Lecture #3: Hardness for Max-Coverage

We prove that 1-vs-$(\frac{3}{4}+\varepsilon)$ decision problem for Max Coverage is NP-hard.

i.e. An algo for ↗ that runs in poly time $\Rightarrow$ P = NP.

Build on a theory of hardness of approximations

Cook-Levin theorem says 1-vs-$(1-1/m)$ decision problem for Max-3SAT is NP-hard

$\Downarrow$

PCP theorem says 1-vs-$(1-\varepsilon)$ " " for Max3SAT — for some $\varepsilon > 0$

Raz's Parallel Repetition Thm says 1-vs-η " " for Max-Label-Cover — for every constant $\eta > 0$.

This is where we catch the story.

Max-Label-Cover: Bipartite graph $G = (U, V, E)$

~~Labels~~ Keys K for vertices in U, Labels L for vertices in V.

Each edge has a "projection" function $\pi_{uv} : L \to K$. (π_{uv})

Objective: assign keys to U, labels to V. i.e. $f: U \to K$, $V \to L$

s.t. $\#\{(u,v) \mid f(u) = \pi_{uv}(f(v))\}$ is maximized.

Thm: $\forall \eta > 0$, $\exists q = \mathrm{poly}(1/\eta)$ s.t. ~~the~~ the 1-vs-η decision problem for Max Label Cover is NP-hard with $|K|, |L| \le q$.

Moreover: $|U| = |V|$ for these instances, and bipartite graph is regular.

Reduction:

Take instances of MaxLC $\mathcal{G}$ $\longrightarrow$ instances $\mathcal{H}$ of Max Coverage s.t.

(Completeness) $\quad val(\mathcal{G}) = 1 \quad \Rightarrow \quad val(\mathcal{H}) = 1$

Soundness. $\quad val(\mathcal{G}) < \eta \quad \Rightarrow \quad val(\mathcal{H}) < \frac{3}{4} + \varepsilon. \qquad \varepsilon = \varepsilon(\eta).$

How?

① Elements: Q_{uv} cube for each edge. $\quad Q_{uv} = \{0,1\}^K.$

$K = |K| = \#$ keys.

$\Rightarrow |E|.2^K$ nodes. $\qquad < |L|.$

~~|E|~~ $d.2^K$

$|\text{Elements}|$ "

② Sets: for each $u \in \mathcal{U}$, $\forall a \in K$,

$$S_{u,a} = \bigcup_{v \sim u} \{x \in Q_{uv} \mid x_a = 0\}.$$

"left half of cubes along the a^{th} ~~compon~~ dimension".

$|S_{u,a}| = 2^{K-1} d$

$$S_{v,\alpha} = \bigcup_{u \sim v} \{x \in Q_{uv} \mid x_{\pi_{uv}(\alpha)} = 1\}$$

$\Rightarrow K.|\mathcal{U}| + \ell.|V| = O(q)(|\mathcal{U}| + |V|)$ sets.

③ Target $T = |\mathcal{U}| + |V| = 2|\mathcal{U}|$ $\qquad$ ($\Rightarrow$ is exact cover)

Note: Completeness! if $\mathcal{G}$ satisfiable $\Rightarrow$ $\mathcal{H}$ has coverage sol^n of value 1.

Pf: sps. f is labeling that sat's all edges.

then choose $S_{u,f(u)}$ and $S_{v,f(v)}$ sets, one per vertex.

Note: $\forall u,v$ edge $f(u) = \pi_{uv}(f(v))$ so $S_{u,f(u)}$ covers all $x \in Q_{uv}$ with 0 along that dimension

and $S_{v,f(v)}$ covers the other Q_{uv} points.

Q: What about soundness?

Theorem: if $val(H) \geq \frac{3}{4}+\varepsilon$ then $val(\mathcal{G}) > \varepsilon^3/2000 := \eta(\varepsilon)$

[Contrapositive of Claimed Soundness]

Main Idea: suppose $\mathcal{S}^*$ is a solution of value $\geq \frac{3}{4}+\varepsilon$.

- **And assume**: $\mathcal{S}^*$ chooses exactly one set per vertex. ie. both $S_{u,a}$ & $S_{u,b}$ not in $\mathcal{S}^*$.

big if

then consider the map f s.t. $S_{u,f(u)} \in \mathcal{S}^*$, $S_{v,f(v)} \in \mathcal{S}^*$ for each u,v.

Claim: value of this map $f \geq \Omega(\varepsilon)$.

Why? for each unsatisfied edge u,v,

$S_{u,f(u)}$ and $S_{v,f(v)}$ cover $\frac{3}{4}$ of cube.

So in order to get $\frac{3}{4}+\varepsilon$ coverage,

$\Omega(\varepsilon)$ fraction of edges must be ~~covered~~ satisfied! ☺

$$\left(p \cdot 1 + (1-p)\frac{3}{4} = \frac{3}{4}+\varepsilon \Rightarrow p = 4\varepsilon\right)$$

But solution could choose multiple sets per vertex (or none at all)

- How to show that $\exists$ good value labeling?
- **Hope1**: on average one set chosen per vertex, so maybe not so bad.
- **Hope2**: picking many sets per cube (that are not complementary) gives diminishing returns.

Now we make these ideas precise

Recall $\mathcal{S}^* =$ ~~[scribble]~~ coverage solution of value $\frac{3}{4}+\varepsilon$.

$Suggest(u) = \{a \in K \mid S_{u,a} \in \mathcal{S}^*\}$ $\quad$ $Suggest(v) = \{\alpha \in L \mid S_{v,\alpha} \in \mathcal{S}^*\}$.

