Chapter 1
Set Cover

Advanced Algorithms

SS 2019

Fabian Kuhn

Approximation Algorithms

An approximation algorithm is an algorithm that computes a solution for an optimization problem with an objective value that is provably within a bounded factor of the optimal objective value.

Formally:

- $\mathrm{OPT} \geq 0$: optimal objective value
  $\mathrm{ALG} \geq 0$: objective value achieved by the algorithm
- **Approximation Ratio α:**

$$\textbf{Minimization: } \alpha := \max_{\text{input instances}} \frac{\mathrm{ALG}}{\mathrm{OPT}} \geq 1$$

$$\textbf{Maximization: } \alpha := \min_{\text{input instances}} \frac{\mathrm{ALG}}{\mathrm{OPT}} < 1$$

Set Cover

Input: A set of elements E and a collection $\mathcal{S}$ of subsets E, i.e., $\mathcal{S} \subseteq 2^E$

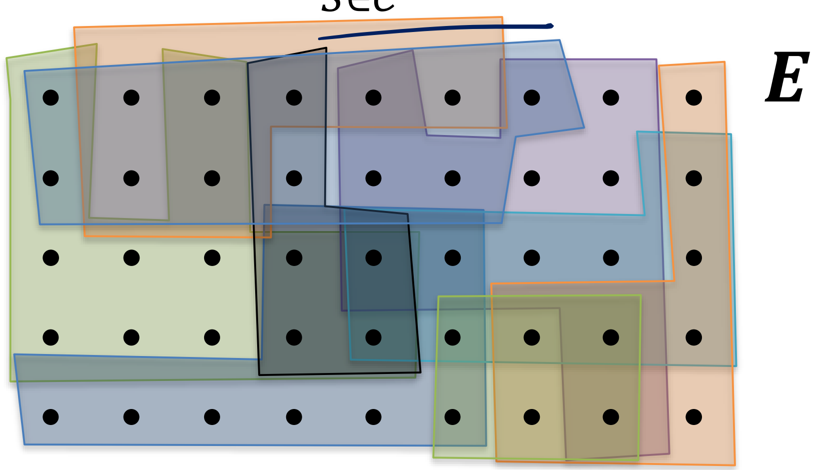
- such that $\bigcup_{S \in \mathcal{S}} S = E$, $|E| = n$
- Maximum set size $\Delta := \max_{S \in \mathcal{S}} |S|$
- Maximum element frequency $f := \max_{e \in E} |\{S \in \mathcal{S} : e \in S\}|$

$(E, \mathcal{S})$: set system

Set Cover: A set cover $\mathcal{C}$ of $(E, \mathcal{S})$ is a subset of the sets $\mathcal{S}$ which covers E:

$$\bigcup_{S \in \mathcal{C}} S = E$$

Example:



Minimum (Weighted) Set Cover

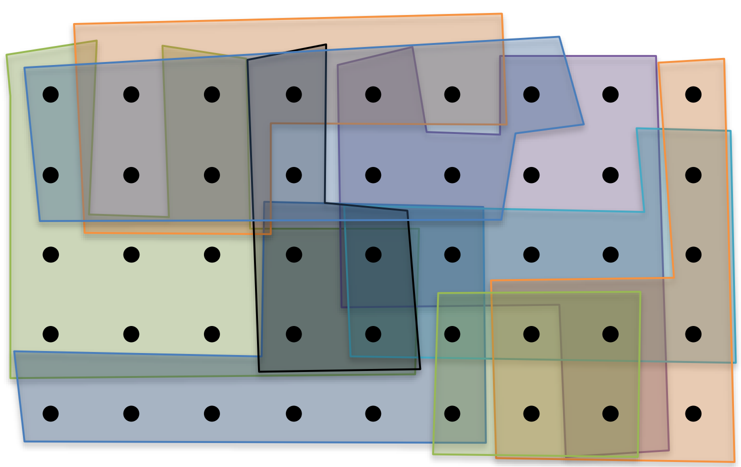
Minimum Set Cover:

- **Goal:** Find a set cover $\mathcal{C}$ of smallest possible size
  - i.e., cover E with as few sets as possible

Minimum Weighted Set Cover:

- Each set $S \in \mathcal{S}$ has a weight $w(S) > 0$
- **Goal:** Find a set cover $\mathcal{C}$ of minimum weight

Example:



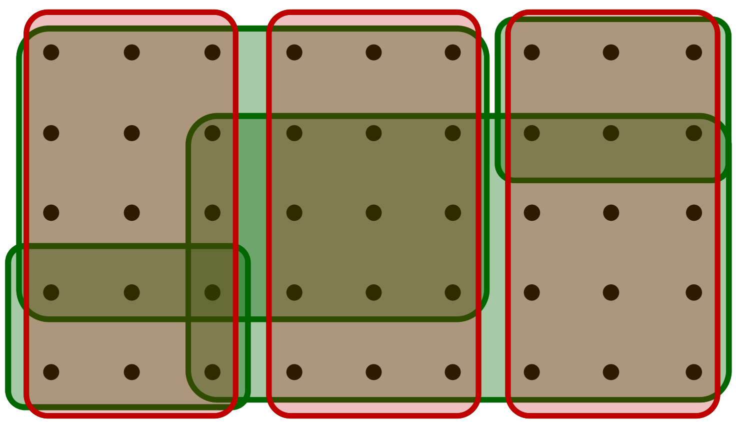
$\boldsymbol{E}$

Minimum Set Cover: Greedy Algorithm

Greedy Set Cover Algorithm:

