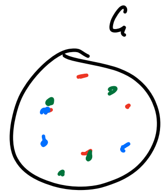
Group Steiner Tree:

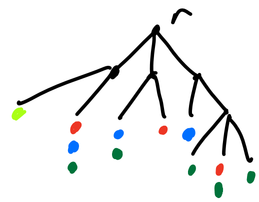
Input: - $G=(V,E)$

- edge costs $c: E \to \mathbb{R}_{\geq 0}$
- vertex r (root)
- k groups $g_1, g_2, \dots, g_k$, each $g_i \subseteq V$



Feasible: Tree $T \subseteq E$ s.t. $\forall i \in [k]$, $\exists v \in g_i$ where T has an r–v path

Objective: $\min c(T) = \sum_{e \in T} c(e)$



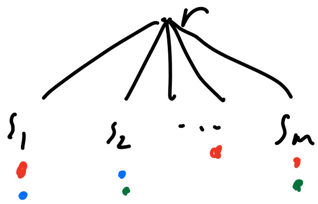
Today: G a tree (GST on trees)

$\Rightarrow$ WLOG, all terminals are leaves

Thm: Set Cover is a special case of GST on trees

Pf: Let $(U, \mathcal{S})$ instance of set cover

$\Rightarrow$



$\forall i \in U$, $g_i = \{S \in \mathcal{S} : i \in S\}$

$c(e) = 1 \quad \forall e \in E$

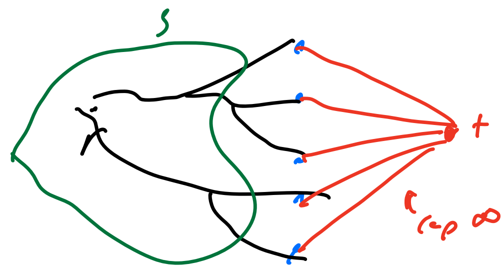
$\Rightarrow$ Set cover is valid GST solution.

any GST is a set cover

Cor: NP-hard to approximate better than $\Omega(\log n)$

Thm [Halperin-Krauthgamer '03]: $\forall$ constant $\varepsilon > 0$, NP-hard to approximate GST on trees better than $\Omega(\log^{2-\varepsilon} n)$

Thm [Garg, Konjevod, Ravi '00]: There is an $O(\log n \log k)$-approximation algorithm for GST on trees.



LP relaxation:

$$\min \sum_{e \in E} c(e)\, x_e$$

$$\text{s.t.} \quad \sum_{e \in (S, \bar{S})} x_e \geq 1 \qquad \forall i \in [k], \forall S \subseteq V \text{ s.t. } r \in S, g_i \cap S = \emptyset$$

$$0 \leq x_e \leq 1$$

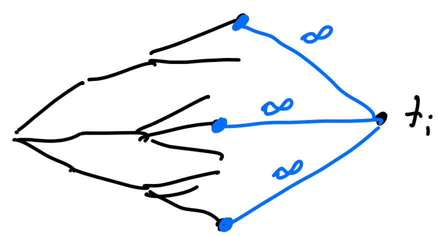
Thm: ILP (add $x_e \in \{0,1\}$ constraints) is an exact formulation

Solve LP:

Ellipsoid + Separation Oracle!

Sep oracle: Given x, is there an S separating r from g_i with $\sum_{e \in \delta(S,\bar{S})} x_e < 1$?

Min cut!

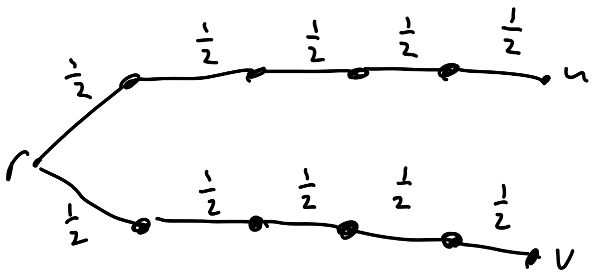


Equivalent interpretations of LP (by max-flow = min-cut):

1) Find x s.t. every cut separating r from g_i has ≥ 1 total x across

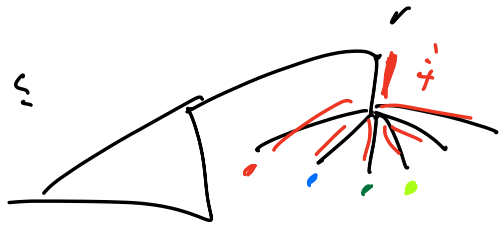
2) Find x s.t. if we think of x as capacities, we can send 1 flow from r to g_i

Rounding: Can't round each edge independently!



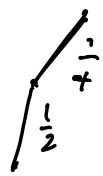
$g_1 = \{u, v\}$

$\Pr[r \text{ connected to } g_1] \leq 2 \cdot \left(\frac{1}{2}\right)^{n/2}$



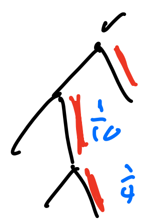
Inflation (like set cover) not helpful.

Def: Let $p(e)$ = parent edge of e

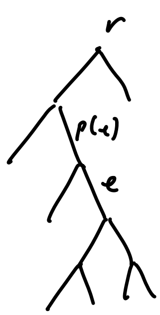


Lemma: $x_{p(e)} \geq x_e \quad \forall e$ in any optimal solution x

Pf sketch:



Obvious when think of capacities



If $x_{p(e)} < x_e$, then cannot utilize x_e capacity at e!

$\Rightarrow$ Decrease x_e to $x_{p(e)}$, get better solution

$\Rightarrow\Leftarrow$ x optimal

Basic Alg (GKR Rounding)

- Solve LP to get x
- For each $e \in E$:
  - mark e with probability $\frac{x_e}{x_{p(e)}}$. If e has no parent edge (incident on r), mark with prob. x_e
- For each $e \in E$: include in T if marked and all ancestors marked.

