Advanced topics on Algorithms

Luciano Gualà
www.mat.uniroma2.it/~guala/

Approximation algorithms: Episode IV

(the final one)

Primal-dual schema

High-level idea of the approach

Algorithm

start with:
- an infeasible integral primal solution, and
- a dual feasible solution

Iteratively:
- improve the dual solution
- improve the feasibility of the integral primal solution

Until a feasible integral primal solution is obtained

analysis: prove the approximation guarantee using the value of the dual solution as a lower bound

minimum Set Cover problem

Input:

- universe U of n elements
- a collection of subsets of U, $\mathcal{S}=\{S_1,\ldots,S_k\}$
- each $S\in\mathcal{S}$ has a positive cost c(S)

Feasible solution:

a subcollection $\mathcal{C}\subseteq\mathcal{S}$ that covers U (whose union is U)

measure (min):

cost of $\mathcal{C}$: $\sum_{S\in\mathcal{C}} c(S)$

frequency of an element e: number of sets e belongs to

f: frequency of the most frequent element

ILP:

$$\text{minimize} \quad \sum_{S \in \mathcal{S}} c(S) x_S$$

$$\text{subject to} \quad \sum_{S: e \in S} x_S \geq 1 \qquad e \in U$$

$$x_S \in \{0,1\} \qquad S \in \mathcal{S}$$

LP-relaxation

$$\text{minimize} \quad \sum_{S \in \mathcal{S}} c(S) x_S$$

$$\text{subject to} \quad \sum_{S: e \in S} x_S \geq 1 \qquad e \in U$$

$$x_S \geq 0 \qquad S \in \mathcal{S}$$

dual program

$$\text{maximize} \quad \sum_{e \in U} y_e$$

$$\text{subject to} \quad \sum_{e: e \in S} y_e \leq c(S) \qquad S \in \mathcal{S}$$

$$y_e \geq 0 \qquad e \in U$$

Given a dual solution y, we say that a set S is tight if $\sum_{e \in S} y_e = c(S)$

idea: pick only tight sets & do not overpack any set.

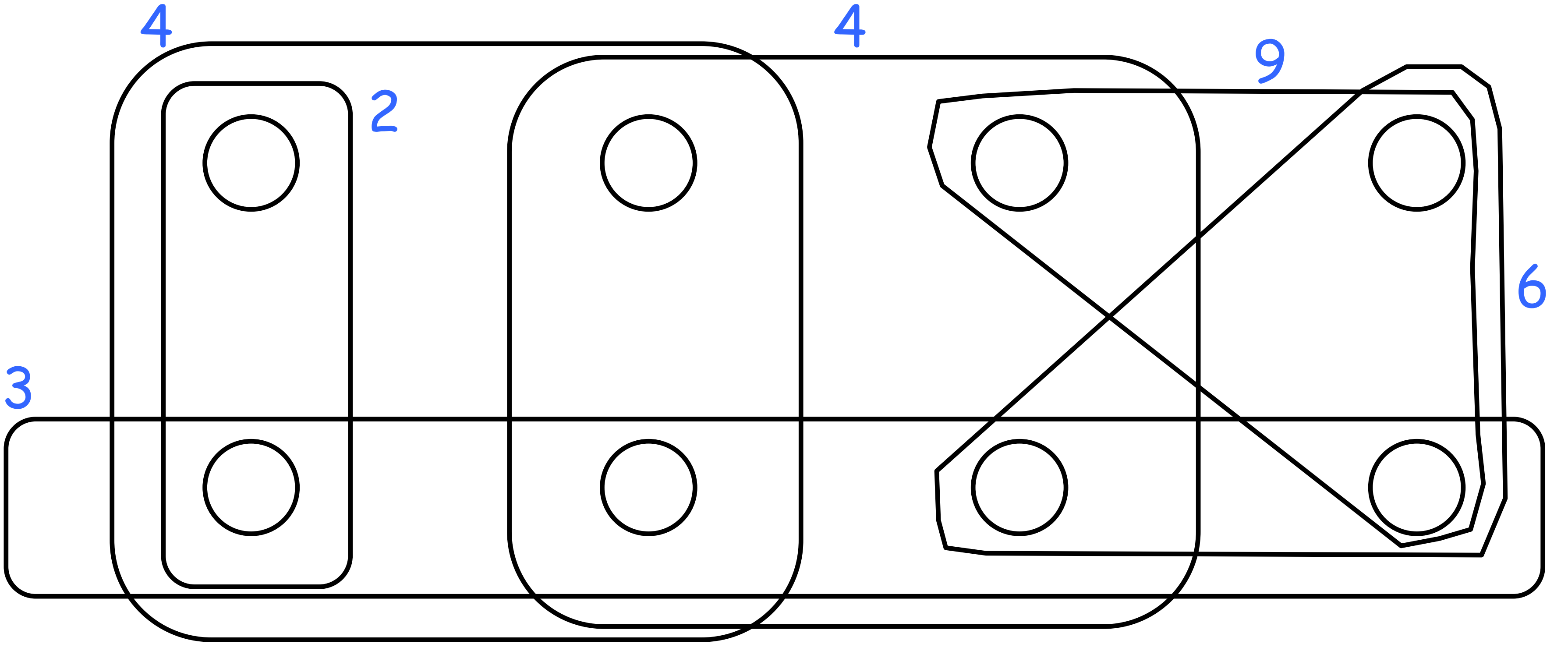
Algorithm 15.2 (Set cover – factor f)

1. **Initialization:** $\boldsymbol{x} \leftarrow \mathbf{0}$; $\boldsymbol{y} \leftarrow \mathbf{0}$
2. Until all elements are covered, do:

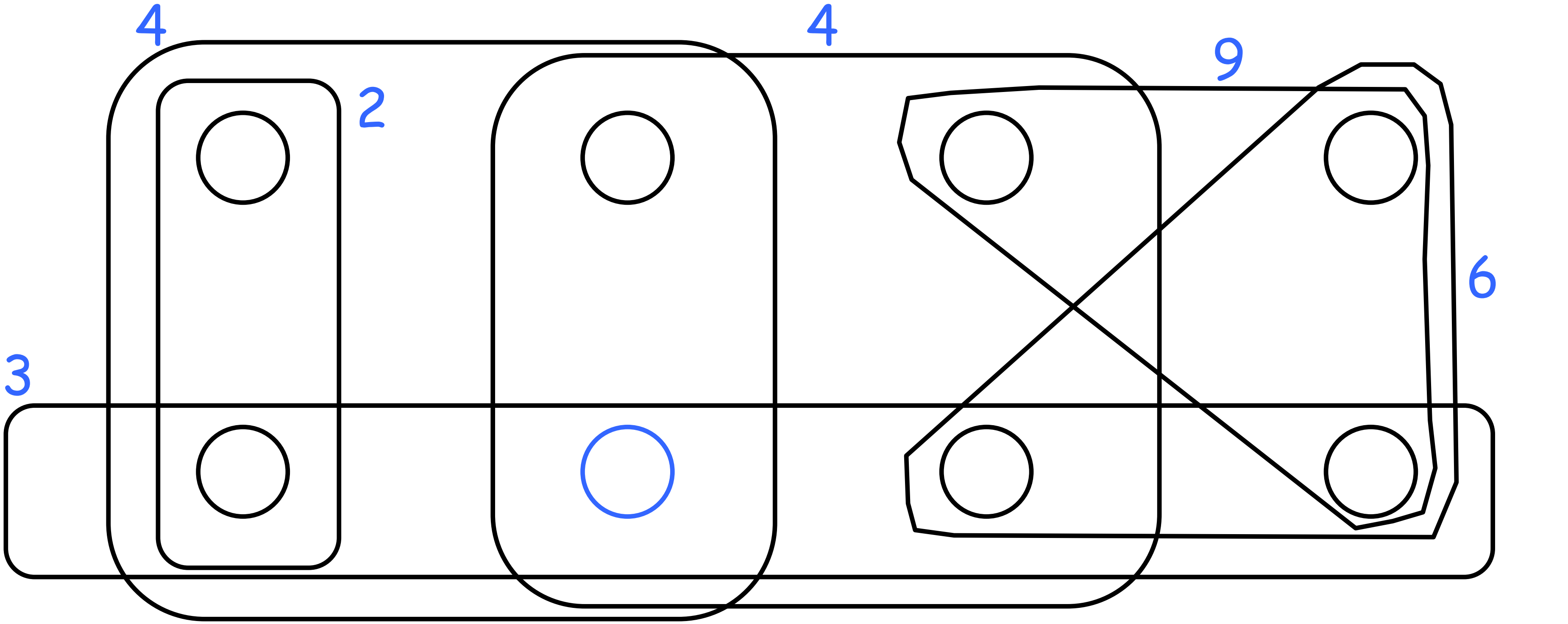
   Pick an uncovered element, say e, and raise y_e until some set goes tight.

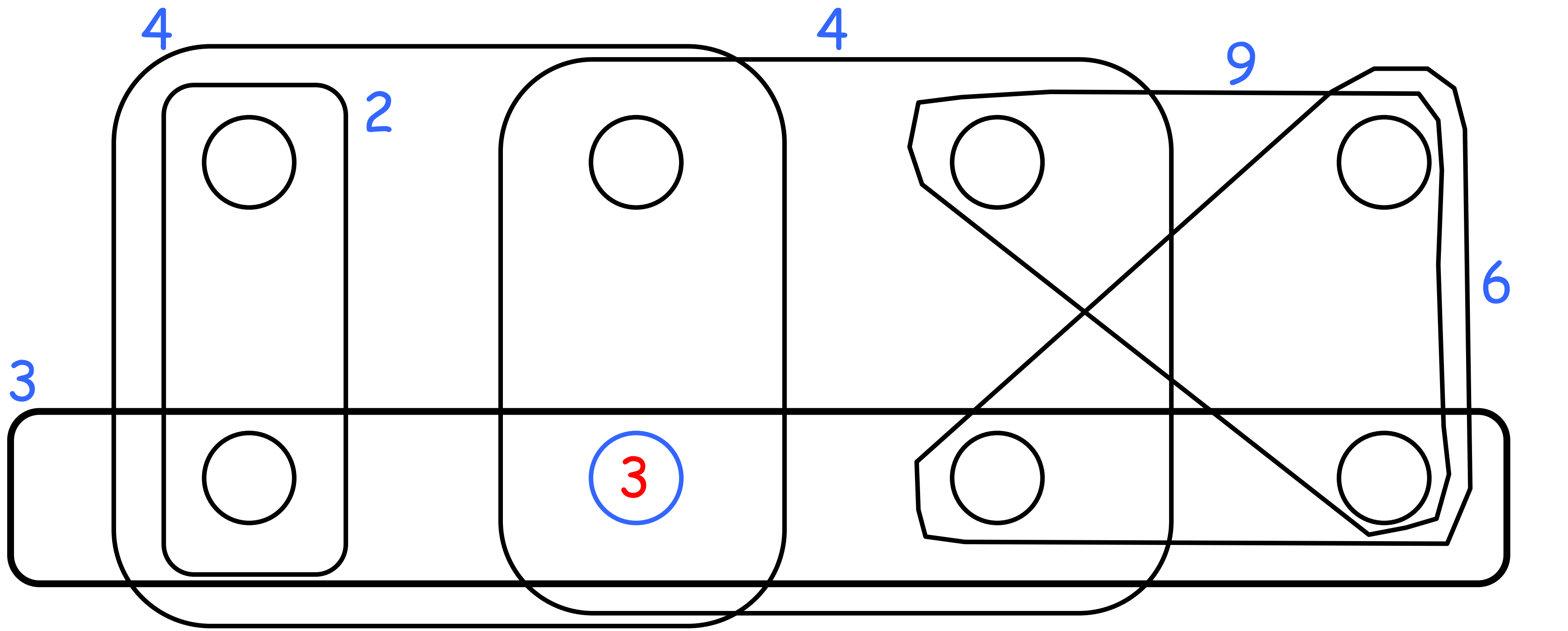
   Pick all tight sets in the cover and update $\boldsymbol{x}$.

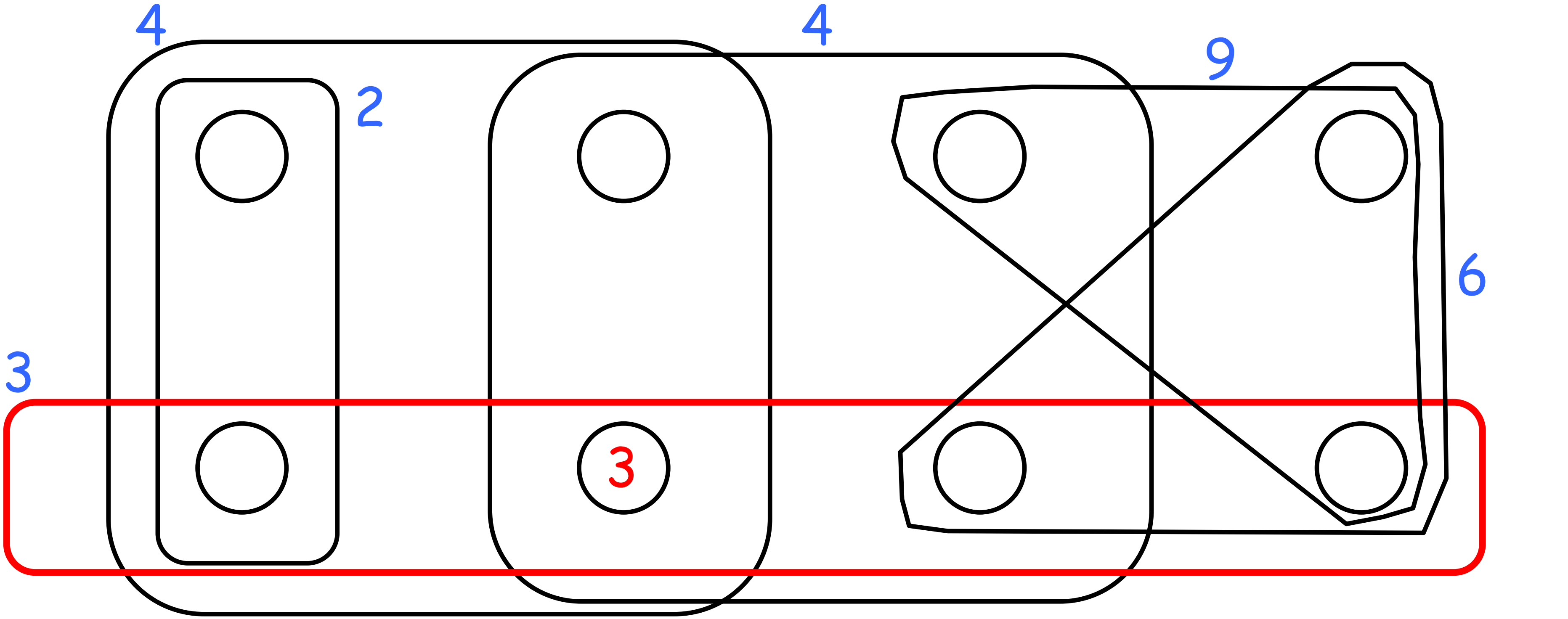
   Declare all the elements occurring in these sets as "covered".
3. Output the set cover $\boldsymbol{x}$.

