A Deterministic Algorithm for the COST-DISTANCE Problem

Chandra Chekuri [*] Sanjeev Khanna [†] Joseph (Seffi) Naor[‡]

October 12, 2000

Abstract

The COST-DISTANCE network design problem is the following. We are given an undirected graph $G = (V, E)$, a designated *root* vertex $r \in V$, and a set of terminals $S \subset V$. We are also given two non-negative real valued functions defined on E, namely, a cost function c and a length function ℓ, and a non-negative weight function w on the set S. The goal is to find a tree T that connects the terminals in S to the root r and minimizes $\sum_{e \in T} c(e) + \sum_{t \in S} w(t)\ell_T(r,t)$, where $\ell_T(r,t)$ is the length of the path in T from t to r.

We give a *deterministic* $O(\log k)$ approximation algorithm for the COST-DISTANCE network design problem, in a sense derandomizing the algorithm given in [4]. Our algorithm is based on a natural linear programming relaxation of the problem and in the process we show that its integrality gap is $O(\log k)$.

Introduction: We study the COST-DISTANCE network design problem, recently defined by Meyerson, Munagala, and Plotkin [4], as a common generalization of several useful problems in network design and facility location (see [4]). In this problem, we are given an undirected graph $G = (V, E)$ with a designated *root* vertex r and a set of terminals S. We are also given two non-negative real valued functions defined on E, namely, a cost function c and a length function ℓ, and a non-negative weight function w on the set S. The goal is to construct a tree T that connects the terminals in S to the root r and minimizes

$$\sum_{e \in T} c(e) + \sum_{t \in S} w(t)\ell_T(r,t),$$

where $\ell_T(r,t)$ is the length of the path in T from t to r. This problem can also be seen as a Lagrangian relaxation of the budgeted versions of network design problems studied earlier in [3, 1, 2]. Building on the basic framework developed in [3], Meyerson *et al* [4] give a *randomized* approximation algorithm with an expected ratio of $O(\log k)$, where $k = |S|$. An interesting question raised in [4] is whether this ratio can be achieved deterministically. In this note we answer the question in the affirmative. In the process we also show that a natural linear programming relaxation has an integrality gap of $O(\log k)$.

Matching Graph: Define a complete graph H on $S \cup \{r\}$ with a weight function b on its edges; $b(u,v)$ is assigned the weight of the shortest path in G connecting u and v with respect to the metric $m(e) = c(e) + \min\{w(u), w(v)\}\ell(e)$. We assume that $w(r) = \infty$ and $w(v) = 0$ for any $v \notin S \cup \{r\}$. The algorithm of [4] is based on finding a minimum weight (near) perfect matching in H, and proceeds as follows:

[*]Bell Labs, 600 Mountain Ave, Murray Hill, NJ 07974. E-mail: chekuri@research.bell-labs.com.

[†]Dept. of Computer & Information Science, University of Pennsylvania. E-mail: sanjeev@cis.upenn.edu. Supported in part by an Alfred P. Sloan Research Fellowship.

[‡]Bell Labs, 600 Mountain Ave, Murray Hill, NJ 07974. On leave from the Computer Science Dept., Technion, Haifa 32000, Israel. E-mail: naor@research.bell-labs.com.

1. Compute a minimum weight (near) perfect matching M in H.
2. For all $(u,v) \in M$, u is chosen to be a center with probability $p = w(u)/(w(u)+w(v))$, otherwise v becomes the center (with probability $1-p$). The weight of the center is set to $w(u)+w(v)$ and the non-center vertex is removed from the terminal set.
3. Continue recursively, where the set of terminals is restricted to be the set of centers computed in the previous step. Connect the non-center vertices from the previous step to the root via their respective centers.

The analysis in [4] is based on the following two claims: a) the expected cost of an optimal integral solution to the new instance is no more than that of the original instance, and b) the expected cost of connecting the non-center vertices to the root via their centers is bounded by some constant factor of the cost of an optimal integral solution. Since the number of terminals is reduced by a constant factor in each iteration this leads to an $O(\log k)$ approximation.

Derandomization: For any terminal t, let $L_t^* = \ell_{T^*}(r,t)$ denote the distance from the r to t in some optimal tree T^*. The *distance cost* of a terminal t in a given tree T is simply $w_t \cdot \ell_T(r,t)$. Then for any edge (u,v) in the matching M, it is shown in [4] that the expected distance cost of the new center is $w(u)L_u^* + w(v)L_v^*$ which is exactly the sum of the distance costs of the two terminals u and v before merging. Thus the expected distance cost of the new instance obtained after the merging is unchanged. Further, the random process works without any knowledge of L_u^* or L_v^*. To derandomize the algorithm using the standard method of conditional probabilities we would require pessimistic estimators for the quantities L_u^* and L_v^* which the algorithm in [4] does not provide.

