601.435 / 601.635 Approximation Algorithms **Spring 2026**
Homework #4 **Due:** April 2, 2026, 9:00am

Reminder: you may work in groups of up to three people, but must write up solutions entirely on your own. Collaboration is limited to discussing the problems – you may not look at, compare, reuse, etc. any text from anyone else in the class. Please include your list of collaborators on the first page of your submission. Many of these problems have solutions which can be found on the internet – please don't look. You can of course use the internet (including the links provided on the course webpage) as a learning tool, but don't go looking for solutions.

Please include proofs with all of your answers, unless stated otherwise.

1 Tree embeddings with total cost (Exercise 8.12) (33 points)

We saw in class that given a metric space (V, d), the FRT tree embedding algorithm is a randomized algorithm that generates a tree metric (V', T) such that 1) $V' \supseteq V$, 2) $d(u, v) \leq d_T(u, v)$ for all $u, v \in V$ and T in the support of the algorithm, and 3) $\mathbf{E}[d_T(u, v)] \leq O(\log n) \cdot d(u, v)$ for all $u, v \in V$ (where the expectation is taken over the randomized choice of T). Note that this is for the distribution over trees generated by FRT – any individual tree might distort some distance quite badly. But sometimes we will want a bound on a single tree, so we might ask for something else: a single tree in which the *total cost* is small.

Suppose that we are given a metric space (V, d) and costs $c : \binom{V}{2} \to R^+$ (where costs have nothing to do with distances, and $\binom{V}{2}$ denotes all unordered pairs of nodes in V). Give a polynomial-time randomized algorithm that returns a tree metric (V', T) for V (so V is the leaves of T) such that:

1) $d(u, v) \leq d_T(u, v)$ for all $u, v \in V$ with probability 1, and
2) $\sum_{u,v \in V} c(u, v) d_T(u, v) \leq O(\log n) \sum_{u,v \in V} c(u, v) d(u, v)$ with very high probability (at least $1 - \frac{1}{2^n}$).

Hint: Try to *use* FRT rather than *modifying* it.

Solution: Suppose that we run FRT on (V, d) to get a tree metric (V', T). Note that we know from the FRT guarantee that (V', T) is a tree metric for V and that $d(u, v) \leq d_T(u, v)$ for all $u, v \in V$. The *expected* total cost is (by linearity of expectations)

$$\begin{aligned}\mathbf{E}\left[\sum_{u,v \in V} c(u, v) d_T(u, v)\right] &= \sum_{u,v \in V} c(u, v) \mathbf{E}[d_T(u, v)] \leq \sum_{u,v \in V} c(u, v) O(\log n) d(u, v) \\ &= O(\log n) \sum_{u,v \in V} c(u, v) d(u, v).\end{aligned}$$

So FRT works well in expectation. By losing an extra factor of 2 (which will be hidden in the $O(\log n)$ anyway), we know from Markov's inequality that the probability the average cost is more than twice this expectation is at most $1/2$. Thus if we simply run FRT n times and return the tree with lowest average cost of those n trials, the probability that the average cost is more

than $2 \cdot O(\log n) \sum_{u,v \in V} c(u,v)d(u,v) = O(\log n) \sum_{u,v \in V} c(u,v)d(u,v)$ is at most $1/2^n$. Hence the probability that we success is at least $1 - \frac{1}{2^n}$.

2 Capacitated Dial-a-Ride (Exercise 8.11) (67 points)

In the Capacitated Dial-a-Ride problem, we are given a metric (V, d), a single vehicle with a given integer capacity $C > 0$ located at a given node $r \in V$, and k source-sink pairs $(s_1, t_1), \ldots, (s_k, t_k)$. At each source s_i there is an item which must be delivered to the sink t_i by the vehicle. The vehicle can carry at most C items at a time. The goal is to find the shortest tour that starts and ends at the r, delivers each item from its source to its destination without exceeding the vehicle capacity, and returns to r. Note that such a tour may visit a node of V multiple times. We assume that the vehicle is allowed to temporarily leave items at any node in V.

(a) (34 points) Suppose that (V, d) is actually a tree metric (V, T) (so T is a tree with vertex set V, not just with leaves V). Give an $O(1)$-approximation algorithm for this case.

