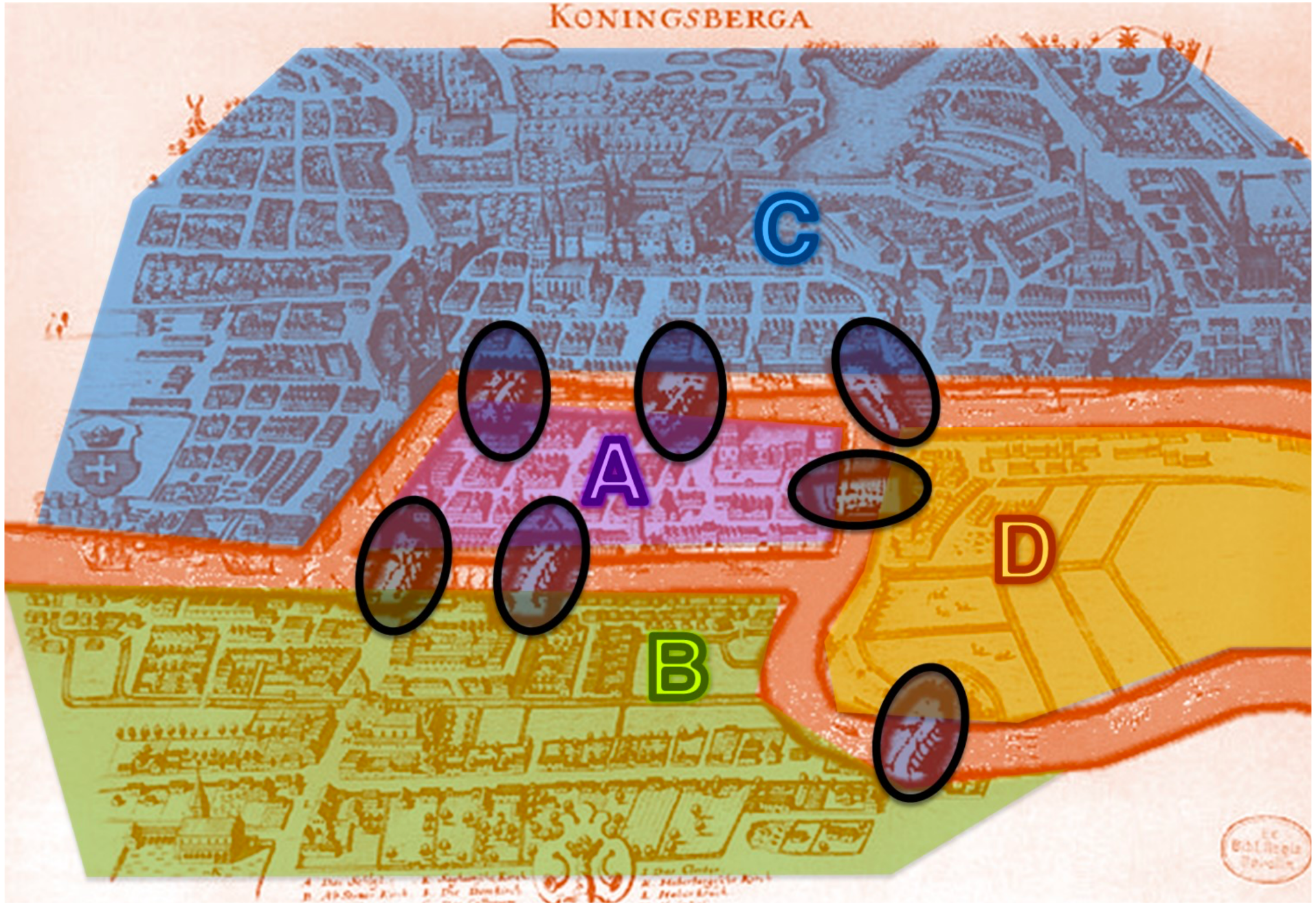
Social Network Analysis

#3 Graphs

© 2020 T. Erseghe

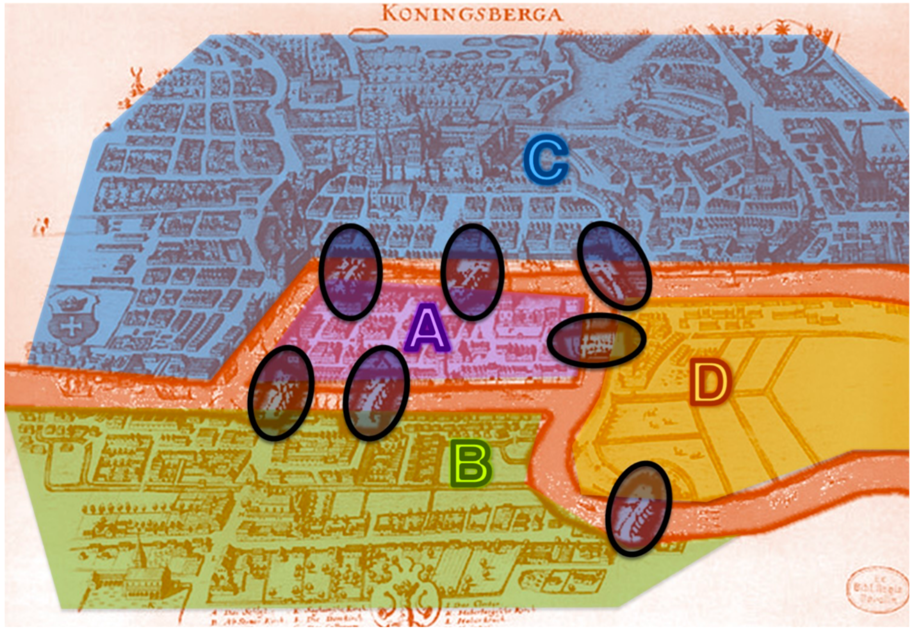
MiME.

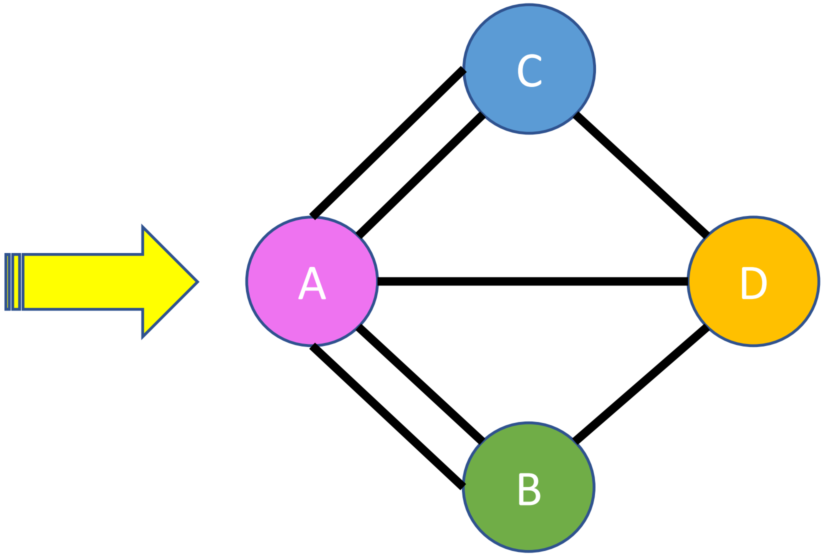
Euler & the 7 bridges of Königsberg (1736)



How to walk through the city by crossing each bridge only once?

Networks as graphs





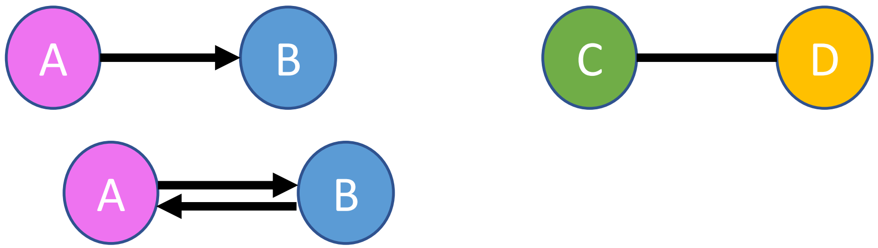
Graph $\mathcal{G}$ ($\mathcal{V}$,$\mathcal{E}$): network

- ❑ Vertices (set $\mathcal{V}$) : nodes, people, concepts
- ❑ Edges (set $\mathcal{E}$): links, relations, associations

