

# Clustering

K-means

---

Haotian Wang and Yiyang Li

School of Software, Shanghai Jiao Tong University

## Introduction to clustering

---

## Definition

A cluster is a collection of objects which are “similar” between them and are “dissimilar” to the objects belonging to other clusters.

## Definition

Clustering is the algorithm that recognizes clusters from a given data set.

## Definition

A cluster is a collection of objects which are “similar” between them and are “dissimilar” to the objects belonging to other clusters.

## Definition

Clustering is the algorithm that recognizes clusters from a given data set.

Part of common application domains in which the clustering problem arises are as follows:

- Multimedia Data Analysis
- Responding to public health crises
- Intermediate Step for other fundamental data mining problems
- Intelligent Transportation

## K-means Algorithm

---

The **k-means clustering** problem is one of the oldest and most important questions in all of computational geometry.

Given an integer  $k$  and a set of  $n$  data points in  $\mathbb{R}^d$ , the goal of this problem is to choose  **$k$  centers** so as to **minimize the total squared distance between each point and its closest center**.

The most common K-means algorithm was first proposed by **Stuart Lloyd** of Bell Labs in 1957.

## Max-Flow and Min-Cut Problems

---

# A Flow network

A **flow network** is a tuple  $G = (V, E, s, t, c)$ .

- Diagraph  $(V, E)$  with **source**  $s \in V$  and **sink**  $t \in V$ .
- Capacity  $c(e) > 0$  for each  $e \in E$ .

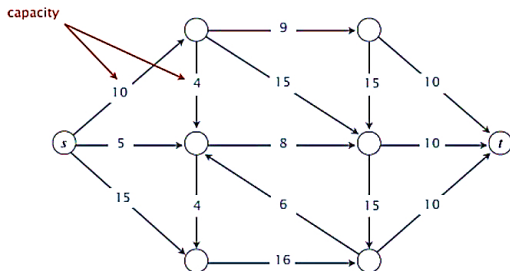


# A Flow network

A **flow network** is a tuple  $G = (V, E, s, t, c)$ .

- Diagraph  $(V, E)$  with **source**  $s \in V$  and **sink**  $t \in V$ .
- Capacity  $c(e) > 0$  for each  $e \in E$ .

**Intuition.** Material flowing through a transportation network, which originates at source and is sent to sink.



### Definition

An *st-cut* (cut) is a partition  $(A, B)$  of the nodes with  $s \in A$  and  $t \in B$ .

# Minimum-cut problem

## Definition

An ***st-cut*** (***cut***) is a partition  $(A, B)$  of the nodes with  $s \in A$  and  $t \in B$ .

## Definition

Its ***capacity*** is the sum of the capacities of the edges from  $A$  to  $B$ .

$$\text{cap}(A, B) = \sum_{e \text{ out of } A} c(e)$$

# Minimum-cut problem

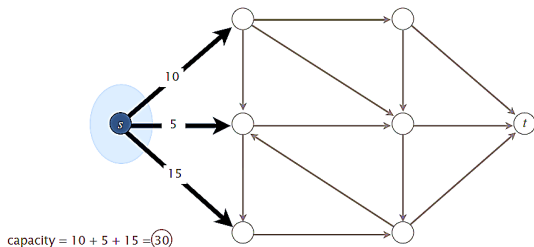
## Definition

An **st-cut (cut)** is a partition  $(A, B)$  of the nodes with  $s \in A$  and  $t \in B$ .

## Definition

Its **capacity** is the sum of the capacities of the edges from  $A$  to  $B$ .

$$\text{cap}(A, B) = \sum_{e \text{ out of } A} c(e)$$



# Minimum-cut problem

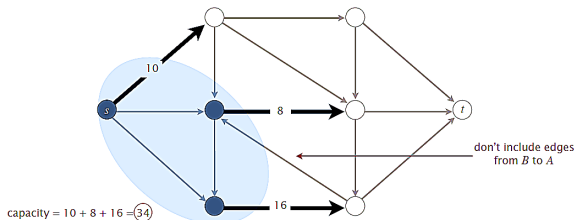
## Definition

An ***st-cut*** (**cut**) is a partition  $(A, B)$  of the nodes with  $s \in A$  and  $t \in B$ .

## Definition

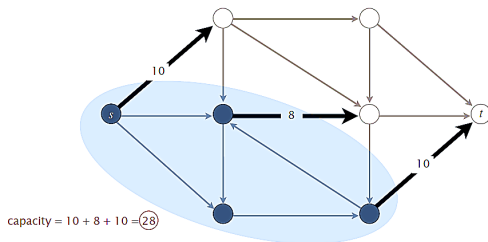
Its **capacity** is the sum of the capacities of the edges from  $A$  to  $B$ .

$$\text{cap}(A, B) = \sum_{e \text{ out of } A} c(e)$$



## Minimum-cut problem

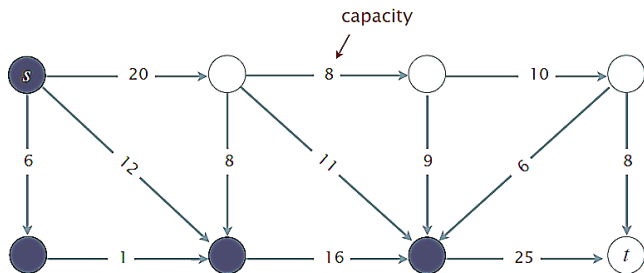
Min-cut problem. Find a cut of minimum capacity.



## Quiz 1

Which is the capacity of the given  $st$ -cut?

- A. 11 ( $20 + 25 - 8 - 11 - 9 - 6$ )
- B. 34 ( $8 + 11 + 9 + 6$ )
- C. 45 ( $20 + 25$ )
- D. 79 ( $20 + 25 + 8 + 11 + 9 + 6$ )

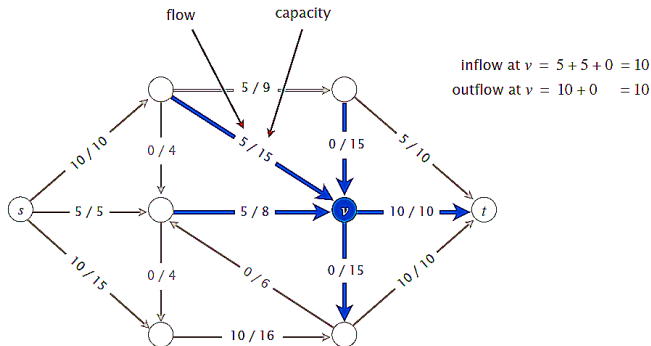


# Maximum-flow problem

## Definition

An *st-flow*(flow)  $f$  is a function that satisfies:

- For each  $e \in E$ :  $0 \leq f(e) \leq c(e)$
- For each  $v \in V - \{s, t\}$ :  $\sum_{e \text{ in to } v} f(e) = \sum_{e \text{ out of } v} f(e)$





# Maximum-flow problem

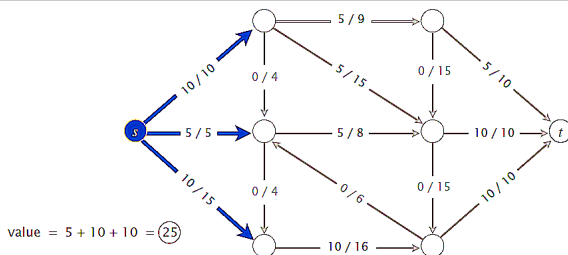
## Definition

An *st-flow*(flow)  $f$  is a function that satisfies:

- For each  $e \in E$ :  $0 \leq f(e) \leq c(e)$
- For each  $v \in V - \{s, t\}$ :  $\sum_{e \text{ in to } v} f(e) = \sum_{e \text{ out of } v} f(e)$

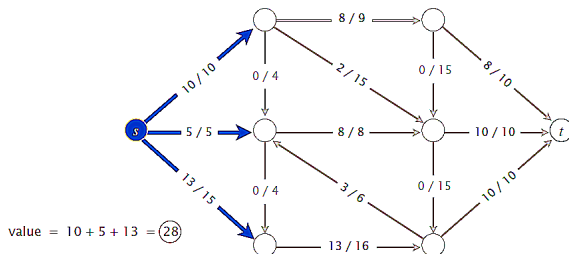
## Definition

The *value* of a flow  $f$  is:  $val(f) = \sum_{e \text{ out of } s} f(e) - \sum_{e \text{ in to } s} f(e)$



# Maximum-flow problem

Max-flow problem. Find a flow of maximum value.



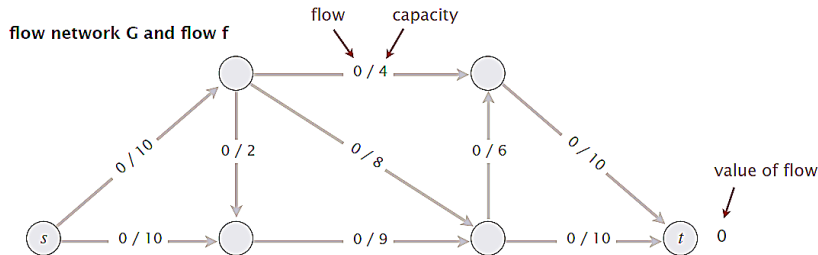
## Ford-Fulkerson Algorithm

---

# Toward a max-flow algorithm

## Greedy algorithm.

- Start with  $f(e) = 0$  for each edge  $e \in E$ .
- Find an  $s \rightsquigarrow t$  path  $P$  where each edge has  $f(e) < c(e)$ .
- Augment flow along path  $P$ .
- Repeat until you get stuck.

