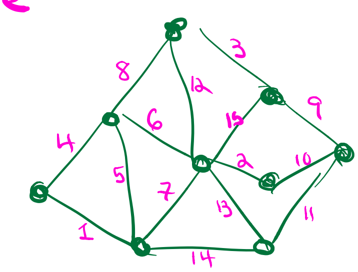
Minimum Spanning Tree

Assume $m \geq n$
graph connected.
all edge wts distinct.

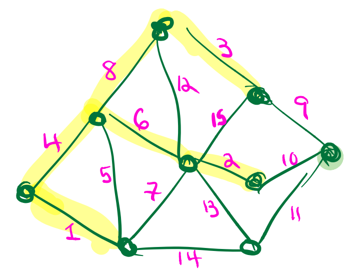
Given undirected, weighted graph $G=(V,E)$, where w_e is the weight of edge e

$n=|V|$, $m=|E|$

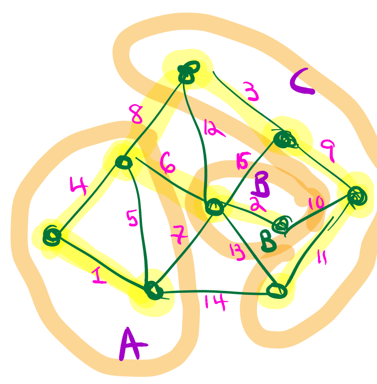
Find minimum spanning tree, i.e. T that minimizes $w(T) = \sum_{e \in T} w_e$



Kruskal: sort edges in ↑ weight. Consider one by one adding if endpoints are in different connected components



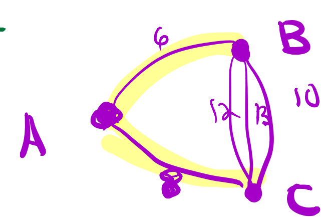
Prim: grow out from connected set, taking min



Boruvka: proceeds in rounds

in a round, take lightest edge from each vertex

contract connected components
(eliminating self loops & all but lightest edge between each pair of vertices



round $O(m)$

$\log n$ rounds

$O(m \log n)$

Fredman/Tarjan $O(m \log^* n)$

$\boxed{\text{Chazelle } O(m\,\alpha(n))}$ inverse Ackermann

$O(m)$

Today: linear time randomized alg [Karger, Klein, Tarjan.]

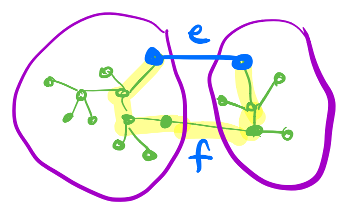
illustrates $\boxed{\text{random sampling}}$

Preliminaries

Cut Rule

$\forall$ cut $(S, \bar{S})$ in graph, the min weight edge crossing that cut must be in MST

suppose for some cut $(S, \bar{S})$
min wt edge e crossing cut is not in ST.
$T - f + e$ cheaper tree.

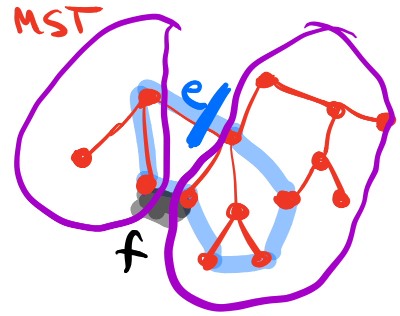


Cycle Rule

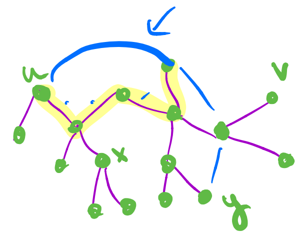
$\forall$ cycle in G, the heaviest edge on that cycle cannot be in MST

suppose $e \in T$

$T - e + f$ cheaper than T

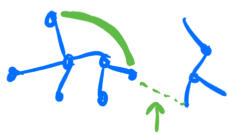


Verification



To check that T is a MST must check that $\forall\, e=(u,v)$ s.t. $e \notin T \Rightarrow w_e > \underbrace{w_T(u,v)}_{\text{weight of heaviest edge on } u \rightsquigarrow v \text{ path in } T}$ (✱)

[Amazingly, this can be done in $O(m)$ time. Returns $\xrightarrow{\text{yes}}$ T is MST; $\xrightarrow{\text{no}}$ $\{e \mid (✱) \text{ is violated}\}$]



Suppose that F is a forest in $G=(V,E)$

Verification alg can be adapted to find all edges e s.t.

