Group Steiner Tree and Tree Embeddings:

Input:
- $G=(V,E)$
- edge costs $c: E \to \mathbb{R}_{\geq 0}$
- vertex r (root)
- k groups $g_1, g_2, \ldots, g_k$, 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

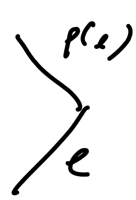
Objective: $\min \; c(T) = \sum_{e \in T} c(e)$

LP relaxation:

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

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

$$0 \leq x_e \leq 1$$



<u>Def</u>: Let $p(e)$ = parent edge of e

<u>Lemma</u>: $x_{p(e)} \geq x_e$ $\forall e$ in any optimal solution x

<u>Basic Alg</u> (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.

<u>Lemma</u>: $\Pr[e \text{ in } T] = x_e$

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

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

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

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

Create x':

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

2) Reduce x values until minimally feasible

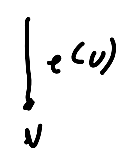
   (exactly 1 flow can be sent from r to g,
   min r-g cut $= 1$)

3) Round each x_e down to power of $\frac{1}{2}$

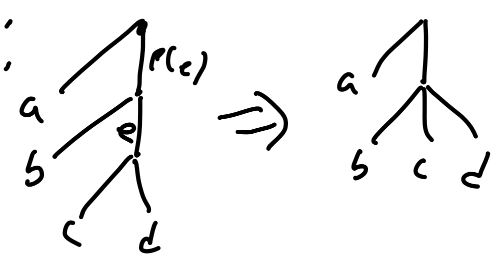
   (can send $\ge \frac{1}{2}$ flow, min cut $\ge \frac{1}{2}$)

4) Delete all edges with $x_e \le \frac{1}{4|g|}$

   (at most $|g|$ leaves, so flow $\ge$
   $\frac{1}{2} - |g| \frac{1}{4|g|} = \frac{1}{4}$)



5) If $x_e = x_{p(e)}$, contract e:

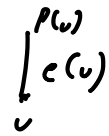


(e will be marked with probability 1)

Lemma: Height of tree $\leq O(\log |g|)$

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

Def: For $v \in V$, let $e(v)$ be edge from v to $p(v)$



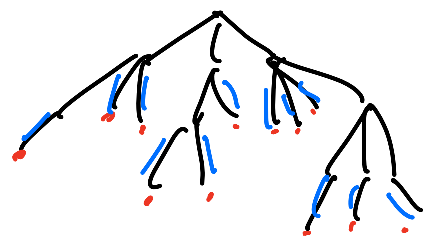
$$E[\,|\{v \in g : v \text{ connected to } r\}|\,] = E\Big[\sum_{v \in g} \mathbb{1}[v \text{ connected to } r]\Big]$$

$$= \sum_{v \in g} \Pr[v \text{ connected to } r]$$

$$= \sum_{v \in g} x_{e(v)}$$

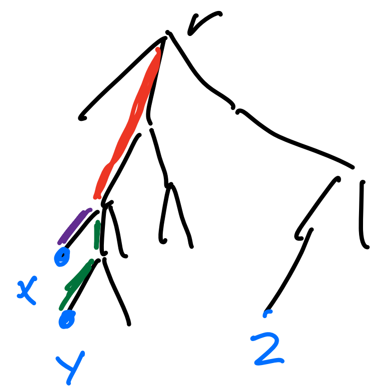
$$= \text{total flow to } g$$

$$\geq 1 \text{ (using } x\text{)} \quad \text{or} \quad \geq \tfrac{1}{4} \text{ (using } x'\text{)}$$



$\Rightarrow$ if concentrated around expectation, would have reasonable probability of connecting g to r

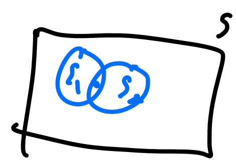
But lots of dependence!



