CSE 632 (Fall 2019): Analysis of Algorithms II : Randomized Algorithms

Lecture 25-26 (11/20/2019, 11/22/2019): Embedding Into Trees

Lecturer: Shi Li *Scriber: Zhun Liu, Shi Li*

1 Embedding of Metrics

Definition 1. *Given a set X, and function $d : X \times X \to \mathbb{R}_{\geq 0}$, (X, d) is called a metric if*

- *$d(x, x) = 0$ for every $x \in X$,*
- *(symmetry) $d(x, y) = d(y, x)$ for every $x, y \in X$, and*
- *(triangle inequalities) $d(x, y) + d(y, z) \geq d(x, z)$ for every $x, y, z \in X$.*

For example, in the k-dimensional Euclidean metric space (X, d), we have $X = \mathbb{R}^k$, and for every $x, y \in X$, $d(x, y) = \|x - y\|_2$.

Definition 2. *Given two metrics (X, d_X) and (Y, d_Y), an embedding of (X, d_X) to (Y, d_Y) is a function $f : X \to Y$.*

In the lecture, we assume the embeddings are *non-contractive*:

Definition 3. *A embedding f of (X, d_X) to (Y, d_Y) is said to be non-contractive if for every $x, y \in X$, we have $d_Y(f(x), f(y)) \geq d(x, y)$.*

Definition 4. *For a non-contractive embedding f of (X, d_X) to (Y, d_Y), we say f has distortion (at most) D, for some $D \geq 1$, if for every $x, y \in D$, we have*

$$d_Y(f(x), f(y)) \leq D \cdot d_X(x, y).$$

We use function f to embed (X, d_x) to a host metric space: $(Y, d_Y)$is a function f : $X \to Y s.t.$ $d_x(x, y) \leq d_Y(f(x), f(y)) \leq Dd_x(x, y)$ D is called the distortion of the embedding.

Recall that the following embedding result is implied by the Johnson-Lindenstrauss dimension reduction lemma. Assume (X, d) is a metric on n points in the Euclidean space. If $k = O(\frac{\log n}{\epsilon^2})$ is big enough, then there is an embedding of (X, d) to $(\mathbb{R}^k, L_2)$ with distortion $1 + \epsilon$, where $(\mathbb{R}^k, L_2)$ is the k-dimensional Euclidean space.

2 Embedding of Metrics to Distribution of Tree Metric

Definition 5. *A metric (X, d) is said to be a tree-metric, if we can construct a weighted tree $T = (V, E, w)$ such that $X \subseteq V$ such that $d(u, v)$ for every $u, v \in X$ is the length of the unique path between u and v in T.*

Given a tree T, we use d_T to denote the shorted-path distance function between points in T.

Many NP-hard problems on general metrics can be solved efficiently since trees are often friendly to the dynamic programming technique. Then it is appealing if to embed any metric into a tree metric with a small distortion. Unfortunately, the following metric shows that this is impossible: Consider the metric C_n induced by the cycle on n-points. One can indeed show that any embedding of the metric into a tree metric has distortion at least $\Omega(n)$. It is tricky to prove the statement, but we can consider a restricted way to construct the tree-metric: We remove any edge from the cycle to obtain a path and let the tree metric to be induced by the path. Then the distortion of the metric is $n - 1$ since the distance between the two end points of the removed edge changes from 1 to $n - 1$.

However, we can show that there is a small-distortion of metrics into tree-metrics, if we relax the requirement slightly. Consider the cycle example. If we remove an edge *uniform at random*, then the *expected distortion* of any pair of points will be at most 2. We just focus on two points u, v that are adjacent in C_n; one can easily show that it suffices to consider such pairs. Then with probability $1/n$, the edge between u and v will be removed and the distance between u and v becomes $n - 1$

in the case. With the remaining $1 - 1/n$ probability, the distance remains 1. Thus the expected distortion is $\frac{1}{n} \times (n-1) + (1 - \frac{1}{n}) \times 1 \leq 2$.

This motivates us to embed any metric into *a distribution of tree metrics*. Building on the work of Bartal, Fakcharoenphol, Rao, and Talwar shows that any n-point metric (X, d) can be embedded into a distribution of non-contractive tree metrics with expected distortion $O(\log n)$.

Theorem 6 (Fakcharoenphol, RaoTalwar)**.** *Let (X, d) be a n-point metric. Then there is a distribution π over trees T, such that*

- *for every T in the support of π, and every $u, v \in X$, we have $d_T(x, y) \geq d(x, y)$, and*
- *for every $x, y \in X$, we have $\mathbb{E}_{T \sim \pi}\, d_T(x, y) \leq O(\log n) d(x, y)$.*

We are not going to prove the theorem. Instead we show a building block in the proof of the theorem, which we call the *padded partitioning* of a metric. With this building block, one can easily obtain an embedding with a slightly distortion of $O(\log^2 n)$. The $O(\log n)$ distortion can be obtained by a more careful use of the padded partitioning technique.

3 Padded Partitioning

In this section, we fix the metric (X, d) with $n = |X|$. For every non-empty set $S \subseteq X$, we define the *diameter* of S to be $\mathsf{diam}(S) := \max_{u,v \in S} d(u, v)$.

