COMP261
Algorithms and Data Structures
2024 Tri 1

Jyoti Sahni
jyoti.sahni@ecs.vuw.ac.nz
Office Hours (COMP261): AM414, Thursday 10:00 – 12:00

Next 3 weeks

- More on graphs and related algorithms

Assessment Items	Weight (%)
1 Assignment – Assignment 3 (released week 7; due week 10)	10
Mid-Term Test 3 (due week 9)	15

Introduction

- Graphs are one of the unifying themes of computer science.
- There are many different structures that can be modelled using graphs:
  - Transportation systems
  - Human interactions in a social network
  - Telecommunications network
- Today's most pressing data challenges centre around relationships, not just tabulating discrete data.
  - Relationships may evolve over time: e.g. social network of a person, his buying pattern, taste of music etc.
- Graphs help model relationships

Introduction

- Graphs as a way to model relationships.
- Nodes / vertices: Objects that make up a graph
- Links / Edges: Relationship

- But modelling graphs is only half the story.

- We also want to process the graphs to reveal insight that isn’t immediately obvious.

Three levels of Graph analytics

Graph query - Graph queries, for the most part, attempt to identify an explicit pattern within the graph database.

What graph queries usually do is find a known subset of the important nodes in the graph haystack.

```
Find me all friends Dan and Kevin share
```

```
Find:
- The set of friends of Dan
- The set of friends of Kevin
- Find the intersection of the two sets
```

Three levels of Graph analytics

Graph algorithm - Graph algorithms, still begins from a declared subsection or a local graph, but in addition to selecting and grouping nodes, it also involves categorizing them or using some other processing technique to sort them out and learn something from them.

```
Find me the shortest path between Dan and Kevin
```

```
Find:
– Of all the paths between Dan and Kevin, the
  shortest path
```

A query may find you the needles in the haystack you knew you were looking for; an algorithm may tell you which needles are most interesting.

Three levels of Graph analytics

Graph analytics - tells us something about the graph in general.

In this case, nearly all nodes in a graph will be inspected as a part of the calculation.

```
Who are the most influential nodes in the network
```

```
Find:
- Of all the nodes in the network, nodes with
  highest  number of followers
```

Graph Analytics

- Three areas of analytics that are at the heart of graph based systems:
- Path Finding
  - Shortest path from point A to point B
  - Average shortest path – average distance between stations.
  - Longest path – which stations are farthest apart
  - Minimum spanning tree
  - The maximum flow that can take place between two nodes in a transportation network – Flow network
- Centrality: Finding which nodes are more important in a network.
- Community Detection (hubs/ hierarchies/ tendencies): identifying common attributes

Next 3 weeks…

Study some complex problems that can be modelled using graphs

- Pathfinding
  - Single source shortest path
  - All pair-shortest paths
  - Transportation flow paths
  - Cycle Finding / Detection / Avoidance
  - Spanning Trees
- Impact of nodes in a network
  - Centrality
  - Page Ranking
- Finding group behaviours
  - Community Detection

We'll cover these

Flow Networks

Flow Network problems

Graph related problems which involves flow rates:

- Road Network: Flow of vehicles
  - How fast can I evacuate a city in emergency
- Distribution network : Network of pipelines
  - Flow of oil / gas / electricity etc.
  - What is the maximum amount that can be transferred over the network from point A to point B.
- Computer network: Flow of packets
  - At what rate can I transfer data from point A to point B.

Flow Network problems

Just as we can model a road map as a directed graph, to find the shortest path from one point to another, we can also interpret a directed graph as a "flow network" – and use it to answer questions about material flows.

- Flow – Intuitively, the flow of material at any point in the system is the rate at which material flows.

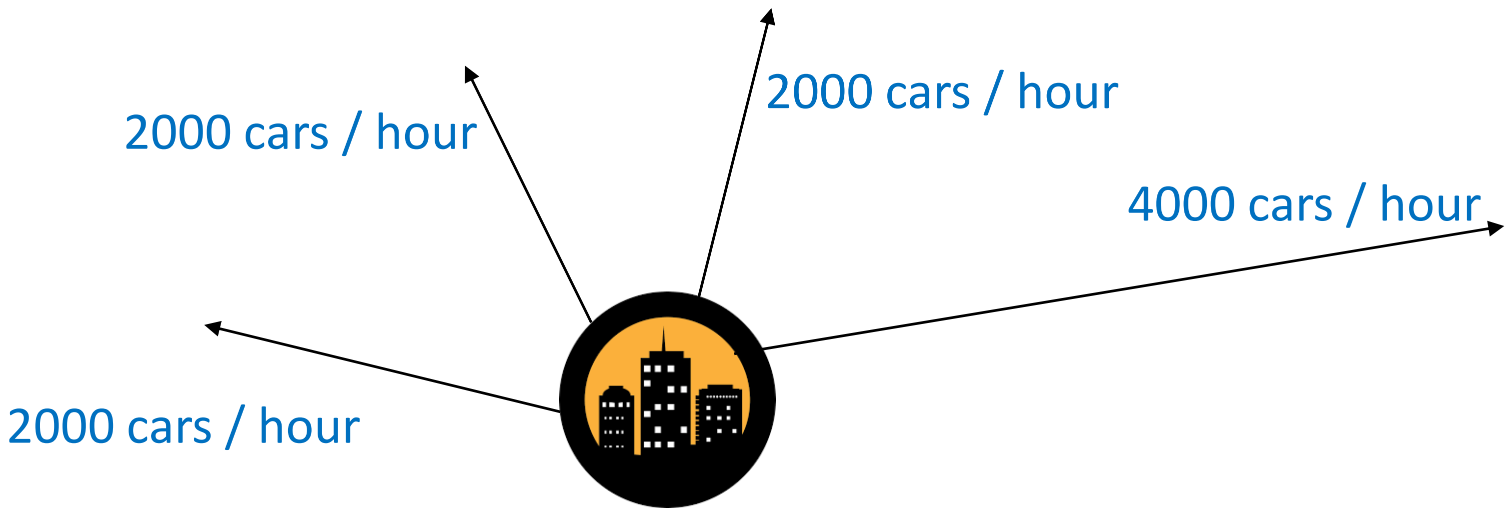
Flow Network problems

How are these problems different from Single source shortest paths involved edges with cost?

