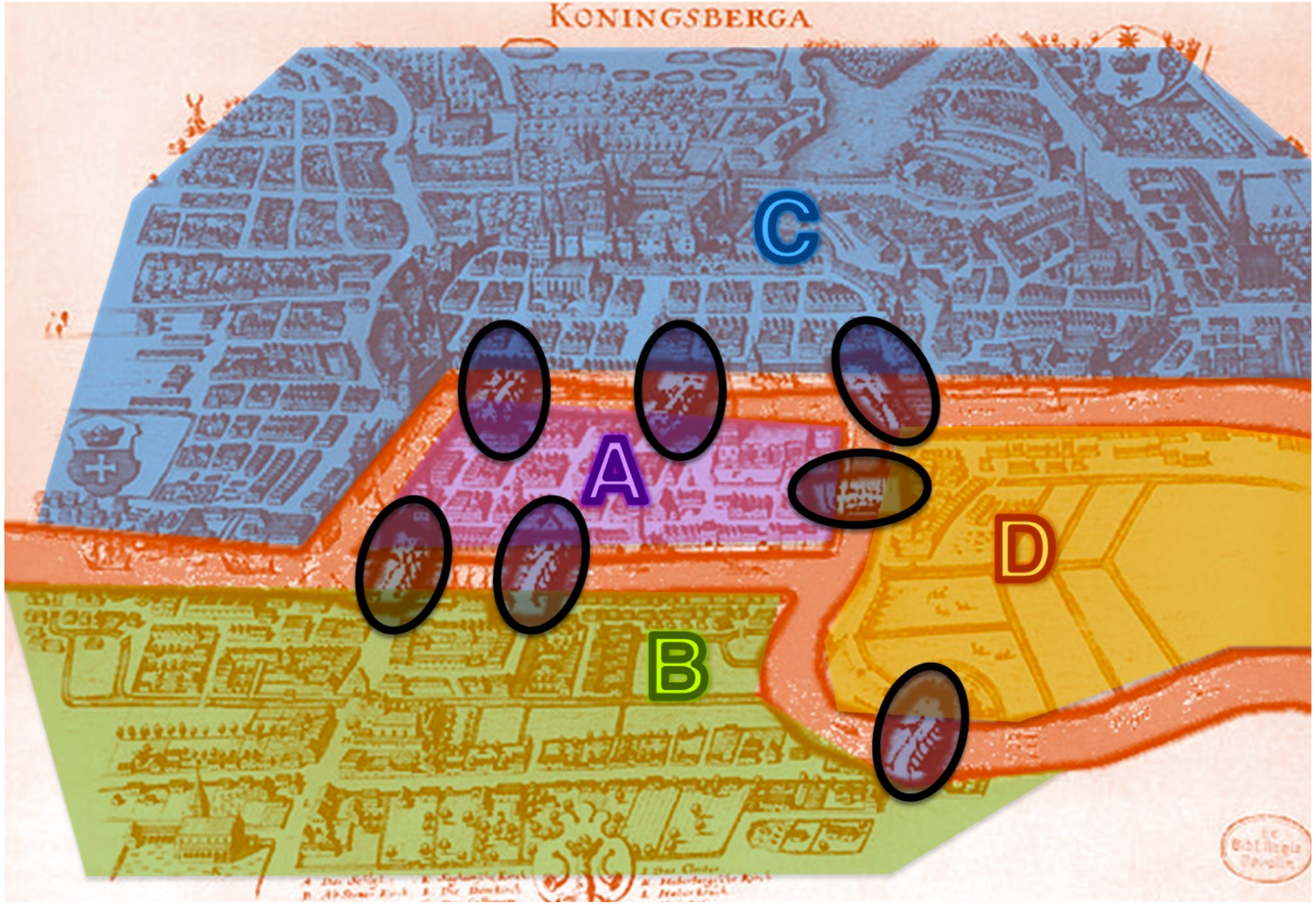
Network Science

#2 Graphs

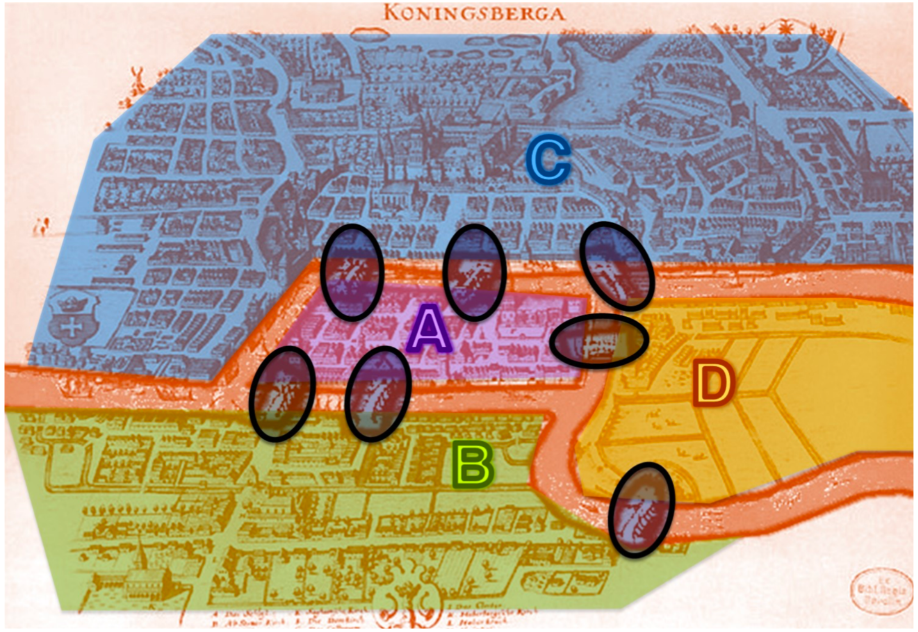
© 2020 T. Erseghe

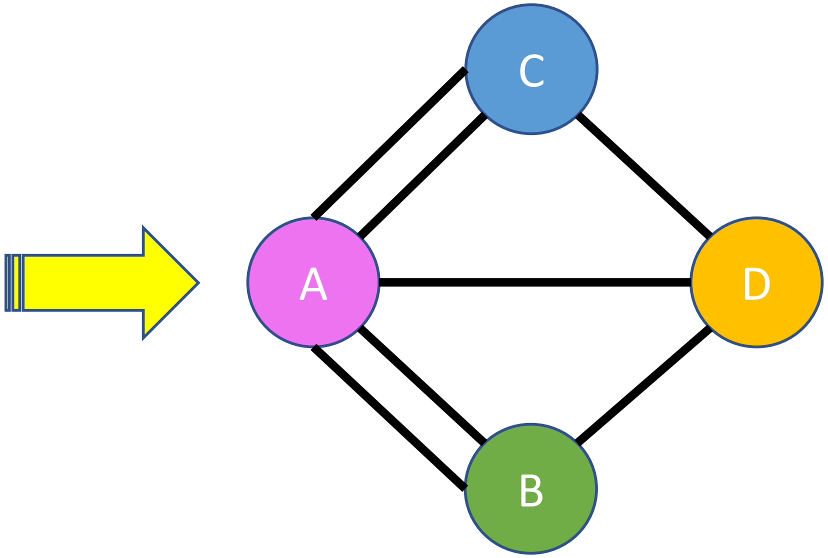
Euler & the 7 bridges of Königsberg (1736)



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

Networks as graphs





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

- Vertices (set $\mathcal{V}$) : nodes, users, elements
- Edges (set $\mathcal{E}$): links, arcs, hops, connections

Modelling aspects

- ☐ "Network, nodes, links" = technology
- ☐ "Graph, vertices, edges" = mathematics

- ☐ Design choices for what nodes and links are
  - ☐ graph structure can be a given
  - ☐ or it can be the focus of the model itself

Directed versus undirected

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

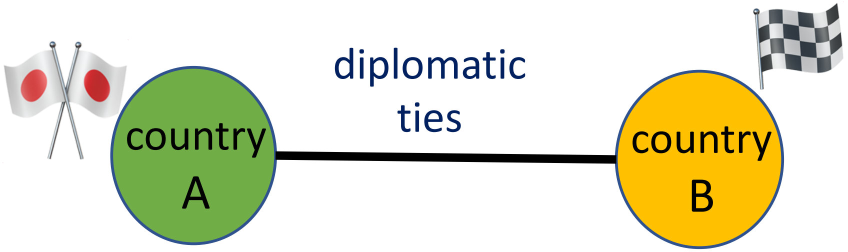
- If the network has only (un)directed links, it is also called itself (un)directed network
  - Certain networks can have both types

