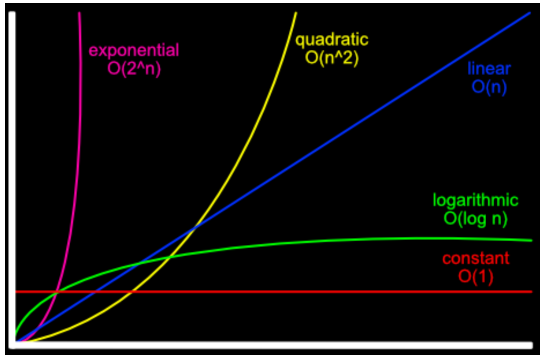
Density and Sparsity

$$\rho = \frac{\langle k \rangle}{n} \quad \langle k \rangle = \rho n$$



- The average degree $\langle k \rangle$ is related to the density:
  - In a **dense** network, where ρ is approximately constant as $n \to \infty$,
    - The mean degree $\langle k \rangle$ grows linearly with n
  - In a **sparse** network:
    - The mean degree $\langle k \rangle$ grows sublinearly with n ($e.g.$, $\log n$)
  - In an ***extremely sparse*** network:
    - The mean degree $\langle k \rangle$ remains constant
    - Ex) Friendship networks (the maintenance of friendships is likely independent of world population)

More details in later lectures on 'network models'

Real-world networks are far from complete

Most real-world networks are sparse

$$(|E| \ll |E|_{\max}) \text{ or } (\langle k \rangle \ll n-1)$$

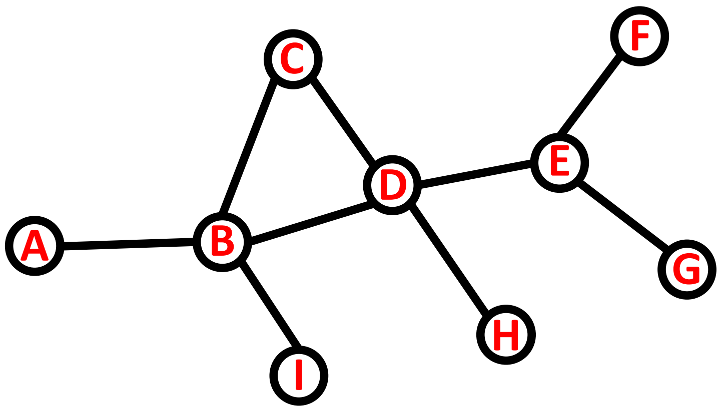
Social networks (LinkedIn):	N=6,946,668	⟨k⟩=8.87
Communication (MSN IM):	N=242,720,596	⟨k⟩=11.1
Coauthorships (DBLP):	N=317,080	⟨k⟩=6.62
Internet (AS-Skitter):	N=1,719,037	⟨k⟩=14.91
Roads (California):	N=1,957,027	⟨k⟩=2.82
Proteins (S. Cerevisiae):	N=1,870	⟨k⟩=2.39

(Source: *Leskovec et al., Internet Mathematics, 2009*)

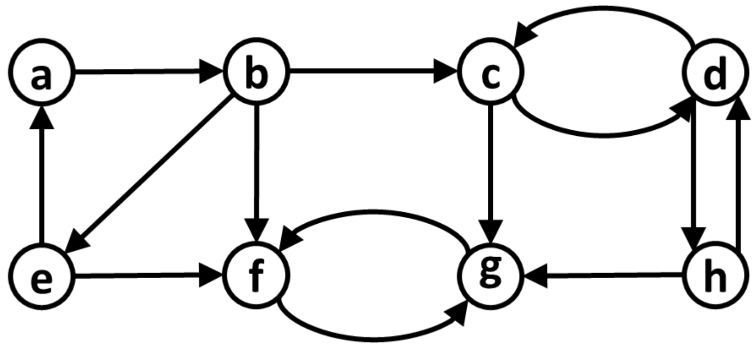
Consequence: Adjacency matrix is filled with zeros!

