CMPUT 675: Approximation Algorithms **Winter 2005**

Lecture 26: April 15

Lecturer: Mohammad R. Salavatipour *Scribe: Zhuang Guo*

Today we complete the proof of $\Omega(\log n)$-hardness of set cover. In the last lecture, we introduced the following theorem:

Theorem 26.1 *There is a reduction mapping an instance ϕ of SAT to an instance $\mathcal{L}(G(V, W, E), M = [7^k], N = [2^k], \{\Pi_{v,w}\})$ of label cover, such that:*

- *If ϕ is a yes instance, then $opt(\mathcal{L}) = 1$.*
- *If ϕ is a no instance, then $opt(\mathcal{L}) \leq 2^{-\delta k}$, for some $\delta > 0$, $|\mathcal{L}| = n^{O(k)}$.*

To prove the hardness of approximation of set cover, we need the following set system.

Definition 26.2 *A set system $(U, C_1, \cdots, C_m, \overline{C_1}, \cdots, \overline{C_m})$ with parameters m and l, where U is the a universe of elements of size $O(l \cdot logm \cdot 2^l)$ and $C_1, \cdots, C_m$ are subsets of U. This set system has the property that any collection of $\leq l$ subsets from $C_i's$ that cover U must contain a set and its complement.*

There are explicit construction of such (m, l)-set systems. There are also easy probabilisitic constructions. Consider a label cover instance $\mathcal{L}(G(V, W, E), M = [7^k], N = [2^k], \{\Pi_{v,w}\})$. We can assume $|V| = |W|$ (e.g. if not, we can create copies of the vertices in V with the same neighbours). We build an instance of set cover $\mathcal{S}$ such that:

- If $opt(\mathcal{L}) = 1$, then $opt(S) \leq |V| + |W|$.
- If $opt(\mathcal{L}) < \frac{2}{l^2}$, then $opt(\mathcal{S}) > \frac{l}{16}(|V| + |W|)$.

Consider a set system with $m = N = 2^k$ and l to be specified later. For every edge $e = (v, w) \in G$, we have a (disjoint) (m, l)-set system with universe U_e. Let $C_1^{vw}, \cdots, C_{N=m}^{vw}$ be the subsets of U_e. The union of all $U_e's$ (for all the edges e) is the universe of the set cover instance, denoted as

$$U = \bigcup_{(v,w) \in G} U_{vw}.$$

Now we define the subsets in our set cover instance. For every $v \in V$ $(w \in W)$ and every label $i \in [2^k]$ $(j \in [7^k])$, we have a set

$$S_{v,i} = \bigcup_{w:(v,w) \in E} C_i^{vw} \qquad S_{w,j} = \bigcup_{v:(v,w) \in E} \overline{C_{\Pi_{vw}(j)}^{vw}}$$

This completes the construction of $\mathcal{S}$ from $\mathcal{L}$.

Lemma 26.3 *If $opt(\mathcal{L}) = 1$, then $opt(\mathcal{S}) \leq |V| + |W|$.*

Proof: Consider an optimal labeling $l : V \to [2^k], W \to [7^k]$ for $\mathcal{L}$. Because it is covering every edge $(v, w) \in E$, $\Pi_{vw}(l(w)) = l(v)$. This labeling defines a label for every vertex and every pair of vertex/label corresponds to a set in $\mathcal{S}$. From $C_{l(v)}^{vw} \subseteq S_{v,l(v)}$ and $\overline{C_{l(v)}^{vw}} = \overline{C_{\Pi_{vw}(l(w))}^{vw}} \subseteq S_{w,l(w)}$, we have that $S_{v,l(v)} \cup S_{w,l(w)} \supseteq U_{vw}$. Because all U_e's for $e \in E$ are covered, U is covered. So we have a set cover of size $|V| + |W|$. ■

Lemma 26.4 *if opt($\mathcal{S}$) $\leq \frac{l}{16}(|V|+|W|)$, then opt($\mathcal{L}$) $\geq \frac{2}{l^2}$.*

Proof: From the set cover solution, we assign labels (maybe more than one label) to the vertices. If $S_{v,i}$ is in the solution, v gets label i. Since there are at most $\frac{l}{16}(|V|+|W|)$ sets and $|V|+|W|$ vertices, the average number of labels per vertex is $\leq \frac{l}{16}$. We discard vertices with more than $\frac{l}{2}$ labels. Afterwords, at most $\frac{|V|}{4}$ vertices from each side of V and W are discarded. Let V' and W' be the vertices remaining in V and W respectively. Then, $|V'| > \frac{3}{4}|V|$ and $|W'| > \frac{3}{4}|W|$. Pick an edge $e = (v, w)$ from G randomly. Then $\Pr[v \in V' \textit{ and } w \in W'] \geq 1 - (\frac{1}{4}+\frac{1}{4}) = \frac{1}{2}$. This means at least half of the edges of G are between V' and W'. Let $T_v = \{S_{v,i} : i \textit{ is a label of } v\}$ and $T_w = \{S_{w,j} : j \textit{ is a label of } w\}$. We have $|T_v| \leq \frac{l}{2}$ and $|T_w| \leq \frac{l}{2}$. Note that sets in $T_v \cup T_w$ cover U_{vw}. To be more precise, sets in $X_1 = \{C_i^{vw} : i \textit{ is a label of } v\} \cup X_2 = \{\overline{C_{\Pi_{vw}(j)}^{vw}} : j \textit{ is a label of } w\}$ cover universe U_{vw} ($|X_1 \leq \frac{l}{2}$ and $|X_2 \leq \frac{l}{2}$). Because U_{vw} is covered by at most l sets (i.e. $|X_1|+|X_2| \leq l$), there must be a set $C_i^{vw} \in X_1$ and $\overline{C_{\Pi_{vw}(j)}^{vw}} \in X_2$, such that they are complement, i.e. $i = \Pi_{vw}(j)$. Because we pick labels of v and w randomly, with probability $(\frac{2}{l})^2 = \frac{4}{l^2}$ we have set C_i^{vw} for v and $\overline{C_j^{vw}}$ for w, i.e. the labels i for v and j for w cover edge $e \in E$. Thus the expected fraction of edges between V' and W' that are covered is $\geq \frac{4}{l^2}$. Therefore, at least a fraction of $\frac{2}{l^2}$ of edges of G are covered. ■

This lemma is equivalent to saying that if $opt(\mathcal{L}) < \frac{2}{l^2}$ then $opt(\mathcal{S}) > \frac{l}{16}(|V|+|W|)$.

Let $l \in \Theta(2^{\frac{\delta k}{2}})$. Then, $l^2 \in \Theta(2^{\delta k})$. We get a hardness of $\Omega(l)$ for $\mathcal{S}$. The size of $\mathcal{S}$ is $n^{O(k)} \cdot O(l \cdot \log m \cdot 2^l)$. If $k = c \log\log n$ for sufficiently large c, $l = O(2^{O(\log\log n)}) \geq \log n \log\log n$. $\log|\mathcal{S}| = O(\log\log n \cdot \log n + \log l + \log\log\log n + l) = \Theta(l)$.

We have the following hardness result for set cover:

Theorem 26.5 *Unless $NP \subseteq DTIME(n^{O(\log\log n)})$, set cover has no $\Omega(\log n)$-approximation algorithm.*