

Connectivity in graphs

Social Networks Analysis and Graph Algorithms

Prof. Carlos “ChaTo” Castillo — <https://chato.cl/teach>



Universitat
Pompeu Fabra
Barcelona

Contents

- Sparsity
- Paths and distances
- Connected components

Sources

- A. L. Barabási (2016). Network Science - [Chapter 01](#)
- F. Menczer, S. Fortunato, C. A. Davis (2020). A First Course in Network Science – Chapter 02
- URLs cited in the footer of specific slides

Sparsity

Real networks are sparse

- Theoretically $L_{\max} = \binom{N}{2} = \frac{N(N-1)}{2}$
- Most real networks are sparse, i.e., $L \ll L_{\max}$

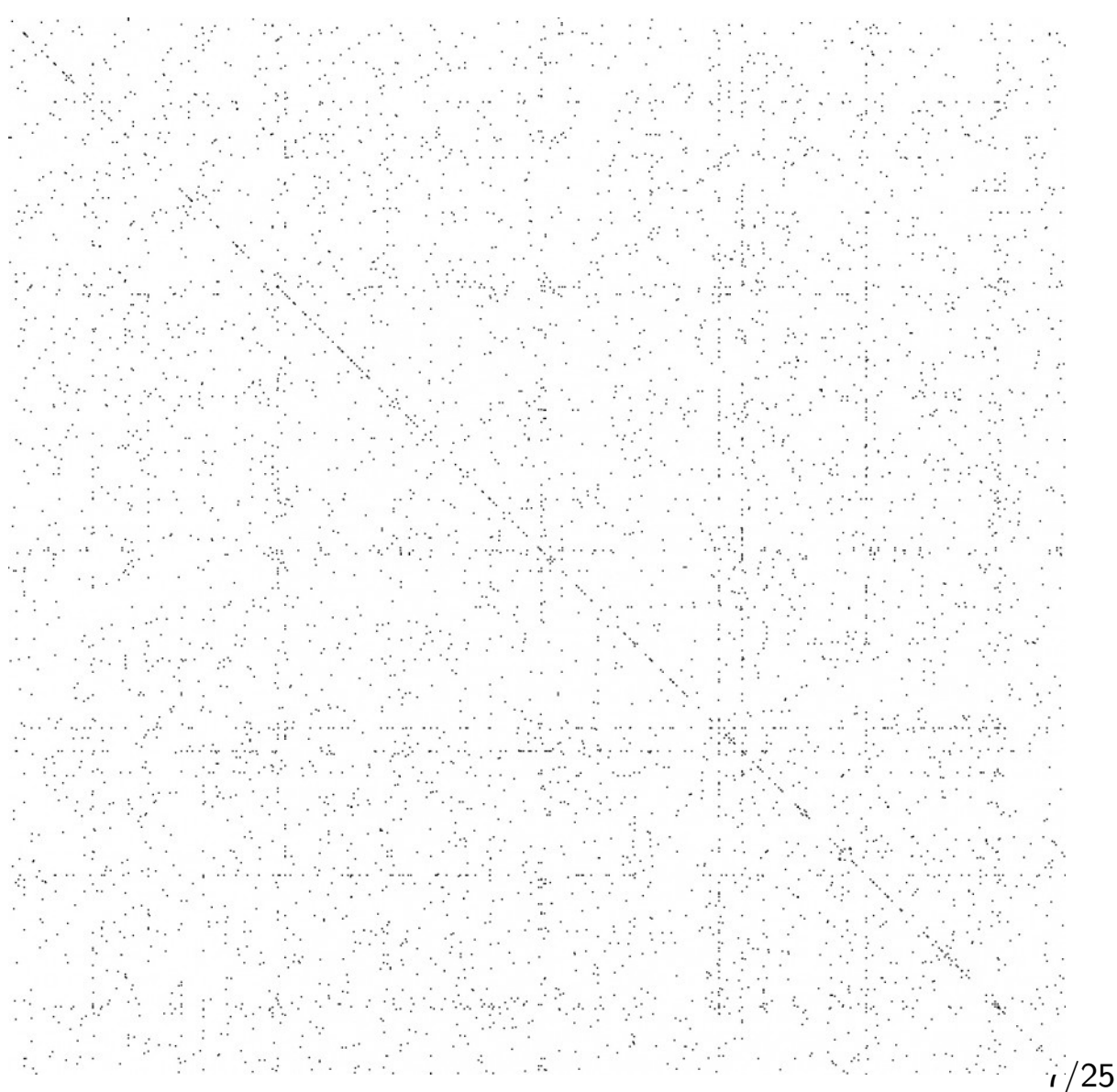
L is the number of links in the network, N is the number of nodes on it

How sparse are some networks?

