## Section 1.6: A Greedy Algorithm

A natural greedy choice would be to „pay" as little as possible for each additional covered element:

**Alg 1.2 for Set Cover: Greedy**

$I \leftarrow \emptyset$

For $j \leftarrow 1$ to $m$

  $\hat{S}_j \leftarrow S_j$  (uncovered part of $S_j$)

While $\{S_j \mid j \in I\}$ is not a set cover

  $\ell \leftarrow \underset{j \,:\, \hat{S}_j \neq \emptyset}{\arg\min} \dfrac{w_j}{|\hat{S}_j|}$  ($S_\ell$: set with smallest cost per uncovered element)

  $I \leftarrow I \cup \{\ell\}$

  For $j \leftarrow 1$ to $m$

    $\hat{S}_j \leftarrow \hat{S}_j - S_\ell$

The greedy alg. is an $H_n$-approx. alg

Recall: $H_n = 1 + \frac{1}{2} + \frac{1}{3} + \ldots + \frac{1}{n} \approx \ln(n)$

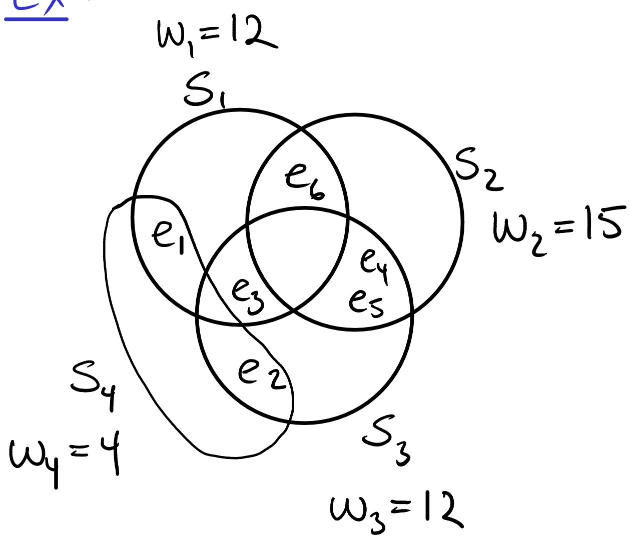
It is „likely" that no significantly better approx. ratio can be obtained:

Thm 1.13:

Approx. factor $\frac{\ln n}{c}$, $c > 1$, for unweighted Set Cover

$\Rightarrow$ $n^{O(\log \log n)}$-approx alg. for NPC

Ex:




$$\frac{w_1}{|S_1|} = \frac{12}{3} = 4$$

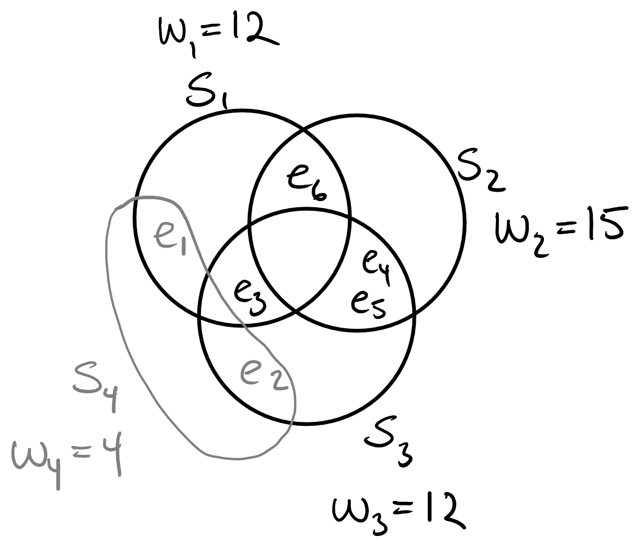
$$\frac{w_2}{|S_2|} = \frac{15}{3} = 5$$

