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.

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 $(s_1, t_1), \dots, (s_k, t_k)$. 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.

(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.