- In network flow problems:
  - edges have maximum capacities
  - Source node generates traffic
  - Sink nodes absorb traffic as it arrives
  - Traffic is transmitted across the edges

Example

How fast can we evacuate a city in emergency ?



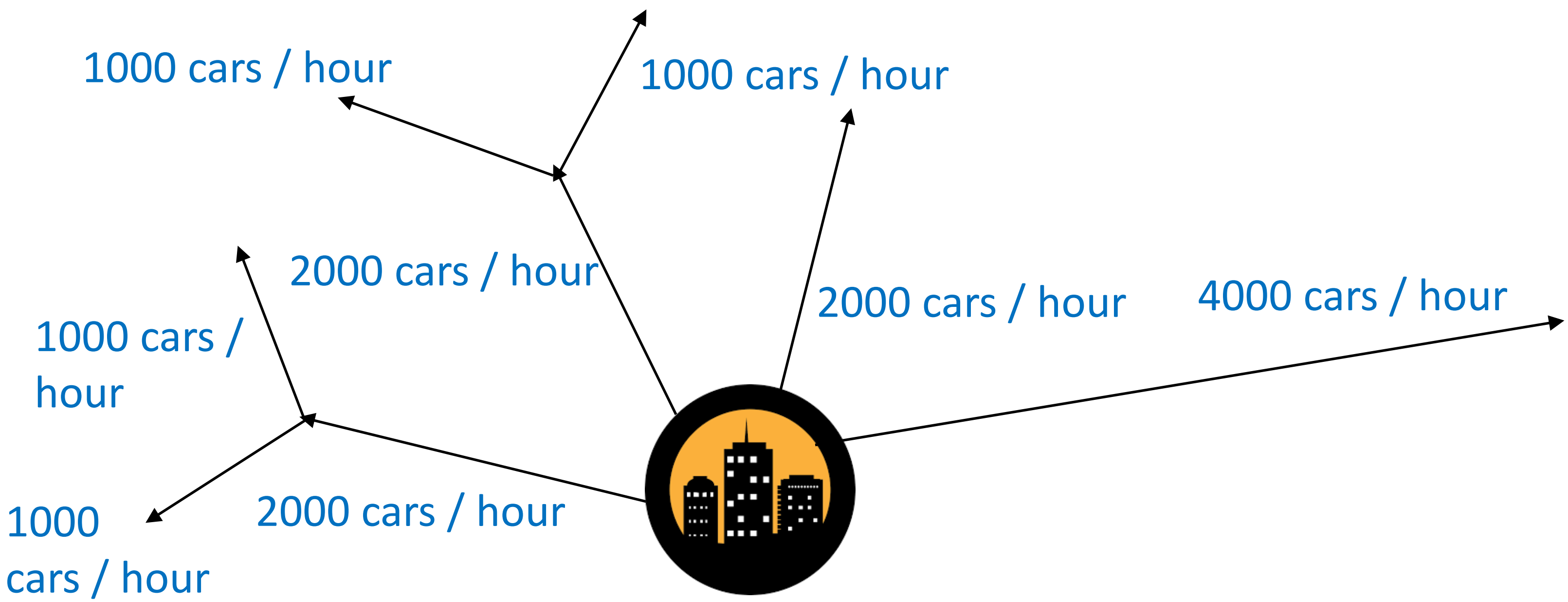
Three smaller roads can handle 2000 cars / hour

One main highway can handle 4000 cars / hour

Example

How fast can we evacuate a city in emergency ?

Two of the three smaller roads split into two halves

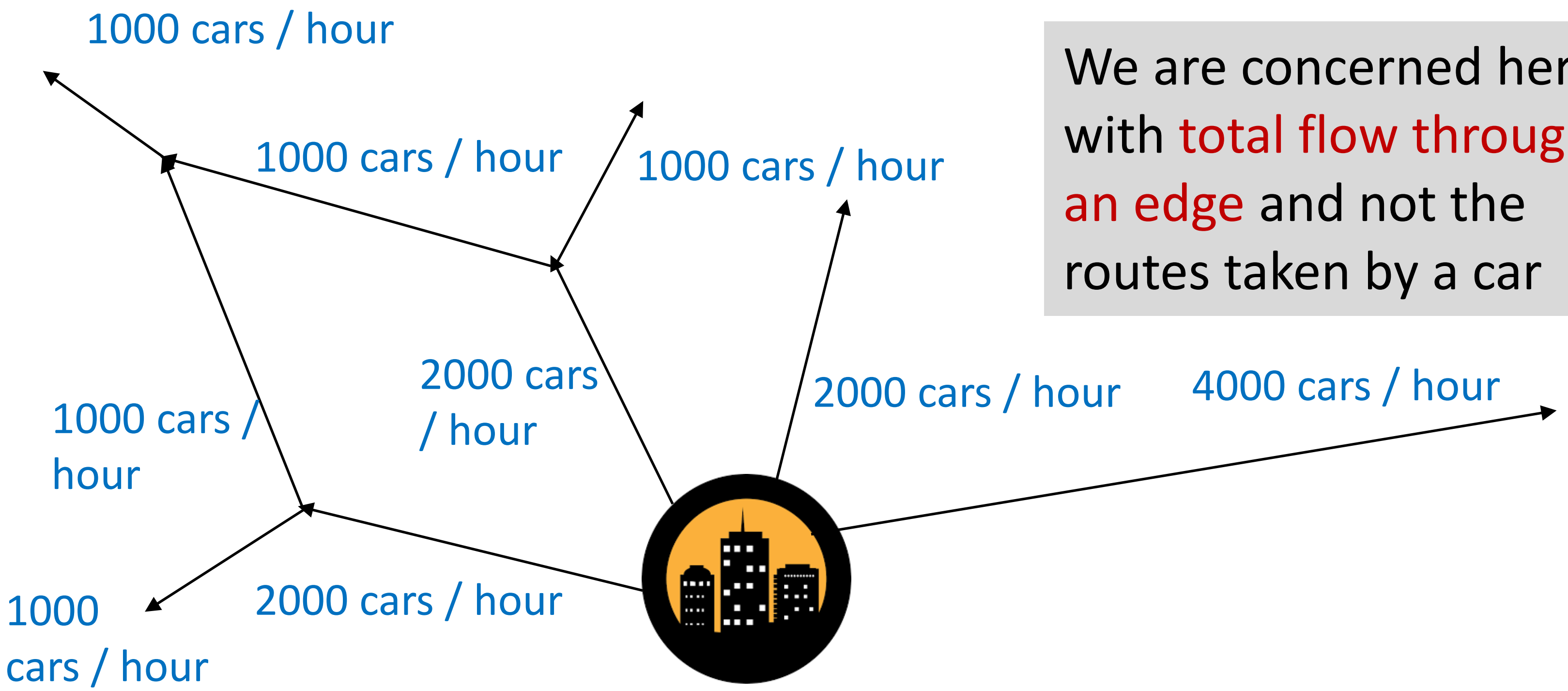


One main highway can handle 4000 cars / hour

Example

How fast can we evacuate a city in emergency ?