↑ *mathematics* ↑ *technology* ↑ *social psychology* ↑ *social cognition*

Directed versus undirected

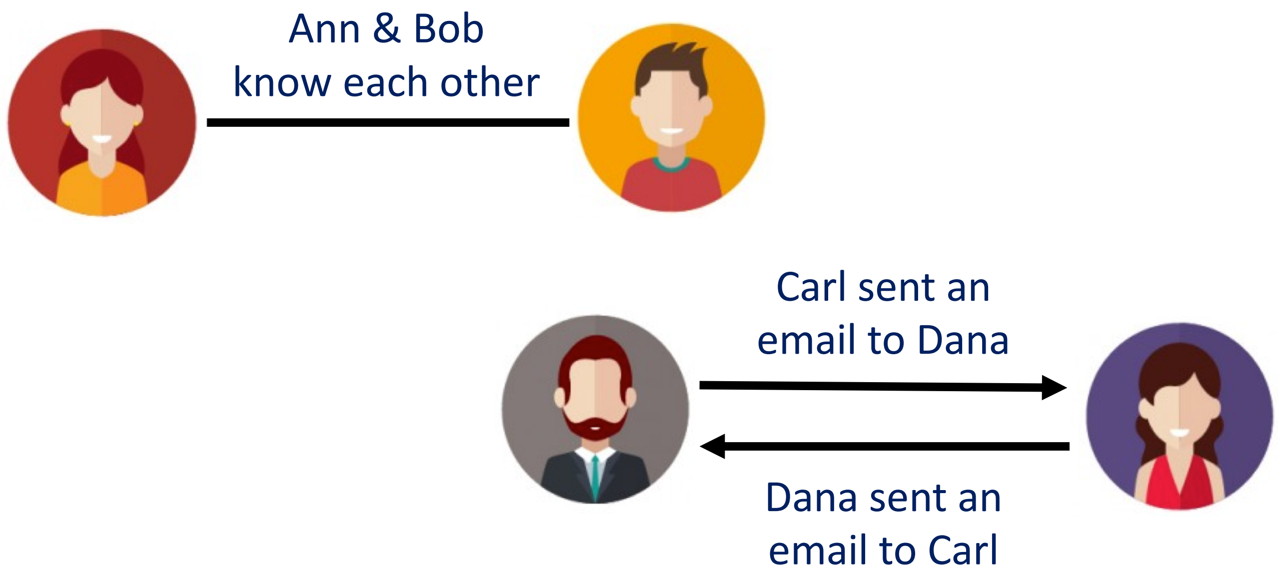
- A connection relationship can have a privileged direction or can be mutual
  - Either a directed or an undirected link



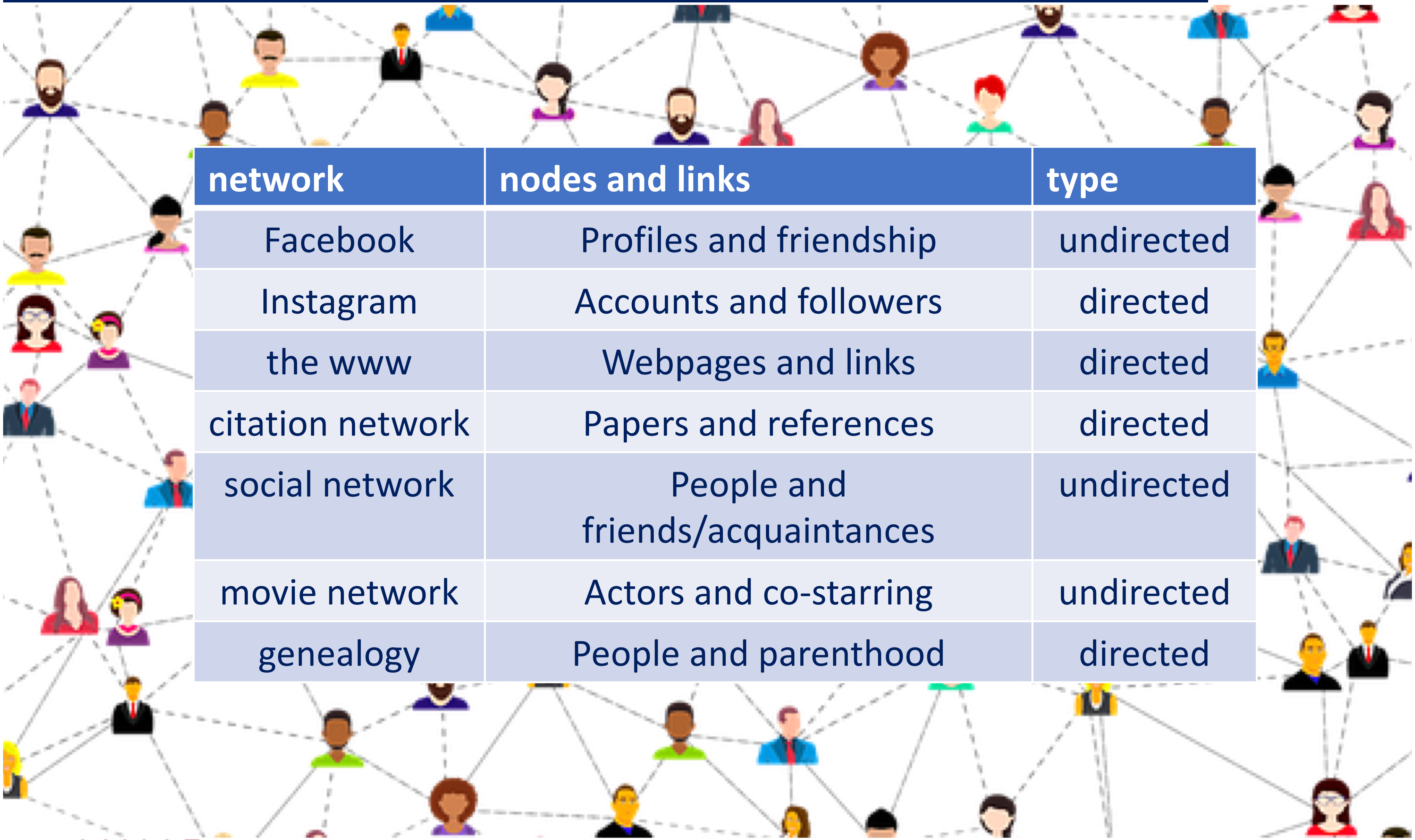
- If the network has only (un)directed links, it is also called itself (un)directed network
  - Certain networks can have both types

Directed versus undirected

- At first glance undirected → directed by duplicating links, but not necessarily quite the same though



Some examples

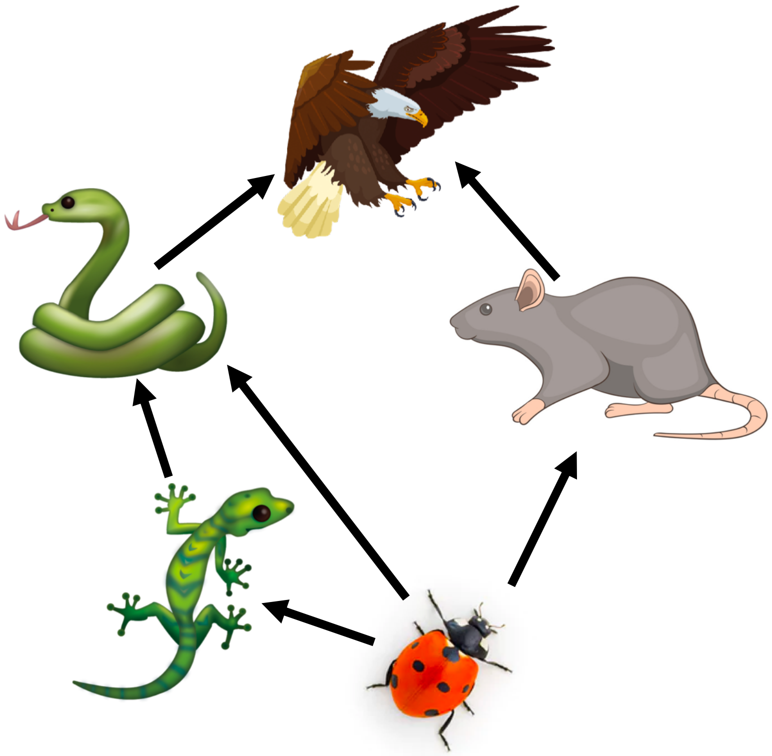


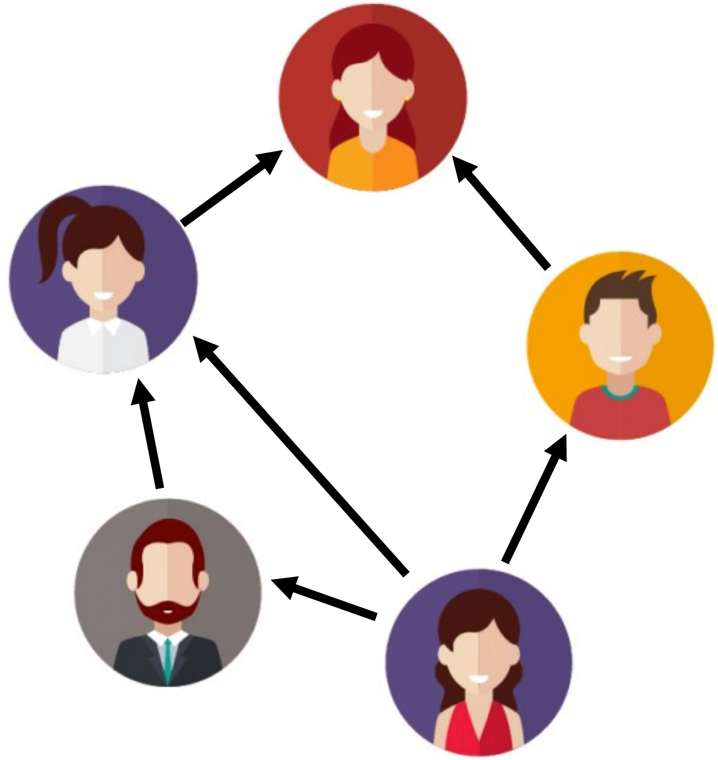
network	nodes and links	type
Facebook	Profiles and friendship	undirected
Instagram	Accounts and followers	directed
the www	Webpages and links	directed
citation network	Papers and references	directed
social network	People and friends/acquaintances	undirected
movie network	Actors and co-starring	undirected
genealogy	People and parenthood	directed

