An Approximation for Set Cover

V_C

Select from $v_1 \cdots v_m$
s.t. $e_1 \cdots e_n$ are covered

- Select an arbitrary edge e_i
- add endnodes of e_i to selected set of nodes
- remove edges connected to selected nodes
- Continue until all edges are covered
  ↳ Approx-ratio $= 2$

Set_Cover

Select from $S_1 \cdots S_m$
s.t. $u_1, \cdots, u_n$ are covered

- select an element u_i
- add all sets including u_i to selected sets
- remove all elements that belong to a selected set
- Continue until all elements are covered
  ↳ Approx-ratio $= X$?

X is the max # of sets an element belong to

X

Greedy Approximation Alg. for Set-Cover

$O(m(n+m))$

- $A = \{\ \}$
- Until all elements are covered
  - Select the set S_i that covers the max # of uncovered elements
  - Add S_i to A
  - Mark all elements in S_i as covered
  - $\forall u_j \in S_i \setminus \text{Covered} : e_j = \frac{1}{|S_i \setminus \text{Covered}|}$
- return A

Observation: $|A| = \sum_{\forall u_j \in U} e_j$

- G-Set-Cover has the approx. ratio of $\log(n)$.

Order $e_1 \cdots e_n$ s.t. e_1 is the smallest

$$\overbrace{e_1 \cdots e_{i-1}}^{i-1} \Big\{ \overbrace{e_i \cdots e_n}^{n-(i-1)=n-i+1}$$

optimum selects OPT sets

optimum pays $\frac{OPT}{n-i+1}$ to cover $e_i \cdots e_n$

$$e_i \le \frac{OPT}{n-i+1}$$

⇐ because Greedy picks the most cost-effective set

$$|A| = \sum_{i=1}^{n} e_i \le \sum_{i=1}^{n} \frac{opt}{n-i+1} = OPT \sum_{i=1}^{n} \frac{1}{n-i+1}$$

$$= OPT\left[\frac{1}{n-1} + \frac{1}{n-2} \cdots + \frac{1}{1}\right]$$

$$= OPT \sum_{i=1}^{n-1} 1/i = OPT\, H_n$$

$$\boxed{\sum_{i=1}^{n} 1/i \le \int_{i=1}^{n} 1/i \, di = \ln i \Big|_{i=1}^{n} = \ln(n)}$$

$$\Rightarrow |A| = OPT\, H_n = O(\log n)\, Opt$$

$$\Rightarrow \frac{|A|}{opt} = O(\log n)$$