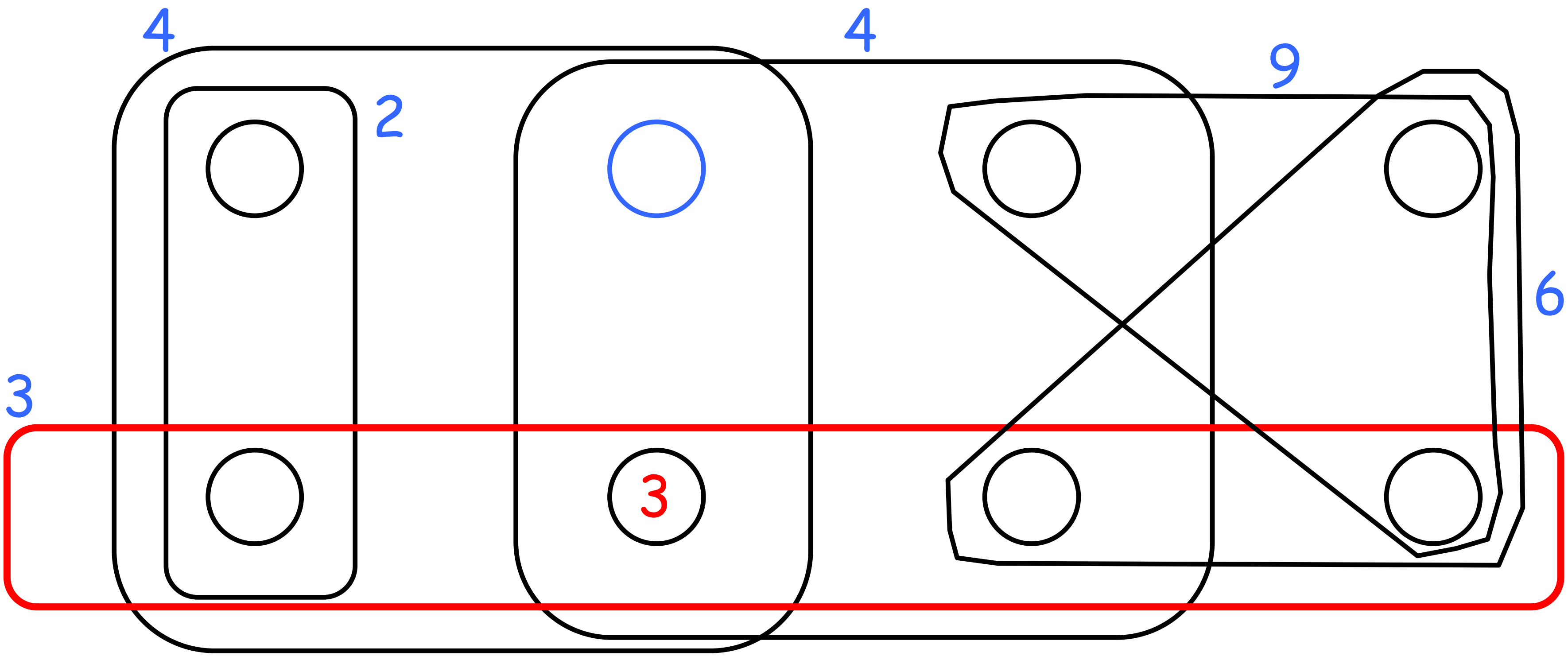


4
4
9
2
6
3

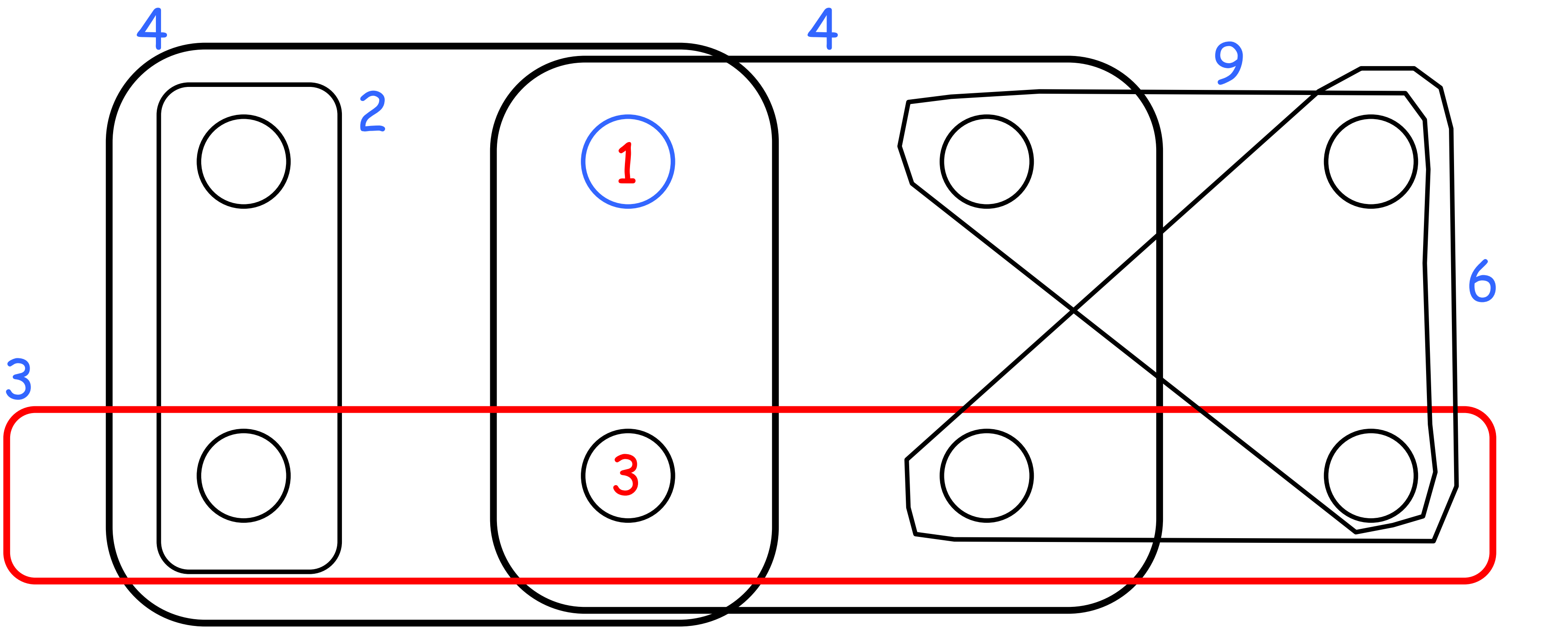

4
4
9
2
6
3

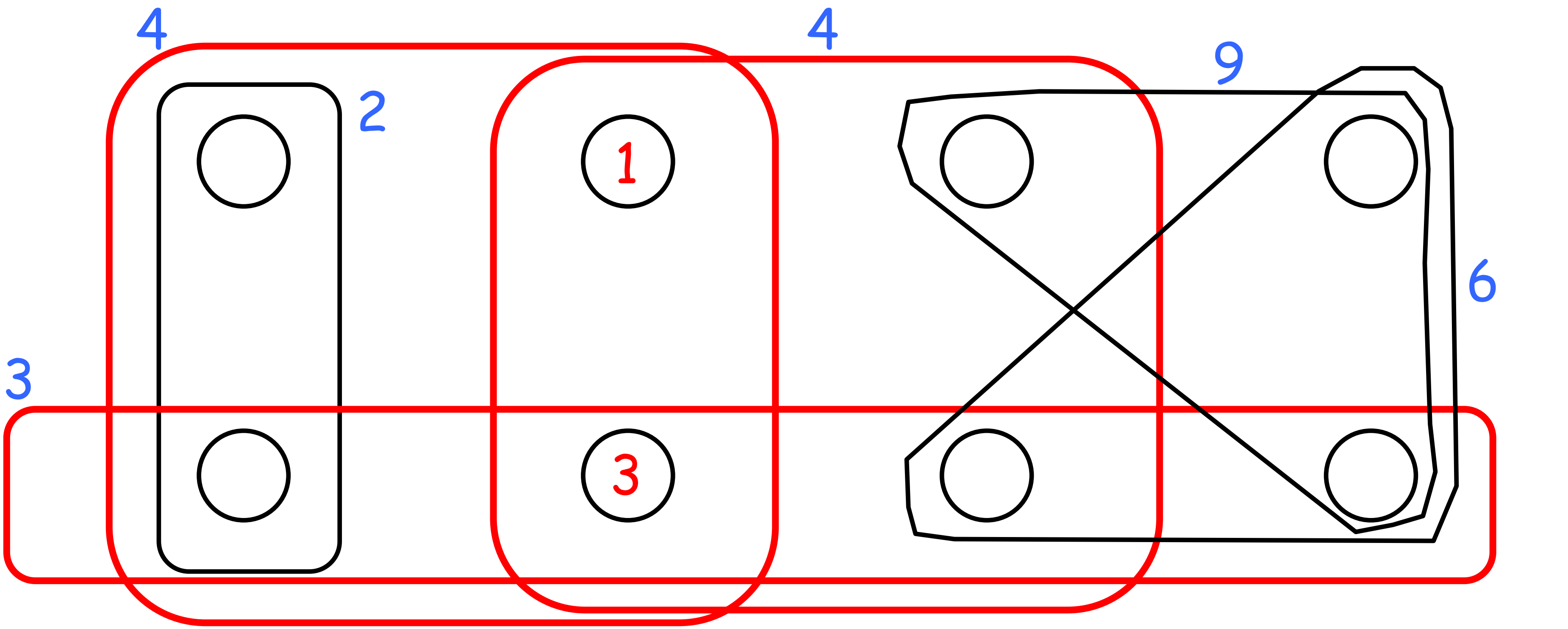

4
4
9
2
6
3
3

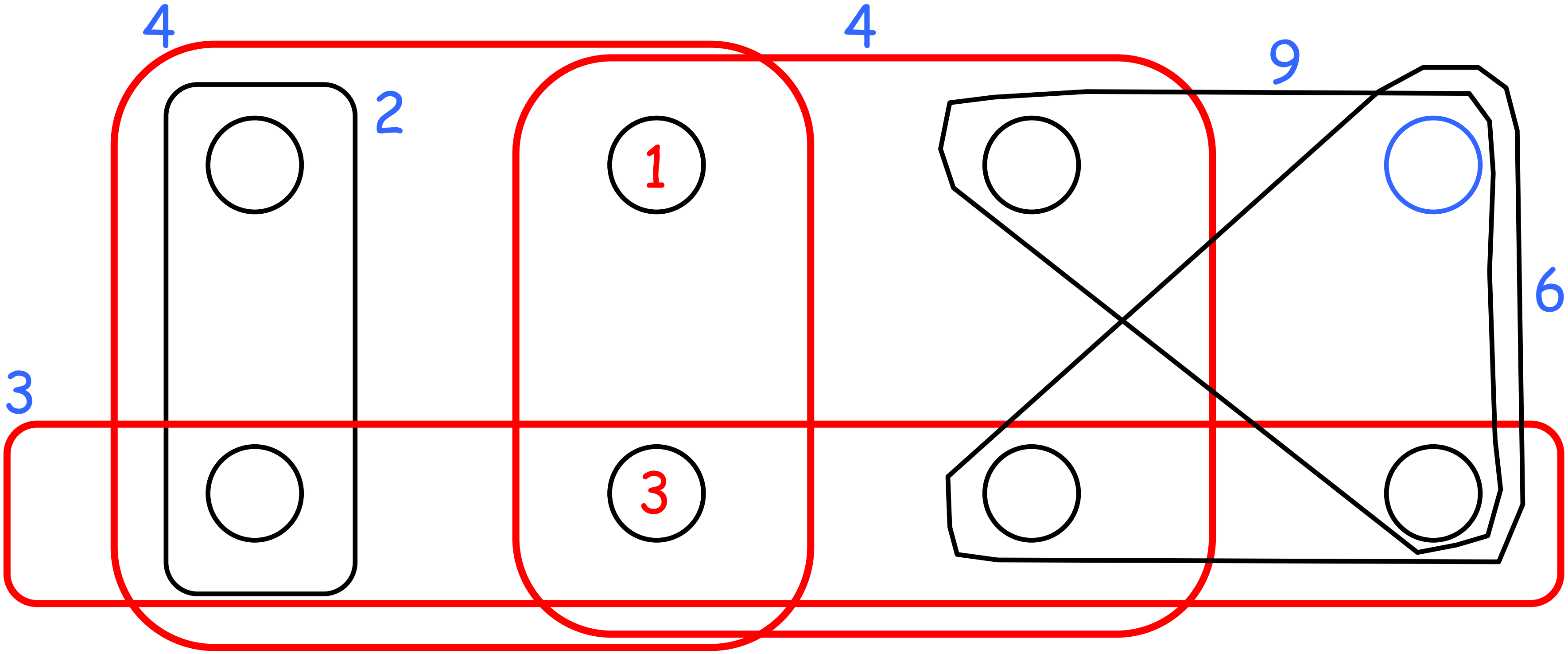

4
4
9
2
6
3
3

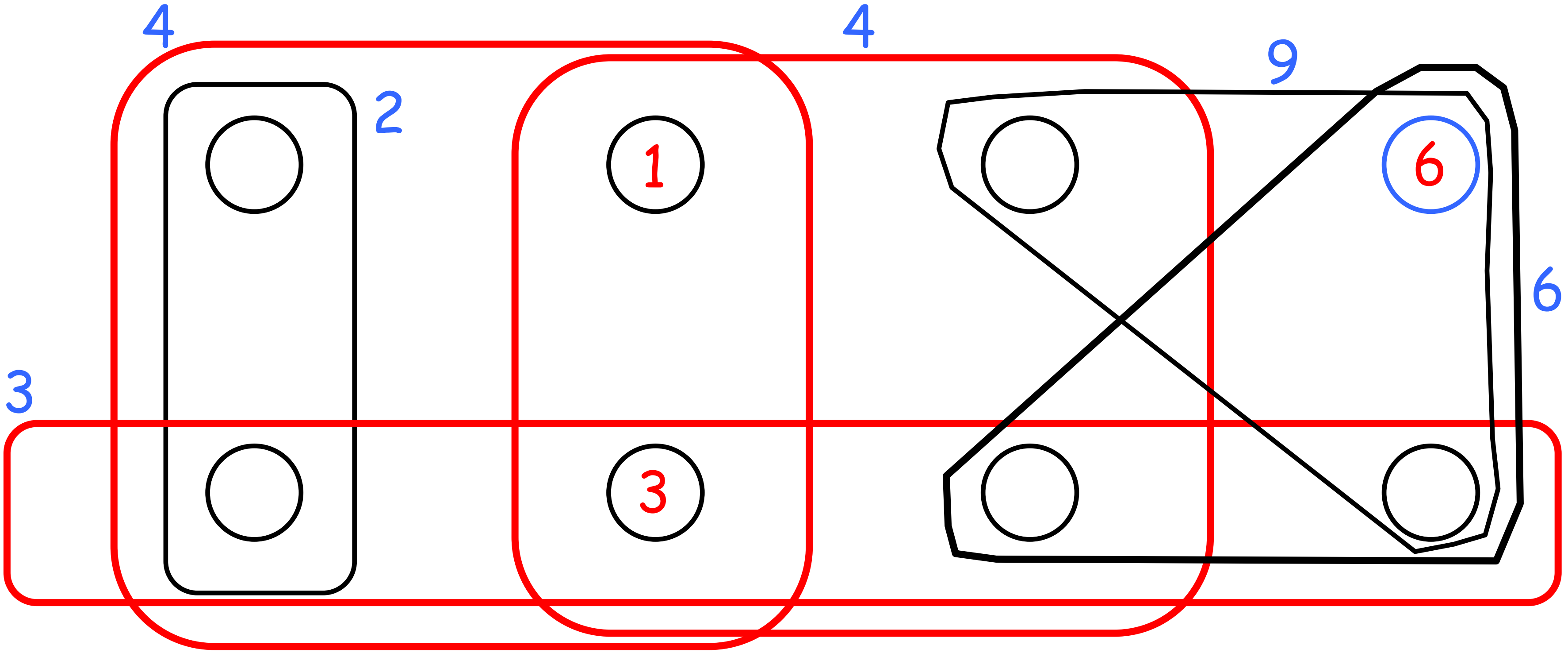


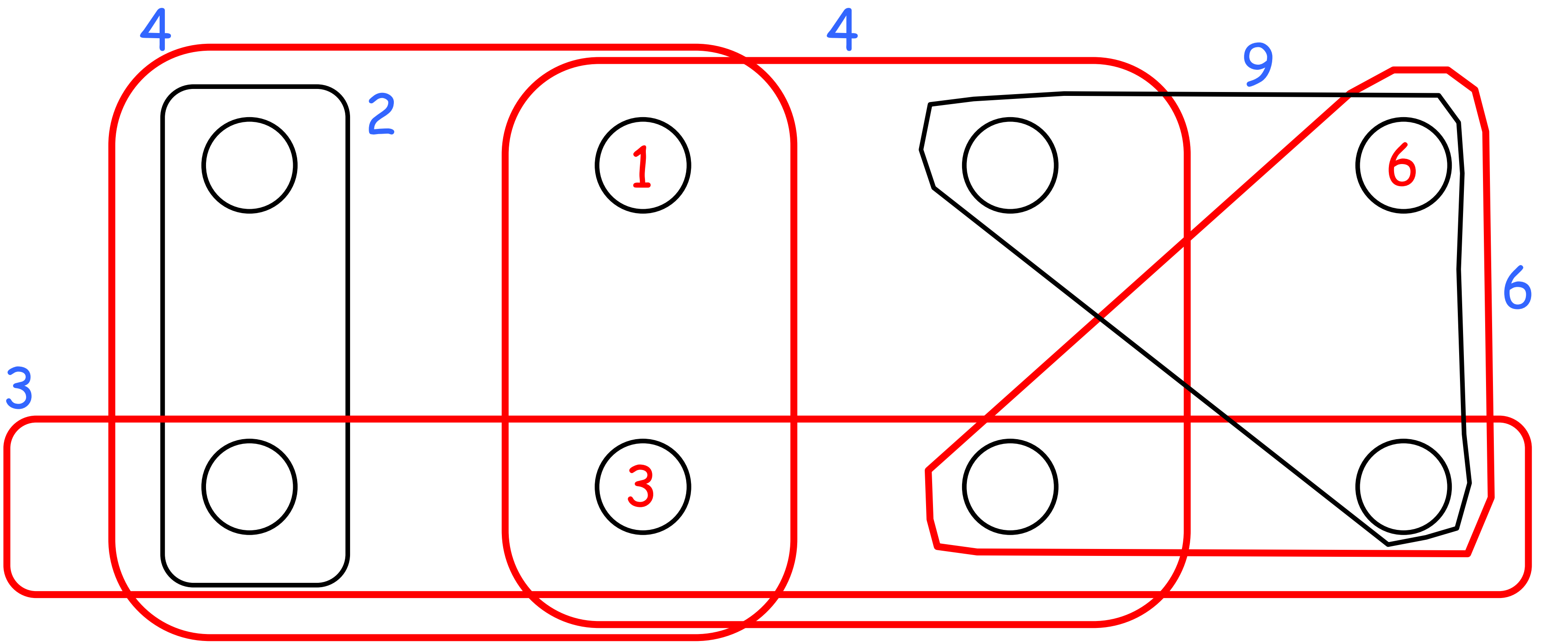
4
4
9
2
6
3
3


4
4
9
2
1
6
3
3


4
4
9
2
1
6
3
3


4
4
9
2
1
6
3
3


4
4
9
2
1
6
6
3
3


4
2
4
9
6
1
6
3
3

Theorem

The algorithm is an f-approximation algorithm for the SC problem.

proof

the computed cover is clearly feasible.

we claim that: $$\sum_{S\in\mathcal{S}} c(S)x_S \leq f \underbrace{\sum_{e\in U} y_e}$$

each element e:
- has $f \cdot y_e$ amount of money
- pays y_e for each picked set S containing e

Think of it as money you can use to buy the picked primal solution

since each e is in at most f sets, e has enough money for its payments

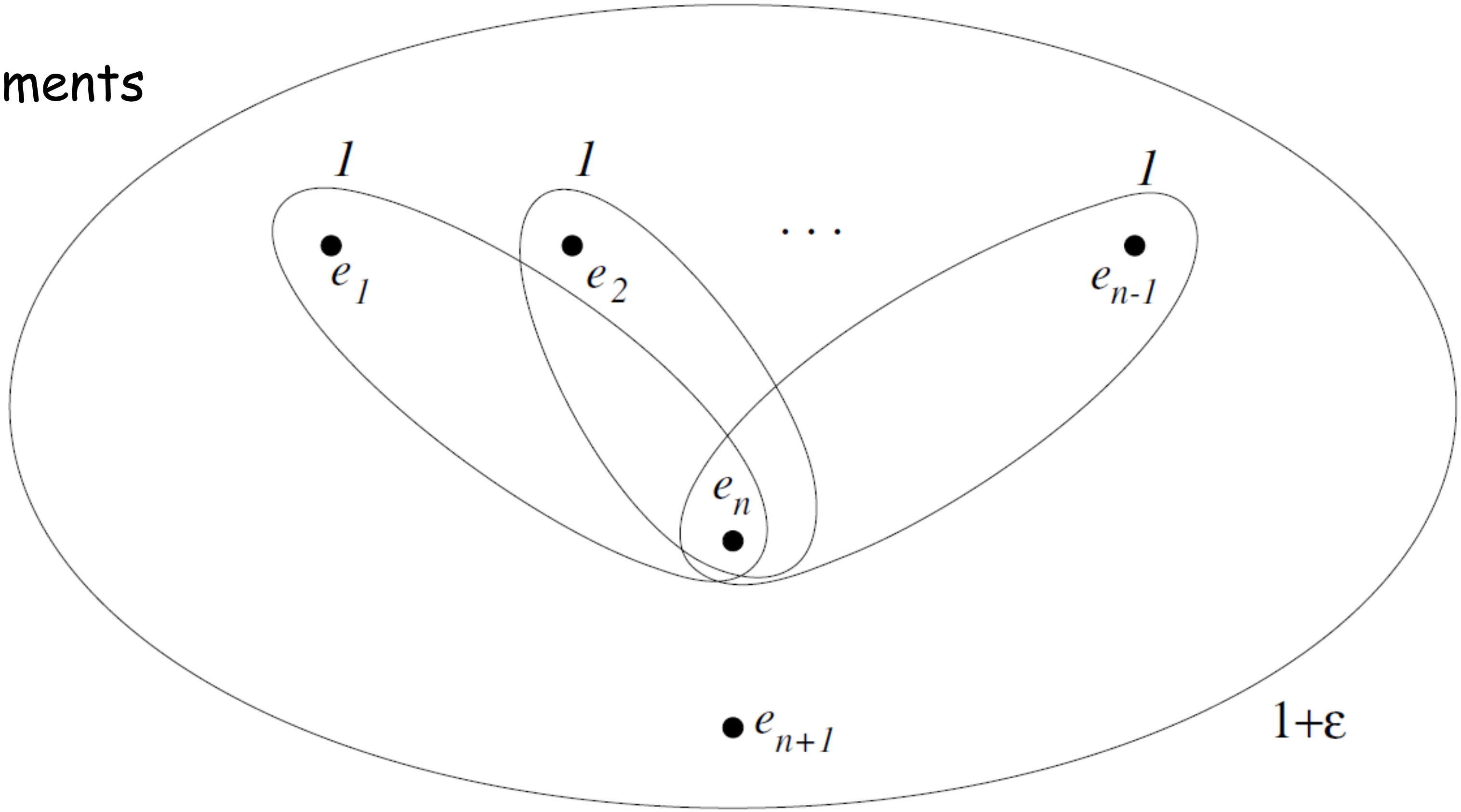
since each picked set S is tight, S is fully paid for by the elements it contains

since y is feasible: $$\sum_{e\in U} y_e \leq OPT$$

tight example

n+1 elements

$f=n$



suppose the algorithm raises first variable y_{e_n}

returned solution has cost $n+\varepsilon$

$OPT=1+\varepsilon$

The Steiner Forest problem

minimum Steiner Forest problem

Input:

- undirected graph $G=(V,E)$ with non-negative edge costs
- collection of disjoint subsets of V, $S_1,\dots,S_k$

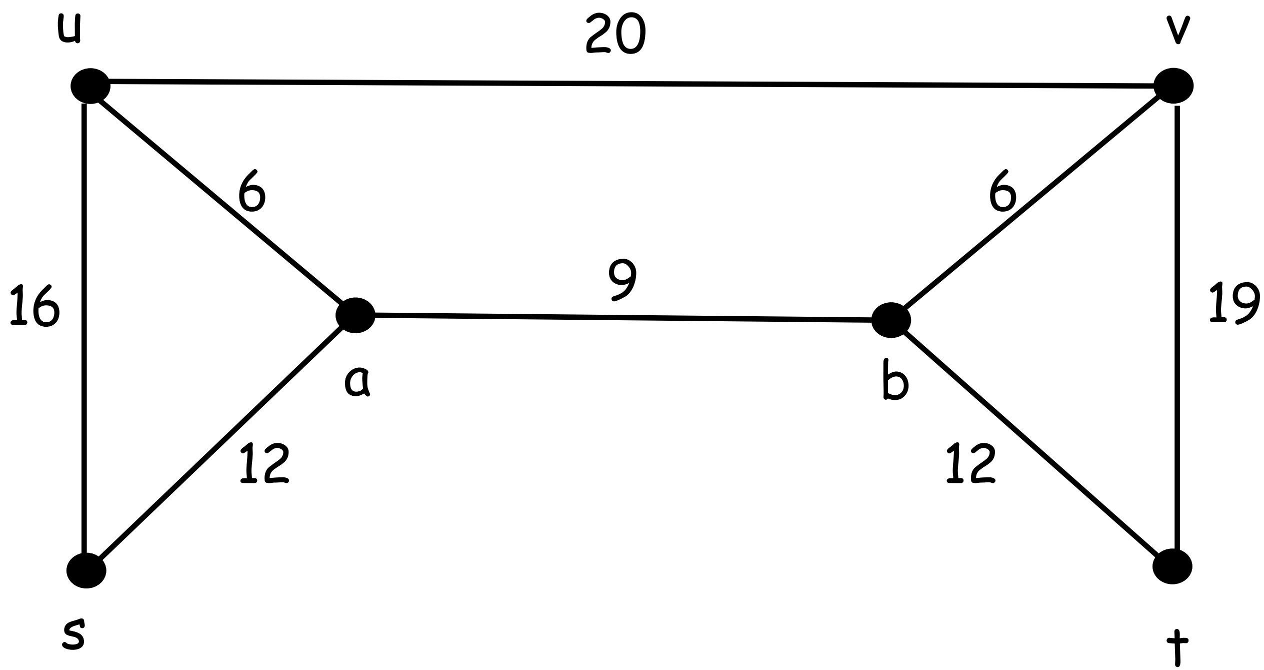
Feasible solution:

a forest F in which each pair of vertices belonging to the same set S_i is connected

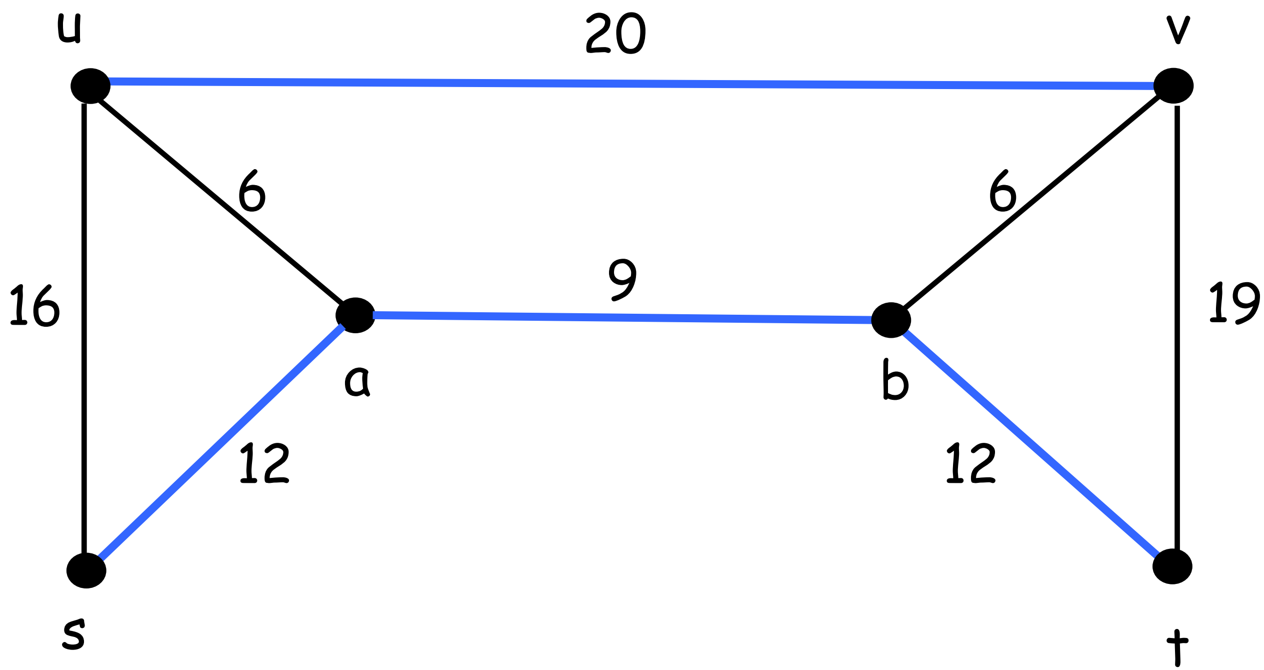
measure (min):

cost of F : $\sum_{e \in E(F)} c(e)$

$S_1=\{u,v\}$ $S_2=\{s,t\}$

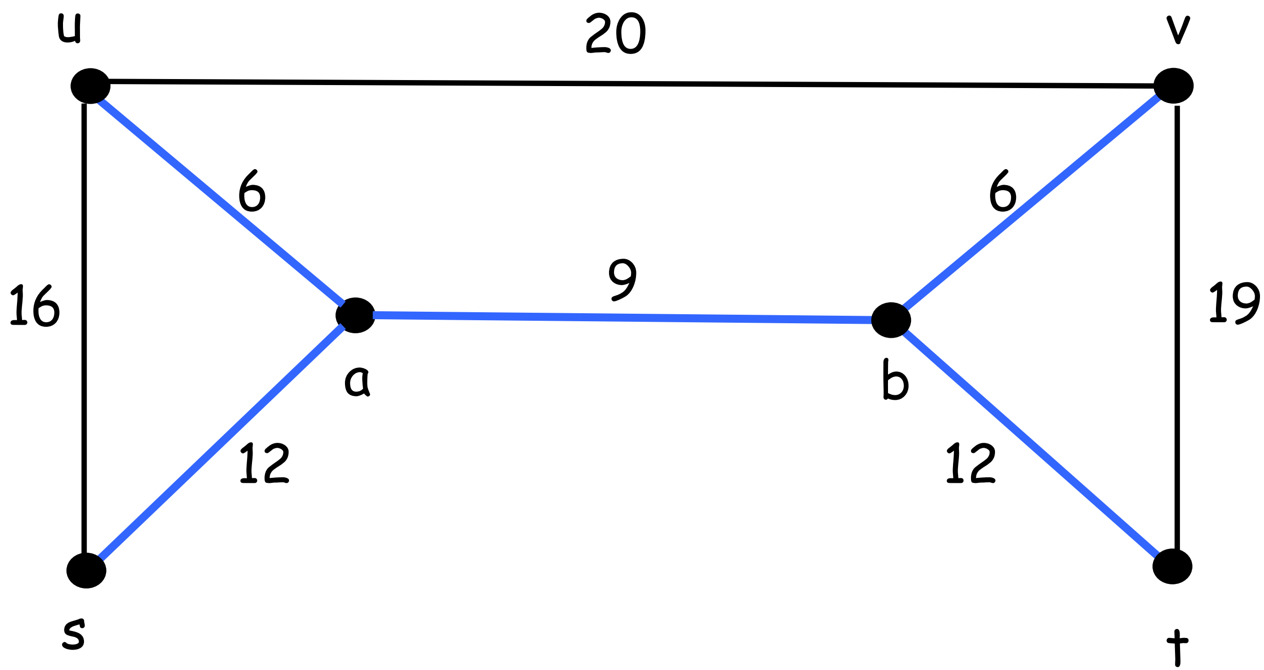


$S_1=\{u,v\}$ $S_2=\{s,t\}$



a Steiner forest of cost 53

$S_1=\{u,v\}$ $S_2=\{s,t\}$



a better Steiner forest of cost 45

minimum Steiner Forest problem

Input:

- undirected graph $G=(V,E)$ with non-negative edge costs
- collection of disjoint subsets of V, $S_1,\dots,S_k$

Feasible solution:

a forest F in which each pair of vertices belonging to the same set S_i is connected

measure (min):

cost of F : $\sum_{e \in E(F)} c(e)$

metric Steiner forest problem:

- G is complete, and
- edge costs satisfy the *triangle inequality*
  for every u,v,w : $c(u,v) \leq c(u,w)+c(w,v)$

minimum Steiner Forest problem

Input:

- undirected graph $G=(V,E)$ with non-negative edge costs
- collection of disjoint subsets of V, $S_1,\ldots,S_k$