We are concerned here with total flow through an edge and not the routes taken by a car

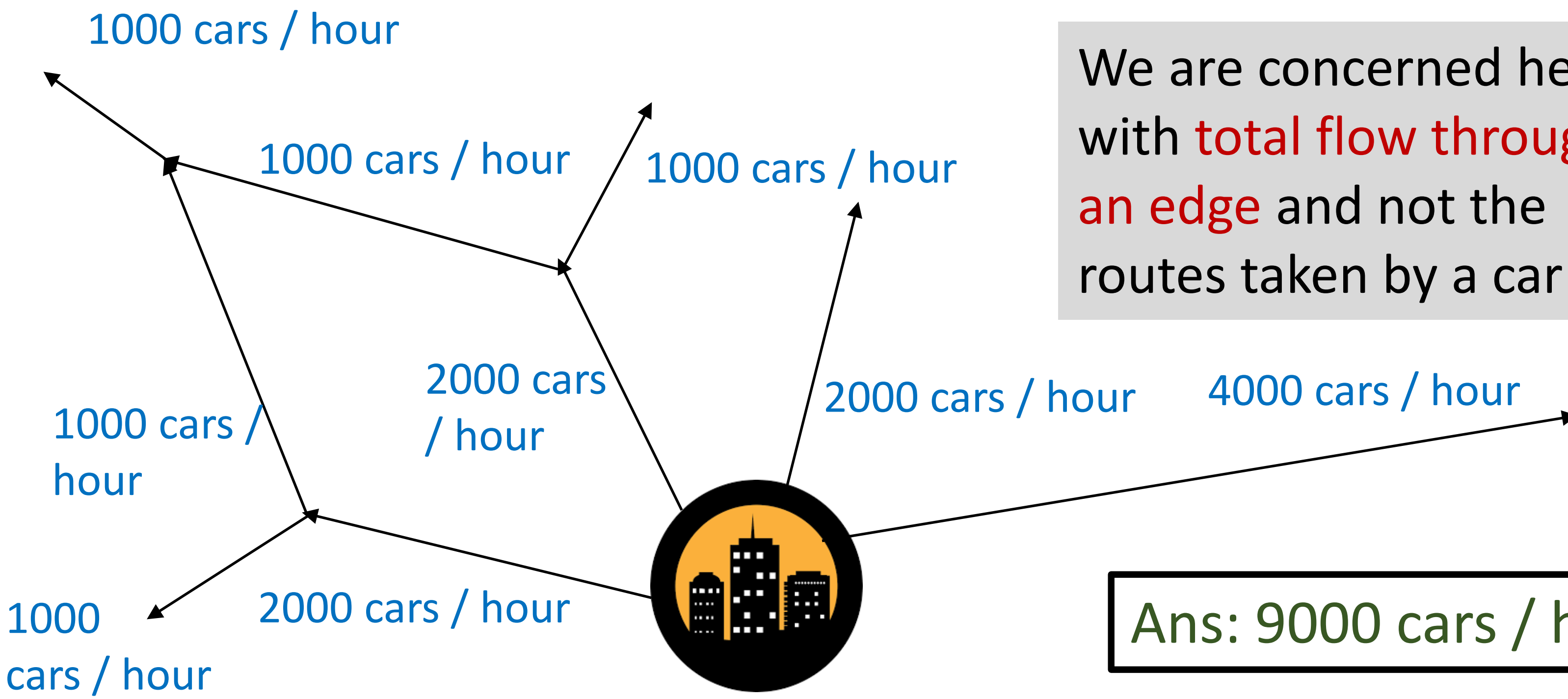


Two halves merge together but still can handle 1000 cars / hour

One main highway can handle 4000 cars / hour

Example

How fast can we evacuate a city in emergency ?



We are concerned here with total flow through an edge and not the routes taken by a car

Ans: 9000 cars / hour

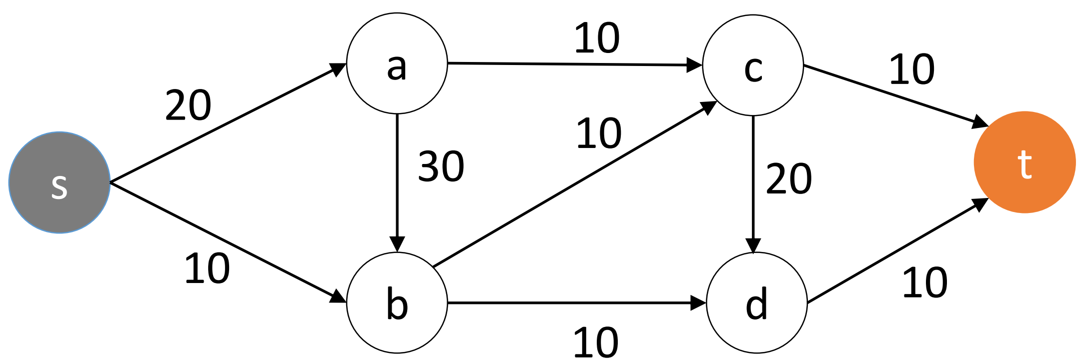
Two halves merge together but still can handle 1000 cars / hour

One main highway can handle 4000 cars / hour

Definition

A flow network is a connected, directed graph $G = (V, E)$ where:

- Each edge, e, has a non-negative, integer capacity c_e.
- One (or more) vertex is labelled as a source $s \in V$.
- One (or more) vertex is labelled as sink t $\in V$.
- No edge enters the source and no edge leaves the sink.