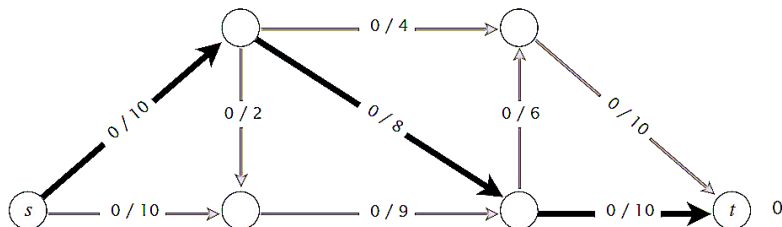


## Toward a max-flow algorithm

### Greedy algorithm.

- Start with  $f(e) = 0$  for each edge  $e \in E$ .
- Find an  $s \rightsquigarrow t$  path  $P$  where each edge has  $f(e) < c(e)$ .
- Augment flow along path  $P$ .
- Repeat until you get stuck.

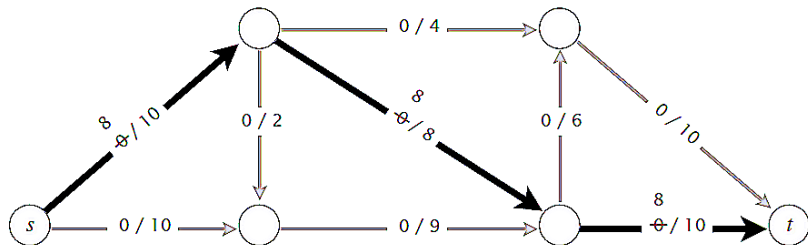
flow network  $G$  and flow  $f$



## Toward a max-flow algorithm

### Greedy algorithm.

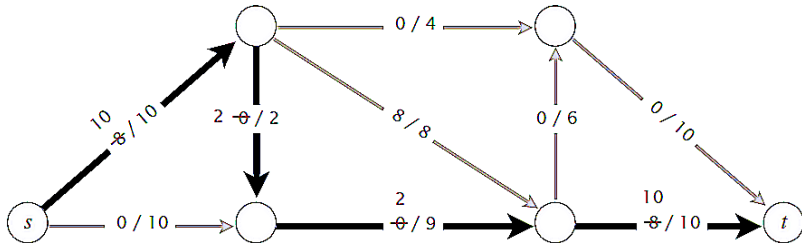
- Start with  $f(e) = 0$  for each edge  $e \in E$ .
- Find an  $s \rightsquigarrow t$  path  $P$  where each edge has  $f(e) < c(e)$ .
- Augment flow along path  $P$ .
- Repeat until you get stuck.



## Toward a max-flow algorithm

### Greedy algorithm.

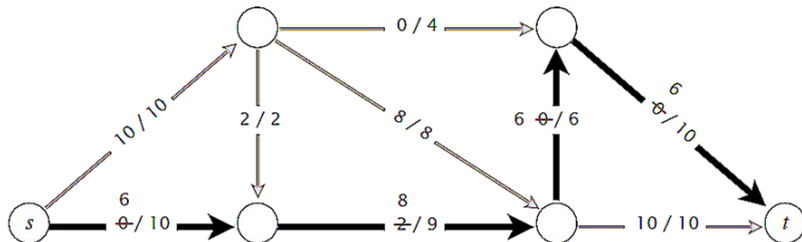
- Start with  $f(e) = 0$  for each edge  $e \in E$ .
- Find an  $s \rightsquigarrow t$  path  $P$  where each edge has  $f(e) < c(e)$ .
- Augment flow along path  $P$ .
- Repeat until you get stuck.



## Toward a max-flow algorithm

### Greedy algorithm.

- Start with  $f(e) = 0$  for each edge  $e \in E$ .
- Find an  $s \rightsquigarrow t$  path  $P$  where each edge has  $f(e) < c(e)$ .
- Augment flow along path  $P$ .
- Repeat until you get stuck.

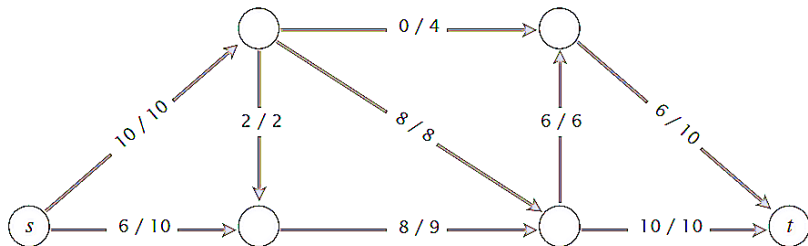




## Toward a max-flow algorithm

### Greedy algorithm.

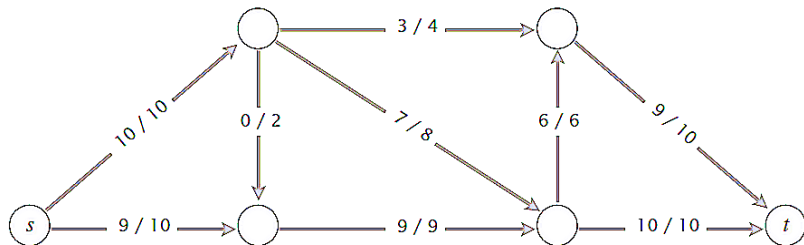
- Start with  $f(e) = 0$  for each edge  $e \in E$ .
- Find an  $s \rightsquigarrow t$  path  $P$  where each edge has  $f(e) < c(e)$ .
- Augment flow along path  $P$ .
- Repeat until you get stuck.



## Toward a max-flow algorithm

### Greedy algorithm.

- Start with  $f(e) = 0$  for each edge  $e \in E$ .
- Find an  $s \rightsquigarrow t$  path  $P$  where each edge has  $f(e) < c(e)$ .
- Augment flow along path  $P$ .
- Repeat until you get stuck.



## Why the greedy algorithm fails

Q. Why does the greedy algorithm fail?

## Why the greedy algorithm fails

Q. Why does the greedy algorithm fail?

A. Once greedy algorithm increases flow on an edge, it never decreases it.

## Why the greedy algorithm fails

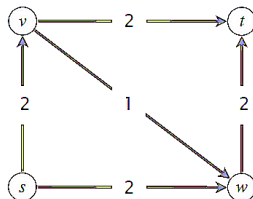
Q. Why does the greedy algorithm fail?

A. Once greedy algorithm increases flow on an edge, it never decreases it.

Ex. Consider flow network  $G$ .

- The unique max flow has  $f^*(v, w) = 0$ .
- Greedy algorithm could choose  $s \rightarrow v \rightarrow w \rightarrow t$  as first augmenting path.

flow network  $G$



# Why the greedy algorithm fails

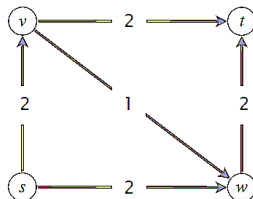
Q. Why does the greedy algorithm fail?

A. Once greedy algorithm increases flow on an edge, it never decreases it.

Ex. Consider flow network  $G$ .

- The unique max flow has  $f^*(v, w) = 0$ .
- Greedy algorithm could choose  $s \rightarrow v \rightarrow w \rightarrow t$  as first augmenting path.

flow network  $G$



Bottom line. Need some mechanism to **undo** a bad decision.

# Residual network

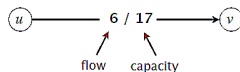
Original edge  $e = (u, v) \in E$ .

- Flow  $f(e)$ .
- Capacity  $c(e)$

Reverse edge  $e^{\text{reverse}} = (v, u)$

- Undo flow sent.

original flow network G

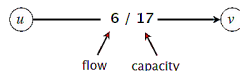


# Residual network

Original edge  $e = (u, v) \in E$ .

- Flow  $f(e)$ .
- Capacity  $c(e)$

original flow network  $G$



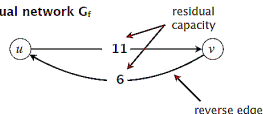
Reverse edge  $e^{\text{reverse}} = (v, u)$

- Undo flow sent.

Residual capacity

$$c_f(e) = \begin{cases} c(e) - f(e) & \text{if } e \in E \\ f(e) & \text{if } e^{\text{reverse}} \in E \end{cases}$$

residual network  $G_f$



Residual network  $G_f = (V, E_f, s, t, c_f)$

- $E_f = \{e : f(e) < c(e)\} \cup \{e^{\text{reverse}} : f(e) > 0\}$ .
- Key property:  $f'$  is a flow in  $G_f$  iff  $f + f'$  is a flow in  $G$



### Definition

An **augmenting path** is a simple  $s \rightsquigarrow t$  path in the residual network  $G_f$ .

### Definition

An **augmenting path** is a simple  $s \rightsquigarrow t$  path in the residual network  $G_f$ .

### Definition

The **bottleneck capacity** of an augmenting path  $P$  is the minimum residual capacity of any edge in  $P$ .

## Augmenting path

**Key Property.** Let  $f$  be a flow and let  $P$  be an augmenting path in  $G_f$ . After calling  $f' \leftarrow \text{Augment}(f, c, P)$ , the resulting  $f'$  is a flow and  $\text{val}(f') = \text{val}(f) + \text{bottleneck}(G_f, P)$ .

## Augmenting path

**Key Property.** Let  $f$  be a flow and let  $P$  be an augmenting path in  $G_f$ . After calling  $f' \leftarrow \text{Augment}(f, c, P)$ , the resulting  $f'$  is a flow and  $\text{val}(f') = \text{val}(f) + \text{bottleneck}(G_f, P)$ .

```
Augment( $f, c, P$ )
```

```
 $\delta \leftarrow$  bottleneck capacity of augmenting path  $P$ ;
```

```
for each edge  $e \in P$  do
```

```
    if ( $e \in E$ ) then  $f(e) \leftarrow f(e) + \delta$ ;
```

```
    else
```

```
         $f(e^{\text{reverse}}) \leftarrow f(e^{\text{reverse}}) - \delta$ 
```

```
    end
```

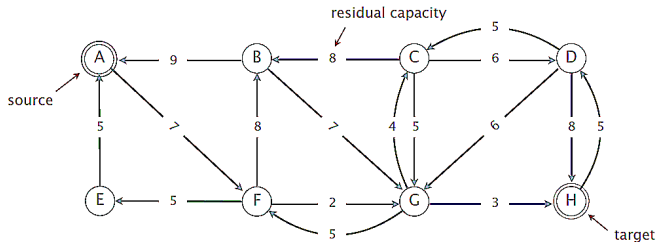
```
end
```

```
Return  $f$ ;
```

## Network flow: quiz 2

Which is the augmenting path of highest bottleneck capacity?