Feasible solution:

a forest F in which each pair of vertices belonging to the same set S_i is connected

measure (min):

cost of F : $\sum_{e \in E(F)} c(e)$

a connectivity requirement function r

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

a function f on all cuts in G, for each $S \subseteq V$ – cut $(S, S'=V \setminus S)$:

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

an Integer Linear Programming (ILP) formulation of SF

LP-relaxation

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

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

$$x_e \in \{0,1\} \qquad e\in E$$

relax with
$x_e \geq 0$ & $\underbrace{x_e \leq 1}_{\text{redundant}}$

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

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

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

$\delta(S)$: edges crossing the cut $(S, S'=V\setminus S)$

LP-relaxation

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

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

$$x_e \geq 0 \quad e \in E$$

dual program

$$\text{maximize} \quad \sum_{S \subseteq V} f(S)\, y_S$$

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

$$y_S \geq 0 \quad S \subseteq V$$

edge e *feels* dual y_S if $y_S > 0$ and $e \in \delta(S)$

S has been *raised* in a dual solution if $y_S > 0$

obs: raising S or S′ has the same effect

obs: no advantage in raising a set S with $f(S)=0$

⇒ assume never raise such sets

edge e is *tight* if the total amount of dual it feels equals its cost

obs: dual program tries to maximize the sum of the duals subject to no edge is *overtight* (i.e., feels more than its cost) at any point, the currently picked edges form a forest F

S is unsatisfied if $f(S)=1$ but there is no picked edge crossing the cut (S,S')

S is active if it is a minimal (w.r.t. inclusion) unsatisfied set in F

obs: if F is not feasible then there must be an active set

Lemma

Set S is active iff it is a connected component in the currently picked forest and $f(S)=1$.

proof

Let S be an active set

S cannot contain part of a connected component because otherwise there will already be a picked edge in the cut (S,S')

⇒ S is the union of connected components

Since $f(S)=1$, there is a vertex $u \in S$ and $v \in S'$ such that $r(u,v)=1$

Let S' be the connected component containing u

minimality of S implies $S'=S$.

