Cormen section 35.3

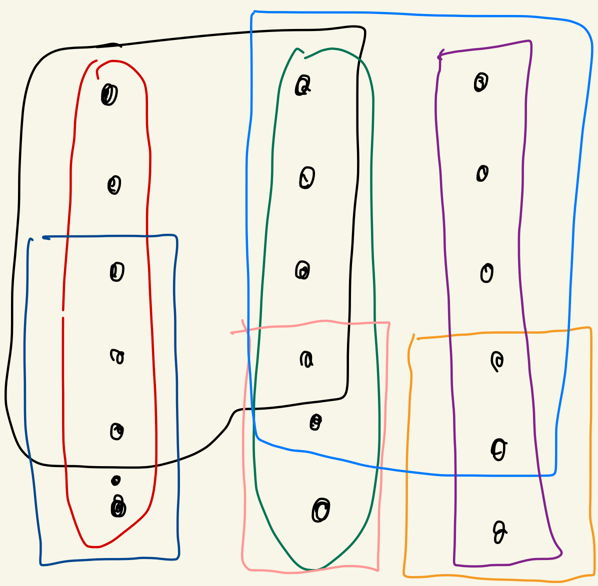
Set cover Problem:

<u>Input</u> $(X, \mathcal{F})$ where X is a finite set and $\mathcal{F}$ is a family of subsets of X

<u>Output:</u> A subfamily $\mathcal{F}' \subseteq \mathcal{F}$ s.t. $\bigcup_{S \in \mathcal{F}'} S = X$

and $|\mathcal{F}'|$ is minimized

<u>Example</u>



cover X

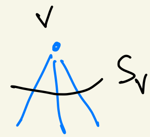
no two sets cover X

so optimum $= 3$

Optimization version of Vertex cover is a special case of Set cover

Given $G=(V,E)$ construct instance $(X,\mathcal{F})$ of set cover

$X = E$, $\mathcal{F} = \{S_v \mid v \in V\}$ where



Decisission version of Vertex cover is NPC so the decission version of Set cover [input $(X,\mathcal{F},K)$
Question: is there a solution $\mathcal{F}'$ with $|\mathcal{F}'| \le K$]
is NP-complete

Greedy set cover: (algorithm)

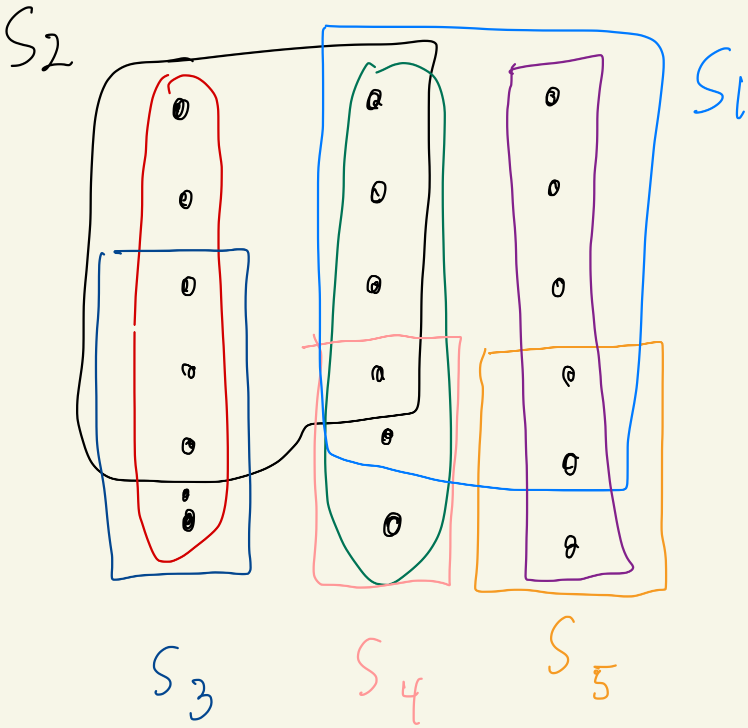
$$Z \leftarrow X;\ \mathcal{F}' \leftarrow \emptyset$$

while $Z \neq \emptyset$
  pick $S \in \mathcal{F}$ s.t. $|S \cap Z|$ is maximized
  $Z \leftarrow Z \setminus S$
  $\mathcal{F}' \leftarrow \mathcal{F}' \cup \{S\}$
end

assumption $\bigcup_{S \in \mathcal{F}} S = X$

Greedy set cover finds a solution

How good is it?