$$\frac{w_3}{|S_3|} = \frac{12}{4} = 3$$

$$\frac{w_4}{|S_4|} = \frac{4}{2} = 2$$

← price per element in first iteration

Pick $S_4$



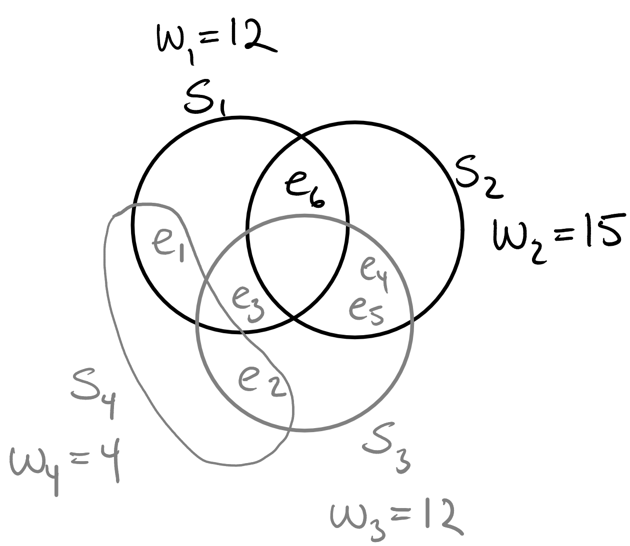

$$\frac{w_1}{|\hat{S}_1|} = \frac{12}{2} = 6$$

$$\frac{w_2}{|\hat{S}_2|} = \frac{15}{3} = 5$$

$$\frac{w_3}{|\hat{S}_3|} = \frac{12}{3} = 4$$

← price per element in second it.

Pick $S_3$

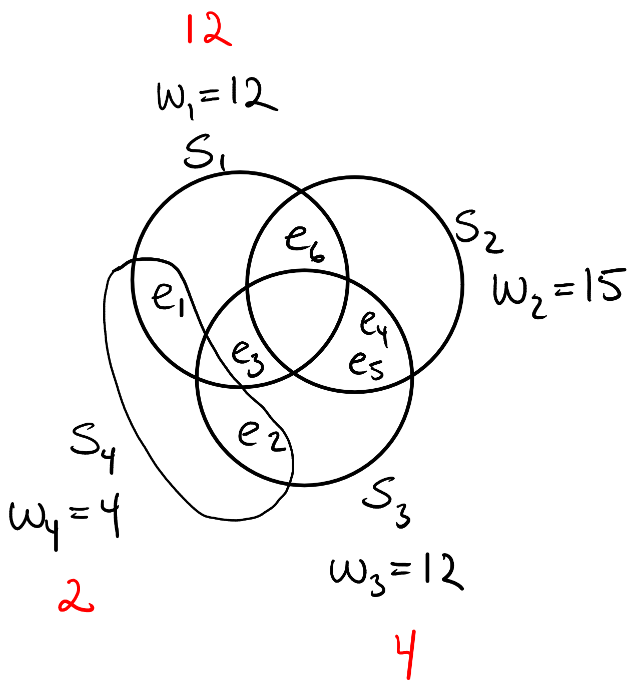



$$\frac{w_1}{|\hat{S}_1|} = \frac{12}{1} = 12$$

← price per element in third it.

$$\frac{w_2}{|\hat{S}_2|} = \frac{15}{1} = 15$$

Pick $S_1$



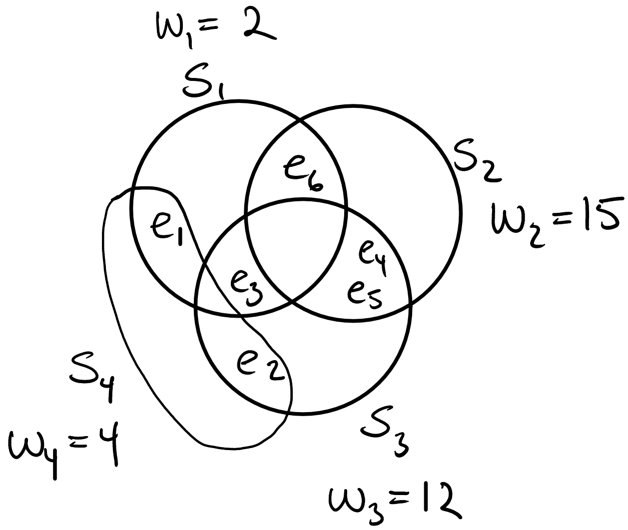

$$
\begin{aligned}
\text{Greedy} &= 28 \\
&= w_4 + w_3 + w_1 \\
&= 4 + 12 + 12 \\
&= 2+2 + 4+4+4 + 12 \\
&= \sum_{i=1}^{6} \text{price}(e_i)
\end{aligned}
$$

$$
\begin{aligned}
\text{OPT} &= 24 \\
&= w_3 + w_1 \\
&= 12 + 12 \\
&= 4+4+4+4 + 6+6
\end{aligned}
$$