∎

Algorithm 22.3 (Steiner forest)

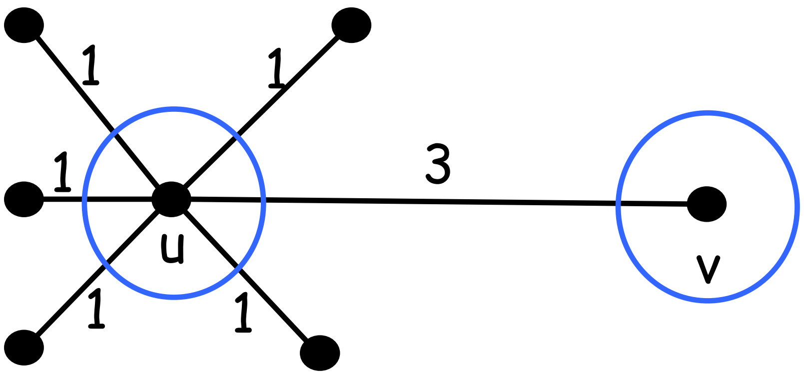
1. **(Initialization)** $F \leftarrow \emptyset$; for each $S \subseteq V$, $y_S \leftarrow 0$.
2. **(Edge augmentation)** while there exists an unsatisfied set do:
   simultaneously raise y_S for each active set S, until some edge e goes tight;
   $F \leftarrow F \cup \{e\}$.
3. **(Pruning)** return $F' = \{e \in F \mid F - \{e\}$ is primal infeasible$\}$

discard all redundant edges

an edge $e \in F$ is redundant if $F-\{e\}$ is also a feasible solution

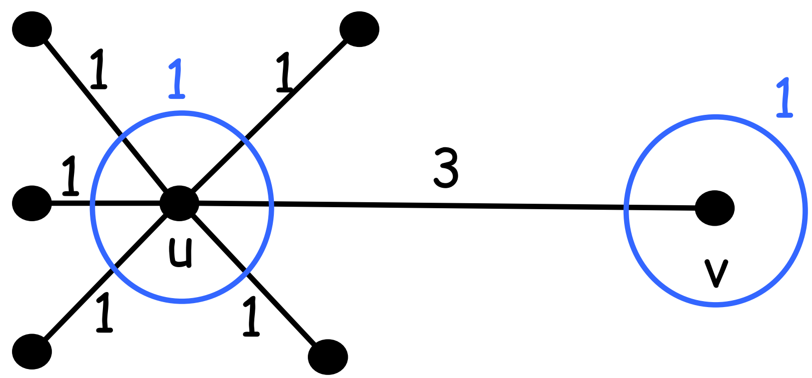
example: pruning step is needed

$S_1=\{u,v\}$



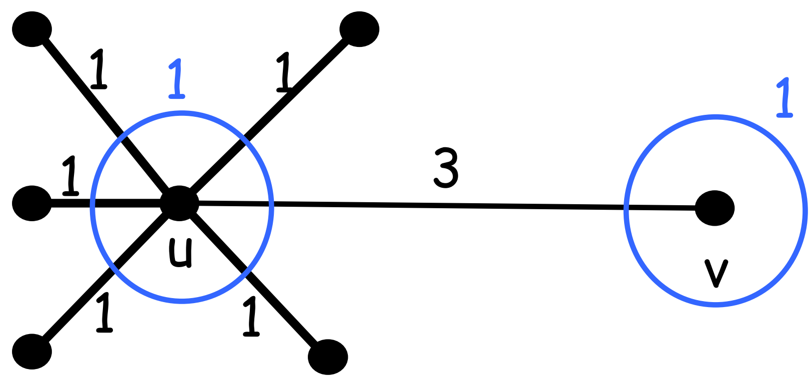
example: pruning step is needed

$S_1=\{u,v\}$



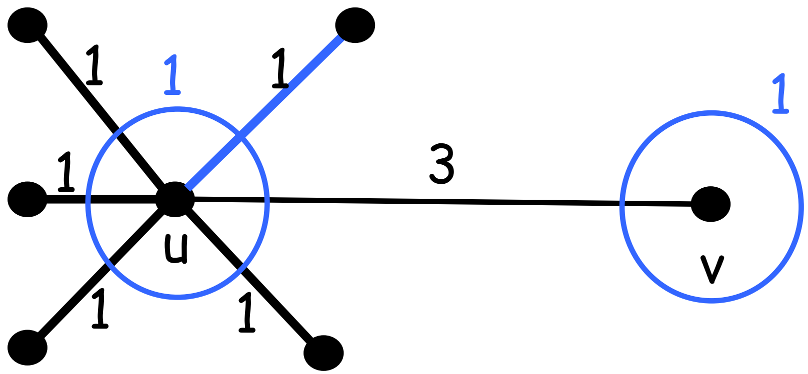
example: pruning step is needed

$S_1=\{u,v\}$



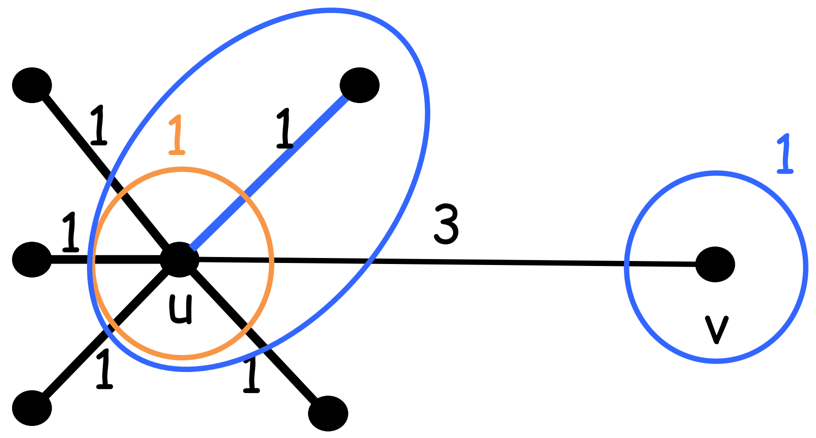
example: pruning step is needed

$S_1=\{u,v\}$



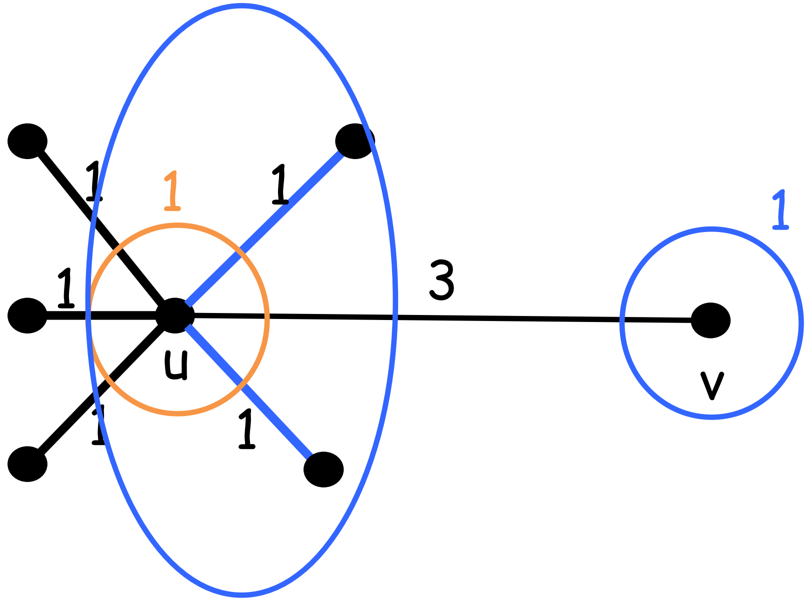
example: pruning step is needed

$S_1=\{u,v\}$



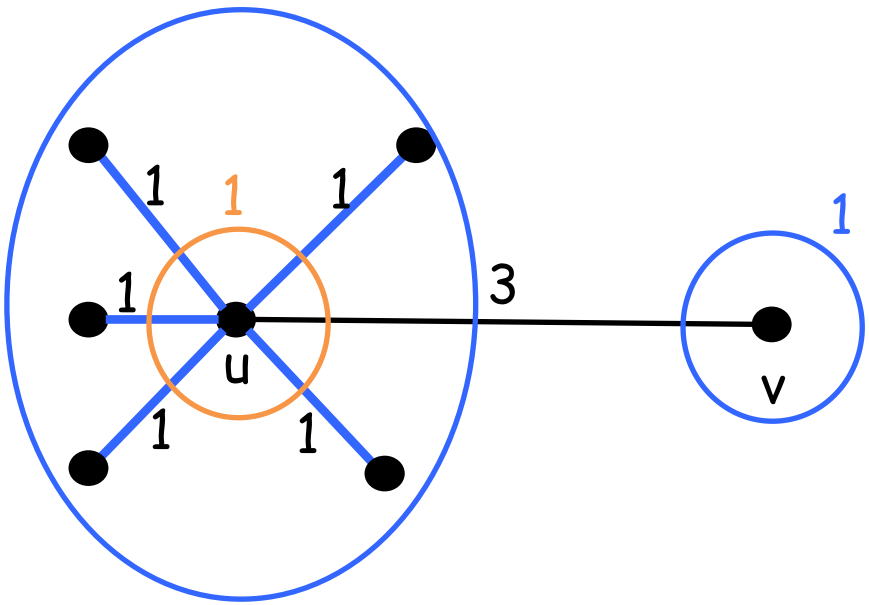
example: pruning step is needed

$S_1=\{u,v\}$



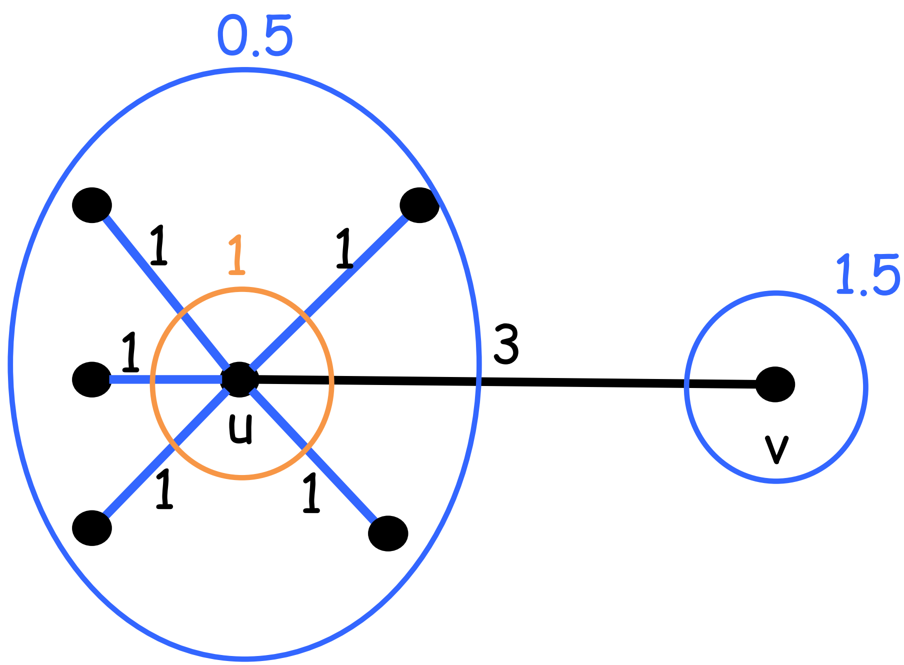
example: pruning step is needed

$S_1=\{u,v\}$



example: pruning step is needed

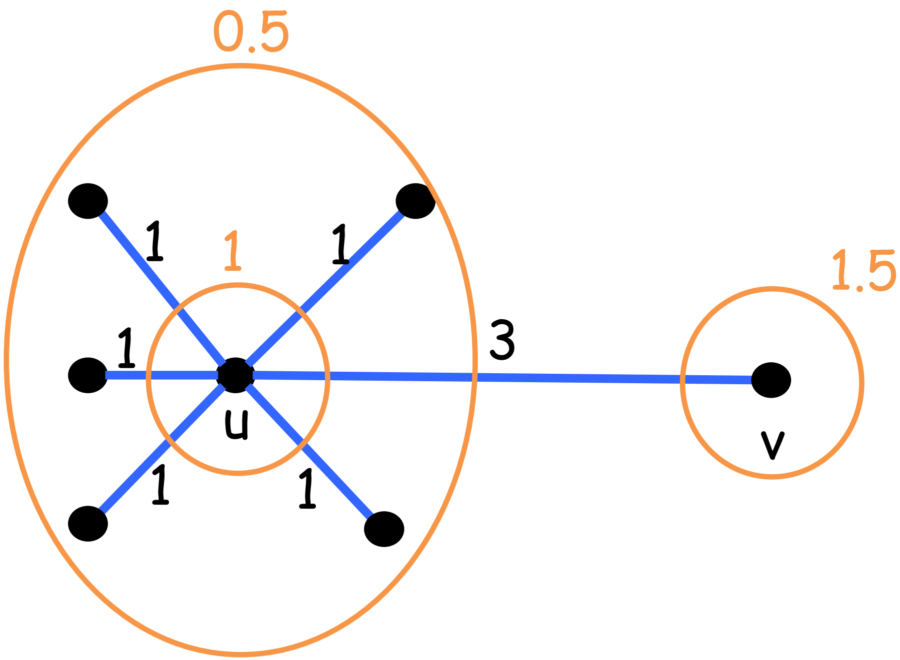
$S_1=\{u,v\}$



example: pruning step is needed

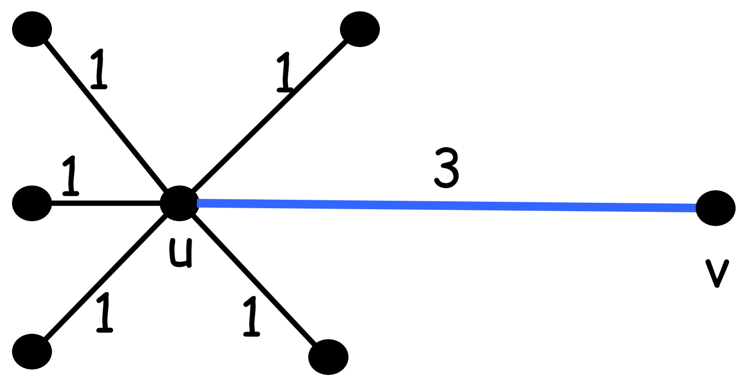
$S_1=\{u,v\}$

...pruning...