Network	$ V $	$ E $	Max $ E $
Zachary's Karate Club	34	78	561
Game of Thrones	84	216	3496
US companies ownership	1351	6721	911K
Marvel comics	6K	167K	17M

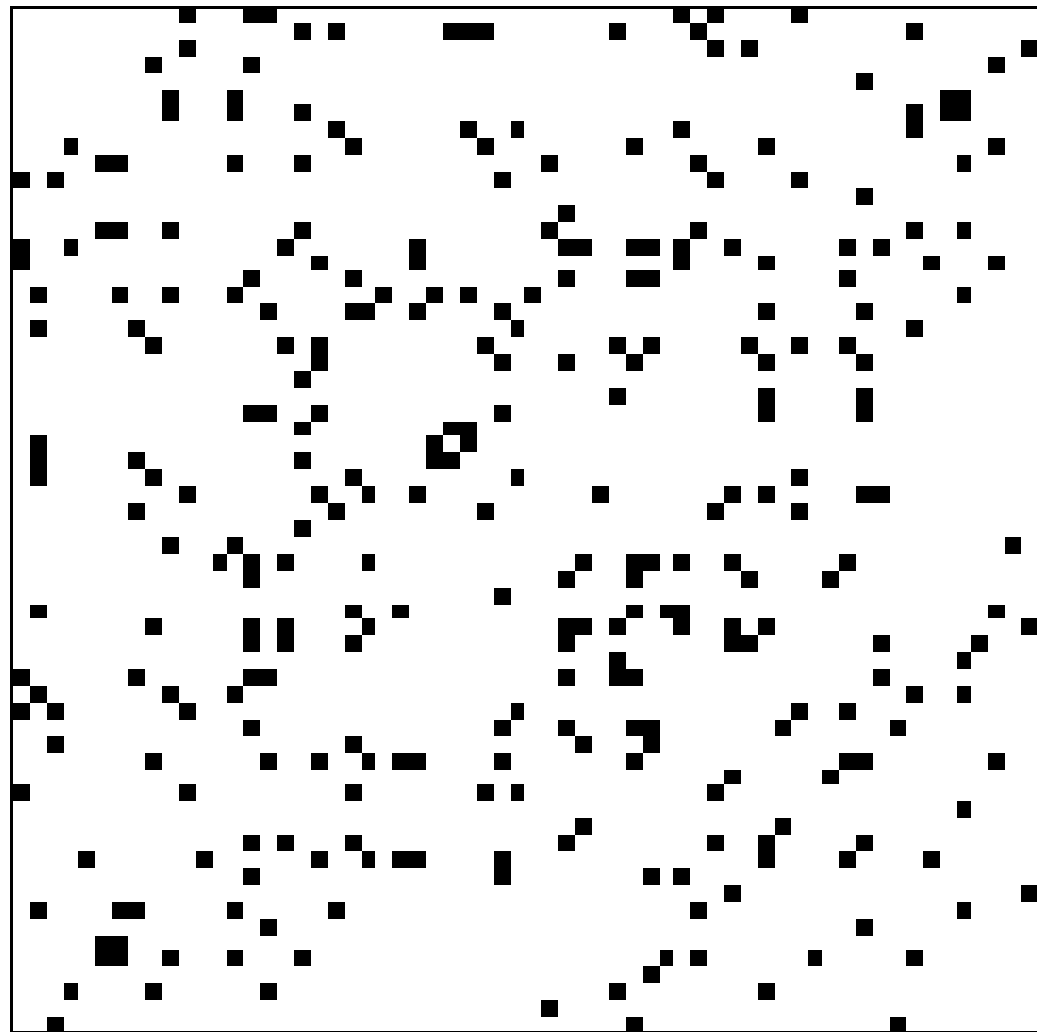
Example: protein interaction network

($N=2K$, $L=3K$)



