Steiner Forest Problem

Approximation Algorithms

Aristotelis – Emmanouil Thanos – Filis
NTUA

The problem

Given an undirected graph G = (V, E), an edge cost function c : E $\rightarrow Q^{+}$, and a collection of disjoint subsets of V, $S_1, S_2, \ldots, S_k$, find a minimum cost subgraph in which each pair of vertices belonging to the same set S_i is connected

Alternatively we can define a connectivity requirement function r that maps unordered pairs of vertices to {0,1} as follows:

$$r(u,v) = \begin{cases} 1 \text{ if } u \text{ and } v \text{ belong to the same set } S_i \\ 0 \quad \text{otherwise} \end{cases}$$

The problem now is to find a minimum cost subgraph that contains a u-v path for each pair (u,v) with r(u,v)=1

Relationship to Steiner Tree

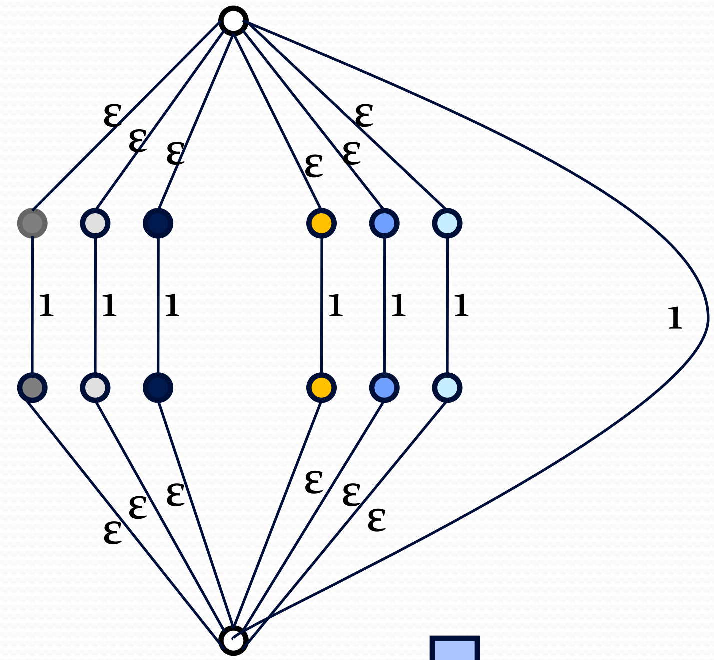
Minimum Steiner tree is a special case in which k = 1 and S_1 is an arbitrary subset of V.

Since Steiner Tree is NP-hard, Steiner forest is also NP-Hard.

There is a very simple 2-approximation algorithm for Steiner Tree.

G. Robins, A. Zelikovsky (2005): The best known approximation factor is: $1+\frac{\ln 3}{2}\approx 1,55$

Why can’t we use a simple algorithm for Steiner Tree problem to solve the Steiner Forest problem? We would just have to merge the Steiner trees for every set S_i ...



This example shows that the above technique could lead us to a k-approximation algorithm

LP

- Function $f : 2^V \rightarrow \{0,1\}$ specifies the minimum number of edges that must cross each cut in any feasible solution:

$$f(S) = \begin{cases} 1 & \text{if } \exists u \in S \text{ and } v \in \bar{S} \text{ such that } r(u,v) = 1 \\ 0 & \text{otherwise} \end{cases}$$

- $\delta(S)$ denotes the set of edges crossing the cut $(S, \bar{S})$
- x_e will be set to **1** iff e is picked, else will be set to **0**

Then the problem is:

$$\begin{aligned} \text{minimize} \quad & \sum_{e \in E} c_e x_e \\ \text{subject to} \quad & \sum_{e:\ e \in \delta(S)} x_e \geq f(S), \quad S \subseteq V \\ & x_e \in \{0,1\} \qquad e \in E \end{aligned}$$

LP relaxation

$$\text{minimize} \quad \sum_{e \in E} c_e x_e$$

$$\text{subject to} \quad \sum_{e:\ e \in \delta(S)} x_e \geq f(S), \qquad S \subseteq V$$

$$x_e \geq 0, \qquad e \in E$$

Integrality gap

Consider a cycle on n vertices and with each edge of cost 1. We require all the vertices to be connected to each other. The minimum Steiner forest has cost n-1 as we choose n-1 edges arbitrarily. However, the LP can be solved by setting $x_e = 0,5$ for all e (it satisfies the constraints of the LP) , giving a value of $\frac{n}{2}$

This leads to the result that the integrality gap **is greater than**

$$\frac{n-1}{\frac{n}{2}} = \frac{2n-2}{n} = 2 - \frac{2}{n}$$

Dual Program

$$\text{maximize} \quad \sum_{e \in E} f(S)\, y_S$$

$$\text{subject to} \quad \sum_{S:\ e \in \delta(S)} y_S \le c_e \,, \qquad e \in E$$

$$y_S \ge 0, \qquad S \subseteq V$$

Algorithm (Steiner Forest)