Definition 7. *Given a metric (X, d), two parameters $\Delta > 0$ and $\alpha \geq 1$, we say a random partition $\mathcal{P}$ of X into subsets is said to be an (Δ, α)-padded partitioning if*

- *with probability 1, for every $S \in \mathcal{P}$, we have $\mathsf{diam}(S) \leq \Delta$, and*
- *for every $u, v \in X$, we have*

$$\Pr_{\mathcal{P}}[x, y \text{ are separated in } \mathcal{P}] \leq \alpha \cdot \frac{d(u, v)}{\Delta}.$$

In this section, we show that for every Δ, there exists a $(\Delta, O(\log n))$-padded partition of (X, d). For simplicity, we can assume $\Delta = 1$ by scaling the distances.

3.1 Construction of $(1, O(\log n))$-Padded Partition

For every $X' \subseteq X, w \in X$ and $r \in \mathbb{R}_{\geq 0}$, we define $\mathsf{ball}_{X'}(w, r) := \{v \in X' : d(w, v) \leq r\}$ to be the ball of points in X' with distance at most r to w. The random partition $\mathcal{P}$ is produced using the following algorithm: It is easy to see that $\mathcal{P}$ is always a partition of X. Moreover, every set S has

1: let r be a number selected uniformly at random from $[1/4, 1/2]$
2: let π be a random permutation over X
3: $X' \leftarrow X, \mathcal{P} \leftarrow \emptyset$
4: **for** every $w \in X$, in the order of π **do**
5:     **if** $\mathsf{ball}_{X'}(w, r) \neq \emptyset$ **then**
6:         $\mathcal{P} \leftarrow \mathcal{P} \cup \{\mathsf{ball}_{X'}(w, r)\}, X' \leftarrow X' \setminus \mathsf{ball}_{X'}(w, r)$
7: **return** $\mathcal{P}$

diameter at most $2r \leq 1$ since it is a ball of radius r. It suffices to prove the second property in the definition of padded partition.

Lemma 8. *For every set $X' \subseteq X$, every three points $u, v \in X'$ and $w \in X$, we have*

$$\Pr[u \in \mathsf{ball}_{X'}(w, r), v \notin \mathsf{ball}_{X'}(w, r)] \leq 4d(u, v).$$

Proof. The event happens only if $d(u, w) \leq r < d(v, w)$. If $d(u, w) < d(v, w)$, the probability that the event happens is at most $\frac{d(v,w) - d(u,w)}{1/4} \leq 4d(u, v)$ by triangle inequalities, since r is uniformly chosen from an interval of length $1/4$. □

Definition 9. *Focus on an execution of the algorithm. We say w separates u from v if at the beginning of the iteration of Loop 4 for w, we have $u, v \in X', u \in \mathsf{ball}_{X'}(w, r)$ and $v \notin \mathsf{ball}_{X'}(w, r)$.*

Fix a u, we sort the vertices in X using the non-decreasing order of distances to u. For every $w \in X$ we define $\mathsf{rank}_u(w)$ denote the rank of w in the ordering. That is, if $\mathsf{rank}_u(w) = i$, then w is the i-th nearest point in X to u. The following lemma can be proven:

Lemma 10. $\Pr[w$ *separates* u *from* $v] \leq \frac{4d(u,v)}{\mathsf{rank}_w(u)}$.

Proof. Suppose there is a $w' \in X$ with $\mathsf{rank}_u(w') < \mathsf{rank}_u(w)$ and w' appears before w in the ordering π. Then, we claim that w can not separate u from v. Assume otherwise. Then u and v must both be in X' at the beginning of the iteration for w'. If $d(w', u) \leq r$, then u will be removed from X' in the iteration and thus u will not be in X' at the iteration for w. Thus, we must have $d(w', u) > r$. However, in this case, we also have that w can not separate u from v.

Thus, for w to be able to separate u from v, we must have that for every w' with $\mathsf{rank}_u(w') < \mathsf{rank}_u(w)$, w' must be after w in the permutation π. The probability that this event happens is exactly $\frac{1}{\mathsf{rank}_u(w)}$. Under the condition that this happens, the probability that w separates u from v is at most $4d(u, v)$ by Lemma 8, since r and π are independently sampled. Thus, overall, the probability that w separates u from v is at most $\frac{4d(u,v)}{\mathsf{rank}_w(u)}$. □

Thus, the probability that some $w \in X$ separates u from v is at most $\sum_{w \in X} \frac{4d(u,v)}{\mathsf{rank}_w(u)} = 4d(u, v) \sum_{i=1}^{n} \frac{1}{n} = O(\log n) d(u, v)$. The same can be prove for the probability that some $w \in X$ separates v from u. Notice that if u and v are separated in $\mathcal{P}$, then either some w separated u from v, or some w separated v from u. Thus, the probability that u and v are separated is at most $O(\log n) d(u, v)$. Thus, we proved the second property in the definition of padded partition.

In the homework 5, we shall see an example of applying the FRT tree embedding result to reduce an instance of some problem on general metrics to some instance on tree metrics.