Set Cover:

Input: - Universe U with $|U| = n$

- Family of sets $S_1, S_2, \dots, S_m$ with each $S_i \subseteq U$

Feasible solution: $I \subseteq \underbrace{[m]}_{\{1,2,\dots,m\}}$ s.t. $\bigcup_{i \in I} S_i = U$

Objective: $\min |I|$

Generalizes Vertex Cover!

- $U = E$ - $S_v = N(v)$

neighborhood of v: $\{e \in E : v \in e\}$

Greedy Algorithm:

Init $I = \emptyset$, $X = U$

while ($X \neq \emptyset$) {

Let $i \in [m]$ maximizing $|X \cap S_i|$

Add i to I, remove S_i from X

}

return I

<u>Thm</u>: If OPT uses k sets, greedy uses $\leq k(\ln \frac{n}{k} + 1)$

<u>Cor</u>: Greedy is $O(\log n)$-approximation

<u>Cor</u>: If $|S_i| \leq B$ $\forall i \in [m]$, then $O(\log B)$-approx

<u>Pf</u>: $k \geq \frac{n}{B} \Rightarrow$ greedy $\leq k(\ln B + 1)$

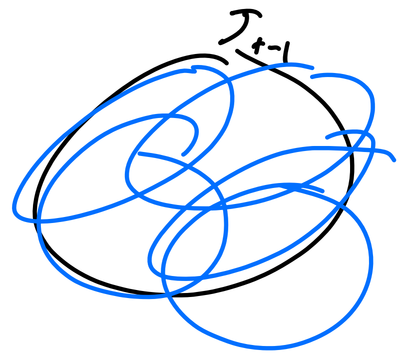
<u>Pf of thm</u>:

- Let I_t be sets selected in first t iterations

$J_t = U \setminus \left(\bigcup_{i \in I_t} S_i\right) =$ el'ts uncovered after t iterations

$n_t = |J_t| = n - \left|\bigcup_{i \in I_t} S_i\right|$

$\Rightarrow I_0 = \emptyset$, $J_0 = U$, $n_0 = n$



<u>Claim</u>: $n_t \leq (1 - \frac{1}{k}) n_{t-1}$ $\forall t$

<u>Pf</u>: Consider iteration t

OPT covers U with k sets

$\Rightarrow$ OPT covers J_{t-1} with $\leq k$ sets

$\Rightarrow \exists$ set covering $\geq \frac{n_{t-1}}{k}$ el'ts of J_{t-1}

$\Rightarrow$ greedy in iteration t covers $\geq \frac{n_{t-1}}{k}$ elements of J_{t-1}

$$\Rightarrow n_t \leq n_{t-1} - \frac{n_{t-1}}{k} = \left(1 - \frac{1}{k}\right) n_{t-1}$$ ✓

Claim $\Rightarrow$ $$n_t \leq \left(1-\frac{1}{k}\right)^t n_0 = \left(1-\frac{1}{k}\right)^t n$$

$\Rightarrow$ when $t = k \ln \frac{n}{k}$:

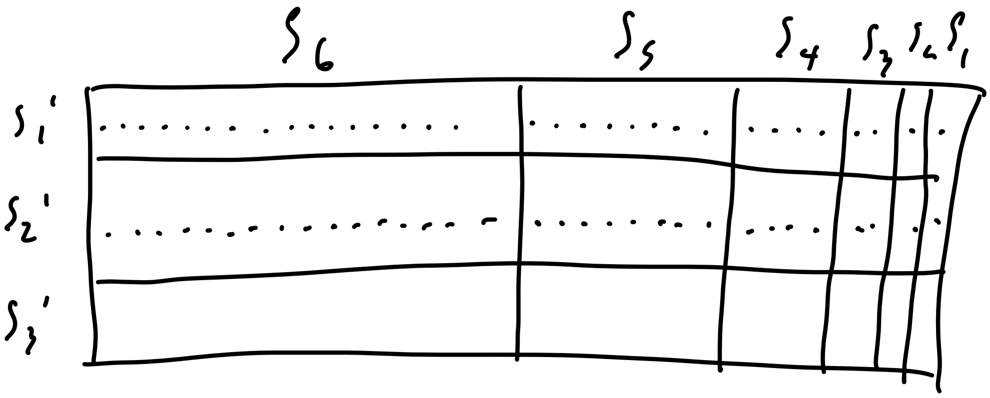
$$n_t \leq \left(1-\frac{1}{k}\right)^{k \ln \frac{n}{k}} n \leq \left(e^{-\frac{1}{k}}\right)^{k \ln \frac{n}{k}} n$$

