COM
CITY UNIVERSITY OF HONG KONG
DEPARTMENT OF MEDIA AND COMMUNICATION (COM),
COLLEGE OF LIBERAL ARTS AND SOCIAL SCIENCES (CLASS)
CityU

# GE2234 Social Networks

for Media, Business and Technological Applications

# Lecture Note 2: Basic Concepts in Social Networks

By Dr. Wang Xiaohui, Vincent

# Outline

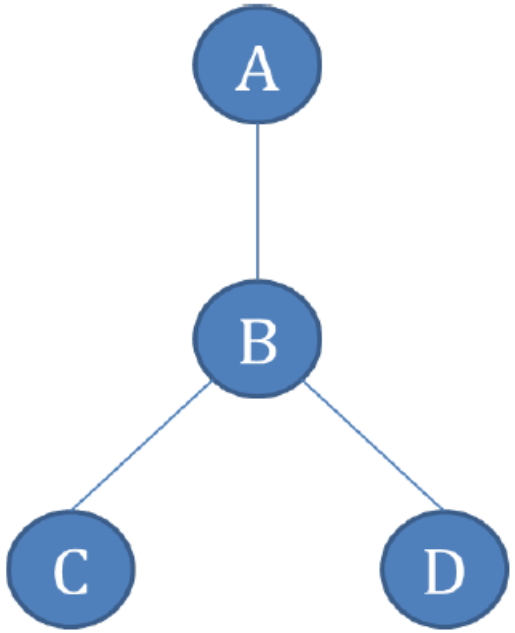
- Network Representation
- Network Measures

# Introduction to Network Science: Fundamental Concepts

# Nodes

- Nodes = Actors = Agents = Vertices
  - Individual or social structures (e.g., teams, organizations)
  - Content (e.g., webpage, article, keyword, video)
  - Physical locations or events

- Individual attributes attached to nodes help data analysis and visualization



# Edges

- Edges = Ties = Connections = Links

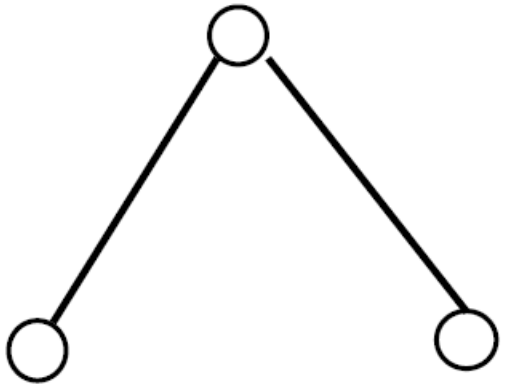
| Similarities | | | Social Relations | | | | Interactions | Flows |
|---|---|---|---|---|---|---|---|---|
| **Location** | **Membership** | **Attribute** | **Kinship** | **Other role** | **Affective** | **Cognitive** | e.g., | e.g., |
| e.g., | e.g., | e.g., | e.g., | e.g., | e.g., | e.g., | Sex with | Information |
| Same spatial and temporal space | Same clubs | Same gender | Mother of | Friend of | Likes | Knows | Talked to | Beliefs |
| | Same events | Same attitude | Sibling of | Boss of | Hates | Knows about | Advice to | Personnel |
| | etc. | etc. | | Student of | etc. | | Helped | Resources |
| | | | | Competitor of | | Sees as happy | Harmed | etc. |
| | | | | | | etc. | etc. | |

**Fig. 3.** A typology of ties studied in social network analysis.

# Undirected vs. Directed

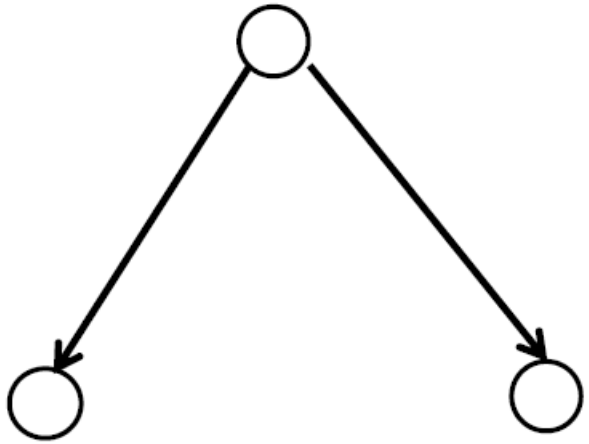
Undirected (= Symmetric)

- No clear origin and destination



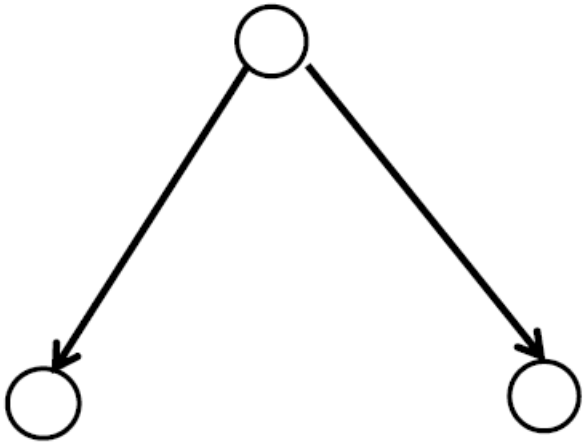
Directed (= Asymmetric)

- Follower vs. Followee
- Email sender vs. Receiver

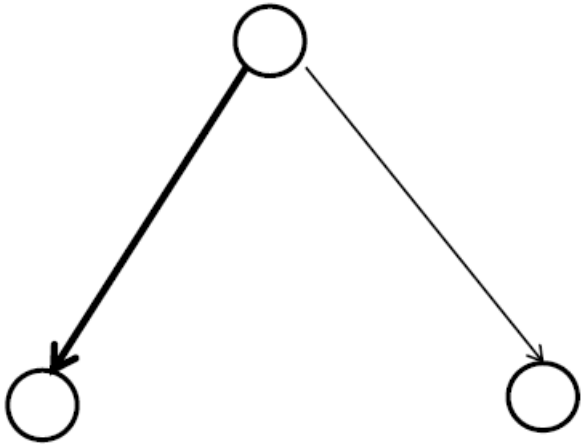


# Unweighted vs. Weighted

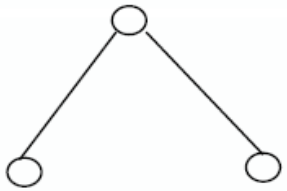
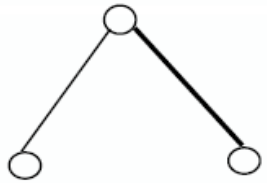
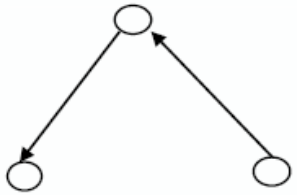
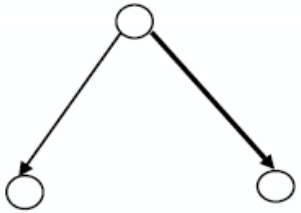
- Unweighted Edge (= Binary Edge)
  - If an edge exists or not