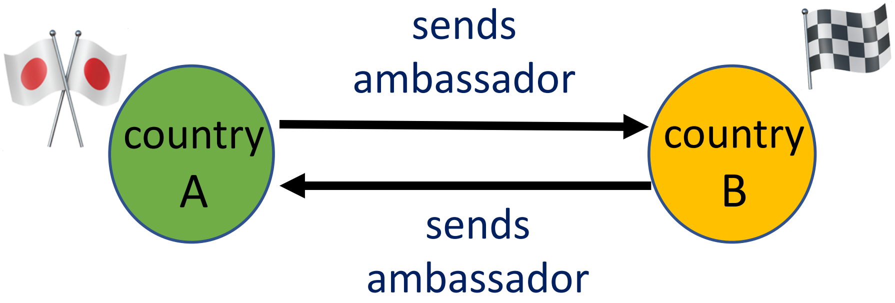
Some examples

network	nodes and links	type
the Internet	Hosts and connections	undirected
the web	Webpages and links	directed
electrical grid	Power stations and cables	undirected
social network	Users and friendship	undirected
citation network	Papers and references	directed
movie network	Actors and co-starring	undirected
metabolism	Compounds and reactions	directed
protein network	Proteins and bindings	undirected
genealogy	People and parenthood	directed

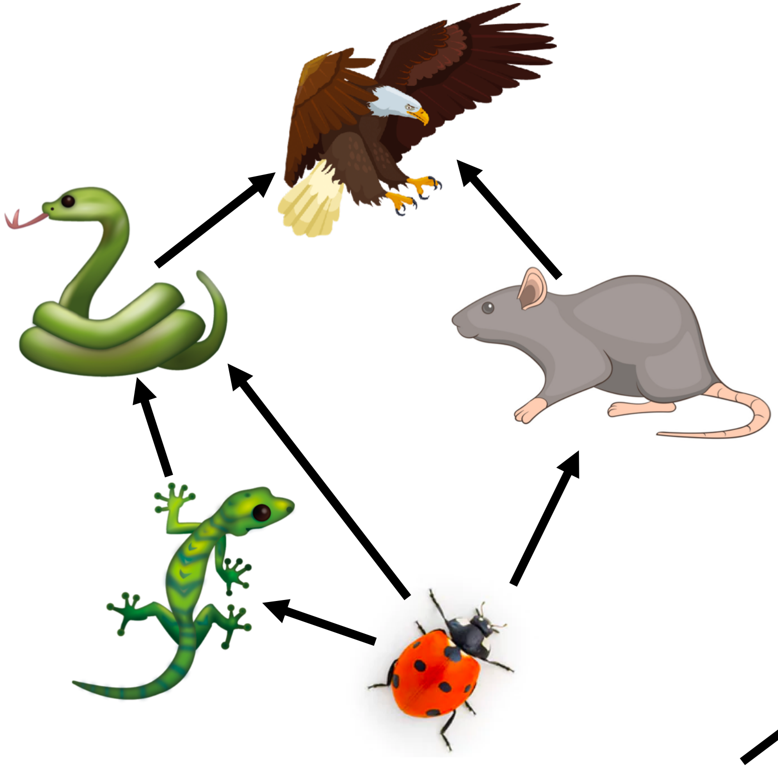
Directed versus undirected

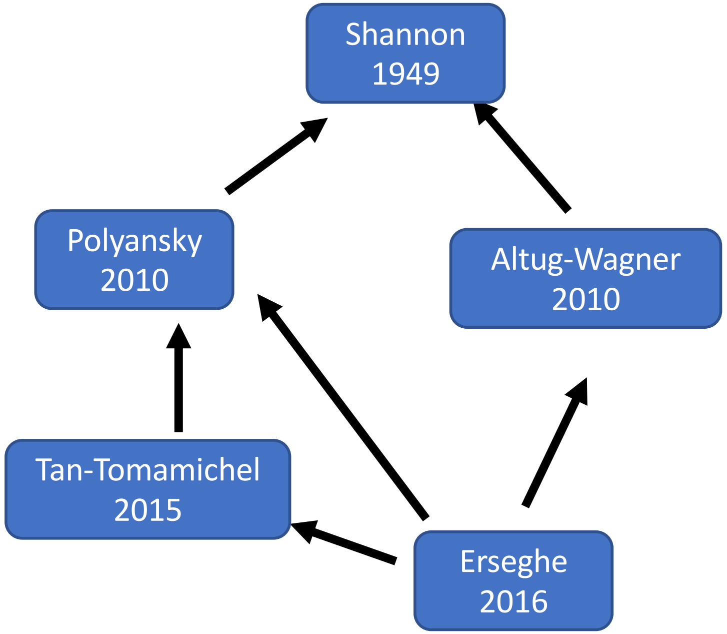
- ❑ At first glance undirected → directed by duplicating links, but not necessarily quite the same though





Generality of representation





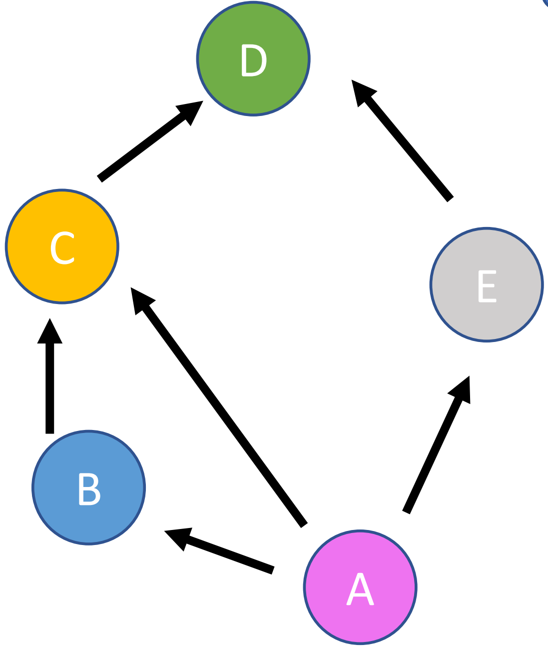
Shannon
1949

Polyansky
2010

Altug-Wagner
2010

Tan-Tomamichel
2015

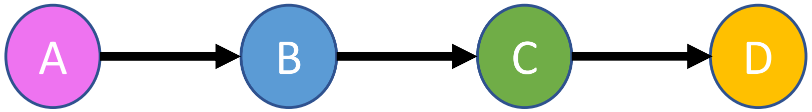
Erseghe
2016



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*)

- Shortest path (between any two nodes)

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

- Diameter (of the network)

  the highest distance in the network

Useful terms

- ☐ **Algorithms**

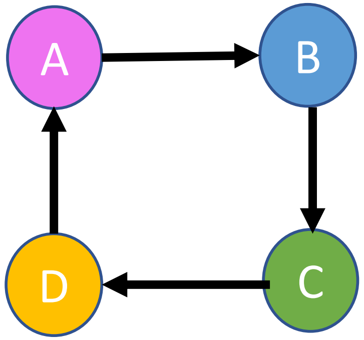
  available to compute distances: Dijkstra, Bellman-Ford, BFS

- ☐ **Average path length**

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

- ☐ **Cycle**

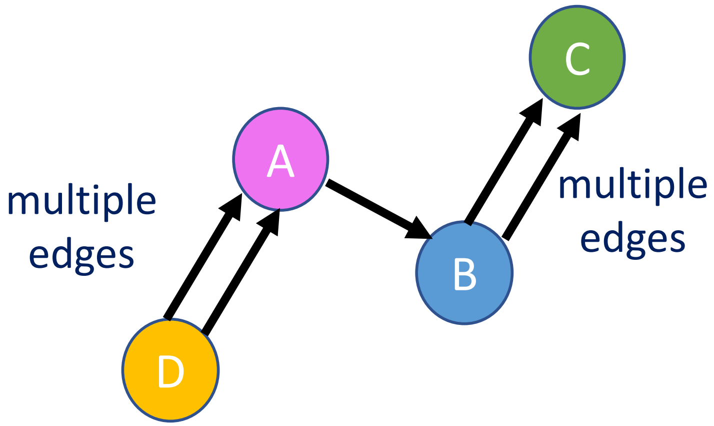
  path where starting and ending nodes coincide



Multi-graphs

- ❑ Multi-graphs (or pseudo-graphs)

