COM S 783 / OR&IE 634 Approximation Algorithms **Fall Semester, 2003**

Lecture 12: October 15, 2003

Lecturer: Yuval Rabani

In today's lecture, we will present an $O(\log n)$-approximation algorithm for the Minimum Linear Arrangement ("MLA") problem through the use of spreading metrics.

1 Problem Definition

In MLA, the input consists of an undirected graph $G = (V, E)$, with non-negative edge weights, $w : E \to \mathbb{Q}^+$. For notational purposes, let $n := |V|$ and $V = \{1, \ldots, n\}$.

Formally, we want to find a bijection $\sigma : V \to \{1, \ldots, n\}$ that minimizes the function

$$\sum_{(i,j)\in E} |\sigma(i) - \sigma(j)| w_{i,j}$$

Informally, an MLA of graph G is an embedding of G in the linear array such that:

(i) we have a one-to-one mapping from the nodes of G to the nodes of the linear array, and

(ii) the weighted sum of the edge lengths is minimized, where the length of an edge in the embedding is given by the distance between its endpoints on the linear array.

2 Finding a Spreading Metric to MLA

A spreading metric on the graph G for a given MLA problem instance is an assignment, $\delta : E \to \mathbb{Q}$, of lengths to the edges of the graph that has the following properties:

- The volume of a spreading metric, $\sum_{(i,j)\in E} \delta_{i,j} w_{i,j}$, of graph G in the MLA problem instance provides a lower bound on the value of the optimal solution of the MLA problem.
- It "spreads apart" (w.r.t. metric lengths) the nontrivial subgraphs. This translates to "The diameter of every nontrivial connected subgraph U of V is $\Omega(|U|)$."

Claim: For any spreading metric, $\delta_{i,j}$,

$$\sum_{i\in U} \delta_{i,j} \geq \frac{|U|^2}{4}, \forall j \in U, U \subseteq V, |U| \geq 1$$

This follows from the case where node j^* is in the "middle" of all the other nodes in U with respect to the linear arrangement. i.e. There are $\frac{|U|}{2}$ nodes each to the "left" and to the "right" of node j^* in context of the set U. Hence,

$$\sum_{i \in U} \delta_{i,j} \geq 2 \sum_{i=1}^{|U|/2} i \geq \frac{|U|^2}{4}, \forall j \in U, U \subseteq V, |U| \geq 1$$

Now, consider a solution δ to the following LP:

$$W := \min \sum_{(i,j) \in E} w_{i,j} \delta_{i,j}$$

s.t.

$$\sum_{i \in U} \delta_{i,j} \geq \frac{|U|^2}{4}, \forall j \in U, U \subseteq V, |U| \geq 1$$

δ is a metric

Remark: Implicit from the first constraint is that $\delta(i,j) \geq 1, \forall i, j$.

Claim: δ is a spreading metric for the MLA problem instance.

Lemma 1: [2] $\forall U \subseteq V, |U| > 1$, $\forall v \in U, \exists u \in U$ s.t. $\delta_{u,v} \geq \frac{1}{4}(|U| - 1)$.

Lemma 2: [2] The volume of the spreading metric δ is a lower bound to the value of the optimal solution of MLA.

Using the spreading metric δ found by the LP, we can apply divide-and-conquer in which the divide step is guided by the volumes of subproblems rather than traditional methods that divide according to the sizes of the subproblems. We'll be using Lemma 1 in the analysis of the approximation guarantee of the approximation algorithm.

3 The Approximation Algorithm

Assume WLOG that G is connected $\forall e \in E, w(e) \geq 1$. Before describing the algorithm, let's introduce some notation and definitions:

- For integers x and y: $[x, y] : \{x, \dots, y-1\}$ and $[x] := \{0, \dots, x-1\}$
- For a node $v \in V$, an edge (x, y) is at *level* i with respect to v iff $\delta(v, x) \leq i$ and $\delta(v, y) > i$. Note that an edge may be at more than one level.
- The weight of level $i := \rho_i := \sum_{(x,y) \text{ in level } i} w_{x,y}$

- WLOG, we will assume that $\log W$ is an integer. Let $\alpha_k = 2^k, \forall k \in [(\log W) + 1]$. Level i has *index* k, $k \in [\log W]$, iff $\rho_i \in I_k = (\alpha_k, \alpha_{k+1}]$.

Claim: There are at least $n/4$ distinct levels with non-zero weight.

Proof: Follows from the first constraint of the LP by rewriting it as

$$\frac{1}{|U|} \sum_{i \in U} \delta_{i,j} \geq \frac{|U|}{4}$$

Corollary: $\exists$ index k st. at least $n/(4 \log W)$ levels have index k.

Let r be the exact number of levels of index k.

In each recursive step of the algorithm, cut along the sequence of r levels of index k. i.e., remove the edges that are at those levels, even if they are also at some other level of index different from k.