New tool: <u>Janson's Inequality</u>

Setup:

- S ground set of items
- $S_1, \dots, S_k$ subsets of S
- $p_e \in [0,1]$ $\forall e \in S$
- S': set obtained by adding each $e \in S$ independently with probability p_e



- $\mathcal{E}_i$ = event that $S_i \subseteq S'$
- $\mu = \sum_{i=1}^{k} \underbrace{\Pr[\mathcal{E}_i]}_{\prod_{e \in S_i} p_e}$ (expected # events that occur)
- $\Delta = \sum_{i \sim j} \Pr[\mathcal{E}_i \cap \mathcal{E}_j]$, where $i \sim j$ if $S_i \cap S_j \neq \emptyset$
  ($\mathcal{E}_i, \mathcal{E}_j$ dependent)

<u>Thm</u> [Janson's Inequality]: If $\mu \le \Delta$, then probability that none of the events occur is

$$\Pr\left[\bigcap_{i=1}^{k} \bar{\mathcal{E}}_i\right] \le e^{-\frac{\mu^2}{2\Delta}}$$

Use Janson for Group Steiner Tree:

- $S = E$
- $p_e = \Pr[e \text{ marked}] = \frac{x'_e}{x'_{p(e)}}$
- S_i = path from r to $v_i \in g$
- $\mathcal{E}_i$ = event that all edges on S_i are marked
  (v_i connected to r, so g connected to r)

<u>Claim</u>: $1 \ge \mu = \sum_i \Pr[\mathcal{E}_i] \ge \frac{1}{4}$

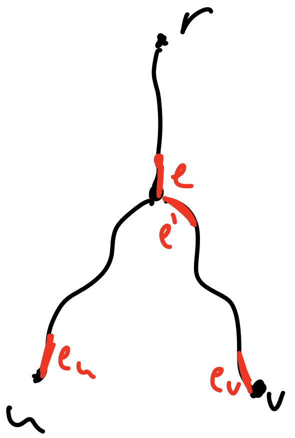
<u>Pf</u>: Already showed $\mu \ge \frac{1}{4}$

$\mu \le 1$ since x' minimally feasible for g

Claim: $\Delta = O(\log |g|)$

Pf: Let $H = O(\log |g|)$ be height of tree

$$\Delta = \sum_{i \sim j} \Pr[\mathcal{E}_i \cap \mathcal{E}_j] = \sum_{u \in g} \underbrace{\sum_{\substack{v \in g: \\ LCA(u,v) \neq r}} \Pr[\mathcal{E}_u \cap \mathcal{E}_v]}_{\Delta_u}$$



e = lowest edge shared by S_u, S_v

$$\Pr[\mathcal{E}_u] = x'_{e_u}$$

$$\Pr[\mathcal{E}_v \mid \mathcal{E}_u] = \frac{x'_{e_v}}{\cancel{x'_{p(e_v)}}} \cdot \frac{\cancel{x'_{p(e_v)}}}{\cancel{x'_{p(p(e_v))}}} \cdots \frac{\cancel{x'_{e'}}}{x'_e}$$

$$= \frac{x'_{e_v}}{x'_e}$$

$$\Rightarrow \Pr[\mathcal{E}_u \cap \mathcal{E}_v] = \Pr[\mathcal{E}_v \mid \mathcal{E}_u] \cdot \Pr[\mathcal{E}_u] = \frac{x'_{e_u} x'_{e_v}}{x'_e}$$

Fix $u \in g$, let $\Delta_u = \sum_{\substack{v \in g: \\ LCA(u,v) \neq r}} \Pr[\mathcal{E}_u \cap \mathcal{E}_v]$

$$\Rightarrow \Delta = \sum_{u \in g} \Delta_u$$

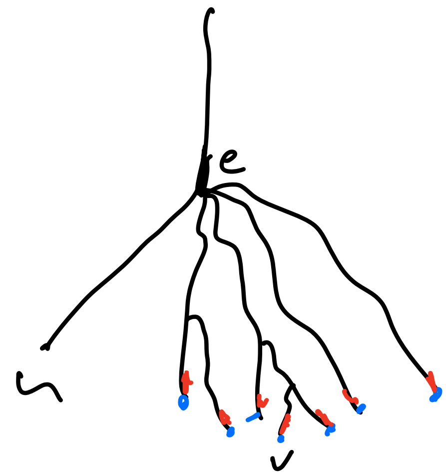
Let $F(e) = \{ v \in g : e \text{ lowest edge in } S_u \cap S_v \}$