- (**Initialization**) $F \leftarrow \emptyset$; for each $S \subseteq V$, $y_s \leftarrow 0$

- (**Edge augmentation**) while there exists an unsatisfied set do:
  simultaneously raise y_s for each active set S, until some edge *e* goes tight;
  $F \leftarrow F \cup \{ \text{e} \}$

- (**Pruning**) return $F' = \{e \in F \mid F - \{e\}$ is primal infeasible}

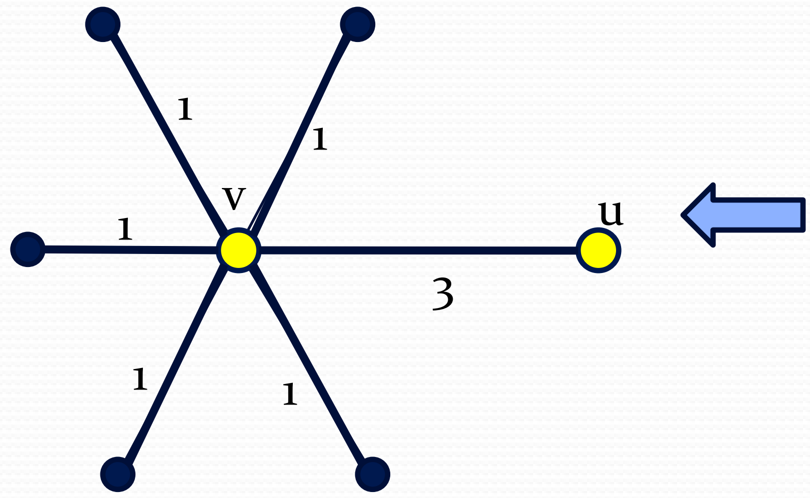
Definitions

•Edge e **feels dual** y_s if $y_s > 0$ and $e \in \delta(S)$

•Edge e is **tight** if the total amount of dual it feels equal its cost

•Set S is **unsatisfied** if $f(S)=1$ but there is no picked edge crossing the cut $(S,\overline{S})$

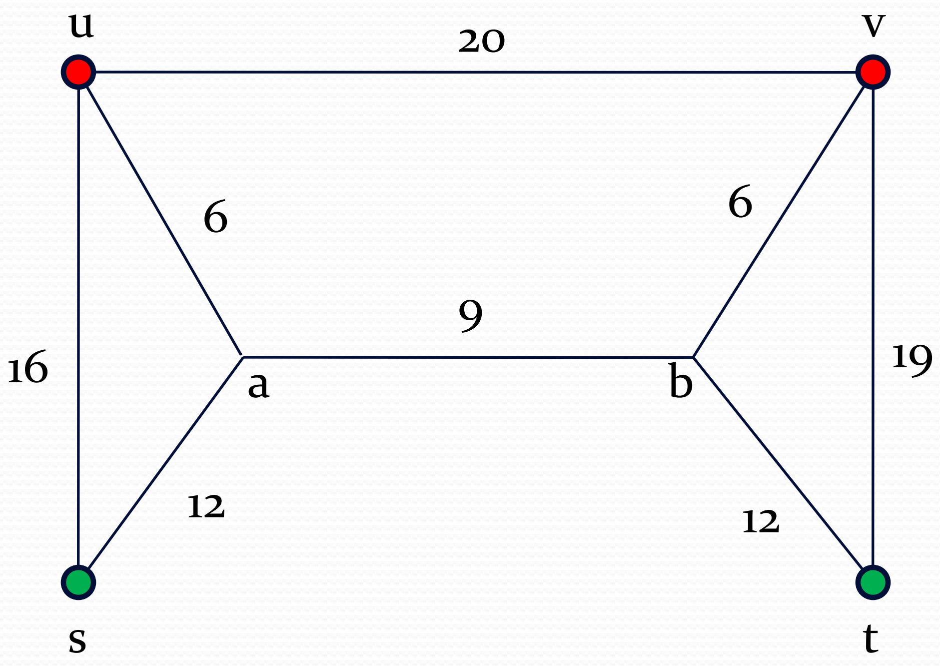
•Set S is **active** if it is a minimal unsatisfied set in the current iteration