- Weighted Edge
  - Values assigned to edges indicating the strength/frequency of a tie
  - Thicker or darker lines



# 2 × 2 Combination

| | Unweighted | Weighted |
|---|---|---|
| Undirected |  |  |
| Directed |  |  |

# Examples

| | Unweighted | Weighted |
|---|---|---|
| **Undirected** | MTR network / Road Network | Co-authorship Network / Actor Network |
| **Directed** | Hyperlinks on WWW / Follower-followee network on Twitter | Communication network / Retweet network on Twitter |

# Beyond direction and strength, edges differ between:

- **Observed** (directly measured, e.g., hyperlinks) vs. **hidden** (inferred , e.g., co-occurrences)
- **Formal** (institutionally arranged) vs. **informal** (self-organized)
- **Strong-tie** (close friend) vs. **weak-tie** (occasional acquaintance)
- **Static** (unchanged over time) vs. **dynamic** (evolving)
- **Positive** (e.g., friending) vs. **negative** (e.g., de-friending)

- Innovative network analysis focuses on unobvious relationships such as hidden, informal, weak, and evolving edges.

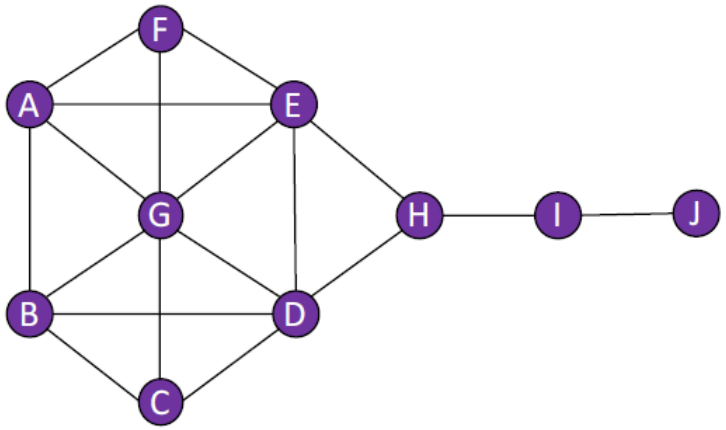
# Network Representation

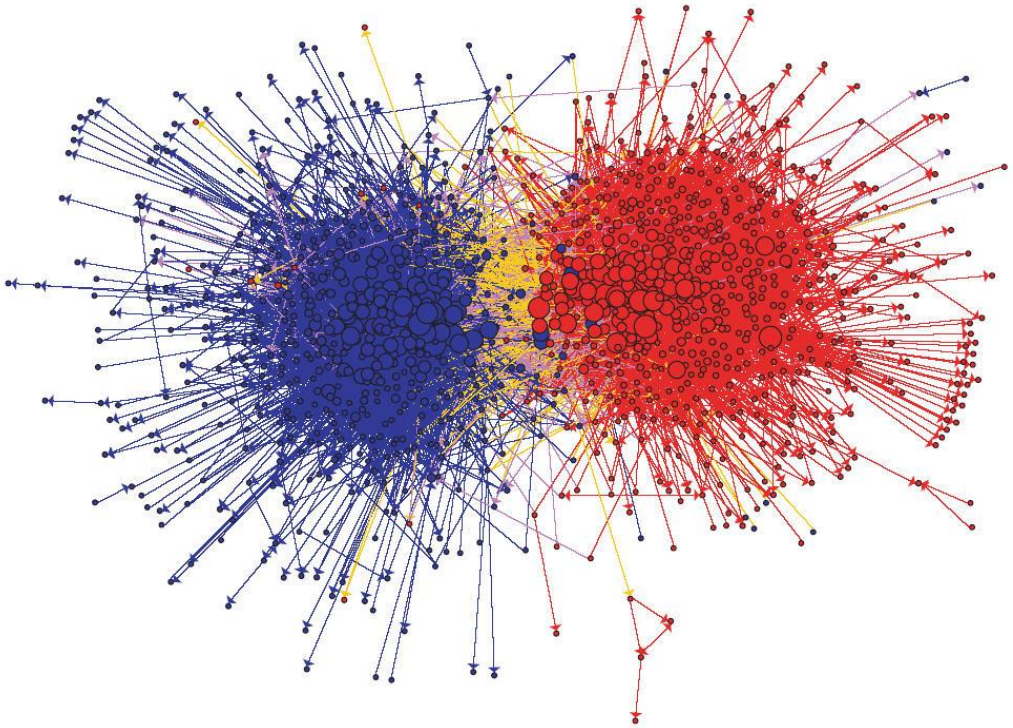
# Three Types of Representation

- Graph
- Edge List
- Adjacency Matrix

# Network Representation as Diagram (Graph)

A graph is a diagrammatical representation of some physical structure



# Network Representation as Edge Lists

- A list, or array, of edges

| From | To | Weight |
|---|---|---|
| A | B | 1 |
| A | C | 1 |
| B | C | 1 |

# Network Representation as Matrix

## Attribute-based data

| ID | Sex | Race | Education | … |
|---|---|---|---|---|
| id01 | 0 | Malay | Diploma | … |
| id02 | 1 | Indian | Postgraduate | … |
| id03 | 1 | Chinese | A-level | … |
| id04 | 0 | Eurasian | O-level | … |
| … | … | … | … | … |