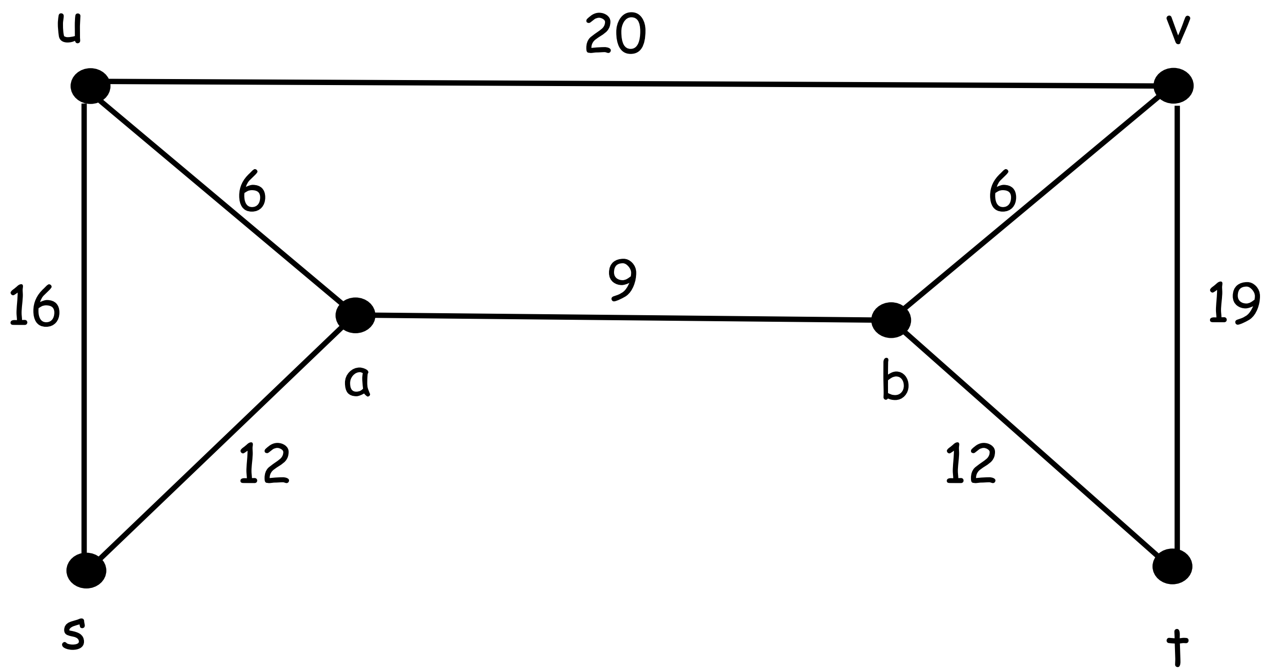
example: pruning step is needed

$S_1=\{u,v\}$

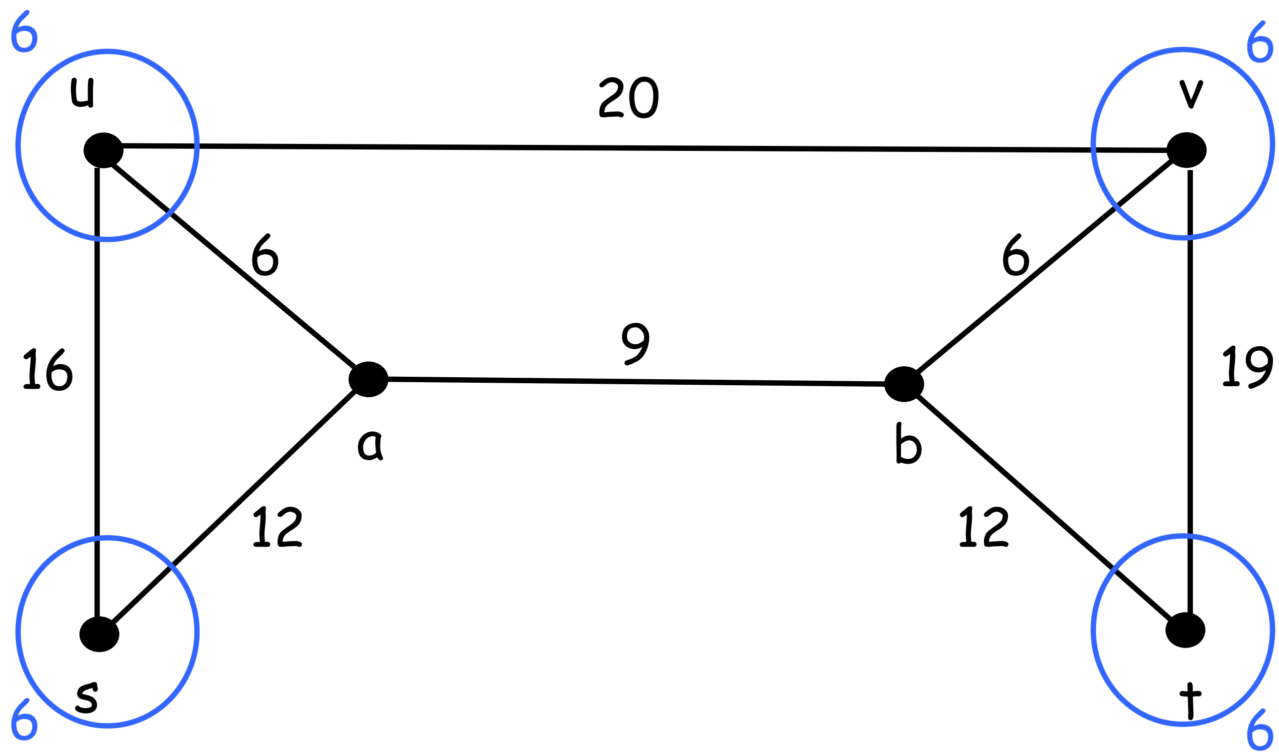


computed
solution

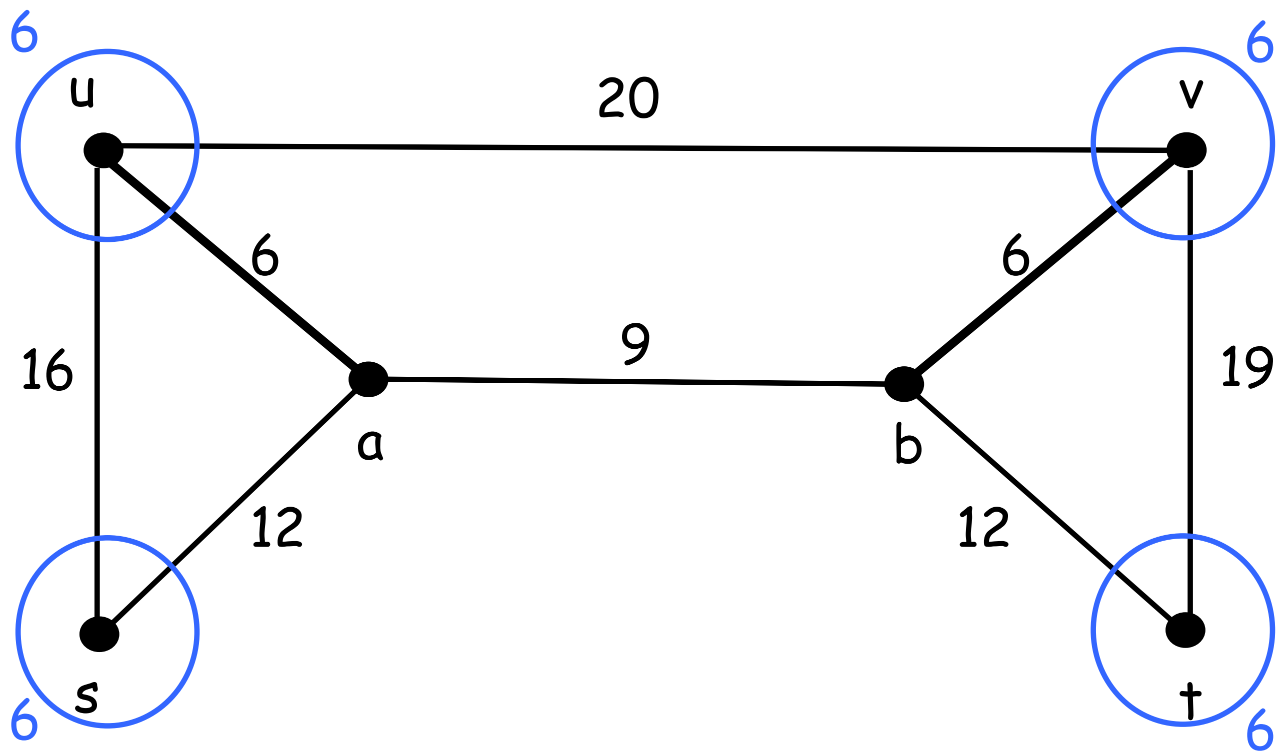
$S_1=\{u,v\}$ $S_2=\{s,t\}$



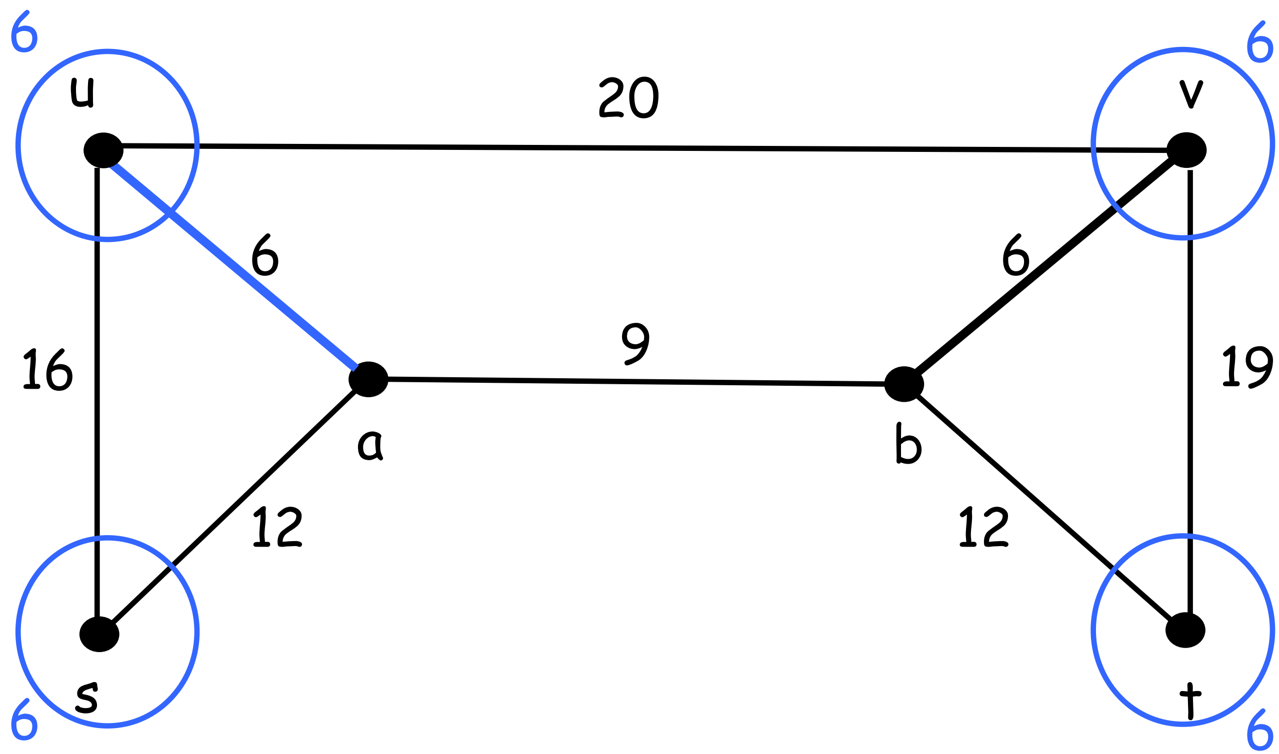
$S_1=\{u,v\}$ $S_2=\{s,t\}$

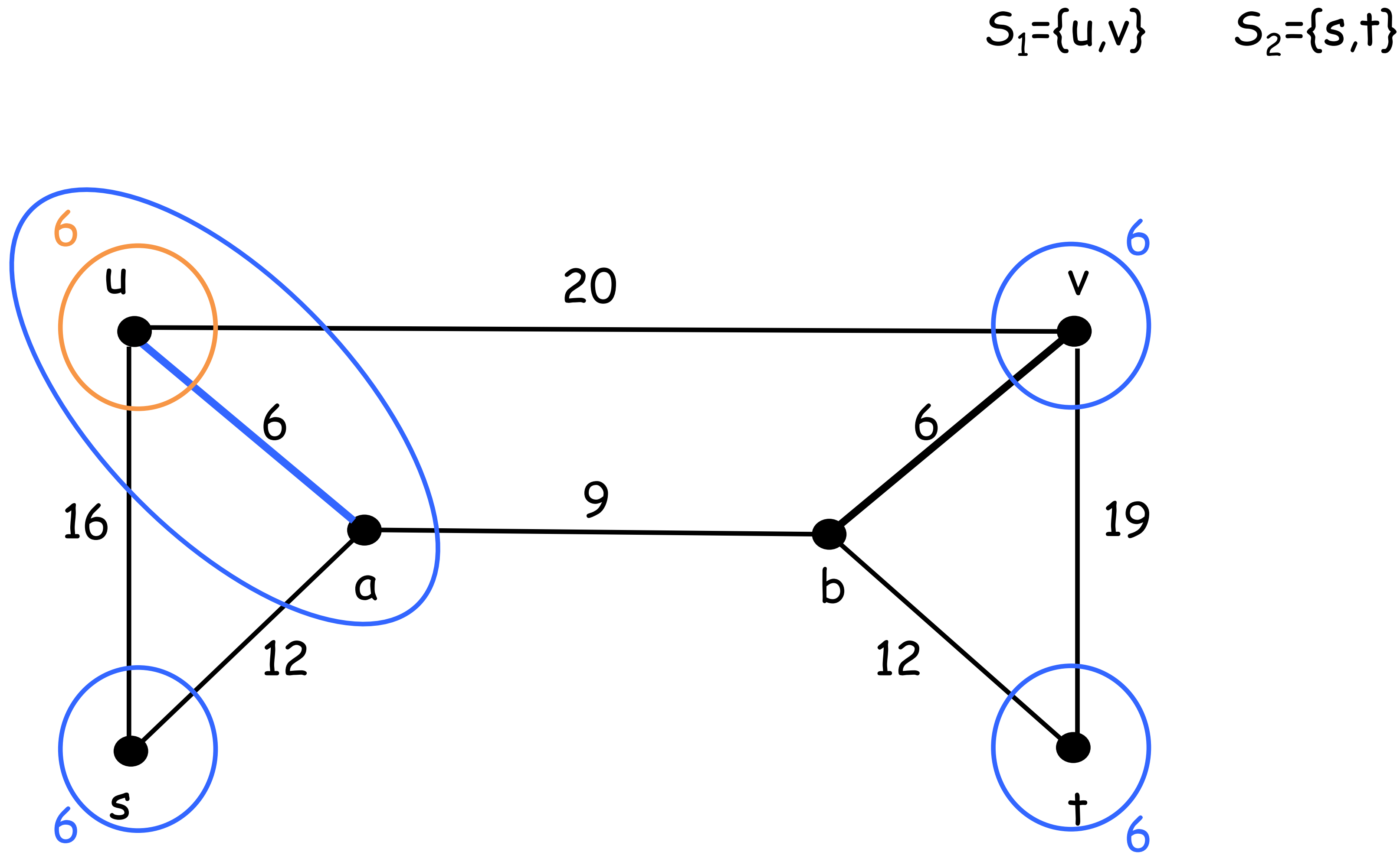


$S_1=\{u,v\}$ $S_2=\{s,t\}$