Example: dolphins

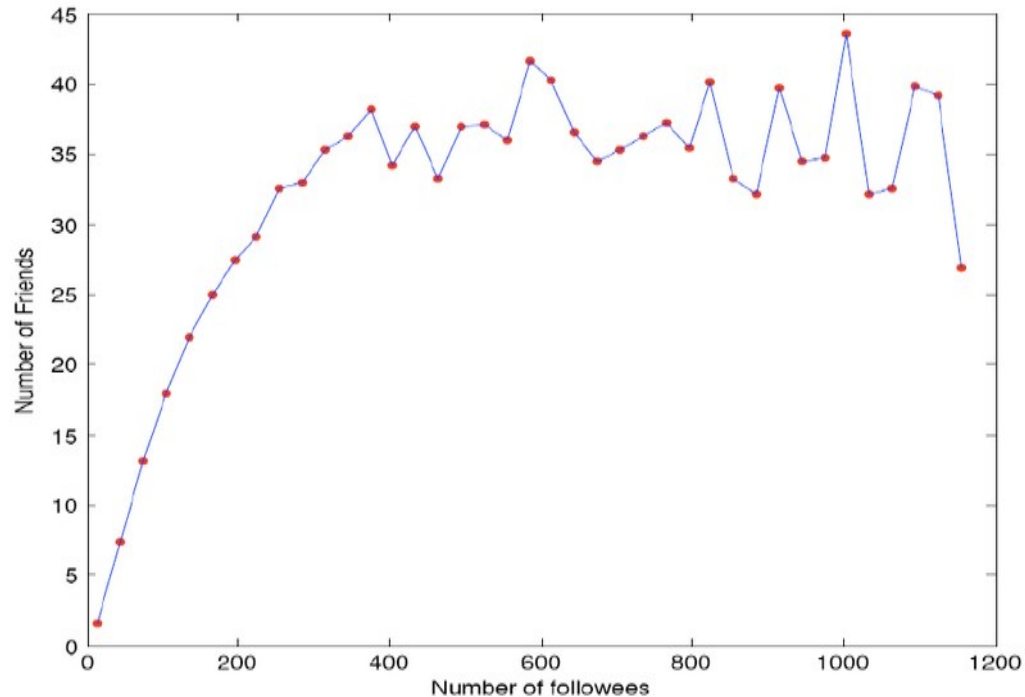
($N=62$, $L=318$)



Why are networks sparse?

- Different mechanisms, think about it from the node perspective:
 - How many items **could** the node be connected to
 - Would it be **realistic** to connect to a large fraction of them?
- In social networks, Dunbar's number (≈ 150)

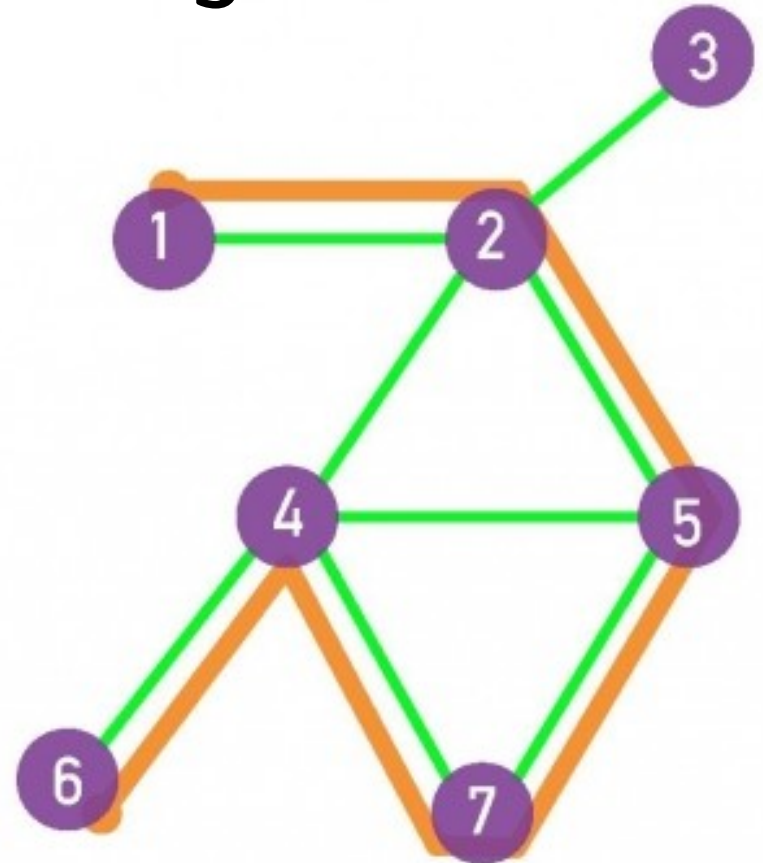
Example: actual friends in Twitter vs people you follow in Twitter