- Start with $\mathcal{C} = \emptyset$
- In each step, add set $S \in \mathcal{S} \setminus \mathcal{C}$ to $\mathcal{C}$ s.t. S covers as many uncovered elements as possible

Example:



Weighted Set Cover: Greedy Algorithm

Greedy Weighted Set Cover Algorithm:

- Start with $\mathcal{C} = \emptyset$
- Price-per-element ratio of $S \in \mathcal{S} \setminus \mathcal{C}$:

$$\underline{\mathrm{ppe}(S)} := \frac{w(S)}{\left|S \setminus \bigcup_{T \in \mathcal{C}} T\right|}$$

$|S \setminus \bigcup_{T \in \mathcal{C}} T|$: # newly covered elem. when adding S

- In each step, add set $S \in \mathcal{S} \setminus \mathcal{C}$ with minimum $\mathrm{ppe}(S)$

Analysis of Greedy Algorithm:

- Assign a $\underline{\mathrm{price}(e)}$ to each element $e \in E$:
  (price-per-element when covering the element)
- If covering e with $\underline{\text{set } S}$ and partial cover is $\mathcal{C}$ before adding S:

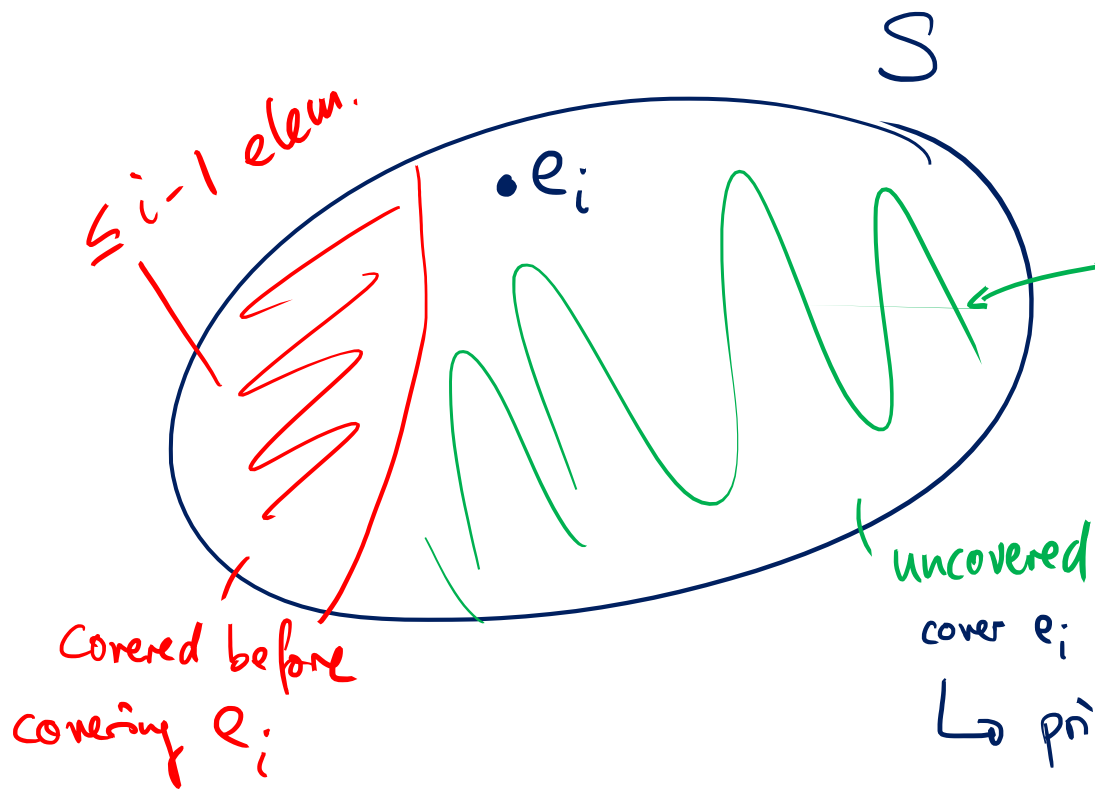
$$\underline{\mathrm{price}(e)} = \mathrm{ppe}(S)$$

property: $\sum_{e \text{: covered}} \mathrm{price}(e) = \sum_{S \in \mathcal{C}} w(S)$

Weighted Set Cover: Greedy Algorithm

Lemma: Consider a set $S = \{e_1, e_2, \dots, e_k\} \in \mathcal{S}$ and assume that the elements are covered in the order $e_1, e_2, \dots, e_k$ by the greedy algorithm (ties broken arbitrarily).

Then, the price of element e_i is at most $\text{price}(e_i) \leq \frac{w(S)}{k-i+1}$



$\geq k-(i-1)$

$= k-i+1$ elem.

$\text{ppe}(S) \leq \frac{w(S)}{k-i+1}$

cover e_i by S' with $\text{ppe}(S') \leq \text{ppe}(S)$

$\hookrightarrow \text{price}(e_i) \leq \frac{w(S)}{k-i+1}$

Weighted Set Cover: Greedy Algorithm

Lemma: Consider a set $S = \{e_1, e_2, \dots, e_k\} \in \mathcal{S}$ and assume that the elements are covered in the order $e_1, e_2, \dots, e_k$ by the greedy algorithm (ties broken arbitrarily).

