Cocitation
bibliographic coupling

SYSM 6302

CLASS 2

More mathematically:

Graph $G = (V, E)$

A finite vertex set:

$V = \{v_1, v_2, \dots, v_n\}$

A **relation** E on V:

$E = \{\dots, \underbrace{(v_i, v_j)}_{e_k}, \dots\} \underbrace{\subseteq V \times V}$

A subset of the Cartesian product

So far we've discussed undirected graphs, in which the presence of

$(v_i, v_j) \in E \iff (v_j, v_i) \in E \implies$ **symmetric relation**

The cardinality of the vertex and edge sets are:

$|V|$ is the **order** of the graph

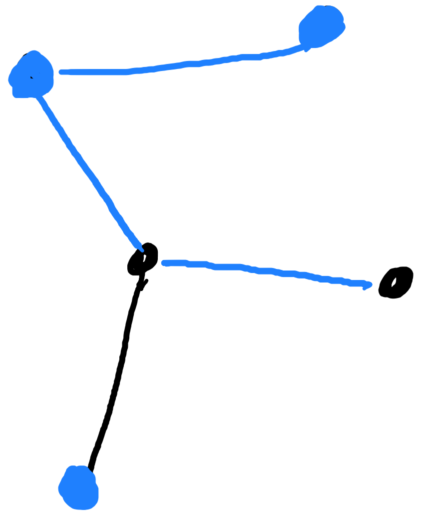
$|E|$ is the **size** of the graph

Nodes u and v are **adjacent** if $\overbrace{(u,v)}^{u \,\&\, v \text{ are end points of this edge}} \in E$.

Edges (t,s) and (u,v) are adjacent if the edges are distinct and share a common node.

A node is $v \in V$ is **incident** to an edge $e \in E$ if v is an endpoint of e.

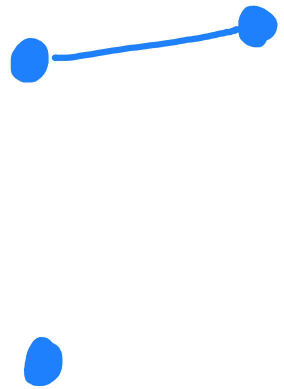
Given $G = (V, E)$



$\tilde{G} = (\tilde{V}, \tilde{E})$, where $\tilde{V} \subseteq V$ and $\tilde{E} \subseteq E$

is a subgraph of G

It is an induced subgraph if $\tilde{E}$ only contains all edges with endpoints in $\tilde{V}$, i.e.,