## Relational data

| | id01 | id02 | id03 | … | $N_k$ |
|---|---|---|---|---|---|
| id01 | 0 | 3 | 6 | … | m |
| id02 | 2 | NA | 0 | … | n |
| id03 | 7 | 5 | 1 | … | … |
| … | … | … | … | … | … |
| $N_k$ | i | j | k | … | |

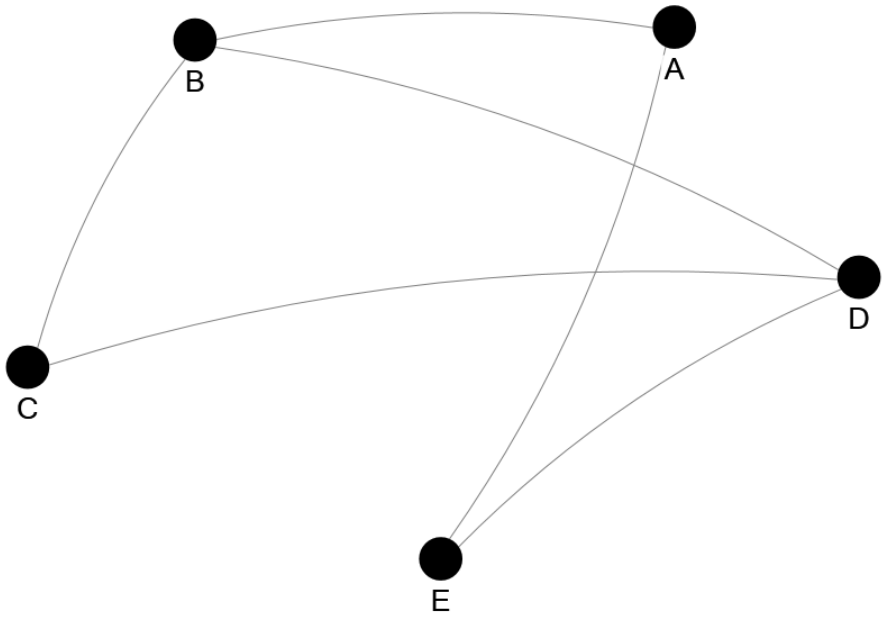
# Adjacency Matrix (AM)

- A square matrix in which the rows and columns represent nodes and an entry in row *i* and column *j* represent a link from *i* to *j*.

- By convention
  - The direction goes from the rows to the columns
  - Values in the entry are all positive
  - 0 indicates no ties
  - If the network is undirected, the matrix will be symmetric
  - If the network is reflexive (self loop via which a node is connected to itself), there will be values down the main diagonal

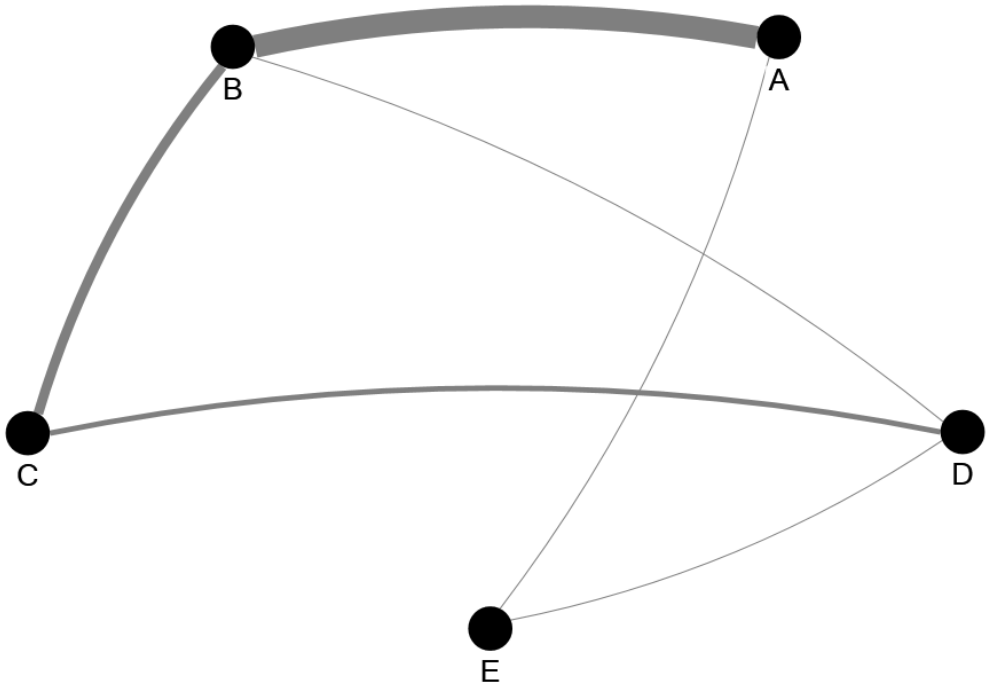
| | id01 | id02 | id03 | … | $N_k$ |
|---|---|---|---|---|---|
| id01 | 0 | 3 | 6 | … | m |
| id02 | 2 | NA | 0 | … | n |
| id03 | 7 | 5 | 1 | … | … |
| … | … | … | … | … | … |
| $N_k$ | i | j | k | … | |

