# CIE 500: Urban Water Systems and Complex Networks

## – Network Characteristics

Xudong Fan

Department of Civil, Structural and Environmental Engineering
University at Buffalo

February 10, 2025

# Overview

1. Recap

2. Network Characteristics

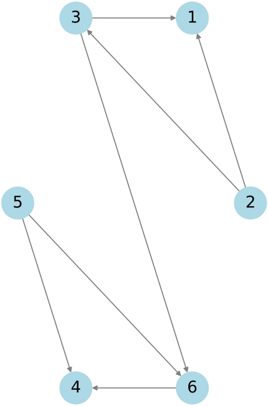
3. Node centrality measures

# Recap

In the last week, we have introduced

- Urban water systems and urban water cycle
- Random networks
- Network mathematical representation

# Node degrees




## undirected networks

degree $d(i) = d_i$ of node $i$ is defined as $d_i := |\{j \in V : (i,j) \in E\}|$

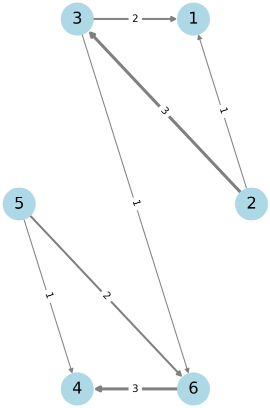
## directed networks

indegree $d_{in}(i)$ is the number of incoming edges, i.e.
$d_{in}(i) := |\{j \in V : (j,i) \in E\}|$
outdegree $d_{\text{out}}(i)$ is the number of outgoing edges, i.e.
$d_{\text{out}}(i) := |\{j \in V : (i,j) \in E\}|$

## example network

$d_{\text{in}}(6) = 2$

$d_{\text{out}}(6) = 1$

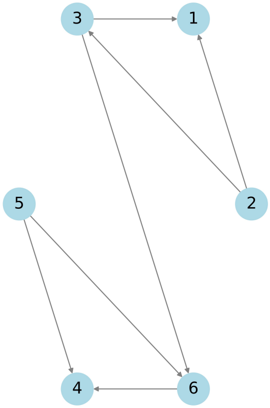



## weighted degrees

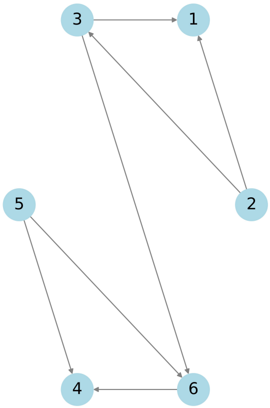
for weighted networks, the **weighted in- or outdegree** of a node is the sum of incoming or outgoing link weights, i.e.

$$w_{\text{in}}(i) := \sum_{j \in v} w(j,i) = \sum_{j=1}^{n} A_{ji}$$

$$w_{\text{out}}(i) := \sum_{j \in v} w(i,j) = \sum_{j=1}^{n} A_{ij}$$




- sequence $(\ p_0, p_1, \ldots, p_l\ )$ of nodes $p_i \in V$ is a **walk** from $p_0$ to $p_l$ if
$$(p_i, p_{i+1}) \in E \text{ for } i = 0, \ldots, l-1$$
- walk $(p_0, p_1, \ldots, p_l)$ is a **(simple) path** if
$$p_i \neq p_j \text{ for } 0 \leq i, j \leq l \text{ and } i \neq j$$
- walk $(p_0, p_1, \ldots, p_l)$ is a **cycle** if 1. $p_0 = p_l$ 2. $p_i \neq p_j$ for $0 < i, j < l$ and $i \neq j$
- **length of path, walk, or cycle**is defined as $\text{len}\,(p_0, \ldots, p_l) := l$

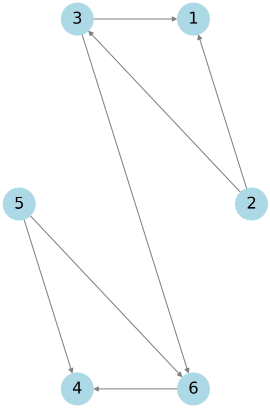



- (directed) network with no cycle is called **directed acyclic graph (DAG)**
- for any DAG $G = (V, E)$ we can find (at least) one **topological ordering of nodes**, i.e. mapping $T : V \rightarrow \mathbb{N}$ such that $T(v) < T(w)$ for all $(v, w) \in E$
- we can compute topological ordering in **linear time**
- DAGs are important concept in **scheduling, path analysis, causal inference, temporal network analysis**