Paths and distances

Paths: sequences of edges

- The destination of each edge is the origin of the next edge
 - In directed graphs, paths follow the direction of the edges
- The length of the path is the number of edges on it
 - Example: path in orange has length 5



Distance

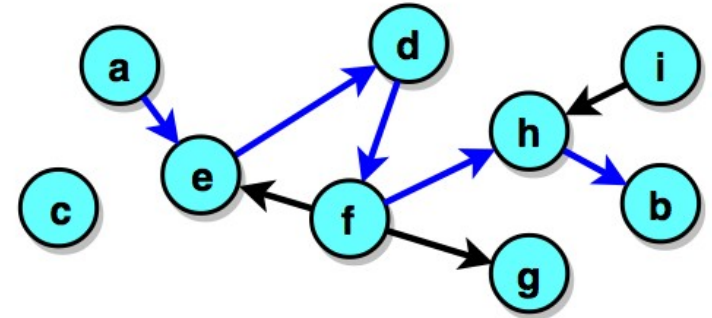
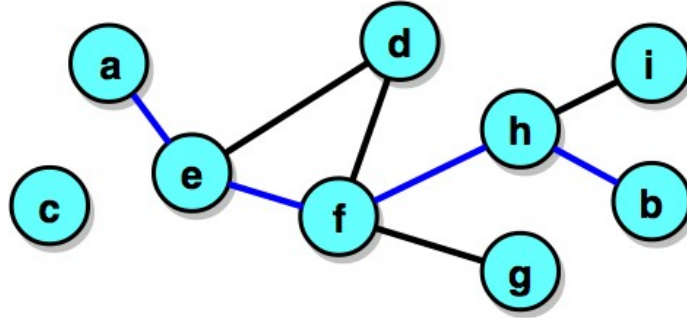
- If two nodes i, j are in the same connected component:
 - the **distance** between i and j , denoted by d_{ij} is the **length of the shortest path** between them
- If they are not in the same connected component, the distance is by definition infinite (∞)

Blue =
shortest
path
between
nodes *a*
and *b*

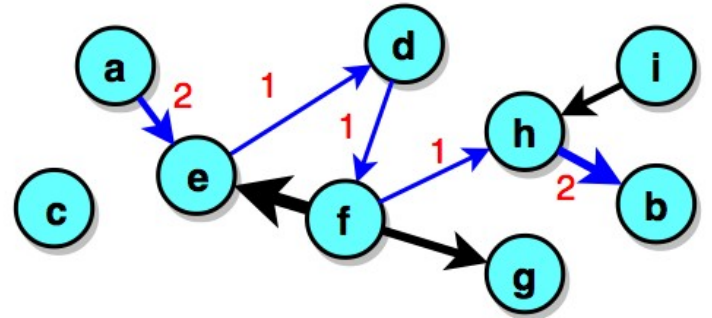
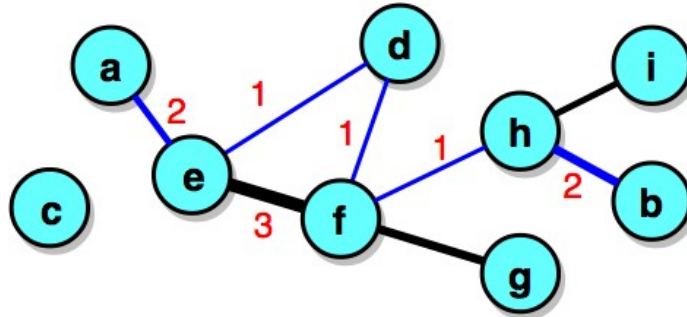
Undirected

Directed

Unweighted



Weighted



Diameter

- The **diameter** of a network is the maximum distance between two nodes on it, d_{\max}
- The **effective diameter** (or **effective-90% diameter**) is a number d such that 90% of the pairs of nodes (i,j) are at a distance smaller than d
- The **average distance** is $\langle d \rangle$, and is measured only for nodes that are in the same connected component