# AM for Undirected and Unweighted Network



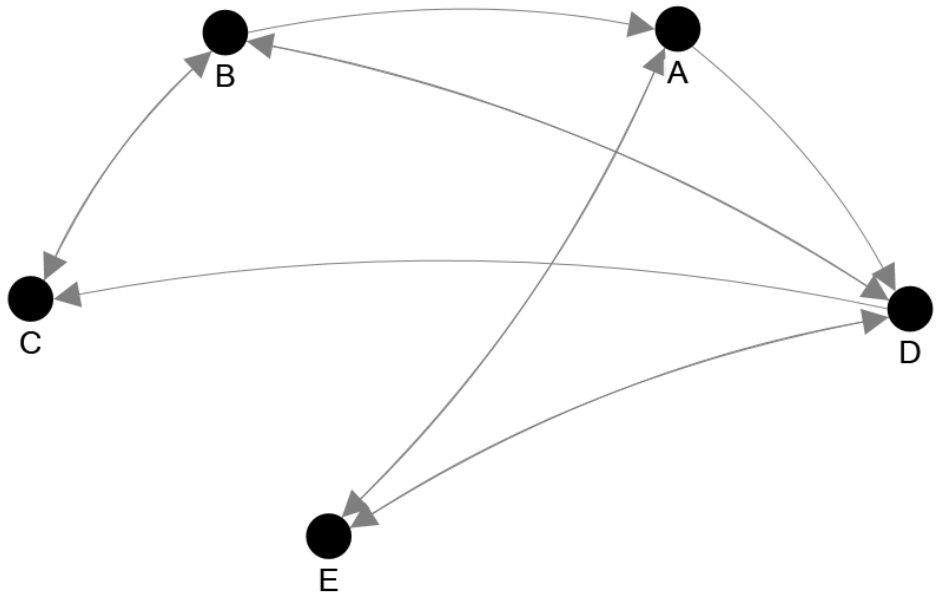

| | A | B | C | D | E |
|---|---|---|---|---|---|
| **A** | 0 | 1 | 0 | 0 | 1 |
| **B** | 1 | 0 | 1 | 1 | 0 |
| **C** | 0 | 1 | 0 | 1 | 0 |
| **D** | 0 | 1 | 1 | 0 | 1 |
| **E** | 1 | 0 | 0 | 1 | 0 |

# AM for Undirected and weighted Network



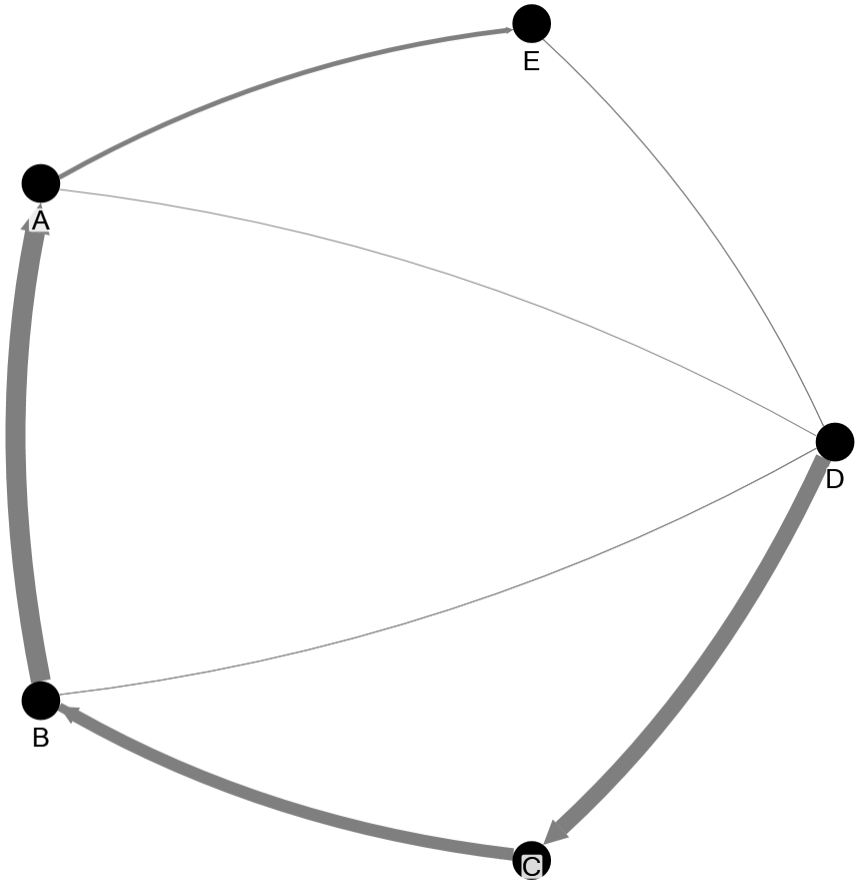

| | A | B | C | D | E |
|---|---|---|---|---|---|
| **A** | 0 | 6 | 0 | 0 | 1 |
| **B** | 6 | 0 | 3 | 1 | 0 |
| **C** | 0 | 3 | 0 | 2 | 0 |
| **D** | 0 | 1 | 2 | 0 | 1 |
| **E** | 1 | 0 | 0 | 1 | 0 |

# AM for directed and unweighted Network



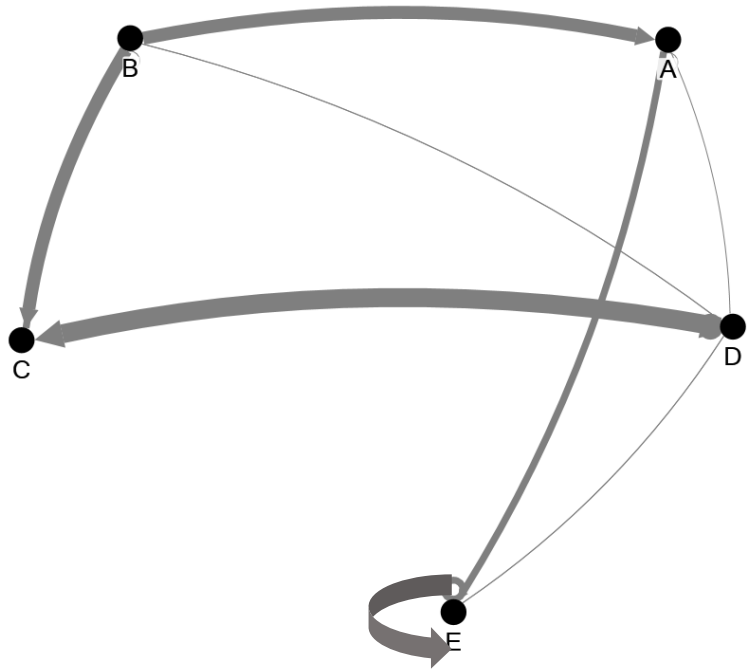

| | A | B | C | D | E |
|---|---|---|---|---|---|
| A | 0 | 0 | 0 | 1 | 1 |
| B | 1 | 0 | 1 | 1 | 0 |
| C | 0 | 1 | 0 | 0 | 0 |
| D | 0 | 1 | 1 | 0 | 1 |
| E | 1 | 0 | 0 | 1 | 0 |