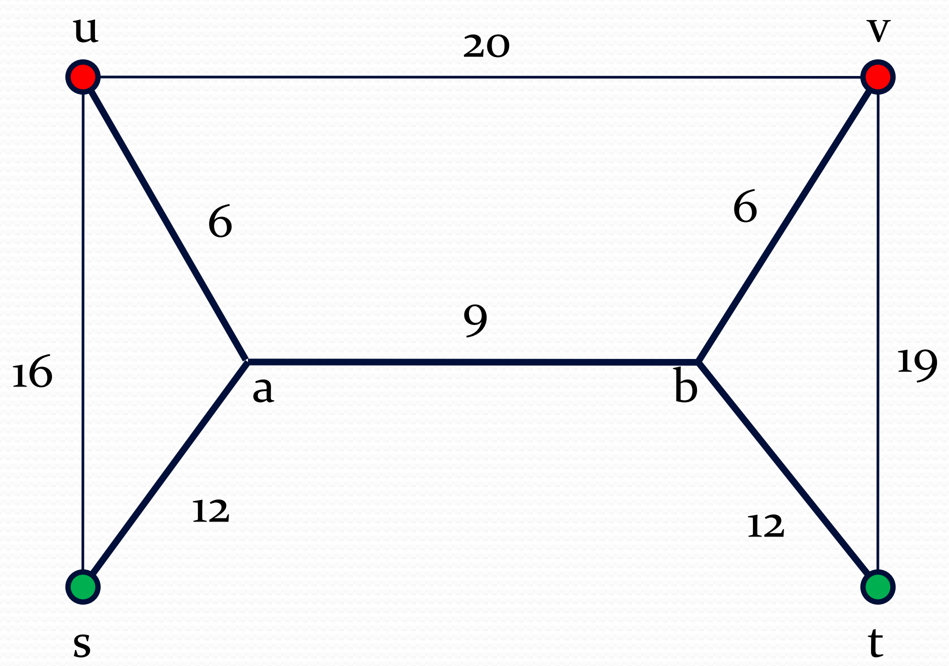


Is the pruning step is necessary?
YES!

A given graph

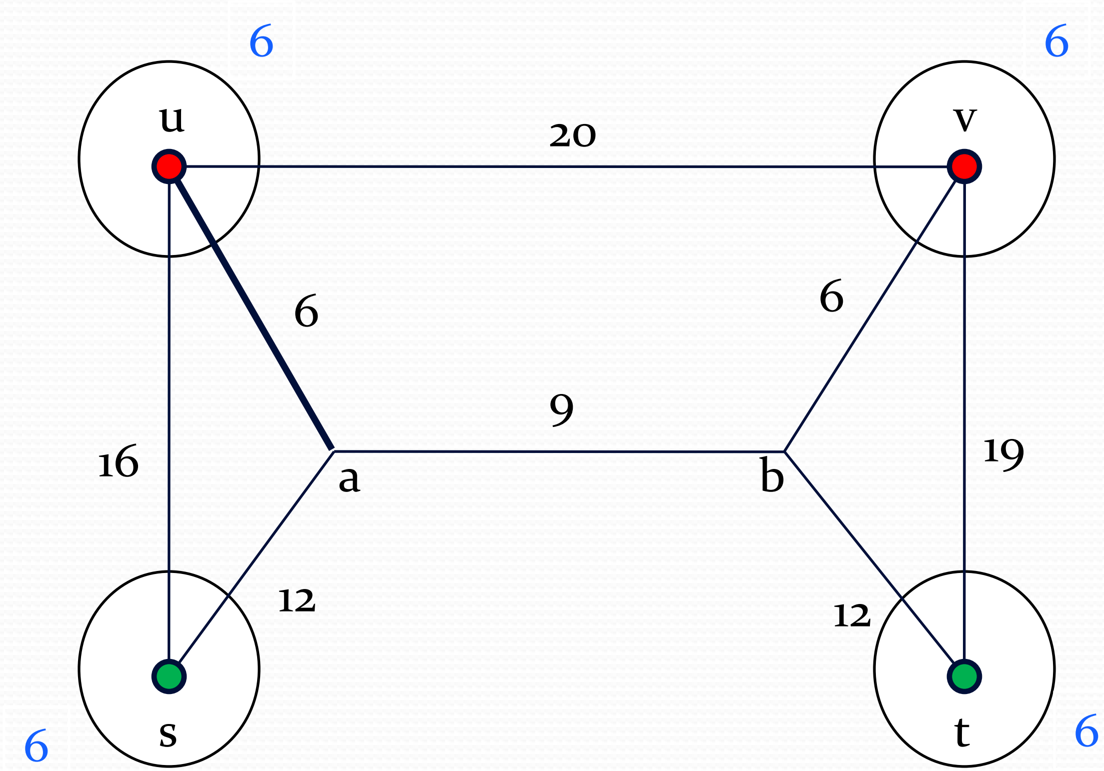
In this graph the connectivity requirements are r(u,v)=1 and r(s,t)=1