Then, the price of element e_i is at most $\mathrm{price}(e_i) \leq \frac{w(S)}{k-i+1}$

Corollary: The total price of a set $S \in \mathcal{S}$ of size $|S| = k$ is

$$\sum_{e \in S} \mathrm{price}(e) \leq w(S) \cdot H_k, \qquad \text{where } H_k = \sum_{i=1}^{k} \frac{1}{i} \leq 1 + \ln k$$

$$\sum_{e\in S} \mathrm{price}(e) = \sum_{i=1}^{k} \mathrm{price}(e_i) \leq \sum_{i=1}^{k} \frac{w(S)}{k-i+1} = \sum_{j=1}^{k} \frac{w(S)}{j} = w(S) \cdot H_k \leq w(S) \cdot H_\Delta$$

Weighted Set Cover: Greedy Algorithm

Corollary: The total price of a set $S \in \mathcal{S}$ of size $|S| = k$ is

$$\sum_{e \in S} p(e) \leq w(S) \cdot H_k, \qquad \text{where } H_k = \sum_{i=1}^{k} \frac{1}{i} \leq 1 + \ln k$$

Theorem: The approximation ratio of the greedy minimum (weighted) set cover algorithm is at most $\boldsymbol{H_\Delta \leq 1 + \ln \Delta}$, where s is the cardinality of the largest set ($\Delta = \max_{S \in \mathcal{S}} |S|$).

C : greedy set cover $\quad C^*$: opt set cover

$$w(C) = \sum_{e \in E} \mathrm{price}(e) \underset{\substack{\uparrow \\ C^* \text{ is a set cover}}}{\leq} \sum_{S \in C^*} \sum_{e \in S} \mathrm{price}(e) \leq \sum_{S \in C^*} w(S) \cdot H_\Delta = w(C^*) \cdot H_\Delta$$

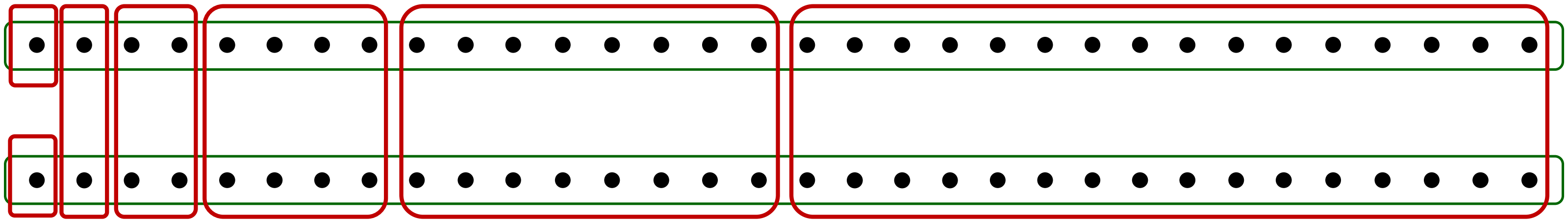
Set Cover Greedy Algorithm

Can we improve this analysis?

No! Even for the unweighted minimum set cover problem, the approximation ratio of the greedy algorithm is $\geq \big(1 - o(1)\big) \cdot \ln \Delta$.

- if Δ is the size of the largest set... (Δ can be linear in n)

Let's show that the approximation ratio is at least $\Omega(\log n)$...



$\mathbf{OPT = 2}$

$\mathbf{GREEDY \geq \log_2 \boldsymbol{n}}$

Set Cover: Better Algorithm?

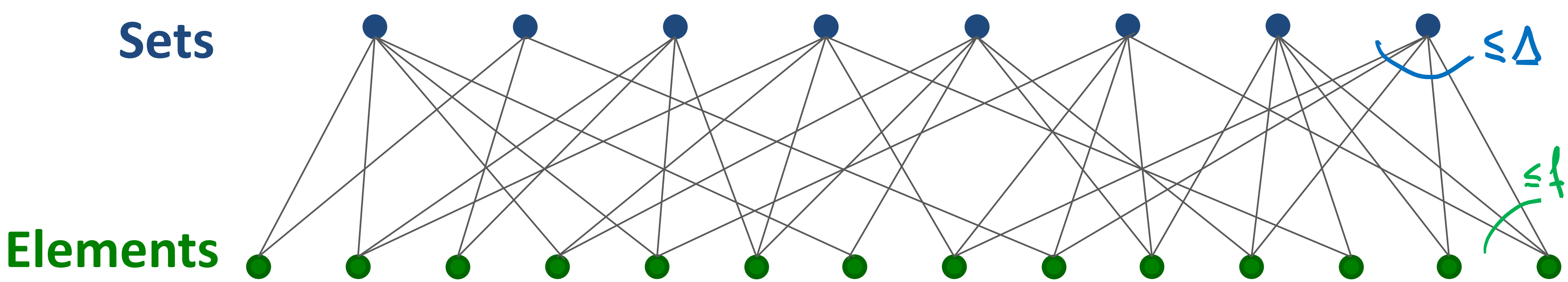
An approximation ratio of $\ln n$ seems not spectacular...

Can we improve the approximation ratio?

No: In a series of work, Lund and Yannakakis (1994), Feige (1998), and Moshkovitz (2015) showed that it is NP-hard to approximate minimum set cover by a factor $(1-\varepsilon) \cdot \ln n$ for any constant $\varepsilon > 0$.