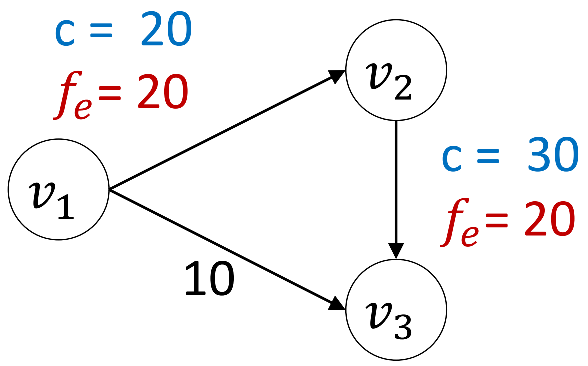


Flow in a flow network

Flow in a flow network concerns with total flow, f_e, through each edge.

- $f_e \equiv$ amount of material carried on the edge $\boldsymbol{e}$.

- (f_e is not the same as capacity)

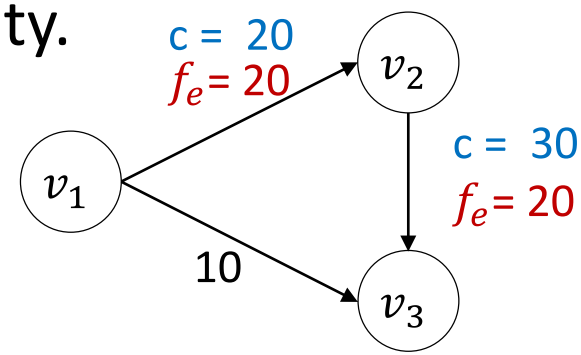


Flow Constraints

Flows have two constraints:

a) Capacity constraint: For each edge, e, the flow at the edge must be less than or equal to its capacity.

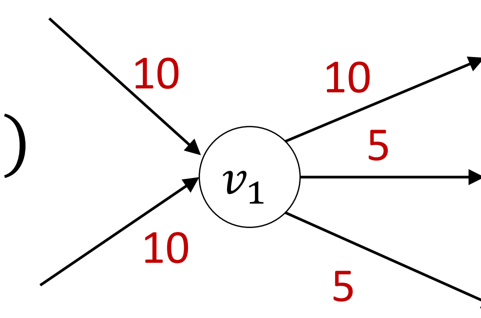
$$0 \leq f_e \leq c_e$$



b) Balance constraint: Rate at which the material enters a vertex must equal the rate at which it leaves the vertex.

- For each node v except s and t, whatever flows in, must flow out; we have:

$$\sum_{e \text{ } into \text{ } v} f(e) = \sum_{e \text{ } out \text{ } of \text{ } v} f(e)$$

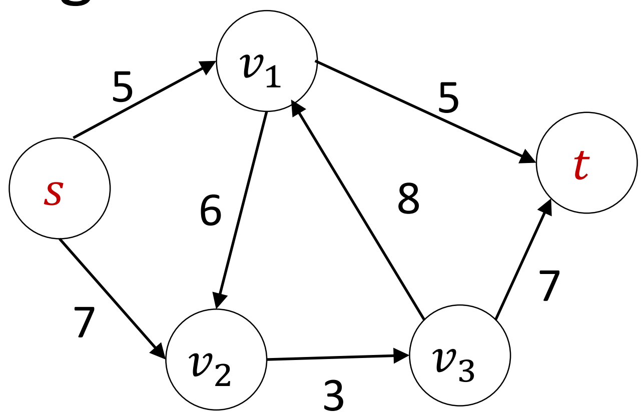


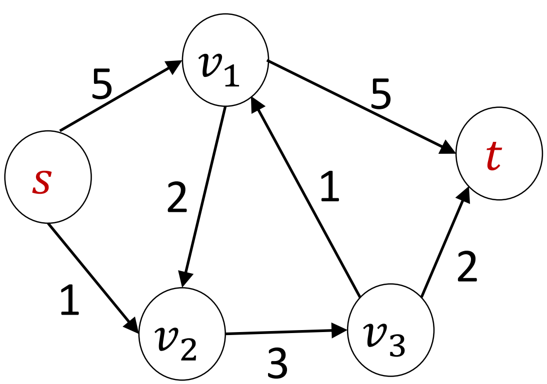
Formal Definition

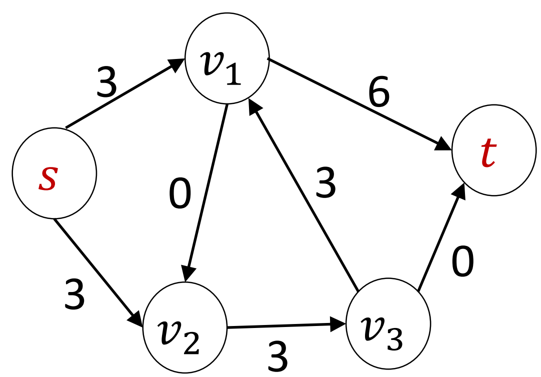
- A flow in a network is an assignment of a real number flow f_e to each edge such that:
  - for all e
$$0 \leq f_e \leq c_e$$
  - For each vertex v except s (source) and t (sink):
$$\sum_{e\ into\ v} f_e = \sum_{e\ out\ of\ v} f_e$$

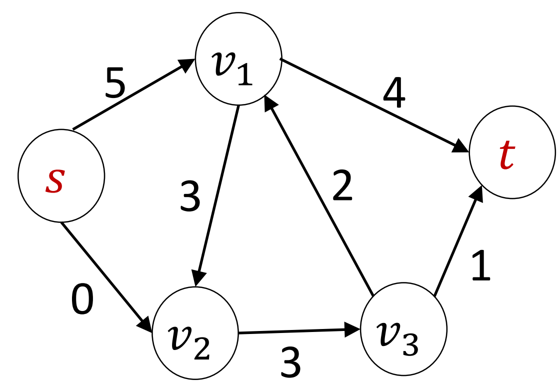
A Quick Recap

Which of the following is a valid flow for the given network ?