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

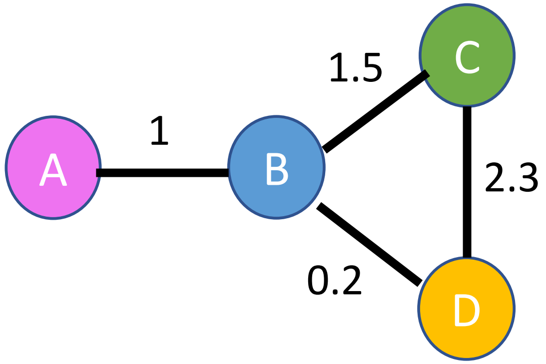


Weighted graph

- ❑ Weighted graph

Sometimes a **weight** w_{ij} is associated to a link $(i,j) \in \mathcal{E}$, 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)

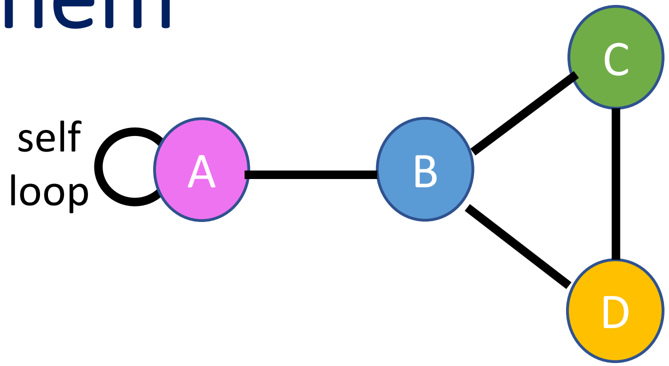


Self-interactions

- ☐ In many networks nodes do not interact with themselves

  if $j \in \mathcal{V}$ then $(j,j) \notin \mathcal{E}$

- ☐ To account for self-interactions, we add loops to represent them



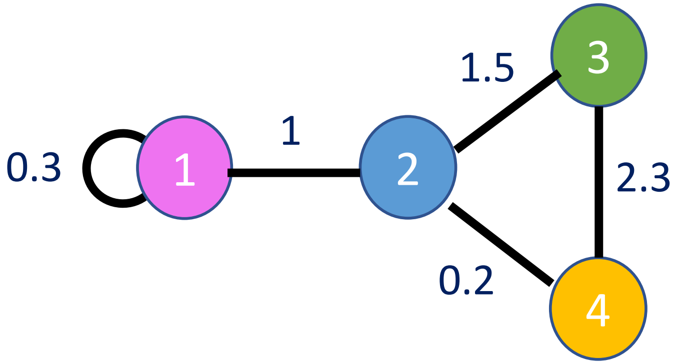
Adjacency matrix

- ☐ An adjacency matrix $A = [a_{ij}]$ associated to graph $\mathcal{G}$ ($\mathcal{V}$,$\mathcal{E}$) has

  entries $a_{ij} = 0$ for $(i,j) \notin \mathcal{E}$ (not a connection)

  if nodes *i* and *j* are connected then $a_{ij} \neq 0$

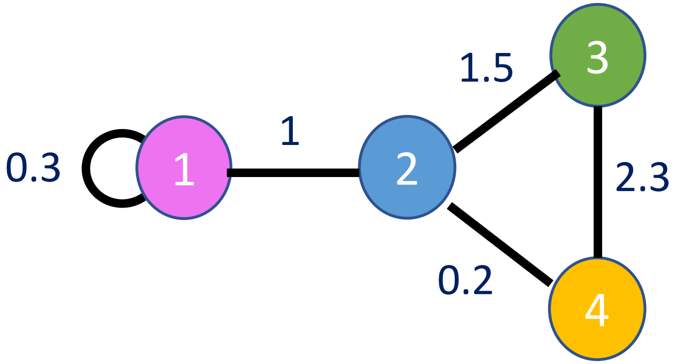
  in plain graphs $a_{ij} = 1$ for $(i,j) \in \mathcal{E}$



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

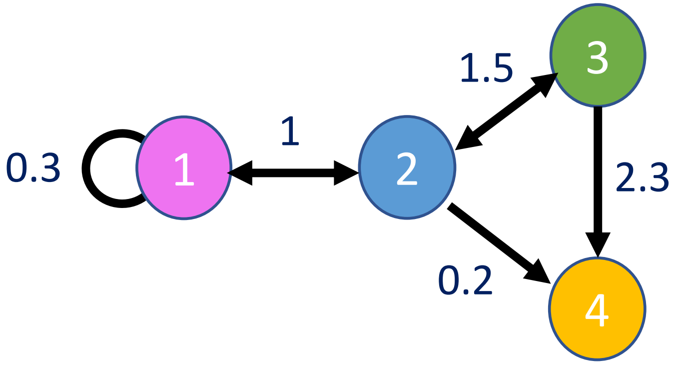
Symmetries

- Undirected graph = symmetric matrix



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

- Directed graph = asymmetric matrix



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

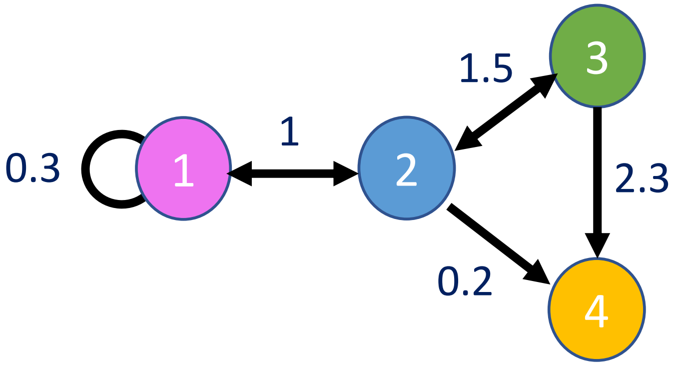
Convention

- The weight a_{ij} is associated to

  i th row

  j th column

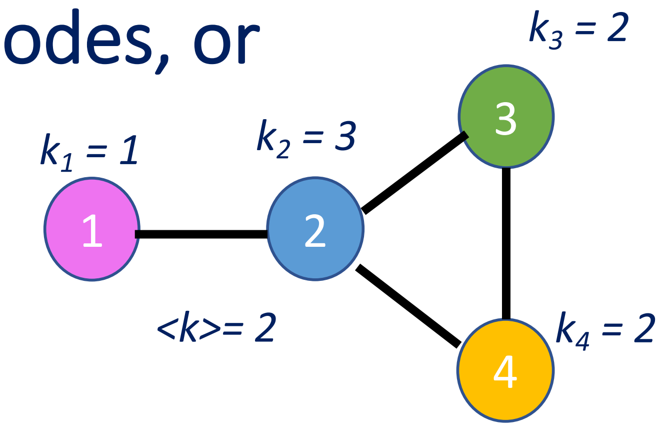
  directed edge *j*→*i* starting from node *j* and leading to node *i*



$$A = \begin{bmatrix} 0.3 & 1 & 0 & 0 \\ 1 & 0 & 1.5 & \overset{a_{24}}{0} \\ 0 & 1.5 & 0 & \underset{a_{34}}{0} \\ 0 & \underset{a_{42}}{0.2} & \underset{a_{43}}{2.3} & 0 \end{bmatrix}$$

Degree

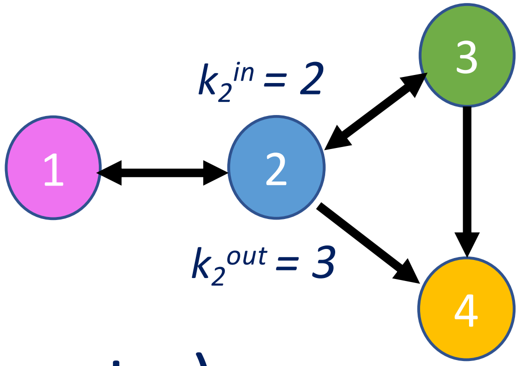
- ❑ The **degree** k_i of node i in an **undirected** networks is
  the # of links i has to other nodes, or
  the # of nodes i is linked to



- ❑ The # of nodes is $N = |\mathcal{V}|$
- ❑ The # of edges is $L = |\mathcal{E}| = ½ \sum_i k_i$
- ❑ The **average** degree is $<k> = \sum_i k_i / N = 2L / N$

Degree

- For directed networks we distinguish between
  - in-degree k_i^{in} = # of entering links
  - out-degree k_i^{out}= # of exiting links
  - total degree $k_i = k_i^{in} + k_i^{out}$

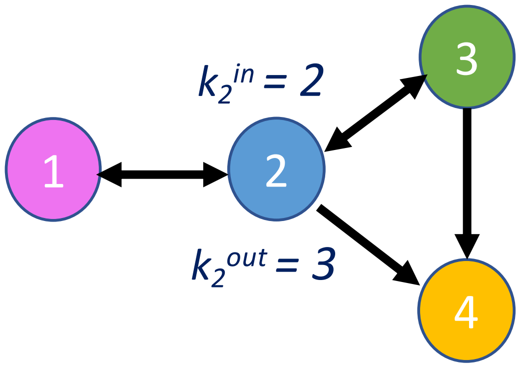


(undirected: $k_i^{in} = k_i^{out}$ due to the symmetry)

- The # of links is $L = \sum_i k_i^{in} = \sum_i k_i^{out}$
  - (no need for factor ½)
  - The average # of links is <k> = L / N

Adjacency matrix & degree

- The in (out) degree can be obtained by summing the adjacency matrix over rows (columns)



no self-loops in this case!!!