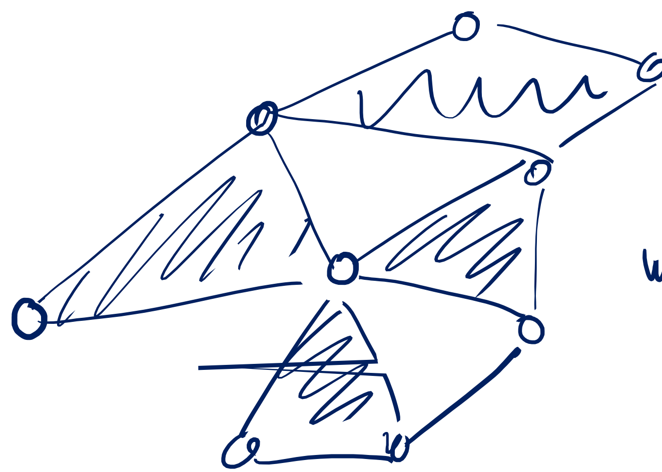
- Proof is based on the so-called PCP theorem
  - PCP theorem is one of the main (relatively) recent advancements in theoretical computer science and the major tool to prove approximation hardness lower bounds
  - Shows that every language in NP has certificates of polynomial length that can be checked by a randomized algorithm by only querying a constant number of bits (for any constant error probability)

Special Case: Small f



Formulation as Minimum Hypergraph Vertex Cover

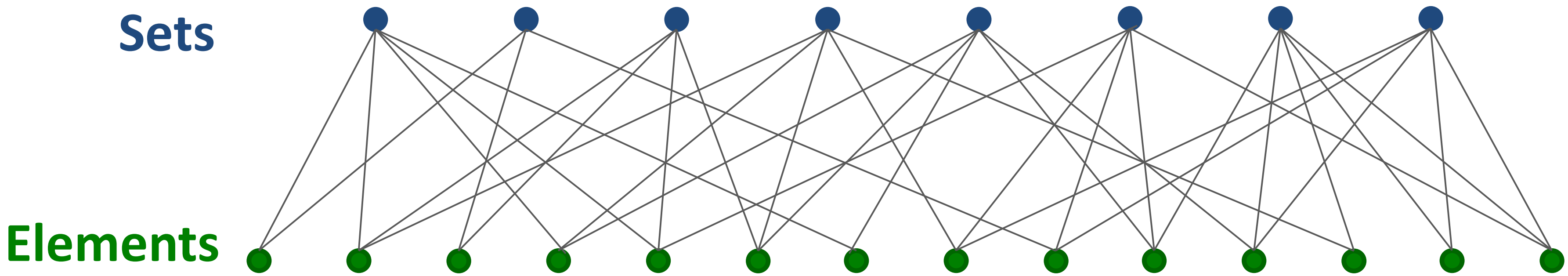
- Hypergraph $H = (V, E)$, $E \in 2^H$ are the hyperedges



rank of H

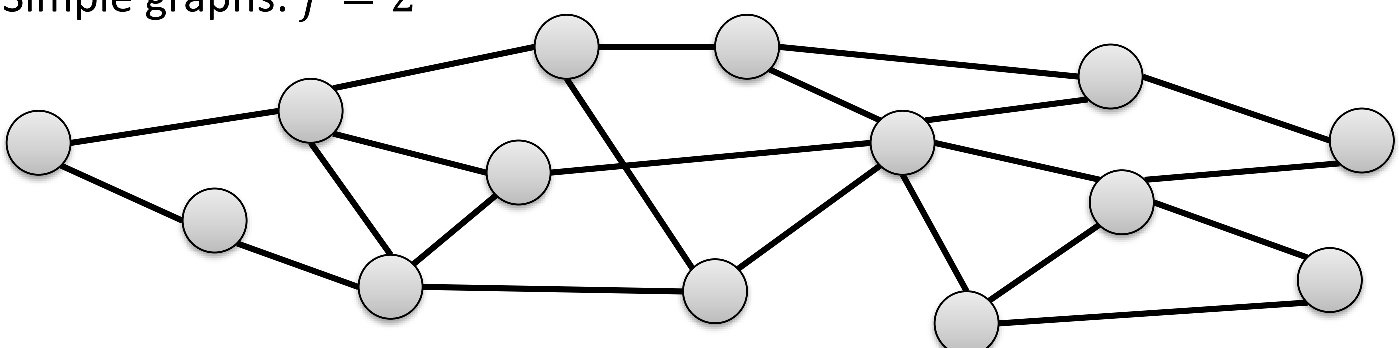
max. ~~hyperedge~~ size

Special Case: Small f



Formulation as Minimum Hypergraph Vertex Cover

- Hypergraph $H = (V, E)$, $E \in 2^H$ are the hyperedges
- Vertex cover: $S \subseteq V$ s.t. $\forall e \in E : S \cap e \neq \emptyset$
  - equivalent to set cover (V: sets, E: elements)
  - Max. frequency f = max. hyperedge size = rank of H
  - Simple graphs: $f = 2$