Connected components

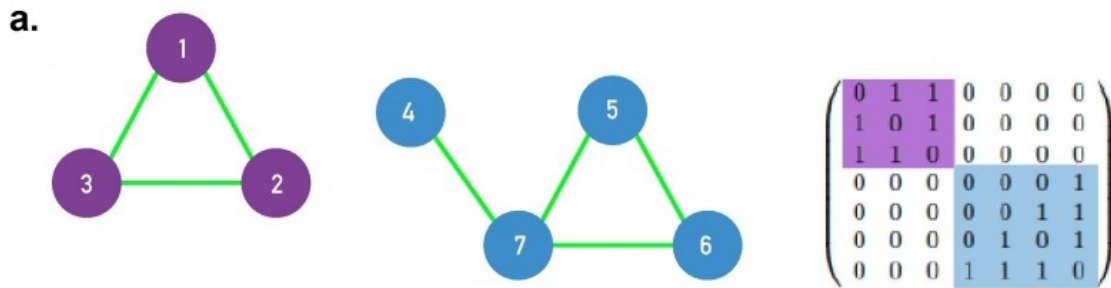
Connectedness

- If a path exists between two nodes i, j : those nodes are part of the same **connected component**
- A **connected graph** has only one connected component
- A **singleton** is a connected component with only one node

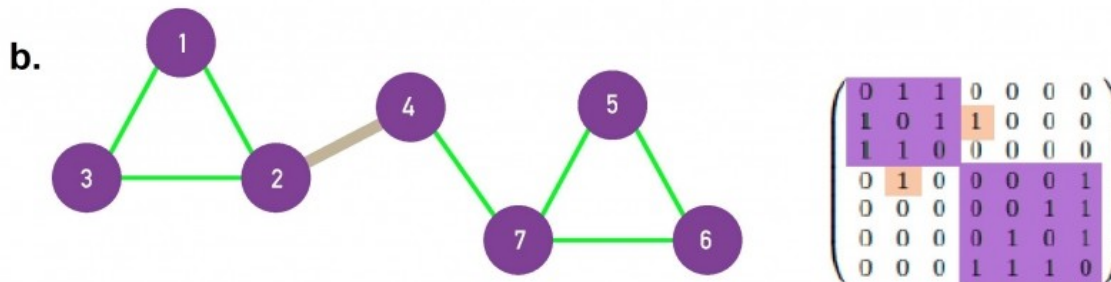
Connected graphs

A **disconnected graph** has an adjacency matrix that can be arranged in block diagonal form

a. disconnected



b. connected

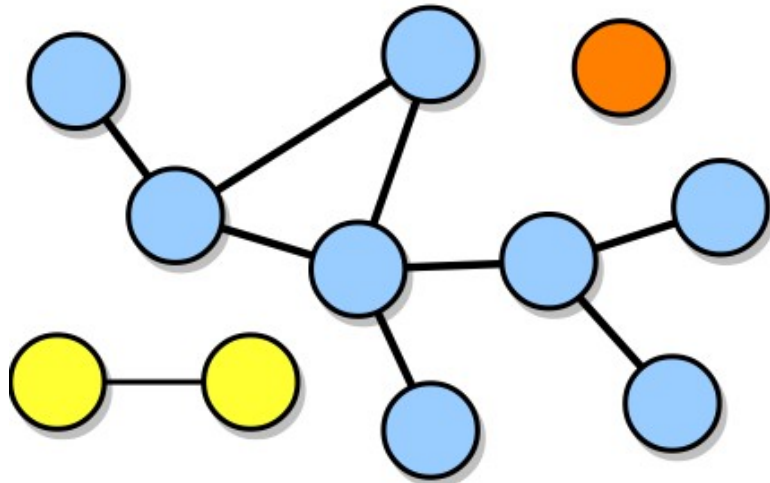


Connectedness in directed graphs

- A directed graph is **strongly connected** if it has only one connected component
- A directed graph is **weakly connected** if, when seen as undirected, has only one connected component

