Krushkal's Algorithm

- From a Graph G ,we add one (the cheapest one) edge so that it joins two trees together. These step is repeated until and unless it cover all the vertex

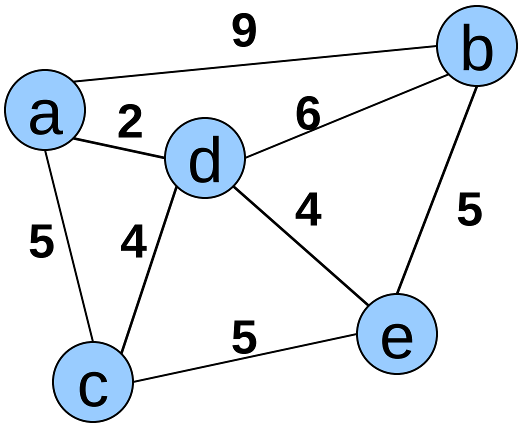
Kruskal's algorithm

Initialization

a. Create a set for each vertex $v \in V$

b. Initialize the set of "safe edges" A comprising the MST to the empty set

c. Sort edges by increasing weight



F = {a}, {b}, {c}, {d}, {e}

A = $\emptyset$

E = {(a,d), (c,d), (d,e), (a,c), (b,e), (c,e), (b,d), (a,b)}

Kruskal's algorithm

For each edge $(\boldsymbol{u,v}) \in \boldsymbol{E}$ **in increasing order while more than one set remains:**

If $\boldsymbol{u}$ and $\boldsymbol{v}$, belong to different sets $\boldsymbol{U}$ and $\boldsymbol{V}$

a. add edge $(\boldsymbol{u,v})$ to the safe edge set

$\boldsymbol{A} = \boldsymbol{A} \cup \{(\boldsymbol{u,v})\}$

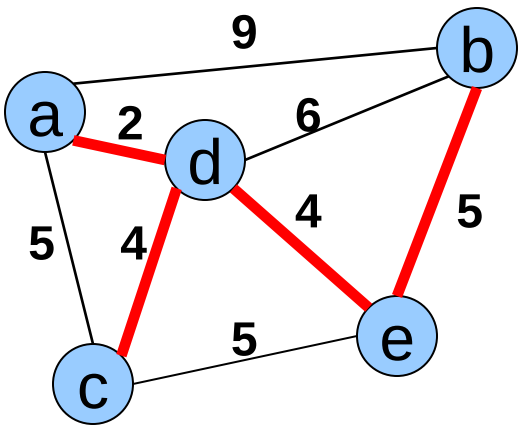
b. merge the sets $\boldsymbol{U}$ and $\boldsymbol{V}$

$\boldsymbol{F} = \boldsymbol{F} - \boldsymbol{U} - \boldsymbol{V} + (\boldsymbol{U} \cup \boldsymbol{V})$

Return $\boldsymbol{A}$

- Running time bounded by sorting (or findMin)
- O($|\boldsymbol{E}|\mathbf{log}|\boldsymbol{E}|$), or equivalently, O($|\boldsymbol{E}|\mathbf{log}|\boldsymbol{V}|$) (***why???***)

Kruskal's algorithm



E = {(a,d), (c,d), (d,e), (a,c),
(b,e), (c,e), (b,d), (a,b)}

Forest	*A*
{a}, {b}, {c}, {d}, {e}	∅
{a,d}, {b}, {c}, {e}	{(a,d)}
{a,d,c}, {b}, {e}	{(a,d), (c,d)}
{a,d,c,e}, {b}	{(a,d), (c,d), (d,e)}
{a,d,c,e,b}	{(a,d), (c,d), (d,e), (b,e)}

Kruskal's Algorithm Invariant

- After each iteration, every tree in the forest is a MST of the vertices it connects