The algorithm of und 5 sets so not optimum

Theorem The Greedy set cover algorithm is an $H(|X|)$-approximation algorithm

Recall input is $(X, \mathcal{F})$

and $H(n) = \sum_{i=1}^{n} \frac{1}{i}$

Proof:

Idea: each set $S_i \in \mathcal{F}'$ contributes one to $|\mathcal{F}'|$
distribute this over the new elements covered by S_i:

Let $\mathcal{F}' = \{S_1, S_2, \ldots S_{|\mathcal{F}'|}\}$

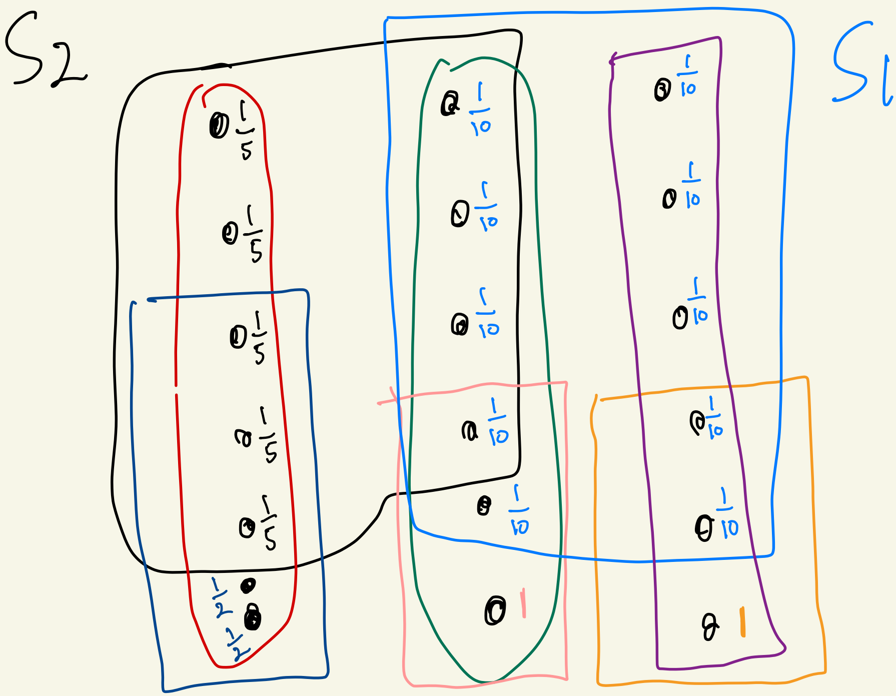
be the output from the algorithm, when S_i is chosen in step i.

For each $x \in X$

let S_i be the first set in $\mathcal{F}'$ which contains x

and assign x the weight $c_x = \frac{1}{|S_i \setminus (S_1 \cup \ldots \cup S_{i-1})|}$

$|S_i \setminus (S_1 \cup \ldots \cup S_{i-1})|$ = # elements covered for the first time by S_i



Now

$$(*)\quad |\mathcal{F}'| = \sum_{x \in X} c_x \le \sum_{S \in \mathcal{F}^*} \sum_{x \in S} c_x \qquad \mathcal{F}^* \text{ optimal sol}$$

Claim

$$\sum_{x \in S} c_x \le H(|S|) \quad \forall S \in \mathcal{F}$$

Suppose this is true. Then $(*)$ becomes

$$\begin{aligned} |\mathcal{F}'| \le \sum_{S \in \mathcal{F}^*} \sum_{x \in S} c_x &\le \sum_{S \in \mathcal{F}^*} H(|S|) \\ &\le \sum_{S \in \mathcal{F}^*} H\left(\max_{S \in \mathcal{F}} |S|\right) \\ &\le \sum_{S \in \mathcal{F}^*} H(|X|) \\ &= |\mathcal{F}^*| \, H(|X|) \end{aligned}$$

Showing that Greedy set cover is a $H(|X|)$-apx algorithm

Claim

$$\sum_{x \in S} c_x \leq H(|S|) \quad \forall S \in \mathcal{F}$$

Proof: Look at a fixed (arbitrary) $S \in \mathcal{F}$ and let k be such that $S \not\subseteq S_1 \cup \dots \cup S_{k-1}$ but $S \subseteq S_1 \cup \dots \cup S_k$

Let $u_i = |S \setminus S_1 \cup \dots \cup S_i|$ for $i = 0, 1, \dots k$

So u_i is # elements of S still uncovered after we have taken $S_1, S_2 \dots S_i$