$$\Rightarrow \Delta_u = \sum_{\substack{v \in g: \\ LCA(u,v) \neq r}} \Pr[\mathcal{E}_u \wedge \mathcal{E}_v]$$

$$= \sum_{e \in S_u} \sum_{v \in F'(e)} \Pr[\mathcal{E}_u \wedge \mathcal{E}_v]$$

$$= \sum_{e \in S_u} \sum_{v \in F'(e)} \frac{x'_{e_u} \, x'_{e_v}}{x'_e}$$

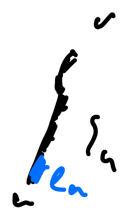
$$= \sum_{e \in S_u} \frac{x'_{e_u}}{x'_e} \sum_{v \in F'(e)} x'_{e_v}$$



$$\leq \sum_{e \in S_u} \frac{x'_{e_u}}{x'_e} \, x'_e$$

(flow!)

$$= \sum_{e \in S_u} x'_{e_u}$$



$$= x'_{e_u} \sum_{e \in S_u} 1 \leq H \, x'_{e_u}$$

$$\Rightarrow \Delta = \sum_{u \in g} \Delta_u \leq H \sum_{u \in g} x'_{e_u} \leq H$$

total flow to $g \leq 1$

Now plug into Janson:

Pr[successfully connect r to g]

$= 1 - \Pr[\text{fail to connect } r \text{ to } g]$

$\geq 1 - e^{-\frac{\mu^2}{2\Delta}}$

$\geq 1 - e^{-\frac{\left(\frac{1}{4}\right)^2}{O(\log|g|)}} = 1 - e^{-\frac{1}{O(\log|g|)}}$

$\geq \dfrac{\frac{1}{O(\log|g|)}}{1 + \frac{1}{O(\log|g|)}}$ $\quad\left(1 - e^{-x} \geq \frac{x}{x+1} \quad \forall x > -1\right)$

$= \dfrac{1}{O(\log|g|)}$

Tree Embeddings:

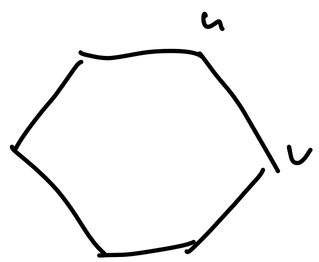
Goal: convert a graph into a tree so we can solve problem (GST) on tree.

Intuition: preserve distances.

Given $G=(V,E)$, is there a tree T s.t.

$$d_G(u,v) \approx d_T(u,v) \quad \forall u,v \in V?$$

C_n:



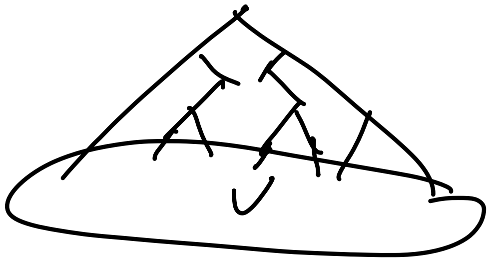
Remove any edge:

distance changes from 1 to $n-1$

Intuitive fix: choose edge to remove randomly

$$\Rightarrow E[d_T(u,v)] = \frac{1}{n}\cdot(n-1) + \frac{n-1}{n}\cdot 1 = 2\left(\frac{n-1}{n}\right) = 2\left(1-\frac{1}{n}\right)$$

Def: A tree metric (V',T) for a set of nodes V is a tree T with vertices V' s.t. $V \subseteq V'$ are the leaves of T, and a nonnegative length for each edge in T



So tree metric for V is tree with V as <u>leaves</u>, gives distances between leaves

<u>Def</u>: Let (V,d) be a metric space, (V',T) a tree metric for V. Then (V,d) embeds into T with distortion α if $d(u,v) \le d_T(u,v) \le \alpha \cdot d(u,v)$ $\forall u,v \in V$

<u>Thm [Fakcharoenphol, Rao, Talwar '03]</u>:

Let (V,d) be a metric space. Then there is a randomized, polytime algorithm that produces a tree metric (V',T) for V s.t.