Vertex Cover vs Matching

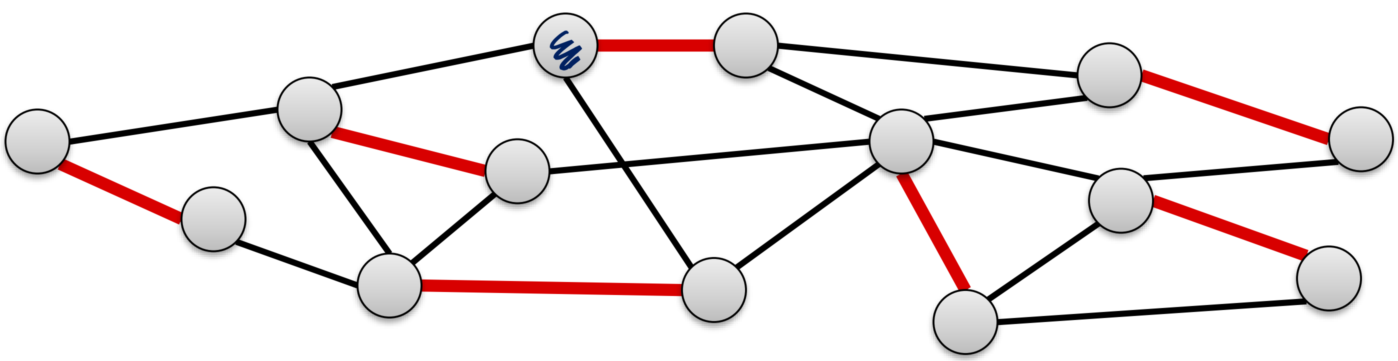
Matching of a hypergraph $\boldsymbol{H = (V, E)}$

- A disjoint set of edges $M \subseteq E$

Lemma: Given a hypergraph $H = (V, E)$, for every matching $M \subseteq E$ and every vertex cover $S \subseteq V$, we have $|M| \leq |S|$.

Proof:

- S is a vertex cover $\Longrightarrow \forall e \in M, \exists v_e \in e \cap S$
- M is a matching $\Longrightarrow v_{e_1} \neq v_{e_2}$ for $e_1 \neq e_2$ (e_1 & e_2 are disjoint)

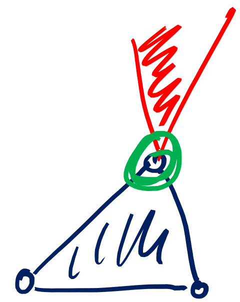


Matching Approximation of Vertex Cover

Vertex Cover Approximation Algorithm

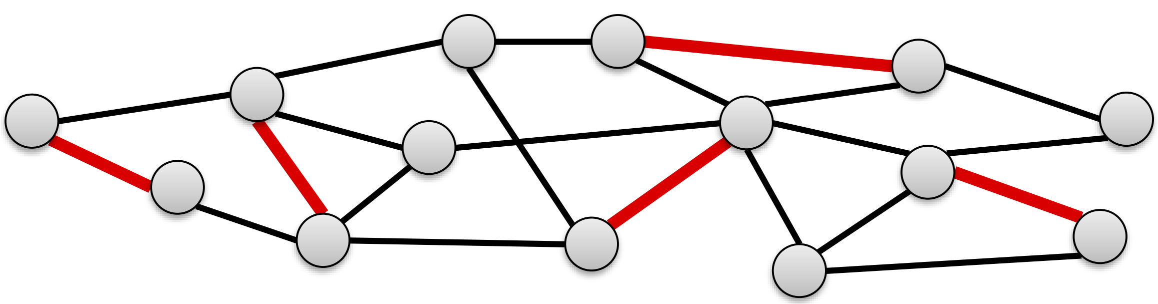
- Let $H = (V, E)$ be a hypergraph of rank $\leq f$
- Compute a maximal matching M of H
- Define vertex cover S as $S := \bigcup_{e \in M} e$

Theorem: The above algorithm computes an f-approximation of the (unweighted) minimum vertex cover problem in H.



Proof:

- M maximal $\Longrightarrow$ S is a vertex cover
  - $\forall \{v_1, \dots, v_k\} \in E$, at least one of vertices $v_1, \dots, v_k$ is matched



- We have $|S| = \sum_{e \in M} |e| \leq f \cdot |M|$ and $|M| \leq |S^*|$

$$\Longrightarrow |S| \leq f \cdot |S^*|$$

Linear Programming-Based Formulation

Linear Program (LP)

- (Continuous) optimization of a linear objective function subject to linear constraints

$$
\begin{aligned}
&\min \boldsymbol{c}^T \boldsymbol{x} \\
&\text{s.t. } A\boldsymbol{x} \geq \boldsymbol{b} \\
&\qquad \boldsymbol{x} \geq \boldsymbol{0}
\end{aligned}
$$

$c = \begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix} \quad x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$

$c^T x = \sum_{i=1}^{n} c_i x_i$

$Ax \geq b$

ineq. holds

comp.-wise

i-th ineq.

$a_{i1} x_1 + \ldots + a_{in} x_n \geq b_i$

LP Duality

- Every LP has a dual LP

Linear Program

$$\min \boldsymbol{c}^T \boldsymbol{x}$$
$$\text{s.t. } A\boldsymbol{x} \geq \boldsymbol{b}$$
$$\boldsymbol{x} \geq \boldsymbol{0}$$

Dual Linear Program

$$\max \boldsymbol{b}^T \boldsymbol{y}$$
$$\text{s.t. } A^T\boldsymbol{y} \leq \boldsymbol{c}$$
$$\boldsymbol{y} \geq \boldsymbol{0}$$

dual

$\max (-c)^T x$

$(-A^T)^T x \leq -b$

$x \geq 0$

dual

$\min (-b)^T y$

$(-A^T) y \geq -c$

$y \geq 0$

- **Weak duality:** For feasible solutions $\boldsymbol{x}$ and $\boldsymbol{y}$: $\boldsymbol{b}^T\boldsymbol{y} \leq \boldsymbol{c}^T\boldsymbol{x}$

$b^T y = y^T b \leq y^T A x \longleftrightarrow y^T A x \leq c^T x$

- **Strong duality:** For optimal solutions $\boldsymbol{x}^*$ and $\boldsymbol{y}^*$: $\boldsymbol{b}^T\boldsymbol{y}^* = \boldsymbol{c}^T\boldsymbol{x}^*$