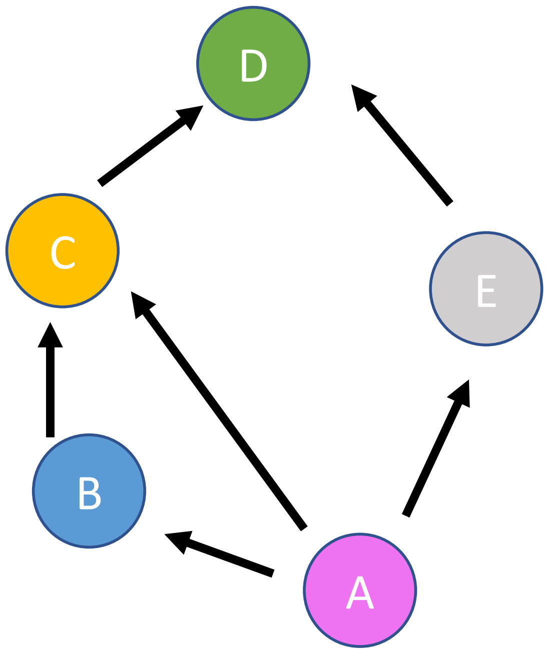
Can U think of other social nets?

network	nodes and links	type
Twitter	followers & tweets	directed
WhatsApp	telephone # & messages	directed
LinkedIn		
WhatsApp	telephone # & connections through groups	undirected
Twitter	people & hashtags	directed/un directed
Twitter	hashtags & tweets	undirected
Twitter	tweets & hashtags	undirected

Generality of representation



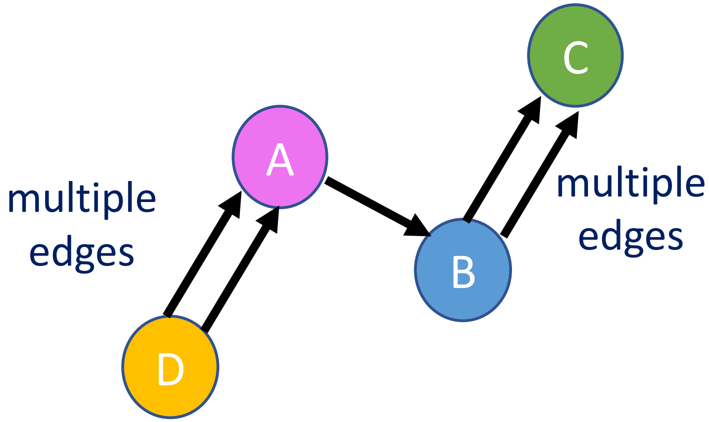




Multi-graphs

- Multi-graphs (or pseudo-graphs)

  Some network representations require multiple links (e.g., number of citations from one author to another)

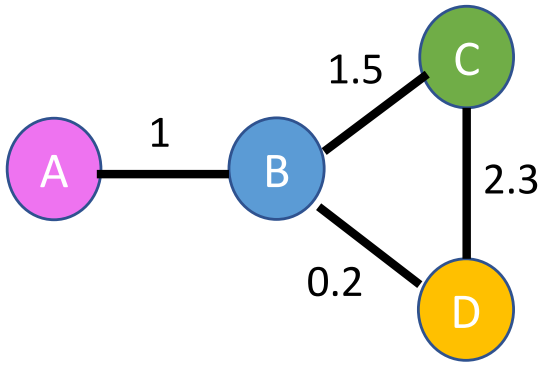


Weighted graph

❑ Weighted graph

Sometimes a weight is associated to a link, e.g., to underline that the links are not identical (strong/weak relationships)

Can be seen as a generalization of multi-graphs (weight = # of links)

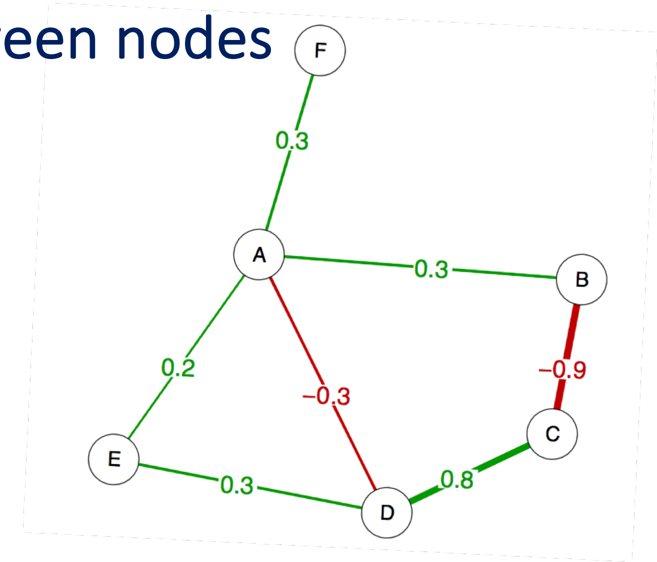


e.g., strength of a tie
- 0.2 = weak (acquaintances)
- 1 = strong (friends)
- 1.5 = stronger (close friends)
- 2.3 = very strong (best friends)

Signed graphs

- Edges can have signed values

  positive if there is an agreement between nodes
  negative if there's a disagreement



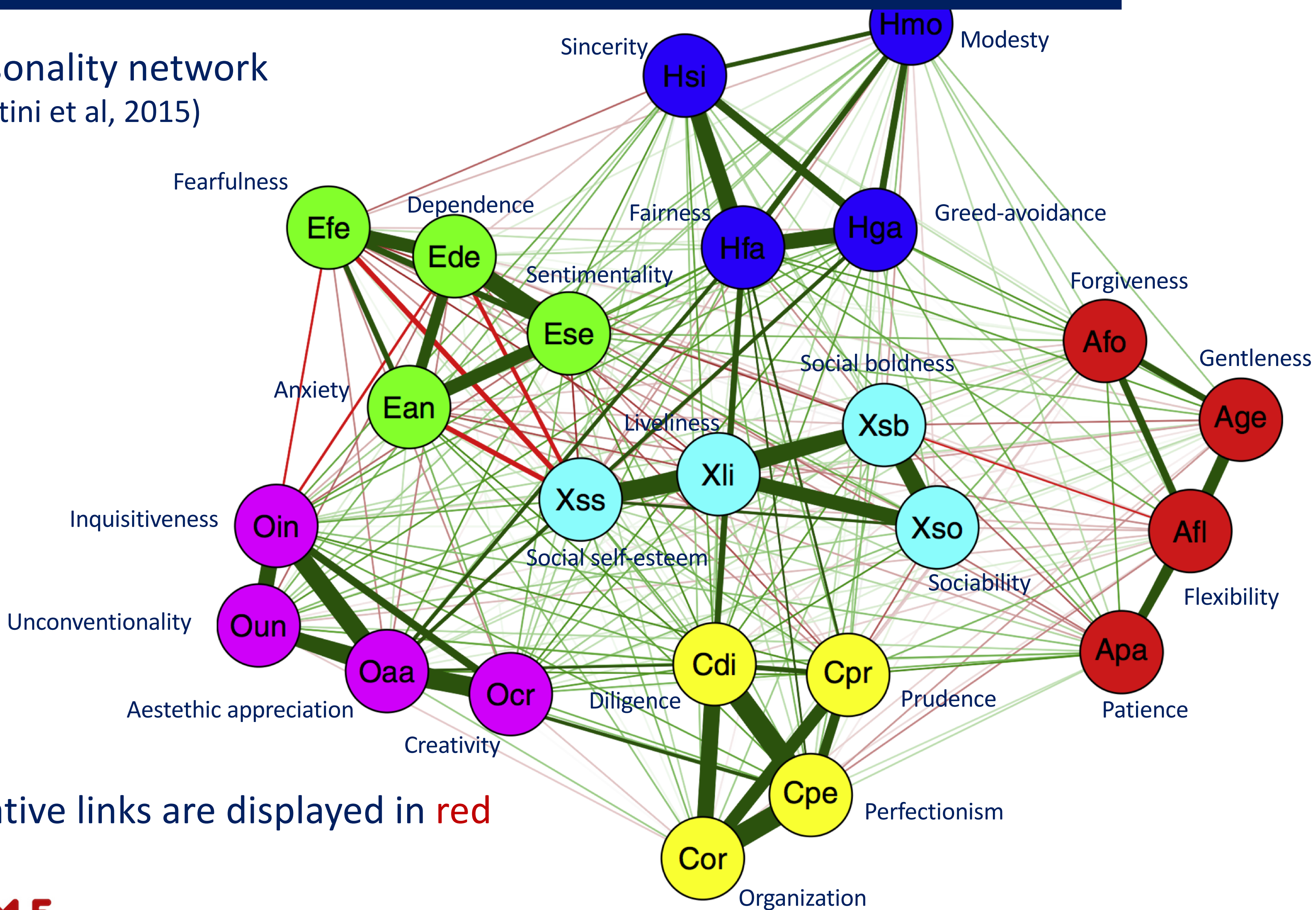
- This is typical of correlation networks

  correlation = a measure of similarity

- More difficult to handle

Example

A personality network
(Costantini et al, 2015)

