Esercises on Graph Analytics

Exercise. Let $G = (V, E)$ be an undirected graph with n nodes and $m = n^{1+c}$ edges, for some constant $c > 0$. The following algorithm computes a minimum spanning forest for G in MapReduce using $O(m)$ aggregate space and $O(n^{1+\epsilon})$ local space, for some fixed $\epsilon \in (0, c]$. Let $M = n^{1+\epsilon}$ and $m_0 = m$. Initially, all edges are called *live edges*. For $i \geq 1$

Round i: if there are $m_{i-1} \leq M$ live edges then return the minimum spanning forest computed on the subgraph induced by these edges. Otherwise

- Partition the live edges into $\ell = m_{i-1}/M$ subsets by assigning each edge to a random subset, chosen with uniform probability and independently of the other edges. Assume for simplicity that m_{i-1}/M is an integer.
- Compute the minimum spanning forest separately for each subset. Call live edges all edges belonging to the ℓ forests and let m_i be their number. All other edges are discarded.

Answer the following questions.

1. Give an upper bound to the number of rounds as a function of c and ϵ.
2. Consider a Round i with $m_{i-1} > M$, and let $E^{(i)}$ be the set of m_{i-1} live edges at the beginning of the round. In the round, $E^{(i)}$ is randomly partitioned into $\ell = m_{i-1}/M$ subsets, say $E_1^{(i)}, E_2^{(i)}, \ldots, E_\ell^{(i)}$. For any fixed j, with $1 \leq j \leq \ell$, give an upper bound to $\Pr(|E_j^{(i)}| \geq 6M)$.

   Hint: Use the Chernoff Bound that states that for a Binomial r.v. Z with $E[Z] = \mu$, $\Pr(Z \geq 6\mu) \leq 2^{-6\mu}$.
3. Using the above points, show that the probability that the algorithm requires local space $< 6M$ tends to 1 as n goes to ∞.

Solution.

1. Since a spanning forest of any subgraph of G has at most $n - 1$ edges we have that $m_i \leq (n-1)m_{i-1}/M < nm_{i-1}/M$, which implies $m_i < m_0(n/M)^i = n^{1+c}(n/M)^i$, for every $i \geq 1$. Hence, the last round is Round $i + 1$, where i is the smallest integer such that $m_i \leq M$. Since for $i + 1 \geq \lceil c/\epsilon \rceil$ we have $n^{1+c}(n/M)^i \leq M$, we conclude that the number of rounds is $\leq \lceil c/\epsilon \rceil$.
2. Let $Z = |E_j^{(i)}|$. Since each live edge of $E^{(i)}$ ends up in $E_j^{(i)}$ with probability $1/\ell$, independently of the other edges, Z can be regarded as the sum of m_{i-1} i.i.d. Bernoulli

variables X_e, one for each live edge e, where $X_e = 1$ if $e \in E_j^{(i)}$ and 0 otherwise, and $\Pr(X_e = 1) = 1/\ell$. Hence, Z is a Binomial r.v. with expectation

$$\mu = \frac{m_{i-1}}{\ell} = m_{i-1}\frac{1}{m_{i-1}/M} = M.$$

By Chernoff Bound we have that

$$\Pr(Z \geq 6M) \leq 2^{-6M}.$$

3. Since there are $\leq \lceil c/\epsilon \rceil$ rounds and $< n^2/M$ subsets of edges created at each round, the probability that exists a round and a subset of live edges, created in that round, with size $\geq 6M$, is at most

$$\left\lceil \frac{c}{\epsilon} \right\rceil \frac{n^2}{M}\frac{1}{2^{6M}} = \left\lceil \frac{c}{\epsilon} \right\rceil n^{1-\epsilon}\frac{1}{2^{6n^{1+\epsilon}}},$$

which tends to 0 as n tends to ∞. Therefore, the algorithm requires local space $< 6M$ with probability that tends to 1 as n goes to ∞.

□

Exercise. Let $G = (V, E)$ be a connected, undirected graph with n nodes. Suppose that a BFS is executed from each of $k > 1$ distinct *pivots* $v_1, v_2, \ldots, v_k \in V$, and that the following two values are computed:

$$\begin{aligned} R &= \max_{u \in V} \min_{1 \leq i \leq k} \text{dist}(u, v_i) \\ \Delta &= \max_{1 \leq i,j \leq k} \text{dist}(v_i, v_j). \end{aligned}$$

Determine a lower and an upper bound to the diameter of G (denoted by Diameter(G)) as functions of R and Δ. Justify your answer.

Solution. By definition,

$$\text{Diameter}(G) = \max_{x,y \in V} \text{dist}(x, y).$$

Since Δ accounts only for a subset of pairs, we have that $\text{Diameter}(G) \geq \Delta$. Let x and y be two nodes such that $\text{dist}(x, y) = \text{Diameter}(G)$, and let v_i be the pivot closest to x, and v_j the pivot closest to y. Since one can go from x to y passing first through v_i and then through v_j, we have that

$$\text{dist}(x, y) \leq \text{dist}(x, v_i) + \text{dist}(v_i, v_j) + \text{dist}(v_j, y) \leq 2R + \Delta.$$

Thus,

$$\Delta \leq \text{Diameter}(G) \leq 2R + \Delta.$$

□

Exercise. Let $G = (V, E)$ be a connected, undirected graph with n nodes. For an integer, with $1 \leq k < n$, let $v_1, v_2, \ldots, v_k$ be k random *pivots* drawn from V independently, with replacement and with uniform probability. According to the Eppstein and Wang's method, For each $v \in V$, its closeness centrality $c(v)$ can be approximated by the estimator

$$\tilde{c}(v) = \frac{k(n-1)}{n} \sum_{i=1}^{k} \mathrm{dist}(v, v_i).$$

Show that $1/\tilde{c}(v)$ is an unbiased estimator of $1/c(v)$, i.e.,

$$\mathrm{E}\left[\frac{1}{\tilde{c}(v)}\right] = \frac{1}{c(v)},$$

Solution. Since the pivots are random, for each $v \in V$ and for each pivot v_i the value $\mathrm{dist}(v, v_i)$ is a random variable. Since the pivots are drawn from V independently, with replacement and with uniform probability, we have that the expectation of $\mathrm{dist}(v, v_i)$ is

$$E[\mathrm{dist}(v, v_i)] = \sum_{u \in V} \left(\frac{1}{n}\mathrm{dist}(v, u)\right).$$

Then, for each $v \in V$

$$\begin{aligned}
E\left[\frac{1}{\tilde{c}(v)}\right] &= E\left[\frac{n}{k(n-1)} \sum_{i=1}^{k} \mathrm{dist}(v, v_i)\right] \\
&= \frac{n}{k(n-1)} E\left[\sum_{i=1}^{k} \mathrm{dist}(v, v_i)\right] \\
&= \frac{n}{k(n-1)} \sum_{i=1}^{k} E[\mathrm{dist}(v, v_i)] \quad \text{(by linearity of expectation)} \\
&= \frac{n}{k(n-1)} \sum_{i=1}^{k} \sum_{u \in V} \left(\frac{1}{n}\mathrm{dist}(v, u)\right) \\
&= \frac{n}{k(n-1)} \frac{k}{n} \sum_{u \in V} \mathrm{dist}(v, u) \\
&= \frac{n}{k(n-1)} \frac{k}{n} \sum_{u \in V, u \neq v} \mathrm{dist}(v, u) \\
&= \frac{1}{c(v)}
\end{aligned}$$

□