$u_0 = |S|$ and $u_k = 0$

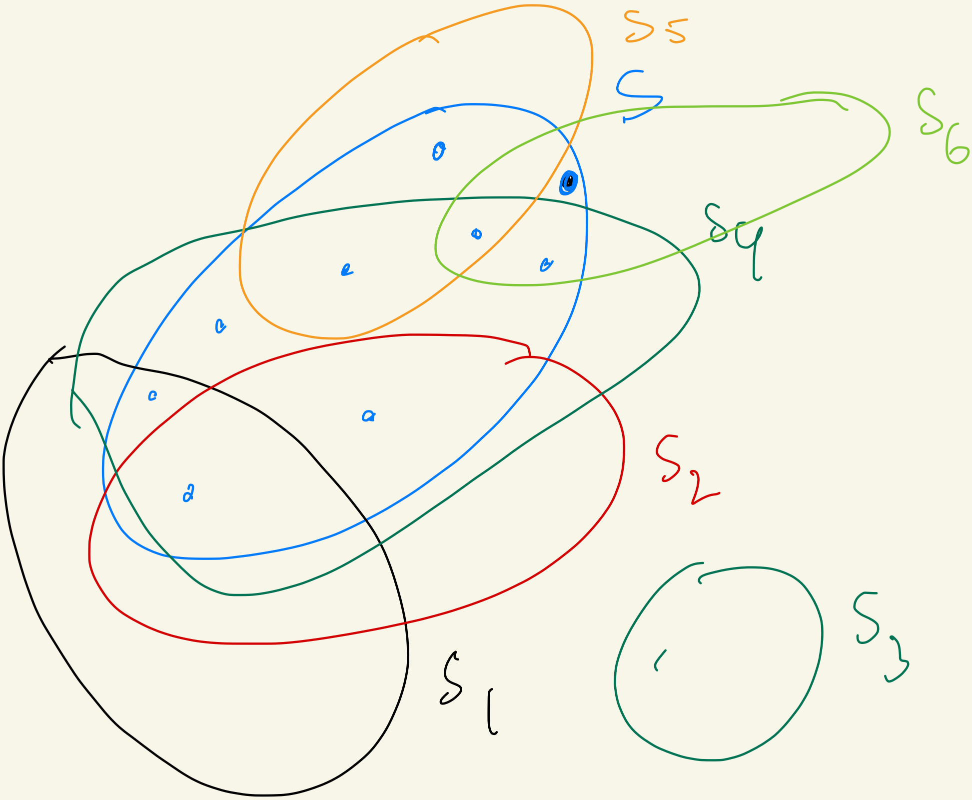
Clearly $u_{i-1} \geq u_i$ for $i = 1, 2, \dots, k$

and $u_{i-1} - u_i$ elements of S are covered for the first time by S_i

Also

$$(\square) \quad |S_i \setminus (S_1 \cup \dots \cup S_{i-1})| \geq |S \setminus (S_1 \cup \dots \cup S_{i-1})|$$

as Alg chose S_i in step i



S5
S
S6
S4
S2
S1
S3

Now $\displaystyle\sum_{x\in S} c_x = \sum_{i=1}^{k} (u_{i-1}-u_i)\cdot \frac{1}{|S_i \setminus (S_1 \cup \dots \cup S_{i-1})|}$

$\displaystyle \leq \sum_{i=1}^{k} (u_{i-1}-u_i)\cdot \frac{1}{|S \setminus (S_1 \cup \dots \cup S_{i-1})|}$ by $(\square)$

$\displaystyle = \sum_{i=1}^{k} \frac{u_{i-1}-u_i}{u_{i-1}}$

Note that for $b \geq a$ $\displaystyle H(b) - H(a) = \sum_{i=a+1}^{b} \frac{1}{i} \geq (b-a)\cdot\frac{1}{b}$

So

$$\sum_{x\in S} c_x \leq \sum_{i=1}^{k} \frac{u_{i-1}-u_i}{u_{i-1}} \leq \sum_{i=1}^{k} \left(H(u_{i-1}) - H(u_i)\right)$$

$$= H(u_0) - H(u_k) \qquad u_k = 0$$

$$= H(u_0) = H(|S|)$$

Thus $\displaystyle\sum_{x\in S} c_x \leq H(|X|)$ proving the claim

<u>NB</u> (not in book)

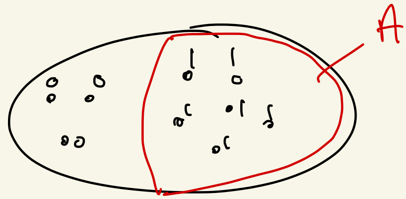
Unless $P=NP$ we cannot find a better approximation algorithm for set cover (up to constants)

