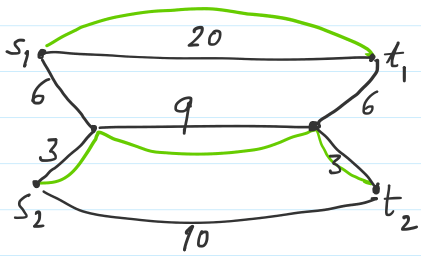
Generalized Steiner Tree (Steiner Forest)

This is an example of a network design problem. We are given an undirected graph $G(V,E)$ with $c: E \to \mathbb{R}^{+}_{\geq 0}$ being the cost/length on the edges (we do not need metric here).
Further, we are given terminal pairs $\{s_i, t_i\} \in V$, $i = 1, 2, \dots, k$

A Steiner Forest is a subgraph that has at least one path between s_i and t_i, $\forall i = 1, 2, \dots k$.
Objective: Find the min cost Steiner Forest.

Green edges form a Steiner forest (Not optimal one!)



Special Cases: i) $s_i = s$ $\forall i = 1, 2, \dots k$: rooted Steiner Tree problem
ii) Every pair of vertex forms a pair : Spanning Tree . . .

A Primal-Dual Algorithm

Primal: $\min \sum_{e \in E} c_e \cdot x_e$

s.t.: $\boxed{\sum_{e \in \delta(S)} x_e \geq 1, \ \forall S \subseteq V, \ S \odot \{s_i, t_i\} \text{ for some } i = 1, 2, \dots, k}$

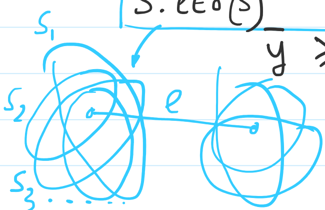
$\bar{x} \geq 0$

Cut constraints.
Use Menger's Thm to prove that any 0/1 solution $\Leftrightarrow$ Steiner Forest

Dual: $\max \sum_S y_S$

s.t.: $\boxed{\sum_{S : e \in \delta(S)} y_S \leq c_e} \ \forall e \in E$

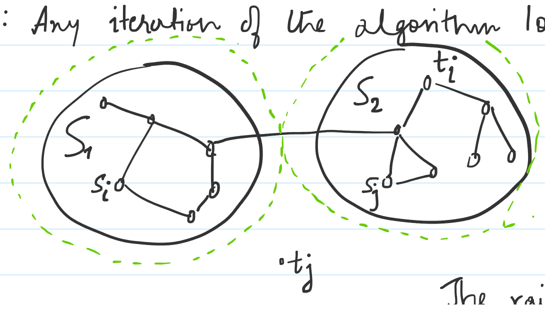
$y \geq 0$



Total "mass" of all cuts that has an endpoint of e should not exceed cost of the edge

Algorithm: [Due to Agrawal, Klein, Ravi / Goemans & Williamson]

Idea: Any iteration of the algorithm looks like this:



We shall raise the dual values of all the "active" cuts simultaneously at the same rate.
"active cut": That separates some $\{s_i, t_i\}$
The raising is continued till two "balls"

The raising is continued till two "balls" collide. In other words, for some edge, total dual from the two ends $= c_e$. Then, S_1 & S_2 are declared to be "dead" balls and the ball $S_1 \cup S_2$ is now "active", provided it separates some $\{s_j, t_j\}$. Note that, collision of S_1 & S_2 creates a path between s_i and t_i. Now we formalize the algorithm a bit.

Algorithm:

$\$ = \{s_i\}, \{t_i\}\ \forall i = 1, 2, \ldots, k$ # Set of active duals

$F = \phi$ # Set of tight edges

$y = 0$

$\ell = 0$

While some $\{s_i, t_i\}$ is not connected in (V, F)

Increase y_S, $S \in \$$ simultaneously until for some $e \in \delta(S)$, $S \in \$$, $\sum_{S: e \in \delta(S)} y_S = c_e$.

$F \leftarrow F \cup \{e\}$

Let S, S' be the active dual balls that collided.

$\$ \leftarrow \$ - \{S, S'\} \cup \{S \cup S'\}$

$F' \leftarrow F$

For $p \leftarrow \ell$ downto 1

if $F' - \{e_\ell\}$ is still feasible,

$F' \leftarrow F' - \{e_\ell\}$

Return F'

Cleanup Step

Also called "Reverse Delete"

A crucial Step for the analysis to work.

Analysis: We need to bound the total cost of edges in F'.

(Convince yourself F' is a feasible solution)

$$c(F') = \sum_{e \in F'} c_e = \underbrace{\sum_{e \in F'} \sum_{S: e \in \delta(S)} y_S}_{\text{Since we include an edge when it is tight}}$$

We would like to show $c(F') \leq 2 \sum y_S$.

Hence, it suffices to prove that $\sum_{e \in F'} \sum_{S: e \in \delta(S)} y_S \leq 2 \sum_S y_S$

or $$\boxed{\sum_S \deg_{F'}(S) \cdot y_S \leq 2 \sum_S y_S}$$

We prove this by induction

Lemma 1. $\sum_S \deg_{F'}(S) y_S \le 2 \sum_S y_S$

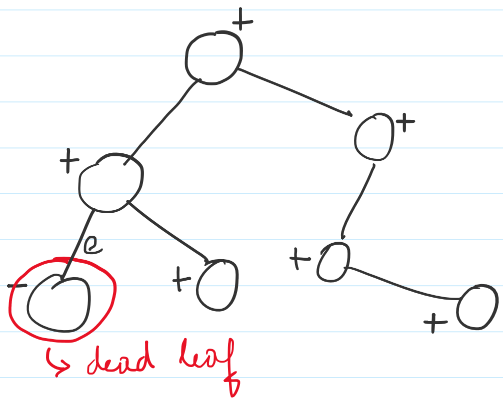
<u>Proof</u>: The base case is trivial since $y_S = 0 \; \forall S$ initially

<u>Induction</u>: Let the increase in all active duals in any iteration be Δ.

Then it suffices to prove that $\Delta \sum_{S \in \$} \deg_{F'}(S) \le 2\Delta |\$|$

or $\sum_{S \in \$} \deg_{F'}(S) \le 2|\$|$ ——— (1)

Let us F', the final set of edges on $\$$ at any iteration



Contract the subgraph induced by the sets in $\$$ into supernodes. Consider any connected component of $(V_\$, F')$, say T'

$V_\$$ = Set of supernodes

<u>Claim 1</u>: T' is a tree

<u>Claim 2</u>: T' does not contain any dead leaf.

Both the claims follow from the Reverse Delete Step.
It is simple to argue that T' cannot contain cycles!
To see claim 2, assume T' contains a dead leaf and e be the solitary edge incident upon it. But then e must have been removed in the Reverse Delete Step.

Now we are ready to finish off the proof of (1).

The average degree of any tree ≤ 2. Since the average degree of all dead nodes is ≥ 2 (since they are non-leaf nodes, Claim 2), the avg. degree of all active nodes is ≤ 2. ∎

<u>Thm</u>: There exists a polynomial time 2-approximation for Steiner Forest.
(The ratio, in fact, is $(2 - 1/k)$, Can you prove this?)

<u>Remarks</u>: 1. The above algorithm in fact works for more general settings. Consider a general network design problem whose constraints can be expressed as

$$\sum x_e \ge f(S) \quad \forall S \subseteq V$$

$$e \in \delta(S)$$

where $f(S)$ is a 0/1 proper function. Then the above algorithm is a 2-approximation for such a problem.