Modern Algorithm Design (Monsoon 2024)
Endsem
Full Marks 40

1. Recall the core idea of computing a perfect matching in an unweighted bipartite graph $G = (L \cup R, E)$ - As long as there exists a free vertex in L, find an augmenting path and update the matching. By Berge's Theorem, this process will always terminate with a perfect matching, given that one exists. Now, suppose, we determine the *augmenting path of shortest possible length* in each iteration.

   (a) (5 points) Show that, after $h \geq 0$ iterations, there will exist an augmenting path of length at most $n/(n-h)$ with respect to the current matching (Hint: Recall the proof of Berge's theorem)

   (b) (3 points) Show that the total length of all augmenting paths found by the above algorithm over n iterations is $\mathcal{O}(n \ln n)$

(a) Let M_h be the matching after iteration h.
Let M^* be a perfect matching for the entire graph.

+2

Obs. In $M_h \oplus M^*$, there are exactly $n-h$ augmenting paths, each of which begins at a free vertex in L and ends at a free vertex in R. These paths are vertex disjoint and hence total length is at most n. Hence $\exists$ one path of length $\leq n/(n-h)$ (Pigeon Hole) +1

+2

(b) Since you are picking shortest aug. path, the length at iteration h is $\leq \frac{n}{n-h}$

$\Rightarrow$ Total $= n \sum_{h=1}^{n} \frac{1}{n-h} \approx \mathcal{O}(n \cdot \ln n)$

2. As discussed in class, the *weighted vertex cover problem* is defined as follows. We are given a graph $G = (V, E)$ with non-negative weights w_v on vertices $v \in V$. The task is to pick the minimum weight subset $V' \subseteq V$ such that for every edge $(uv) \in E, \{u, v\} \cap V' \neq \emptyset$. The following is a natural LP relaxation of the problem

$$\text{minimize} \sum_{v \in V} w_v \cdot x_v$$
$$\text{subject to: } x_u + x_v \geq 1, \forall (uv) \in E$$
$$x_v \geq 0, \forall v \in V$$

(a) (7 points) Show that any *extreme point* of the above polytope has the property that $x_v \in 0, 1/2, 1$ for all $v \in V$ (Hint : Try to show a contradiction, what happens if some extreme point does not have the above property ?)

(b) (3 points) Design a 2-approximation algorithm for weighted vertex cover that runs in polynomial time.

(a) Let x be an extreme point solution that does not satisfy the above property (for contrad.)

Define,

$$\left.\begin{aligned} V_{<1/2} &= \{v \in V \mid 0 < x_v < 1/2\} \\ V_{>1/2} &= \{v \in V \mid 1/2 < x_v < 1\} \end{aligned}\right\} +2$$

Now, define two other vectors x^1 and x^2 as follows

$$x^1_v = \begin{cases} x_v, & \text{if } x_v \in \{0, 1/2, 1\} \\ x_v + \epsilon, & \text{if } v \in V_{<1/2} \\ x_v - \epsilon, & \text{if } v \in V_{>1/2} \end{cases} \quad +1$$

$$x^2_v = \begin{cases} x_v, & \text{if } x_v \in \{0, 1/2, 1\} \\ x_v - \epsilon, & \text{if } v \in V_{<1/2} \\ x_v + \epsilon, & \text{if } v \in V_{>1/2} \end{cases} \quad +1$$

1. Both x^1_v, x^2_v lie inside the polytope (for ϵ small enough) +2

2. $x_v = \frac{1}{2}(x^1_v + x^2_v)$ and hence cannot be an extreme point. +1

(b) 1. Solve the LP : x^* 2. if $x_v \in \{1/2, 1\}$, include v in the solution. For any edge (uv), $x^*_u + x^*_v \geq 1$,

+2

$\Rightarrow$ one of $x_u^*, x_v^* \geq 1/2 \Rightarrow$ Solution is feasible. $\}$ +1
Cost can only increase by a 2-factor.

3. (8+2 points) Given a graph $G = (V, E)$, a set $D \subseteq V$ is dominating if for every vertex $v \in V$, either $v \in D$ or some neighbor of v is in D. Suppose the minimum degree of any vertex in G is δ. Design a simple randomized algorithm to find a set D which is dominating with a probability of at least $1 - 1/n^c$ (c is some constant). What is the expected size of D ? (Hint : Note that your solution might not be a dominating set, but you need to prove that this happens with a probability of only $1/n^c$)

<u>Algorithm</u>: Pick each vertex independently w.p. $\min\left\{1, \frac{c \ln n}{\delta+1}\right\}$. (+ 4)