So apx-factor is $\Omega(\log n)$ $\quad n = |X|$

Cormen Section 35.4

Recall the notion of an indicator random variable

$$X_A(s) = \begin{cases} 1 & \text{if } s \in A \\ 0 & \text{if } s \notin A \end{cases}$$



A randomized approximation algorithm $\mathcal{A}$ is a ρ-approximation algorithm if $\max\left(\frac{C}{C^*}, \frac{C^*}{C}\right) \leq \rho$ when C is the expected value of the solution found by $\mathcal{A}$

Max-3-SAT

Input: A 3-SAT instance $f = C_1 \wedge C_2 \wedge \ldots \wedge C_m$ over variables $x_1, x_2, \ldots, x_n$

Goal: Find a truth assignment $\varphi : \{x_1, x_2, \ldots, x_n\} \to \{T, F\}^n$ which maximizes # satisfied clauses

IF we could solve Max-3-SAT by a polynomial alg B
THEN we could decide 3-SAT in polynomial time
so $P = NP$ would hold

Randomized algorithm $\mathcal{A}$:

For $i := 1$ to n
assign x_i value T with probability ½
F —||—

Let the random variable X denote the # of satisfied clauses by $\mathcal{A}$

Then $X = X_1 + X_2 + \dots + X_m$, where

$$X_i = \begin{cases} 1 & \text{if } C_i \text{ is satisfied by } \mathcal{A} \\ 0 & \text{else} \end{cases}$$

$$pr(X_i = 1) = 1 - pr(X_i = 0) = 1 - \left(\frac{1}{2}\right)^3 = \frac{7}{8}$$

$$E(X_i) = pr(X_i = 1) = \frac{7}{8}$$

$$C = E(X) = E\left(\sum_{i=1}^{m} X_i\right) = \sum_{i=1}^{m} E(X_i) = \sum_{i=1}^{m} \frac{7}{8} = \frac{7m}{8}$$

Hence $\dfrac{C^*}{C} \le \dfrac{m}{\frac{7m}{8}} = \dfrac{8}{7}$ so

$\mathcal{A}$ is a randomized $\frac{8}{7}$-approx alg for max-3-SAT

Weighted Vertex Cover

Given $G=(V,E)$ and $w: V \to \mathbb{R}_+$

Find vertex cover U^* s.t. $w(U^*) \le w(U)$

for all vertex covers (here $w(X) = \sum_{v \in X} w(v)$)

NP version:

Given $G=(V,E)$, $w: V \to \mathbb{R}_+$ and $K \in \mathbb{R}_+$

decide if $\exists$ VC U s.t. $w(U) \le K$

This is NP-complete as Vertex Cover $\le_p$ this

(take $w(v) = 1 \ \forall v \in V$)

Formulate weighted VC as a 0-1 integer prog problem

variables $X(v)$: $X(v) = 1 \iff v$ is in U

Condition $X(u) + X(v) \ge 1 \quad \forall uv \in E$

objective $\min \sum_{v \in V} X(v) w(v)$

$$Z_I = opt = \min \sum_{v \in V} X(v) w(v)$$

$$X(u) + X(v) \geq 1 \quad \forall uv \in E$$

$$X(u) \in \{0,1\} \quad \forall u \in V$$

LP-relaxation

$$Z_{LP} = \min \sum X(v) w(v)$$

$$(\Delta) \quad X(u) + X(v) \geq 1 \quad \forall uv \in E$$

$$0 \leq X(u) \leq 1 \quad \forall u \in V$$

$Z_{LP} \leq Z_I = opt$

LP-relaxation can be solved in polynomial time

Approximation algorithm B for weighted VC:

1. Solve LP-relaxation and let $\bar{X} = (\bar{X}(v))_{v \in V}$ be an optimal sol to this
2. let $U = \{v \mid \bar{X}(v) \geq \frac{1}{2}\}$
3. Return U

U is a vertex cover as $\bar{X}$ satisfies (Δ)

<u>Claim</u> $w(U) \le 2 \cdot opt = Z_I \cdot 2$

<u>Proof:</u> $opt \ge Z_{LP} = \sum_{v \in V} \bar{x}(v) w(v)$

$$\ge \sum_{\{v \mid \bar{x}(v) \ge \frac{1}{2}\}} \bar{x}(v) w(v)$$

$$\ge \sum_{\{v \mid \bar{x}(v) \ge \frac{1}{2}\}} \frac{1}{2} \cdot w(v)$$

$$= \frac{1}{2} \sum_{v \in U} w(v) = \frac{1}{2} w(U)$$

$\Rightarrow \quad w(U) \le 2 \cdot opt$