1.  $A \rightarrow F \rightarrow G \rightarrow H$
2.  $A \rightarrow B \rightarrow C \rightarrow D \rightarrow H$
3.  $A \rightarrow F \rightarrow B \rightarrow G \rightarrow H$
4.  $A \rightarrow F \rightarrow B \rightarrow G \rightarrow C \rightarrow D \rightarrow H$



# Ford–Fulkerson algorithm

Ford–Fulkerson augmenting path algorithm.

- Start with  $f(e) = 0$  for each edge  $e \in E$ .
- Find an  $s \rightsquigarrow t$  path  $P$  in the residual network  $G_f$ .
- Augment flow along path  $P$ .
- Repeat until you get stuck.

# Ford–Fulkerson algorithm

Ford–Fulkerson augmenting path algorithm.

- Start with  $f(e) = 0$  for each edge  $e \in E$ .
- Find an  $s \rightsquigarrow t$  path  $P$  in the residual network  $G_f$ .
- Augment flow along path  $P$ .
- Repeat until you get stuck.

```
Ford–Fulkerson( $G$ )
```

```
  for each edge  $e \in E$  do
```

```
     $f(e) \leftarrow 0$ 
```

```
  end
```

```
   $G_f \leftarrow$  residual network of  $G$  with respect to flow  $f$ ;
```

```
  while there exists an  $s \rightsquigarrow t$  path  $P$  in  $G_f$  do
```

```
     $f \leftarrow \text{Augment}(f, c, P)$ ;
```

```
     $\text{Update}(G_f)$ ;
```

```
  end
```

```
  Return  $f$ ;
```

## Max-Flow Min-Cut Theorem

---



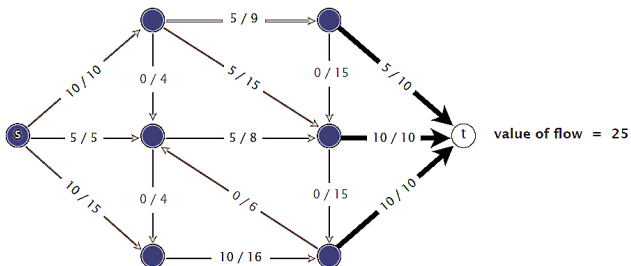
# Relationship between flows and cuts

## Lemma

Let  $f$  be any flow and let  $(A, B)$  be any cut. Then, the value of the flow  $f$  equals the net flow across the cut  $(A, B)$ .

$$\text{val}(f) = \sum_{e \text{ out of } A} f(e) - \sum_{e \text{ in to } A} f(e)$$

net flow across cut =  $5 + 10 + 10 = 25$



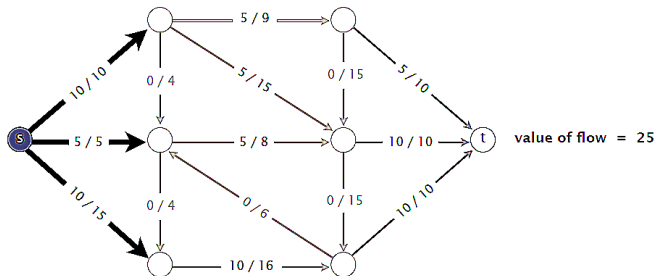
# Relationship between flows and cuts

## Lemma

Let  $f$  be any flow and let  $(A, B)$  be any cut. Then, the value of the flow  $f$  equals the net flow across the cut  $(A, B)$ .

$$\text{val}(f) = \sum_{e \text{ out of } A} f(e) - \sum_{e \text{ in to } A} f(e)$$

$$\text{net flow across cut} = 10 + 5 + 10 = 25$$



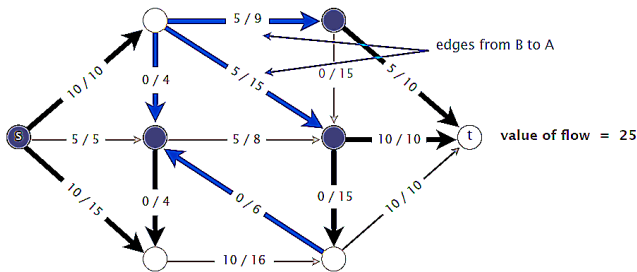
# Relationship between flows and cuts

## Lemma

Let  $f$  be any flow and let  $(A, B)$  be any cut. Then, the value of the flow  $f$  equals the net flow across the cut  $(A, B)$ .

$$\text{val}(f) = \sum_{e \text{ out of } A} f(e) - \sum_{e \text{ in to } A} f(e)$$

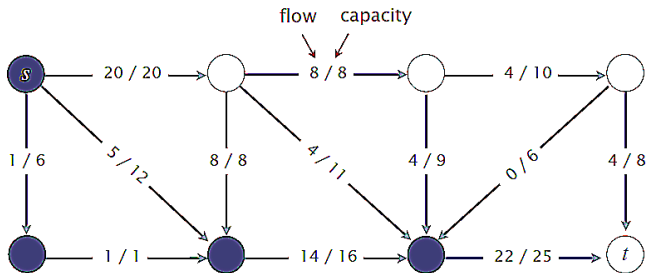
$$\text{net flow across cut} = (10 + 10 + 5 + 10 + 0 + 0) - (5 + 5 + 0 + 0) = 25$$



## Network flow: quiz 3

Which is the net flow across the given cut?

1. 11 ( $20 + 25 - 8 - 11 - 9 - 6$ )
2. 26 ( $20 + 22 - 8 - 4 - 4$ )
3. 42 ( $20 + 22$ )
4. 45 ( $20 + 25$ )



### Lemma

Let  $f$  be any flow and let  $(A, B)$  be any cut. Then, the value of the flow  $f$  equals the net flow across the cut  $(A, B)$ .

$$\text{val}(f) = \sum_{\text{out of } A} f(e) - \sum_{e \text{ in to } A} f(e)$$

## Relationship between flows and cuts

### Lemma

Let  $f$  be any flow and let  $(A, B)$  be any cut. Then, the value of the flow  $f$  equals the net flow across the cut  $(A, B)$ .

$$\text{val}(f) = \sum_{e \text{ out of } A} f(e) - \sum_{e \text{ in to } A} f(e)$$

*Proof.*

$$\begin{aligned} \text{val}(f) &= \sum_{e \text{ out of } s} f(e) - \sum_{e \text{ in to } s} f(e) \\ &= \sum_{v \in A} \left( \sum_{e \text{ out of } v} f(e) - \sum_{e \text{ in to } v} f(e) \right) \\ &= \sum_{e \text{ out of } A} f(e) - \sum_{e \text{ in to } A} f(e). \end{aligned}$$

## Relationship between flows and cuts

### Theorem

*Weak Duality* Let  $f$  be any flow and  $(A, B)$  be any cut. Then,  $val(f) \leq cap(A, B)$ .

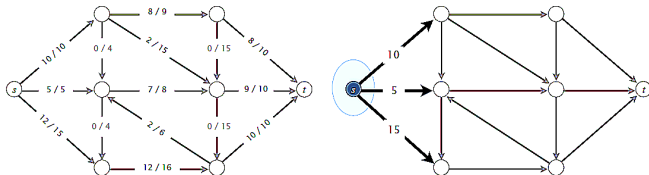
# Relationship between flows and cuts

## Theorem

*Weak Duality* Let  $f$  be any flow and  $(A, B)$  be any cut. Then,  $val(f) \leq cap(A, B)$ .

*Proof.*

$$\begin{aligned} val(f) &= \sum_{e \text{ out of } A} f(e) - \sum_{e \text{ in to } A} f(e) \\ &\leq \sum_{e \text{ out of } A} f(e) \\ &\leq \sum_{e \text{ out of } A} c(e) \\ &= cap(A, B) \end{aligned}$$





### Corollary

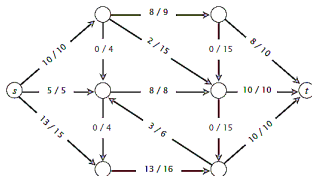
Let  $f$  be a flow and let  $(A, B)$  be any cut. If  $\text{val}(f) = \text{cap}(A, B)$ , then  $f$  is a max flow and  $(A, B)$  is a min cut.

## Corollary

Let  $f$  be a flow and let  $(A, B)$  be any cut. If  $\text{val}(f) = \text{cap}(A, B)$ , then  $f$  is a max flow and  $(A, B)$  is a min cut.

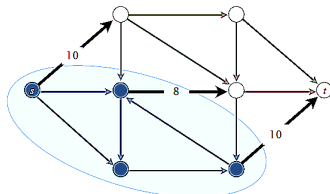
*Proof.*

- For any flow  $f'$  :  $\text{val}(f') \leq \text{cap}(A, B) = \text{val}(f)$ .
- For any cut  $(A', B')$  :  $\text{cap}(A', B') \geq \text{val}(f) = \text{cap}(A, B)$



value of flow = 28

=



capacity of cut = 28

# Max-flow min-cut theorem

## Max-Flow Min-Cut Theorem

Value of a max flow = capacity of a min cut.