# AM for directed and weighted Network




| | A | B | C | D | E |
|---|---|---|---|---|---|
| **A** | 0 | 0 | 0 | 1 | 2 |
| **B** | 6 | 0 | 3 | 1 | 0 |
| **C** | 0 | 4 | 0 | 0 | 0 |
| **D** | 0 | 1 | 5 | 0 | 1 |
| **E** | 1 | 0 | 0 | 1 | 0 |

# AM for Reflexive Network



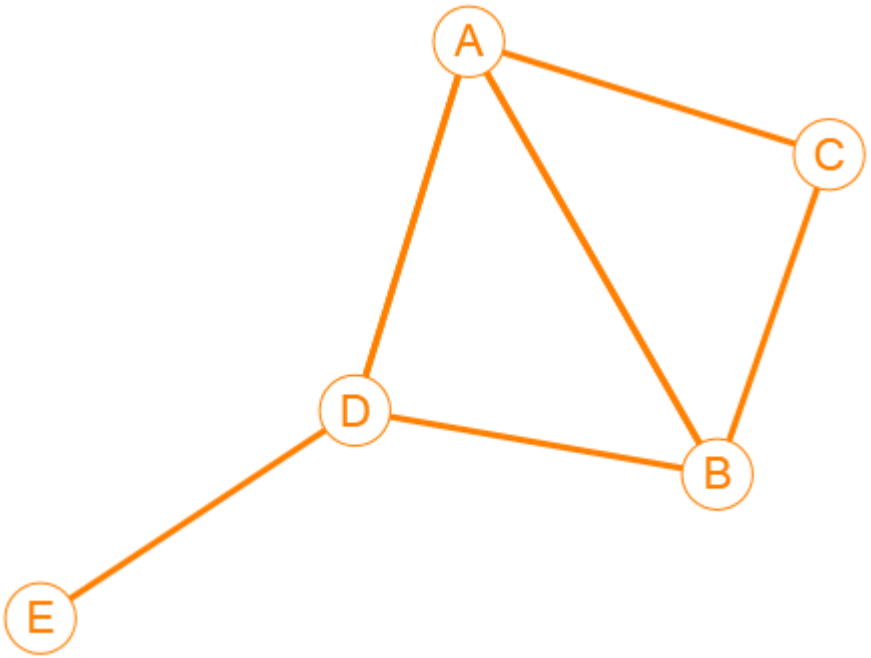

| | A | B | C | D | E |
|---|---|---|---|---|---|
| **A** | 0 | 0 | 0 | 1 | 2 |
| **B** | 3 | 0 | 3 | 1 | 0 |
| **C** | 0 | 2 | 0 | 0 | 0 |
| **D** | 0 | 1 | 4 | 0 | 1 |
| **E** | 1 | 0 | 0 | 1 | 3 |

# Which is better?

- Diagram
  - Direct sense of overall structure
  - Direct sense of nodes' positions in small network ($n < 30$)