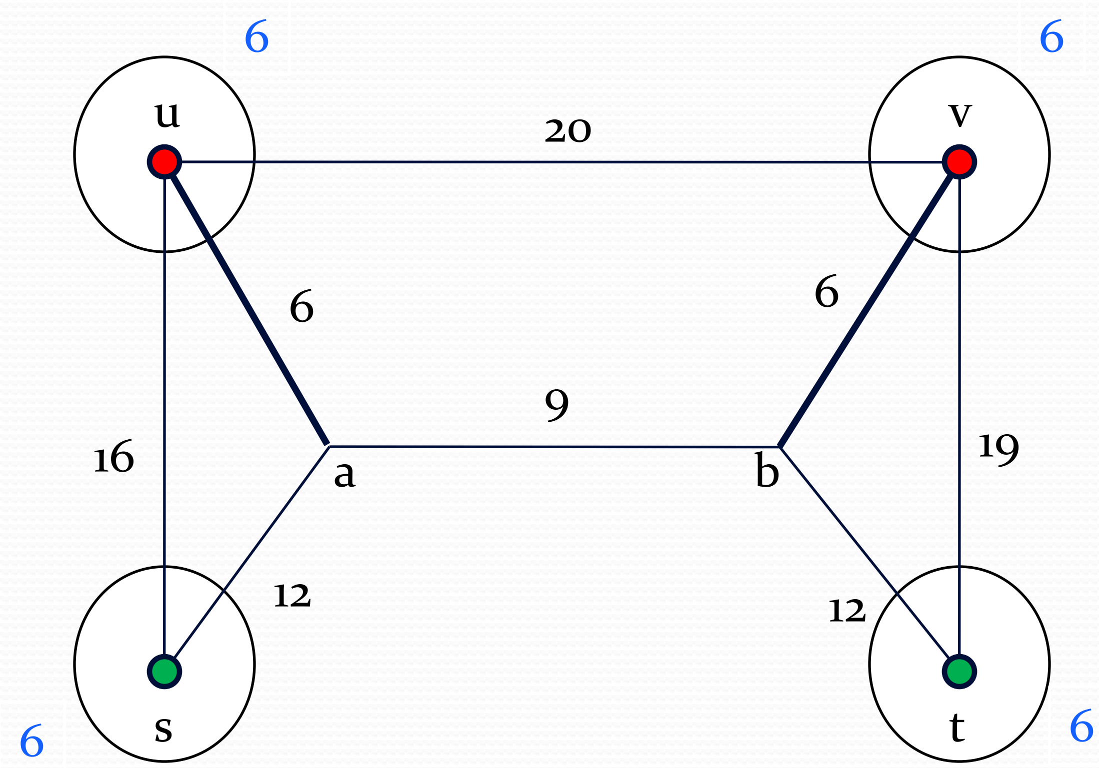
The optimal solution to the above graph of cost 45