LP-Based Approximation Algorithms

Important Technique to Design Approximation Algorithms

- LPs can be solved optimally in polynomial time
  - Using interior-point methods [Khachiyan '79], [Karmarkar '84]
- Many combinatorial optimization problems can be phrased as an integer linear program (ILPs):
  - LP with additional constraint that variables have to take integer values

Basic idea of many approximation algorithms:

1. Formulate given problem as an ILP
2. Relax integer constraints to get an LP
   - known as the LP relaxation of the given ILP
3. Solve the LP
4. Convert (fractional) LP solution to an integer solution
   - typically the hard part …

Minimum Set Cover as an ILP

Given: set system $(E, \mathcal{S})$ and weight $w(S) > 0$ for all S

need to determine for each $S \in \mathcal{S}$, need to determine whether $S \in \mathcal{C}$

variable $\underline{x_S \in \{0,1\}} \iff x_S = 1 \leftrightarrow S \in \mathcal{C}$

$$\min \sum_{S \in \mathcal{S}} w(S) \cdot x_S$$

$$\text{s.t. } \forall e \in E: \underline{\sum_{S : e \in S} x_S \geq 1}$$

Fractional Set Cover

$x_S \in [0,1]$

- LP relaxation gives variables $x_S \geq 0$ for each $S \in \mathcal{S}$, s.t.

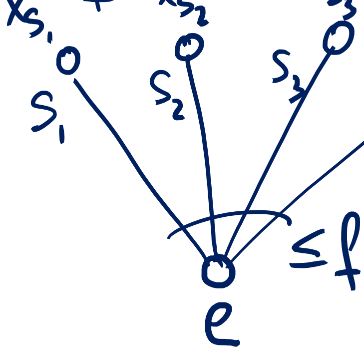
$$\forall e \in E : \sum_{S : e \in S} x_S \geq 1$$

opt. sol. for LP

and s.t. $\sum_{S \in \mathcal{S}} x_S \cdot w(S) \leq w(\mathcal{C}^*)$, where $\mathcal{C}^*$ is an optimal set cover.

- How can we turn this fractional solution into an integer one?
  - i.e., we need to round the fractional values $x_S \in [0,1]$ to $\hat{x}_S \in \{0,1\}$
- First consider the setting with bounded element frequency f

$x_{S_1} + x_{S_2} + x_{S_3} + x_{S_4} \geq 1$



for every $e \in E$: $\exists S : e \in S$ s.t. $x_S \geq \frac{1}{f}$

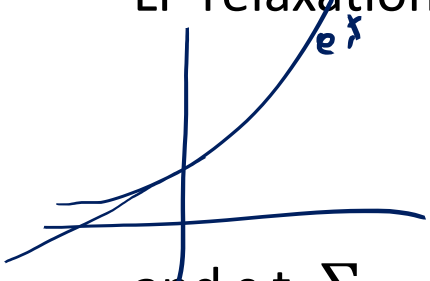
$$\hat{x}_S := \begin{cases} 1 & \text{if } x_S \geq 1/f \\ 0 & \text{otherwise} \end{cases}$$

$\hat{x}_S \leq f \cdot x_S$

$$\sum_S w(S) \cdot \hat{x}_S \leq f \cdot \sum w(S) x_S \leq f \cdot w(C^*)$$

Fractional Set Cover $\mathbb{E}\left[\sum w(S)\cdot \hat{x}_S\right] = \sum w(S)\cdot x_S$

- LP relaxation gives variables $x_S \geq 0$ for each $S \in \mathcal{S}$, s.t.

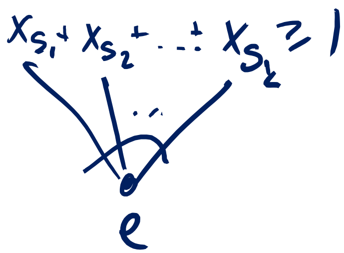


$$\forall e \in E : \sum_{S:e\in S} x_S \geq 1 \qquad \mathbb{E}[\hat{x}_S] = x_S$$

and s.t. $\sum_{S\in\mathcal{S}} x_S \cdot w(S) \leq w(\mathcal{C}^*)$, where $\mathcal{C}^*$ is an optimal set cover.

- How can we turn this fractional solution into an integer one?
  - i.e., we need to round the fractional values $x_S \in [0,1]$ to $\hat{x}_S \in \{0,1\}$ $\quad 1+x \leq e^{+x}$

$x_S \to \hat{x}_S \in \{0,1\}$



<u>idea</u>: use randomization: use x_S as a prob.

$$\hat{x}_S = \begin{cases} 1 & \text{with prob. } x_S \\ 0 & \text{with prob. } 1-x_S \end{cases}$$

$$\mathbb{P}(\text{elem. } e \text{ is \underline{not} covered}) = \prod_{S:e\in S} (1-x_S) \leq$$

Set Cover: Randomized Rounding

Set Cover Rounding Algorithm:

1. Set $p_S \coloneqq \min\{1, x_S \cdot \ln \Delta\}$
2. Add each set S to set cover $\mathcal{C}$ with probability p_S (independently)
3. For each $e \in E$: If e is not covered, add min-weight set cont. e

Theorem: Given an optimal fractional weighted set cover solution, the set cover rounding algorithm computes a set cover $\mathcal{C}$ of expected weight

$$\mathbb{E}[w(\mathcal{C})] \leq w(\mathcal{C}^*) \cdot (1 + \ln \Delta)$$

Proof:

