MINIMUM SPANNING TREE (MST)

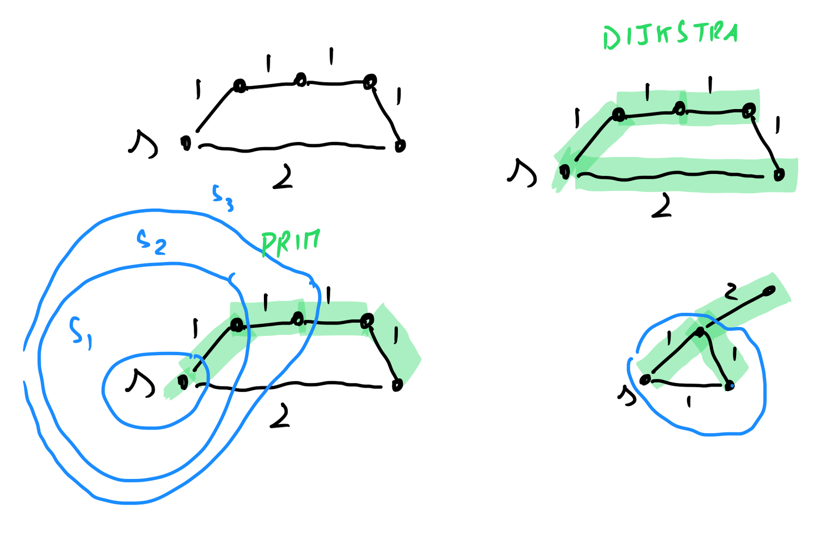
PROBLEM: GIVEN A CONNECTED WEIGHTED GRAPH $G(V,E)$, c (WHERE WEIGHTS ARE NON-NEGATIVE, I.E., $\forall e \in E: c(e) \geq 0$)

FIND A SUBSET T $T \subseteq E$ OF ITS EDGES SUCH THAT $G(V,T)$ IS CONNECTED AND $\sum_{e \in T} c(e)$ IS MINIMUM.

(THE GRAPH $G(V,T)$ IS A TREE).

① PRIM'S ALGORITHM

- SELECT ARBITRARILY A SOURCE NODE s
- $S_1 \leftarrow \{s\}$
- $T \leftarrow \emptyset$
- FOR $i=1$ TO $n-1$
  - LET e_i BE A MINIMUM COST EDGE AMONG THOSE THAT HAVE EXACTLY ONE END POINT IN S_i
  - SUPPOSE THAT $e_i = \{v, w\}$ AND THAT $v \in S_i$
  - $T \leftarrow T \cup \{e_i\}$
  - $S_{i+1} \leftarrow S_i \cup \{w\}$



② KRUSKAL'S ALGORITHM

- SORT THE EDGES INCREASINGLY BY COST
- LET $T \leftarrow \emptyset$
- FOR EACH e:
  IF e CAN BE ADDED TO T W/O CREATING CYCLES:
  $T \leftarrow T \cup \{e\}$

WHEN IS IT "SAFE" TO ADD AN EDGE TO THE TREE?

"CUT PROPERTY OF MST"

L: ASSUME THAT THE EDGE COSTS ARE ALL DISTINCT, AND THAT G IS CONNECTED.

LET $\emptyset \subset S \subset V$ BE A SUBSET OF NODES OF $G(V,E)$.

LET $e \in E$ BE THE EDGE OF $G(V,E)$ THAT HAS SMALLEST COST AMONG THE EDGES THAT CUT ACROSS S (THAT IS, AMONG THE EDGES THAT HAVE ONE END POINT IN S, AND ONE IN $V-S$).

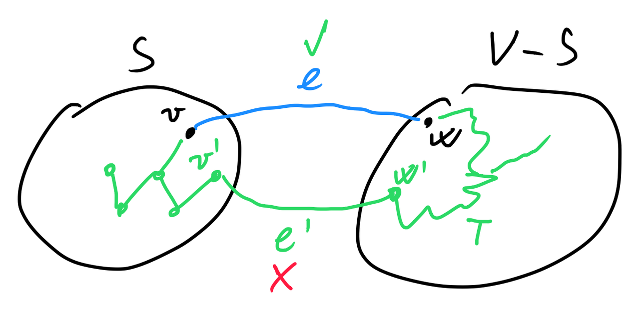
THEN, <u>EACH</u> MST OF $G(V,E)$ CONTAINS e.

P: LET T BE A SPANNING TREE THAT DOES NOT CONTAIN e. WE SHOW THAT T IS NOT A MST.

LET $e = \{v, w\}$, AND $v \in S$. THEN, $w \in V-S$.

SINCE T IS A SPANNING TREE THERE MUST EXIST A PATH π IN $G(V,T)$ FROM v TO w.

LET w' BE THE FIRST NODE OF π THAT IS IN $V-S$; LET v' BE THE NODE PRECEDING w' IN π.



$e \notin T$
$e' \in T$

LET $e' = \{v', w'\}$ BE THE EDGE THAT CONNECTS v' AND w'.

WE HAVE THAT $v' \in S$ AND $w' \in V-S$.

