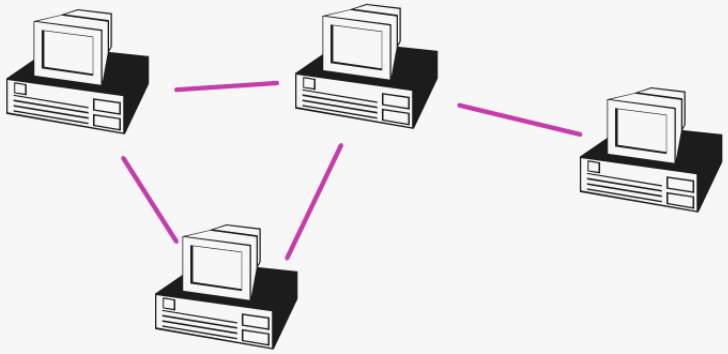
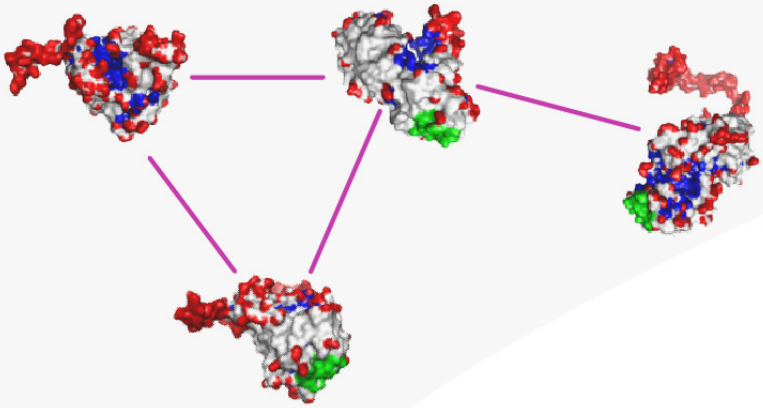
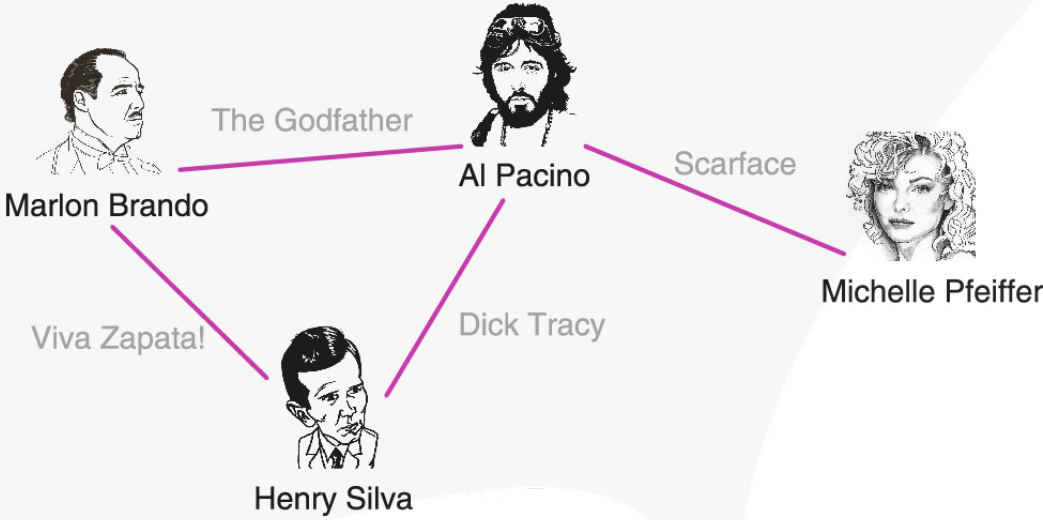
# ISE 599
# Special Topics Applied Predictive Analytics

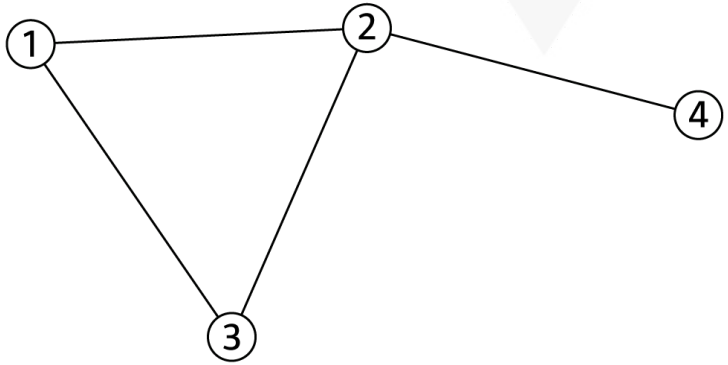
Mayank Kejriwal

# Networks give us a common language to study ‘structure’…








N=4

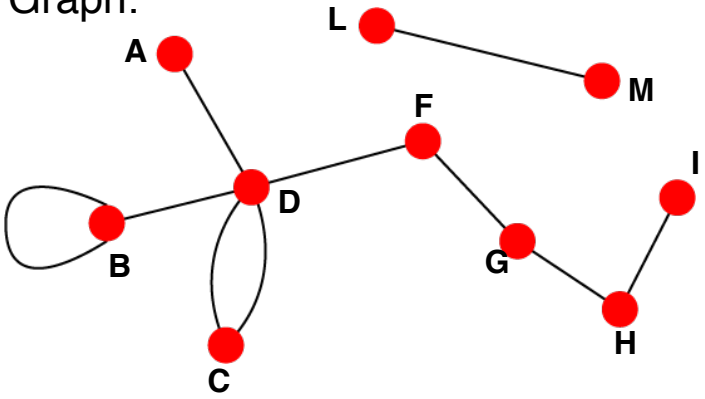
L=4

# Degree, Average Degree and Degree Distribution

# Undirected

Links: undirected (*symmetrical*)