Walks and Paths

- ***Walk*** – any sequence of nodes in a network such that every consecutive pair of nodes in the sequence is connected by an edge
  - A walk can intersect itself (possibly revisiting a node more than once)
  - [In undirected networks]: each edge can be traversed in either direction
  - [In directed networks]: each edge must be traversed in the direction of that edge



Ex) A-B-C-D-B-D-E-F



Ex) a → b → c → d → c → g → f → g

Walks and Paths

- ***Path*** – a walk that does not intersect itself
- ***Length*** of a walk in a network: #edges traversed along the walk (not #nodes)
  - The number of '***hops***' from node to adjacent node

- How to calculate **the number of walks of a given length** γ on a network?
  - The total number of walks of length 2 from i to j via any node is:

$$N_{ij}^{(2)} = \sum_{k=1}^{n} A_{ik} A_{kj} = [\boldsymbol{A}^2]_{ij} \quad ,$$

Where $[\ldots]_{ij}$ denotes the ijth element of the matrix (i-th row, j-th column)

Walks and Paths

- Then, the total number of walks of length 3 from i to j via k and l?

$$N_{ij}^{(3)} = \sum_{k=1}^{n} \sum_{l=1}^{n} A_{ik} A_{kl} A_{lj} = [\boldsymbol{A}^3]_{ij} \quad .$$

- Generalizing to walks of arbitrary length γ:

$$N_{ij}^{(\gamma)} = [\boldsymbol{A}^\gamma]_{ij} \ .$$

- Then, the total number of **loops** of length γ?

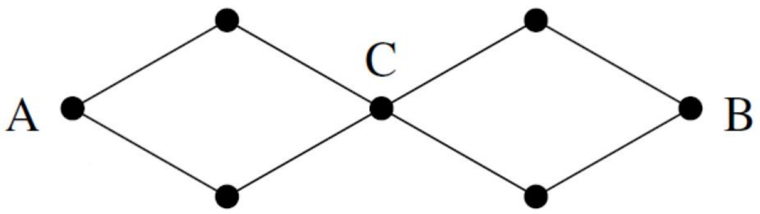
$$L_\gamma = \sum_{i=1}^{n} [\boldsymbol{A}^\gamma]_{ii} = \mathrm{tr}(\boldsymbol{A}^\gamma) \ .$$

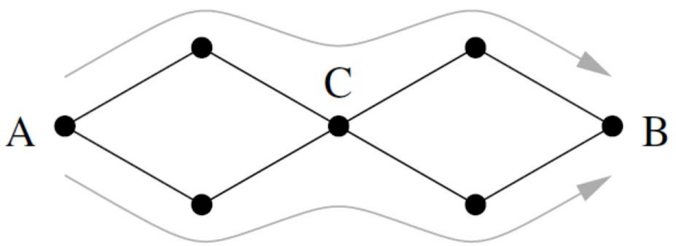
Independent Paths

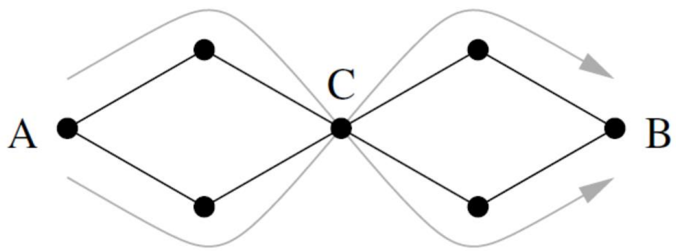
- Two paths connecting a given pair of nodes are ***edge-independent*** if they share no edges.
- Two paths connecting a given pair of nodes are ***node-independent*** if they share no nodes (other than starting and ending nodes).
- If two paths are node-independent then they are also edge-independent (but, the reverse is not true).

Independent Paths

- Can you find edge-independent paths from A to B?



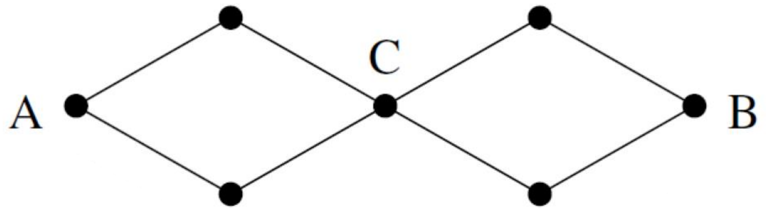




- Edge-independent paths are **not unique**.
- The number of independent paths(edge- or node-independent) from A to B **cannot exceed A's degree nor B's degree**.
- **The smaller of the degrees** of the two nodes gives an **upper bound** on the number of independent paths.