$Suggest(u,v) = Suggest(u) \cup Suggest(v)$.

__Fact__: $\underset{(u,v)}{E}[|Suggest(u,v)|] = 2$. $\quad$ // use fact that bip graph is regular.

__Pf__: $\underset{uv}{E}[|Suggest(uv)|] = \underset{u}{E}[|Suggest(u)|] + \underset{v}{E}[|Suggest(v)|] = \frac{1}{|U|}[\sum |Suggest(u)| + |Suggest(v)|]$

regularity $\quad$ $|U|=|V|$ $\quad = \frac{|U|+|V|}{|U|} = 2.$ ∎

__Random decoding__: pick random element from $Suggest(u)$ to be $f(u)$

if $Suggest(u) = \emptyset$, $f(u) =$ arbitrary.

$$\boxed{\text{Claim 1: } E_f[Coverage(f)] \geq \varepsilon^3/2000. \quad \Rightarrow \exists \text{ good labeling for } \mathcal{G}.}$$

__Def__: (u,v) __consistent__ if $\exists\, a \in Suggest(u)$, $\alpha \in Suggest(v)$ s.t. $\pi_{uv}(\alpha) = a$.

__Def__: u __sparse__ if $|Suggest(u)| \leq 10/\varepsilon$.

(u,v) __sparse__ if u __sparse__ and v __sparse__.

__Def__: (u,v) good if ~~[scribble]~~ if sparse and consistent.

$$\boxed{\text{Claim 2: } \geq \varepsilon/20 \text{ fraction of edges are good.}}$$

__Pf of Claim 1 (assuming claim 2).__

$\varepsilon/20$ fraction are good. for each, $f(u), f(v)$ chosen to be such that $\pi_{uv}(f(v)) = f(u)$

w.p. $(\varepsilon/10)\cdot(\varepsilon/10)$. $\quad \Rightarrow$ total ~~[scribble]~~ E[fraction of sat. edges] $\geq (\varepsilon^3/2000)$.

by sparsity

Now to prove Claim 2 (that $\varepsilon/20$ fraction of edges good).

Sps not. then

$$\Pr[(u,v)\text{ consistent}] \le \underbrace{\Pr[(u,v)\text{ good}]}_{\le \varepsilon/20} + \Pr[u \text{ not sparse}] + \Pr[v \text{ not sparse}]$$

$\le \varepsilon/20$ ← by assumption (for contradiction)

$\le \varepsilon/10$, $\le \varepsilon/10$ ← Markov's inequality since expect 1 key/label per vertex.

$< 5\varepsilon/20 = \varepsilon/4.$

This means $< \varepsilon/4$ ~~edges~~ fraction of edges are ~~good~~ consistent (which give perfect coverage for cube Q_{uv}).

So at least $3/4+\varepsilon - \varepsilon/4 = 3/4 + \frac{3\varepsilon}{4}$ coverage comes from inconsistent edges.

And $> (1-\varepsilon/4)$ fraction of edges are inconsistent.

Now:

- ~~Claim~~ Fact 1: if edge (u,v) has ~~t labels, but not~~ $|\text{Suggest}(u,v)| = t$, but not consistent then coverage of cube $Q_{uv} = 1-2^{-t}$.

- Fact 2 (Concavity Fact). if $t_1+t_2+\dots+t_r = r\cdot\bar{t}$ then $\max_{t_1,t_2..t_r} \sum_{i=1}^{r}(1-2^{-t_i}) = r\cdot(1-2^{-\bar{t}})$.

Pf: $\frac{1}{r}\sum_i 2^{-t_i} \underset{\text{AM-GM}}{\ge} 2^{-\sum_i t_i/r} = 2^{-\bar{t}}$. ∎

Pf of Claim 2 (cont'd).

An average inconsistent edge has $\frac{2}{(1-\varepsilon/4)} \lesssim 2+\varepsilon$ labels.

And gets coverage $\frac{3}{4}+\frac{3\varepsilon}{4}$.

By Fact 2 (concavity), coverage using $2+\varepsilon$ labels on average is at most

$$1-2^{-(2+\varepsilon)} = 1-\frac{1}{4}\cdot 2^{-\varepsilon} \le 1-\frac{1}{4}(1-\varepsilon\ln 2)$$

AG's favorite inequality :)

$$\le \frac{3}{4}+\frac{\varepsilon\ln 2}{4} < \frac{3}{4}+\frac{3}{4}\varepsilon. \quad \text{Contradiction!}$$

:)

To recap:

Hardness of Label Cover says: NP-hard to distinguish between

$$\mathrm{val}(G) = 1 \quad \text{vs.} \quad \mathrm{val}(G) = \eta$$

This reduction says: if we can distinguish $\mathrm{val}(H) = 1$ vs $\mathrm{val}(H) \geq \frac{3}{4} + \varepsilon$ in polytime

H = Max Coverage instance

$\Rightarrow$ can ~~[illegible]~~ distinguish of

$$\mathrm{val}(G) = 1 \quad \text{vs.} \quad \mathrm{val}(G) = \Theta(\varepsilon^3).$$

$\Rightarrow$ 1-vs-$(\frac{3}{4}+\varepsilon)$ decision problem is NP-hard

$\Rightarrow$ cannot approx Max Coverage to better than $\frac{3}{4} = 0.75$

Fancier Reduction gives:

cannot approx Max Coverage to better than $(1 - 1/e) \cong 0.63\ldots$

Uses a k-partite version of Label Cover. [Feige 95]

A different extension:— hardness of $\Omega(\log n)$ for Set Cover. [Lund-Yannakakis, Feige, Dinur-Steurer]

~~[illegible]~~