We will now use this ex. to illustrate the proof of Thm 1.11 stating that Greedy is an $H_n$-approx. alg.:

## $H_6$ - approximation:

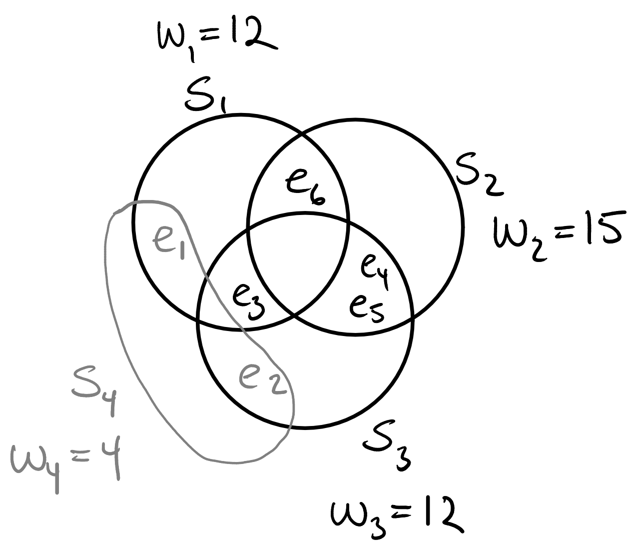
$$\left( H_6 = 1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \frac{1}{5} + \frac{1}{6} = \frac{147}{60} < 2.5 \right)$$




$OPT \geq 6 \cdot price(e_1)$

$OPT \geq 6 \cdot price(e_2)$

since $S_4$ gives the best average weight per element.

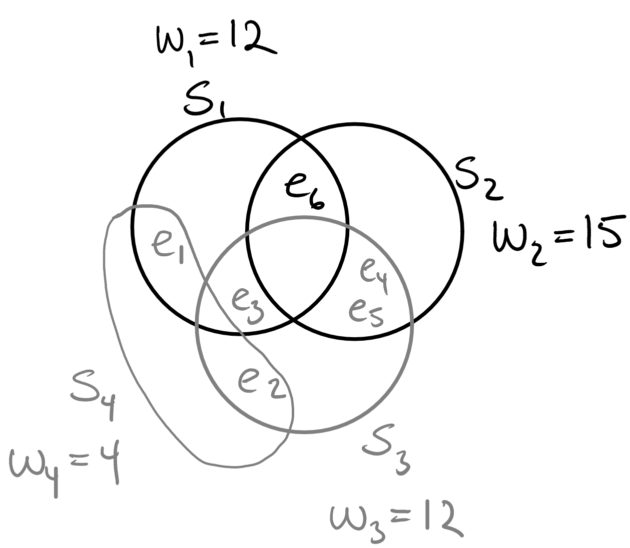



$S_4$ cannot cover any of the elements $e_3, e_4, e_5, e_6$. Thus, the average weight of these elements cannot be lower than $price(e_3)$, even for OPT:

$OPT \geq 4 \cdot price(e_3)$

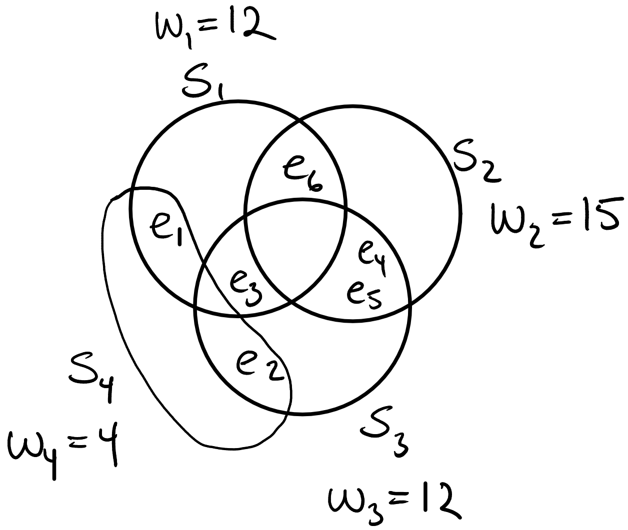
$OPT \geq 4 \cdot price(e_4)$

$OPT \geq 4 \cdot price(e_5)$




Similarly:

