Homework #8
Introduction to Algorithms/Algorithms 1
600.363/463
Spring 2014
Solutions

April 2, 2014

Problem 1 (20 points)

Let $G = (V, E)$ be an undirected graph with distinct non-negative edge weights. Consider a problem similar to the single-source shortest paths problem, but where we define path cost differently. We will define the cost of a simple s–t path $P_{s,t} = \{e_1, e_2, \ldots, e_k\}$ to be

$$c(P_{s,t}) = \max_{e \in P} w_e.$$

That is, the cost of a path is now just the largest weight on that path, rather than the sum of the weights on the path. Give an algorithm that takes as input an undirected graph $G = (V, E)$ with non-negative edge weights, and a vertex $s \in V$, and computes the path from s to every other node in G with the least cost under cost function $c(\cdot)$. That is, for each $t \in V$, find a simple path connecting s to t, $P_{s,t} = \{(s, v_1), (v_1, v_2), \ldots, (v_{k-1}, v_k), (v_k, t)\}$, such that, letting $\mathcal{S}$ denote the set of all simple paths connecting s and t, $c(P_{s,t}) = \min_{P \in \mathcal{S}} c(P)$. Your algorithm should run in $O(|E| + |V| \log |V|)$ time. Prove the correctness of your algorithm and its runtime. Hint: note the similarities between Dijkstra's algorithm and Prim's algorithm.

0.1 Solution

It suffices to change the RELAX function in Dijkstra's algorithm. Nothing else in the analysis changes– our invariant is still that the distance stored for each vertex is always the cost of the best path to that vertex discovered so far.

Specifically, we will change the way we update the $v.d$ field for each vertex v (recall that $v.d$ tracks what we currently believe to be the length of the shortest path to vertex v). Suppose we are currently at vertex u, and $v \in \text{Adj}(u)$. If $\max\{u.d, w(u,v)\} < v.d$, then we will set $v.d = \max\{u.d, w(u,v)\}$. The correctness of our new version of Dijkstra's algorithm follows from the correctness of Dijkstra's algorithm and the fact that our new version of the RELAX function still obeys the invariant used in Dijkstra's, namely that the $v.d$ field is always the cost of the lowest-cost path from the source node to v explored so far.

Problem 2 (20 points)

Let $G = (V, E)$ be a connected undirected graph with unit weights. We will define two similar but distinct *graph statistics*, numbers that summarize information about graph G. The first we'll call the *diameter* of G,

$$\text{diam}(G) = \max_{u,v \in V} \delta(u,v),$$

where $\delta(u,v)$ is the length of the shortest path connecting u and v. The second we'll call the *average distance* of G,

$$\text{avedist}(G) = \frac{1}{\binom{n}{2}} \sum_{\{u,v\} \subseteq V} \delta(u,v).$$

In general, it need not be the case that $\text{diam}(G) = \text{avedist}(G)$. For example, consider an unweighted graph H on three nodes a, b, c, with two edges (a,b) and (b,c). This graph has $\text{diam}(H) = \delta(a,c) = 2$, while $\text{avedist}(H) = \frac{\delta(a,b)+\delta(b,c)+\delta(a,c)}{3} = 4/3$. Prove or disprove the following claim: there exists a positive constant c such that for all connected, undirected graphs G, we have

$$\frac{\text{diam}(G)}{\text{avedist}(G)} \leq c.$$

Hint: start with a fully connected graph, and connect a "path" to it.

0.2 Solution

The claim is false.

Consider undirected graph $G_{k,m}$ on $k+m$ nodes: We have a vertex set $V = \{b_1, b_2, \ldots, b_k, t_1, t_2, \ldots, t_m\}$ and and edge set E. Edge $\{b_i, b_j\} \in E$ for all $1 \leq i < j \leq k$, $\{t_i, t_{i+1}\} \in E$ for all $1 \leq i < m$, and $\{b_1, t_1\} \in E$. This graph can be seen to be a complete connected component (the $\{b_i\}$ nodes), with a linear chain attached to it (the $\{t_j\}$ nodes). The diameter of this graph is $m+1$ (the distance from b_i to t_m is $m+1$ for any $1 < i \leq k$). Thus,

$$\operatorname{diam}(G_{k,m}) = m + 1. \tag{1}$$

To calculate $\operatorname{avedist}(G_{k,m})$, we must resort to a bit of algebra. We have

$$\operatorname{avedist}(G_{k,m}) = \frac{1}{\binom{m+k}{2}} \sum_{\{u,v\} \in V} \delta(u, v). \tag{2}$$