- For each i, let level a_i be the ith level of index k, in increasing order of distances to v.
- Let H_i be the subgraph induced by nodes $\{q : a_i < \delta(q, v) \leq a_{i+1}\}$.
- Let H_0 be the subgraph induced by nodes $\{q : \delta(q, v) \leq a_1\}$.
- Let H_r be the subgraph induced by nodes $\{q : a_r < \delta(q, v)\}$.
- Let $n_i = |H_i|$.

Recursively find a linear arrangement σ_i on the subgraph H_i. Finally, combine the linear arrangements obtained for each H_i as follows:

$$(\sigma(1), \ldots, \sigma(n)) = (\sigma_0(1), \ldots, \sigma_0(n_0), \ldots, \sigma_r(1), \ldots, \sigma_r(n_r))$$

4 Analysis

We use a charging scheme to account for the length of an edge e in the linear arrangement for G obtained by the algorithm (note that we will account for the length of the edge in the linear arrangement, rather than for the spreading metric length of the edge). If some edge e in level a_i belongs to some other level of index k, say level a_j, then this edge also belongs to every level of index k between a_i and a_j. WLOG, assume that $i < j$. Edge e will be "stretched over" all the nodes in $H_i \cup \ldots \cup H_{j-1}$, and may be "stretched over" some of the nodes in H_{i-1} and H_j, in the linear arrangement produced by the algorithm. Hence, the length of such an edge in the final linear arrangement will be at most $n_{i-1} + \ldots, n_j$. Suppose we charge $n_{p-1} + n_p$ for the portion of the edge that is stretched over the nodes in $H_{p-1} \cup H_p$, when considering level a_p, for all $p \in [i, j+1]$. Then the total charge associated with edge e is equal to $n_{i-1} + 2(n_i + \ldots + n_{j-1}) + n_j \geq n_{i-1} + \ldots + n_j$: That is, edge e will be charged at least as much as its length in a final linear arrangement.

We will now compute an upper bound on the cost of a linear arrangement obtained by the algorithm. Let $C(Z)$ be the maximum cost of a linear arrangement obtain by the algorithm for a subgraph of G whose volume of the spreading metric ℓ is at most Z. Since the sum of the weights of all edges in level a_i is p_{a_i}, and since we charge for the length of an edge as described in the preceding paragraph, we have the following recurrence relation for $C(W)$:

$$C(W) \leq C(W - \sum_{i=1}^{r} p_{a_i}) + \sum_{i=1}^{r} [p_{a_i}(n_{i-1} + n_i)]$$

We now show that $C(W) = O(W \log n)$.

Lemma: $\exists c$ s.t. $C(W) \leq cW \log W$

Proof: WLOG, W is a rational number. Use induction on W. The base case $W = 0$ corresponds to a totally disconnected graph. Therefore, $C(0) = 0$.

Combining the recurrence relation for $C(W)$ with $a_k < p_{a_i} \leq a_{k+1}, \forall i$, we obtain:

$$\begin{aligned}
C(W) &\leq C\left(W - \frac{\alpha_k n}{4 \log W}\right) + \alpha_{k+1} \sum_i (n_{i-1} + n_i) \\
&\leq c \left[W - \frac{\alpha_k n}{4 \log W}\right] \log \left(W - \frac{\alpha_k n}{4 \log W}\right) + 2\alpha_{k+1} n \\
&\leq c \left[W - \frac{\alpha_k n}{4 \log W}\right] \log(W) + 2\alpha_{k+1} n \\
&\leq cW \log W + \alpha_{k+1} n \left[2 - \frac{c}{8 \log W} \log W\right] \\
&\leq cW \log W
\end{aligned}$$

The second inequality follows from the induction hypothesis; the fourth inequality follows since $\alpha_{k+1} = 2\alpha_k$. QED

What remains is to show that we can modify the problem such that $O(\log W) = O(\log n)$, and the cost of the optimal solution to the modified problem only goes up by a factor of W.

Modify the original instance of MLA by rounding the weights of each edge $w_{i,j}$ down to the nearest multiple of $W/(n|E|)$. The error incurred by this rounding procedure is at most W. Furthermore, scale the rounded weights by $W/(n|E|)$, obtaining new weights for the edges that are all integers in the interval $[0, n|E|]$. Note that we've only changed the units in which the weights are expressed.

Theorem: [1] We can modify the original problem instance as above to give an approximation factor of $O(\log W) = O(\log n)$.

References

[1] S. Rao, A.W. Richa. New Approximation Techniques for Some Ordering Problems. *Proceedings of the Ninth Annual ACM-SIAM Symposium on Discrete Algorithms*, 211-218, 1998.

[2] G. Even, J.S. Naor, S. Rao, B., Schiever. *Divide-and-Conquer Approximation Algorithms via Spreading Metrics. Proceedings of the 36th Annual Symposium on Foundations of Computer Science*, 62-71, 1995.