In this note we describe an LP-based approach that obtains a deterministic $O(\log k)$-approximation algorithm for the COST-DISTANCE problem. We obtain a pessimistic estimator for L_t^* via the LP relaxation. As a byproduct we also show that the integrality gap of the LP is $O(\log k)$. Only a constant factor hardness is known for the COST-DISTANCE problem, and the worst LP integrality gap known is also $O(1)$. Our result thus raises the question of whether the LP integrality gap is $o(\log k)$; an affirmative answer would give improved approximation for the COST-DISTANCE problem as well as several other network design problems.

LP Formulation: We will view G as a directed graph (replace each undirected edge by two anti-symmetric arcs) and the solution as a tree rooted at r with directed paths to terminals in S. For all $t \in S$, $\mathcal{P}_t$ denotes the set of directed paths from the root r to t, and for each $e \in E$, a variable $x(e)$ indicates its capacity (it is a binary variable in the integral case indicating whether e is chosen in the tree or not). Note that $\mathcal{P}_t \cap \mathcal{P}_{t'} = \emptyset$ for $t \neq t'$.

$$(P)\ \min \sum_{e\in E} c(e)x(e) + \sum_{t\in S} w(t) \sum_{p\in\mathcal{P}_t} \ell(p)f(p)$$

subject to:

$$\begin{aligned}
\sum_{p\in\mathcal{P}_t | e\in p} f(p) &\leq x(e) && \forall\, e\in E,\ t\in S \\
\sum_{p\in\mathcal{P}_t} f(p) &\geq 1 && \forall\, t\in S \\
f(p), x(e) &\geq 0 && \forall\, e\in E,\ t\in S,\ p\in\mathcal{P}_t
\end{aligned}$$

The LP assigns fractional capacities to edges such that one unit of flow can be shipped from the root r to each terminal $t \in S$. We can view the flow going to different terminals as separate commodities. However, since the flow belonging to separate commodities is non-aggregating, there is no need to explicitly refer to commodities. The dual program is as follows.

$$(D)\ \max \quad \sum_{t \in S} \alpha(t)$$

subject to:

$$\begin{aligned}
\sum_{t \in S} \beta^t(e) \quad &\leq \quad c(e) && \forall\, e \in E \\
\alpha(t) - \sum_{e \in p} \beta^t(e) \quad &\leq \quad w(t) \cdot \ell(p) && \forall\, t \in S,\ p \in \mathcal{P}_t \\
\alpha(t), \beta^t(e) \quad &\geq \quad 0 && \forall\, e \in E,\ t \in S
\end{aligned}$$

Lemma 1 *The graph H contains a matching of size at least $|S|/3$ such that its weight is no more than the optimal value of the linear program (P).*

Proof. Consider the following LP for the minimum weight perfect matching problem in H.

$$(MP)\ \min \sum_{(u,v)} b(u,v) \cdot x(u,v)$$

$$\begin{aligned}
\sum_{v} x(u,v) \quad &= \quad 1 && \forall u \in V(H) \\
x(u,v) \quad &\geq \quad 0 && \forall (u,v) \in E(H)
\end{aligned}$$

Notice that we do not have the *odd-set constraints* that are necessary to obtain an optimal matching. However, it is well known that a basic feasible solution to (MP) is half-integral and that the half-integral edges induce odd cycles. Consider an odd half-integral cycle of length $i \geq 3$. It is easy to see that we can recover at least $\lfloor i/2 \rfloor$ edges of weight no more than that of the fractional cycle weight. Hence we can obtain a matching M of size at least $\lceil k/3 \rceil$ with total weight upper bounded by the optimum of (MP). The dual of (MP) is given below.

$$(MD)\ \max \sum_{u} y(u)$$

$$\begin{aligned}
y(u) + y(v) \quad &\leq \quad b(u,v) && \forall (u,v) \in E(H) \\
y(u) \quad &\geq \quad 0 && \forall u \in V(H)
\end{aligned}$$

Let x^* and y^* be optimum solutions to (MP) and (MD) respectively. It follows that $y^*(u) + y^*(v) \leq b(u,v)$ holds and by duality $\sum_{u \in H} y^*(u)$ is at least the weight of M.

We claim that there is a feasible solution to (D) of value at least $\sum_u y^*(u)$. To prove this set $\alpha(t) = y^*(t)$ for all terminals t that are matched vertices in M and set $\alpha(t) = 0$ otherwise. In the graph G, define the ball B_t centered at t of radius $\alpha(t)$ where distance is measured with respect to the metric $c(e) + w(t)\ell(e)$. The main observation is that the balls $\{B_t\}_{t \in S}$ are disjoint (they do not overlap). This follows from the fact that $y^*(u) + y^*(v) \leq b(u,v)$ for all matched edges where $b(u,v)$ is the distance between u and v in the metric $c(e) + \min\{w(u), w(v)\}\ell(e)$. For each edge e, let $g^t(e)$ be the fraction of the edge contained in B_t. We set $\beta^t(e) = g^t(e) \cdot c(e)$. It can be easily verified that these settings are feasible for (D). Hence, the optimal value of (D) (and (P)) on G is at least weight of the matching M obtained from (MP) as above. $\square$

Our algorithm will be very similar to that of [4]. Let I be the given instance.