Connectivity

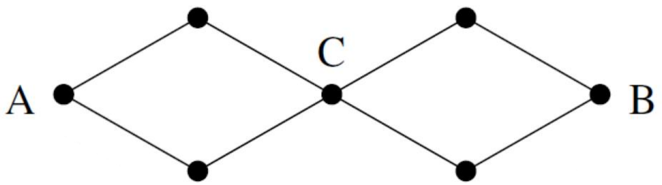
- ***Connectivity***: the number of independent paths between a pair of nodes
  - ***node connectivity***: #node-independent paths between a pair of nodes
  - ***edge connectivity***: # edge -independent paths between a pair of nodes



- edge connectivity 2
- node connectivity 1

The connectivity of a pair of nodes can be thought of as **a measure** of **how strongly connected** those nodes are.

Cut Sets



- Connectivity can also be thought of in terms of ‘***bottlenecks***’ between nodes.
  - Ex) Nodes A and B are connected by only one node-independent path due to node C:
- ***Node cut set***: a set of nodes whose removal will disconnect a specified pair of the nodes
  - Ex) The removal of node C leads to no path from A to B (a cut set of size 1).
  - Ex) Other cut sets for A and B (a cut set larger than size 1)
- ***Edge cut set***: a set of edges whose removal will disconnect a specified pair of nodes
- ***Minimum cut set***: the smallest cut set disconnecting a specified pair of nodes
  - Ex) {C}: a minimum node cut set for nodes A and B

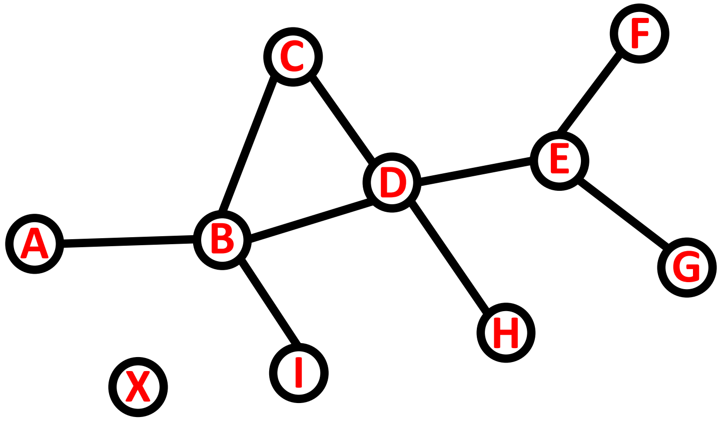
minimum cut set size = #independent paths (b.w. any pair of nodes)

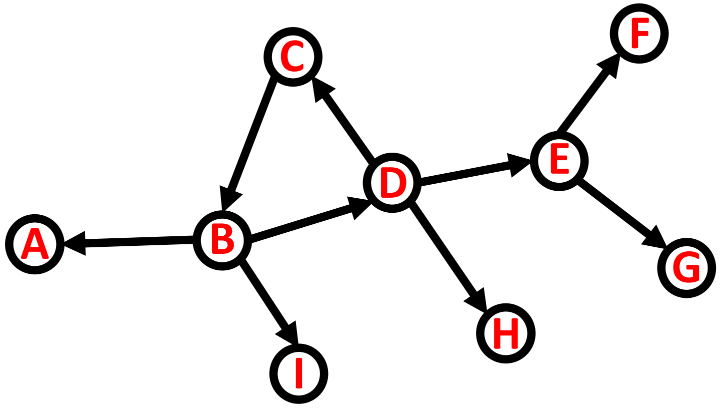
We can obtain minimum cut set size by counting independent paths.