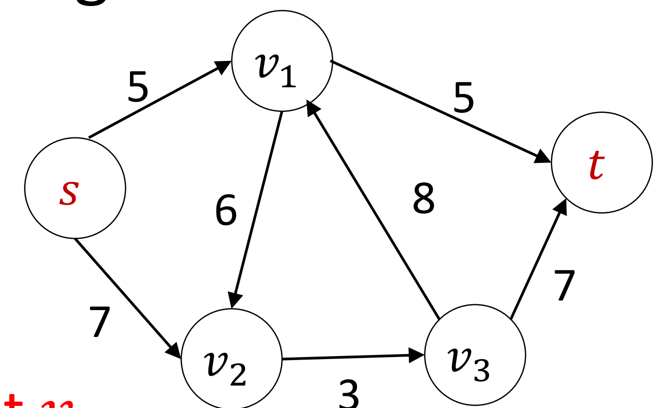




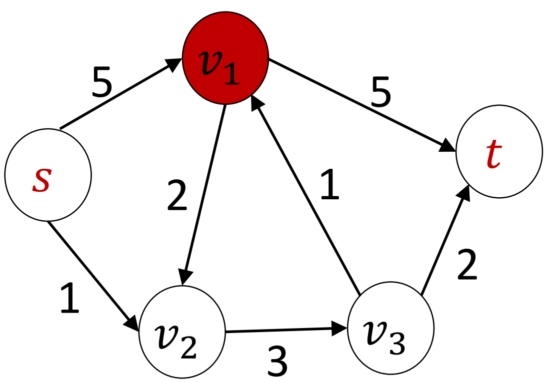


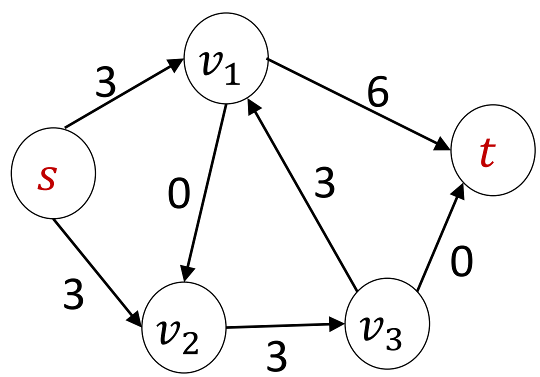
A Quick Recap

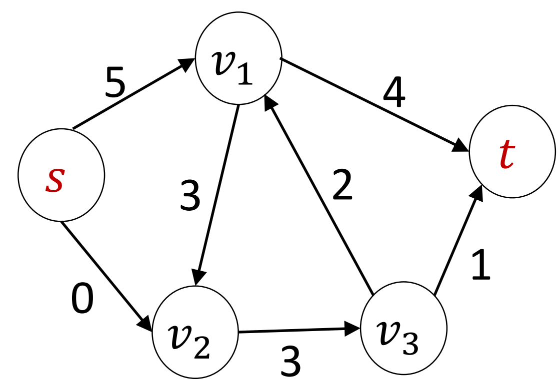
Which of the following is a valid flow for the given network ?



Fails to conserve flow at v_1

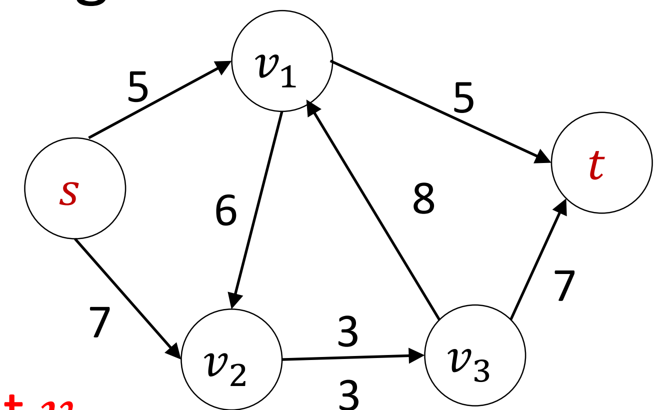




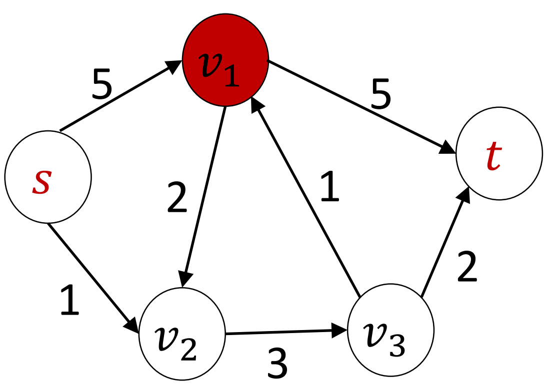


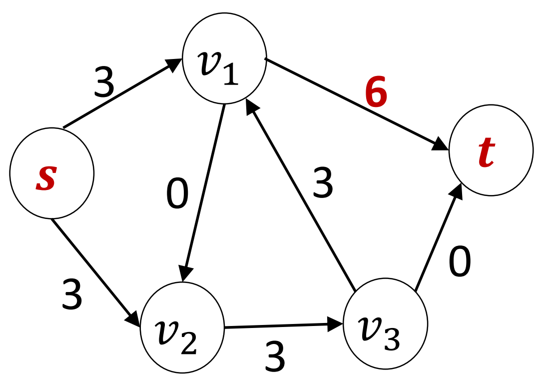
A Quick Recap

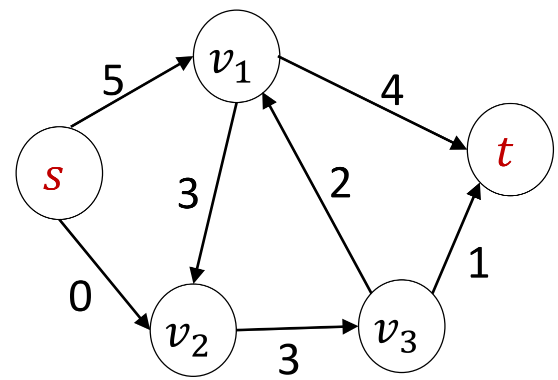
Which of the following is a valid flow for the given network ?