1) $d(u,v) \le d_T(u,v)$ $\forall u,v \in V$
2) $E[d_T(u,v)] \le O(\log n) \cdot d(u,v)$ $\forall u,v \in V$

Embeddings into a distribution of trees (dominating trees)

GST on general graphs using FRT

1) Extend costs to metric space:

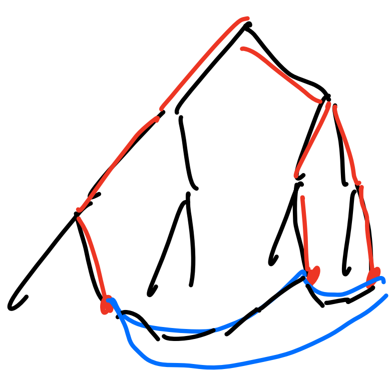
$$c(u,v) = \text{min cost } u\text{-}v \text{ path in } G$$

2) Use FRT to embed into tree $(V;T)$

- distortion $O(\log n)$

3) Use GKR to get tree T' which is $O(\log n \log k)$-approx on T

4) "Shortcut" T' to get cycle C only on terminals



5) Remove arbitrary edge of C to get path, replace each edge of C by min-cost path in G. Return spanning tree H.

Thm: Returns a feasible solution

Pf: T' connects ≥ 1 node from each group

$\Rightarrow C$ has ≥ 1 node from each group

$\Rightarrow H$ has ≥ 1 node from each group ✓

Thm: $E[c(H)] \leq O(\log^2 n \log k) \cdot OPT$

Pf:

Notation:

- S = terminals connected by OPT (so $S \cap g_i \neq \emptyset \ \forall i$)
- Let C_S be cycle on S from shortcutting OPT $\Rightarrow c(C_S) \leq 2 \cdot OPT$
- Let c_T be cost/distance in T (earlier d_T)
- $OPT(T)$ = optimal solution in T
- T_S = subtree of T induced by S (paths from S to $LCA(S)$)

$$E[c(H)] \le E[c(C)]$$ (H spanning tree of C)

$$\le E[c_T(C)]$$ (distances non-decreasing: $c(u,v) \le c_T(u,v)$)

$$\le 2 \cdot E[c_T(T')]$$ (shortcutting costs 2)

$$\le 2 \cdot E[O(\log n \log k) \cdot c_T(OPT(T))]$$ (GKR-approx)

$$= O(\log n \log k) \cdot E[c_T(OPT(T))]$$ (linearity of expectation)

$$\le O(\log n \log k) \cdot E[c_T(T_S)]$$ (def of OPT(T))

$$\le O(\log n \log k) \cdot E[c_T(C_S)]$$ (C_S a cycle on leaves of T_S)

$$= O(\log n \log k) \cdot E\Big[\sum_{(u,v)\in C_S} c_T(u,v)\Big]$$ (def)

$$= O(\log n \log k) \cdot \sum_{(u,v)\in C_S} E[c_T(u,v)]$$ (linearity of expectation; C_S deterministic)

$$\le O(\log n \log k) \cdot \sum_{(u,v)\in C_S} O(\log n) \cdot c(u,v)$$ (FRT)

$$= O(\log^2 n \log k) \cdot \sum_{(u,v)\in C_S} c(u,v)$$ (algebra)

$$\leq O(\log^2 n \log k) \cdot 2 \cdot OPT$$ (Cs shortcutted OPT)

$$= O(\log^2 n \log k) \cdot OPT$$

Metric Embeddings in General (sketch)

Def: (V,d) embeds into (V,d') with distortion α if

$$d(u,v) \le d'(u,v) \le \alpha \cdot d(u,v) \quad \forall u,v \in V$$

Sps have a β-approx for problem in d', but not in d

ALG: Embed into d', solve there

If costs = sums of distances, then

$$c(ALG) = \sum_{(u,v)\in ALG} d(u,v) \le \sum_{(u,v)\in ALG} d'(u,v)$$

$$\le \beta \sum_{(u,v)\in OPT(d')} d'(u,v)$$

$$\le \beta \sum_{(u,v)\in OPT} d'(u,v)$$

$$\le \beta\alpha \sum_{(u,v)\in OPT} d(u,v)$$

$$= \beta\alpha \cdot c(OPT)$$