### MAXIMAL FLOW THROUGH A NETWORK

L. R. FORD, JR. AND D. R. FULKERSON

**Introduction.** The problem discussed in this paper was formulated by T. Harris as follows:

"Consider a rail network connecting two cities by way of a number of intermediate cities, where each link of the network has a number assigned to it representing its capacity. Assuming a steady state condition, find a maximal flow from one given city to the other."

### ON THE MAX FLOW MIN CUT THEOREM OF NETWORKS

G. B. Dantzig  
D. R. Fulkerson

P-826

April 15, 1955

### A Note on the Maximum Flow Through a Network\*

P. ELIAS†, A. FEINSTEIN‡, AND C. E. SHANNON§

*Summary*—This note discusses the problem of maximizing the rate of flow from one terminal to another, through a network which consists of a number of branches, each of which has a limited capacity. The main result is a theorem: The maximum possible flow from left to right through a network is equal to the minimum value among all simple cut-sets. This theorem is applied to solve a more general problem, in which a number of input nodes and a number of output nodes are used.

from one terminal to the other in the original network passes through at least one branch in the cut-set. In the network above, some examples of cut-sets are  $(d, e, f)$ , and  $(b, e, e, g, h)$ ,  $(d, g, h, i)$ . By a *simple cut-set* we will mean a cut-set such that if any branch is omitted it is no longer a cut-set. Thus  $(d, e, f)$  and  $(b, e, e, g, h)$  are simple cut-sets while  $(d, e, h, i)$  is not. When a simple cut set is

### Max-Flow Min-Cut Theorem

Value of a max flow = capacity of a min cut.

## Max-Flow Min-Cut Theorem

Value of a max flow = capacity of a min cut.

## Augmenting path theorem

A flow  $f$  is a max flow iff no augmenting paths.

# Max-flow min-cut theorem

## Max-Flow Min-Cut Theorem

Value of a max flow = capacity of a min cut.

## Augmenting path theorem

A flow  $f$  is a max flow iff no augmenting paths.

*Proof.* The following three conditions are equivalent for any flow  $f$ :

## Max-Flow Min-Cut Theorem

Value of a max flow = capacity of a min cut.

## Augmenting path theorem

A flow  $f$  is a max flow iff no augmenting paths.

*Proof.* The following three conditions are equivalent for any flow  $f$ :

- i. There exists a cut  $(A, B)$  such that  $\text{cap}(A, B) = \text{val}(f)$ .
- ii.  $f$  is a max flow.
- iii. There is no augmenting path with respect to  $f$ .

## Max-Flow Min-Cut Theorem

Value of a max flow = capacity of a min cut.

## Augmenting path theorem

A flow  $f$  is a max flow iff no augmenting paths.

*Proof.* The following three conditions are equivalent for any flow  $f$ :

- i. There exists a cut  $(A, B)$  such that  $\text{cap}(A, B) = \text{val}(f)$ .
- ii.  $f$  is a max flow.
- iii. There is no augmenting path with respect to  $f$ .

$[i \Rightarrow ii]$



## Max-Flow Min-Cut Theorem

Value of a max flow = capacity of a min cut.

## Augmenting path theorem

A flow  $f$  is a max flow iff no augmenting paths.

*Proof.* The following three conditions are equivalent for any flow  $f$ :

- i. There exists a cut  $(A, B)$  such that  $\text{cap}(A, B) = \text{val}(f)$ .
- ii.  $f$  is a max flow.
- iii. There is no augmenting path with respect to  $f$ .

$[i \Rightarrow ii]$  This is the weak duality corollary.

# Max-flow min-cut theorem

## Max-Flow Min-Cut Theorem

Value of a max flow = capacity of a min cut.

## Augmenting path theorem

A flow  $f$  is a max flow iff no augmenting paths.

*Proof.* The following three conditions are equivalent for any flow  $f$ :

- i. There exists a cut  $(A, B)$  such that  $\text{cap}(A, B) = \text{val}(f)$ .
- ii.  $f$  is a max flow.
- iii. There is no augmenting path with respect to  $f$ .

[ii  $\Rightarrow$  iii]

# Max-flow min-cut theorem

## Max-Flow Min-Cut Theorem

Value of a max flow = capacity of a min cut.

## Augmenting path theorem

A flow  $f$  is a max flow iff no augmenting paths.

*Proof.* The following three conditions are equivalent for any flow  $f$ :

- i. There exists a cut  $(A, B)$  such that  $\text{cap}(A, B) = \text{val}(f)$ .
- ii.  $f$  is a max flow.
- iii. There is no augmenting path with respect to  $f$ .

[ii  $\Rightarrow$  iii]      We prove contrapositive:  $\neg \text{iii} \Rightarrow \neg \text{ii}$ .

# Max-flow min-cut theorem

## Max-Flow Min-Cut Theorem

Value of a max flow = capacity of a min cut.

## Augmenting path theorem

A flow  $f$  is a max flow iff no augmenting paths.

*Proof.* The following three conditions are equivalent for any flow  $f$ :

- i. There exists a cut  $(A, B)$  such that  $\text{cap}(A, B) = \text{val}(f)$ .
- ii.  $f$  is a max flow.
- iii. There is no augmenting path with respect to  $f$ .

[ii  $\Rightarrow$  iii] We prove contrapositive:  $\neg \text{iii} \Rightarrow \neg \text{ii}$ .

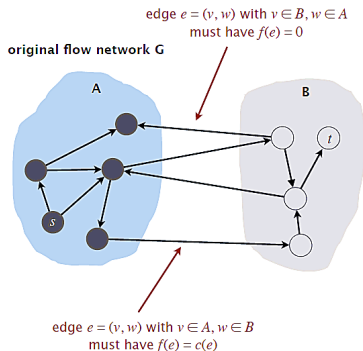
- Suppose that there is an augmenting path with respect to  $f$ .
- Can improve flow  $f$  by sending flow along this path.
- Thus,  $f$  is not a max flow.

# Max-flow min-cut theorem

[ iii  $\Rightarrow$  i ]

- Let  $f$  be a flow with no augmenting paths.
- Let  $A$  be set of nodes reachable from  $s$  in residual network  $G_f$ .
- By definition of  $A$ :  $s \in A$ .
- By definition of flow  $f$ :  $t \notin A$ .

$$\begin{aligned} \text{val}(f) &= \sum_{e \text{ out of } A} f(e) - \sum_{e \text{ in to } A} f(e) \\ &= \sum_{e \text{ out of } A} c(e) - 0 \\ &= \text{cap}(A, B) \end{aligned}$$



## Capacity-Scaling Algorithm

---

## Analysis of Ford–Fulkerson algorithm (when capacities are integral)

**Assumption.** Every edge capacity  $c(e)$  is an integer between 1 and  $C$ .

## Analysis of Ford–Fulkerson algorithm (when capacities are integral)

**Assumption.** Every edge capacity  $c(e)$  is an integer between 1 and  $C$ .

**Integrality invariant.** Throughout Ford-Fulkerson, every edge flow  $f(e)$  and residual capacity  $c_f(e)$  is an integer.



## Analysis of Ford–Fulkerson algorithm (when capacities are integral)

*Assumption.* Every edge capacity  $c(e)$  is an integer between 1 and  $C$ .

*Integrality invariant.* Throughout Ford-Fulkerson, every edge flow  $f(e)$  and residual capacity  $c_f(e)$  is an integer.

*Proof.* By induction on the number of augmenting paths.

## Analysis of Ford–Fulkerson algorithm (when capacities are integral)

**Assumption.** Every edge capacity  $c(e)$  is an integer between 1 and  $C$ .

**Integrality invariant.** Throughout Ford–Fulkerson, every edge flow  $f(e)$  and residual capacity  $c_f(e)$  is an integer.

**Proof.** By induction on the number of augmenting paths.

### Theorem

*Ford–Fulkerson terminates after at most  $\text{val}(f^*) \leq nC$  augmenting paths, where  $f^*$  is a max flow.*

## Analysis of Ford–Fulkerson algorithm (when capacities are integral)

**Assumption.** Every edge capacity  $c(e)$  is an integer between 1 and  $C$ .

**Integrality invariant.** Throughout Ford–Fulkerson, every edge flow  $f(e)$  and residual capacity  $c_f(e)$  is an integer.

**Proof.** By induction on the number of augmenting paths.

### Theorem

*Ford–Fulkerson terminates after at most  $\text{val}(f^*) \leq nC$  augmenting paths, where  $f^*$  is a max flow.*

**Proof.** Each augmentation increases the value of the flow by at least 1.

### Corollary

*The running time of Ford–Fulkerson is  $O(mnC)$ .*

## Analysis of Ford–Fulkerson algorithm (when capacities are integral)

### Corollary

*The running time of Ford–Fulkerson is  $O(mnC)$ .*

*Proof.* Can use either BFS or DFS to find an augmenting path in  $O(m)$  time.

## Analysis of Ford–Fulkerson algorithm (when capacities are integral)

### Corollary

*The running time of Ford–Fulkerson is  $O(mnC)$ .*

*Proof.* Can use either BFS or DFS to find an augmenting path in  $O(m)$  time.

### Integrality Theorem

There exists an integral max flow  $f^*$

## Analysis of Ford–Fulkerson algorithm (when capacities are integral)

### Corollary

*The running time of Ford–Fulkerson is  $O(mnC)$ .*

*Proof.* Can use either BFS or DFS to find an augmenting path in  $O(m)$  time.

### Integrality Theorem

There exists an integral max flow  $f^*$

*Proof.* Since Ford–Fulkerson terminates, theorem follows from integrality invariant.