Every $\{b_i, b_j\}$ pair contributes $\delta(b_i, b_j) = 1$ to the summation in Equation (2). For every pair $\{b_i, t_j\}$, if $1 < i \leq k$ and $1 \leq j \leq m$, we have $\delta(b_i, t_j) = j + 1$. b_1 is "special", and contributes $\delta(b_1, t_j) = j$ for all $1 \leq j \leq m$. Finally, we must compute the total distance contributed by all $\{t_i, t_j\}$ edges for $1 \leq i < j \leq m$. This distance can be expressed as $\delta(t_i, t_j) = j - i$. Summing over all such pairs in the tail of our graph, we have

$$\begin{aligned}
\sum_{i=1}^{m-1} \sum_{j=i+1}^{m} \delta(t_i, t_j) &= \sum_{i=1}^{m-1} \sum_{j=i+1}^{m} (j - i) \\
&= \sum_{i=1}^{m-1} \left(\sum_{j=1}^{m} j - \sum_{j=1}^{i} j \right) - \sum_{i=1}^{m-1} (m-i)i \\
&= \sum_{i=1}^{m-1} \left(\frac{m(m+1)}{2} - \frac{i(i+1)}{2} \right) - \sum_{i=1}^{m-1} \left(mi - i^2 \right) \\
&= \frac{(m-1)m(m+1)}{2} - \left(\sum_{i=1}^{m-1} \frac{i^2}{2} \right) - \left(\sum_{i=1}^{m-1} \frac{i}{2} \right) - m \left(\sum_{i=1}^{m-1} i \right) + \left(\sum_{i=1}^{m-1} i^2 \right) \\
&= \frac{(m-1)m(m+1)}{2} - \frac{1}{2} \frac{(m-1)m(2m-1)}{6} - \frac{1}{2} \frac{(m-1)m}{2} \\
&\qquad - m \frac{(m-1)m}{2} + \frac{(m-1)m(2m-1)}{6} \\
&= \frac{m(m-1)}{4} + \frac{(m-1)m(2m-1)}{12} \\
&= \frac{m(m^2-1)}{6}.
\end{aligned}$$

Combining the above facts, we have

$$\begin{aligned}\text{avedist}(G_{k,m}) &= \frac{1}{\binom{k+m}{2}}\left(\binom{k}{2} + (k-1)\frac{m^2+3m}{2} + \frac{m(m+1)}{2} + \frac{m(m^2-1)}{6}\right)\\ &= \frac{3k^2 - 3k + 3m^2k + 9mk - 7m + m^3}{(k+m)(k+m-1)}.\end{aligned}$$

Now, plugging in Equation (1), we have

$$\begin{aligned}\frac{\text{diam}(G_{k,m})}{\text{avedist}(G_{k,m})} &= \frac{(k+m)(k+m-1)(m+1)}{3k^2 - 3k + 3m^2k + 9mk - 7m + m^3}\\ &= \frac{k^2m + k^2 + mk + 2m^2k + m^3 - k - m}{3k^2 + 9mk + 3m^2k + m^3 + 9mk - 7m}.\end{aligned}$$

If we send $k \to \infty$ and $m \to \infty$ so that $k^2m \to \infty$ but $\frac{k}{m} \to \infty$ (i.e., make both k and m go to infinity, but make k go to infinity at a much faster rate than m), then for any constant c we can make this ratio as large as we please by selecting suitably large k and m, contradicting the claim that the ratio is bounded by some constant.

Let's try and say this another, slightly more hand-wavy way for those of us who haven't taken calculus (or real analysis) recently: Cor any $c > 1$ (we can assume $c > 1$, since we already saw in the example given in the problem description that we can make the ratio greater than 1), we need to choose k and m so that

$$k^2m > k^2(3c-1) + mk(9c-1) + m^2k(3c-2) + m^3(c-1) - k(3c+1) - m(7c+1). \tag{3}$$

We can safely ignore the last two terms, because they're negative (and because $c > 1$ by assumption, so that all the c-related terms inside the parentheses are positive). By taking $m > (3c-1)$, we make $k^2m > k^2(3c-1)$. By taking $k > (9c-1)$, we make $k^2m > mk(9c-1)$. By taking $k > (3c-2)m$, we make $k^2m > m^2k(3c-2)$. Finally, by taking $k > m\sqrt{c-1}$, we make $k^2m > m^3(c-1)$. Since none of these inequalities conflict with one another, we can make k and m as large as we like (provided we obey the $k > \max\{(3c-2)m, m\sqrt{c-1}\}$ inequality), to the point that we can ensure that the "tougher" inequality given in Equation (3) is also true. But if that inequality is true, then we've made the ratio of diameter to average distance larger than c, as we wished to do.

Optional exercises

Solve the following problems and exercises from CLRS: 26-4, 26-5, 26-6.