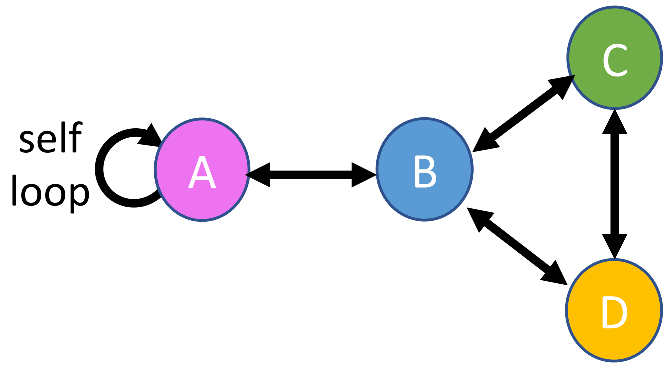


Negative links are displayed in red

MIME

Self-interactions

- In many networks nodes do not interact with themselves
- To account for self-interactions, we add loops to represent them



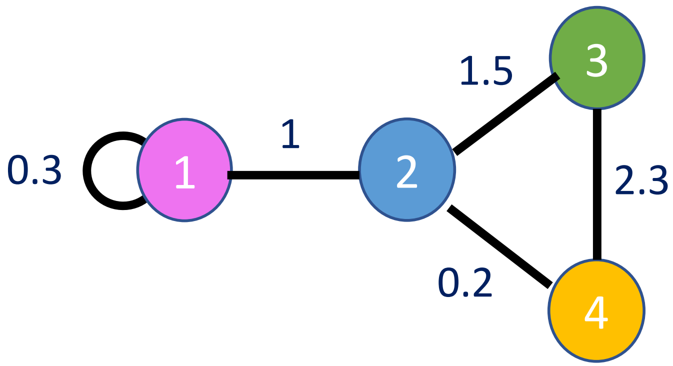
Adjacency matrix

- An adjacency matrix $A = [a_{ij}]$ associated to graph $\mathcal{G}$ has

  i is the row index   *j* is the column index

  entries $a_{ij} = 0$ if nodes *i* and *j* are not connected
  if nodes *i* and *j* are connected then $a_{ij} \neq 0$

  in plain (binary) graphs $a_{ij} = \{1,0\}$



this is a_{12}

$$A = \begin{bmatrix} 0.3 & 1 & 0 & 0 \\ 1 & 0 & 1.5 & 0.2 \\ 0 & 1.5 & 0 & 2.3 \\ 0 & 0.2 & 2.3 & 0 \end{bmatrix}$$

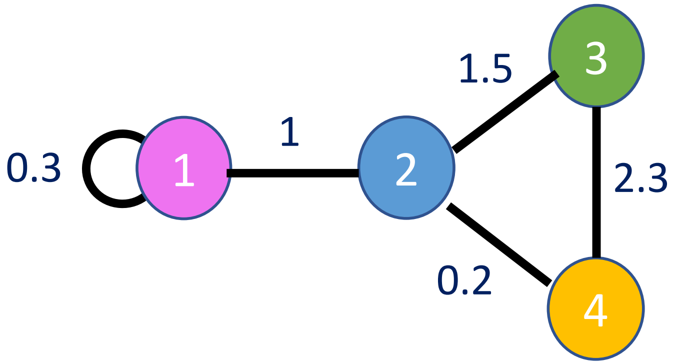
row 1

column 2

MiME.

Symmetries

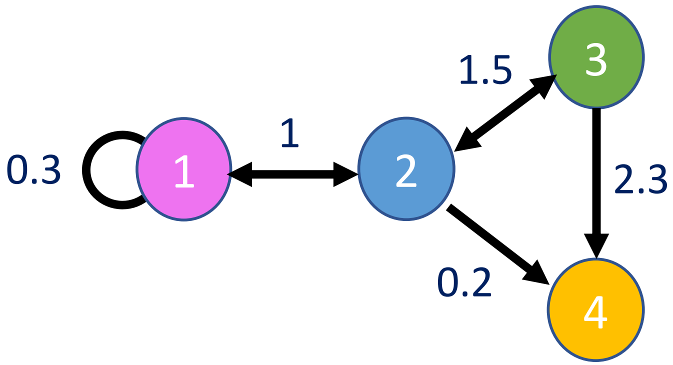
- Undirected graph = symmetric matrix



$A =$

to \ from	1	2	3	4
1	0.3	1	0	0
2	1	0	1.5	0.2
3	0	1.5	0	2.3
4	0	0.2	2.3	0

- Directed graph = asymmetric matrix

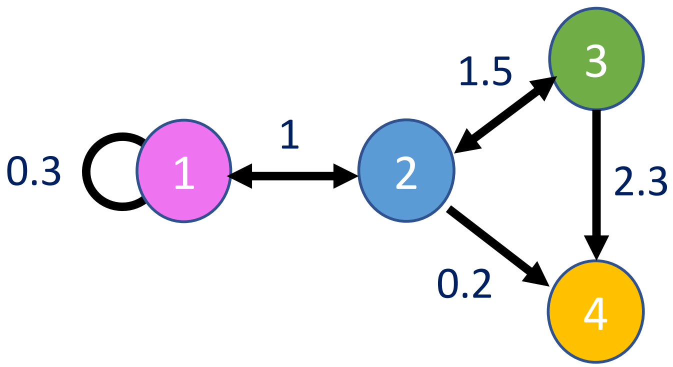


$A =$

to \ from	1	2	3	4
1	0.3	1	0	0
2	1	0	1.5	0
3	0	1.5	0	0
4	0	0.2	2.3	0

Convention

- The weight a_{ij} is associated to
  i th row
  j th column
  directed edge $j \rightarrow i$ starting from node *j* and leading to node *i*

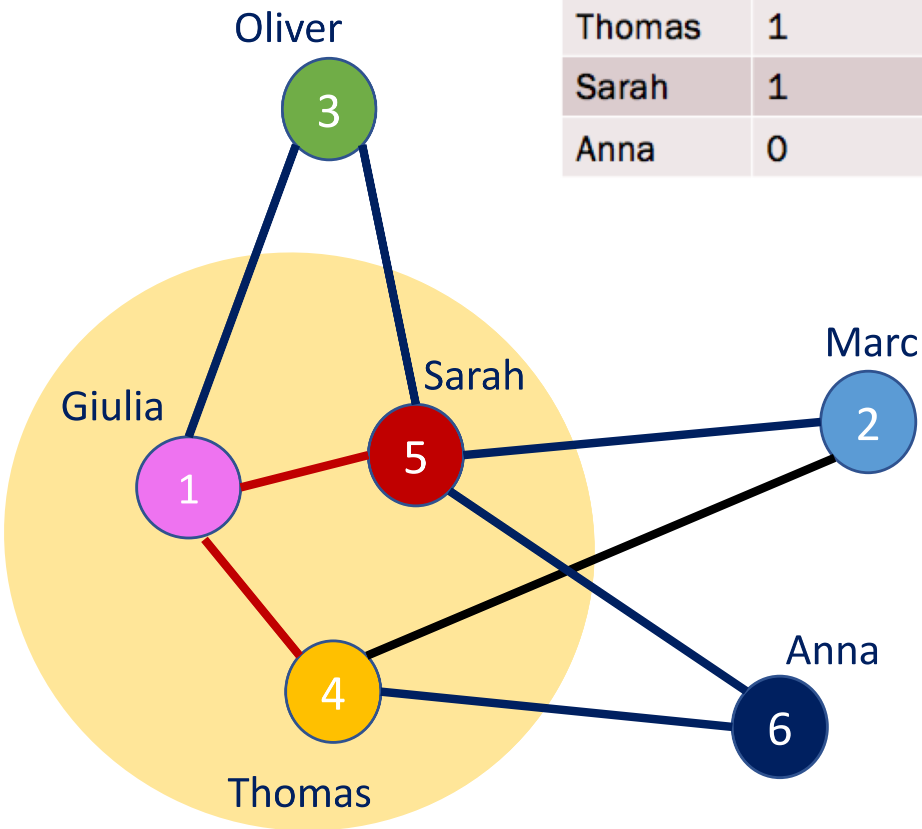


$$A = \begin{bmatrix} 0.3 & 1 & 0 & 0 \\ 1 & 0 & 1.5 & 0 \\ 0 & 1.5 & 0 & 0 \\ 0 & 0.2 & 2.3 & 0 \end{bmatrix}$$