$w_e \leq w_F(u,v)$ where

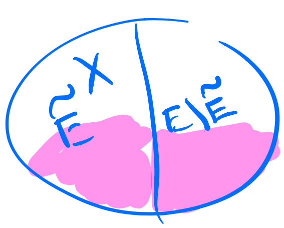
$$w_F(u,v) = \begin{cases} \text{weight of heaviest edge on } u \rightsquigarrow v \text{ path in } F & \text{if such a path exists} \\ \infty & \text{o.w.} \end{cases}$$

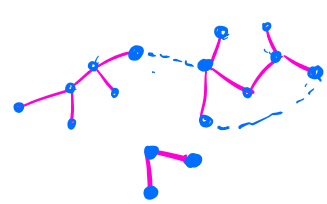
$O(m)$ time

Idea: $E \xrightarrow[\text{with prob } p]{\text{choose each edge } e \in E} \tilde{E}$ Subsample edges.

Find Minimum Spanning Forest F in $(V, \tilde{E})$

Hope that can remove from consideration most of $E \setminus \tilde{E}$





$\tilde{E} \longrightarrow F = \text{MSF}(V, \tilde{E})$ pink

$\boxed{E \setminus \tilde{E} \cup F}$

$\shortparallel$

F min spanning forest for $(V, \tilde{E})$

which are F-light ⇓ candidates for final MST

Definition: An edge $e=(u,v)$ is $\boxed{\text{F-light}}$ for a forest F if

$$w_e \le w_F(u,v)$$

$w_F(u,v)$ → weight of heaviest edge on $u \rightsquigarrow v$ path in F
[if $\nexists$ $u \rightsquigarrow v$ path in F, $w_F(u,v)=\infty$]

Note: all edges in F are F-light.

$$(V, \underbrace{\text{F-light edges}}_{\text{not too many}})$$

$\boxed{\text{Key Lemma}}$ Let F be MSF on $\tilde{G}=(V,\tilde{E})$ where $\tilde{E}$ obtained by sampling each edge independently with probability p.

$p=\frac{1}{2}$

Then $$E(\#\text{ F-light edges}) \le \frac{n-1}{p}$$

$2n$

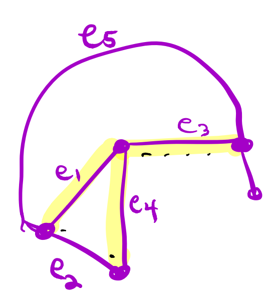
$\boxed{\text{Proof}}$ from the Book!

take sample of edges $\xrightarrow{\text{flip coins}}$ $\tilde{E}$
$H=(V,\tilde{E}) \longrightarrow$ find $F=MSF(H)$

min weight — sorted — max weight

$e_1\ e_2\ e_3\ e_4\ e_5 \cdots e_i \cdots\cdots e_m$

F-light?						
Coin Flip	H	T	H	H	T	H
$\tilde{E}$						
F						



construct $\tilde{E}$ & F
at same time, only flipping coins

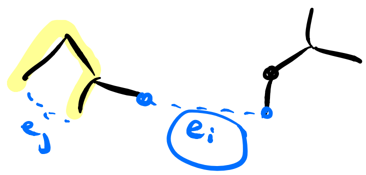
Thought experiment Run Kruskal's alg on <u>all</u> edges

Initialize $\tilde{E} := \emptyset$, $F := \emptyset$

min weight — sorted — max weight

	e_1	e_2	e_3	e_4	e_5 …	e_i	e_j …	e_m
F-light?	✓	✓	✓	✓	×	✓	✗	
Coin Flip	H	T	H	H	H	H	T	
$\tilde{E}$	✓		✓	✓	✓		✓	
F	✓		✓	✓				

$E \setminus \tilde{E}$ F-light



To process e_i:

① Test if both endpts are in same conn comp of current F.

yes: e F-heavy

no: e F-light

② flip a coin with prob p of coming up heads

coin heads ⟹ add e to $\tilde{E}$

if e is F-light, add e to F

F-light edges: edges in F and any edge determined F-light coin toss T

Claim correctly classifies edges as F-light / F-heavy. Correctly computes F (exactly what Kruskal would produce if you just gave it H.)

Partition sequence of F-light edges into phases

phase k starts immediately after $(k-1)^{st}$ edge has been added to F

exp length of each phase $= \frac{1}{p}$

n # vertices

m # edges

phases $\leq n-1$

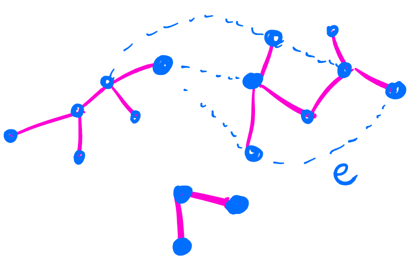
$$E(\#\text{ F-light edges}) \leq (n-1)\,\frac{1}{p}$$

$$\text{Exp } \#\text{ F-light edges} = \frac{\#\text{ phases}}{p} \leq \frac{n-1}{p}.$$

Algorithm (version 1)