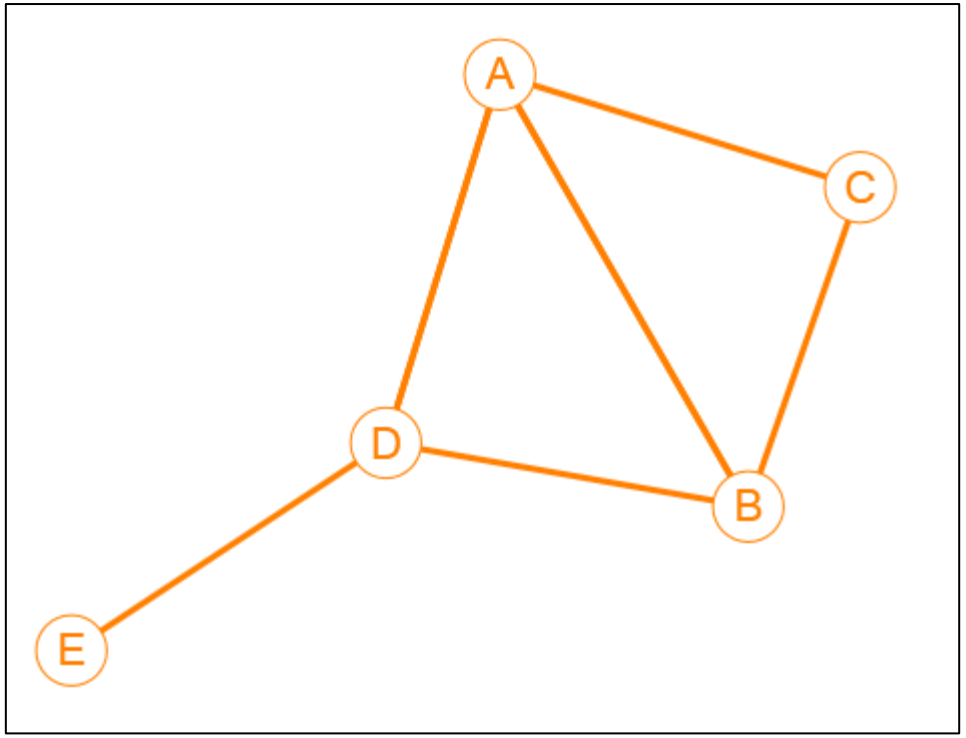
- Matrix/Edge List
  - Computation

# Hands-on exercise: Matrix/Graph Representation of Networks

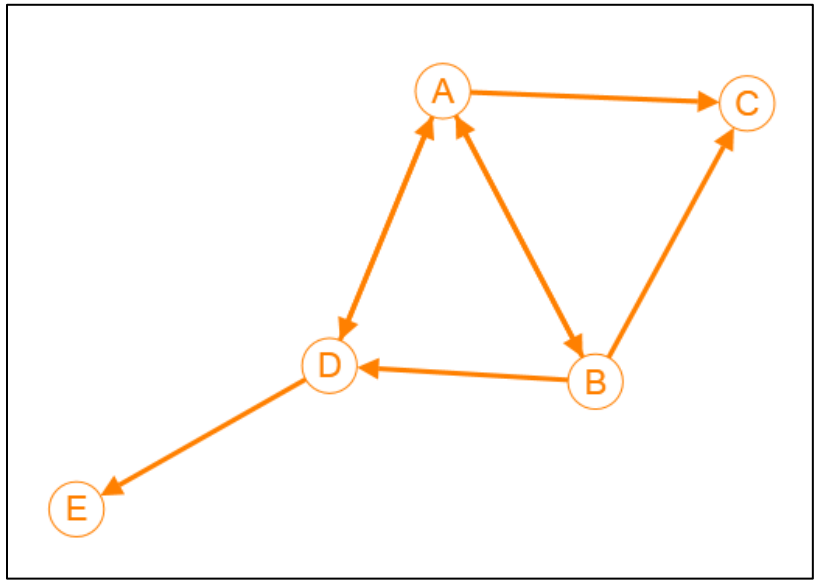



| Vertex 1 | Vertex 2 |
|---|---|
| | |
| | |
| | |
| | |
| | |
| | |

# Hands-on exercise: Matrix/Graph Representation of Networks




| | A | B | C | D | E |
|---|---|---|---|---|---|
| A | | | | | |
| B | | | | | |
| C | | | | | |
| D | | | | | |
| E | | | | | |




| Vertex 1 | Vertex 2 |
|---|---|
| A | B |
| A | C |
| A | D |
| B | A |
| B | C |
| B | D |
| D | A |
| D | E |

| | A | B | C | D | E |
|---|---|---|---|---|---|
| **A** | 0 | 1 | 1 | 1 | 0 |
| **B** | 1 | 0 | 1 | 1 | 0 |
| **C** | 0 | 0 | 0 | 0 | 0 |
| **D** | 1 | 0 | 0 | 0 | 1 |
| **E** | 0 | 0 | 0 | 0 | 0 |

# How to Define Nodes/Edges?

**Node Attributes:**

- Size (degree or another measure of magnitude)
- Color (node categories, e.g., men vs. women)
- Shape (second dimension of node categories, e.g., young, mid-age, and old)

**Edge Attributes:**

- Arrow (tie direction)
- Size (tie strength)
- Color (tie categories, e.g., positive vs. negative)

# Data Format of Node/Edge Attributes

**Node Attribute Data**: a rectangular table of N rows and k columns (k = N of attributes)

| Node | Attr1 | Attr2 | ... |
|---|---|---|---|
| A | 2 | ... | ... |
| B | 1 | ... | ... |
| C | 2 | ... | ... |
| ... | ... | ... | ... |

**Edge Attribute Data**: a rectangular table of up to N(N-1)/2 rows and k columns

| From | To | Attr1 | ... |
|---|---|---|---|
| A | B | 10 | ... |
| A | C | 22 | ... |
| C | D | 5 | ... |
| ... | ... | ... | ... |

# Network Basics: Network Level Metrics

# Network Size

- Network size is an important structural property, reflecting the boundary of the network.
  - Size of a school, size of a company, size of personal social network…
- Size may not be the most interesting network indicator and may only reflect the boundary of the network, but it is the primary network indicator.
- Evidence exists that 150 is an optimal size for a stable, and **cohesive** human group (Dunbar, 1993).

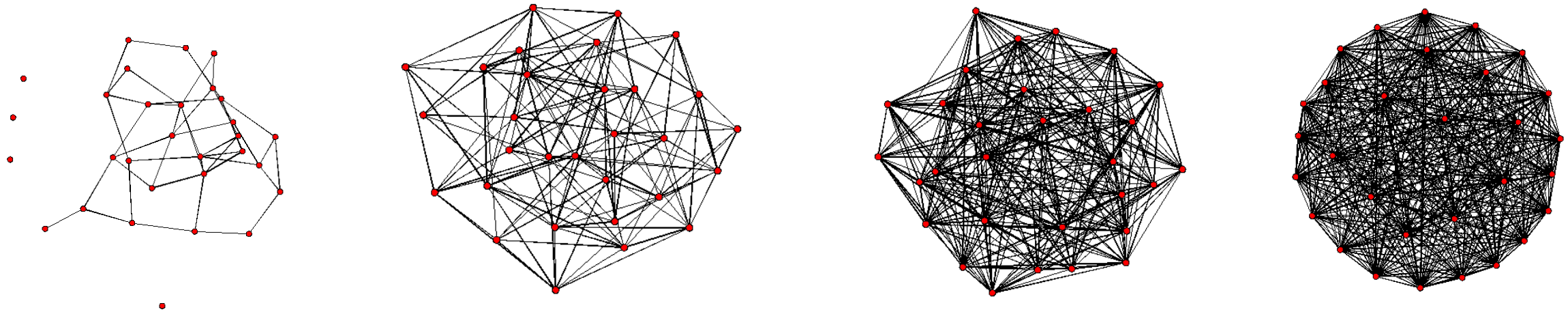
# Density

- Density refers to the proportion of all possible ties that are actually present
- The more nodes are connected to one another, the denser the network will be

$$D = \frac{l}{N(N-1)}$$

where *l* is the number of links in the network and *N* is network size.

# Density



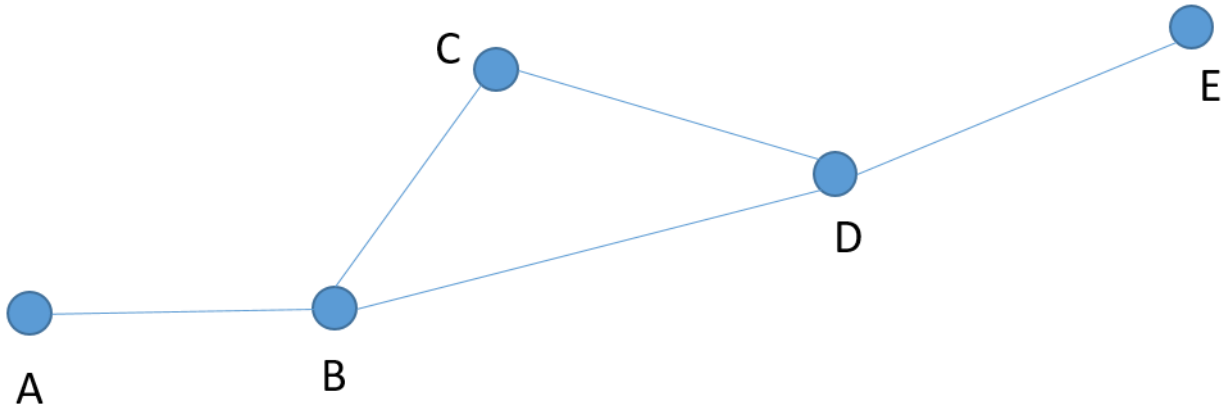
Network Size = 30; Density = 0.1, 0.3, 0.7,  and 1

# Path / Walk

- A **path** is a sequence of nodes and edges starting with one node and ending with another, tracing the indirect connection between the two. On a path, you never go backwards or revisit the same node twice.
  - Example: a -> b -> c -> d

- A **walk** is any sequence of nodes and edges, and may go backwards.
  - Example: a -> b -> c -> b -> c -> d

- There may be multiple paths and walks between two nodes. The length of a walk/path is the number of occurrences of lines in it.




