denote the $V_G = \{v_1, \dots, c_n\}$.
Define $f(v_i) = \frac{\sum_{j=1}^{i} w(v_{j-1}, v_j)}{w(G)}$.
Notice that f is an embedding s.t. $dist(f) = 1$ and embeds into $[0,1)$ and is reversible.
Therefore, for a given partition of $[0,1)$ with $\Delta' = \frac{\Delta}{w(G)}$ denoting $(S_1, \dots, S_{\frac{1}{\Delta'}}) \sim P$, the new partition will be $C_1, \dots, C_{\frac{1}{\Delta'}} = f^{-1}(S_1), \dots, f^{-1}(S_{\frac{1}{\Delta'}})$.
Notice that it is a partition due to it being a partition before and usage of a reversable function and $Diam(C_i) \leq \Delta$.
Let $x \in V$, therefore $\exists i, x \in C_i$

$$\begin{aligned}\mathbb{P}\left(B(x, \eta \cdot \Delta) \not\subseteq P_G(x)\right) &\leq \mathbb{P}\left(f(B(x, \eta \cdot \Delta)) \not\subseteq f(C_i)\right) = \mathbb{P}\left(f(B(x, \eta \cdot \Delta)) \not\subseteq S_i\right) \\ &= \mathbb{P}\left(B(f(x), \eta \cdot \Delta \cdot \frac{1}{w(G)}) \not\subseteq S_i\right) \leq O(1) \cdot \eta\end{aligned}$$

So we proved that $\forall x \in V, \forall 0 < \eta < \frac{1}{16}, \mathbb{P}\left(B(x, \eta \cdot \Delta)\right) \leq 2\eta = O(1) \cdot \eta$.
Therefore, we proved that there exists a Δ- bounded probabilistic partition P with padding parameter $O(1)$ for $X = G$, as required.

2. (a) We will first prove that we can embed k-HST (for $k > 2$) to a tree without Steiner points non-contracting with distortion $\alpha = \frac{2k}{k-2} = O(1)$ by induction of the height.
Base case: height(U)=0, this means the k-HST is a lead and it can embed to itself isometrically.
Induction: Assume that the claim is true for any k-HST with $height(U') \leq height(U)$ and we will prove for U.
Define $U_1, \dots, U_k$ the sub trees of U created by its children.
via the assumption, $\forall 1 \leq i \leq k \exists f_i : U_i \to T_i$ s.t. $dist(f) \leq \alpha$ and non-contracting.
Let $\forall 1 \leq i \leq k, t_i \in V_{T_i}$.
Define $T = \left\langle \bigcup_{i=1}^{k} V_{T_i}, \bigcup_{i=1}^{k} E_{T_i} \cup \{\{t_1, t_i\} \mid 2 \leq i \leq k\} \right\rangle$.
Let $f : U \to T$ be defined in the following way: $f(u) = \sum_{i=1}^{k} \mathbb{1}_{u \in U_i} f_i(u)$.
Define $d_T(e) = \sum_{i=1}^{k} \mathbb{1}_{e \in E_{T_i}} d_{T_i}(e) + \sum_{i=2}^{k} \mathbb{1}_{e = \{t_1, t_i\}} \cdot \Delta(U)$.
Notice that $\forall 1 \leq i \leq k, \forall x \neq y \in U_i$, $d_T(x,y) = d_{T_i}(x,y)$ and therefore non-contracting by default.
Furthermore, notice that $\forall 1 \leq i \neq j \leq k, \forall x \in U_i, \forall y \in U_j$, $d_T(x,y) \geq d_T(t_i, t_j) = \Delta(U) = d_{k-HST}(x,y)$ and therefore non-contracting in this case either.
Notice that $\forall 1 \leq i \leq k, \forall u_1, u_2 \in U_i, d_T(u_1, u_2) = d_{T_i}(u_1, u_2)$ and therefore $dist_f(u_1, u_2) = dist_{f_i}(u_1, u_2) \leq \alpha$.
Notice that $\forall 1 \leq i \leq k, \forall v \in U_i$ it holds that $d_T(v, t_i) \leq \alpha \cdot \Delta(U_i)$ Notice that $\forall 2 \leq i \leq k, \forall u \in U_1, \forall v \in U_i$