The algorithm: First iteration



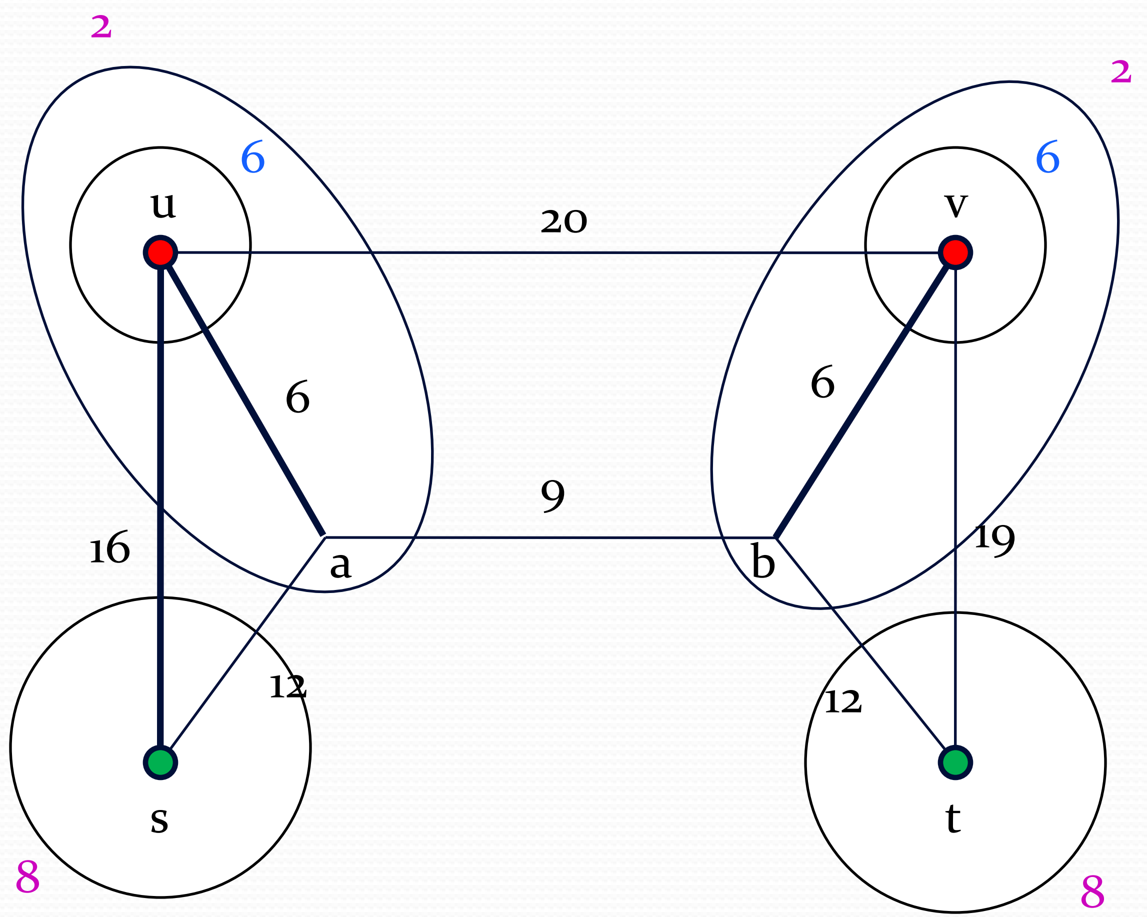
Active sets
- {u}
- {v}
- {s}
- {t}

The algorithm: Second iteration



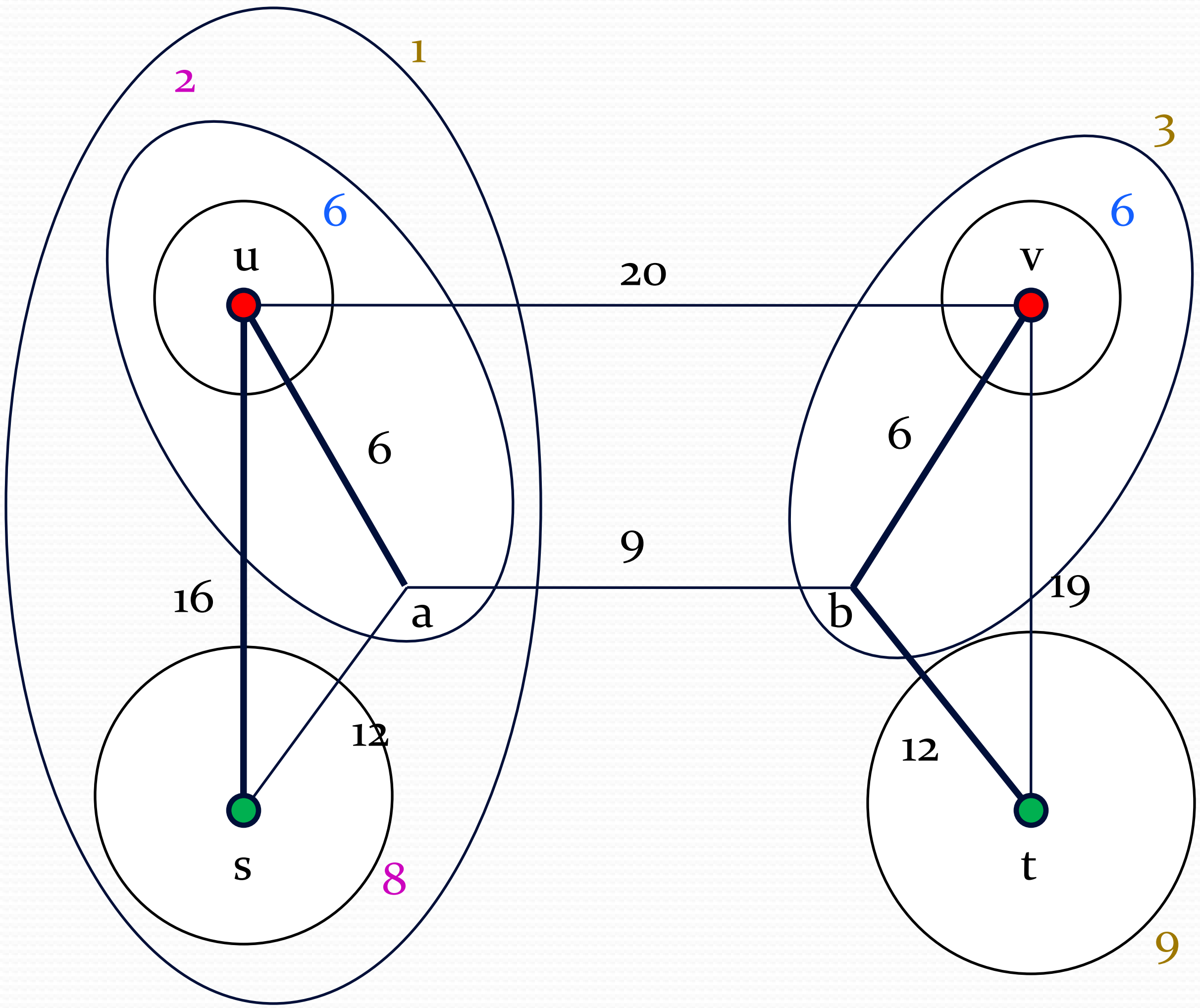