<u>Proof</u>:- Consider any vertex v. The probability that this vertex is neither included in D, nor has any of its neighbors is at most $\left(1, \frac{c \ln n}{\delta+1}\right)^{\delta_v+1}$, where δ_v is the degree of v and hence $\delta_v \geq \delta$. Now,

$$\left(1 - \frac{c \ln n}{\delta+1}\right)^{\delta_v+1} < e^{-\frac{c \ln n}{\delta+1}(\delta_v+1)}$$

$$\leq e^{-\frac{c \ln n}{\delta+1}(\delta+1)} \quad [\because \delta_v \geq \delta]$$

$$= e^{\ln n^c} = 1/n^c \quad (+4)$$

Hence, by union bound, the probability of any vertex not being "covered" by D is $\leq 1/n^{c-1}$ ☑

Expected size of D is $\leq \frac{c \cdot n \cdot \ln n}{\delta+1}$ as desired (+2)

[I have given 6+1 for "almost" correct

probabilities]

4. Consider a scenario where you are seeing elements 'on the fly' and you want to keep an approximate count of how many elements you have seen, but you have limited memory. Here is a way of maintaining an approximate counter. Start with $X = 0$. When an element arrives, increment X by 1 with probability 2^{-X}. When queried, return $N := 2^X - 1$.

   (a) (6 points) Suppose the actual count is n, show that $E[N] = n$.

   (b) (6 points) Assume $Var(N) = n(n-1)/2$.(You do not need to prove this). Imagine that you have k of these counters which are working *independently* and you have estimates $N_1, N_2, \cdots N_k$ from them. Define $\widehat{N} = \frac{1}{k}\sum_{i=1}^{k} N_i$. Show that $\mathbf{Pr}[\widehat{N} \notin (1 \pm \varepsilon)n] \leq \frac{1}{2\varepsilon^2 k}$.

(a) This can be proved by induction on n.
Base case $(n=1)$: $\mathbb{E}[2^X] = 2^1 \cdot \mathbb{P}r[2^X = 2^1]$

$$= 2 \cdot \mathbb{P}r[X=1] = 2 \cdot \frac{1}{2^0} = 2.$$

Hence $\mathbb{E}[N] = \mathbb{E}[2^X - 1] = 1.$

S'pose this is true for $n = i$. Let us consider $n = i+1$. Let us also introduce a notation X_i to denote value of X after ith element is seen.

$$\mathbb{E}[2^{X_{i+1}}] = \sum_{j=0}^{\infty} 2^j \cdot \mathbb{P}r[X_{i+1} = j]$$

$$= \sum_{j=0}^{\infty} 2^j \cdot \left(\mathbb{P}r[X_i = j] \cdot \left(1 - \frac{1}{2^j}\right) + \mathbb{P}r[X_i = j-1] \cdot \left(\frac{1}{2^{j-1}}\right) \right) \quad (i)$$

$$= \sum_{j=0}^{\infty} 2^j \cdot \left(\mathbb{P}r[X_i = j]\right) + \sum_{j=0}^{\infty} \left(2 \mathbb{P}r[X_i = j-1] - \mathbb{P}r[X_i = j]\right).$$

Observe that the first term is $\mathbb{E}[2^{X_i}]$ which is $i+1$ by induction hypothesis. The second term, if you look at it well is just

$$\sum_{j=0}^{\infty} \Pr[X_i = j] = 1 \text{ (by definition of probability dist.)}$$

Why is (i) true? Well X_{i+1} can be j under two conditions: either $X_i = j$ and no increment happened at $(i+1)^{th}$ iteration or $X_i = j-1$ and increment did happen at $(i+1)$.

Hence, $\mathbb{E}[2^{X_{i+1}}] = i+2$

$\Rightarrow \mathbb{E}[2^{X_{i+1}} - 1] = i+1$ as desired.

(b) :- Obs. 1 :- $\mathbb{E}[\hat{N}] = n$ (by linearity)

$\text{var}[\hat{N}] = \frac{1}{k} \frac{n(n-1)}{2}$ [using independence]

[$\text{var}[\hat{N}] = \frac{1}{k^2} \sum_{i=1}^{k} \text{var}(N_i) = $ etc. . .]

Now applying Chebyshev,

$$\Pr\left[\,|\hat{N} - \mathbb{E}[\hat{N}]| > \epsilon \cdot n\right] \le$$

$$\frac{\epsilon \cdot n \cdot \sqrt{\text{var}(\hat{N})}}{\sqrt{\text{var}(\hat{N})}} \le \frac{\text{Var}(\hat{N})}{\epsilon^2 \cdot n^2}$$

$$= \frac{n(n-1)}{2k\epsilon^2 n^2} \le \frac{1}{2k\epsilon^2}$$

[Rubric is highly subjective]