Assignment 3

Problem 1. On-line Algorithm to Greedy On-line Algorithm.

Solution. Let the on-line inputs be $u_1, ..., u_n$. We will convert an on-line algorithm $\mathcal{A}$ to a greedy on-line algorithm $\mathcal{B}$ as follows: $\mathcal{B}$ randomly chooses an order σ on the fixed vertices of V. On input u_i, $\mathcal{B}$ runs $\mathcal{A}$ on inputs $u_1, ..., u_i$. Suppose $\mathcal{A}$ matches u_i to v_i. If v_i is not currently matched by $\mathcal{B}$, then matched u_i to v_i. Otherwise match u_i to an available neighbor v with smallest $\sigma(v)$. Further, if $\mathcal{A}$ does not match u_i, $\mathcal{B}$ matches u_i to an available neighbor v with smallest $\sigma(v)$.

Let $M_{\mathcal{A}}$ be the matching produced by algorithm $\mathcal{A}$ and $M_{\mathcal{B}}$ be the matching produced by the greedy algorithm $\mathcal{B}$. We will show that $|M_{\mathcal{A}}| \leq |M_{\mathcal{B}}|$ by transforming $M_{\mathcal{B}}$ into a matching which contains $M_{\mathcal{A}}$. Let (u_i, v_j) be an edge with smallest index i which is in $M_{\mathcal{A}}$ and not in $M_{\mathcal{B}}$. There must exist a matching (u_k, v_j) in $M_{\mathcal{B}}$ otherwise $\mathcal{B}$ would have added edge (u_i, v_j). Observe that $k < i$ since the edge (u_k, v_j) was already added when $\mathcal{B}$ considered u_i. u_k cannot be matched by $\mathcal{A}$ since u_i is the vertex with smallest index whose matching edge differed in $M_{\mathcal{A}}$ and $M_{\mathcal{B}}$. Thus we can remove (u_k, v_j) and add (u_i, v_j) to $M_{\mathcal{B}}$ without changing the number of edges in $M_{\mathcal{B}}$. After finitely many such modifications, $M_{\mathcal{A}} \subseteq M_{\mathcal{B}}$ so $\mathcal{B}$ has at least the same approximation ratio as $\mathcal{A}$.

Upon further inspection, it seems unnecessary to fix an order on V, but doing so does not hurt.

Problem 2. ϵ-Approximating Median.

Solution. Let $X_1, ..., X_m$ be i.i.d. random variables in $\{0, 1\}$ where X_i indicates whether or not the input at index i was sampled. Let

$$X_L = \sum_{i: a_i \leq m/2 - \epsilon m} X_i \qquad \text{and} \qquad X_H = \sum_{i: a_i \geq m/2 + \epsilon m} X_i.$$

We will use the Chernoff bound to bound the probability that more than half of the sampled inputs are in S_L or more than half of the sampled inputs are in S_H. First observe that $E[X_L] = E[X_H] = t \cdot \left(\frac{1}{2} - \epsilon\right)$ and $t \cdot \left(1 + \frac{2\epsilon}{1-2\epsilon}\right) \cdot \left(\frac{1}{2} - \epsilon\right) = \frac{t}{2}$. Let $\gamma = \frac{2\epsilon}{1-2\epsilon}$.

$$\Pr\left[X_L \geq \frac{t}{2}\right] = \Pr\left[X_L \geq t\left(1 + \gamma\right) \cdot \left(\frac{1}{2} - \epsilon\right)\right] \leq e^{-\frac{\gamma^2 \cdot t\left(\frac{1}{2} - \epsilon\right)}{3}}.$$

Further note that $\Pr[X_L \geq t/2] = \Pr[X_S \geq t/2]$. The probability that the algorithm does not return the ϵ-median is: $(1 - \Pr[X_L \geq t/2]) \cdot (1 - \Pr[X_S \geq t/2]) = (1 - P)^2$ where $P = \Pr[X_L \geq$