$OPT \geq price(e_6)$




$$\mathrm{OPT} \geq 6 \cdot \mathrm{price}(e_1) \iff \mathrm{price}(e_1) \leq \frac{\mathrm{OPT}}{6}$$

$$\mathrm{OPT} \geq 6 \cdot \mathrm{price}(e_2) \iff \mathrm{price}(e_2) \leq \frac{\mathrm{OPT}}{6}$$

$$\mathrm{OPT} \geq 4 \cdot \mathrm{price}(e_3) \iff \mathrm{price}(e_3) \leq \frac{\mathrm{OPT}}{4}$$

$$\mathrm{OPT} \geq 4 \cdot \mathrm{price}(e_4) \iff \mathrm{price}(e_4) \leq \frac{\mathrm{OPT}}{4}$$

$$\mathrm{OPT} \geq 4 \cdot \mathrm{price}(e_5) \iff \mathrm{price}(e_5) \leq \frac{\mathrm{OPT}}{4}$$

$$\mathrm{OPT} \geq \mathrm{price}(e_6) \iff \mathrm{price}(e_6) \leq \mathrm{OPT}$$

$$\begin{aligned}
\mathrm{Greedy} &= \sum_{i=1}^{6} \mathrm{price}(e_i) \\
&\leq \frac{\mathrm{OPT}}{6} + \frac{\mathrm{OPT}}{6} + \frac{\mathrm{OPT}}{4} + \frac{\mathrm{OPT}}{4} + \frac{\mathrm{OPT}}{4} + \frac{\mathrm{OPT}}{1} \\
&\leq \frac{\mathrm{OPT}}{6} + \frac{\mathrm{OPT}}{5} + \frac{\mathrm{OPT}}{4} + \frac{\mathrm{OPT}}{3} + \frac{\mathrm{OPT}}{2} + \frac{\mathrm{OPT}}{1} \\
&= H_6 \cdot \mathrm{OPT}
\end{aligned}$$

**Thm 1.11**

Alg. 1.2 is an $H_n$-approx. alg. for Set Cover

Proof:

$n_k$: # uncovered elements at the beginning of the k'th iteration

Above ex.: $n_1=6$, $n_2=4$, $n_3=1$, $n_4=0$
$n_1-n_2=2$, $n_2-n_3=3$, $n_3-n_4=1$

Any algorithm, including OPT, has to cover these $n_k$ elements using only sets in $\mathcal{S}-\{S_j \mid j\in I\}$, since none of them are contained in $\{S_j \mid j\in I\}$.

Hence, there must be at least one element with a price of at most $OPT/n_k$. Otherwise, OPT would not be able to cover the $n_k$ elements (and certainly not all $n$ elements) at a cost of only OPT.
Hence, the $n_k-n_{k+1}$ elements covered in iteration $k$ cost at most $(n_k-n_{k+1})\,OPT/n_k$ in total.

Thus, the cost of the set cover produced by the greedy alg. is

# iterations

$$
\begin{aligned}
\sum_{j \in I} w_j &\leq \sum_{k=1}^{r} \frac{n_k - n_{k+1}}{n_k} \, OPT \\
&= OPT \sum_{k=1}^{r} (n_k - n_{k+1}) \cdot \frac{1}{n_k} \\
&\leq OPT \sum_{k=1}^{r} \Bigg( \overbrace{\frac{1}{n_k} + \frac{1}{n_k - 1} + \dots + \frac{1}{n_{k+1} + 1}}^{n_k - n_{k+1} \text{ terms that are each } \geq \frac{1}{n_k}} \Bigg) \\
&= OPT \sum_{s=1}^{n} \frac{1}{s} \\
&= OPT \cdot H_n \qquad \square
\end{aligned}
$$

Let $g = \max\{\, |S_i| \mid S_i \in \mathcal{S} \,\}$.

**Thm 1.12**

Alg. 1.2 is an $H_g$-approx. alg. for Set Cover

**Proof:** By Dual Fitting:

Consider the dual $D$ of the LP for Set Cover.