Active sets
- {u, a}
- {v}
- {s}
- {t}

The algorithm: Third iteration



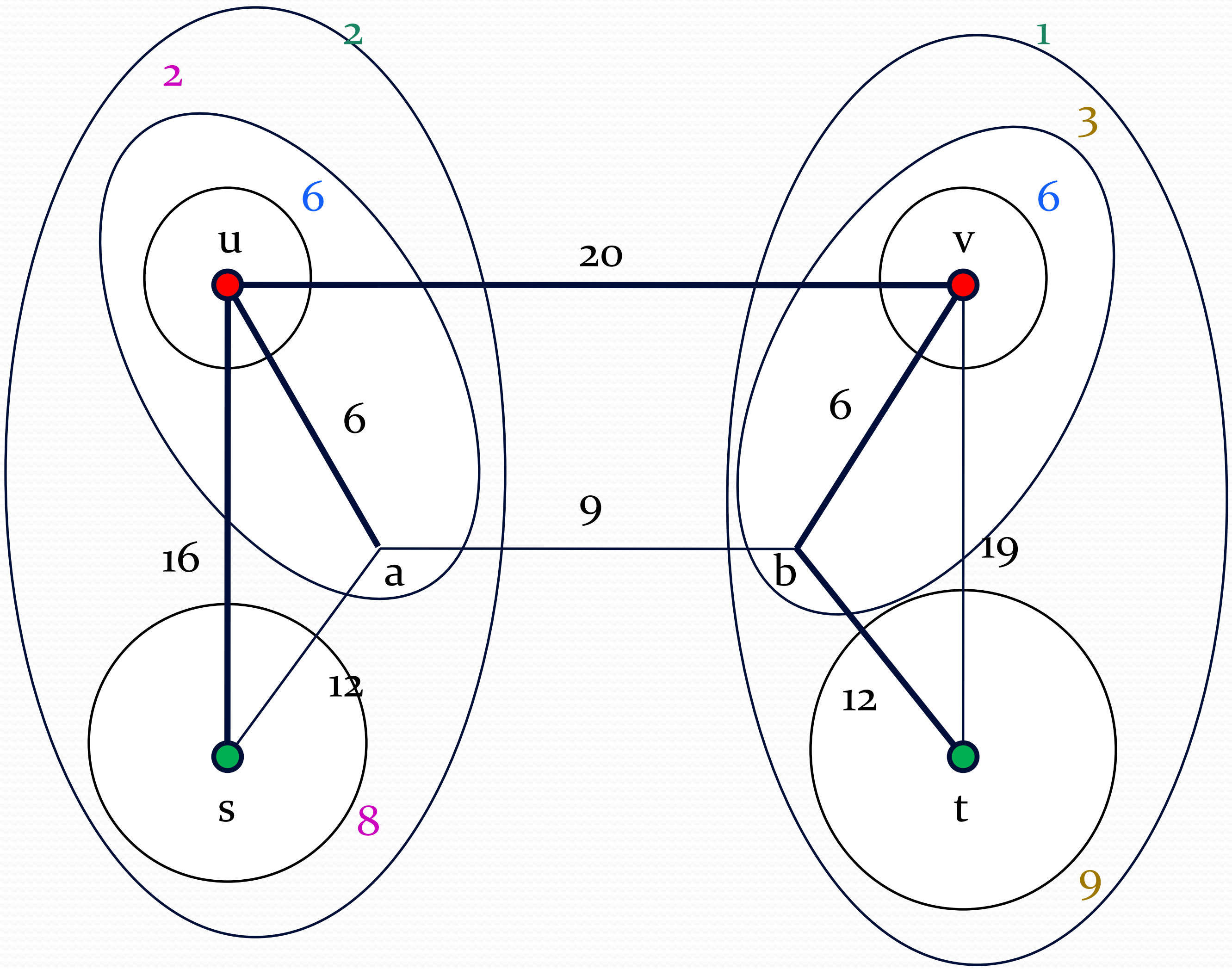
Active sets
- {u, a}
- {v, b}
- {s}
- {t}