$t/2] = \Pr[X_S \geq t/2]$. Since we want this value to be greater than $1-\delta$:

$$\begin{aligned}
(1-P)^2 = \left(1 - e^{-\frac{\left(\frac{2\epsilon}{1-2\epsilon}\right)^2 \cdot t\left(\frac{1}{2}-\epsilon\right)}{3}}\right)^2 &\geq 1-\delta \\
1-\sqrt{1-\delta} &\geq e^{-\frac{\left(\frac{2\epsilon}{1-2\epsilon}\right)^2 \cdot t\left(\frac{1}{2}-\epsilon\right)}{3}} \\
\ln\left(1-\sqrt{1-\delta}\right) &\geq -\frac{\left(\frac{2\epsilon}{1-2\epsilon}\right)^2 \cdot t\left(\frac{1}{2}-\epsilon\right)}{3} \\
\left(\frac{2\epsilon}{1-2\epsilon}\right)^2 \cdot t\left(\frac{1}{2}-\epsilon\right) &\geq 3\ln\left(\frac{1}{1-\sqrt{1-\delta}}\right) \\
t &\geq \frac{3(1-2\epsilon)}{2\epsilon^2} \cdot \ln\left(\frac{1}{1-\sqrt{1-\delta}}\right)
\end{aligned}$$

serves as a bound for t.

Problem 3. Graph Connectivity.

1. If a graph is ϵ-far from being connected, it has at least $\epsilon m + 1$ connected components.

   Proof. A graph is ϵ-far from being connected if we need to add at least ϵm edges before the graph becomes connected. Suppose for a contradiction, that graph G is ϵ-far but has $k < \epsilon m + 1$ components $C_1, ..., C_k$. We can connect G by adding edges $e_1, ..., e_{k-1}$ where e_i connects components C_i and C_{i+1}. Since $k - 1 < \epsilon m$, G is not ϵ-far. □

2. If a graph is ϵ-far from being connected, then at least $\frac{\epsilon m}{2}$ connected components have size at most $\frac{4}{\epsilon d}$.

   Proof. Since $\sum_{v \in V} deg(v) = 2m$, the average degree $d = \frac{2m}{n}$. Suppose for a contradiction that there are $t < \frac{\epsilon m}{2}$ connected components of size at most $\frac{4}{\epsilon d}$. By part 1 we know that there are at least $\epsilon m + 1$ components, so there must be at least $\epsilon m + 1 - t$ components of size $> \frac{4}{\epsilon d}$. The total number of vertices in all the components is more than

   $$\begin{aligned}
   (\epsilon m + 1 - t) \cdot \left(\frac{4}{\epsilon d}\right) + t &= (\epsilon m + 1 - t) \cdot \left(\frac{2n}{\epsilon m}\right) + t \\
   &= 2n + \frac{2n}{\epsilon m} - \frac{2nt}{\epsilon m} + t \\
   &\geq 2n - \frac{2nt}{\epsilon m} \\
   &> n
   \end{aligned}$$

   This is a contradiction so the claim must hold. □

3. The algorithm is indeed a valid connectivity tester.

Proof. If G is connected, then the algorithm will always be able to discover s vertices no matter where it starts. Thus the algorithm will always return 'Yes'.

If G is ϵ-far from being connected, then G must have at least $\epsilon m + 1$ connected components by part 1 and at least $\frac{\epsilon m}{2}$ of these components are "small" (size at most $\frac{4}{\epsilon d}$ where d is the average degree) by part 2. If the algorithm choses any vertex in a small component, then it will be unable to find s distinct vertices and will return 'No'. It remains to show that if we sample $\Theta\left(\frac{1}{\epsilon d}\right)$ vertices at random the probability of picking a vertex in a small component is greater than $\frac{2}{3}$. We will sample $\frac{3}{\epsilon d} = \frac{3n}{2\epsilon m} \in \Theta\left(\frac{1}{\epsilon d}\right)$ vertices. Observe that the probability of picking a vertex in a small component is at least $\frac{\epsilon m}{2n}$. Let $k = \frac{2n}{\epsilon m}$. Then the probability that none of the vertices we picked are in small components is

$$\left(1 - \frac{1}{k}\right)^{\frac{3}{2k}} = \left(\left(1 - \frac{1}{k}\right)^{k}\right)^{\frac{3}{2}}.$$

As $\epsilon \to 0$, $k \to \infty$ so we can approximate $\left(1 - \frac{1}{k}\right)^k$ by $\frac{1}{e}$. Thus the probability that at least one vertex is in a small components is $1 - \left(\frac{1}{e}\right)^{\frac{3}{2}} \approx 0.776 > \frac{2}{3}$. □