Fails to conserve flow at v_1



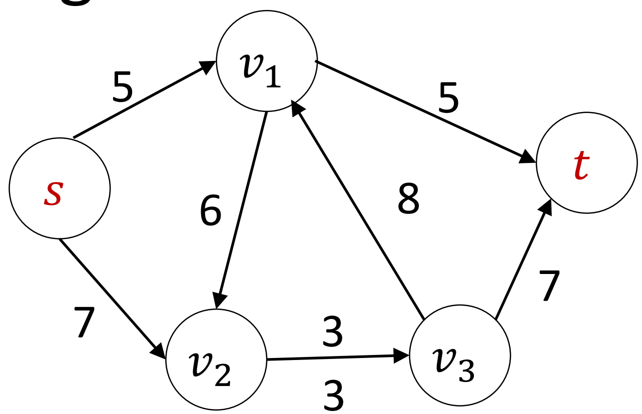


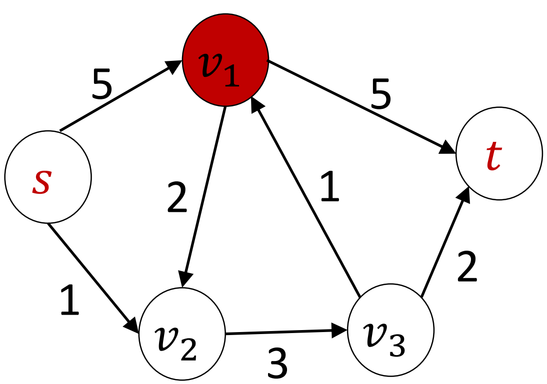


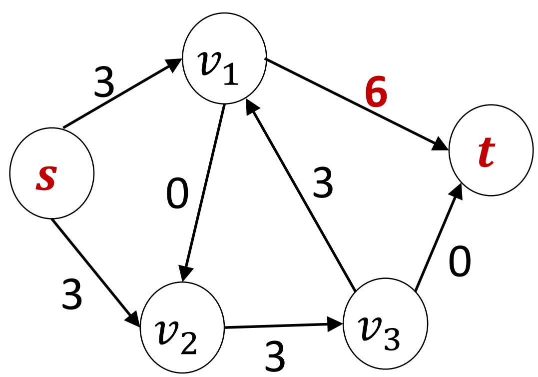
Exceeds the capacity constraint at the edge between v_1 and t

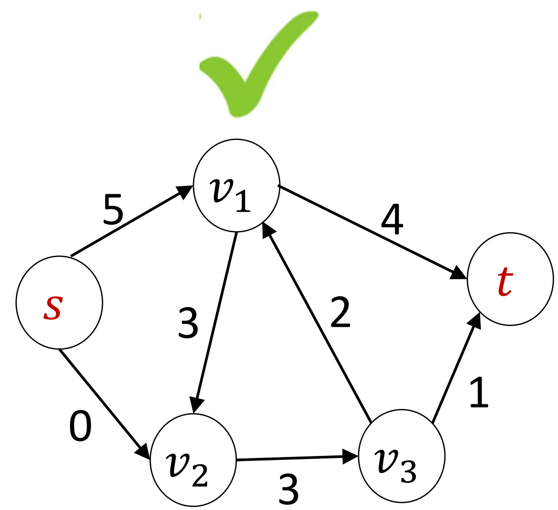
A Quick Recap

Which of the following is a valid flow for the given network ?





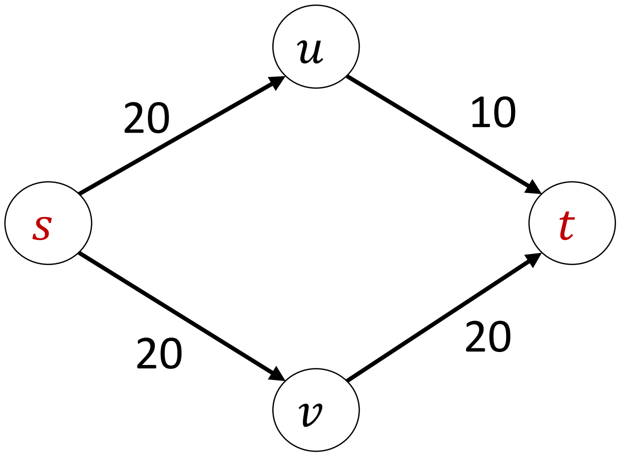


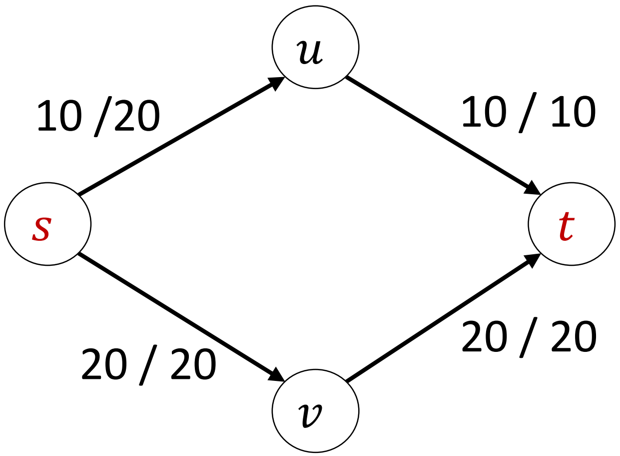


$$Note: \sum_{(s,u)\ \in E} f((s,u)) = \sum_{(u,t)\ \in E} f((u,t))$$

Solving Maximum flow problem

How should we go about solving this problem ?



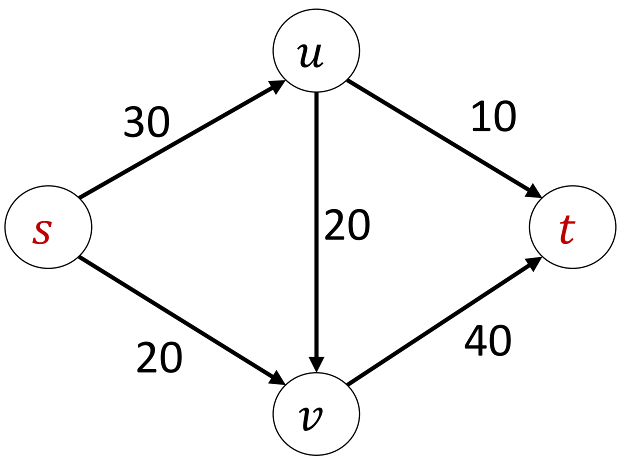


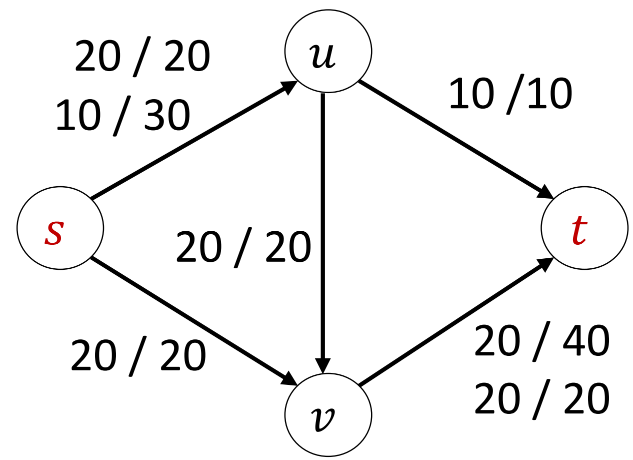
Maximum possible flow: 30

a / b = f_e / C_e

Solving Maximum flow problem

Another example !!