We will construct

- an infeasible solution $\vec{y}$ and
- a feasible solution $\vec{y'}$

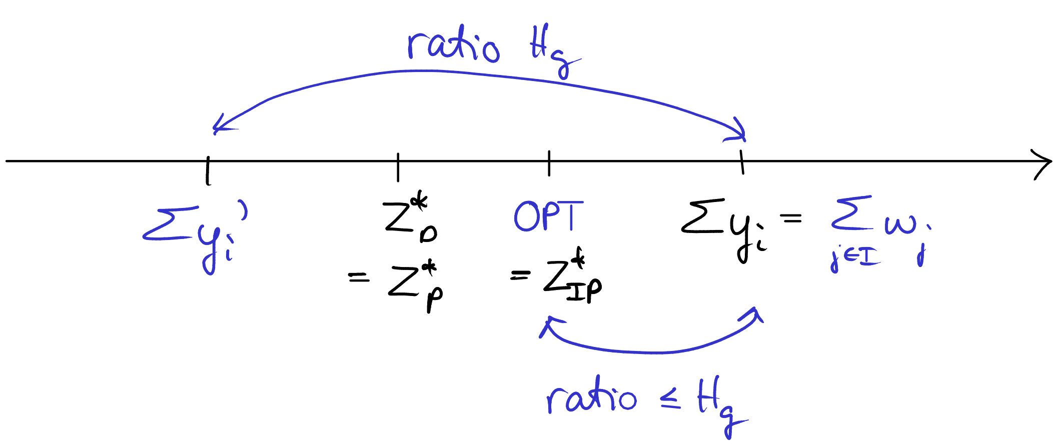
such that

- $\sum_{i=1}^{n} y_i = \sum_{j \in I} w_j$ (obtained, if $y_i = \text{price}(e_i)$)
- $y_i' = \frac{1}{H_g} \cdot y_i$

Then,

$$\sum_{j \in I} w_j = \sum_{i=1}^{n} y_i = H_g \sum_{i=1}^{n} y_i' \le H_g z_D^* \le H_g \cdot OPT,$$

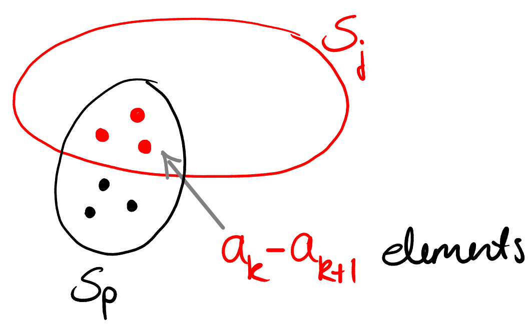
proving the claimed approximation factor. $\square$



For $1 \le i \le n$, let $y_i = price(e_i)$. Then, $\sum_{1 \le i \le n} y_i = \sum_{j \in I} w_j$.

Hence, we just need to show that $\vec{y}$ is feasible:

Consider an arbitrary set $S_j$.
Let $a_k$ be #uncovered elements in $S_j$ at the beginning of the k'th iteration.
Let $S_p$ be the set chosen by Greedy in the k'th iteration.
$S_p$ covers $a_k - a_{k+1}$ previously uncovered elements in $S_j$




The price per elem. in $S_j$ covered in the k'th it. is

$$\frac{w_p}{|\hat{S}_p|} \le \frac{w_j}{|\hat{S}_j|} \le \frac{w_j}{a_k}$$

since otherwise $S_j$ would be a <u>more</u> greedy choice. ↯

Thus,

$$\sum_{e_i \in S_j} y_i \leq \overbrace{\sum_{k=1}^{r} (a_k - a_{k+1}) \frac{w_j}{a_k}}^{\text{Total \#terms} = |S_j|,\ \text{since } a_1 = |S_j| \text{ and } a_{r+1} = 0}$$

$$\leq w_j \sum_{i=1}^{|S_j|} \frac{1}{i},$$ by the same arguments as in the proof of Thm 1.12.

$$\leq w_j \sum_{i=1}^{g} \frac{1}{i}$$

$$= w_j \cdot H_g$$

Hence,

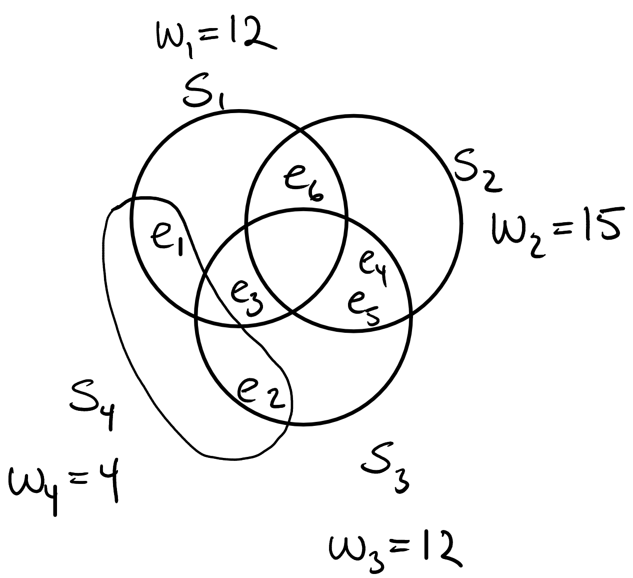
$$\sum_{e_i \in S_j} y_i' = \frac{1}{H_g} \sum_{e_i \in S_j} y_i \leq w_j$$