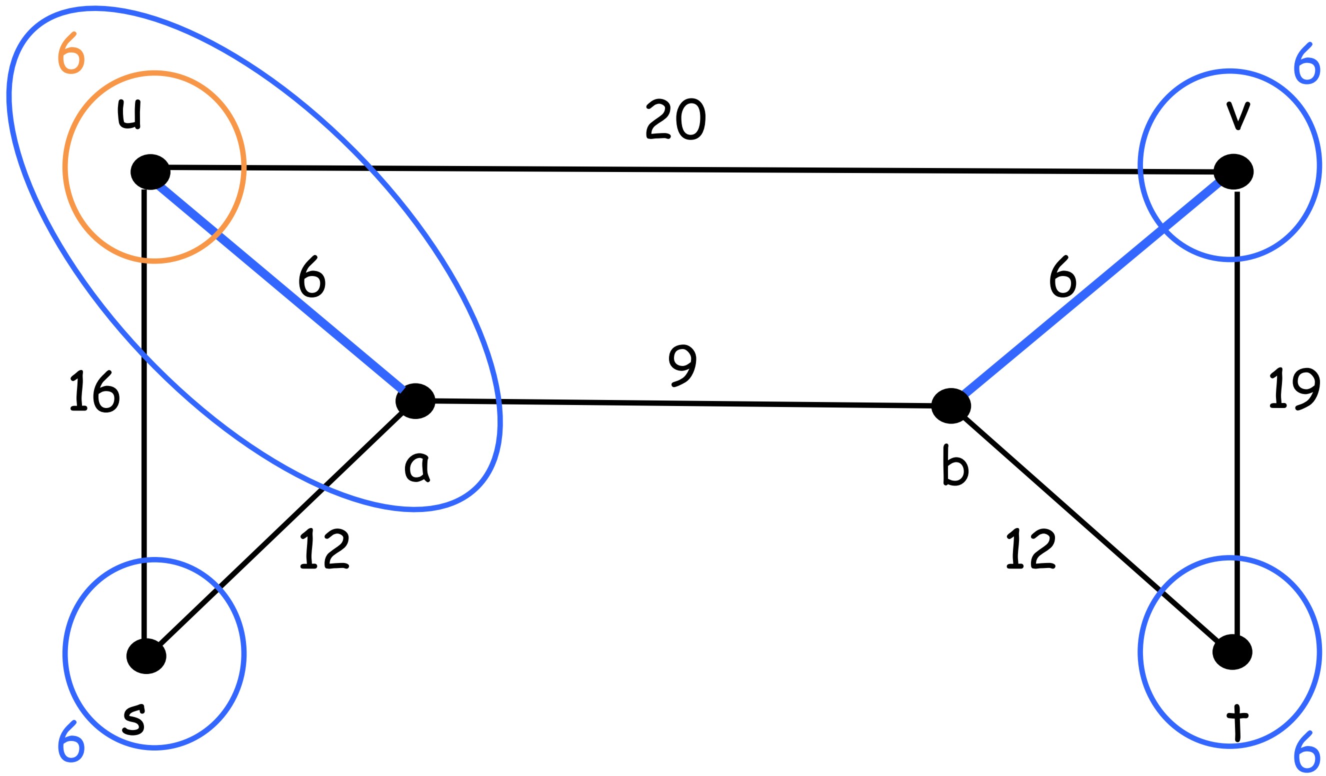
$S_1=\{u,v\}$ $S_2=\{s,t\}$



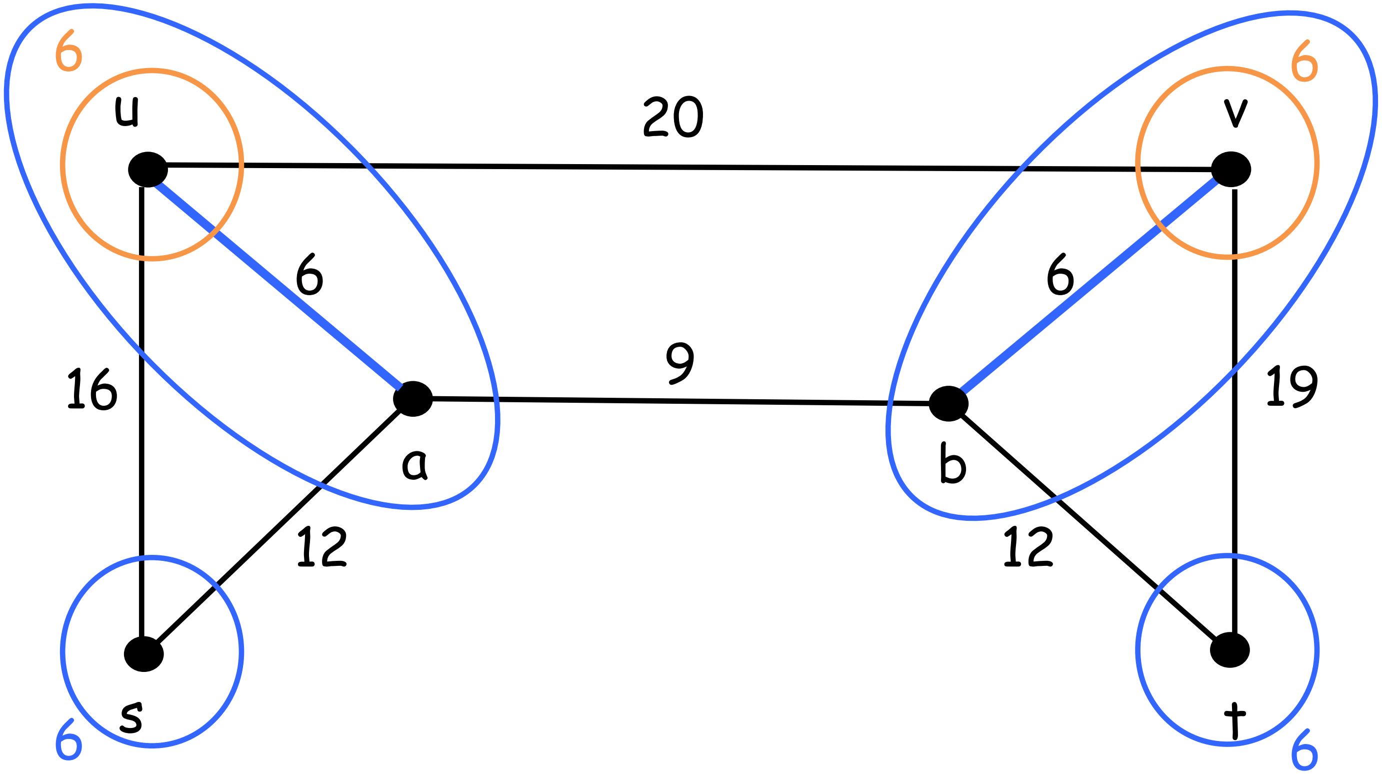


$S_1=\{u,v\}$
$S_2=\{s,t\}$
6
u
20
v
6
6
6
16
9
19
a
b
12
12
s
t
6
6

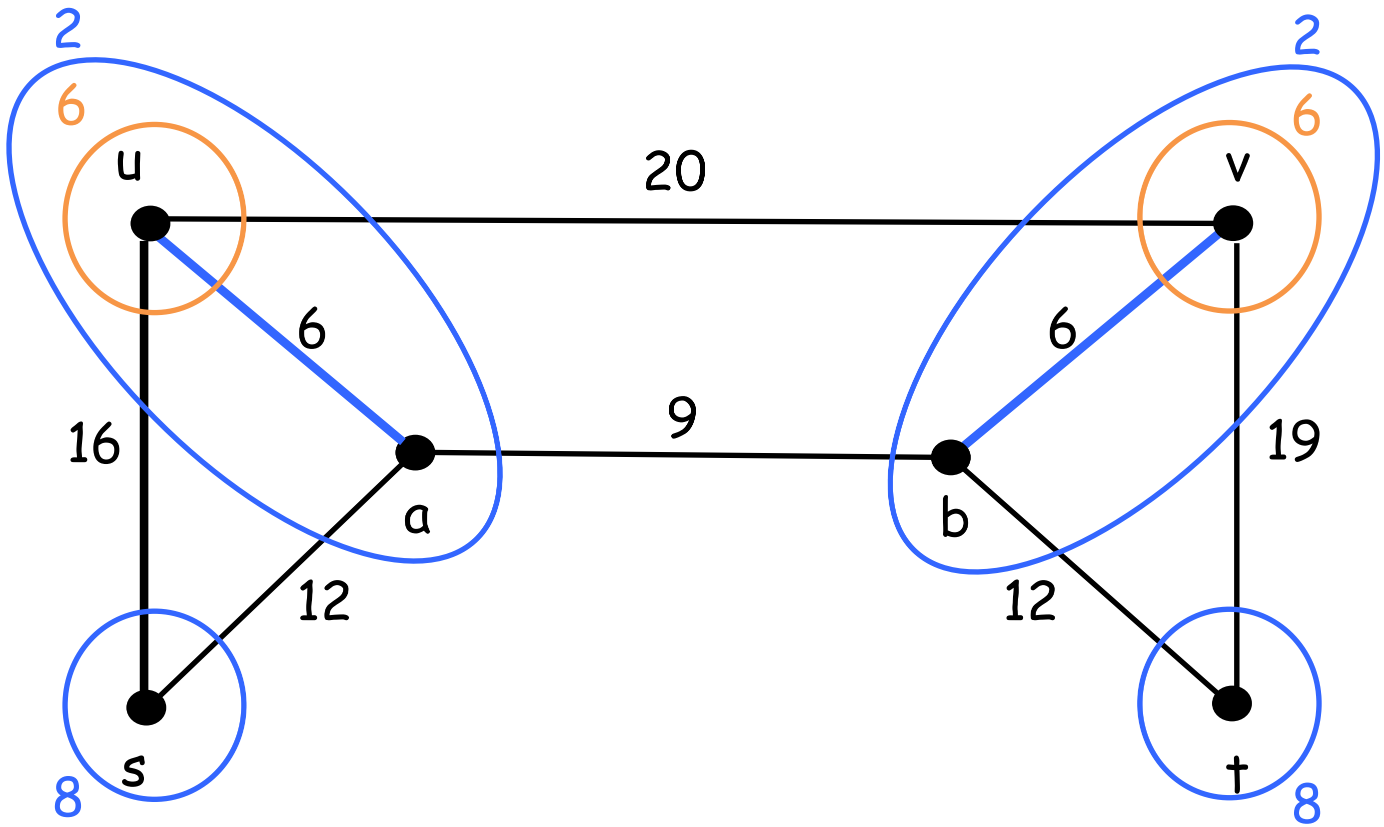
$S_1=\{u,v\}$ $S_2=\{s,t\}$



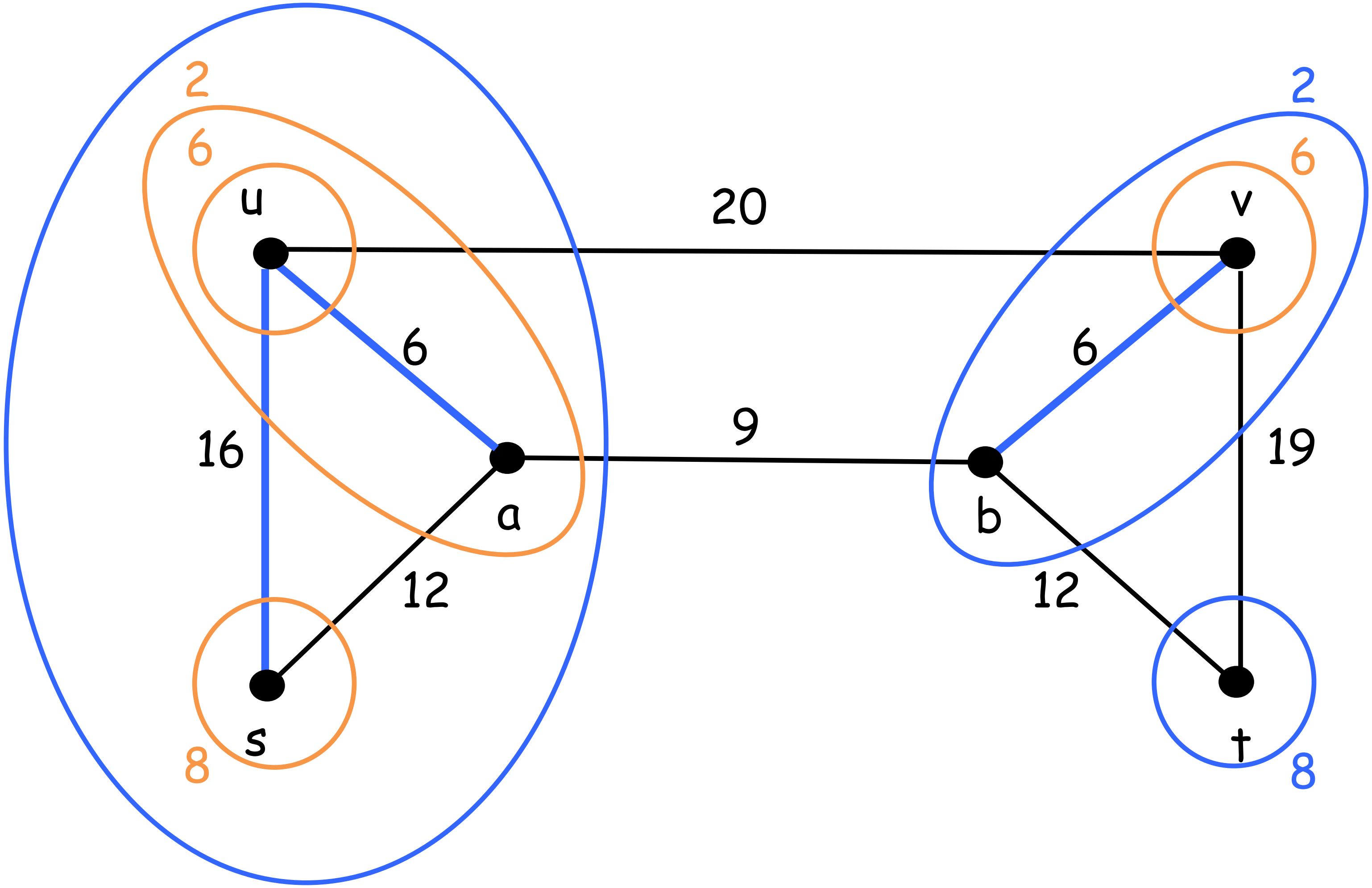
$S_1=\{u,v\}$ $S_2=\{s,t\}$



