MST = minimum spanning tree

Given undirected $G=(V,E)$ with positive edge weights $w(e)>0$

find min weight connected subgraph.

For $S \subseteq E$, $w(S) = \sum_{e \in S} w(e)$.

min weight connected subgraph is a tree, hence called MST problem.

---

## Basic graph theory:

1) Tree on $n$ vertices has $n-1$ edges.
2) Any connected graph $G=(V,E)$ with exactly $|V|-1$ edges is a tree
3) In a tree there is exactly one path between every pair of vertices.

Note, forest = acyclic graph.
thus a tree is a connected forest.

Two algorithms: Kruskal's & Prim's

greedy (Kruskal's)    similar to Dijkstra's (Prim's)

## Kruskal's alg.:

For input $G=(V,E)$:

1. Sort $E$ by $\uparrow$ weight
2. Let $X=\emptyset$.
3. Go through $E$ in sorted $\uparrow$ order:

   For edge $e=(y,z)$

   If $X \cup e$ is acyclic then add $e$ into $X$.

---

How do we test if $X \cup e$ has a cycle?

Check if in graph $(V,X)$, $y$ & $z$ are connected.

Do this using union/find data structure, we'll cover it next class.

Why is Kruskal's alg. correct?

Let's prove by induction.

Suppose current $X$ is correct

means there is a MST $T$ where $X \subset T$.

Consider next edge $e=(y,z)$ that Kruskal's adds to $X$.

Need to show there is a MST $T'$ where $X \cup e \subset T'$.

Let $G'$ be the graph $G' = (V, X)$ on edges $X$.

let $c(y)$ be the component containing $y$

& $c(z)$ be the cmpt. for $z$.

We know $c(y) \neq c(z)$ since $y$ & $z$ are disconnected in $G'$.

We also know $e$ is the min weight edge

from $c(y) \longleftrightarrow V - c(y)$ (rest of graph)

& $c(z) \longleftrightarrow V - c(z)$.

Why? if there is a smaller weight edge $e'$ crossing these cuts (either one) then we would have added $e'$ earlier.

We'll use the following cut property to prove that $X \cup e \subset T'$ for a MST $T'$.

## Cut property:

For $G=(V,E)$, consider $X \subset E$ where $X \subset T$ for a MST $T$.

Take any subset $S$ of vertices where no edge of $X$ crosses $S \leftrightarrow \bar{S}$.

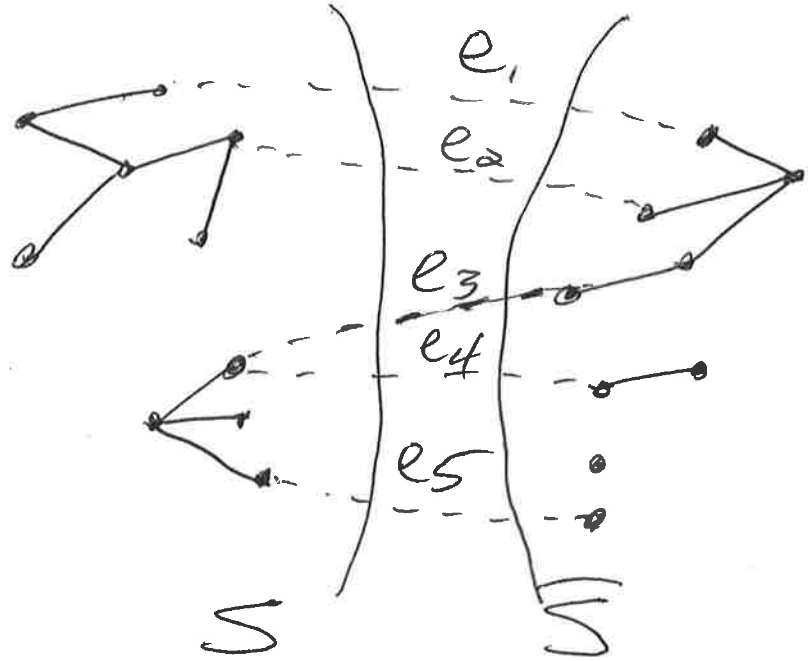
Let $e^*$ be the minimum weight edge of $E$ between $S$ & $\bar{S}$

so $e^*$ is the min in the whole graph.

Then, $X \cup e^* \subset T'$ for a MST $T'$.

Example: $X =$ —— solid edges

other edges of $E =$ - - - - dashed edges




If $w(e_5) \leq w(e_1), w(e_2), w(e_3), w(e_4)$

~~then~~ and $X \subset T$ for a MST $T$

then $X \cup e_5 \subset T'$ for a MST $T'$.

(might be that $T \neq T'$)

For Kruskal's set $S = c(y)$
or $S = c(z)$ (both work)

Why is it called the <u>cut</u> property?

A cut is a partition of $V = S \cup \overline{S}$ into 2 sets $S \longleftrightarrow \overline{S}$

The lemma says that the minimum weight across a cut is part of a MST.

## Proof of the cut property:

We know $X \subset T$ where $T$ is a MST.