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \end{bmatrix} \quad k_2^{in} = 2$$

$k_2^{out} = 3$

- A few useful linear algebra expressions

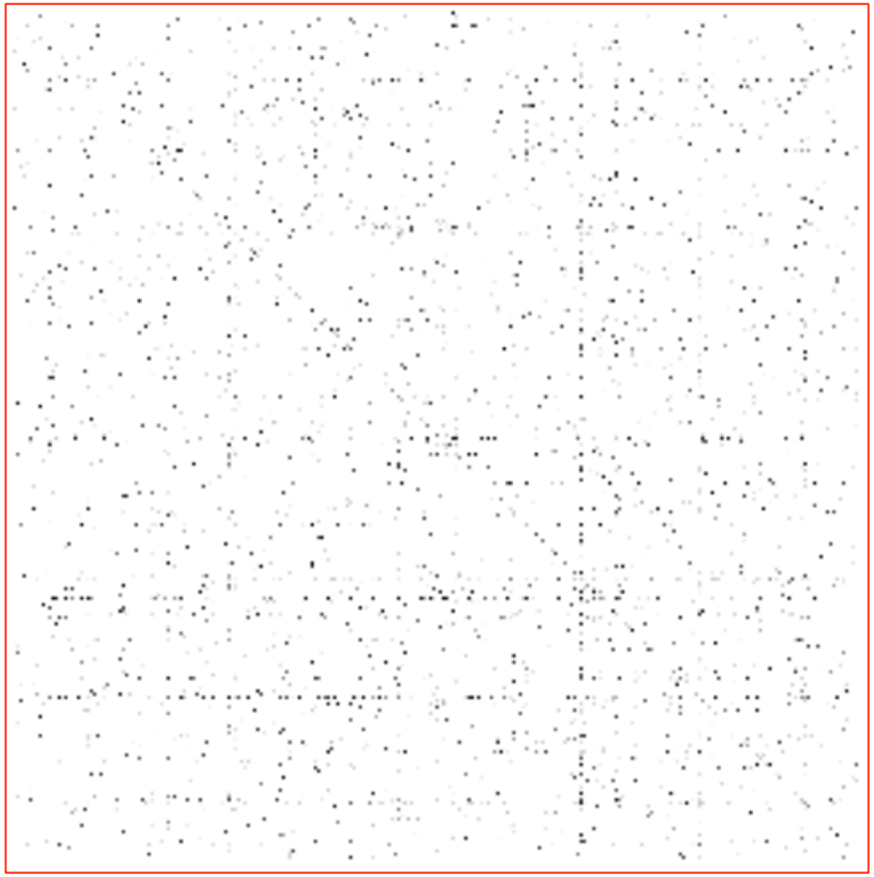
$$\boxed{k^{in} = A \cdot 1} \quad \boxed{k^{out} = A^T \cdot 1 = (1^T \cdot A)^T}$$

Real networks are sparse

- The adjacency matrix is typically sparse

  good for tractability !

$A =$



protein
interaction
network

Real networks are sparse

- Complete graphs: the maximum # of links out of N nodes is

$$L_{max} = ½N(N-1) \quad \text{undirected}$$
$$N(N-1) \quad \text{directed}$$

- The maximum average degree is

$$<k>_{max} = N-1 = 2\, L_{max}/N \quad \text{undirected}$$
$$L_{max}/N \quad \text{directed}$$

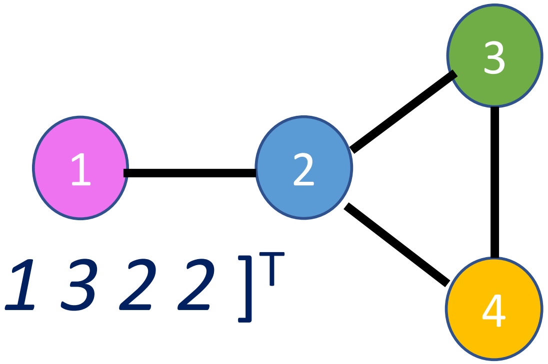
- In real networks $L \ll L_{max}$ and $<k> \ll N-1$

network	**type**	**N**	**L**	**<k>**
www	directed	3.2×10^5	1.5×10^6	4.60
Protein	directed	1870	4470	2.39
Co-authorships	undirected	23133	93439	8.08
Movie actors	undirected	7×10^5	29×10^6	83.7

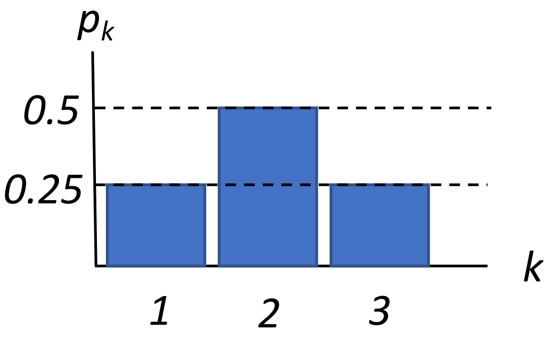
Degree distribution

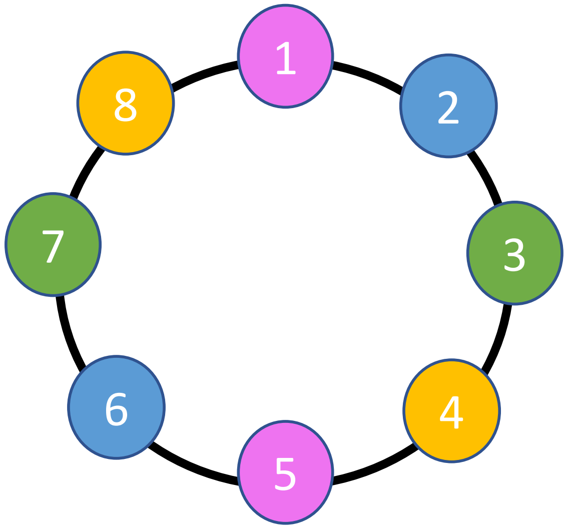
- Degree distribution p_k, a probability distr.

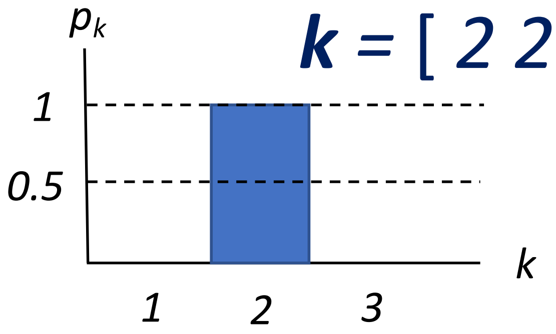
  p_k is the fraction of nodes that have degree exactly equal to *k* (i.e., # of nodes with that degree / *N*)