a_{24}, a_{34}, a_{42}, a_{43}

Example

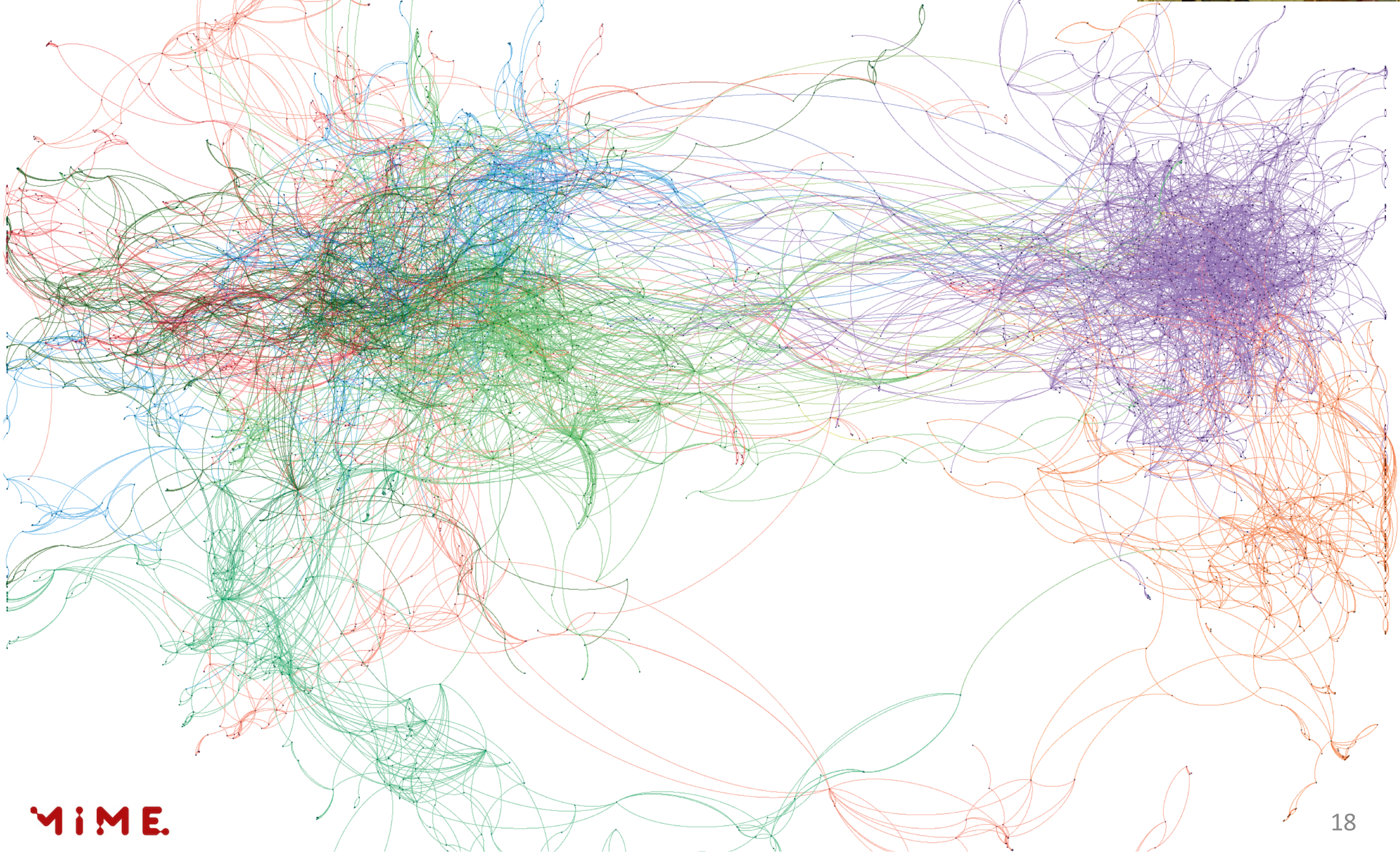
	Giulia	Marc	Oliver	Thomas	Sarah	Anna
Giulia	X					
Marc	0	X				
Oliver	1	0	X			
Thomas	1	1	0	X		
Sarah	1	1	1	0	X	
Anna	0	0	0	1	1	x



$$A = \begin{bmatrix} 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 & 0 \end{bmatrix}$$

which of these representations do you like best?

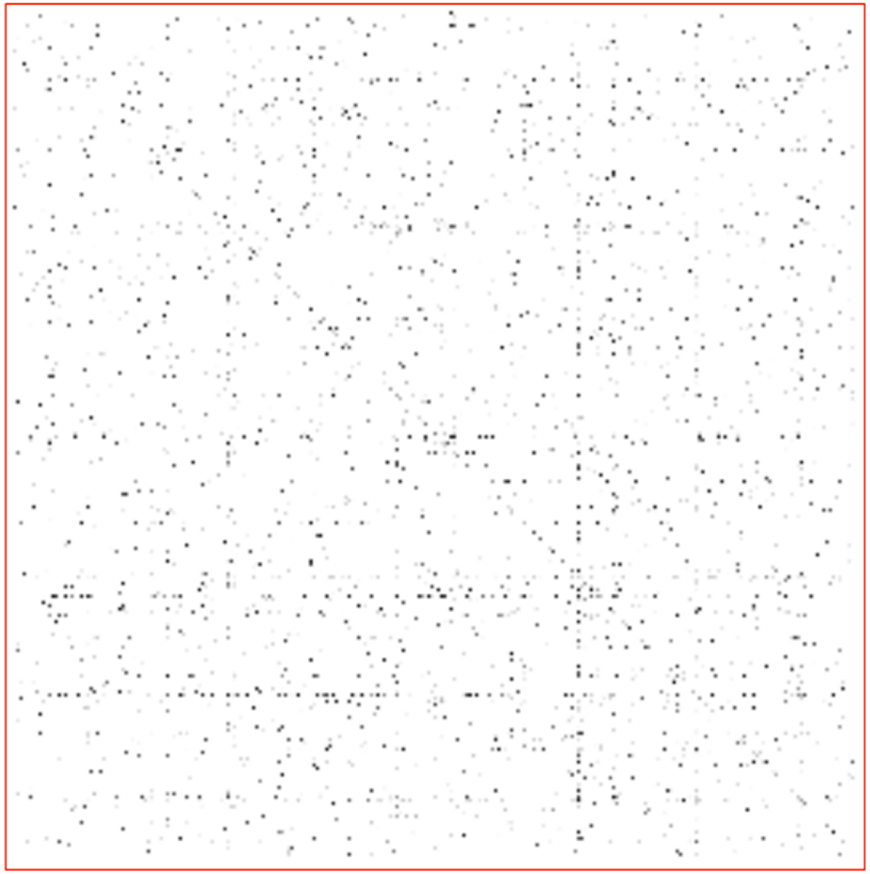
Graph plots do not always carry relevant info



Real networks are sparse

- The adjacency matrix is typically sparse

good for tractability !

$A =$ 

A question 4 u

- So, what’s the take-away so far?

Storing network data

Adjacency matrix versus edge list

N^2 entries

$A =$

0	0	1	1	1	0
0	0	0	1	1	0
1	0	0	0	1	0
1	1	0	0	0	1
1	1	1	0	0	1
0	0	0	1	1	0

L entries

1 → 3
1 → 4
1 → 5
2 → 4
2 → 5
3 → 5
4 → 6
5 → 6

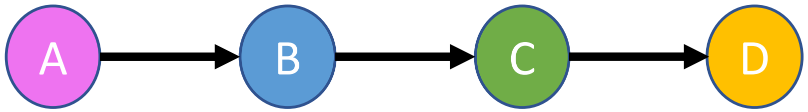
Which one do U think is better?

MIME

Useful terms

- **Path**

  a sequence of interconnected nodes (meaning each pair of nodes adjacent in the sequence are connected by a link)

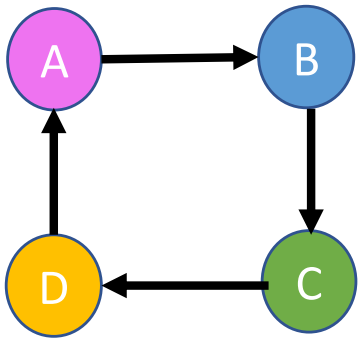


- **Path length**

  # of links involved in the path (if the path involves *n* nodes then the path link is *n-1*)

- **Cycle**

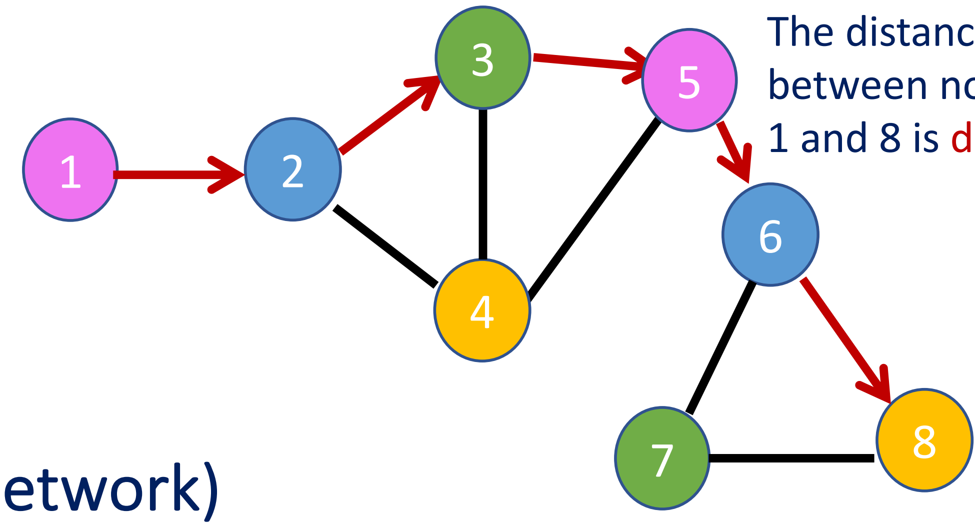
  path where starting and ending nodes coincide



Useful terms

- **Shortest path** (between any two nodes)

  the path with the minimum length, which is called the distance



The distance between nodes 1 and 8 is $d_{18}=5$

  it is not unique!