Q. Is generic Ford–Fulkerson algorithm poly-time in input size?



## Ford–Fulkerson: exponential example

Q. Is generic Ford–Fulkerson algorithm poly-time in input size?

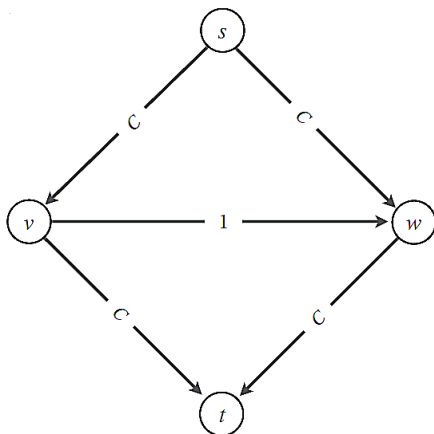
A. No. If max capacity is  $C$ , then algorithm can take  $\geq C$  iterations.

## Ford–Fulkerson: exponential example

Q. Is generic Ford–Fulkerson algorithm poly-time in input size?

A. No. If max capacity is  $C$ , then algorithm can take  $\geq C$  iterations.

- $s \rightarrow v \rightarrow w \rightarrow t$
- $s \rightarrow w \rightarrow v \rightarrow t$
- $s \rightarrow v \rightarrow w \rightarrow t$
- $s \rightarrow w \rightarrow v \rightarrow t$
- ...
- $s \rightarrow v \rightarrow w \rightarrow t$
- $s \rightarrow w \rightarrow v \rightarrow t$



The Ford–Fulkerson algorithm is guaranteed to terminate if the edge capacities are ...

- A. Rational numbers.
- B. Real numbers.
- C. Both A and B.
- D. Neither A nor B.

## Choosing good augmenting paths

Use care when selecting augmenting paths.

- Some choices lead to exponential algorithms.
- Clever choices lead to polynomial algorithms.

## Choosing good augmenting paths

Use care when selecting augmenting paths.

- Some choices lead to exponential algorithms.
- Clever choices lead to polynomial algorithms.

**Pathology.** When edge capacities can be irrational, no guarantee that Ford–Fulkerson terminates (or converges to a maximum flow)!

## Choosing good augmenting paths

Use care when selecting augmenting paths.

- Some choices lead to exponential algorithms.
- Clever choices lead to polynomial algorithms.

**Pathology.** When edge capacities can be irrational, no guarantee that Ford–Fulkerson terminates (or converges to a maximum flow)!

**Goal.** Choose augmenting paths so that:

- Can find augmenting paths efficiently.
- Few iterations.

## Choosing good augmenting paths

Choose augmenting paths with:

## Choosing good augmenting paths

Choose augmenting paths with:

- Max bottleneck capacity( “fattest” ).



## Choosing good augmenting paths

Choose augmenting paths with:

- Max bottleneck capacity( “fattest” ).
- Sufficiently large bottleneck capacity.

# Choosing good augmenting paths

## Choose augmenting paths with:

- Max bottleneck capacity ("fattest").
- Sufficiently large bottleneck capacity.
- Fewest edges.

### Theoretical Improvements in Algorithmic Efficiency for Network Flow Problems

JACK EDMONDS

*University of Waterloo, Waterloo, Ontario, Canada*

AND

RICHARD M. KARP

*University of California, Berkeley, California*

**ABSTRACT.** This paper presents new algorithms for the maximum flow problem, the Hitchcock transportation problem, and the general minimum-cost flow problem. Upper bounds on the numbers of steps in these algorithms are derived, and are shown to compare favorably with upper bounds on the numbers of steps required by earlier algorithms.

**Edmonds-Karp 1972 (USA)**

Dokl. Akad. Nauk SSSR  
Tom 194 (1970), No. 4

Soviet Math. Dokl.  
Vol. 11 (1970), No. 5

### ALGORITHM FOR SOLUTION OF A PROBLEM OF MAXIMUM FLOW IN A NETWORK WITH POWER ESTIMATION

UDC 518.5

E. A. DINIC

Different variants of the formulation of the problem of maximal stationary flow in a network and its many applications are given in [1]. There also is given an algorithm solving the problem in the case where the initial data are integers (or, what is equivalent, commensurable). In the general case this algorithm requires preliminary rounding off of the initial data, i.e. only an approximate solution of the problem is possible. In this connection the rapidity of convergence of the algorithm is inversely proportional to the relative precision.

**Dinitz 1970 (Soviet Union)**

## Capacity-scaling algorithm

**Overview.** Choosing augmented paths with **large** bottleneck capacity.

## Capacity-scaling algorithm

**Overview.** Choosing augmented paths with **large** bottleneck capacity.

- Maintain scaling parameter  $\Delta$ .

# Capacity-scaling algorithm

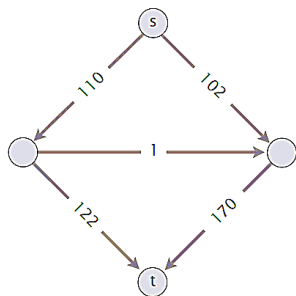
**Overview.** Choosing augmented paths with **large** bottleneck capacity.

- Maintain scaling parameter  $\Delta$ .
- Let  $G_f(\Delta)$  be the part of the residual network containing only those edges with capacity  $\geq \Delta$ .

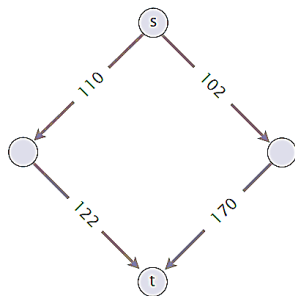
## Capacity-scaling algorithm

**Overview.** Choosing augmented paths with **large** bottleneck capacity.

- Maintain scaling parameter  $\Delta$ .
- Let  $G_f(\Delta)$  be the part of the residual network containing only those edges with capacity  $\geq \Delta$ .
- Any augmenting path in  $G_f(\Delta)$  has bottleneck capacity  $\geq \Delta$ .



$G_f$



$G_f(\Delta), \Delta = 100$

Capacity-Scaling( $G$ )

**for** each edge  $e \in E$  **do**

$f(e) \leftarrow 0$

**end**

$\Delta \leftarrow$  largest power of 2  $\leq C$ ;

**while**  $\Delta \geq 1$  **do**

$G_f(\Delta) \leftarrow \Delta$ -residual network of  $G$  with respect to flow  $f$ ;

**while** there exists an  $s \rightsquigarrow t$  path  $P$  in  $G_f(\Delta)$  **do**

$f \leftarrow \text{Augment}(f, c, P)$ ;

        Update( $G_\Delta(f)$ );

**end**

$\Delta = \Delta/2$ ;

**end**

Return  $f$ ;

*Assumption.* All edge capacities are integers between 1 and  $C$ .



## Capacity-scaling algorithm: proof of correctness

**Assumption.** All edge capacities are integers between 1 and  $C$ .

**Invariant.** The scaling parameter  $\Delta$  is a power of 2.

## Capacity-scaling algorithm: proof of correctness

*Assumption.* All edge capacities are integers between 1 and  $C$ .

*Invariant.* The scaling parameter  $\Delta$  is a power of 2.

*Proof.* Initially a power of 2; each phase divides  $\Delta$  by exactly 2.

## Capacity-scaling algorithm: proof of correctness

*Assumption.* All edge capacities are integers between 1 and  $C$ .

*Invariant.* The scaling parameter  $\Delta$  is a power of 2.

*Proof.* Initially a power of 2; each phase divides  $\Delta$  by exactly 2.

*Integrity invariant.* Throughout the algorithm, every edge flow  $f(e)$  and residual capacity  $c_f(e)$  is an integer.

## Capacity-scaling algorithm: proof of correctness

*Assumption.* All edge capacities are integers between 1 and  $C$ .

*Invariant.* The scaling parameter  $\Delta$  is a power of 2.

*Proof.* Initially a power of 2; each phase divides  $\Delta$  by exactly 2.

*Integrality invariant.* Throughout the algorithm, every edge flow  $f(e)$  and residual capacity  $c_f(e)$  is an integer.

*Proof.* Same as for generic Ford–Fulkerson.

### Theorem

*If capacity-scaling algorithm terminates, then  $f$  is a max flow.*

### Theorem

*If capacity-scaling algorithm terminates, then  $f$  is a max flow.*

*Proof.*

### Theorem

*If capacity-scaling algorithm terminates, then  $f$  is a max flow.*

*Proof.*

- By integrality invariant, when  $\Delta = 1 \Rightarrow G_f(\Delta) = G_f$
- Upon termination of  $\Delta = 1$  phase, there are no augmenting paths.
- Result follows augmenting path theorem.

## Capacity-scaling algorithm: analysis of running time

### Lemma 1

There are  $1 + \lfloor \log_2 C \rfloor$  scaling phases.

### Lemma 2

There are  $\leq 2m$  augmentations per scaling phase.

### Lemma 3

Let  $f$  be the flow at the end of a  $\Delta$ -scaling phase.  
Then, the max-flow value  $\leq \text{val}(f) + m\Delta$ .



### Theorem

*The capacity-scaling algorithm takes  $O(m^2 \log C)$  time.*

### Theorem

*The capacity-scaling algorithm takes  $O(m^2 \log C)$  time.*

*Proof.*

### Theorem

*The capacity-scaling algorithm takes  $O(m^2 \log C)$  time.*

*Proof.*

- Lemma 1+ Lemma 2  $\Rightarrow O(m \log C)$  augmentations.
- Finding an augmenting path takes  $O(m)$  time.

### Lemma 1

There are  $1 + \lfloor \log_2 C \rfloor$  scaling phases.

### Lemma 1

There are  $1 + \lfloor \log_2 C \rfloor$  scaling phases.

*Proof.* Initially  $C/2 < \Delta \leq C$ ;  $\Delta$  decreases by a factor of 2 in each iteration.

### Lemma 2

There are  $\leq 2m$  augmentations per scaling phase.

### Lemma 2

There are  $\leq 2m$  augmentations per scaling phase.

*Proof.*

- Let  $f$  be the flow at the beginning of a  $\Delta$ -scaling phase.
- Lemma 3  $\Rightarrow$  max-flow value  $\leq \text{val}(f) + m(2\Delta)$ .
- Each augmentation in a  $\Delta$ -phase increases  $\text{val}(f)$  by at least  $\Delta$ .