Hint: for each $i \in [k]$ there is a unique path $p(i)$ in T from s_i to t_i. First show that without loss of generality, you can restrict your attention to the subtree induced by r and the source-sink pairs. For each edge e which remains, let $\ell(e)$ denote the number of these paths which use edge e. Prove that OPT needs to traverse e at least $\max(1, \ell(e)/C)$ times. Design an algorithm which traverses each edge at most $O(1) \cdot \max(1, \ell(e)/C)$ times.

Solution: Let T' be the Steiner tree induced by the $p(i)$ paths. Then T' consists of edges that are in at least one $p(i)$, together with possibly some paths from the root to connect the paths into one tree. Note that every edge $e \in T'$ must be traversed at least once, either to deliver items (if $e \in p(i)$ for some i) or to move between paths.

Let $e \in T'$. Then by definition, $\ell(e)$ items need to cross e. Since our vehicle has capacity C, this takes at least $\ell(e)/C$ traversals. Hence $OPT \geq \sum_{e \in T'} \max\{1, \ell(e)/C\}$.

Our algorithm will essentially perform two DFS traversals of T'. For each $i \in [k]$, let m_i denote the LCA of s_i and t_i. In the first traversal, we bring every item from s_i to m_i. We do this in the straightforward way: for each v we see on the DFS, when all of the calls to its children have returned, by induction we know that any item i such that v is on the $s_i - m_i$ path is at the child of v whose subtree contains s_i. So we visit each child of v and bring to v the items which still need to move up to the tree to m_i (possibly taking multiple trips if there are more than C items). Clearly every edge in T' is traversed $\max\{2\ell(e)/C), 1\}$ times.

In the second phase, we do another DFS to bring every item from m_i to t_i. By the same reasoning, each edge $e \in T'$ is also traversed $\max\{2\ell(e)/C, 1\}$ times. Hence the total length of the algorithm is

$$\sum_{e \in T'} \max\{4\ell(e)/C, 2\} = O\left(\sum_{e \in T'} \max\{\ell(e)/C, 1\}\right) = O(OPT),$$

so it is an $O(1)$-approximation.

(b) (33 points) Give a randomized $O(\log n)$-approximation for the general Capacitated Dial-a-Ride problem, where $n = |V|$. Please do this formally – it's not enough to say "FRT embeds with distortion $O(\log n)$, so we lose an $O(\log n)$ factor". Hint: you might want to use the steiner point removal result discussed in the lecture notes and (briefly) discussed in class.

Solution: We can use FRT to embed the input metric (V, d) into a tree metric (V', T'). We showed in class how to remove steiner points from such a tree embedding while losing only another factor of 4 in the distortion, so we do this to get a tree metric (V, T) in which $d(u, v) \leq d_T(u, v)$ and $\mathbf{E}[d_T(u, v)] \leq O(\log n)d(u, v)$ for all $u, v \in V$. So we perform such an embedding, and then use the algorithm from part (a) to get a tour P which is close to optimal for T.

First note that since T is a tree on V (not a superset V') and we start with a metric space, P is actually a solution to the problem in the original metric (V, d). That is, we can follow P, picking up and dropping off items according to it, in the original metric rather than the tree and in the end every item does get from s_i to t_i and we obey the vehicle capacity. So it simply remains to bound the length of this tour in the original metric. Consider the optimal tour P' for (V, d). Note that while P is a random variable, P' is a fixed tour. Then

$$\mathbf{E}[ALG] = \mathbf{E}\left[\sum_{e \in P} d(e)\right] \leq \mathbf{E}\left[\sum_{e \in P} d_T(e)\right] \leq \mathbf{E}\left[O(1) \cdot \sum_{e \in P'} d_T(e)\right] \leq O(1) \cdot \sum_{e \in P'} \mathbf{E}[d_T(e)]$$
$$\leq O(1) \cdot \sum_{e \in P'} O(\log n)d(e) = O(\log n) \sum_{e \in P'} d(e) = O(\log n) \cdot OPT.$$

Hence this algorithm is an $O(\log n)$-approximation.