$\mathbf{k} = [\ 1\ 3\ 2\ 2\ ]^T$





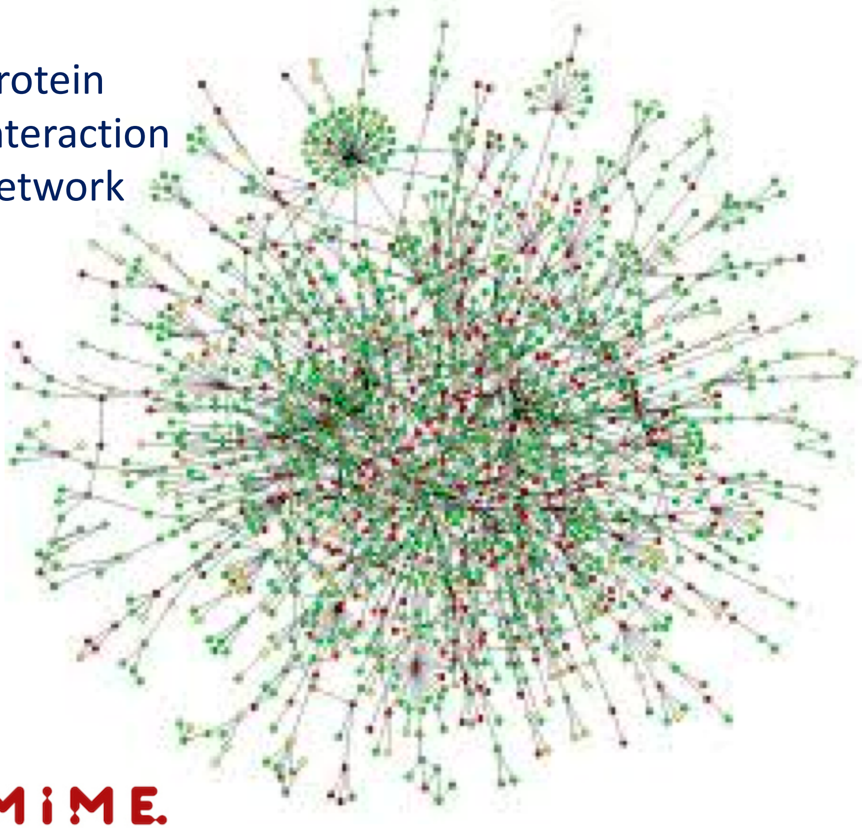


$\mathbf{k} = [\ 2\ 2\ 2\ 2\ 2\ 2\ 2\ 2\ ]^T$

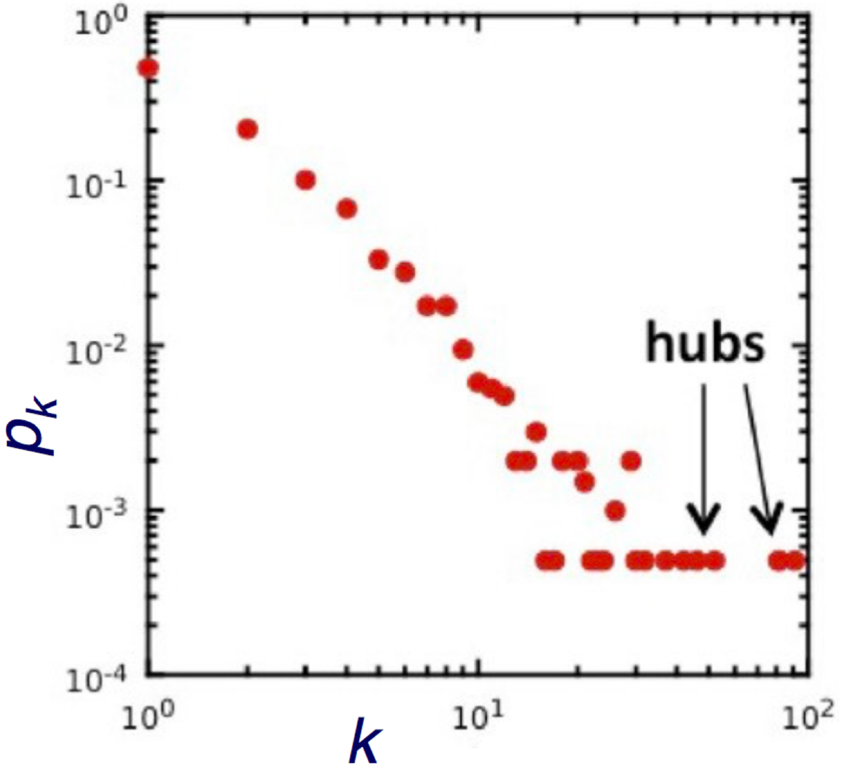
Degree distribution

- In real world (large) networks, degree distribution is typically heavy-tailed

nodes with high degree = hubs



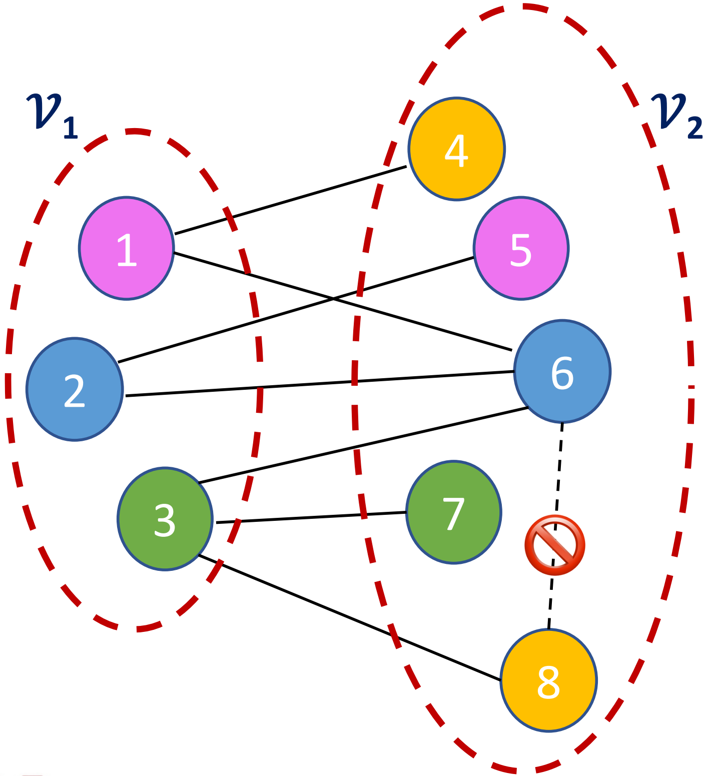
protein interaction network

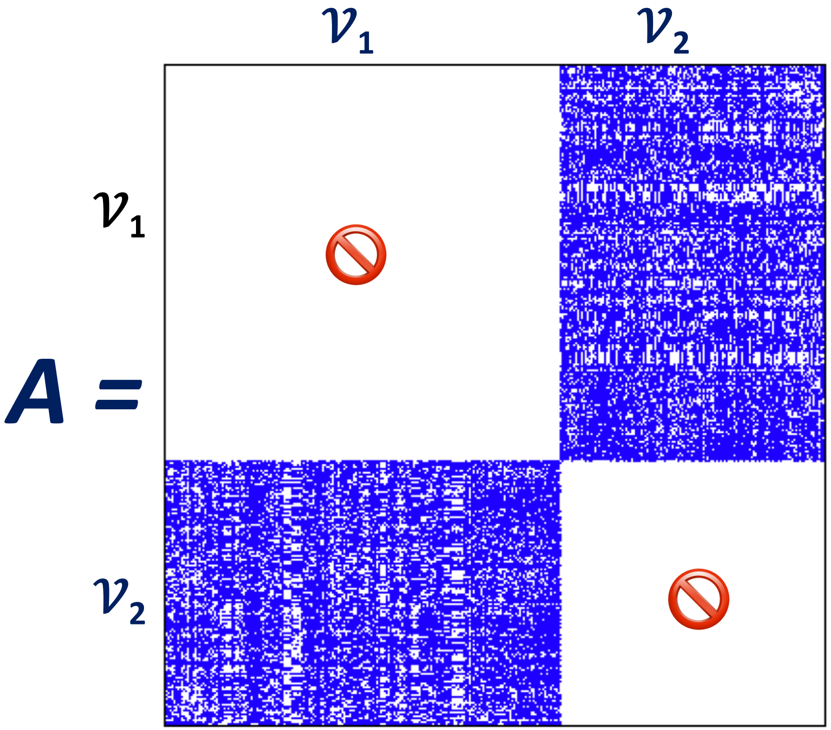


Bipartite graphs

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

$(i,j) \in \mathcal{E}$ if and only if $i \in \mathcal{V}_1$ and $j \in \mathcal{V}_2$, or $j \in \mathcal{V}_1$ and $i \in \mathcal{V}_2$