### Lemma 3

Let  $f$  be the flow at the end of a  $\Delta$ -scaling phase.

Then, the max-flow value  $\leq \text{val}(f) + m\Delta$ .



### Lemma 3

Let  $f$  be the flow at the end of a  $\Delta$ -scaling phase.  
Then, the max-flow value  $\leq \text{val}(f) + m\Delta$ .

*Proof.*

### Lemma 3

Let  $f$  be the flow at the end of a  $\Delta$ -scaling phase.

Then, the max-flow value  $\leq \text{val}(f) + m\Delta$ .

*Proof.*

- We show there exists a cut  $(A, B)$  such that  $\text{cap}(A, B) \leq \text{val}(f) + m\Delta$ .
- Choose  $A$  to be the set of nodes reachable from  $s$  in  $G_f(\Delta)$ .
- By definition of  $A : s \in A$ .
- By definition of flow  $f : t \notin A$ .

# Capacity-scaling algorithm: analysis of running time

## Lemma 3

Let  $f$  be the flow at the end of a  $\Delta$ -scaling phase.

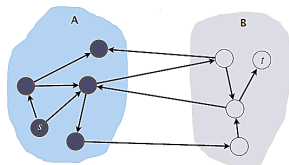
Then, the max-flow value  $\leq \text{val}(f) + m\Delta$ .

*Proof.*

- We show there exists a cut  $(A, B)$  such that  $\text{cap}(A, B) \leq \text{val}(f) + m\Delta$ .
- Choose  $A$  to be the set of nodes reachable from  $s$  in  $G_f(\Delta)$ .
- By definition of  $A : s \in A$ .
- By definition of flow  $f : t \notin A$ .

$$\begin{aligned}\text{val}(f) &= \sum_{e \text{ out of } A} f(e) - \sum_{e \text{ in to } A} f(e) \\ &\geq \sum_{e \text{ out of } A} (c(e) - \Delta) - \sum_{e \text{ in to } A} \Delta \\ &\geq \sum_{e \text{ out of } A} c(e) - \sum_{e \text{ out of } A} \Delta - \sum_{e \text{ in to } A} \Delta \\ &\geq \text{cap}(A, B) - m\Delta\end{aligned}$$

original flow network



## Shortest Augmenting Paths

---

## Shortest augmenting path

Q. How to choose next augmenting path in Ford–Fulkerson?

## Shortest augmenting path

Q. How to choose next augmenting path in Ford–Fulkerson?

A. Pick one that uses the **fewest** edges.

## Shortest augmenting path

Q. How to choose next augmenting path in Ford–Fulkerson?

A. Pick one that uses the **fewest** edges.

Shortest-Augmenting-Path( $G$ )

**for** *each edge*  $e \in E$  **do**

$f(e) \leftarrow 0$

**end**

$G_f \leftarrow$  residual network of  $G$  with respect to flow  $f$ ;

**while** *there exists an*  $s \rightsquigarrow t$  *path in*  $G_f$  **do**

$P \leftarrow \text{BFS}((G_f));$

$f \leftarrow \text{Augment}(f, c, P);$

  Update( $G_f$ );

**end**

Return  $f$ ;

### Lemma 1

The length of a shortest augmenting path never decreases.



## Shortest augmenting path: overview of analysis

### Lemma 1

The length of a shortest augmenting path never decreases.

### Lemma 2

After at most  $m$  shortest-path augmentations, the length of a shortest augmenting path strictly increases.

# Shortest augmenting path: overview of analysis

## Lemma 1

The length of a shortest augmenting path never decreases.

## Lemma 2

After at most  $m$  shortest-path augmentations, the length of a shortest augmenting path strictly increases.

## Theorem

*The shortest-augmenting-path algorithm takes  $O(m^2n)$  time.*

# Shortest augmenting path: overview of analysis

## Lemma 1

The length of a shortest augmenting path never decreases.

## Lemma 2

After at most  $m$  shortest-path augmentations, the length of a shortest augmenting path strictly increases.

## Theorem

*The shortest-augmenting-path algorithm takes  $O(m^2n)$  time.*

*Proof.*

## Shortest augmenting path: overview of analysis

### Lemma 1

The length of a shortest augmenting path never decreases.

### Lemma 2

After at most  $m$  shortest-path augmentations, the length of a shortest augmenting path strictly increases.

### Theorem

*The shortest-augmenting-path algorithm takes  $O(m^2n)$  time.*

*Proof.*

- $O(m)$  time to find a shortest augmenting path via BFS.

# Shortest augmenting path: overview of analysis

## Lemma 1

The length of a shortest augmenting path never decreases.

## Lemma 2

After at most  $m$  shortest-path augmentations, the length of a shortest augmenting path strictly increases.

## Theorem

*The shortest-augmenting-path algorithm takes  $O(m^2n)$  time.*

*Proof.*

- $O(m)$  time to find a shortest augmenting path via BFS.
- There are  $\leq mn$  augmentations

# Shortest augmenting path: overview of analysis

## Lemma 1

The length of a shortest augmenting path never decreases.

## Lemma 2

After at most  $m$  shortest-path augmentations, the length of a shortest augmenting path strictly increases.

## Theorem

*The shortest-augmenting-path algorithm takes  $O(m^2n)$  time.*

*Proof.*

- $O(m)$  time to find a shortest augmenting path via BFS.
- There are  $\leq mn$  augmentations
  - at most  $m$  augmenting paths of length  $k$

# Shortest augmenting path: overview of analysis

## Lemma 1

The length of a shortest augmenting path never decreases.

## Lemma 2

After at most  $m$  shortest-path augmentations, the length of a shortest augmenting path strictly increases.

## Theorem

*The shortest-augmenting-path algorithm takes  $O(m^2n)$  time.*

*Proof.*

- $O(m)$  time to find a shortest augmenting path via BFS.
- There are  $\leq mn$  augmentations
  - at most  $m$  augmenting paths of length  $k \leftarrow \text{Lemma 1} + \text{Lemma 2}$
  - at most  $n - 1$  different lengths

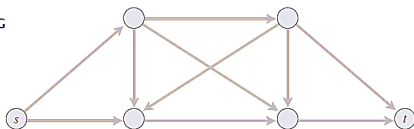
## Shortest augmenting path: analysis

### Definition

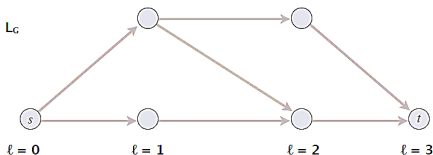
Given a **digraph**  $G = (V, E)$  with source  $s$ , its **level graph** is defined by:

- $\ell(v)$  = number of edges in shortest  $s \rightsquigarrow v$  path.
- $L_G = (V, E_G)$  is the subgraph of  $G$  that contains only those edges  $(v, w) \in E$  with  $\ell(w) = \ell(v) + 1$ .

graph  $G$



level graph  $L_G$

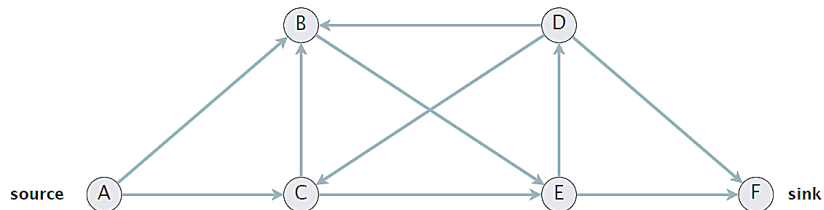




## Network flow: quiz 5

Which edges are in the level graph of the following digraph?

- A.  $D \rightarrow F$
- B.  $E \rightarrow F$
- C. Both A and B.
- D. Neither A nor B.



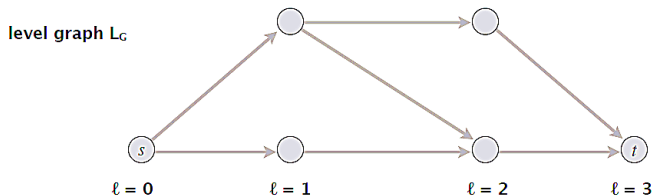
## Shortest augmenting path: analysis

### Definition

Given a **digraph**  $G = (V, E)$  with source  $s$ , its **level graph** is defined by:

- $\ell(v) =$  number of edges in shortest  $s \rightsquigarrow v$  path.
- $L_G = (V, E_G)$  is the subgraph of  $G$  that contains only those edges  $(v, w) \in E$  with  $\ell(w) = \ell(v) + 1$ .

**Key property.**  $P$  is a shortest  $s \rightsquigarrow v$  path in  $G$  iff  $P$  is an  $s \rightsquigarrow v$  path in  $L_G$ .



### Lemma 1

The length of a shortest augmenting path never decreases.

### Lemma 1

The length of a shortest augmenting path never decreases.

*Proof.*

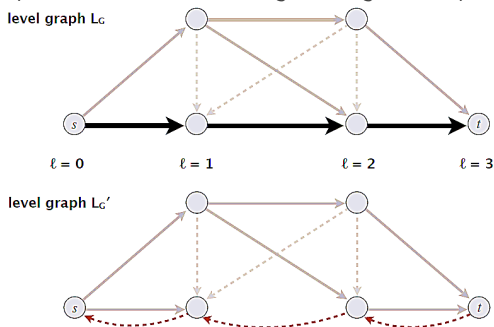
# Shortest augmenting path: analysis

## Lemma 1

The length of a shortest augmenting path never decreases.

*Proof.*

- Let  $f$  and  $f'$  be flow before and after a shortest-path augmentation.
- Let  $L_G$  and  $L_{G'}$  be level graphs of  $G_f$  and  $G_{f'}$ .
- Only back edges added to  $G_{f'}$   
(any  $s \rightsquigarrow t$  path that uses a back edge is longer than previous length)



### Lemma 2

After at most  $m$  shortest-path augmentations, the length of a shortest augmenting path strictly increases.