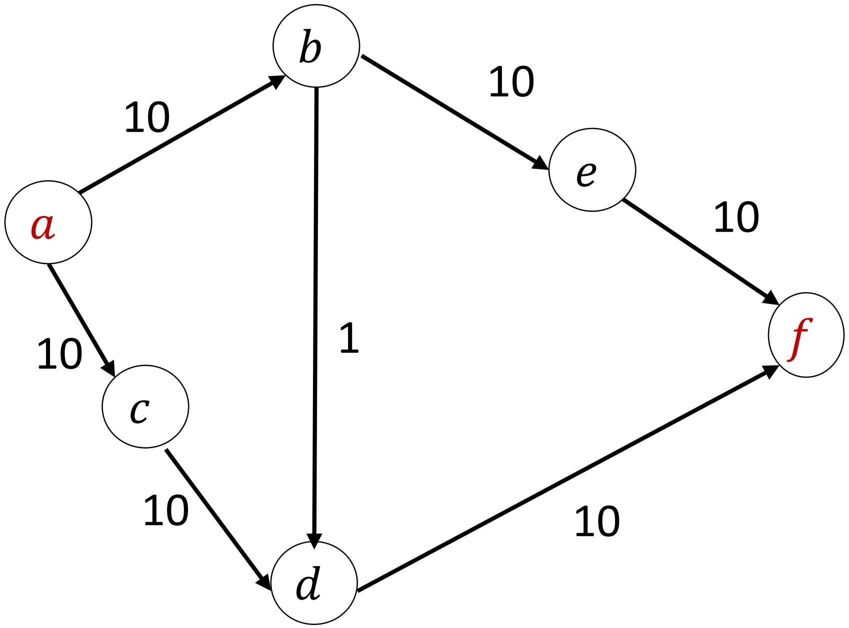
Maximum possible flow: 50

a / b = f_e / C_e

1. Look for paths from source to destination and count them if they have capacities left
2. Repeat until no paths from to destination is found with capacities left

Solving Maximum flow problem

Does the rules we created work for all scenarios?



$a \dashrightarrow b \dashrightarrow d \dashrightarrow f$ 1

$a \dashrightarrow c \dashrightarrow d \dashrightarrow f$ 9

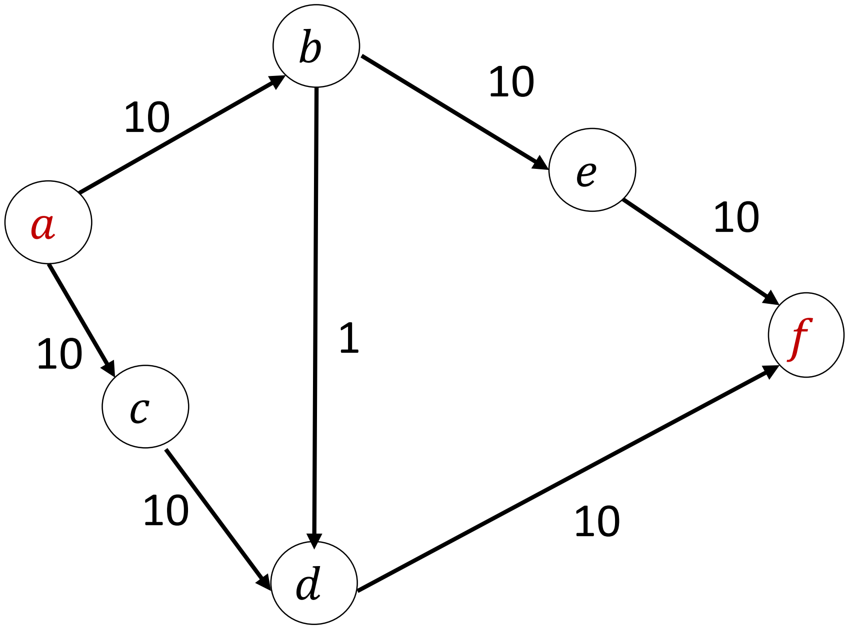
$a \dashrightarrow b \dashrightarrow e \dashrightarrow f$ 9

Max flow: 19. Is this maximum?

Correct answer : 20

Solving Maximum flow problem

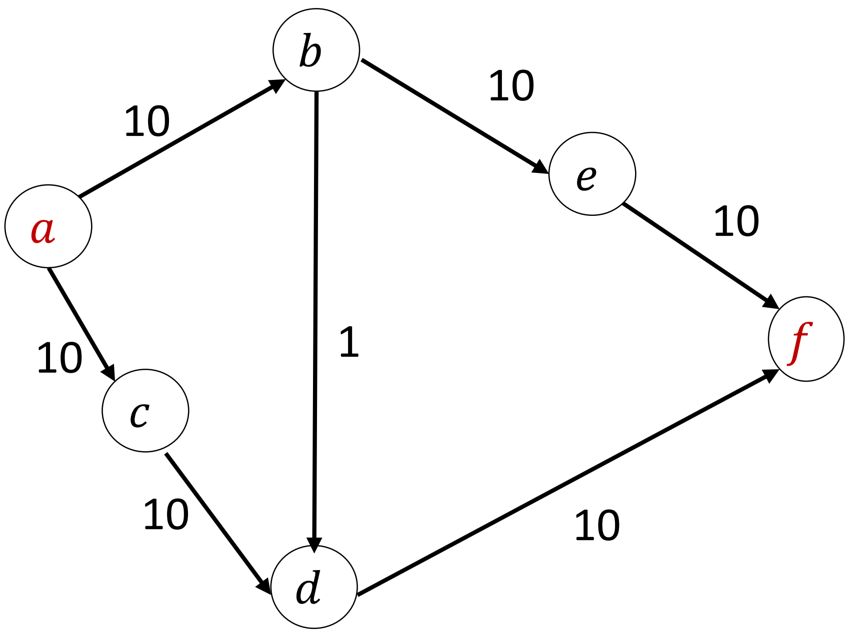
How can we solve this problem ?