**Topological ordering algorithm**

- Kahn's algorithm
- DFS with edge classification

One topological order of this network is $[2, 5, 3, 1, 6, 4]$




- **distance** dist$(v, w)$ between nodes $v$ and $w$ is the minimum length of any path between $v$ and $w$
- $\text{dist}(v, w) := \infty \Leftrightarrow \nexists$ path from $v$ to $w$
- path $(p_0, \ldots, p_l)$ is **shortest path** if $\text{len}\,(p_0, \ldots, p_l) = \text{dist}\,(p_0, p_l)$
- for weighted network $(p_0, \ldots, p_l)$ is **cheapest path** if $\sum_{i=1}^{l} w\,(p_{i-1}, p_i)$ is minimal

**Example**

- $dist(2, 4) = 3$
- the shortest path: $(2, 3, 6, 4)$

# Finding the shortest paths

## Dijstra's algorithm

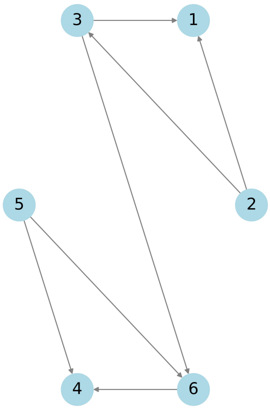
- single-source shortest paths for graphs with positive edge weights
- repeatedly relax path length for neighbors of first node in priority queue
- worst-case time complexity $\mathcal{O}(m + n \cdot \log n)$

## Bellman-Ford algorithm

- single-source shortest paths for graphs with real edge weights (no negative cycles)
- repeatedly relax path length of node v for all edges (v, w)

## Floyd-Warshall algorithm

- all-pairs shortest paths for graphs with real edge weights (no negative cycles)
- test triangle inequality for all triples of nodes
- worst-case complexity $\mathcal{O}\left(n^3\right)$




- diameter $\mathrm{diam}(G)$ of network $G = (V, E)$ is length of the longest shortest path

$$\mathrm{diam}(G) := \max_{v,w \in V} \mathrm{dist}(v, w)$$

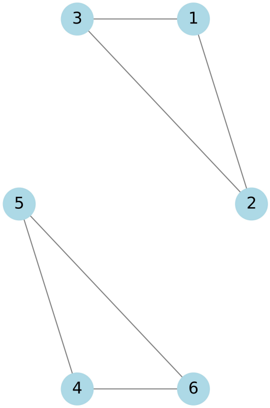
- average shortest path length $\langle l \rangle$ is the average distance between nodes, i.e.

$$\langle l \rangle := \frac{1}{|V|^2} \sum_{(v,w) \in v \times v} \mathrm{dist}(v, w)$$

**Example**

- $diam(G) = 3$

# Connected components

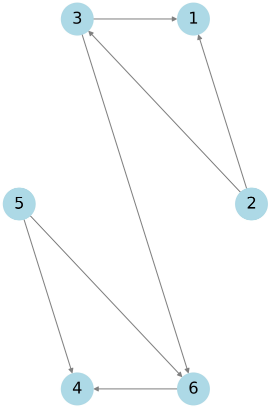



- undirected network $G = (V, E)$ is connected if $\text{dist}(v, w) < \infty$ for all $v, w \in V$
- connected components of $G = (V, E)$ are maximally connected subgraphs $G' = (V', E')$ with $V' \subseteq V$ and $E' \subseteq E$
- size of connected component $G' = (V', E')$ is $|V'|$
- largest connected component $G'$ is called giant connected component if