Bipartite graph example

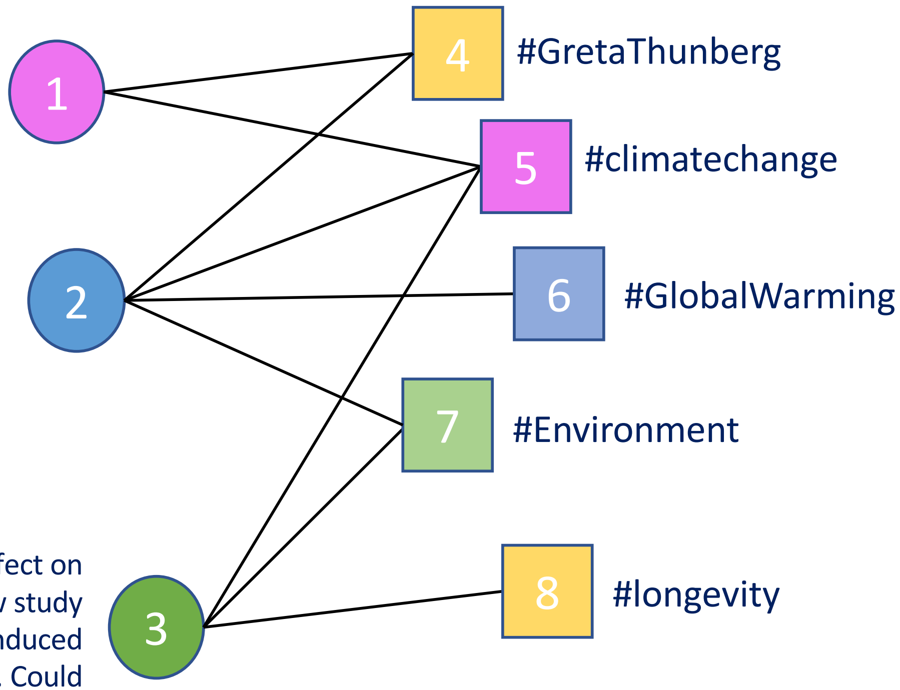
Tweets

Hashtags

those who think they are crazy enough to change the world eventually do. #climatechange #ClimateCrisis #ClimateAction #GretaThunberg #Greta

Hopefully these kids will succeed where past generations have failed. #TheResistance #FBR #ClimateChange #Environment #GlobalWarming #GretaThunberg

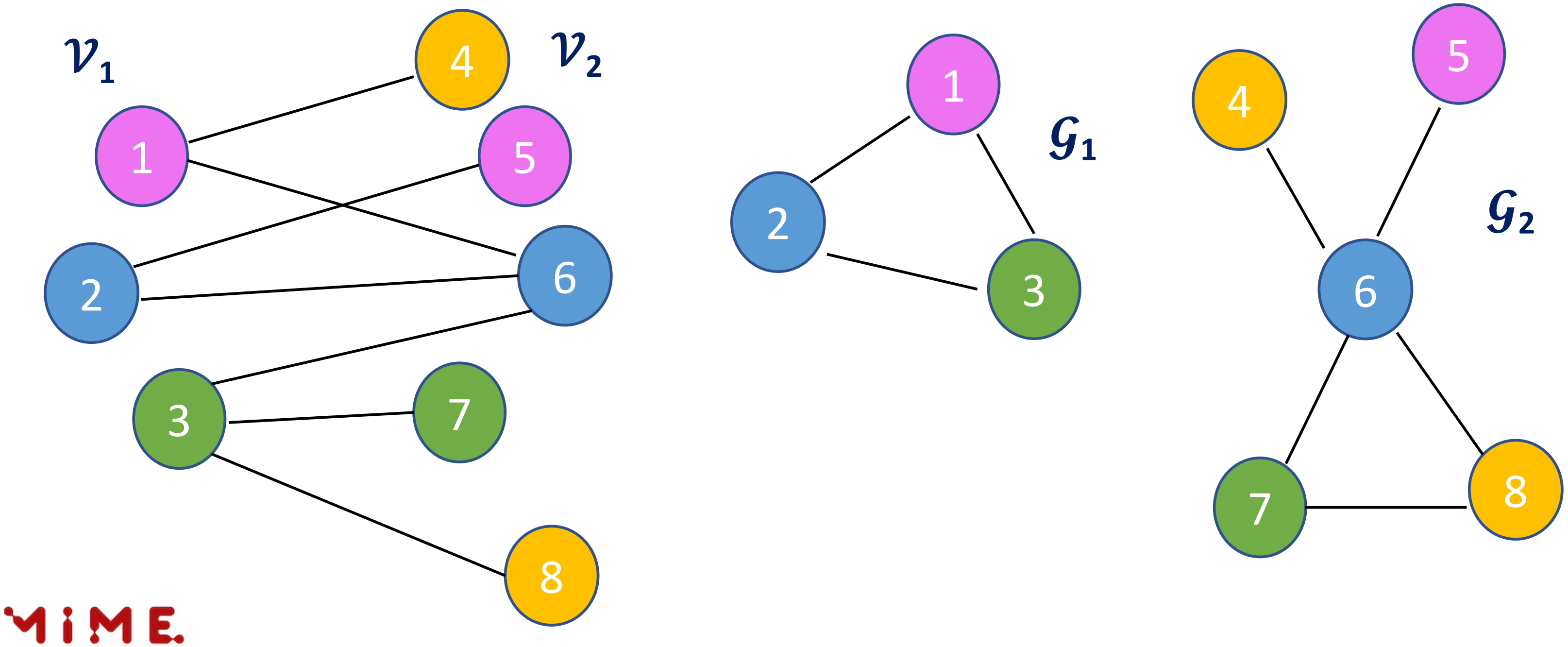
The #environment can have a major effect on the human cardiovascular system. A new study has found an increase in heat-induced #heartattack risk in recent years. Could #ClimateChange be a risk factor? #longevity



Projections

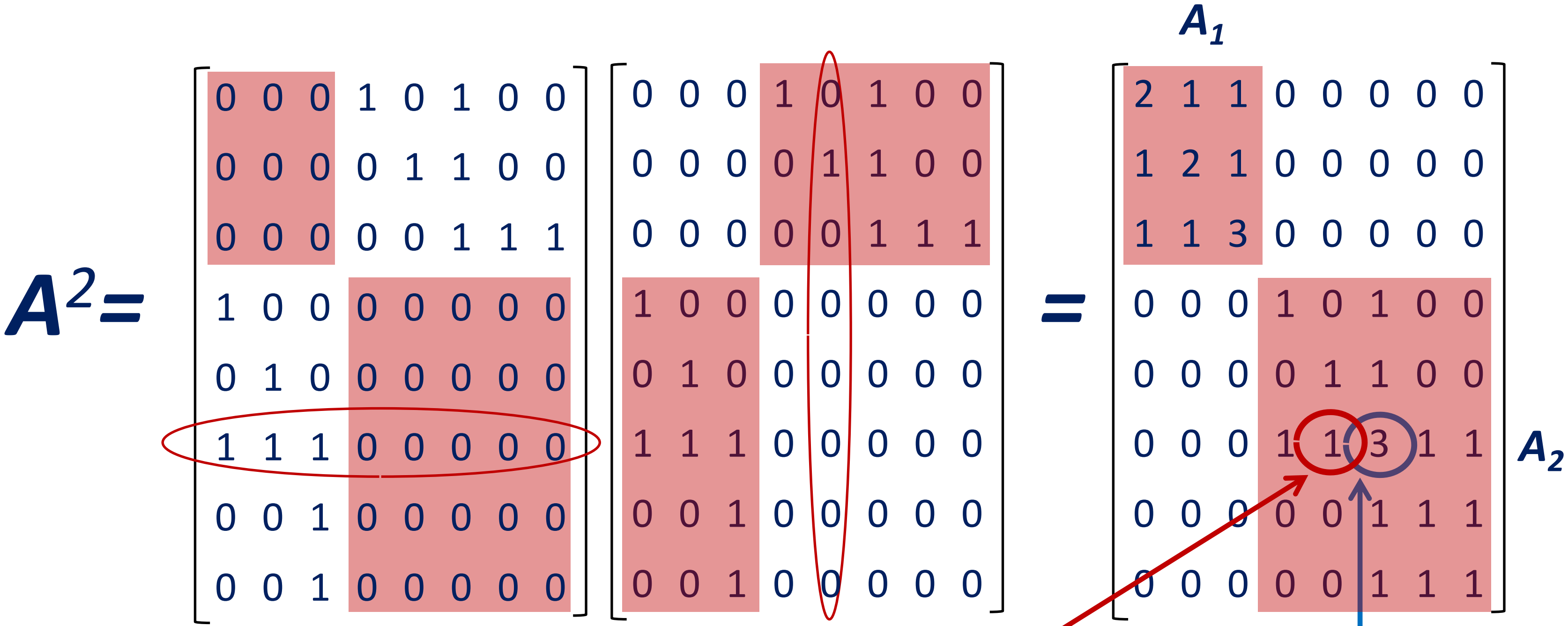
- For a bipartite graph $\mathcal{G}(\mathcal{V}_1 \cup \mathcal{V}_2, \mathcal{E})$, the projection on $\mathcal{V}_1$ is the graph $\mathcal{G}_1(\mathcal{V}_1, \mathcal{E}_1)$ where $(i,j) \in \mathcal{E}_1$ if and only if i and j have a common neighbour k

i.e., a node $k \in \mathcal{V}_2$ such that $(i,k) \in \mathcal{E}$ and $(k,j) \in \mathcal{E}$