Lemma: $\Pr[e \text{ in } T] = x_e$

Pf:

$$\Pr[e \text{ in } T] = \frac{x_e}{\cancel{x_{p(e)}}} \cdot \frac{\cancel{x_{p(e)}}}{\cancel{x_{p(p(e))}}} \cdot \frac{\cancel{x_{p(p(e))}}}{\cancel{x_{p^2(e)}}} \cdot \ldots \cdot \frac{\cancel{x_{p^{i-1}(e)}}}{\cancel{x_{p^i(e)}}} \cdot \cancel{x_{p^i(e)}}$$

$$= x_e$$

Corollary: $E[ALG] \leq LP$

Pf:

$$E[ALG] = E\Big[\sum_{e \in E} c(e) \cdot \mathbb{1}[e \in T]\Big] = \sum_{e \in E} c(e)\, x_e = LP$$

Thm: $\forall i \in [k]$,

$$\Pr[g_i \text{ connected to } r \text{ in } T] \geq \frac{1}{O(\log |g_i|)} \geq \frac{1}{O(\log n)}$$

First: use thm to get $O(\log n \log k)$-approx.

- Run basic alg $\Theta(\log n \log k)$ times, take union
- If g_i not connected to r, add cheapest r-g_i path P_i

Note: $c(P_i) \leq OPT$

$\Rightarrow$ Feasible with probability $\underline{1}$

$\Pr[g_i \text{ not connected to } r \text{ in union}] \leq$

$$\leq \left(1 - \frac{1}{O(\log n)}\right)^{\Theta(\log n \log k)} \leq e^{-\Theta(\log k)} \leq \frac{1}{k}$$

$$\Rightarrow E[c(ALG)] \leq \Theta(\log n \log k) \cdot LP + \sum_{i=1}^{k} \frac{1}{k} \cdot c(P_i)$$

$$\leq O(\log n \log k) \cdot OPT + \sum_{i=1}^{k} \frac{1}{k} \cdot OPT$$

$$\leq O(\log n \log k) \cdot OPT$$

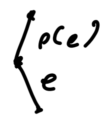
Rest of class: proof of thm

<u>Thm</u>: $\forall i \in [k]$,

$$\Pr[g_i \text{ connected to } r \text{ in } T] \ge \frac{1}{O(\log |g_i|)} \ge \frac{1}{O(\log n)}$$

Let $g = g_i$ be a group

Let FAIL be event that r not connected to g



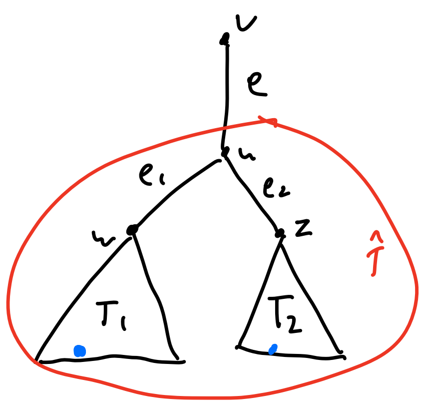
<u>Lemma</u>: If $x'_e \le x_e$ $\forall e$, then

$$\Pr[\text{FAIL using } x'] \ge \Pr[\text{FAIL using } x]$$

<u>Pf sketch</u>:

One edge at a time, induction.

Single edge e with $x'_e < x_e$, all other $\hat{e}$ have $x'_{\hat e} = x_{\hat e}$



$e = \{u, v\}$

$\hat{T}$: subtree rooted at u

Nothing outside of subtree rooted at u changes

$p_1 = \Pr[\text{fail to connect } T_1 \cap g \text{ to } w]$

$p_2 = \Pr[\text{fail to connect } T_2 \cap g \text{ to } z]$

$\Pr[\text{fail to connect } \hat{T} \cap g \text{ to } v \text{ using } x] =$

$$= \underbrace{(1-x_e)}_{\Pr[e \text{ not added}]} + x_e\left(\left(\underbrace{\left(1-\frac{x_{e_1}}{x_e}\right)}_{\Pr[e_1 \text{ not marked}]} + \frac{x_{e_1}}{x_e}\underbrace{p_1}_{\text{fail below } e_1}\right)\left(\underbrace{\left(1-\frac{x_{e_2}}{x_e}\right)}_{\Pr[e_2 \text{ not marked}]} + \frac{x_{e_2}}{x_e}\underbrace{p_2}_{\text{fail below } e_2}\right)\right)$$

$$= 1 - x_{e_1}(1-p_1) - x_{e_2}(1-p_2) + \frac{x_{e_1}x_{e_2}(1-p_1)(1-p_2)}{x_e}$$

increases as x_e decreases!

So we can use smaller x' in analysis:

if $\Pr[g \text{ connected to } r \text{ in } T \text{ using } x'] \geq \frac{1}{\log|g|}$, then

$\Pr[g \text{ connected to } r \text{ in } T \text{ using } x] \geq \frac{1}{\log|g|}$

Create x':

1) Remove all leaves not in g, all unnecessary edges

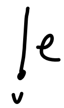
2) Reduce x values until minimally feasible

(exactly 1 flow can be sent from r to g,

min r-g cut $= 1$)

In modified instance, WTS $\Pr[r \text{ connected to } g] \geq \frac{1}{O(\log|g|)}$

$\Pr[\text{connect } v \in g \text{ to } r] = x_{e(v)}$



$\Rightarrow E[\#\text{ nodes in } g \text{ connected to } r] = \sum_{v \in g} x_{e(v)} \geq \frac{1}{4}$

But lots of dependence!