X : total weight added to $\mathcal{C}$ in step 2

Y : total weight " " " " 3

$$\mathbb{E}[w(\mathcal{C})] = \mathbb{E}[X] + \mathbb{E}[Y]$$

$$\mathbb{E}[X] = \sum_S p_S \cdot w(S) \leq \ln \Delta \cdot \underbrace{\sum_S x_S w(S)}_{\text{obj. value of LP}} \leq \ln \Delta \cdot w(\mathcal{C}^*)$$

q_e: prob. that e uncovered after step 2

Set Cover: Randomized Rounding

Theorem: Given an optimal fractional weighted set cover solution, the set cover rounding algorithm computes a set cover $\mathcal{C}$ of expected weight

$$\mathbb{E}[\boldsymbol{w}(\boldsymbol{\mathcal{C}})] \le \boldsymbol{w}(\boldsymbol{\mathcal{C}}^*) \cdot (\mathbf{1} + \mathbf{ln}\,\boldsymbol{\Delta})$$

Proof: We already know that

$$\mathbb{E}[X] \le w(\mathcal{C}^*) \cdot \ln \Delta \qquad \text{and} \qquad \forall e \in E : q_e \le \frac{1}{\Delta}$$

$$q_e = \prod_{S:e\in S}(1-p_S) \le e^{-\sum p_S} \le e^{-\ln\Delta} = \frac{1}{\Delta}$$

↳ if some $p_S = 1 \rightarrow q_e = 0$

↳ otherwise $\sum_{S:e\in S} p_S = \sum_{S:e\in S} x_S \cdot \ln\Delta \ge \ln\Delta$

$$\mathbb{E}[Y] = \sum_{e\in E} \underbrace{\frac{1}{\Delta} \min_{S:e\in S} w(S)}_{q_e} \overset{!}{\le} w(C^*)$$

Set Cover Dual LP

$\sum y_e \leq \sum w(S) x_S \leq w(C^*)$

Linear Program

$$\begin{aligned} &\min \boldsymbol{c}^T \boldsymbol{x} \\ &\text{s.t. } A\boldsymbol{x} \geq \boldsymbol{b} \\ &\quad\quad \boldsymbol{x} \geq \boldsymbol{0} \end{aligned}$$

Dual Linear Program

$$\begin{aligned} &\max \boldsymbol{b}^T \boldsymbol{y} \\ &\text{s.t. } A^T\boldsymbol{y} \leq \boldsymbol{c} \\ &\quad\quad \boldsymbol{y} \geq \boldsymbol{0} \end{aligned}$$

$$\min \sum_{S} w(S) \cdot x_S$$

$$\forall e \in E: \sum_{S: e \in S} x_S \geq 1$$

$$x_S \geq 0$$

$$\boxed{\begin{aligned} &\max \sum_{e \in E} y_e \\ &\forall S: \sum_{e: e \in S} y_e \leq w(S) \\ &\quad\quad y_e \geq 0 \end{aligned}}$$

Set Cover Dual LP

Linear Program

$$\min \boldsymbol{c}^T \boldsymbol{x}$$
$$\text{s.t. } A\boldsymbol{x} \geq \boldsymbol{b}$$
$$\boldsymbol{x} \geq \boldsymbol{0}$$

Dual Linear Program

$$\max \boldsymbol{b}^T \boldsymbol{y}$$
$$\text{s.t. } A^T\boldsymbol{y} \leq \boldsymbol{c}$$
$$\boldsymbol{y} \geq \boldsymbol{0}$$

Set Cover: Randomized Rounding

Theorem: Given an optimal fractional weighted set cover solution, the set cover rounding algorithm computes a set cover $\mathcal{C}$ of expected weight

$$\mathbb{E}[\boldsymbol{w}(\boldsymbol{\mathcal{C}})] \leq \boldsymbol{w}(\boldsymbol{\mathcal{C}}^*) \cdot (\mathbf{1} + \ln \boldsymbol{\Delta})$$

Proof:

- It remains to show that

$$\sum_{e \in E} \underbrace{\frac{1}{\Delta} \cdot \min_{S: e \in S} w(S)}_{y_e} \leq w(\mathcal{C}^*)$$

$$\boxed{\forall S: \sum_{e \in S} y_e \leq w(S)}$$

$$\frac{1}{\Delta} \cdot \sum_{e \in S} \underbrace{\min_{S': e \in S'} w(S')}_{\leq w(S)} = \frac{1}{\Delta} \cdot |S| \cdot w(S) \leq w(S)$$

Approximating Weighted Vertex Cover

Recall maximal matching approximation for the unweighted case

- Vertex cover $S =$ all matched vertices of a maximal matching M
- S is a vertex cover because of the maximality of M
- Edges in M need to be covered by different nodes in $S^* \Longrightarrow |M| \leq |S^*|$

Generalization to Weighted Vertex Cover?