- **Diameter** (of the network)

  the highest distance in the network

The diameter is $d=5$

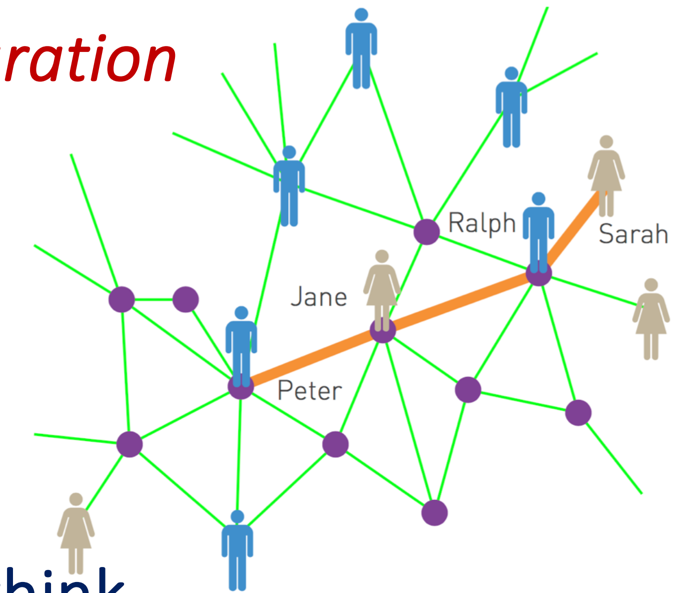
- **Algorithms**

  available to compute distances: Dijkstra, Bellman-Ford, BFS

Small world

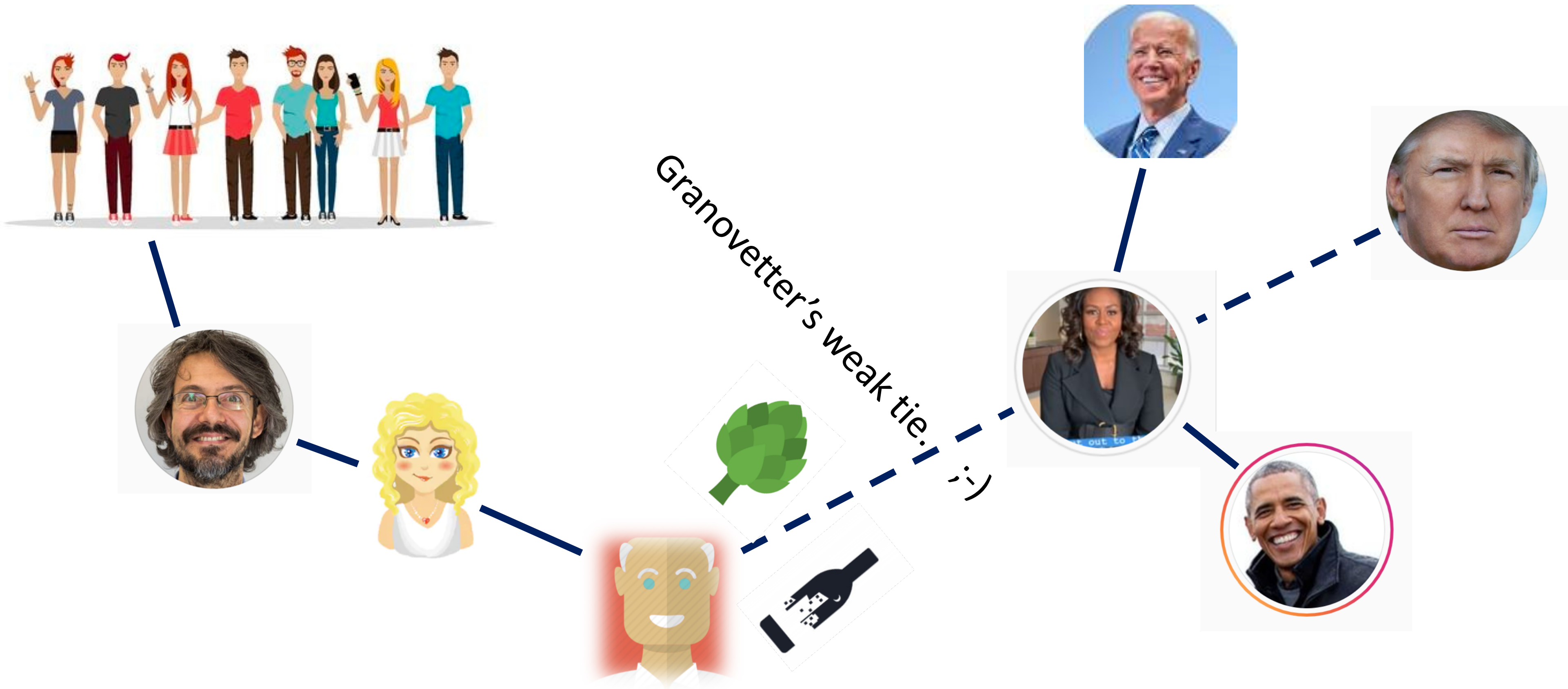
- Average path length
  average distance between all nodes pairs (apply an algorithm to all node couples, and take the average)

- In real networks distance between two randomly chosen nodes is generally short
- Milgram [1967]: *6 degrees of separation*



- What does this mean?
  We are more connected than we think

We & the US



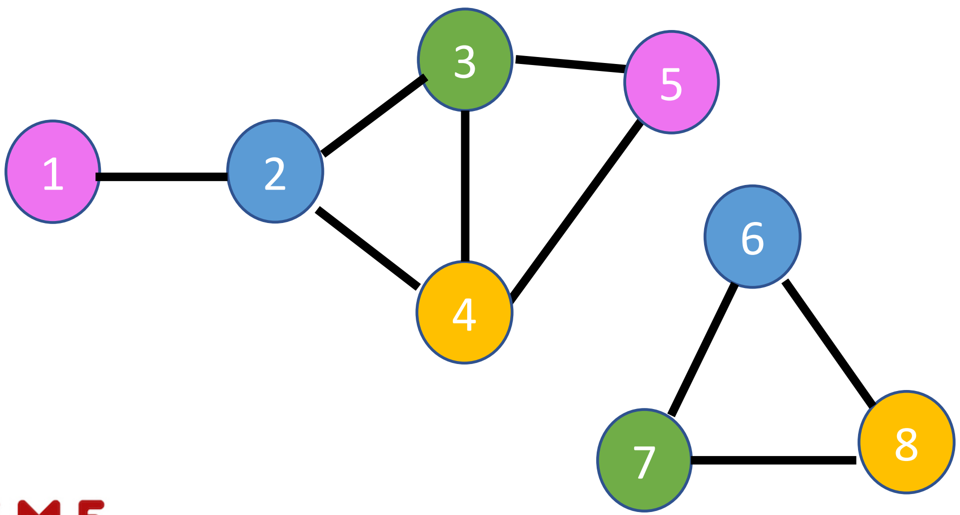
Connectivity

- **Connected graph (undirected)**

  for all couples (*i,j*) there exists a path connecting them

  if disconnected, we count the # of connected components (e.g., use BFS and iterate)

- **Giant component (the biggest one)**
- **Isolates (the other ones)**



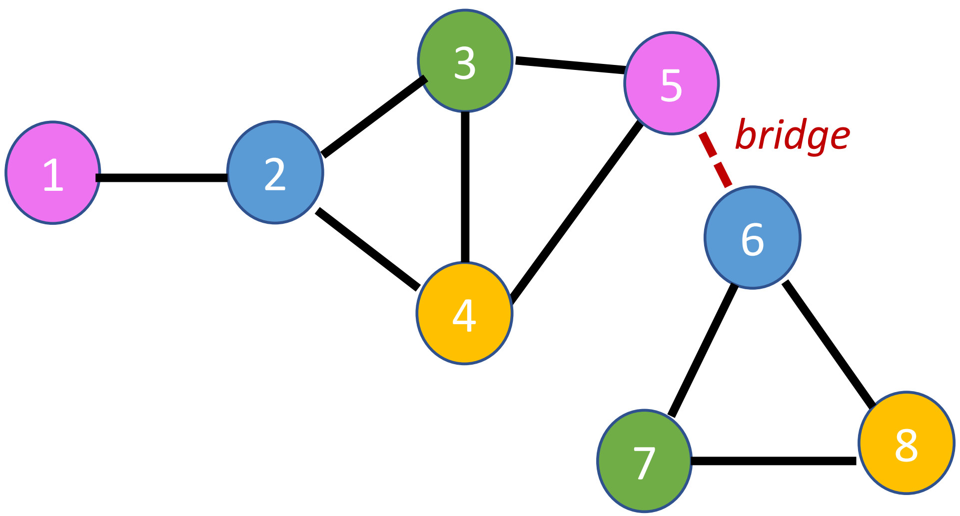
$$A=\begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \end{bmatrix}$$

block-diagonal matrix

Bridges (ideal definition)

