

## Lecture 24: December 1, 2014

CS 430 Introduction to Algorithms  
Fall Semester, 2014

In the *nearest-neighbor* approximation to the traveling salesman problem, we begin by selecting an arbitrary city as a starting point. From the set of cities not yet visited, select as the next city the one closest to the last city added to the tour, or, if all cities have been visited, return to the origin. Let the tour found by the nearest neighbor algorithm be  $NN$  of cost  $|NN|$ , and let the optimal tour be  $OPT$  of cost  $|OPT|$ .

Relabel the edges of tour  $NN$  so that their lengths satisfy  $l_1 \geq l_2 \geq \dots \geq l_n$ , where  $\sum_{i=1}^n l_i = |NN|$ . Next, let the city we exited via edge  $l_i$  be labeled  $c_i$ . Note that, because we are choosing the labels so that the sequence of edge lengths is in decreasing order, we do not know anything about the other endpoint of edge  $l_i$  or the order in which the cities are visited. Let  $C_{a,b}$  be the cost of the edge from city  $a$  to city  $b$ .

We can begin bounding the heuristic's performance by observing that

$$|OPT| \geq 2l_1 \quad (1)$$

by the triangle inequality, since the endpoints of  $l_1$  must be visited sometime during the optimal tour.

Next we show that

$$|OPT| \geq 2 \sum_{i=k+1}^{2k} l_i \quad (2)$$

for all values of  $k$ . To prove this, we define  $T_j$  as the optimal tour of cities  $c_1$  through  $c_j$  and let  $|T_j|$  be its length. Since this subset of cities must be visited in the optimal tour, the triangle inequality tells us that  $|OPT| \geq |T_{2k}|$ . Now, consider two cities  $c_i$  and  $c_j$  such that  $(c_i, c_j)$  is an edge in the tour  $T_{2k}$ . If  $c_i$  precedes  $c_j$  in the heuristic tour, then  $C_{c_i, c_j} \geq l_i$  since  $c_j$  had not been visited so edge  $(c_i, c_j)$  could have been chosen for the heuristic tour. On the other hand, if  $c_j$  precedes  $c_i$  in the heuristic tour, then  $C_{c_j, c_i} \geq l_j$  by the same reasoning. Since  $C_{c_i, c_j} = C_{c_j, c_i}$ , in either case,  $C_{c_i, c_j} \geq \min\{l_i, l_j\}$ .

Summing this inequality over the edges of  $T_{2k}$  gives  $|T_{2k}| \geq \sum \min\{l_i, l_j\}$ . Each  $l_i$  appears in this list at most twice. Further, because the edges were labeled in decreasing order, we can replace edges in the first  $k$   $l_i$  with members of the last  $k$  edges  $l_i$ . Since there are  $2k$  edges in  $T_{2k}$ , this process yields  $|T_{2k}| \geq 2 \sum_{i=k+1}^{2k} l_i$ . Using our previous observation that  $|OPT| \geq |T_{2k}|$  gives the inequality (2).

Finally, with a similar proof, we have

$$|OPT| \geq 2 \sum_{i=\lceil \frac{n}{2} \rceil + 1}^n l_i \quad (3)$$

Summing both sides of (1), (3), and (2) for  $k = 1, 2, 2^2, \dots, 2^{\lceil \lg n \rceil - 2}$ , we conclude that

$$\begin{aligned} (\lceil \lg n \rceil + 1) |OPT| &\geq 2 \sum_{i=1}^n l_i \\ &= 2|NN| \end{aligned}$$

or, alternatively,

$$\frac{|NN|}{|OPT|} \leq \frac{\lceil \lg n \rceil + 1}{2}$$

This result comes from D. J. Rosenkrantz, R. E. Stearns, and P.M. Lewis II, “An analysis of several heuristics for the traveling salesman problem,” *SIAM Journal on Computing*, 6(3):563-581, September 1977.