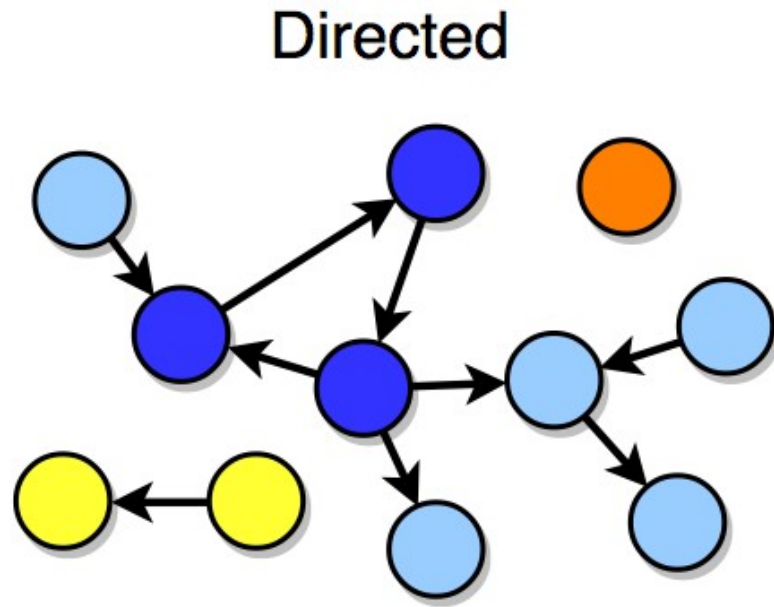
Connectedness example (directed)

Undirected



- Is not connected
- Has 3 connected components
- One of the connected components is a singleton

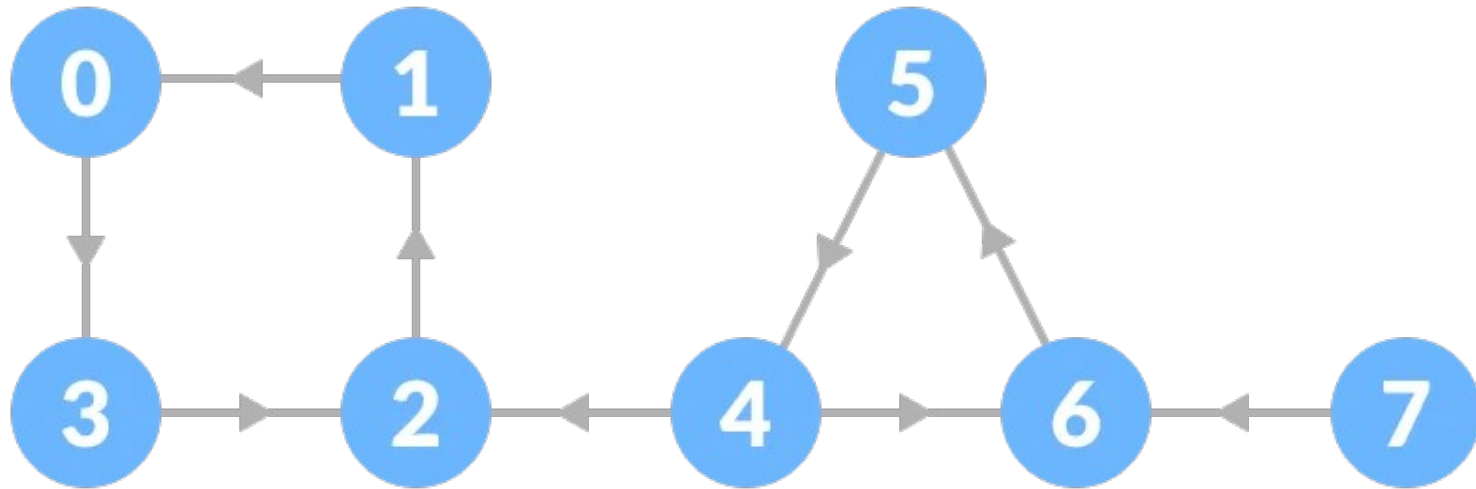
Connectedness example (directed)



- Is not strongly connected
- Is not weakly connected
- Has 3 connected components

Quick exercise

Find the strongly connected components



Summary

Things to remember

- Sparse vs dense graph
- Distance, diameter, effective diameter
 - In directed and undirected graphs
- Connected components
 - In directed and undirected graphs

Practice on your own

- Measure the sparsity of a graph L/L_{\max}
- Compute the distance between two nodes
- Compute the diameter of a graph
- Identify connected components