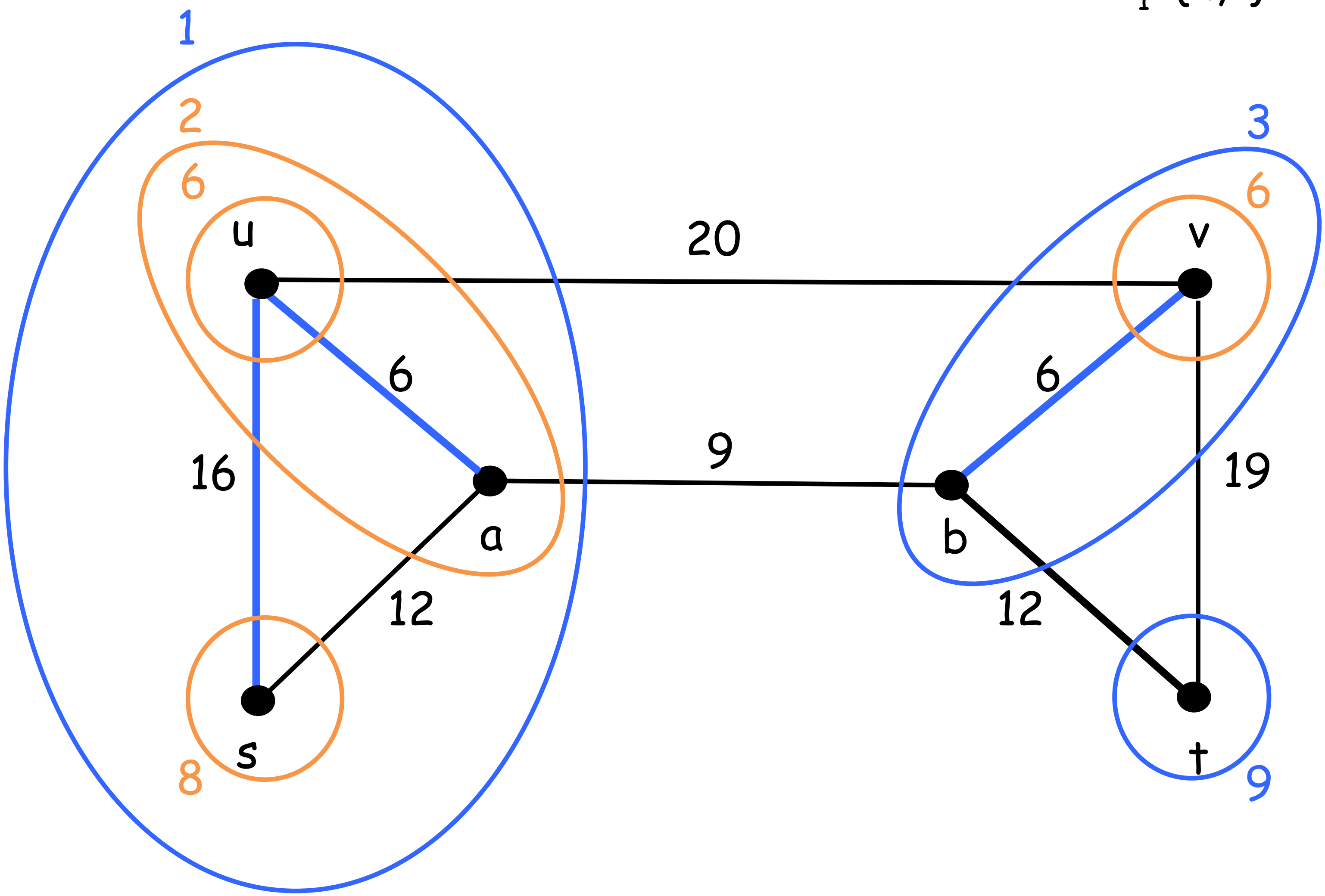
$S_1=\{u,v\}$ $S_2=\{s,t\}$

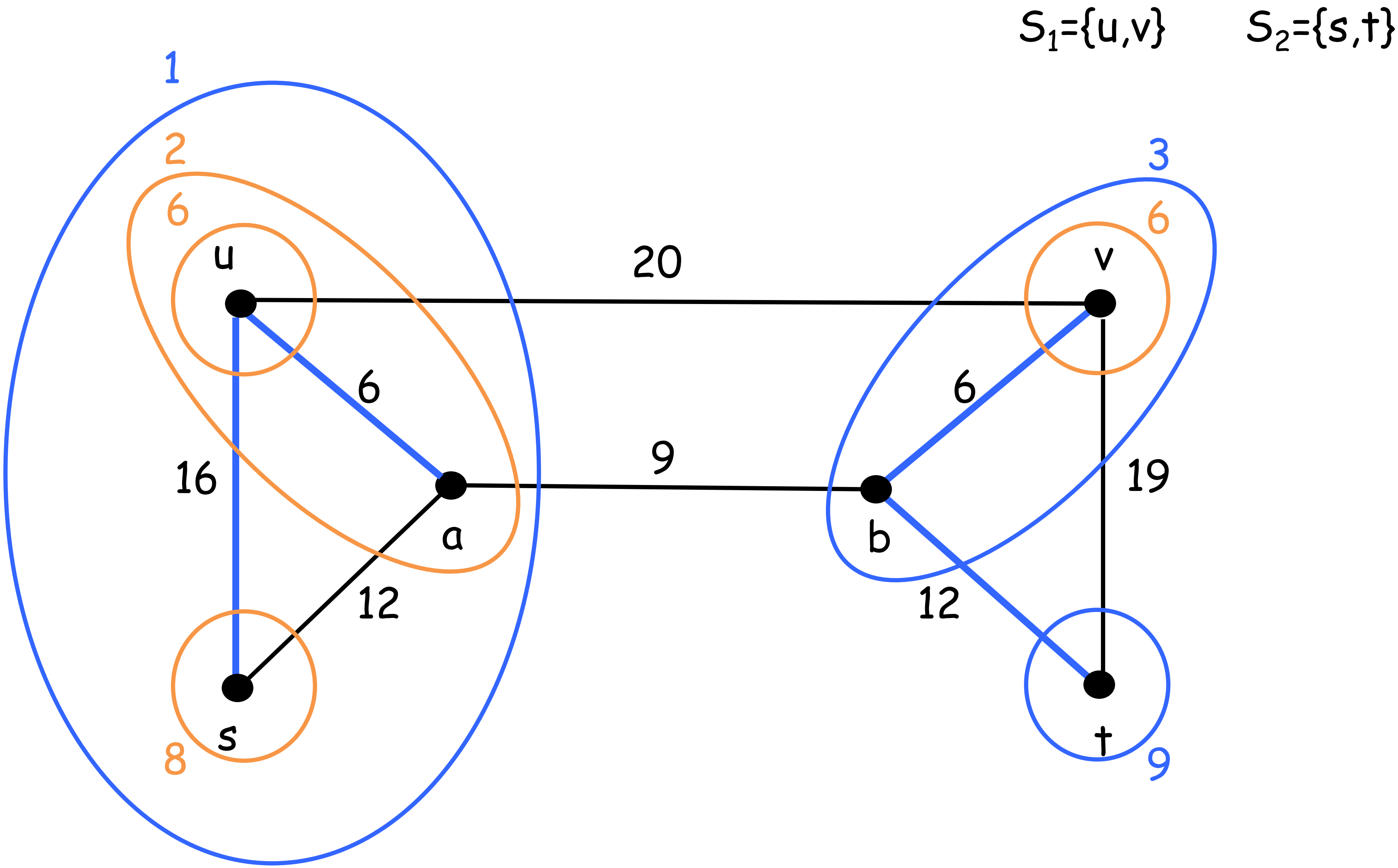


$S_1=\{u,v\}$ $S_2=\{s,t\}$

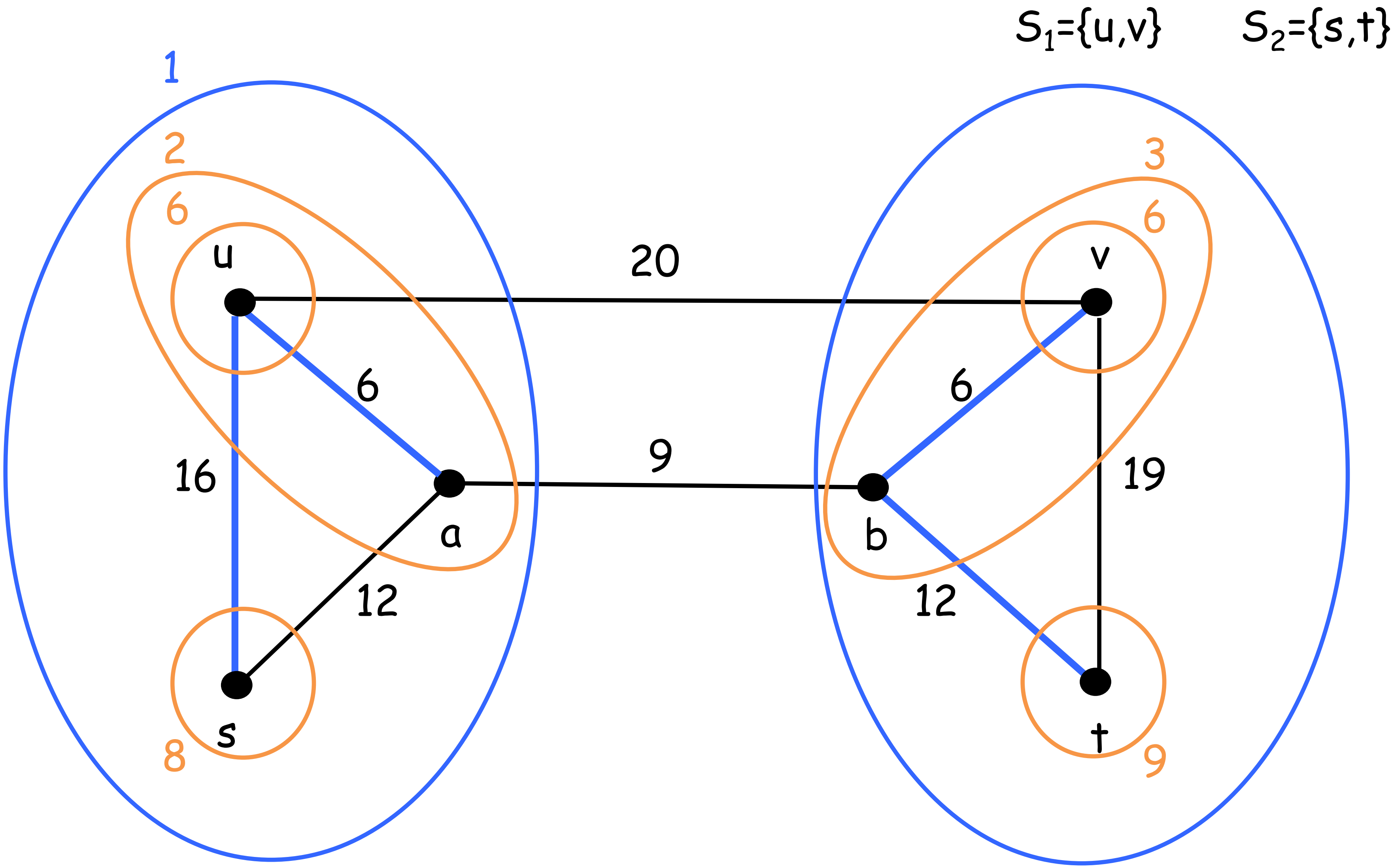


$S_1=\{u,v\}$ $S_2=\{s,t\}$

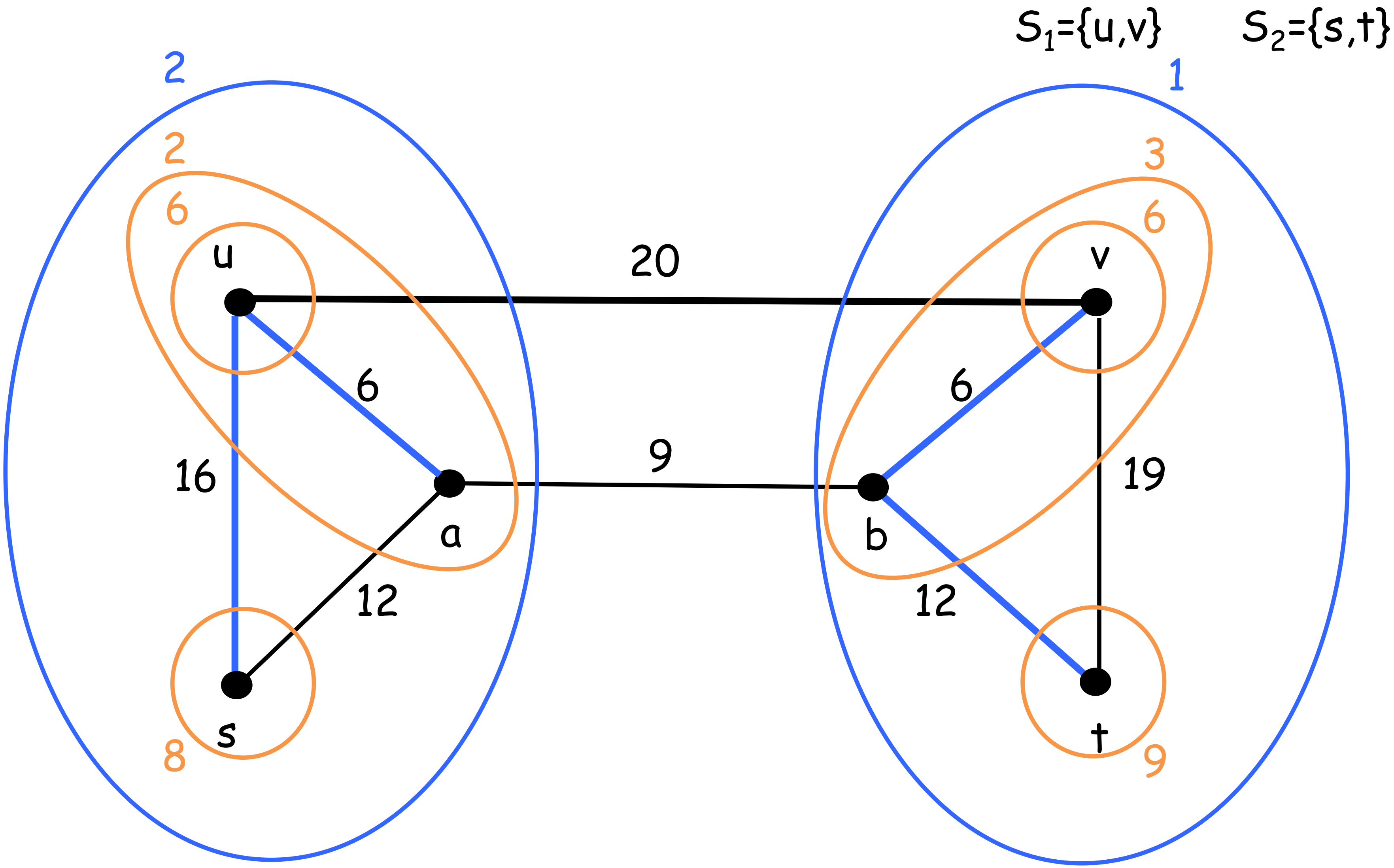




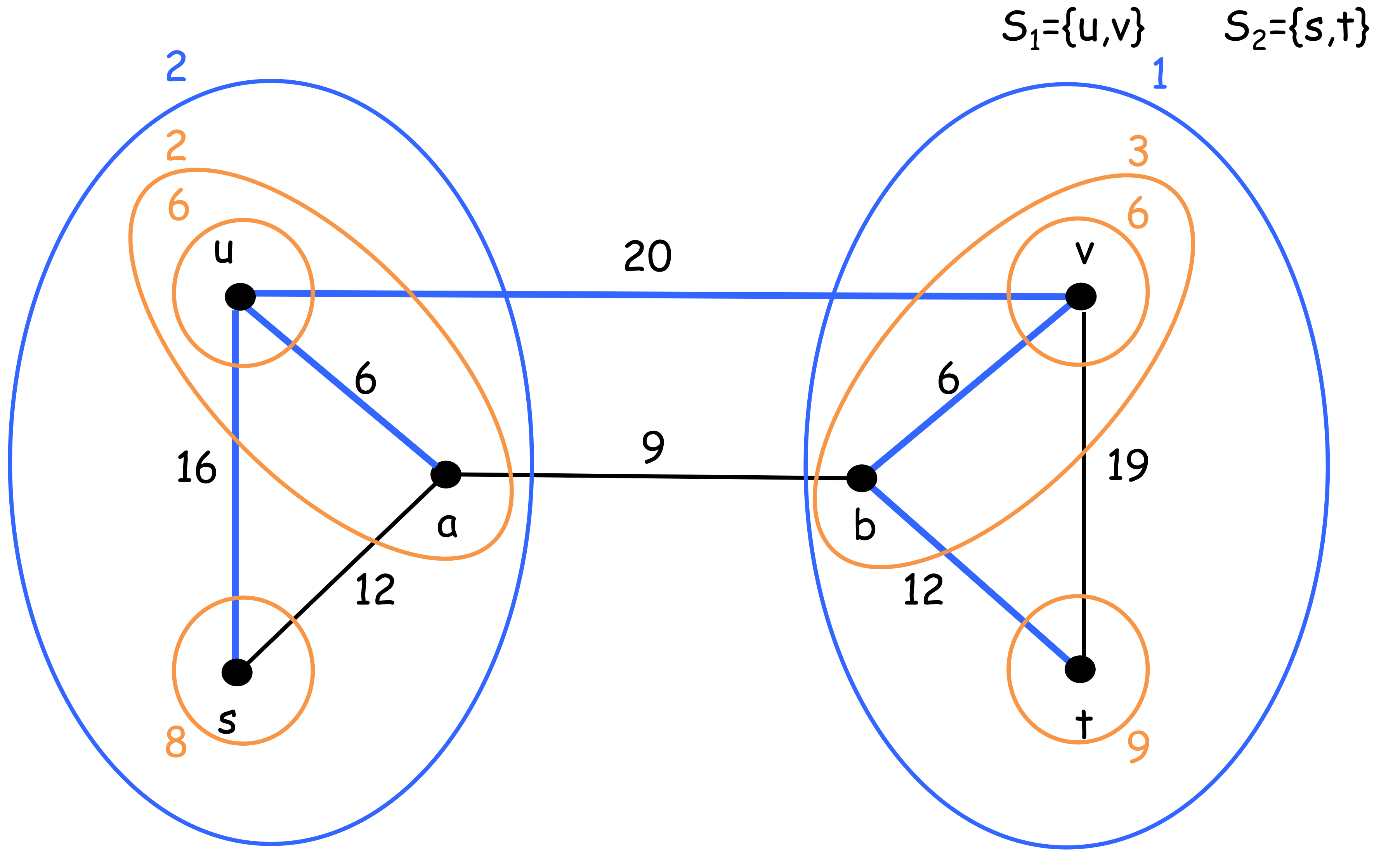
$S_1=\{u,v\}$
$S_2=\{s,t\}$
1
2
6
u
20
3
6
v
6
6
16
9
19
a
b
12
12
s
8
t
9