The algorithm: Fourth iteration



Active sets
- {u, a, s}
- {v, b}
- {t}

The algorithm: Fifth iteration

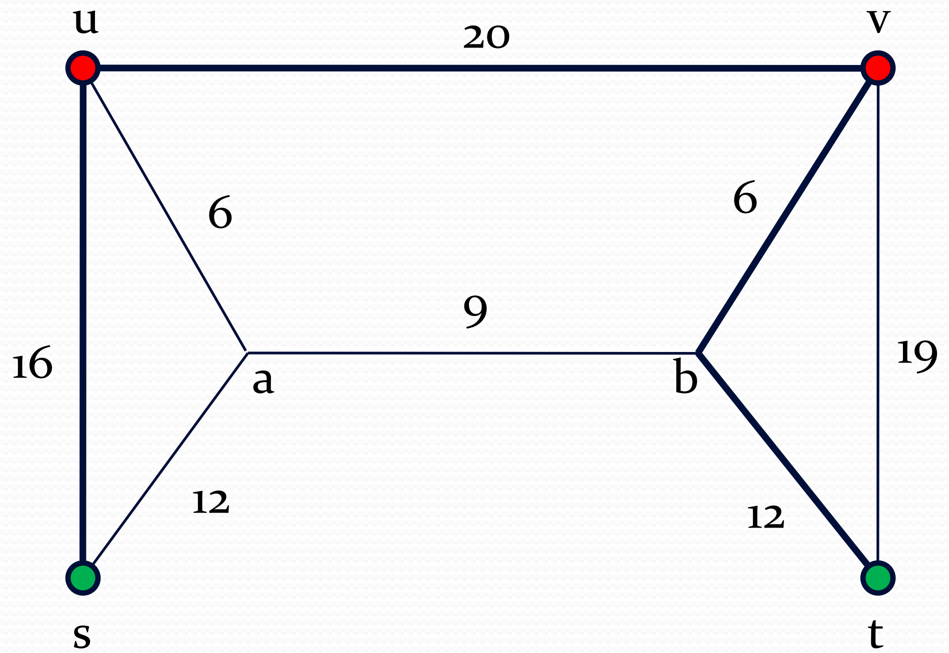


Active sets
- {u, a, s}
- {v, b, t}

The algorithm: Pruning step

Finally we get a solution of cost 54 while the cost of the optional solution was 45

Active sets
There aren't any



Analysis(1)

Lemmata

(i) At the end of the algorithm, F' and y are primal and dual feasible solutions, respectively

(ii) $$\sum_{e \in F'} c_e \leq 2 \sum_{S \subseteq V} y_s$$

Those two lemmata give us the proof, that the algorithm we described, achieves an approximation guarantee of factor 2 for the Steiner Forest problem

Analysis(2)

Proof

(i) By design, F is acyclic because no edge running within the same component, can go tight. Moreover, at the end of the algorithm if r(u,v)=1, there is a unique u-v path in F. Thus, each edge on this path is not redundant and it is not deleted on the pruning step. Hence F' is primal feasible

When an edge becomes tight, the active sets are redifined. As a result, the edge that had just been tight, is a part of the connected component and it can't be overtightened. Hence, y is dual feasible

Analysis(3)

Proof

(ii) **Notation:** $\deg_{F'}(S)$ denotes the number of edges of F’ crossing the cut $(S, \bar{S})$

Since every picked edge is tight:

$$\sum_{e \in F'} c_e = \sum_{e \in F'} \left(\sum_{S:\, e \in \delta(S)} y_s \right)$$

Changing the order of summation we get:

$$\sum_{e \in F'} c_e = \sum_{S \subseteq V} \left(\sum_{e \in \delta(S) \cap F'} y_S \right) = \sum_{S \subseteq V} \deg_{F'}(S)\, y_S$$

Analysis(4)

Thus, we need to show that

$$\sum_{S \subseteq V} \deg_{F'}(S)\, y_S \leq 2 \sum_{S \subseteq V} y_S$$

Let Δ be the extent to which active sets were raised in the last iteration. Then we need to show:

$$\Delta \times \left(\sum_{S \ active} \deg_{F'}(S) \right) \leq 2\Delta \times (\#\ \text{of active sets}) \quad \Rightarrow$$

$$\frac{\sum_{S\ \text{active}} \deg_{F'}(S)}{\#\ \text{of active sets}} \leq 2$$