$$\begin{aligned}d_T(v,u) &= d_T(v,t_i) + d_T(t_i,t_1) + d_T(t_1,u) \leq \alpha \cdot \Delta(U_i) + \Delta(U) + \alpha \cdot \Delta(U_i) \leq \Delta(U) \cdot \left(\frac{2\alpha}{k} + 1\right) \\ &= d_{k-HST}(v,u) \cdot \frac{2\alpha + k}{k} = d_{k-HST}(v,u) \cdot \frac{2\frac{2k}{k-2} + k}{k} = d_{k-HST}(v,u) \cdot \frac{\frac{4k+k^2-2k}{k-2}}{k} \leq d_{k-HST}(v,u) \cdot \frac{k}{k-2} \\ &\leq d_{k-HST}(v,u) \cdot \alpha\end{aligned}$$

and therefore $dist_f(u,v) \leq \alpha$.
Notice that $\forall 2 \leq i \neq j \leq k, \forall u \in U_j, \forall v \in U_i$

$$\begin{aligned}d_T(v,u) &= d_T(v,t_i) + d_T(t_i,t_1) + d_T(t_1,t_j) + d_T(t_j,u) \leq \alpha \cdot \Delta(U_i) + \Delta(U) + \Delta(U) + \alpha \cdot \Delta(U_i) \leq \Delta(U) \cdot \left(\frac{2\alpha}{k} + 2\right) \\ &= d_{k-HST}(v,u) \cdot \frac{2\alpha + 2k}{k} = d_{k-HST}(v,u) \cdot \frac{2\frac{2k}{k-2} + 2k}{k} = d_{k-HST}(v,u) \cdot \frac{\frac{4k+2k^2-4k}{k-2}}{k} \\ &= d_{k-HST}(v,u) \cdot \frac{2k}{k-2} = d_{k-HST}(v,u) \cdot \alpha\end{aligned}$$

and therefore $dist_f(u,v) \leq \alpha$.
Therefore, we proved that in all cases $dist_f(u,v) \leq \alpha$ and therefore $dist(f) \leq \alpha$ and that f is non-contracting.
Now we proved in class (Theorem 11.7) that exists a probabilistic embedding into $k-HST$ trees with expected distortion $O(k \cdot \log(n))$.
Using this claim with $k = 3$, we get that exists a probabilistic embedding into $k-HST$ non contracting trees with expected distortion $O(\log(n))$.

Denote the k-HST distribution of embeddings with D_g.

let $g \sim D_g$ s.t. $g : G \to k$, therefore, due to the what we proved $\exists h : k \to T$ non contracting with distortion at most $\alpha = O(1)$.

Define the following distribution D_T, to sample $g \sim D_g$ and apply $h(g)$. we know that both functions are non contracting and therefore $f = h(g)$ is not contracting.

Therefore, we get that exists a probabilistic embedding into trees D_T with expected distortion $O(\log(n))$ because f only scaled the distortion up to a constant factor from what g previously scaled.

Sampling $g \sim D_G$ and f be the corresponding $f = h(g)$ s.t. $g : G \to k$ and $h : k \to T$, we know that $|V_G| = |Vert(k)| = |V_T|$, therefore there is an isomorphisim between $f' : V_T \to V_g$.

Therefore, T is a spanning tree of G (maybe with different weights).

Define the following distribution D_G, to sample $f \sim D_f$ s.t. $f : V \to T$ and the item chosen from the distribution will be $a(x) = (f'(f(x)), d_G)$. Notice that a is non-contracting because it is a spanning tree of G and $\forall \{x, y\} = e \in E_G$, $d_G(x, y) \leq d_T(f'^{-1}(x), f'^{-1}(y))$ because T is non-contracting of G. Therefore because both are non contracting trees and one is with weights less than the other, therefore $\forall x \neq y, dist_a(x, y) \leq dist_f(x, y)$.