Graph:

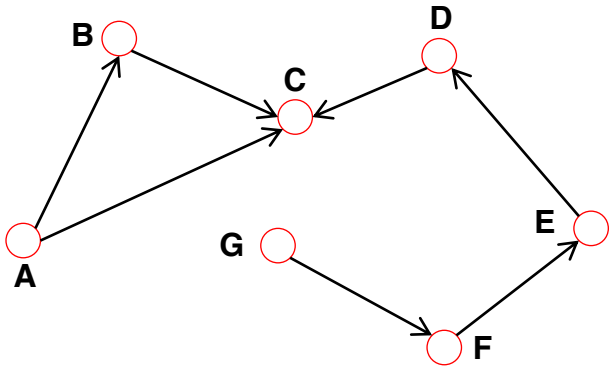



**Undirected links :**
coauthorship links
Actor network
protein interactions

# Directed

Links: directed (*arcs*).

Digraph = directed graph:




*An undirected link is the superposition of two opposite directed links.*

**Directed links :**
URLs on the www
phone calls
metabolic reactions

## BRIEF STATISTICS REVIEW

Four key quantities characterize a sample of $N$ values $x_1, \ldots, x_N$:

*Average (mean)*:

$$\langle x \rangle = \frac{x_1 + x_2 + \ldots + x_N}{N} = \frac{1}{N} \sum_{i=1}^{N} x_i$$

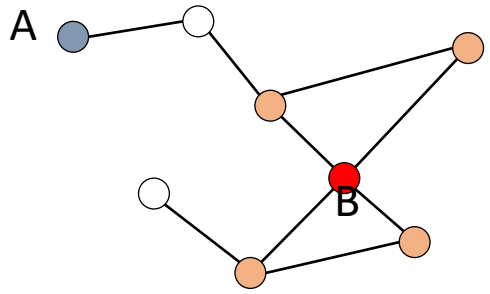
*The $n^{th}$ moment*:

$$\langle x^n \rangle = \frac{x_1^n + x_2^n + \ldots + x_N^n}{N} = \frac{1}{N} \sum_{i=1}^{N} x_i^n$$

*Standard deviation*:

$$\sigma_x = \sqrt{\frac{1}{N} \sum_{i=1}^{N} \left( x_i - \langle x \rangle \right)^2}$$

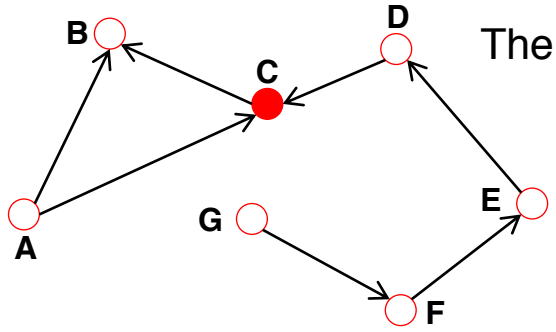
## Undirected




Node degree: the number of links connected to the node.

$$k_A = 1 \qquad k_B = 4$$

## Directed



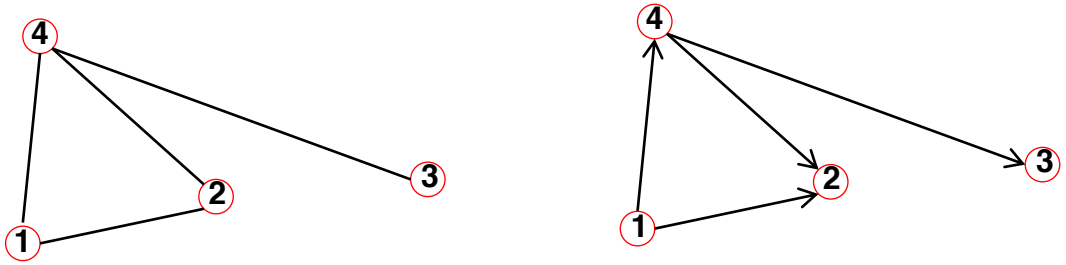

In *directed networks* we can define an in-degree and out-degree.

The (total) degree is the sum of in- and out-degree.

$$k_C^{in} = 2 \qquad k_C^{out} = 1 \qquad k_C = 3$$

Source: a node with $k^{in} = 0$; Sink: a node with $k^{out} = 0$.