- A **bridge** is a link between two connected components

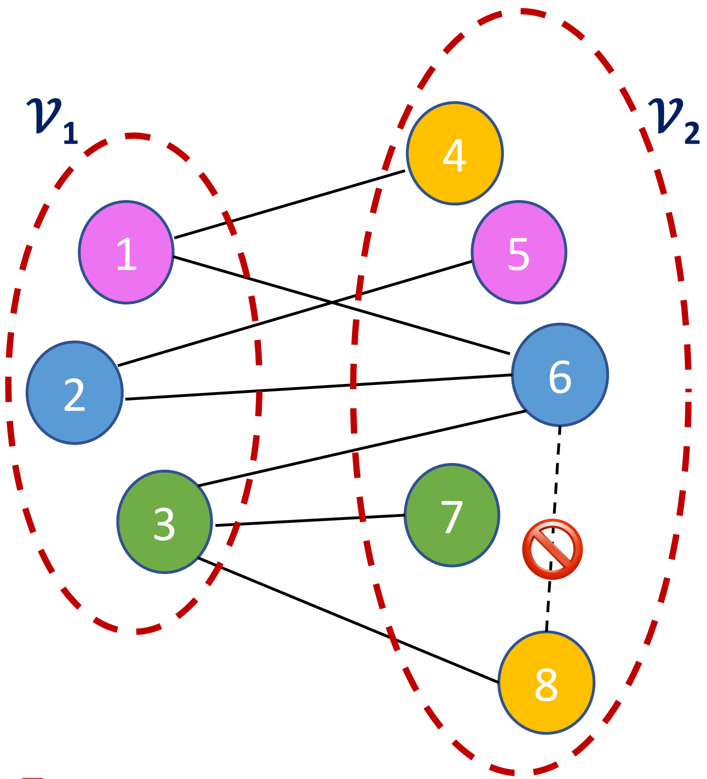
its removal would make the network disconnected

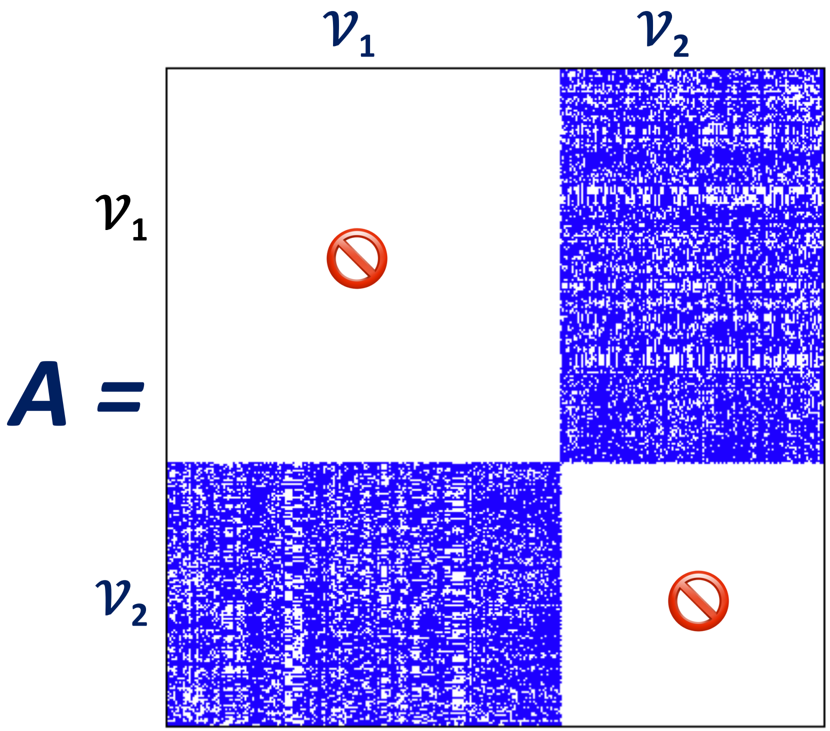


$$
A = \begin{bmatrix}
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 \\
0 & 1 & 1 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 1 & 1 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 1 & 1 & 0
\end{bmatrix}
$$

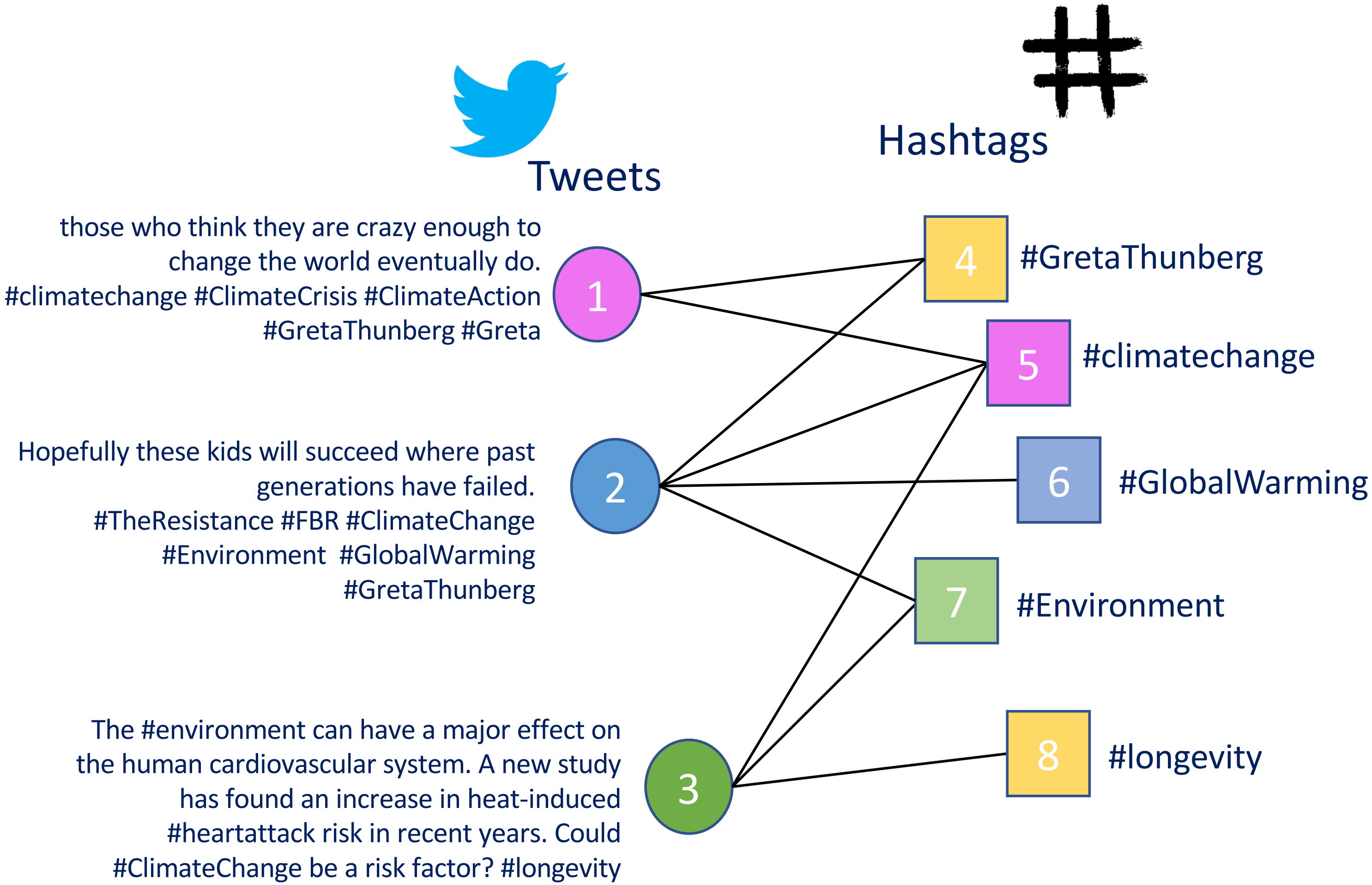
Bipartite graphs

Connections are available only between the groups $\mathcal{V}_1$ and $\mathcal{V}_2$





Example


Tweets
Hashtags
those who think they are crazy enough to change the world eventually do. #climatechange #ClimateCrisis #ClimateAction #GretaThunberg #Greta
1
Hopefully these kids will succeed where past generations have failed. #TheResistance #FBR #ClimateChange #Environment #GlobalWarming #GretaThunberg
2
The #environment can have a major effect on the human cardiovascular system. A new study has found an increase in heat-induced #heartattack risk in recent years. Could #ClimateChange be a risk factor? #longevity
3
4
#GretaThunberg
5
#climatechange
6
#GlobalWarming
7
#Environment
8
#longevity