1. Solve the LP (P) on I.
2. Compute a near perfect matching M in H of size at least $\lceil k/3 \rceil$ using (MP).
3. For all $(u, v) \in M$, choose either u or v *deterministically* (algorithm to be specified later) to be the center. Set the weight of the center to $w(u)+w(v)$ and remove the non-center vertex from the terminal list. Let I' be the new instance.
4. Recursively compute a solution for I' and connect the non-center vertices from the previous step to the root via their counterparts from the matching.

Let f^* be an optimal solution to (P) on I and let O and O' be the cost of the optimal solution to (P) on the instances I and I' respectively. Let C_M be the cost of connecting the non-center vertices to their counterparts after the recursive step. Let k' be the number of vertices in I'. From the size of the matching M we have that $k' \leq 2k/3$.

Lemma 2 *There exists a deterministic algorithm to choose the centers such that the resulting instance satisfies the following equation for some fixed constant a.*

$$a \log k \cdot O \geq a \log k' \cdot O' + C_M.$$

Before we prove the lemma we state the consequence.

Theorem 1 *The linear program (P) yields a* deterministic $O(\log k)$ *approximation algorithm for the* COST-DISTANCE *problem.*

Proof of Lemma 2. Suppose we choose the center for each matched pair of terminals exactly as in the randomized algorithm of [4] described earlier. Let X be the expected cost of (P) on the random instance created by the algorithm and let Y be the expected cost of connected the non-center vertices to their matching counterparts. We claim that $X \leq O$ and $Y \leq 2O$. Assuming these claims are true it follows that

$$\begin{aligned} a \log k' \cdot X + Y &\leq a \log(2k/3) \cdot O + 2O \\ &\leq a \log k \cdot O - a \log(3/2) + 2O \\ &\leq a \log k \cdot O \qquad \text{for sufficiently large } a. \end{aligned}$$

We first prove that $X \leq O$. Observe that f^* is a feasible solution for I' because only the weights of the terminals have been changed. We will analyse the cost of f^* on I' (which is an upperbound on O'). For $t \in S$ let $\gamma_t = \sum_{p \in \mathcal{P}_t} \ell(p) \cdot f^*(p)$. Consider an edge $(u, v) \in M$. The distance cost incurred by u and v in I is $w(u)\gamma_u + w(v)\gamma_v$. The expected distance cost of the random center chosen from $\{u, v\}$ is exactly $w(u)\gamma_u + w(v)\gamma_v$. By linearity of expectations it is easy to see that the claim follows. Observe that we are using γ_t instead of L_t^*, hence we can fact compute the expectations explicitly.

Consider a pair $(u, v) \in M$ and its contribution to Y. If u is chosen as the center then the cost is bounded by the shortest path between u and v in the metric $c(e) + w(v)\ell(e)$, similarly if v is chosen it is the shortest path in the metric $c(e) + w(u)\ell(e)$. Therefore the expected cost is bounded by the shortest path cost in the metric $c(e) + 2\frac{w(u)w(v)}{w(u)+w(v)}\ell(e)$. It is easy to see that $2\frac{w(u)w(v)}{w(u)+w(v)} \leq 2\min\{w(u), w(v)\}$ and hence the expected cost is no more than $2b(u, v)$. From Lemma 1 we know that $\sum_{(u,v)\in M} b(u, v) \leq O$. Hence $Y \leq 2O$.

All that remains is to obtain a deterministic algorithm to choose the centers while preserving the claim. We achieve this by derandomizing the above algorithm using the standard method of conditional probabilities. We process the matching pairs in some arbitrary order choosing a center for each pair as we go along. When processing the ith pair let X_i be the expected cost of (P) on the instance with the first $(i-1)$ centers fixed as chosen earlier and the rest being chosen according to the randomized algorithm. Y_i is defined similarly. We chose the center in the ith pair so that the quantity $a \log k' \cdot X_{i+1} + Y_{i+1}$ is minimized. This can be easily done by trying both vertices in the ith pair as centers, all the expectations (or their pessimistic estimators) can be explicitly computed. From induction it follows that the $a \log k' \cdot X'_k + Y'_k \leq a \log k' \cdot X + Y \leq a \log k \cdot O$. □

References

[1] G. Kortsarz and D. Peleg, "Approximating shallow-light trees", Proceedings of 8th SODA, 103-110, 1997.

[2] S. Khuller, B. Raghavachari, and N. Young, "Balancing minimum spanning trees and shortest path trees", Algorithmica, Vol. 14, pp. 305-321, 1995.

[3] M. V. Marathe, R. Ravi, R. Sundaram, S. S. Ravi, D. J. Rosenkrantz, and H. B. Hunt, Bicriteria network design problems. *J. of Algorithms*, 28, 142–171, 1998.

[4] A. Meyerson, K. Munagala, and S. Plotkin. Cost-Distance: Two Metric Network Design. To appear in *41st FOCS*, 2000.