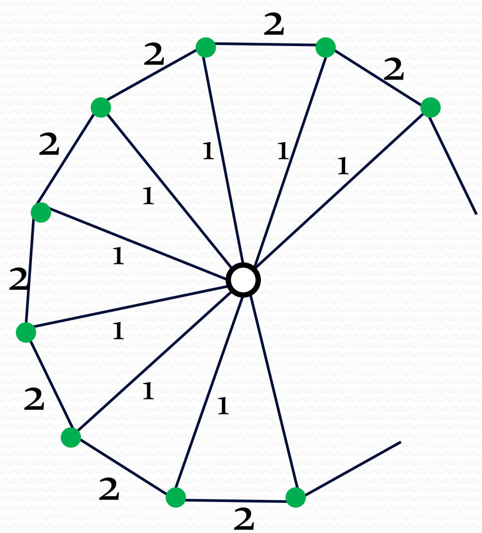
Analysis(5)

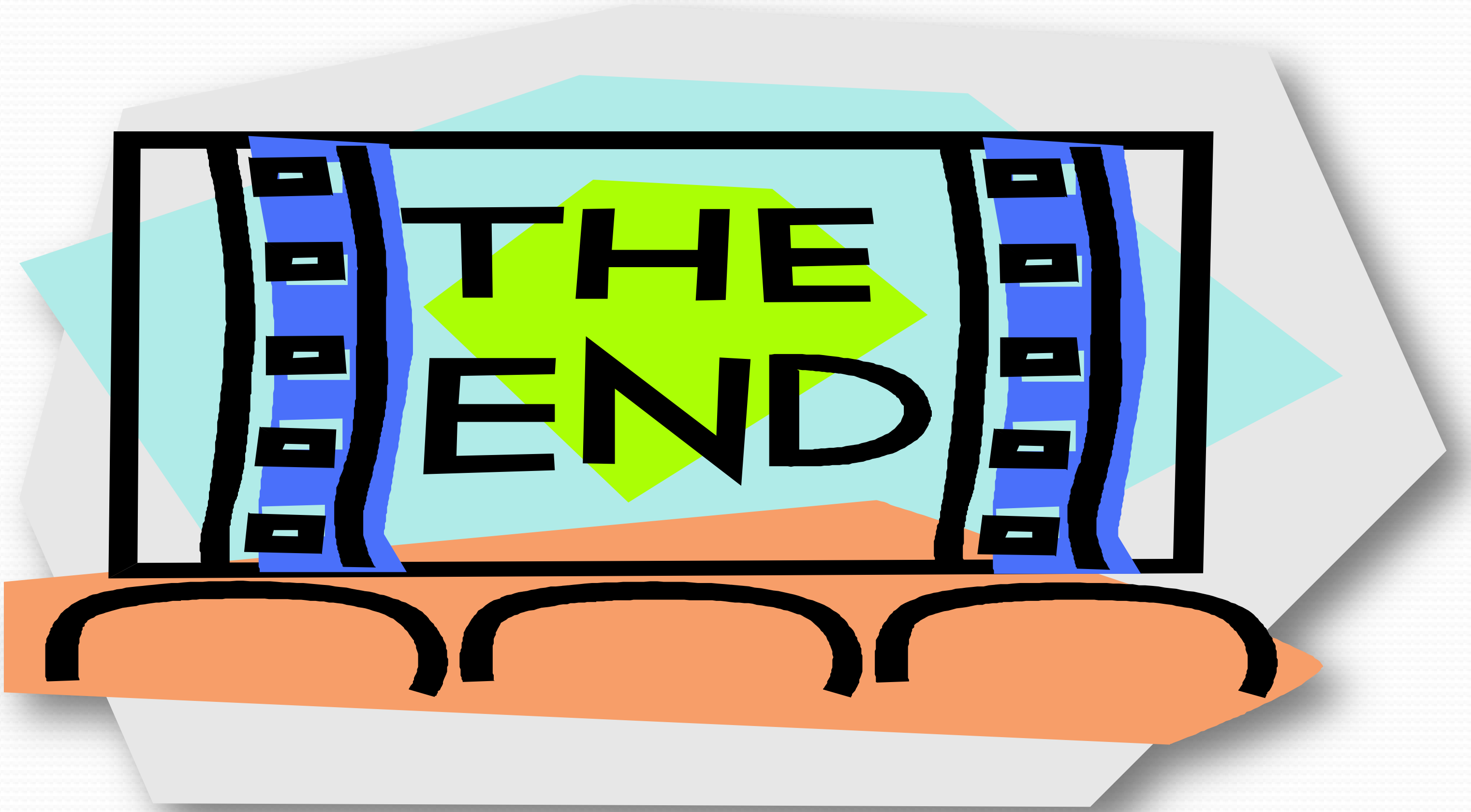
So we need to show that in this iteration , the average degree of active sets is at most 2

Imagine F (the final forest) with each current connected component collapsed into a single node. In this revised F, none of the inactive connected components will be leaves, because we would have removed edges connecting these components to the rest of the forest during the pruning step.

Therefore, the average degree of all active connected components is the average degree of a subset of the nodes in a tree, including all of the leaves. So the average degree is at most 2.

Tight example




THE
END