Maximum flow = (#edge-independent paths) X (edge capacity)
(when the same maximum for every edge capacity is assumed)

Distance in a Network

- ***Distance*** in a network
  - *Shortest* or *geodesic path* between two nodes
  - The length of the shortest path (#edges along the shortest path)
  - Directed Networks: distance is not symmetric
- ***Diameter*** of a network
  - The maximum distance (shortest path) between two nodes in a network (the longest shortest path)



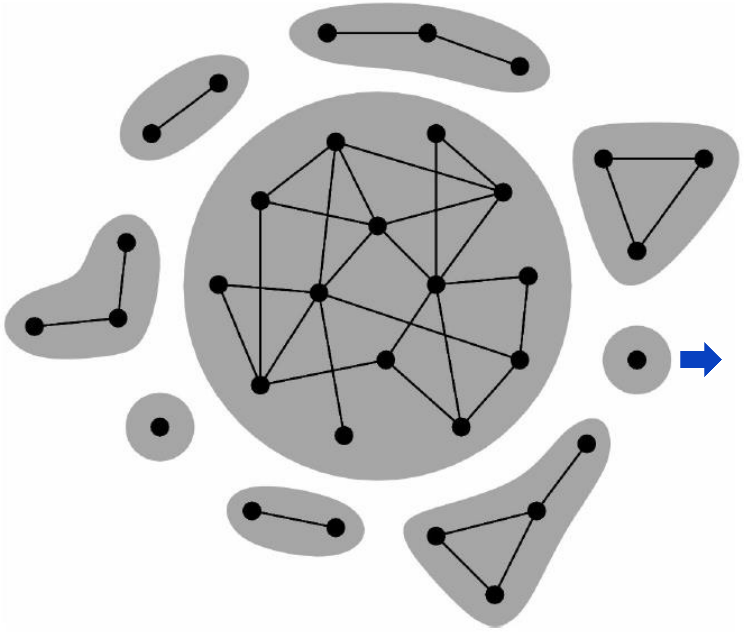


$\boldsymbol{h_{\mathrm{B,D}} \neq h_{\mathrm{D,B}}}$

$h_{\mathrm{A,D}} = \mathbf{2}$ $\quad h_{\mathrm{A,E}} = \mathbf{3}$ $\quad h_{\mathrm{A,X}} = \infty$ $\quad h_{\mathrm{B,D}} = \mathbf{1}$ $\quad h_{\mathrm{D,B}} = \mathbf{2}$

Components

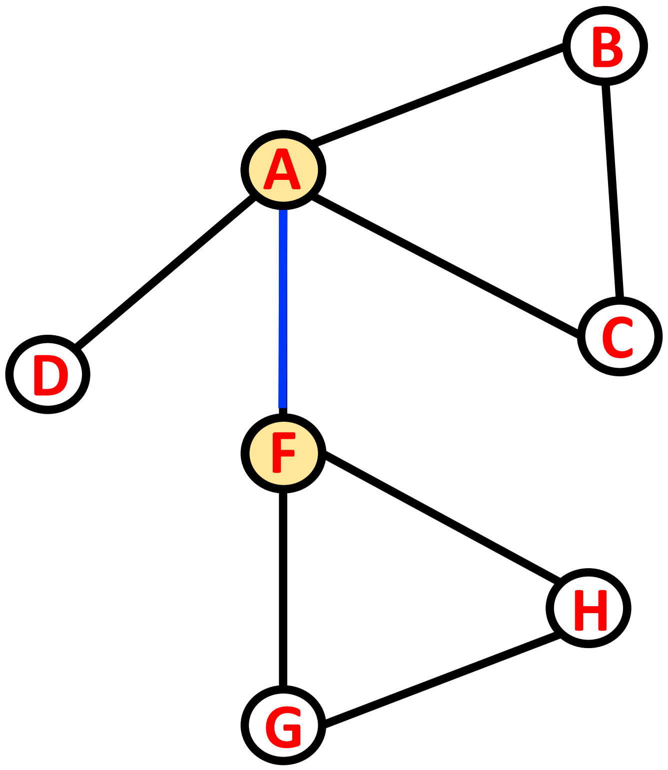
- ***Component***: a subset of nodes of a network, where exists at least one path between any two nodes
  - No path between any pair of nodes in different components
- The majority of nodes in the real-world networks are connected, while the others are fragmented.



➡ [singleton node]: a component of size one

Components – Undirected Networks

- ***Connected Network***
  - Any two vertices can be joined by a path
- ***A disconnected graph*** is made up by two or more connected components



Bridge Edge:

- If erased, the graph becomes disconnected
- Which one?