2 cases: $e^* \in T$ or $e^* \notin T$.

if $e^* \in T$ then $X \cup e^* \subset T$ so we're done.

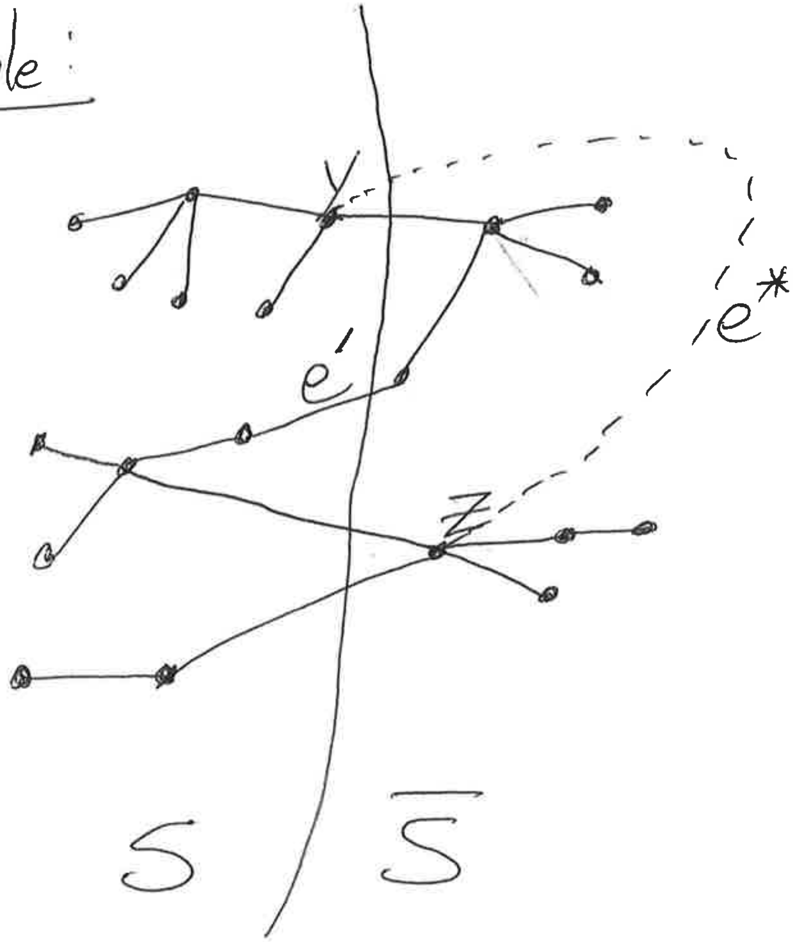
Suppose $e^* \notin T$. Say $e^* = (y, z)$.

Add $e^*$ to $T$.

$T$ is a tree so there's 1 path b/w $y$ & $z$, call it $P$.

Then ~~$\not\in$~~ $T \cup e^*$ has a cycle $C = P \cup e^*$.

Example:




Note, $y \in S$ & $z \in \bar{S}$ because $e^* = (y, z)$ crosses $S \leftrightarrow \bar{S}$.

Thus, $P$ crosses $S \leftrightarrow \bar{S}$ at least once (the above example crosses 3 times)

Take one of the edges of $P$ crossing $S \leftrightarrow \bar{S}$, call it $e'$. (choose any one of the 3.) in above example.

Set $T' = T \cup e^* - e'$.

**Claim:** (i) $X \cup e^* \subset T'$

(ii) $T'$ is a tree

(iii) $w(T') \leq w(T)$

Note, (ii)+(iii) implies $T'$ is a MST (since $T$ is a MST)

& since (i) says $X \cup e^* \subset T'$ then we're done with the proof.

**Proof of claim:**

(i): No edge of $X$ crosses $S \Leftrightarrow \bar{S}$.

Since $e'$ crosses $S \Leftrightarrow \bar{S}$ thus $e' \notin X$.

~~So~~ Since $X \subset T$ & ~~also~~ then $X \subset T - e'$,

and thus $X \cup e^* \subset \underbrace{T \cup e^* - e'}_{=T'}$ ∎

(iii): $e^*$ has min weight of edges crossing $S \leftrightarrow \bar{S}$.

Thus, $w(e^*) \le w(e')$,

and so: $w(e^*) - w(e') \le 0$.

Therefore,

$$w(T') = w(T) + \underbrace{w(e^*) - w(e')}_{\le 0} \le w(T).$$

■

(ii): $T'$ has $n-1$ edges so we just need to show it's connected & that implies it's a tree ($n-1$ edges & connected).

Take a pair of vertices $a$ & $b$.
There's 1 path $P'$ between $a$ & $b$ in $T$.

If $e' \notin P'$ then $P'$ is still in $T'$ & we're done.

If $e' \in P'$ then look at cycle $C = P \cup e^*$.
Say $e' = (u, v)$. Then, $C - e'$ is a path between $u$ & $v$.
So replace $e'$ in $P'$ by $C - e'$ & we get a path in $T'$ between $a$ & $b$. ■

Aside: how quickly can we implement the algorithm embedded in the proof?

For $T \cup e^*$, how quickly can we find the cycle $C$?