Therefore, we get that exists a probabilistic embedding into spanning trees of G D_G with expected distortion $O(\log(n))$ (because it is better than D_T that has $O(\log(n))$ distortion), as required.

(b) Let $T = \langle V, E_T \rangle$ be an MST and $f \sim P$ where P is a probabilistic embedding into spanning trees of G with expected distortion $O(\log(n))$.

Let $\{x, y\} = e \in E_T$, notice that $\mathbb{E}\left[d_f(x, y)\right] \leq d_G(x, y) \cdot O(\log(n))$.

Notice that due to the fact that G satisfies the triangle inequality, $d_G(x, y) = w(x, y)$.

Therefore, $\mathbb{E}\left[d_f(x, y)\right] \leq d_G(x, y) \cdot O(\log(n)) = w(x, y) \cdot O(\log(n)) = w(e) \cdot O(\log(n))$.

Therefore, $\sum_{e \in E_T} \mathbb{E}\left[d_f(x, y)\right] \leq \sum_{e \in E_T} w(e) \cdot O(\log(n)) = O(\log(n)) \cdot \sum_{e \in E_T} w(e) = O(\log(n)) \cdot w(T)$.

Define the spanning Tree created by f to be $T' = \langle V, E_{T'} \rangle$.

Assume by contradiction that $\exists e' \in E_T$ s.t. $\forall \{x, y\} = e \in E_T$, the path between x, y in T' doesn't contain e'. Therefore, given x', y', there is a path in E_T, denote it as $\{x_1, y_1\} = e_1, \ldots, \{x_k, y_k\} = e_k$ and use the assumption to create a path between x', y' that doesn't contain e' (by replacing one edge at a time inductively). Therefore, we can remove e' and still have a spanning graph in contradiction to the fact that T' is a tree.

Therefore, the sum $\sum_{e \in E_T} \mathbb{E}\left[d_f(x, y)\right]$ contains each edge of T' at least once. Therefore $w(T') \leq \sum_{e \in E_T} d_f(x, y)$, therefore $\mathbb{E}\left[w(T')\right] \leq \sum_{e \in E_T} \mathbb{E}\left[d_f(x, y)\right] \leq O(\log(n)) \cdot w(T)$.

Therefore, $\boxed{\mathbb{E}\left[w(T')\right] = w(MST) \cdot O(\log(n))}$, as required.

Bonus 1: We will prove first a helper lemma, let $y_1, \ldots, y_l$ s.t. $B_1(y_1, r), \ldots, B_l(y_l, r) \subseteq B(x, R)$ disjoint balls and let t be the number of balls of radius r to cover $B(x, R)$, we will prove that $l \leq t$.

Due to the assumption, $\exists x_1, \ldots, x_t \in X$ s.t. $B_1(y_1, r), \ldots, B_l(y_l, r) \subseteq B(x, R) \subseteq \bigcup_{i=1}^{t} B(x_i, r)$. Therefore, $\forall 1 \leq i \leq l, \exists 1 \leq j \leq t$ s.t. $y_i \in B(x_j, r)$. Therefore, due to symmetry, $\forall 1 \leq i \leq l, \exists 1 \leq j \leq t$ s.t. $x_j \in B(y_i, r)$.

Assume by contradiction that $l > t$, therefore, due to the pigeon principle, $\exists x_j$ which is contained in $B(y_{i_1}, r)$ and $B(y_{i_2}, r)$. This is a contradiction to the fact that the balls are disjoint.

Notice that to cover a ball of radius 64Δ with balls of radius 32Δ we need at most $2^{\dim(X)}$ balls. We can repeat this process iteratively and get that to over a ball of radius 64Δ with balls of radius $\frac{\Delta}{64}$ we need at most $\left(2^{dim(X)}\right)^{12} = 2^{12\dim(X)}$ balls.