Projections

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



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

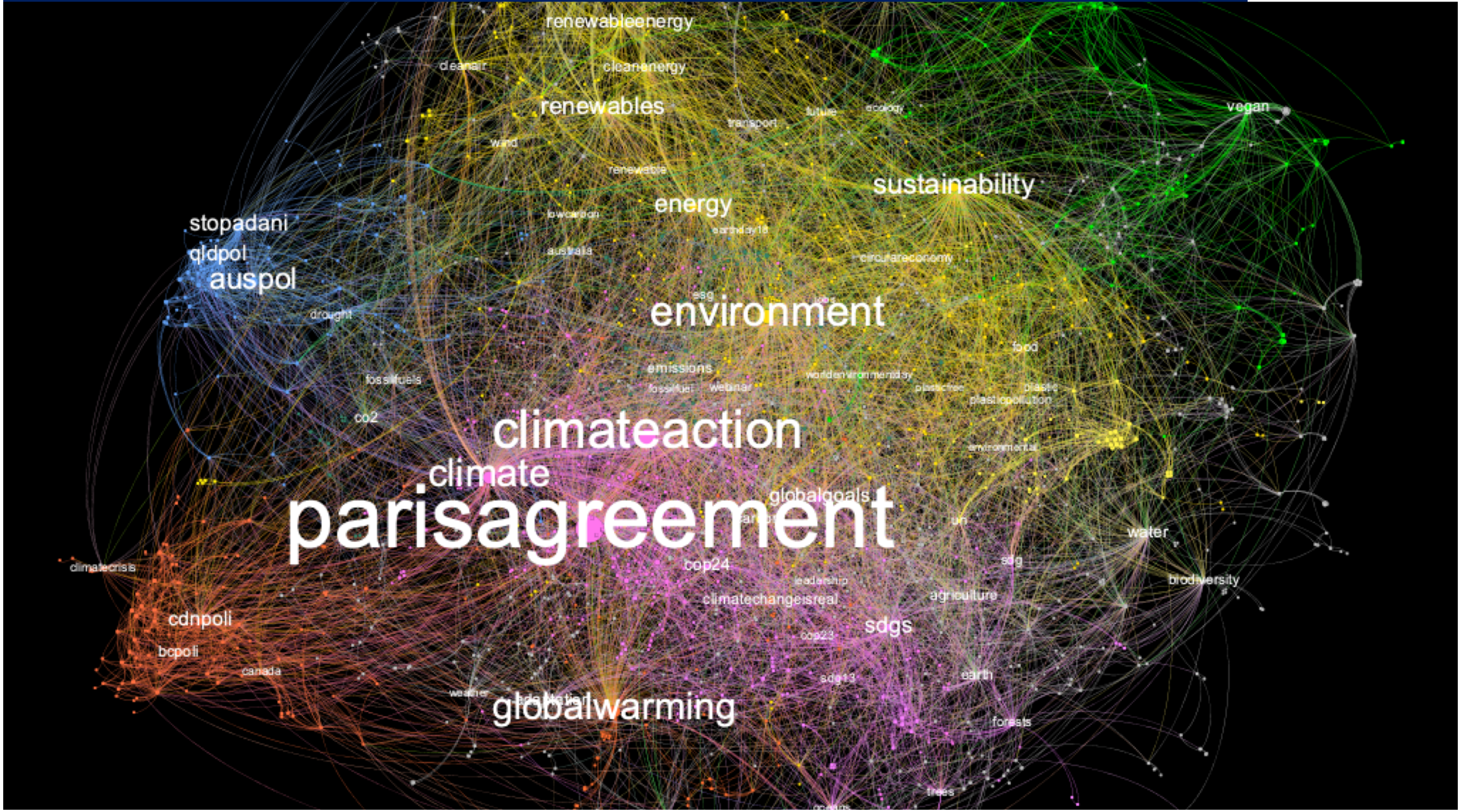
A_1

A_2

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

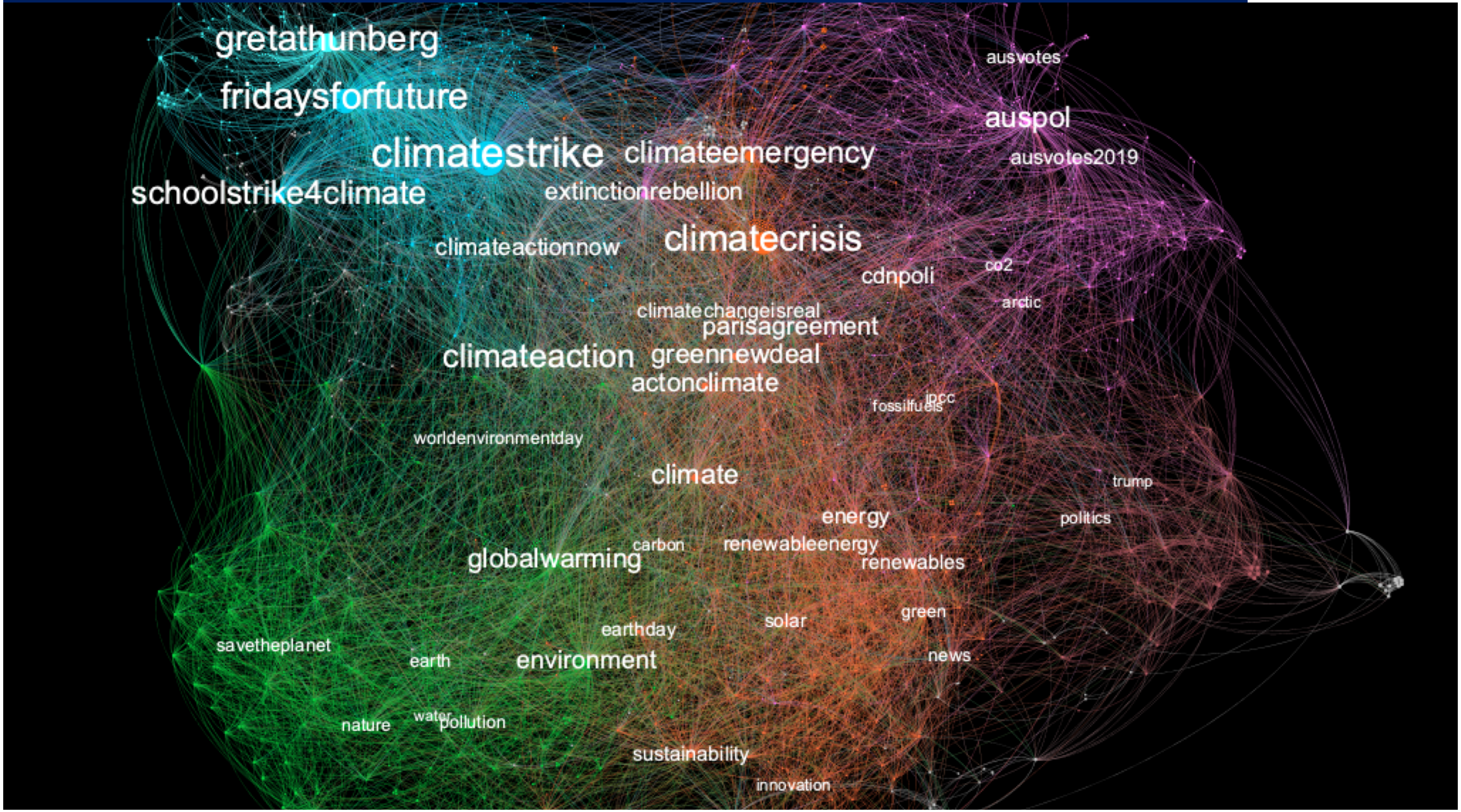
\# of neighbors of *i=6*

Projection example



#climateaction tweets before Greta Thunberg

Projection example



#climateaction tweets after Greta Thunberg

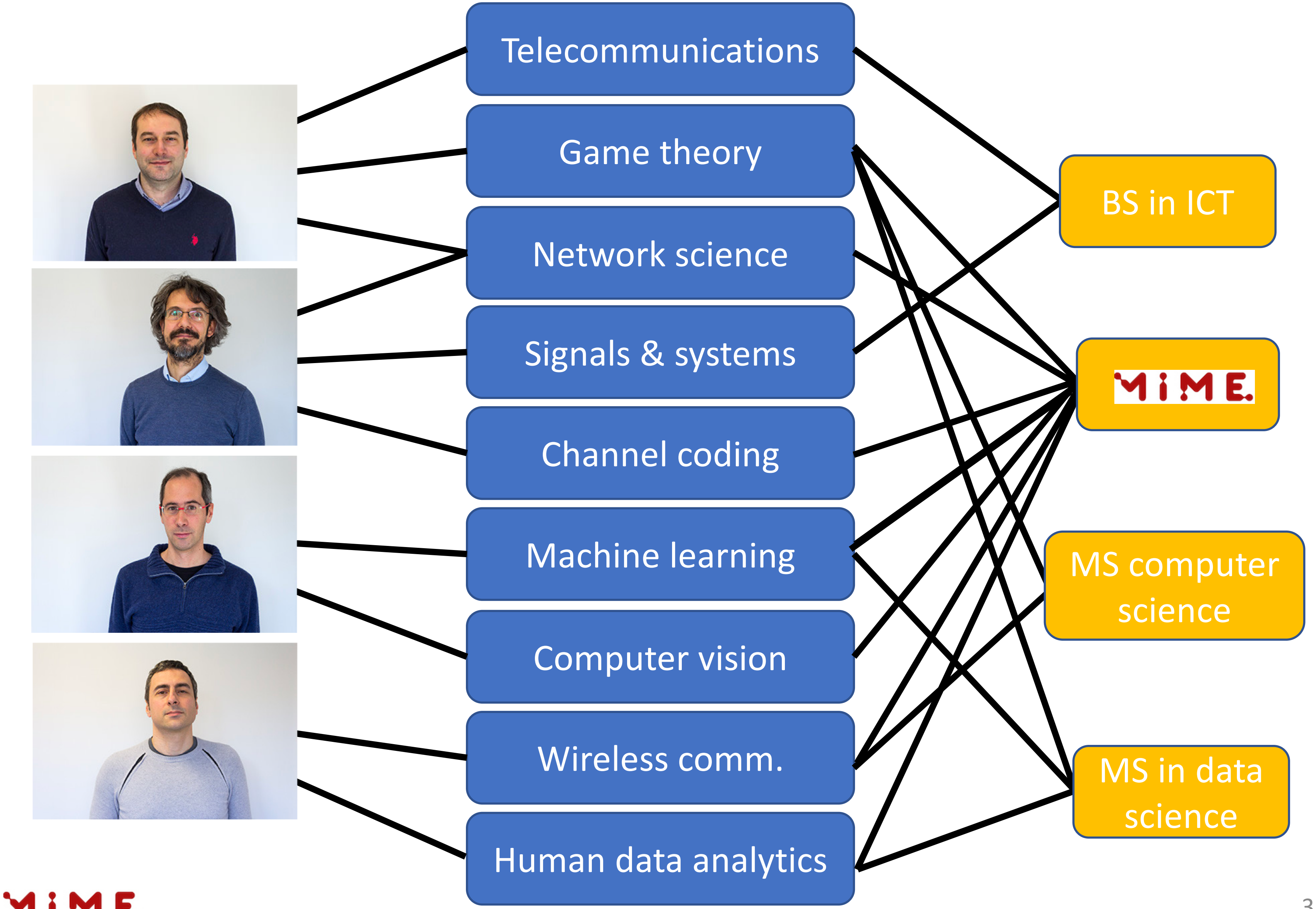
Meaning of projections

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

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