(A, F)

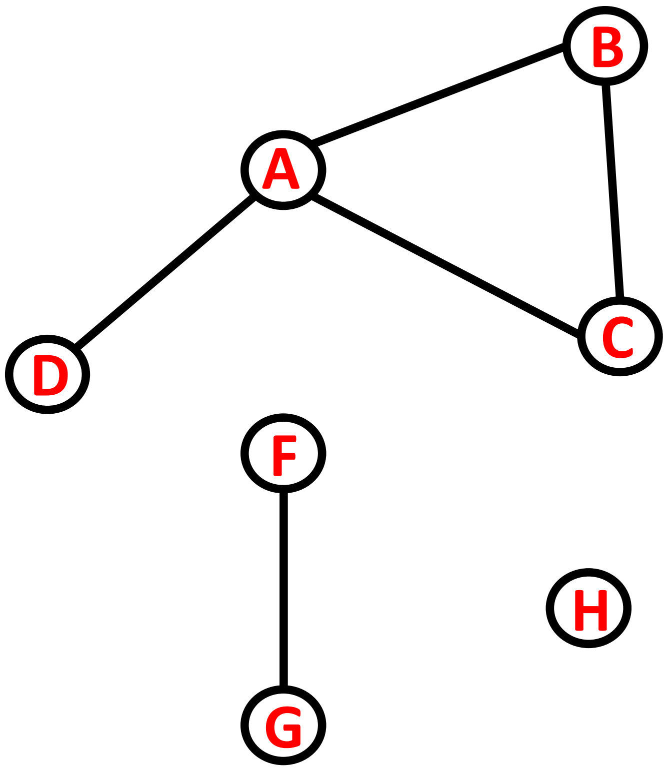
Articulation Node:

- If erased, the graph becomes disconnected
- Which one?

A, F

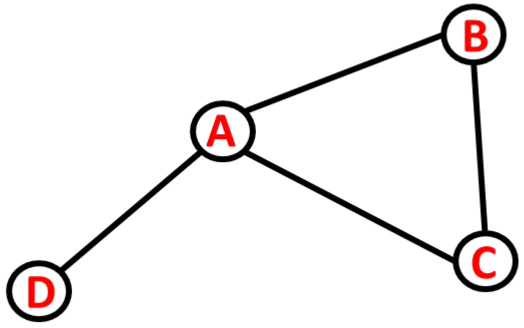
Components – Undirected Networks

- Connected Network
  - Any two vertices can be joined by a path
- A disconnected graph is made up by two or more connected components



***Giant Component*:**

- Largest Component
- Which one?



***Isolated Node*:**

- Degree is zero
- Which one?