We will go over the proof 12.8 again as stated in the lecture. We choose the same clusters. We define $\forall i, \chi_i = 2^{12 \cdot \dim(X)}$. Notice that χ_i is monotonic and $\chi_i \geq 2$ because $\dim(X) \geq 1$, therefore, the defined χ_i satisfy the requirements. If we prove $\sum_{j \in T} \frac{1}{\chi_j} \leq 1$, we can use lemma 12.7 and get that there exists a Δ-probabilistic partition with padding parameter $O\left(128 \log(\chi_i)\right) = O(\log(\chi_i)) = O(\dim(X))$.

Therefore, to prove the claim, we just need to show that $\sum_{j \in T} \frac{1}{\chi_j} \leq 1$.

Notice that $\sum_{j \in T} \frac{1}{\chi_j} = \frac{|T|}{\chi_1}$.

Now notice that in the original lemma of 12.8, we proved that $\forall j, B\left(v_j, \frac{\Delta}{64}\right) \subseteq B\left(x, 63\Delta\right) \subseteq B\left(x, 64\Delta\right)$ and that all the balls are disjoint.

Therefore, due to what we proved before, we know that to cover a ball of radius 64Δ with balls of radius $\frac{\Delta}{64}$ we need at most $2^{12\dim(X)}$ balls.

Due to the other lemma that we proved about disjoint balls, we get that $|T| \leq 2^{12\dim(X)} = \chi_1$.

Therefore, $\sum_{j \in T} \frac{1}{\chi_j} = \frac{|T|}{\chi_1} \leq \frac{\chi_1}{\chi_1} = 1$.

Therefore, we proved that there exists a Δ-probabilistic partition with padding parameter $O\left(128 \log(\chi_i)\right) = O(\log(\chi_i)) = O(\dim(X))$ as required.

Bonus 3: We proved that there exists a distribution P of probabilistic embeddings into ultrametrics s.t. $\forall x, y \in X$ $\mathbb{E}\left[dist_f(x, y)\right] = O(\log(n))$ and non-contractive. therefore, $\exists c > 0$, s.t. $\forall x, y \in X$, $\mathbb{E}\left[dist_f(x, y)\right] \leq c \cdot \log(n)$. Let

$X_1,\dots,X_k \sim P$. Let $x,y\in X$. Define $X_{i,x,y}=\frac{d_{X_i}(x,y)}{d_X(x,y)}$, notice that $|X_{i,x,y}|\in[1,c\cdot\log(n)]$. Let $\varepsilon>0$ Notice that they are independent and therefore from the Hoeffding's inequality

$$\mathbb{P}\left(\sum_{i=1}^{k}X_{i,x,y}-\mathbb{E}\left[\sum_{i=1}^{k}X_{i,x,y}\right]\geq\varepsilon\cdot\mathbb{E}\left[\sum_{i=1}^{k}X_{i,x,y}\right]\right)\leq\exp\left[\frac{-2\left(\varepsilon\cdot\mathbb{E}\left[\sum_{i=1}^{k}X_{i,x,y}\right]\right)^2}{k(c\log(n)-1)^2}\right]\leq\exp\left\{\frac{-2\left(\varepsilon\cdot\sum_{i=1}^{k}\mathbb{E}\left[X_{i,x,y}\right]\right)^2}{k(c\log(n)-1)^2}\right\}$$

$$\leq\exp\left\{\frac{-2\left(\varepsilon\cdot\sum_{i=1}^{k}\mathbb{E}\left[X_{i,x,y}\right]\right)^2}{k(c\log(n)-1)^2}\right\}=\exp\left\{\frac{-2\varepsilon^2\cdot k\left(\mathbb{E}\left[X_{1,x,y}\right]\right)^2}{(c\log(n)-1)^2}\right\}$$