CONSIDER THE SET OF EDGES $T' = T - \{e'\} \cup \{e\}$

WE WANT TO PROVE THAT T' IS A SPANNING TREE, AND THAT ITS COST IS SMALLER THAN THE COST OF T. (THEN, T IS <u>NOT</u> A MST).

OBSERVE THAT $G(V,T')$ IS CONNECTED: $G(V,T)$ IS CONNECTED, AND ANY PATH IN $G(V,T)$ THAT USED THE EDGE e' CAN BE REROUTED IN $G(V,T')$ THROUGH THE EDGES OF $T' = T - \{e'\} \cup \{e\}$:

- WE CAN FIRST GO THROUGH THE PORTION OF THE PATH THAT REACHES v',
- THEN (INSTEAD OF GOING THROUGH e') WE CAN TAKE THE PATH FROM v' TO v, THEN THE EDGE e, AND THEN THE PATH FROM w TO w',
- WE CAN CONTINUE TRAVERSING THE ORIGINAL PATH FROM w'!

THEN $G(V,T')$ IS CONNECTED (AND ACYCLIC — INDEED THE ONLY CYCLE WE CREATED BY ADDING e WAS DESTROYED BY REMOVING e'). THUS $G(V,T')$ IS A TREE.

WE PROVE THAT ITS COST IS STRICTLY SMALLER THAN THE COST OF T.

$$COST(T') = \sum_{f \in T'} COST(f) = \left(\sum_{f \in T} COST(f)\right) - COST(e') + COST(e)$$

$$= COST(T) - COST(e') + COST(e)$$

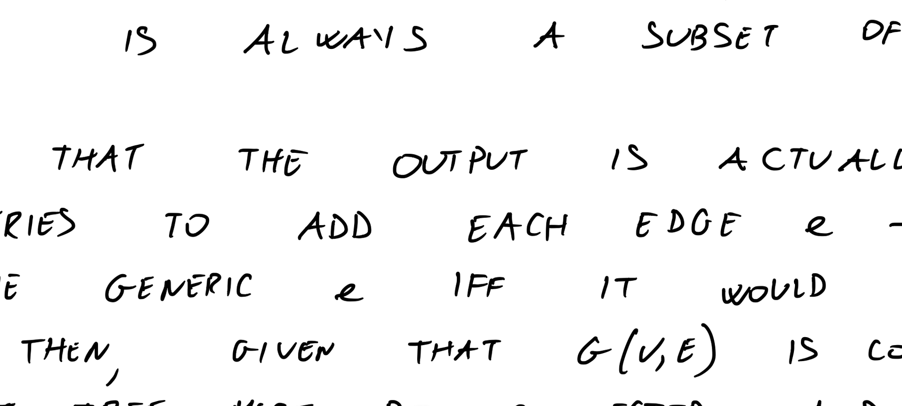
BUT $COST(e) < COST(e')$, SINCE BOTH e AND e' ARE IN THE CUT FROM S TO $V-S$ AND e IS THE EDGE OF SMALLEST COST IN THAT CUT.

GIVEN THAT $COST(T') = COST(T) - (COST(e') - COST(e))$, WE HAVE THAT $COST(T') < COST(T)$. THUS T IS NOT A MST. ∎

THM: ASSUME THAT EDGE COSTS ARE PAIRWISE DIFFERENT AND THAT $G(V,E)$ IS CONNECTED. KRUSKAL'S ALGORITHM PRODUCES A MST.

P: SUPPOSE THAT IN AN ITERATION KRUSKAL ADDS $\{v, w\}$ TO T.

LET S BE THE SET OF NODES REACHABLE FROM v IN T, BEFORE ADDING $\{v, w\}$ TO T.



THEN $v \in S$, AND $w \in V-S$. (O/W $\{v, w\}$ WOULD CREATE A CYCLE, AND KRUSKAL NEVER CREATES CYCLES).

MOREOVER $\{v, w\}$ IS THE CHEAPEST EDGE IN THE $S, V-S$ CUT (SINCE KRUSKAL CONSIDERS EDGES IN INCREASING ORDER OF COST).

THEN, BY THE PREVIOUS LEMMA (THE "CUT PROPERTY") THE EDGE $\{v, w\}$ IS PART OF EACH MST. THUS $T \cup \{\{v, w\}\}$ IS STILL A SUBSET OF A MST (A PARTIAL OPTIMAL SOLUTION). THEN, THE OUTPUT OF KRUSKAL IS ALWAYS A SUBSET OF A MST.

WE PROVE THAT THE OUTPUT IS ACTUALLY A MST. KRUSKAL TRIES TO ADD EACH EDGE e — IT AVOIDS ADDING THE GENERIC e IFF IT WOULD INDUCE A CYCLE. THEN, GIVEN THAT $G(V,E)$ IS CONNECTED, THE OUTPUT TREE MUST BE CONNECTED AND IS THUS A (MINIMUM) SPANNING TREE. ∎

THM: ASSUME THAT THE EDGE COSTS ARE PAIRWISE DIFFERENT. PRIM RETURNS A MST.

P: APPLY THE CUT PROPERTY TO $S_1, S_2, \ldots, S_{n-1}$. ∎