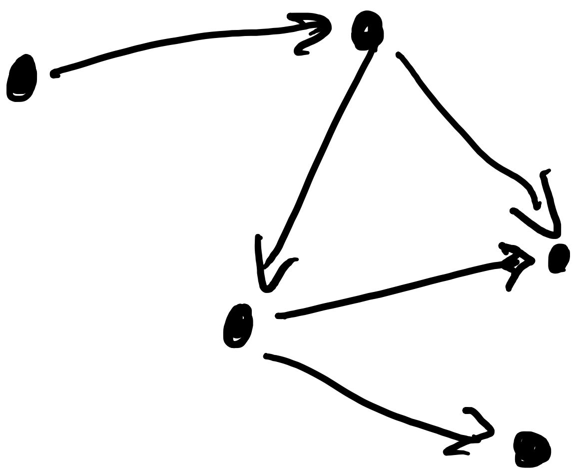


$$\tilde{E} = \{(u,v) \mid u, v \in \tilde{V}\}$$

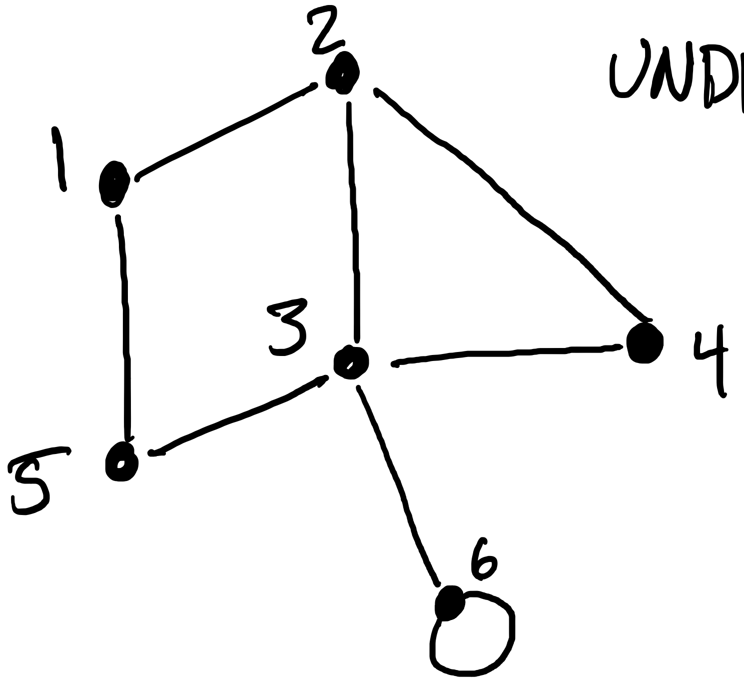
A directed graph assigns orientation to the edges, thus

$G = (V, E)$, but E is generally not symmetric

A directed edge is often called an arc.



Note that undirected graphs are a special class of directed graphs



UNDIRECTED

$$\begin{pmatrix} 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 2 \end{pmatrix}$$

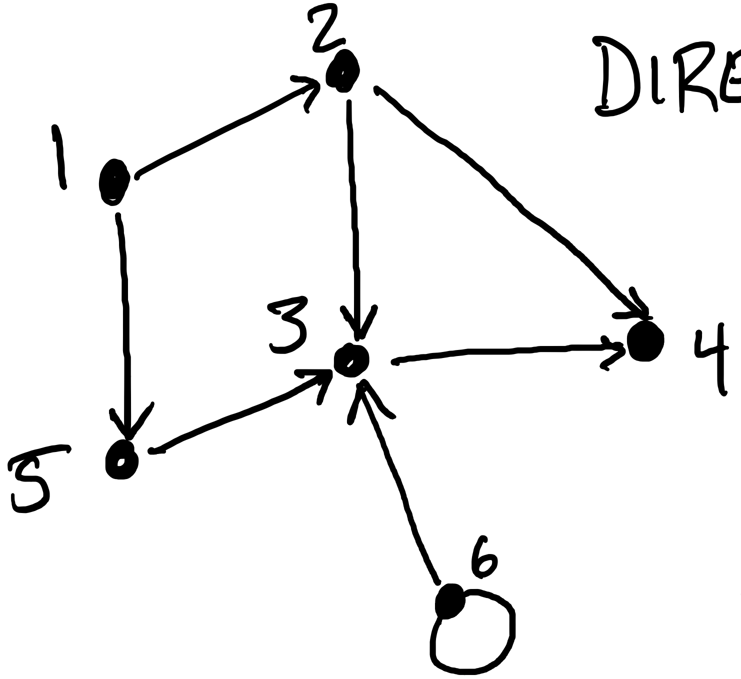
Symmetric Matrix!

GRAPH

$G = (V, E)$

ADJACENCY MATRIX

$$A_{ij} = \begin{cases} 1 & (i,j) \in E \\ 0 & (i,j) \notin E \end{cases}$$



DIRECTED

$$
\begin{array}{cc}
 & \text{FROM} \\
\text{TO} &
\begin{array}{c|cccccc}
 & 1 & 2 & 3 & 4 & 5 & 6 \\
\hline
1 & & & & & & \\
2 & & & & & & \\
3 & & & & & & \\
4 & & & & & & \\
5 & & & & & & \\
6 & & & & & &
\end{array}
\end{array}
\qquad
\begin{array}{c}
 \\
\begin{array}{c}
1 \leftarrow \\ 2 \leftarrow \\ 3 \leftarrow \\ 4 \leftarrow \\ 5 \leftarrow \\ 6 \leftarrow
\end{array}
\end{array}
$$