Path A - E:

A -> B -> D -> E

A -> B -> C -> D -> E

Walk A - E:

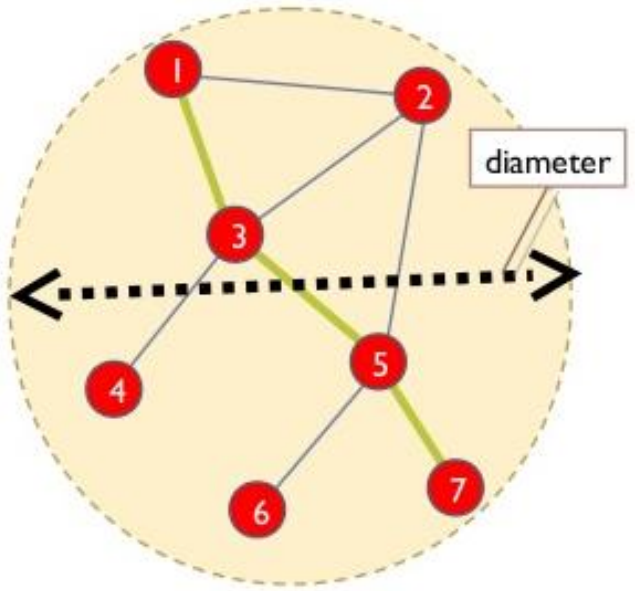
A -> B -> C -> D -> B -> D -> E

# Distance / Diameter / Average Path Length

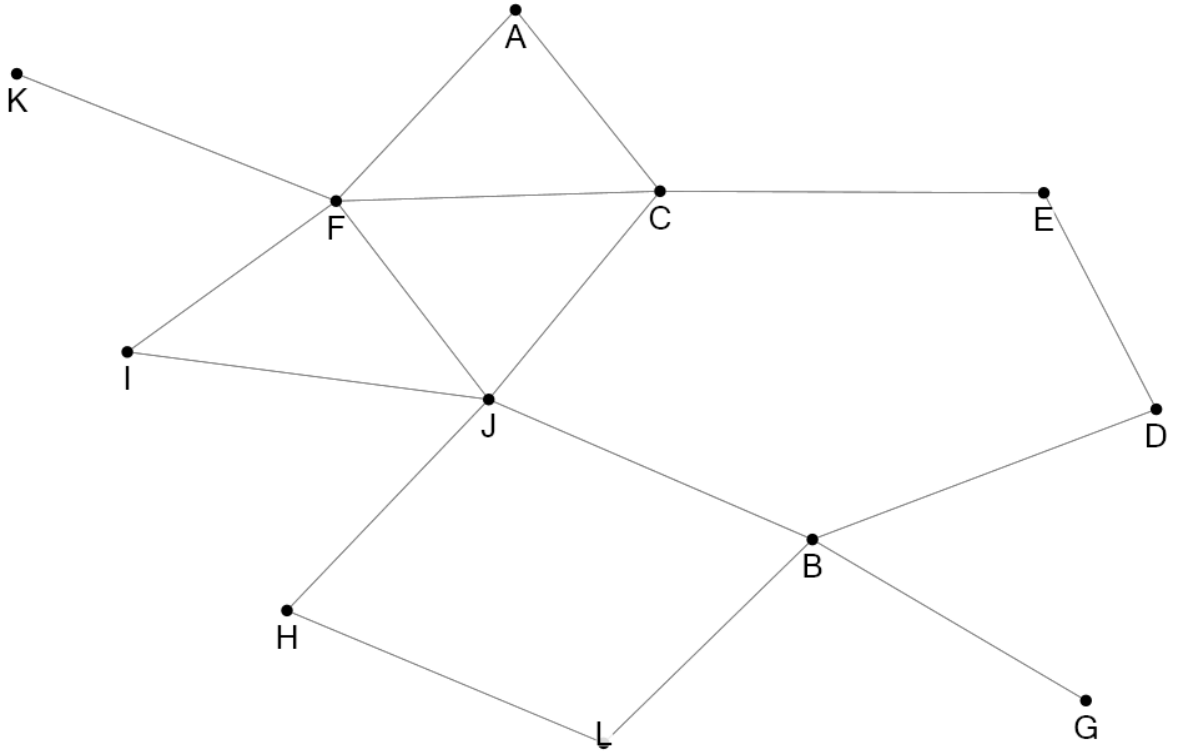
- The **Geodesic distance** or simply the **Distance** is the number of edges in the *shortest* path from one node to another.

- **Diameter** of a connected network is the *maximum* geodesic distance in the network
  - Networks with the same number of nodes, and even the same density (number of links), can have different diameters

- **Average Path Length:** Average distance between all the nodes in a network.