- The same algorithm does obviously not work
- Different view of above algorithm:

Maximal matching M is a maximal feasible solution of the dual LP

dual set cover LP:

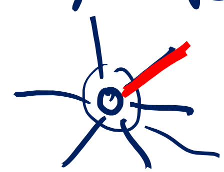
$\forall e \in E: \; y_e \geq 0$

$\forall S: \; \sum_{e \in S} y_e \leq w(S)$

vertex cover

elements: edges

dual sol. assigns $y_e \geq 0$ to each edge



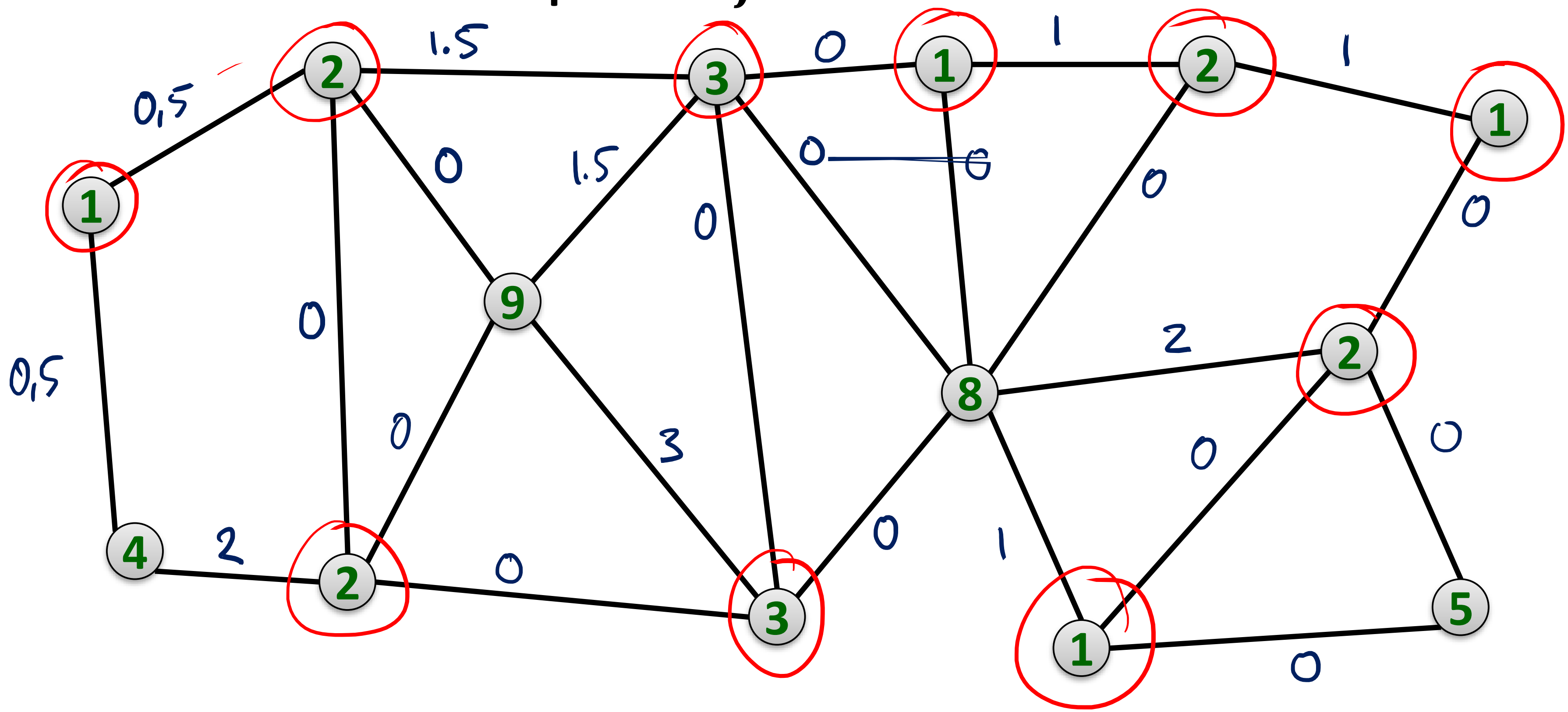
$\sum y_e \leq 1$

Approximating Weighted Vertex Cover

Theorem: Let $y = \{y_e \geq 0 : e \in E\}$ be a maximal feasible solution of the dual weighted (hypergraph) vertex cover LP. Define the vertex set S as $S \coloneqq \{v \in V : \sum_{e:v\in e} y_e = w(v)\}$. Then, S is a vertex cover of weight

$$w(S) \leq f \cdot w(S^*).$$

Let's start with an example with $f = 2$:



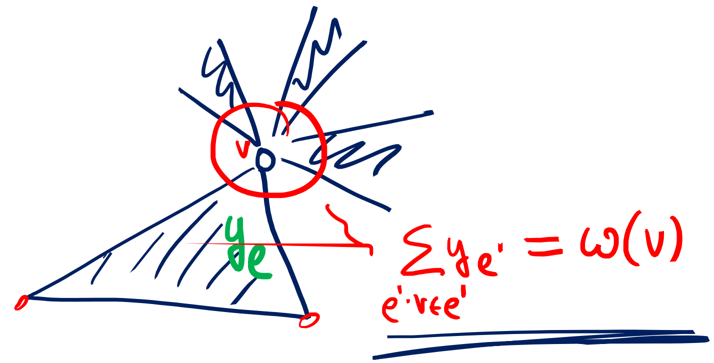
Approximating Weighted Vertex Cover

Theorem: Let $y = \{y_e \geq 0 : e \in E\}$ be a maximal feasible solution of the dual weighted (hypergraph) vertex cover LP. Define the vertex set S as $S \coloneqq \{v \in V : \sum_{e:v\in e} y_e = w(v)\}$. Then, S is a vertex cover of weight

$$w(S) \leq f \cdot w(S^*).$$

Proof:

- S is a vertex cover ✓



- total weight of S

$$w(S) = \sum_{v\in S} w(v) = \sum_{v\in S} \sum_{e:v\in e} y_e \leq f \cdot \sum_{e\in E} y_e \leq f \cdot w(S^*)$$

nodes in S are tight

y_e : dual feasible
weak duality