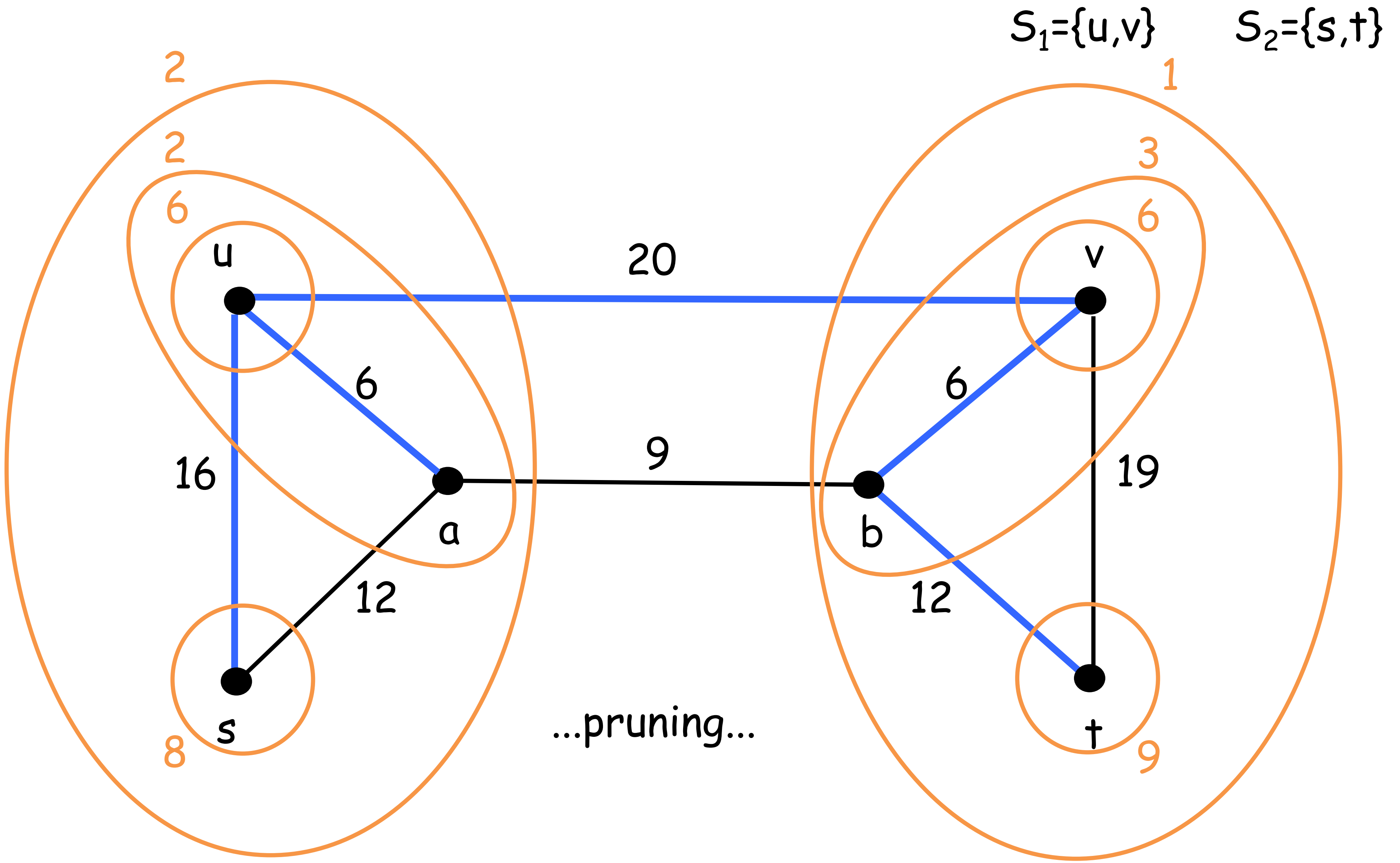
$S_1=\{u,v\}$
$S_2=\{s,t\}$
1
2
3
6
6
8
9
u
v
a
b
s
t
20
6
6
9
16
19
12
12



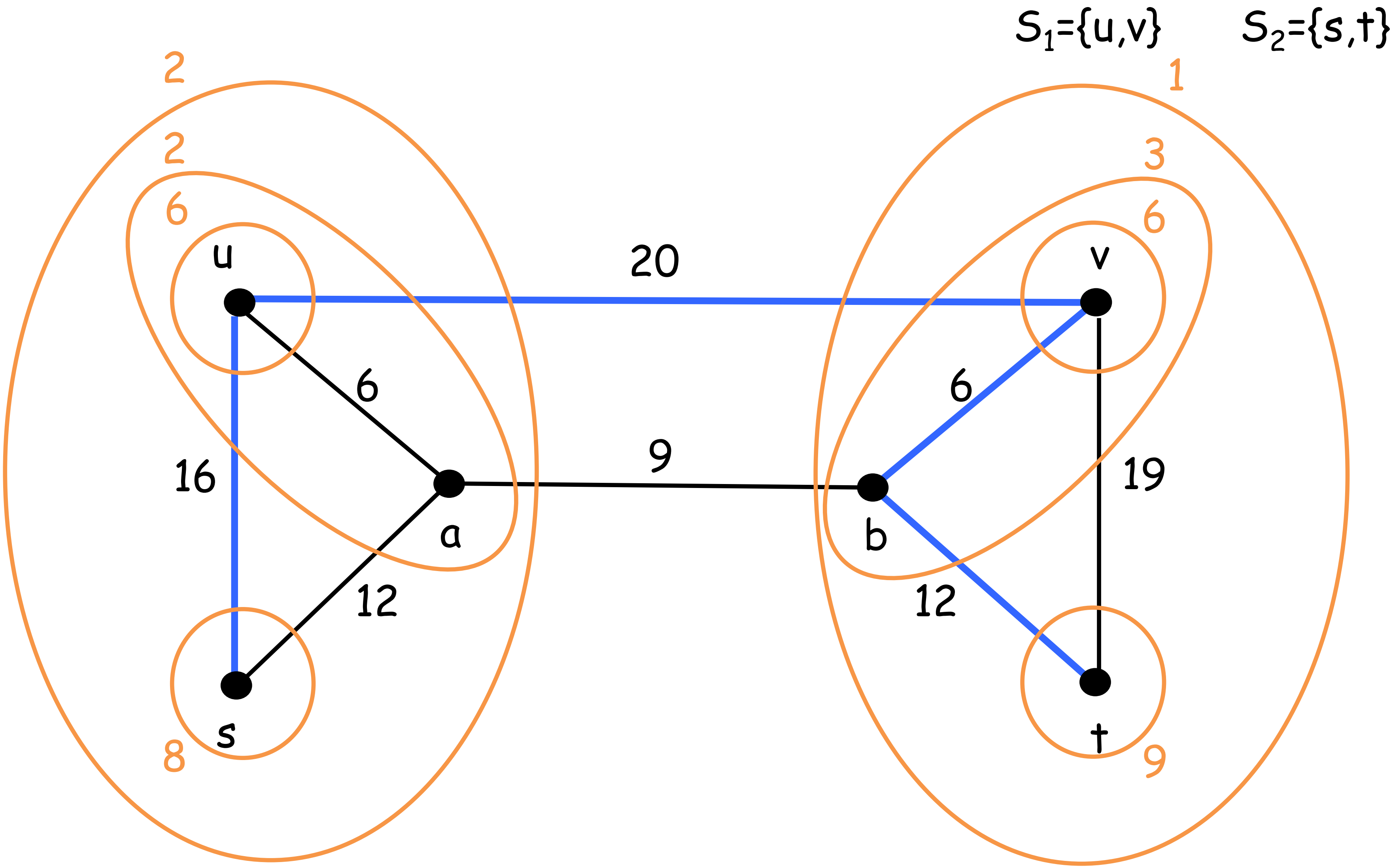
$S_1=\{u,v\}$ $S_2=\{s,t\}$
2
1
2
3
6
6
u
v
20
6
6
16
9
19
a
b
12
12
s
t
8
9



S_1={u,v}
S_2={s,t}
2
1
2
3
6
6
u
v
20
6
6
16
9
19
a
b
12
12
s
t
8
9

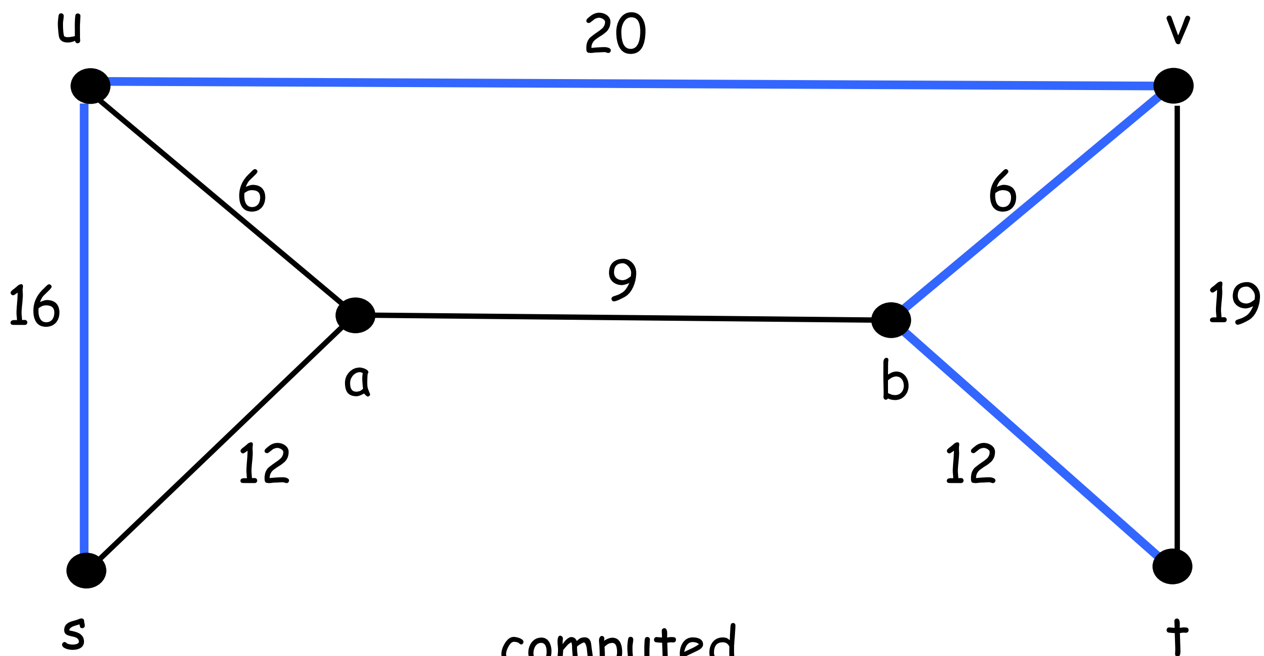


S_1={u,v}
S_2={s,t}
2
1
2
3
6
6
u
v
20
6
6
16
9
19
a
b
12
12
s
t
8
9
...pruning...



S1={u,v}
S2={s,t}
2
2
6
u
20
6
16
9
a
12
s
8
1
3
6
v
6
19
b
12
t
9

$S_1=\{u,v\}$ $S_2=\{s,t\}$



computed solution

Theorem

The algorithm is a 2-approximation algorithm for the SF problem.

proof

The primal computed solution F' is feasible

The dual solution is feasible, since there is no overtight edge

We claim that: $\sum_{e \in F'} c_e \le 2 \sum_{S \subseteq V} y_S$

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

since every picked edge is tight

changing the order of summation

$\deg_{F'}(S)$= # of picked edges crossing the cut (S,S'=V\S) we need to show that: $\sum_{S \subseteq V} \deg_{F'}(S)\, y_S \le 2 \sum_{S \subseteq V} y_S$

We prove a stronger claim:
- in each iteration the increase in the l.h.s. ≤ the increase of in r.h.s.

Consider an iteration, and let Δ be the extent to which active sets were raised in this iteration.

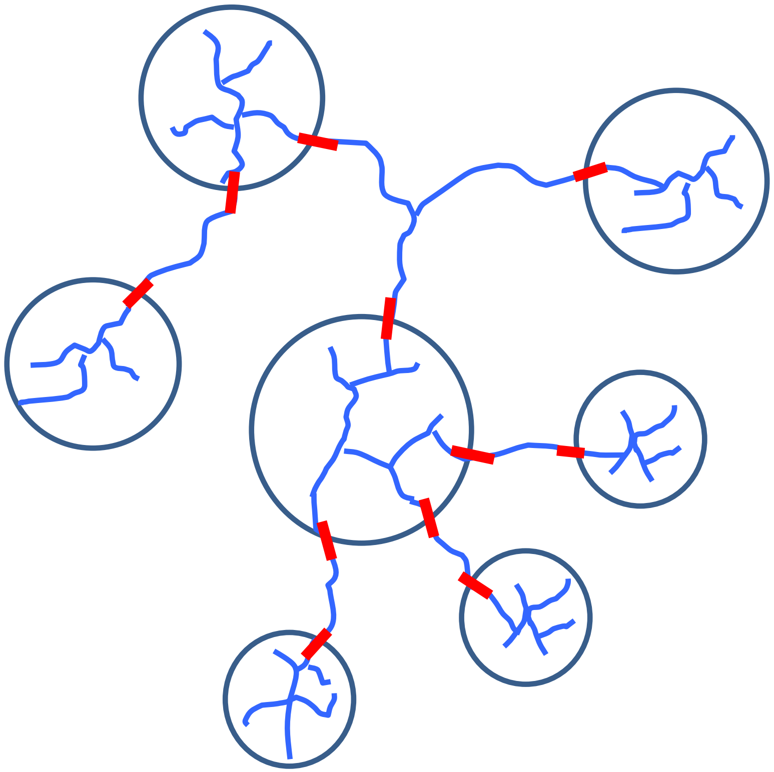
we need to show that:

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

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

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

F' is a forest with no redundant edges



#of red sticks ≤ 2 #of blue circles

shrink blue circles and root the obtained tree arbitrarily

every shrunk circle pays for:

- its red stick towards its parent
- parent's red stick towards it

Thus:

$$\sum_{e \in F'} c_e \leq 2 \sum_{S \subseteq V} y_S \leq 2 \text{ OPT}$$