- Algorithm terminates when all vertices are connected into one tree

Correctness of Kruskal's

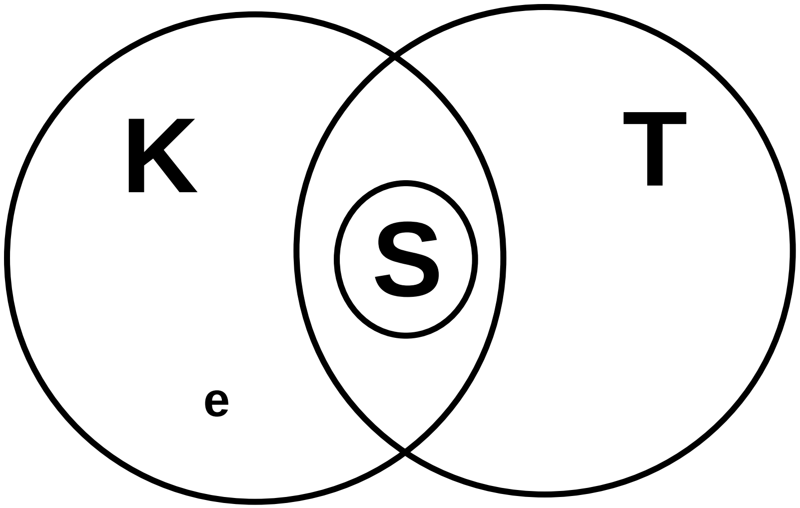
- This algorithm adds *n-1* edges without creating a cycle, so clearly it creates a spanning tree of any connected graph (*you should be able to prove this*).

But is this a *minimum* spanning tree?

Suppose it wasn't.

- There must be point at which it fails, and in particular there must a single edge whose insertion first prevented the spanning tree from being a minimum spanning tree.

Correctness of Kruskal's



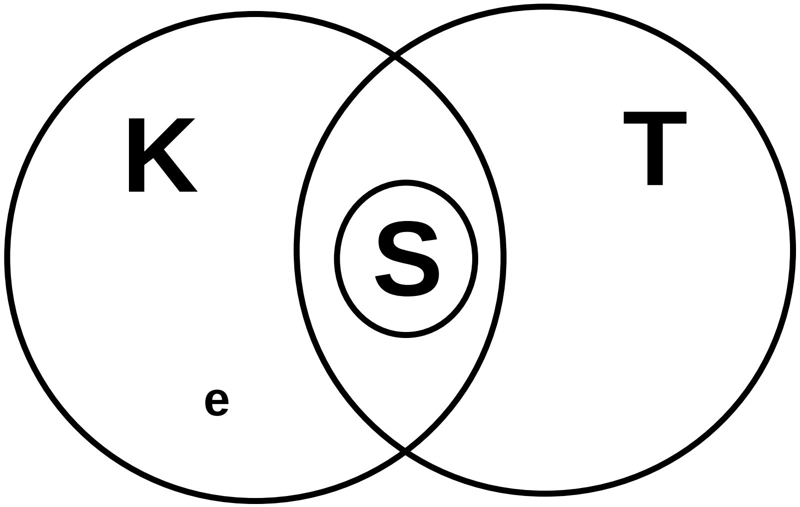
- Let **e** be this first errorful edge.
- Let **K** be the Kruskal spanning tree
- Let **S** be the set of edges chosen by Kruskal's algorithm *before* choosing **e**
- Let **T** be a MST containing all edges in **S**, but not **e**.

Correctness of Kruskal's

Lemma: w(**e'**) >= w(**e**) for all edges **e'** in **T - S**

Proof (*by contradiction*):

- Assume there exists some edge **e'** in **T - S**, w(**e'**) < w(**e**)
- Kruskal's must have considered **e'** before **e**



- However, since **e'** is not in **K** (*why??*), it must have been discarded because it caused a cycle with some of the other edges in **S**.
- But **e' + S** is a subgraph of **T**, which means it cannot form a cycle **...Contradiction**