# Adjacency matrix




**$A_{ij}=1$** if there is a link between node *i* and *j*

**$A_{ij}=0$** if nodes *i* and *j* are not connected to each other.

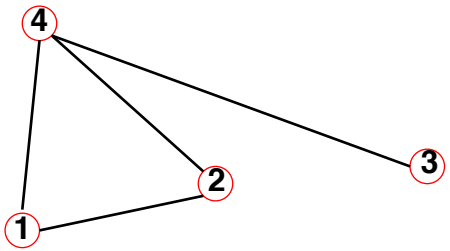
$$A_{ij} = \begin{pmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix} \qquad A_{ij} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{pmatrix}$$

Note that for a directed graph (right) the matrix is not symmetric.

$A_{ij} = 1$ if there is a link pointing from node *j* and *i*

$A_{ij} = 0$ if there is no link pointing from *j* to *i*.

## Undirected




$$A_{ij} = \begin{pmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix}$$

$$A_{ij} = A_{ji}$$

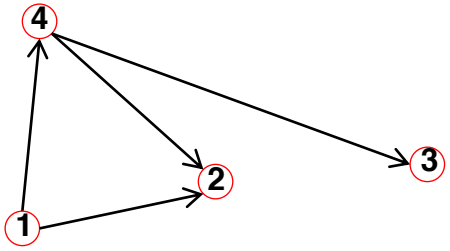
$$A_{ii} = 0$$

$$k_i = \sum_{j=1}^{N} A_{ij}$$

$$k_j = \sum_{i=1}^{N} A_{ij}$$

$$L = \frac{1}{2}\sum_{i=1}^{N} k_i = \frac{1}{2}\sum_{ij} A_{ij}$$

## Directed




$$A_{ij} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{pmatrix}$$

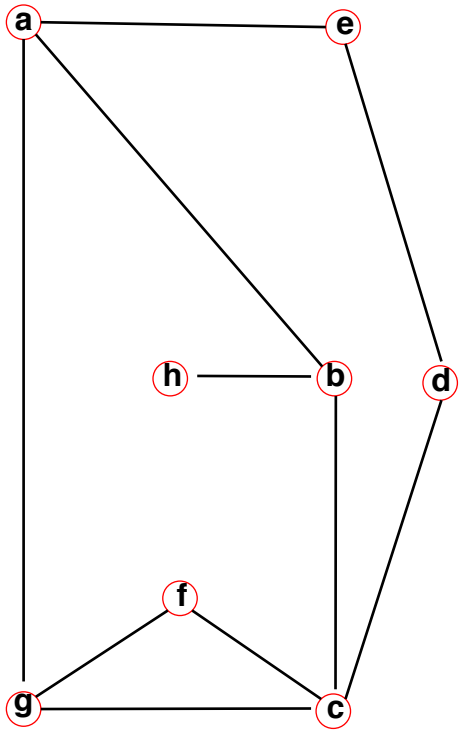
$$A_{ij} \neq A_{ji}$$

$$A_{ii} = 0$$

$$k_i^{in} = \sum_{j=1}^{N} A_{ij}$$

$$k_j^{out} = \sum_{i=1}^{N} A_{ij}$$

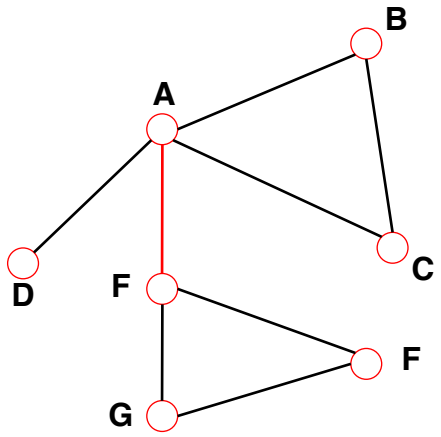
$$L = \sum_{i=1}^{N} k_i^{in} = \sum_{j=1}^{N} k_j^{out} = \sum_{i,j} A_{ij}$$

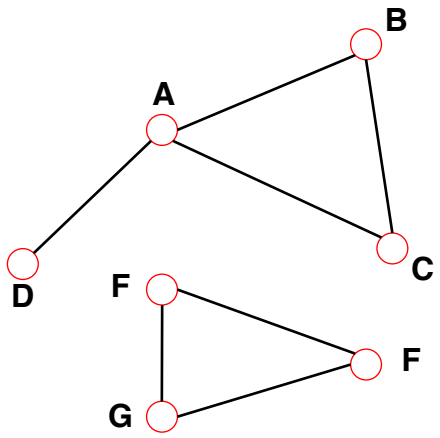



| | a | b | c | d | e | f | g | h |
|---|---|---|---|---|---|---|---|---|
| a | 0 | 1 | 0 | 0 | 1 | 0 | 1 | 0 |
| b | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 1 |
| c | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 0 |
| d | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 |
| e | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| f | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 |
| g | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| h | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |

# CONNECTEDNESS

Connected (undirected) graph: any two vertices can be joined by a path.
A disconnected graph is made up by two or more connected components.




Largest Component:
**Giant Component**

The rest: **Isolates**

Bridge: if we erase it, the graph becomes disconnected.