### Lemma 2

After at most  $m$  shortest-path augmentations, the length of a shortest augmenting path strictly increases.

*Proof.*

# Shortest augmenting path: analysis

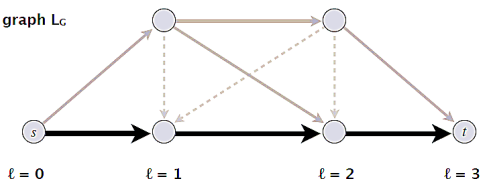
## Lemma 2

After at most  $m$  shortest-path augmentations, the length of a shortest augmenting path strictly increases.

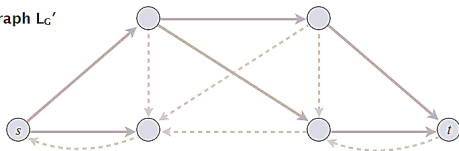
*Proof.*

- At least one (bottleneck) edge is deleted from  $L_G$  per augmentation.
- No new edge added to  $L_G$  until shortest path length strictly increases.

level graph  $L_G$



level graph  $L_{G'}$





## Shortest augmenting path: review of analysis

### Lemma 1

The length of a shortest augmenting path never decreases.

### Lemma 2

After at most  $m$  shortest-path augmentations, the length of a shortest augmenting path strictly increases.

### Theorem

*The shortest-augmenting-path algorithm takes  $O(m^2n)$  time.*

## Shortest augmenting path: improving the running time

Note.  $\Theta(mn)$  augmentations necessary for some flow networks.

## Shortest augmenting path: improving the running time

Note.  $\Theta(mn)$  augmentations necessary for some flow networks.

- Try to decrease time per augmentation instead.
- Simple idea  $\Rightarrow O(mn^2)$  [Dinitz 1970]

# Shortest augmenting path: improving the running time

Note.  $\Theta(mn)$  augmentations necessary for some flow networks.

- Try to decrease time per augmentation instead.
- Simple idea  $\Rightarrow O(mn^2)$  [Dinitz 1970]
- Dynamic trees  $\Rightarrow O(mn \log n)$  [Sleator–Tarjan 1983]

## A Data Structure for Dynamic Trees

DANIEL D. SLEATOR AND ROBERT ENDRE TARJAN

*Bell Laboratories, Murray Hill, New Jersey 07974*

Received May 8, 1982; revised October 18, 1982

A data structure is proposed to maintain a collection of vertex-disjoint trees under a sequence of two kinds of operations: a *link* operation that combines two trees into one by adding an edge, and a *cut* operation that divides one tree into two by deleting an edge. Each operation requires  $O(\log n)$  time. Using this data structure, new fast algorithms are obtained for the following problems:

- (1) Computing nearest common ancestors.
- (2) Solving various network flow problems including finding maximum flows, blocking flows, and acyclic flows.
- (3) Computing certain kinds of constrained minimum spanning trees.
- (4) Implementing the network simplex algorithm for minimum-cost flows.

The most significant application is (2); an  $O(mn \log n)$ -time algorithm is obtained to find a maximum flow in a network of  $n$  vertices and  $m$  edges, beating by a factor of  $\log n$  the fastest algorithm previously known for sparse graphs.

## Dinitz' Algorithm

---

Two types of augmentations.

- **Normal**: length of shortest path does not change.
- **Special**: length of shortest path strictly increases.

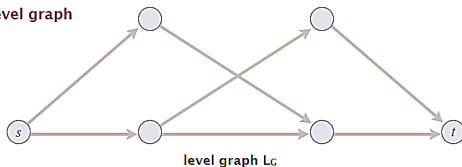
## Two types of augmentations.

- **Normal**: length of shortest path does not change.
- **Special**: length of shortest path strictly increases.

## Phase of normal augmentations.

- Construct level graph  $L_G$ .
- Start at  $s$ , advance along an edge in  $L_G$  until reach  $t$  or get stuck.
- If reach  $t$ , augment flow; update  $L_G$ ; and restart from  $s$ .
- If get stuck, delete node from  $L_G$  and retreat to previous node.

construct level graph



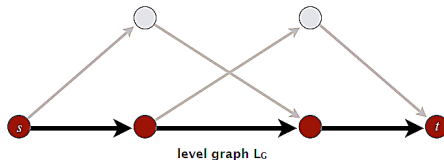
## Two types of augmentations.

- **Normal**: length of shortest path does not change.
- **Special**: length of shortest path strictly increases.

## Phase of normal augmentations.

- Construct level graph  $L_G$ .
- Start at  $s$ , advance along an edge in  $L_G$  until reach  $t$  or get stuck.
- If reach  $t$ , augment flow; update  $L_G$ ; and restart from  $s$ .
- If get stuck, delete node from  $L_G$  and retreat to previous node.

advance





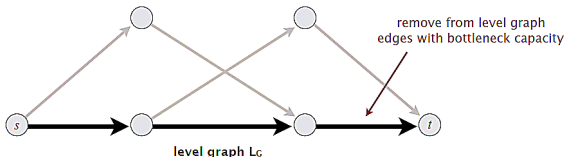
## Two types of augmentations.

- **Normal**: length of shortest path does not change.
- **Special**: length of shortest path strictly increases.

## Phase of normal augmentations.

- Construct level graph  $L_G$ .
- Start at  $s$ , advance along an edge in  $L_G$  until reach  $t$  or get stuck.
- If reach  $t$ , augment flow; update  $L_G$ ; and restart from  $s$ .
- If get stuck, delete node from  $L_G$  and retreat to previous node.

augment



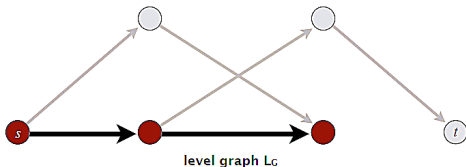
## Two types of augmentations.

- **Normal**: length of shortest path does not change.
- **Special**: length of shortest path strictly increases.

## Phase of normal augmentations.

- Construct level graph  $L_G$ .
- Start at  $s$ , advance along an edge in  $L_G$  until reach  $t$  or get stuck.
- If reach  $t$ , augment flow; update  $L_G$ ; and restart from  $s$ .
- If get stuck, delete node from  $L_G$  and retreat to previous node.

advance



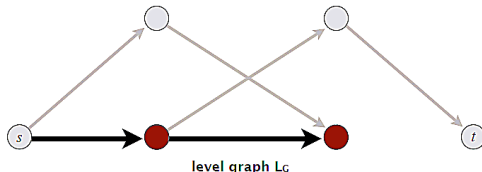
## Two types of augmentations.

- **Normal**: length of shortest path does not change.
- **Special**: length of shortest path strictly increases.

## Phase of normal augmentations.

- Construct level graph  $L_G$ .
- Start at  $s$ , advance along an edge in  $L_G$  until reach  $t$  or get stuck.
- If reach  $t$ , augment flow; update  $L_G$ ; and restart from  $s$ .
- If get stuck, delete node from  $L_G$  and retreat to previous node.

retreat



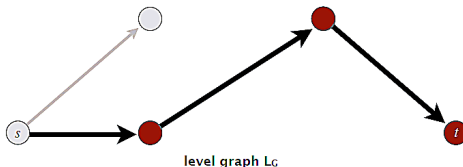
## Two types of augmentations.

- **Normal**: length of shortest path does not change.
- **Special**: length of shortest path strictly increases.

## Phase of normal augmentations.

- Construct level graph  $L_G$ .
- Start at  $s$ , advance along an edge in  $L_G$  until reach  $t$  or get stuck.
- If reach  $t$ , augment flow; update  $L_G$ ; and restart from  $s$ .
- If get stuck, delete node from  $L_G$  and retreat to previous node.

advance

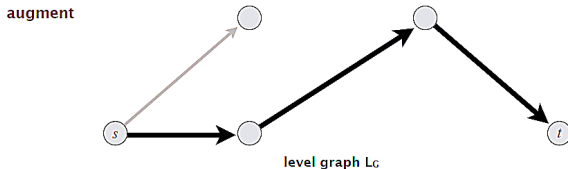


## Two types of augmentations.

- **Normal**: length of shortest path does not change.
- **Special**: length of shortest path strictly increases.

## Phase of normal augmentations.

- Construct level graph  $L_G$ .
- Start at  $s$ , advance along an edge in  $L_G$  until reach  $t$  or get stuck.
- If reach  $t$ , augment flow; update  $L_G$ ; and restart from  $s$ .
- If get stuck, delete node from  $L_G$  and retreat to previous node.



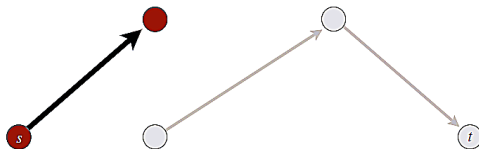
## Two types of augmentations.

- **Normal**: length of shortest path does not change.
- **Special**: length of shortest path strictly increases.

## Phase of normal augmentations.

- Construct level graph  $L_G$ .
- Start at  $s$ , advance along an edge in  $L_G$  until reach  $t$  or get stuck.
- If reach  $t$ , augment flow; update  $L_G$ ; and restart from  $s$ .
- If get stuck, delete node from  $L_G$  and retreat to previous node.

advance



level graph  $L_G$

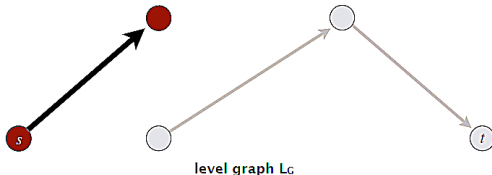
## Two types of augmentations.

- **Normal**: length of shortest path does not change.
- **Special**: length of shortest path strictly increases.