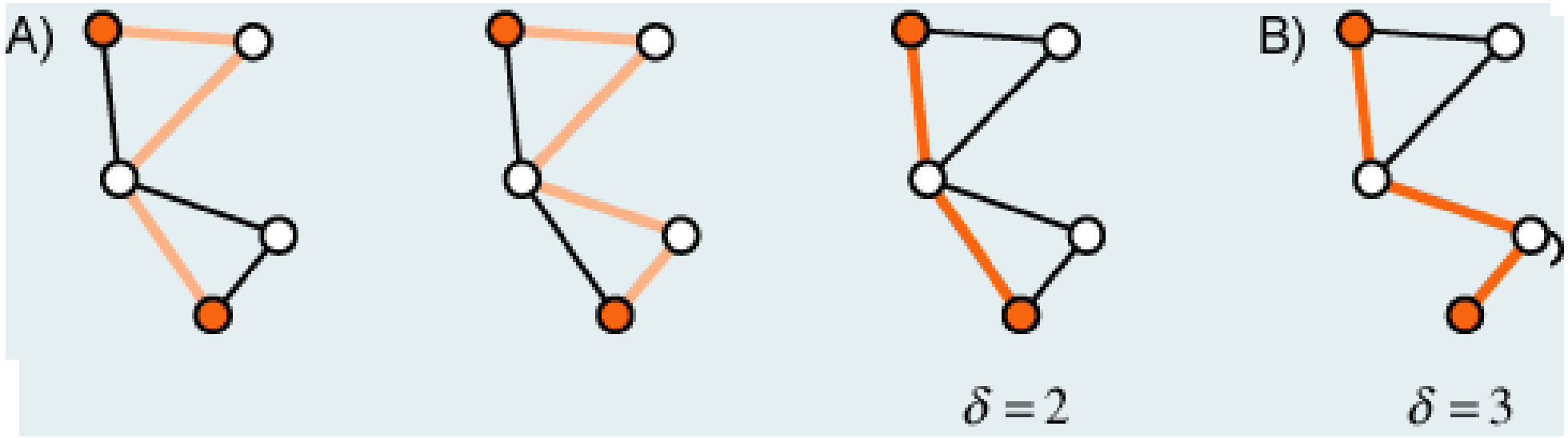


# An Illustration on geodesic distance

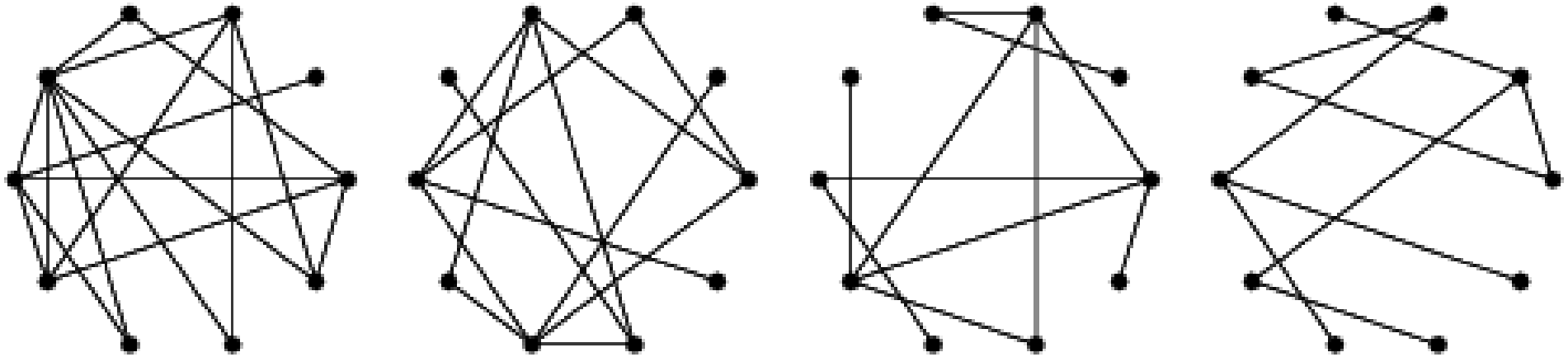



- The geodesic distance between K and B is:_____ (3)
  - K -> F -> J -> B
  - K -> F -> I -> J -> B
  - ……
  - K -> F -> J -> C -> E -> D -> B
  - K -> F -> A -> C -> J -> H -> L -> B

# An Illustration on Diameter



Diameter and **Average Path Length** provide an overall indication of the structure of the network. Low diameter or low average path length indicate a cohesive network with more clustering.

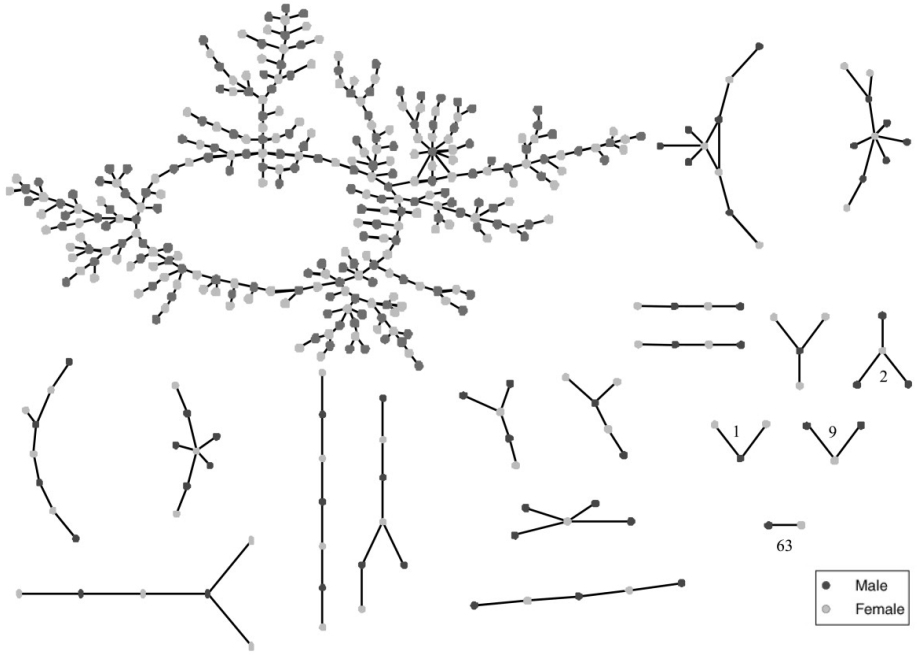


# Connectivity

- Whether every node can reach every other node by path in a network.
- A graph is connected if there is a path between every pair of nodes in the graph

## Component

- A component consists of a subgroup of individuals, whereby all the individuals are connected to one another by at least one path.
- Number of components
- Isolates




Components in a high school network
Source: Bearman et al., 2004