$(1+x) \leq e^x \quad \forall x \in \mathbb{R}$

$$= e^{-\ln \frac{n}{k}} n = \frac{k}{n} \cdot n = k$$

$\Rightarrow$ greedy takes $\leq k \ln \frac{n}{k} + k = k\left(\ln \frac{n}{k} + 1\right)$ iterations

Is analysis tight, or is greedy even better?



Thm [Dinur-Steurer '14]: Unless $P = NP$, there is a $c \ln n$-approx for Set Cover for any constant $c < 1$

Weighted Set Cover:

Each set S_i also has cost c_i

$\min \sum_{i \in I} c_i$ (instead of $|I|$)

Greedy:

maximize "bang-for-the-buck": $\frac{|X \cap S_i|}{c_i}$ (density of S_i)

Thm: Greedy is an $H_n = \Theta(\log n)$-approximation

$H_n = 1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \dots + \frac{1}{n}$

Pf: I_t, J_t, n_t as before.

OPT optimal solution, $C^* = \sum_{i \in OPT} c_i$

g_t index of set picked by greedy in iteration t

Claim: $\frac{|J_{t-1} \cap S_{g_t}|}{c_{g_t}} \geq \frac{n_{t-1}}{c^*}$

Pf: WTS $\exists$ $i \in OPT$ s.t. $\frac{|J_{t-1} \cap S_i|}{c_i} \geq \frac{n_{t-1}}{c^*}$

(greedy choice has at least as good density)

Sps false $\Rightarrow$ $\frac{c_i}{|J_{t-1} \cap S_i|} > \frac{c^*}{n_{t-1}}$ $\forall i \in OPT$

$$\Rightarrow c^* = \sum_{i \in OPT} c_i = \sum_{i \in OPT} \frac{c_i}{|J_{t-1} \cap S_i|} |J_{t-1} \cap S_i|$$

$$> \sum_{i \in OPT} \frac{c^*}{n_{t-1}} |J_{t-1} \cap S_i| = \frac{c^*}{n_{t-1}} \sum_{i \in OPT} |J_{t-1} \cap S_i|$$

$$\geq \frac{c^*}{n_{t-1}} |J_{t-1}| \quad \text{(OPT covers } J_{t-1}\text{)}$$

$$= c^*$$

$\Rightarrow \Leftarrow$

Let $x_t = |S_{g_t} \cap J_{t-1}|$ be # el't covered by greedy in iteration t

$\Rightarrow$ claim is $\dfrac{x_t}{c_{g_t}} \geq \dfrac{n_{t-1}}{c^*}$

$$\Rightarrow c_{g_t} \leq x_t \frac{c^*}{n_{t-1}}$$

$$c(\text{greedy}) = \sum_{t=1}^{k} c_{g_t} \leq \sum_{t=1}^{k} x_t \frac{c^*}{n_{t-1}}$$

$$= c^* \frac{x_1}{n} + c^* \frac{x_2}{n-x_1} + c^* \frac{x_3}{n-x_1-x_2} + \dots + c^* \frac{x_k}{n-x_1-\dots-x_{k-1}}$$

$$= c^* \Bigg(\underbrace{\frac{1}{n} + \frac{1}{n} + \dots + \frac{1}{n}}_{x_1} + \underbrace{\frac{1}{n-x_1} + \frac{1}{n-x_1} + \dots + \frac{1}{n-x_1}}_{x_2} + \dots + \underbrace{\frac{1}{n-x_1-\dots-x_{k-1}} + \dots + \frac{1}{n-x_1-\dots-x_{k-1}}}_{x_k} \Bigg)$$

$$\leq c^* \left(\frac{1}{n} + \frac{1}{n-1} + \frac{1}{n-2} + \dots \frac{1}{n-x_1+1} + \frac{1}{n-x_1} + \frac{1}{n-x_1-1} + \dots + \frac{1}{n-x_1-x_2+1} + \frac{1}{n-x_1-x_2} + \dots + 1 \right)$$

$$= c^* H_n$$