$\square$

Compare the proof of Thm 1.12 (dual fitting) to the proof of Thm 1.11:

- Simpler: Compare prices to $w_j$ instead of OPT
- Stronger result: $H_g$ instead of $H_n$ (could also have been obtained with the technique of the proof of Thm 1.11)

## Ex from before:




$g = 4 \Rightarrow H_g = 1 + \frac{1}{2} + \frac{1}{3} + \frac{}{4} = \frac{}{12} < 2.1$

$y_1 = y_2 = 2$

$y_3 = y_4 = y_5 = 4$

$y_6 = 12$

$y_1' = y_2' = \frac{24}{25}$

$y_3' = y_4' = y_5' = \frac{48}{25}$

$y_6' = \frac{144}{25}$

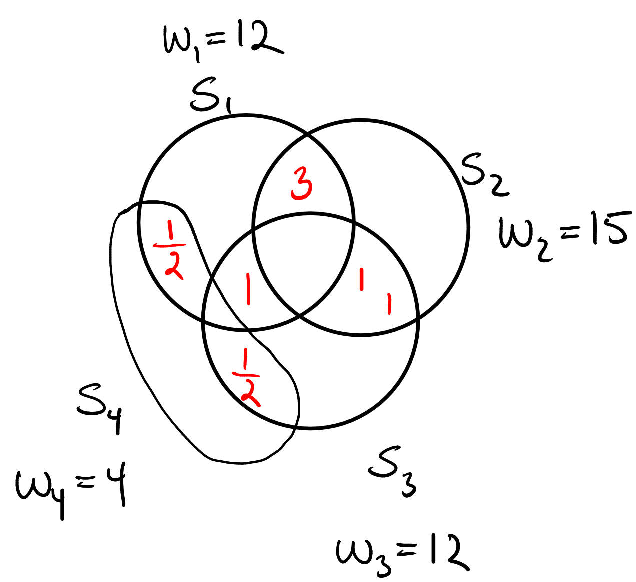
## Dual constraints:

$y_1' + y_3' + y_6' \leq 12 \qquad \frac{24}{25} + \frac{48}{25} + \frac{144}{25} = \frac{216}{25} < 12$

$y_4' + y_5' + y_6' \leq 15 \qquad \frac{48}{25} + \frac{48}{25} + \frac{144}{25} = \frac{240}{25} < 15$

$y_2' + y_3' + y_4' + y_5' \leq 12 \qquad \frac{24}{25} + \frac{48}{25} + \frac{48}{25} + \frac{48}{25} = \frac{168}{25} < 12$

$y_1' + y_2' \leq 4 \qquad \frac{24}{25} + \frac{24}{25} < 4$

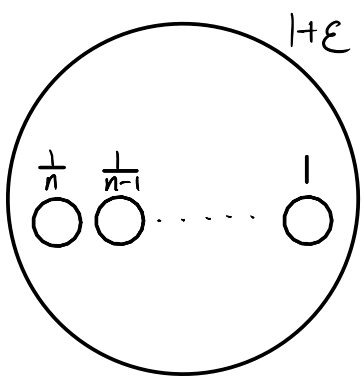


## Is the upper bound of $H_n$ tight?

If it is, the matching lower bound must come from an instance with

- one set containing all elements
  (follows from the upper bound of $H_g$)
- only one additional element covered in each it.
  (otherwise, some of the terms in $\frac{1}{n} + \frac{1}{n-1} + \dots + 1$ would be replaced by smaller terms.)

Ex:



# Summary

## Greedy

$H_n$-approx.:

$$\text{price}(e_i) \leq \frac{OPT}{n-i}, \quad i = 0, 1, \dots, n-1$$

$H_g$ - approx: ($g$: size of largest set)

Dual fitting:

$$y_i \leftarrow \frac{\text{price}(e_i)}{H_g}$$ is a feasible sol. to dual