$$\frac{|v'|}{|v|} \approx 1$$

# Connected components in directed graph





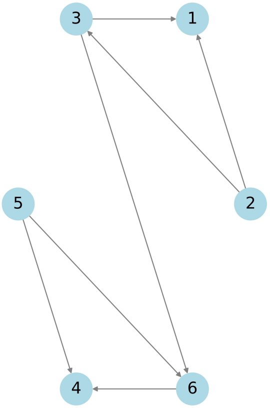

- we distinguish between **strongly and weakly connected directed networks**
- directed network is **weakly connected** if corresponding undirected network is connected
- directed network $G = (V, E)$ is **strongly connected** if $\mathrm{dist}(v, w) < \infty \quad \forall v, w \in V$
- **strongly connected components**of $G = (V, E)$ are maximal strongly connected subgraphs $G' = (V', E')$ with $V' \subseteq V$ and $E' \subseteq E$

# Node centrality measures

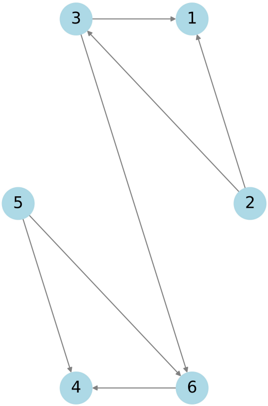
- important basic task in network analysis is to **identify important nodes**

**node centrality measures**

- For network $G = (V, E)$ node centrality $c : V \to \mathbb{R}$ is a function or measure that can be used to assess the importance of nodes in a network.
- For $v \in V$, centrality indicators $c(v)$ provide a total order than can be used to rank nodes by importance.



- we can define centrality of node $v \in V$ based on the number of incident links
- **degree centrality** $d(v) = d_v$ for undirected networks
- **in- or out-degree centrality**$d_{\text{in}}(v)$ or $d_{\text{out}}(v)$ for directed networks
- degree-based centralities are **local measures of importance**



# Betweenness centrality

- **betweenness centrality** of node $v$ is defined as

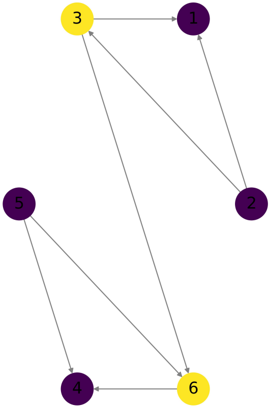
$$C_B(v) := \sum_{s,t \in v - \{v\}, s \neq t} \frac{N_{st}(v)}{N_{st}}$$

where $N_{st}$ is number of **shortest paths** from $s$ to $t$ and $N_{st}(v)$ is number of shortest paths from $s$ to $t$ passing through $v$

- in general we have $0 \leq C_B(v) \leq n^2$
- **normalized betweenness** centrality can be defined as

$$C_{\bar{B}}(v) := \frac{C_B(v) - \min_i C_B(i)}{\max_i C_B(i) - \min_i C_B(i)} \in [0, 1]$$

- the normalized betweenness centrality is (2: 0), (1: 0.0), (3: 0.1), (5: 0.0), (4: 0.0), (6: 0.1)



- **closeness centrality** of node $v$ is the inverse of the average shortest path distance to all other nodes

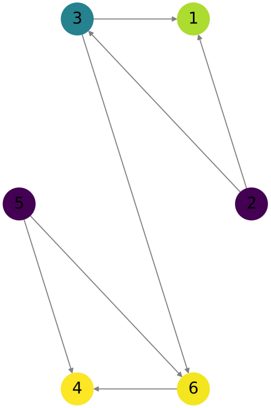
$$c_c(v) = \frac{n-1}{\sum_{w \in v - \{v\}} \text{dist}(v, w)} \in [0, 1]$$

where $n$ is the number of nodes

- for disconnected networks we have

$$c_c(v) = \frac{n-1}{\infty} := 0 \quad \forall v \in v$$

- betweenness and closeness are **path-based centrality measures** that depend on the **topology** of links
- the normalized betweenness centrality is (2: 0.0), (1: 0.4), (3: 0.2), (5: 0.0), (4: 0.457), (6: 0.45)



Conduct the Network Characteristics for the random network that you created in your previous code. Consider the following:

- If the network is too large, show the node degree/centrality distribution.
- Provide a description about the characteristics of the graph, the nodes, and the paths.