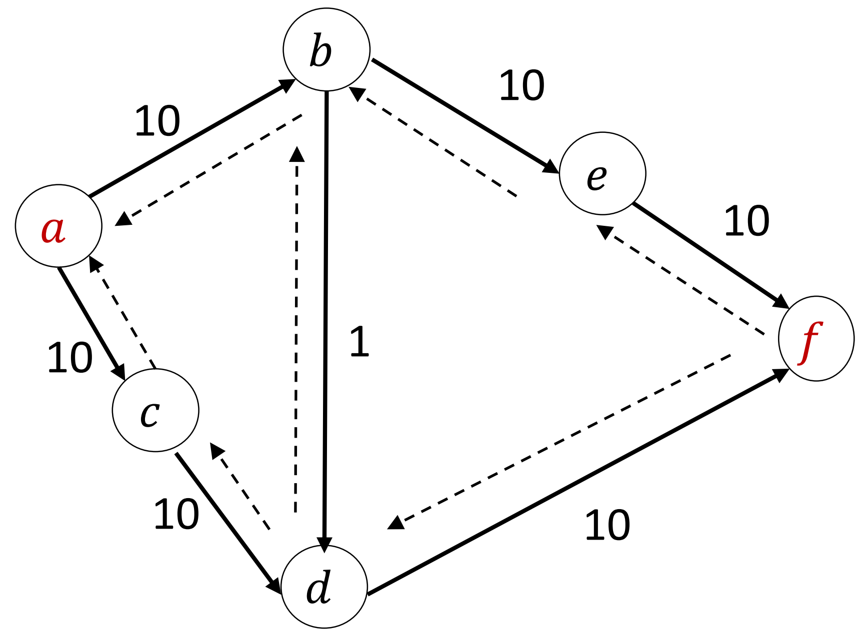


By having an option of reversing our decisions.

Solving Maximum flow problem

How can we solve this problem ?

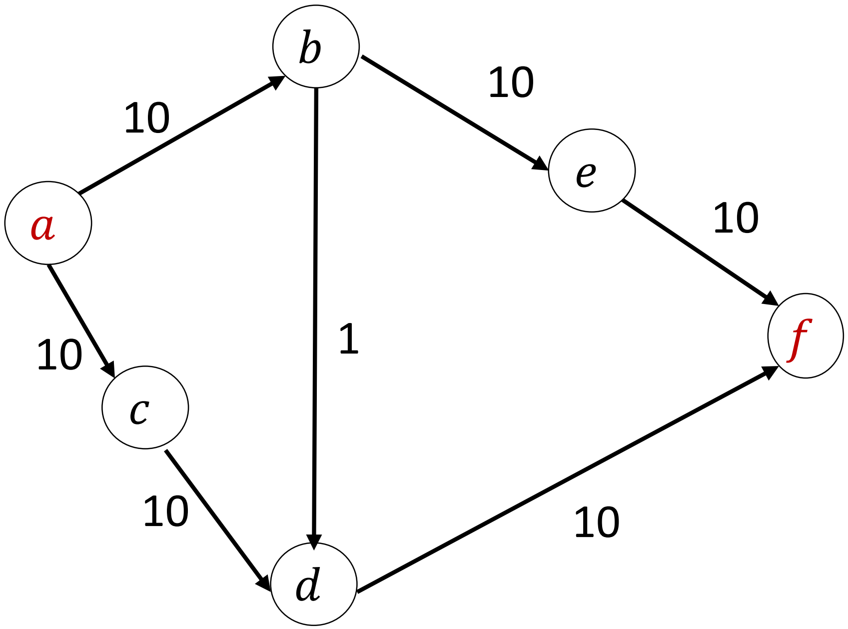


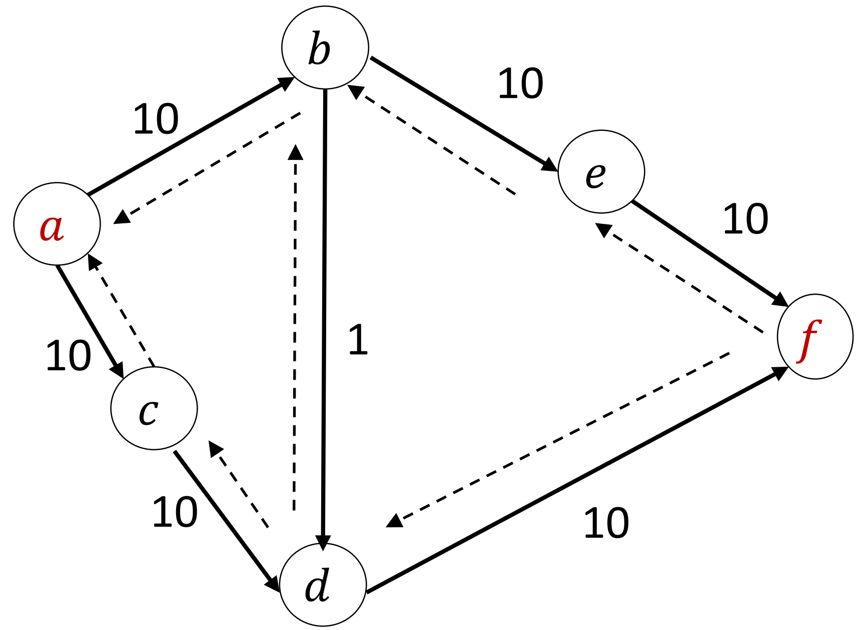


By having an option of reversing our decisions.

For every edge in the original graph, we add an “imaginary” reverse edge – this allows us to reverse our decisions (in a way)

Solving Maximum flow problem





$a \dashrightarrow b \dashrightarrow d \dashrightarrow f$ 1

$a \dashrightarrow c \dashrightarrow d \dashrightarrow f$ 9

$a \dashrightarrow b \dashrightarrow e \dashrightarrow f$ 9

$a \dashrightarrow c \dashrightarrow d \dashrightarrow b \dashrightarrow e \dashrightarrow f$ 1

Next lecture

- Ford-Fulkerson method for solving network flow problems