Components – Adjacency Matrix

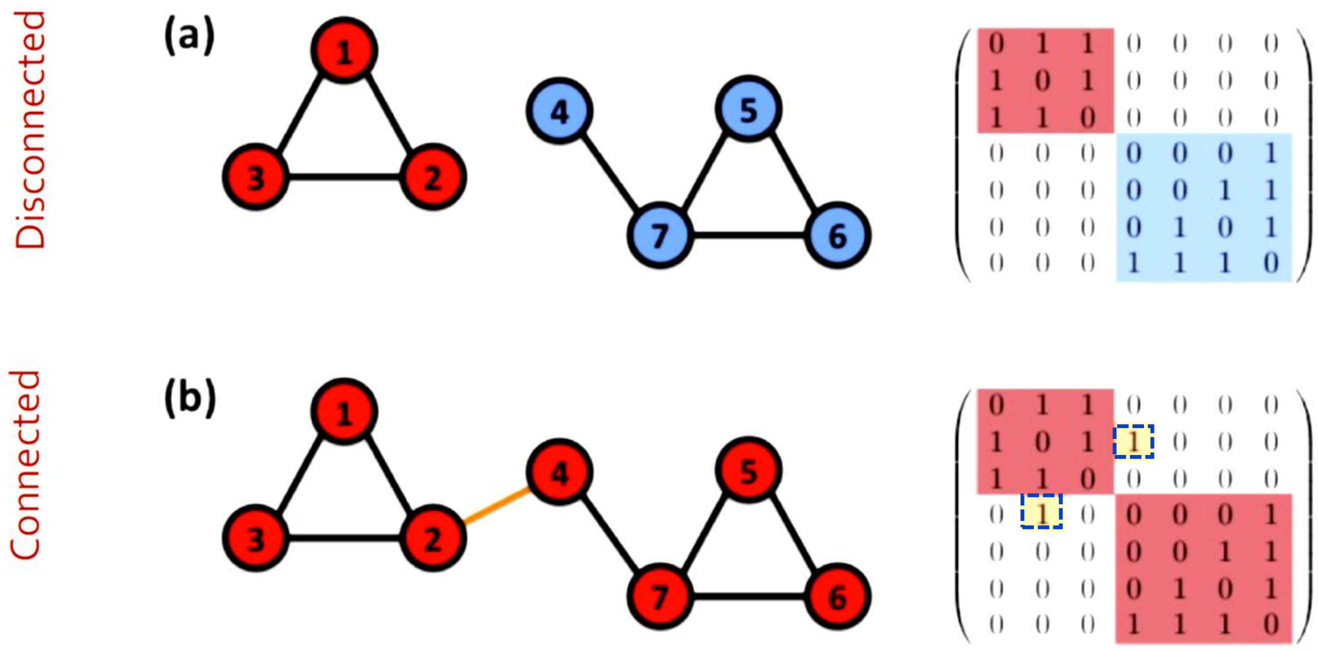
- Adjacency matrix in a block-diagonal form:
  - Nonzero elements are in a block
  - All other elements are zero

$$
\mathbf{A} = \begin{pmatrix} \Box & 0 & \cdots \\ 0 & \Box & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix}.
$$

components

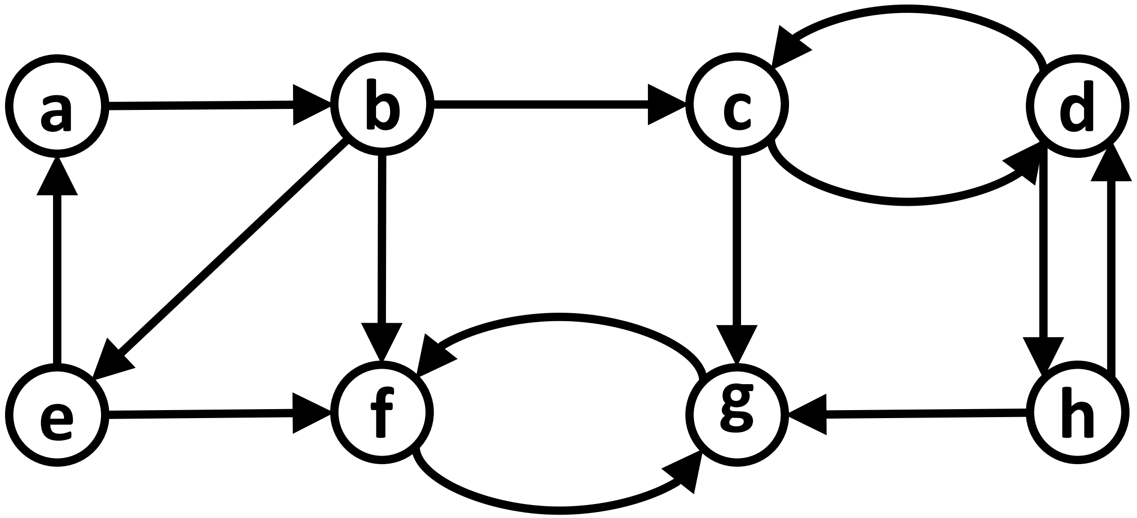
Components – Adjacency Matrix

- Adjacency matrix in a block-diagonal form:
  - Nonzero elements are in a block
  - All other elements are zero