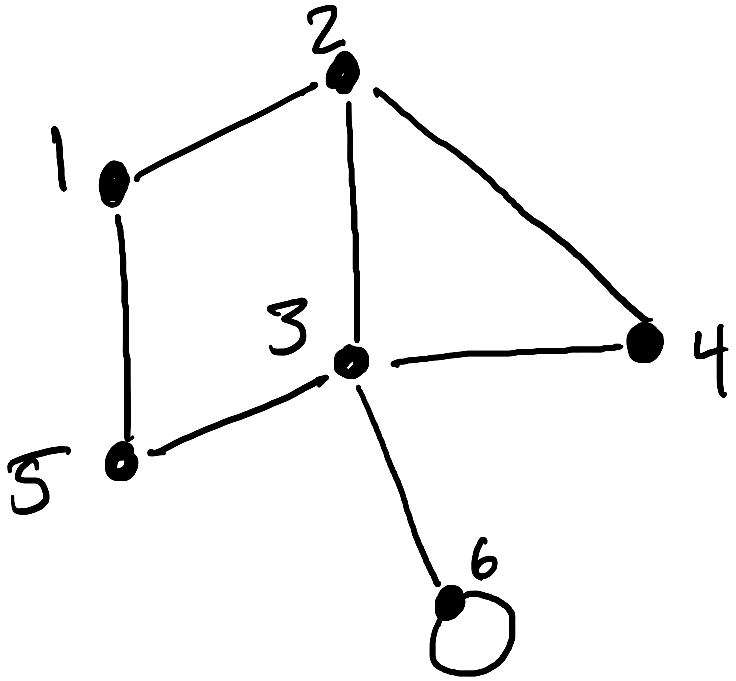
$$
\begin{array}{cc}
 & \begin{array}{cccccc} 1 & 2 & 3 & 4 & 5 & 6 \\ \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow \end{array} \\
\begin{array}{c} 1 \leftarrow \\ 2 \leftarrow \\ 3 \leftarrow \\ 4 \leftarrow \\ 5 \leftarrow \\ 6 \leftarrow \end{array} &
\begin{pmatrix}
0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 1 & 1 \\
0 & 1 & 1 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1
\end{pmatrix}
\end{array}
$$

GRAPH

$$G = (V, E)$$

ADJACENCY MATRIX

$$A_{ij} = \begin{cases} 1 & (j,i) \in E \\ 0 & (j,i) \notin E \end{cases}$$



GRAPH

$G = (V, E)$

$$\begin{bmatrix} [2, 5] \\ [1, 3, 4] \\ [2, 4, 5, 6] \\ [2, 3] \\ [1, 3] \\ [3, 6] \end{bmatrix}$$

ADJACENCY LIST

provides a compact represent for large, sparse graphs

$|V| \sim |E|$

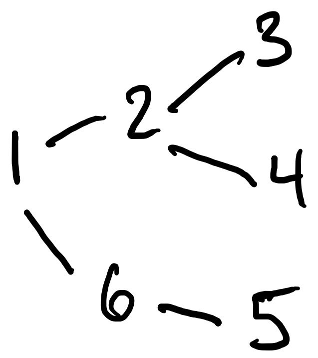
INCIDENCE MATRIX

$$B \in \mathbb{R}^{|V| \times |E|}, \quad B_{ij} = \begin{cases} 1 & \text{if node } i \text{ is incident to edge } j \\ 0 & \text{otherwise} \end{cases}$$

