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

Lecture 14: Oct 22, 2003

Lecturer: Yuval Rabani   *Scribe: Ashwin Kumar Machanavajjhala*

1 Embedding Metrics into distributions of dominating tree metrics[1]

The last lecture, we saw that

$$Pr[(u,v) \text{ is at level } i] \leq \frac{1}{2^{i-1}} d(u,v) \left(\mathcal{H}_{|K_i^u|} + \mathcal{H}_{|K_i^v|}\right)$$

where K_i^u denotes the set of nodes at distance at most 2^i from u. Since $|K_i^u| \leq |V|$, we get that

$$Pr[(u,v) \text{ is at level } i] \leq \frac{1}{2^{i-1}} d(u,v) O(log\, n)$$

If we sum over all the levels, we would get $E(d_T(u,v)) \leq d(u,v)O(log^2 n)$ since there are a logarithmic number of levels. However, if we are more careful, we can bound the distortion by only a logarithmic factor in n.

We now do a more careful analysis and show that except for a constant number of levels, we can find a much smaller upper bound on the value of $Pr[(u,v)$ is at level $i]$. Let j^* denote the smallest level i such that $d(u,v) \leq 2^{i+2}$. Consider a level ℓ, $\ell \geq j^*+4$. Consider $y \in K_\ell^u$. Can y separate (u,v) ?

Since $y \in K_\ell^u$, $d(u,y) \leq 2^{\ell-2}$. By the definition of j^*, $d(u,v) \leq 2^{j^*+2} \leq 2^{\ell-2}$. Hence we have, $d(y,v) \leq 2^{\ell-1}$. Since $\beta_\ell \in [2^{\ell-1}, 2^\ell]$ and both $d(u,y), d(y,v) \leq 2^{\ell-1} \leq \beta_\ell$, such a y cannot separate u and v.

Hence, at levels $\ell \geq j^*+4$ we can discard all points in K_ℓ^u.

$$\begin{aligned}
Pr[(u,v) \text{ is at level } \ell] &\leq \Sigma_{j=|K_{\ell-2}^u|+1}^{|K_\ell^u|} \frac{1}{j}\frac{1}{2^{\ell-1}} d(u,v) \\
&+ \Sigma_{j=|K_{\ell-2}^v|+1}^{|K_\ell^v|} \frac{1}{j}\frac{1}{2^{\ell-1}} d(u,v) \\
&\leq \frac{1}{2^{\ell-1}} d(u,v) \left(\mathcal{H}_{|K_\ell^u|} - \mathcal{H}_{|K_{\ell-2}^u|} + \mathcal{H}_{|K_\ell^v|} - \mathcal{H}_{|K_{\ell-2}^v|}\right) \\
\text{Hence, } E(d_T(u,v)) &\leq \Sigma_{\ell=j^*}^m Pr[(u,v) \text{ is at level } \ell] \cdot 2^{\ell+2} \\
&\leq \Sigma_{\ell=j^*}^{j^*+3} d(u,v) \cdot O(log\, n) \\
&+ \Sigma_{\ell=j^*+4}^m 8d(u,v) \cdot \left(\mathcal{H}_{|K_\ell^u|} - \mathcal{H}_{|K_{\ell-2}^u|} + \mathcal{H}_{|K_\ell^v|} - \mathcal{H}_{|K_{\ell-2}^v|}\right) \\
\mathbf{E(d_T(u,v))} &\leq \mathbf{d(u,v)O(\log n)}
\end{aligned}$$

2 Metric Labeling Problem[2]

<u>INPUT</u>:

- a graph $G = (V, E)$
- positive rational weights on edges $w : E \rightarrow \mathcal{Q}^+$
- set of labels $L = \{1, 2, \ldots, k\}$
- (L, d) a metric space
- assignment cost $c : V \times L \rightarrow \mathcal{Q}^+ \cup \{\infty\}$; i.e. $c(u, j)$ is the cost of assigning label j to node u

<u>PROBLEM</u> Find an assignment ϕ from the set of nodes to the set of labels that minimizes

$$\Sigma_{x \in V} c(x, \phi(x)) + \Sigma_{xy \in E} d(\phi(x), \phi(y)) w(x, y)$$

We solve this problem in the following two steps:

1. Embed the metric d into a distribution over dominating tree metrics $\{t_i\}$
2. pick a random tree t from the distribution and solve the problem with t replacing d.

Note that by embedding the metric in a distribution of tree metrics, no distance shrinks. The second step can be solved optimally by using linear programming techniques.

Why does it work? Let ϕ^* be an optimal solution (to the original problem). Let $cost(\phi^*, t_i)$ denote the cost of the solution measured using t_i.

$$\begin{aligned}
cost(\phi^*, t_i) &= \Sigma_{x \in V} c(x, \phi^*(x)) + \Sigma_{x,y \in E} t_i(\phi^*(x), \phi^*(y)) w(x, y) \\
E[cost(\phi^*, t_i)] &= \Sigma_{x \in V} c(x, \phi^*(x)) + \Sigma_{x,y \in E} E[t_i(\phi^*(x), \phi^*(y))] w(x, y) \\
&\leq \Sigma_{x \in V} c(x, \phi^*(x)) + \Sigma_{x,y \in E} O(log\, k) d(\phi^*(x), \phi^*(y)) w(x, y) \\
&\leq O(log\, k) cost(\phi^*, d)
\end{aligned}$$

On input t_i, the algorithm computes ϕ_i which is the optimal cost for that t_i; i.e. $\phi_i = \phi_i^*$. Since the tree metrics are dominating(non contracting), we have

$$\begin{aligned}
cost(\phi_i, d) &\leq cost(\phi_i, t_i) \\
E[cost(\phi_i, d)] &\leq E[cost(\phi_i, t_i)] \\
&\leq E[cost(\phi^*, t_i)] \text{ since } \phi^* \text{ is the optimal} \\
&\leq O(log\, k) cost(\phi^*, d)
\end{aligned}$$

References

[1] Jittat Fakcharoenphol, Satish Rao and Kunal Talwar. A tight bound on approximating arbitrary metrics by tree metrics. In *Proceedings of the 35nd Annual ACM Symposium on Theory of Computing (STOC 2003), 448-455 .*

[2] Jon M. Kleinberg and Éva Tardos. Approximation Algorithms for Classification Problems with Pairwise Relationships: Metric Labeling and Markov Random Fields. In *Proceedings of FOCS 1999, 14-23 .*