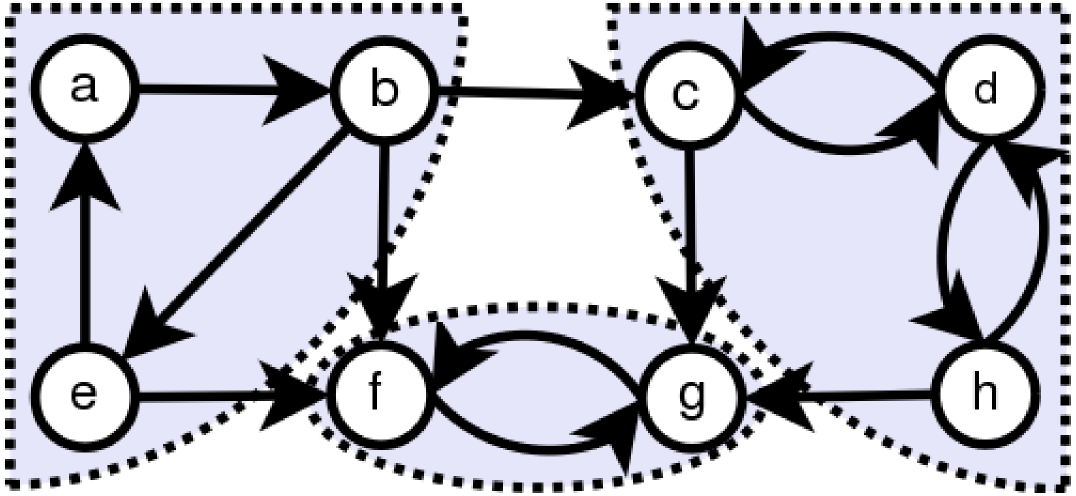
Components – Directed Networks

- ***Strongly connected component*** **(SCC)**
  - Any two nodes can be reachable by paths
- ***Weakly connected component*** **(WCC)**
  - Connected but not strongly connected



Components – Directed Networks

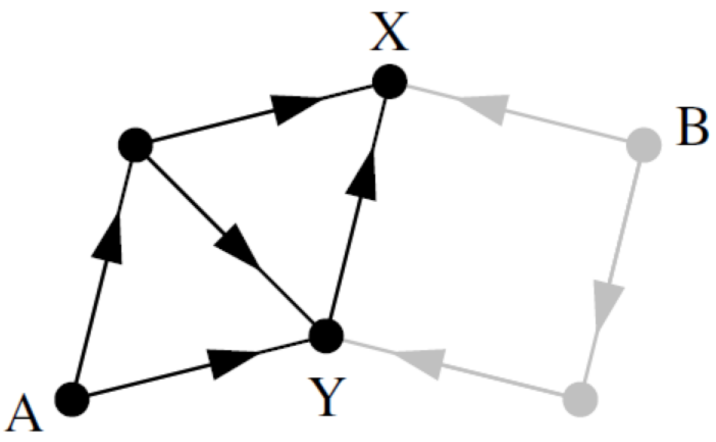
- Strongly connected component (SCC)
  - Any two nodes can be reachable by paths
- Weakly connected component (WCC)
  - Connected but not strongly connected



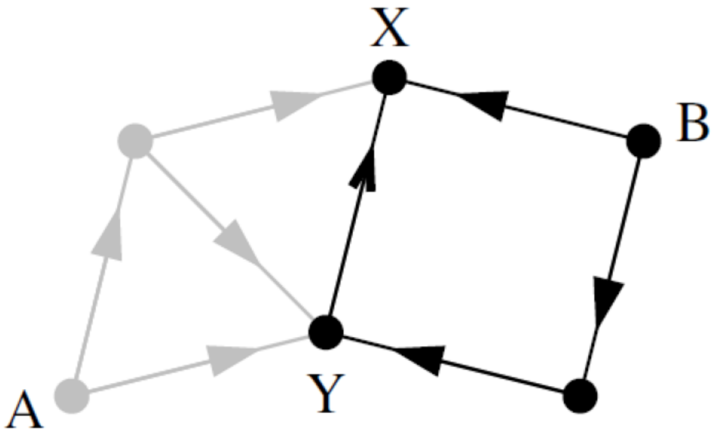
➡ **Every SCC with more than one node must contain at least one cycle.**

Components – Directed Networks

- Out-component (of an arbitrary node V)
  - a set of nodes that are reachable via directed paths starting from a given node V
- In-component (of an arbitrary node V)
  - a set of nodes from which there is a directed path to a given node V



[Out-component of node A]



[Out-component of node B]