Notice that

$$\exp\left\{\frac{-2\varepsilon^2\cdot k\left(\mathbb{E}\left[X_{1,x,y}\right]\right)^2}{(c\log(n)-1)^2}\right\}<\frac{1}{\binom{n}{2}}\iff\frac{-2\varepsilon^2\cdot k\left(\mathbb{E}\left[X_{1,x,y}\right]\right)^2}{(c\log(n)-1)^2}\geq\ln\left(\frac{1}{\binom{n}{2}}\right)\iff\frac{\varepsilon^2\cdot k\left(\mathbb{E}\left[X_{1,x,y}\right]\right)^2}{(c\log(n)-1)^2}\leq-\ln\left(\frac{1}{\binom{n}{2}}\right)$$

$$\iff k\leq\ln\binom{n}{2}\cdot\frac{(c\log(n)-1)^2}{\varepsilon^2\cdot\mathbb{E}\left[X_{1,x,y}\right]^2}\Leftarrow k\leq\ln\binom{n}{2}\cdot\frac{(c\log(n)-1)^2}{\varepsilon^2}=\Theta(\ln(n)^3)$$

So we choose $k=\Omega(\ln(n)^3)$ (note if we knew that $\log(n)$ is tight for each pair we could have gotten $\Omega(\log(n))$ instead). Therefore we get that,

$$\mathbb{P}\left(\exists x,y,\frac{1}{k}\cdot\sum_{i=1}^{k}X_{i,x,y}-\frac{1}{k}\cdot\mathbb{E}\left[\sum_{i=1}^{k}X_{i,x,y}\right]\geq\varepsilon\cdot\frac{1}{k}\cdot\mathbb{E}\left[\sum_{i=1}^{k}X_{i,x,y}\right]\right)=\mathbb{P}\left(\exists x,y,\sum_{i=1}^{k}X_{i,x,y}-\mathbb{E}\left[\sum_{i=1}^{k}X_{i,x,y}\right]\geq\varepsilon\cdot\mathbb{E}\left[\sum_{i=1}^{k}X_{i,x,y}\right]\right)$$

$$\leq\sum_{x\neq y}\mathbb{P}\left(\sum_{i=1}^{k}X_{i,x,y}-\mathbb{E}\left[\sum_{i=1}^{k}X_{i,x,y}\right]\geq\varepsilon\cdot\mathbb{E}\left[\sum_{i=1}^{k}X_{i,x,y}\right]\right)\leq\frac{1}{\binom{n}{2}}\cdot\mathbb{P}\left(\sum_{i=1}^{k}X_{i,x,y}-\mathbb{E}\left[\sum_{i=1}^{k}X_{i,x,y}\right]\geq\varepsilon\cdot\mathbb{E}\left[\sum_{i=1}^{k}X_{i,x,y}\right]\right)<1$$

Therefore, $\exists X_1,\dots,X_k$ s.t. $\forall x\neq y$, $\frac{1}{k}\cdot\sum_{i=1}^{k}X_{i,x,y}-\frac{1}{k}\cdot\mathbb{E}\left[\sum_{i=1}^{k}X_{i,x,y}\right]<\varepsilon\cdot\frac{1}{k}\cdot\mathbb{E}\left[\sum_{i=1}^{k}X_{i,x,y}\right]$.

Therefore $\forall x\neq y, 1\leq\frac{1}{k}\cdot\sum_{i=1}^{k}X_{i,x,y}\leq(1+\varepsilon)\cdot\frac{1}{k}\cdot\mathbb{E}\left[\sum_{i=1}^{k}X_{i,x,y}\right]\leq(1+\varepsilon)\cdot c\cdot\log(n)=O(\log(n))$.

Choose the distribution $X\sim Uni(X_1,\dots,X_k)$. notice that $\forall x\neq y,\mathbb{E}\left[X_{x,y}\right]=\frac{1}{k}\cdot\sum_{i=1}^{k}X_{i,x,y}=O(\log(n))$.

So we found a distribution of probabilistic embeddings into ultrametrics of size $k=\Omega(\log^3(n))$ s.t. $\forall x,y\in X\ \mathbb{E}\left[dist_f(x,y)\right]=O(\log(n))$ and non-contractive because each X_i is non-contractive. Therefore we proved that X probabilistically embeds into ultrametrics with distortion $O(\log(n))$, such that the distribution is over (a support of) $poly(n)=\Theta(\ln^3(n))$ ultrametrics.

Notice that this method can be applied to any space X.