Meaning

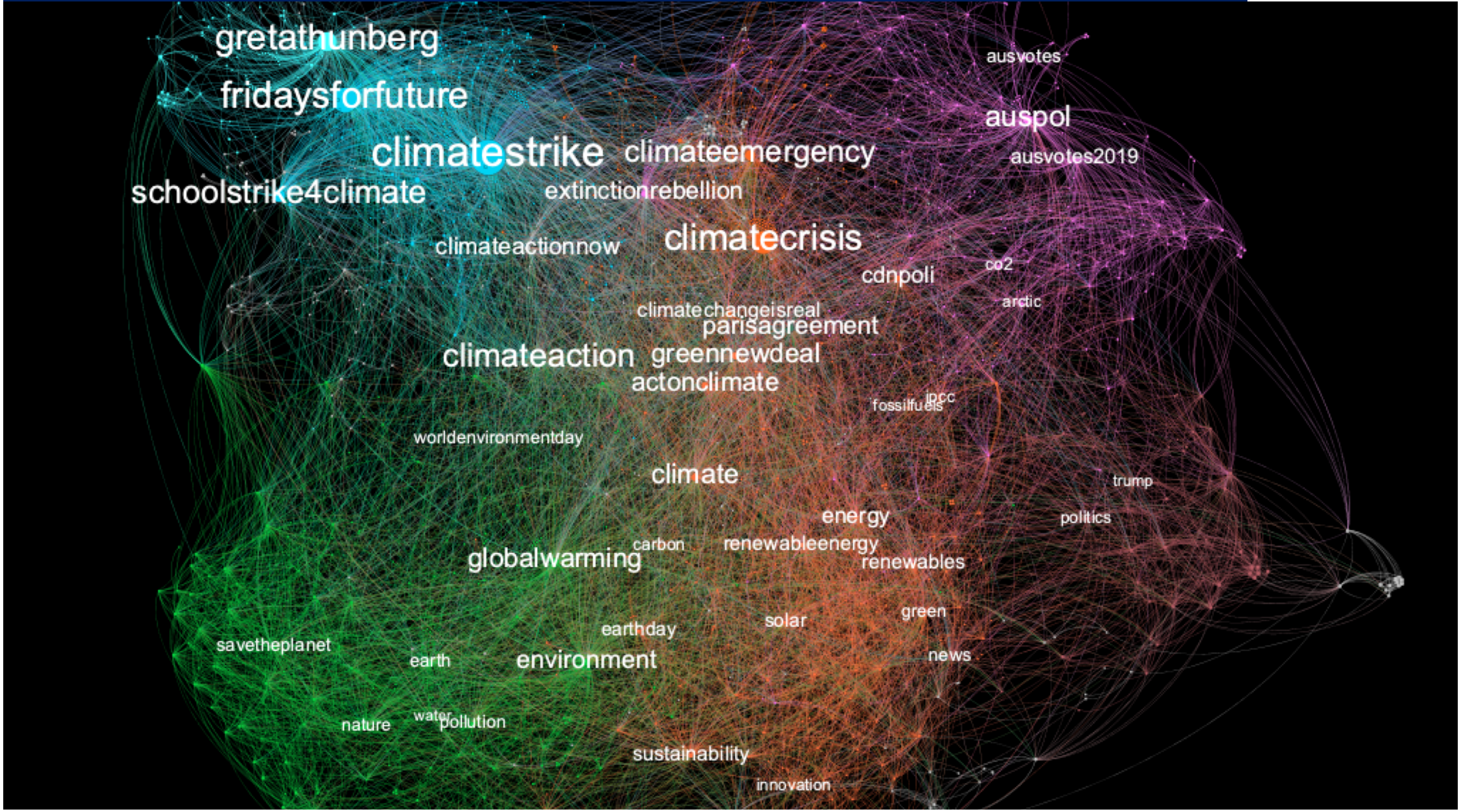
- Bipartite graphs are useful to represents memberships/relationships, e.g., groups ($\mathcal{V}_1$) to which people ($\mathcal{V}_2$) belong

  examples: movies/actors, classes/students, conferences/authors

- We can build separate networks (projections) for $\mathcal{V}_1$ and $\mathcal{V}_2$ (sometimes this is useful)

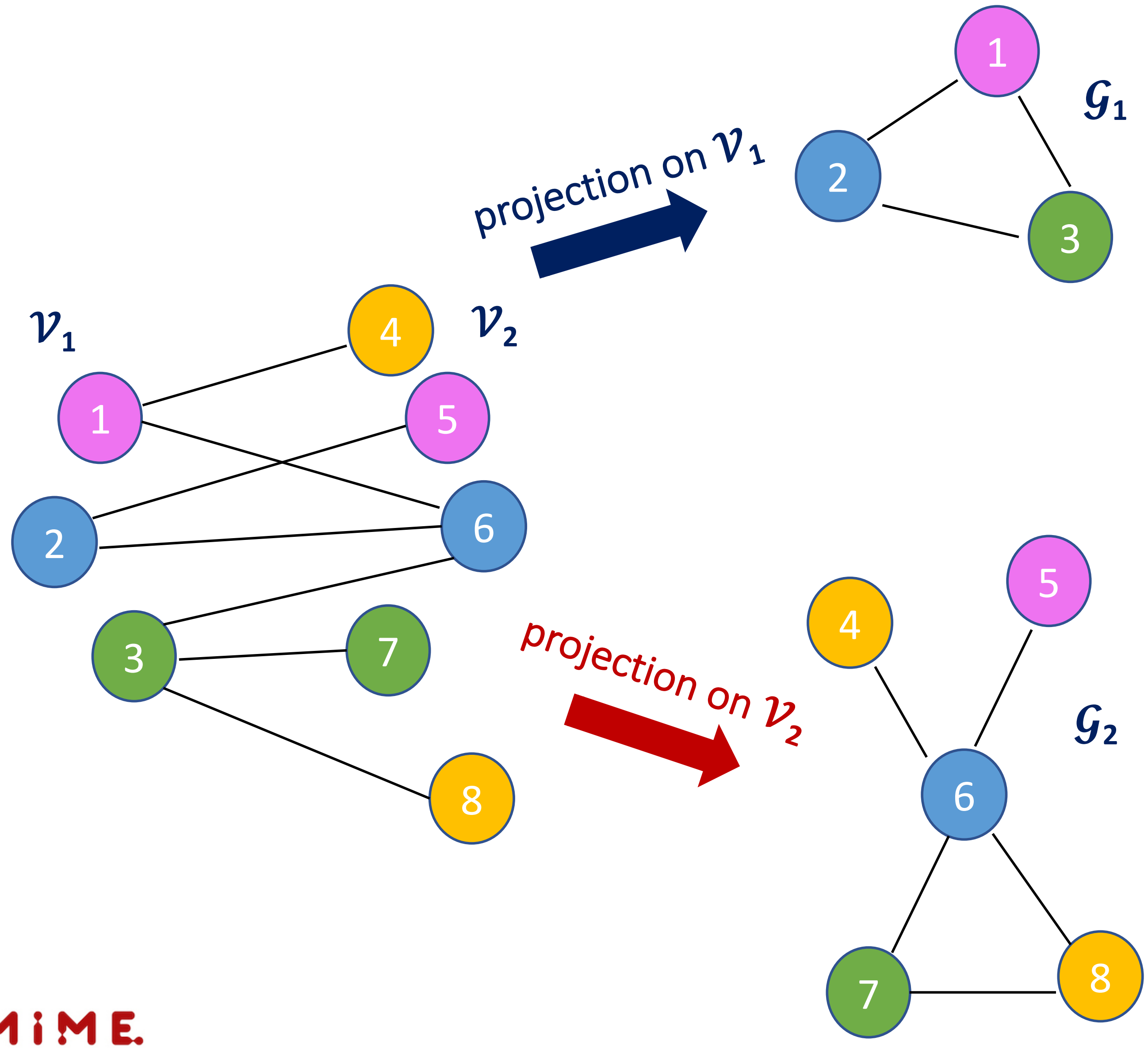
  in the movies/actors example being linked can be interpreted in two ways: "actors in the same movie" (projection on $\mathcal{V}_2$), or "movies sharing the same actor" (projection on $\mathcal{V}_1$)

Projection on



#climateaction tweets after Greta Thunberg

Example



Nodes are linked if they have a common neighbour in $\mathcal{V}_2$

PS: we say that nodes *i* and *j* have a common neighbour *k* if both *i* and *j* are connected to *k*

Nodes are linked if they have a common neighbour in $\mathcal{V}_1$

A bit of maths

The two projections on $\mathcal{V}_1$ and $\mathcal{V}_2$ can be obtained by inspecting the squared adjacency matrix A^2

$$
A^2 = \begin{bmatrix}
0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0
\end{bmatrix}
\begin{bmatrix}
0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0
\end{bmatrix}
=
\begin{bmatrix}
2 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\
1 & 2 & 1 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 3 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\
0 & 0 & 0 & 1 & 1 & 3 & 1 & 1 \\
0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\
0 & 0 & 0 & 0 & 0 & 1 & 1 & 1
\end{bmatrix}
$$

row 6 → (row 6 of the first matrix)

column 5 (column 5 of the second matrix)

A_1 (upper-left block of the result), A_2 (lower-right block of the result)

# of common neighbors of *i=6* and *j=5*

# of neighbors of *i=6*

← row 6

column 5

Today take-aways

- (un)Directed graphs
- Weighted and signed graphs
- Adjacency matrix
- Giant component, isolates, bridges
- Bipartite graphs and projections