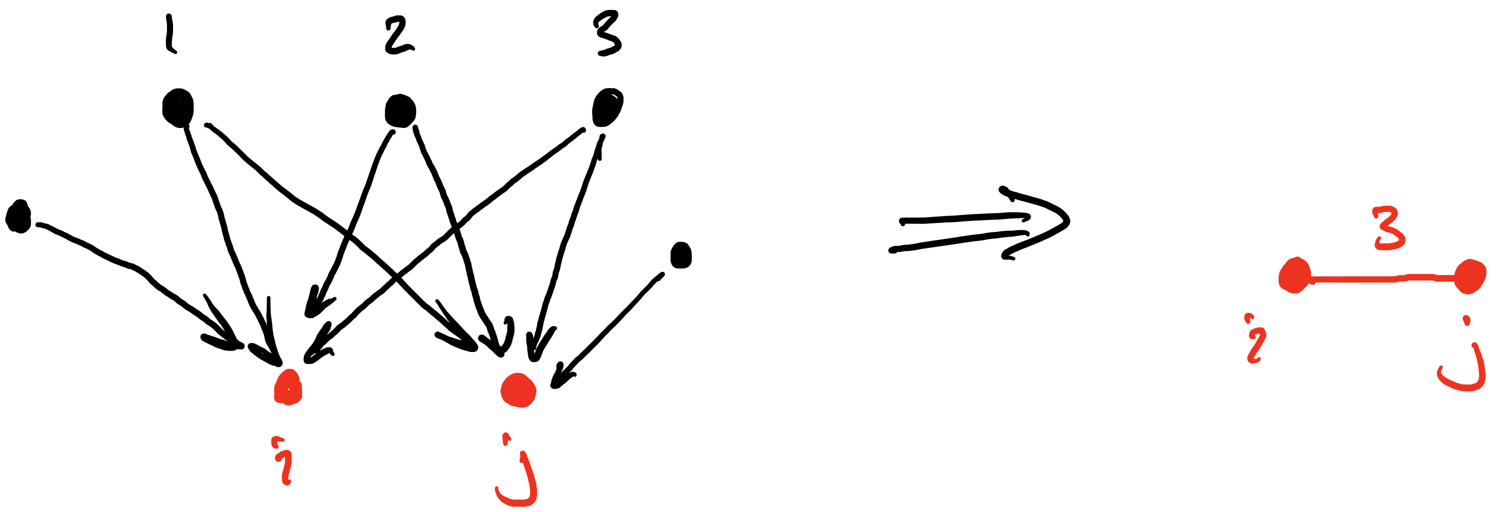
Many file formats:

<u>EDGELIST:</u>

1	2
2	3
2	4
5	6
6	1



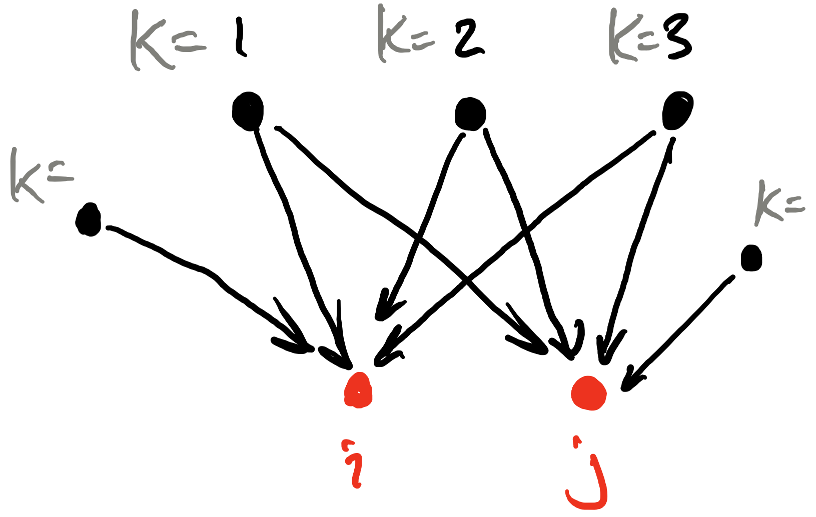
Cocitation: Build a network (undirected) that captures the similarity of connectivity in the original (directed) graph



Nodes i and j are similar because there are "many" nodes that point to both of them.

*for unweighted graphs

Cocitation: Build a network (undirected) that captures the similarity of connectivity in the original (directed) graph



cocitation adjacency

$$C_{ij} = \sum_{\substack{\text{over} \\ \text{nbrs of} \\ i \,\&\, j}} 1 \overset{*}{=} \sum_{k=1}^{n} A_{ik} A_{jk}$$

sum over all nodes

does k point to i?

does k also point to j?

$$= \sum_{k=1}^{n} A_{ik} [A^T]_{kj} \Rightarrow C = AA^T$$

there is one step this is leaving out! Check book.

Example