1. Let $\tilde{E}$ be a random sample of E where each edge included with prob p. $O(m)$
2. Let $F := MSF(V, \tilde{E})$ $O(mp \log n)$
3. Find all $E'' \subseteq E$ that are F-light $O(m)$
4. Find $MST(V, E'')$ $O\left(\frac{n}{p} \log n\right)$

$n, \frac{n}{p}$



Overall expected running time

$$= O(mp \log n) + O\left(\frac{n}{p} \log n\right) + O(m)$$

$$mp = \frac{n}{p} \Rightarrow p = \sqrt{\frac{n}{m}}$$

$$T(n,m) = O\left(\sqrt{mn} \log n + m\right)$$

$m \geq n \log^2 n$

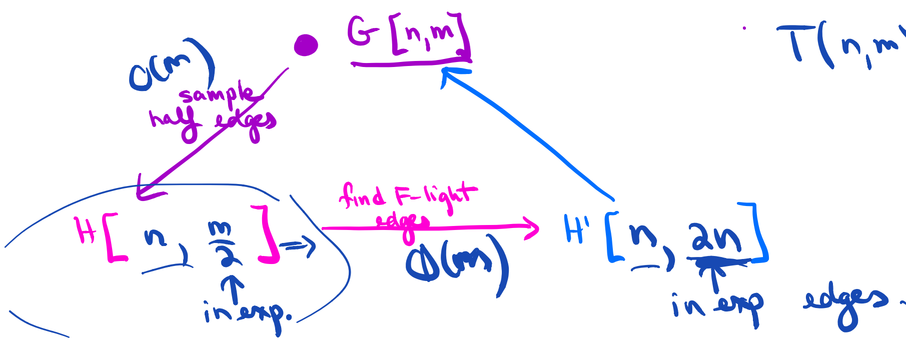
Recursive version:

① Let $\tilde{E}$ be a random sample where each edge included with prob ~~p~~ $\frac{1}{2}$

② Let $F := MSF(V, \tilde{E})$

③ Find all $E'' \subseteq E$ that are F-light

④ Run alg recursively to get $MST(V, E'')$



$T(n,m) \le c_1 n + c_2 m$

$T(n,m)$ expected runtime on graph with n vertices, m edges

$T(n,m) \le T(n, \frac{m}{2}) + T(n, 2n) + O(m)$ X

need to reduce n too

Final version

MSF(G)

1. Run 3 Boruvka steps $\Rightarrow H = (V', E')$

$(m, n) \to \boxed{(m, \frac{n}{8})}$ $O(m)$

$\boxed{\text{2 Boruvka steps}}$

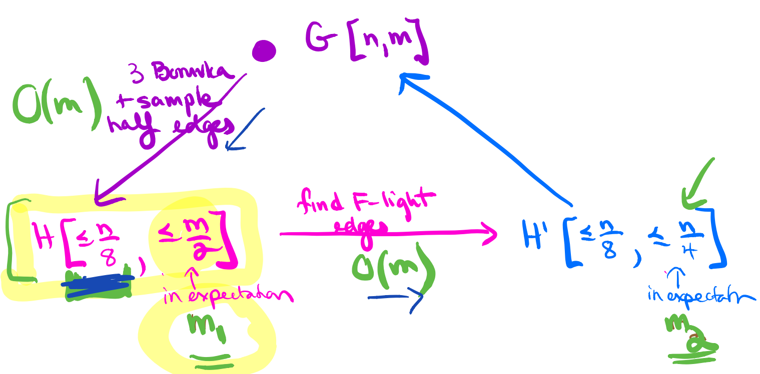
$\leq \frac{n}{8}$ vertices $\leq m$

2. Let $\tilde{E}$ be a random sample of E' where each edge included with prob $\frac{1}{2}$

3. Let $F := MSF(V', \tilde{E})$ recursively

4. Find all $E'' \subseteq E'$ that are F-light

Run alg recursively to get $MST(V', E'')$



$\boxed{\text{Theorem}}$ Exp run time $O(n+m)$

$T(n,m) \triangleq \max_{\substack{G=(V,E) \\ |V| \leq n,\ |E| \leq m}} E[T_G]$

T_G running time on graph G r.v.

$$T(n,m) \leq cm + E_{m_1}\left[T\left(\frac{n}{8}, m_1\right)\right] + E_{m_2}\left[T\left(\frac{n}{8}, m_2\right)\right]$$

Prove by induction on $n+m$ $\quad T(n,m) \leq c(n+2m)$

$$\leq cm + E\left[c\frac{n}{8} + 2cm_1\right] + E\left[c\frac{n}{8} + 2cm_2\right]$$

$E[m_1] \le \frac{n}{2}$

$E[m_2] \le \frac{n}{4}$

$\le c(n+2m)$

$E(\text{runtime}) = O(m)$

Can show runtime $O(m)$

w.p. $\geq 1 - e^{-\Omega(m)}$

<u>Chernoff.</u>