## Phase of normal augmentations.

- Construct level graph  $L_G$ .
- Start at  $s$ , advance along an edge in  $L_G$  until reach  $t$  or get stuck.
- If reach  $t$ , augment flow; update  $L_G$ ; and restart from  $s$ .
- If get stuck, delete node from  $L_G$  and retreat to previous node.

retreat



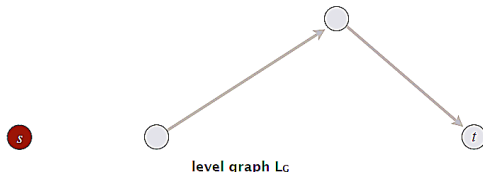
## Two types of augmentations.

- **Normal**: length of shortest path does not change.
- **Special**: length of shortest path strictly increases.

## Phase of normal augmentations.

- Construct level graph  $L_G$ .
- Start at  $s$ , advance along an edge in  $L_G$  until reach  $t$  or get stuck.
- If reach  $t$ , augment flow; update  $L_G$ ; and restart from  $s$ .
- If get stuck, delete node from  $L_G$  and retreat to previous node.

retreat





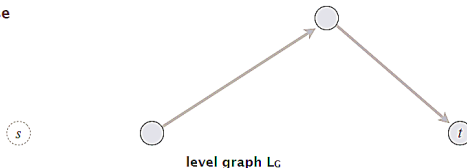
## Two types of augmentations.

- **Normal**: length of shortest path does not change.
- **Special**: length of shortest path strictly increases.

## Phase of normal augmentations.

- Construct level graph  $L_G$ .
- Start at  $s$ , advance along an edge in  $L_G$  until reach  $t$  or get stuck.
- If reach  $t$ , augment flow; update  $L_G$ ; and restart from  $s$ .
- If get stuck, delete node from  $L_G$  and retreat to previous node.

end of phase



## Dinitz' algorithm (as refined by Even and Itai)

Initialize( $G, f$ )

$L_G \leftarrow$  level-graph of  $G_f$

$P \leftarrow \emptyset$

goto Advance( $s$ );

Retreat( $v$ )

**if**  $v = s$  **then** Stop;

**else**

    Delete  $v$  from  $L_G$ ;

    Remove last edge  $(u, v)$   
    from  $P$ ;

**end**

goto Advance( $u$ );

Advance( $v$ )

**if**  $v = t$  **then**

    Augment( $P$ );

    Remove saturated edges  
    from  $L_G$ ;

$P \leftarrow \emptyset$ ;

    goto Advance( $s$ );

**end**

**if** there exists edge  $(v, w) \in L_G$

**then**

        Add edge  $(v, w)$  to  $P$ ;

        goto Advance( $w$ );

**end**

**else**

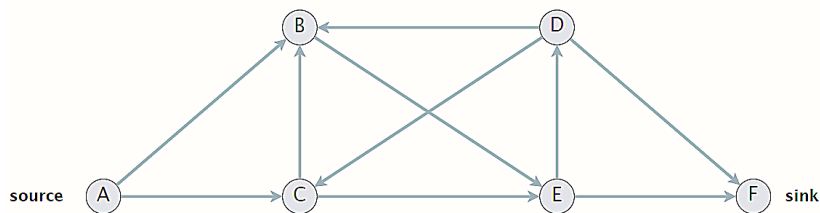
        goto Retreat( $v$ );

**end**

## Network flow: quiz 6

How to compute the level graph  $L_G$  efficiently?

1. Depth-first search.
2. Breadth-first search.
3. Both A and B.
4. Neither A nor B.



### Lemma

*A phase can be implemented to run in  $O(mn)$  time.*

### Lemma

*A phase can be implemented to run in  $O(mn)$  time.*

*Proof.*

- Initialization happens once per phase.

### Lemma

*A phase can be implemented to run in  $O(mn)$  time.*

*Proof.*

- Initialization happens once per phase. using BFS

### Lemma

*A phase can be implemented to run in  $O(mn)$  time.*

*Proof.*

- Initialization happens once per phase. using BFS
- At most  $m$  augmentations per phase.

### Lemma

*A phase can be implemented to run in  $O(mn)$  time.*

*Proof.*

- Initialization happens once per phase. using BFS
- At most  $m$  augmentations per phase.  $\longleftarrow O(mn)$  per phase



### Lemma

*A phase can be implemented to run in  $O(mn)$  time.*

*Proof.*

- Initialization happens once per phase. using BFS
- At most  $m$  augmentations per phase.  $\leftarrow O(mn)$  per phase  
(because an augmentation deletes at least one edge from  $L_G$ )

### Lemma

*A phase can be implemented to run in  $O(mn)$  time.*

*Proof.*

- Initialization happens once per phase. using BFS
- At most  $m$  augmentations per phase.  $\leftarrow O(mn)$  per phase  
(because an augmentation deletes at least one edge from  $L_G$ )
- At most  $n$  retreats per phase.

### Lemma

*A phase can be implemented to run in  $O(mn)$  time.*

*Proof.*

- Initialization happens once per phase. using BFS
- At most  $m$  augmentations per phase.  $\leftarrow O(mn)$  per phase  
(because an augmentation deletes at least one edge from  $L_G$ )
- At most  $n$  retreats per phase.  $\leftarrow O(m + n)$  per phase

### Lemma

*A phase can be implemented to run in  $O(mn)$  time.*

*Proof.*

- Initialization happens once per phase. using BFS
- At most  $m$  augmentations per phase.  $\leftarrow O(mn)$  per phase  
(because an augmentation deletes at least one edge from  $L_G$ )
- At most  $n$  retreats per phase.  $\leftarrow O(m + n)$  per phase  
(because a retreat deletes one node from  $L_G$ )

### Lemma

*A phase can be implemented to run in  $O(mn)$  time.*

*Proof.*

- Initialization happens once per phase. using BFS
- At most  $m$  augmentations per phase.  $\leftarrow O(mn)$  per phase  
(because an augmentation deletes at least one edge from  $L_G$ )
- At most  $n$  retreats per phase.  $\leftarrow O(m + n)$  per phase  
(because a retreat deletes one node from  $L_G$ )
- At most  $mn$  advances per phase.

### Lemma

A phase can be implemented to run in  $O(mn)$  time.

*Proof.*

- Initialization happens once per phase. using BFS
- At most  $m$  augmentations per phase.  $\leftarrow O(mn)$  per phase  
(because an augmentation deletes at least one edge from  $L_G$ )
- At most  $n$  retreats per phase.  $\leftarrow O(m + n)$  per phase  
(because a retreat deletes one node from  $L_G$ )
- At most  $mn$  advances per phase.  $\leftarrow O(mn)$  per phase

### Lemma

*A phase can be implemented to run in  $O(mn)$  time.*

*Proof.*

- Initialization happens once per phase. using BFS
- At most  $m$  augmentations per phase.  $\leftarrow O(mn)$  per phase  
(because an augmentation deletes at least one edge from  $L_G$ )
- At most  $n$  retreats per phase.  $\leftarrow O(m + n)$  per phase  
(because a retreat deletes one node from  $L_G$ )
- At most  $mn$  advances per phase.  $\leftarrow O(mn)$  per phase  
(because at most  $n$  advances before retreat or augmentation)

### Theorem (Dinitz 1970)

*Dinitz' algorithm runs in  $O(mn^2)$  time.*



### Theorem (Dinitz 1970)

*Dinitz' algorithm runs in  $O(mn^2)$  time.*

*Proof.*

### Theorem (Dinitz 1970)

*Dinitz' algorithm runs in  $O(mn^2)$  time.*

*Proof.*

- By Lemma,  $O(mn)$  time per phase.

### Theorem (Dinitz 1970)

*Dinitz' algorithm runs in  $O(mn^2)$  time.*

*Proof.*

- By Lemma,  $O(mn)$  time per phase.
- At most  $n - 1$  phases

### Theorem (Dinitz 1970)

*Dinitz' algorithm runs in  $O(mn^2)$  time.*

*Proof.*

- By Lemma,  $O(mn)$  time per phase.
- At most  $n - 1$  phases (as in shortest-augmenting-path analysis).

## Augmenting-path algorithms: summary

year	method	# augmentations	running time
1955	augmenting path	$nC$	$O(mnC)$
1972	fattest path	$m \log(mC)$	$O(m^2 \log n \log(mC))$
1972	capacity scaling	$m \log C$	$O(m^2 \log C)$
1985	improved capacity scaling	$m \log C$	$O(mn \log C)$
1970	shortest augmenting path	$mn$	$O(m^2 n)$
1970	level graph	$mn$	$O(mn^2)$
1983	dynamic trees	$mn$	$O(mn \log n)$

augmenting-path algorithms with  $m$  edges,  $n$  nodes, and integer capacities between 1 and  $C$

# Maximum-flow algorithms: theory highlights

year	method	worst case	discovered by
1951	simplex	$O(mn^2C)$	Dantzig
1955	augmenting paths	$O(mnC)$	Ford–Fulkerson
1970	shortest augmenting paths	$O(mn^2)$	Edmonds–Karp, Dinitz
1974	blocking flows	$O(n^3)$	Karzanov
1983	dynamic trees	$O(mn \log n)$	Sleator–Tarjan
1985	improved capacity scaling	$O(mn \log C)$	Gabow
1988	push–relabel	$O(mn \log(n^2/m))$	Goldberg–Tarjan
1998	binary blocking flows	$O(m^{3/2} \log(n^2/m) \log C)$	Goldberg–Rao
2013	compact networks	$O(mn)$	Orlin
2014	interior-point methods	$\tilde{O}(mm^{1/2} \log C)$	Lee–Sidford
2016	electrical flows	$\tilde{O}(m^{10/7} C^{1/7})$	Madry
20xx		???	

augmenting-path algorithms with  $m$  edges,  $n$  nodes, and integer capacities between 1 and  $C$