  from a mathematical viewpoint being part of the same group can be interpreted in both ways, e.g., “actors in the same movie” or “movies sharing the same actor”

Tri-partite graphs



Telecommunications

Game theory

Network science

Signals & systems

Channel coding

Machine learning

Computer vision

Wireless comm.

Human data analytics

BS in ICT

MS computer science

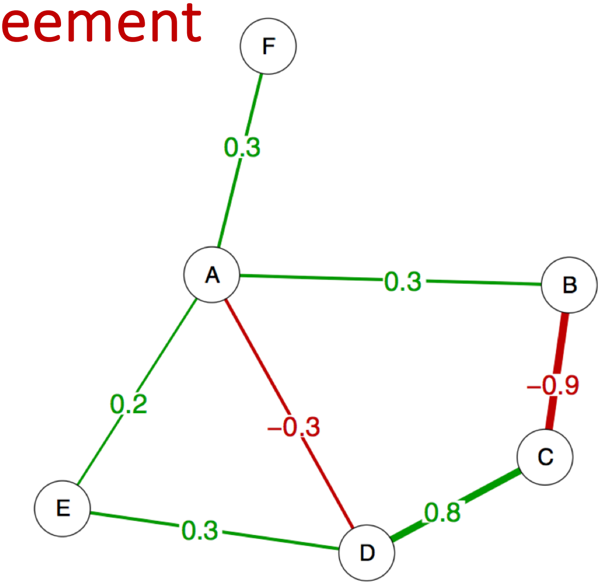
MS in data science

Signed graphs

- ❑ Edges can have signed values

  positive if there is an agreement between nodes

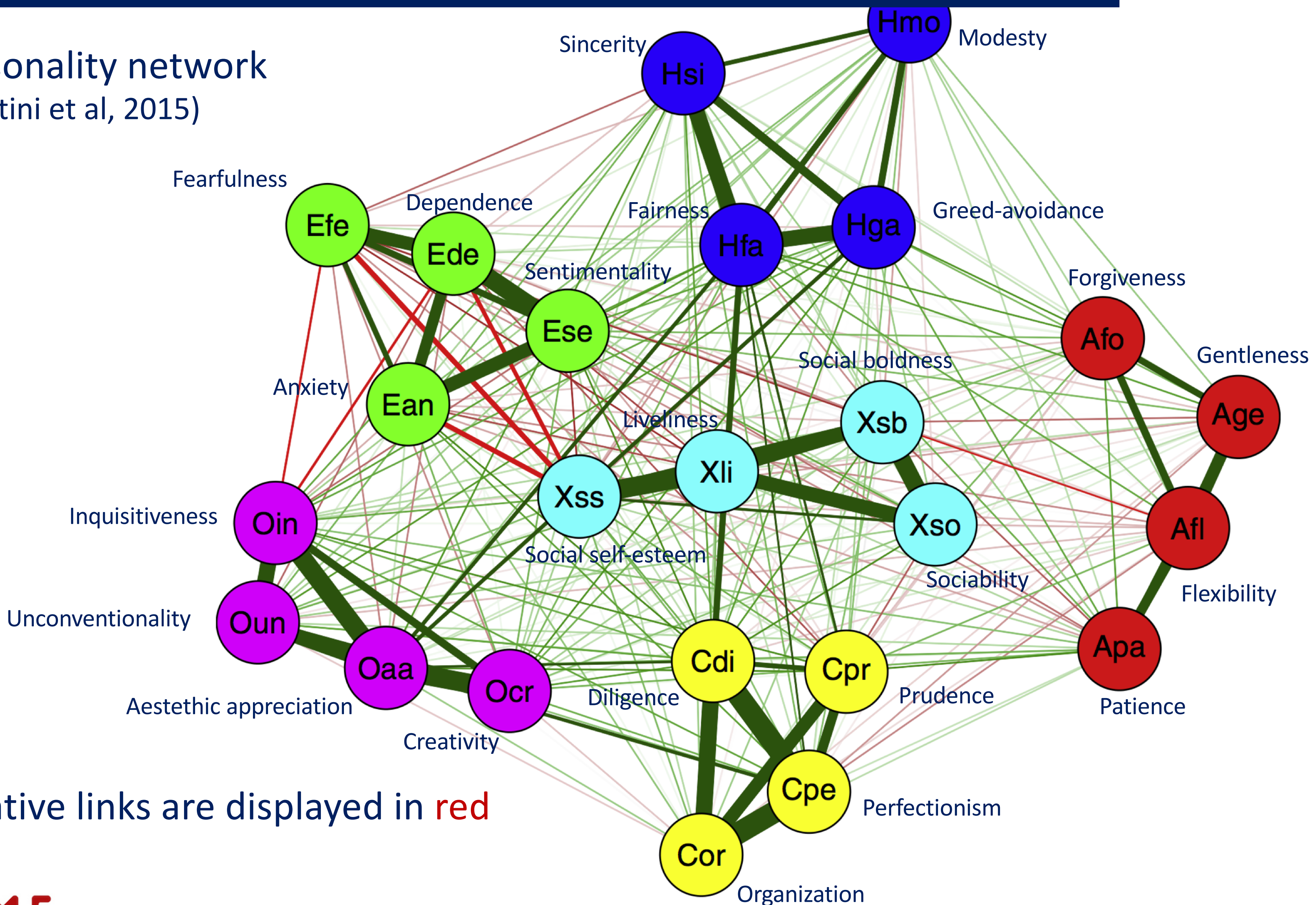
  negative if there's a disagreement



- ❑ This is typical of correlation networks

Signed graph example

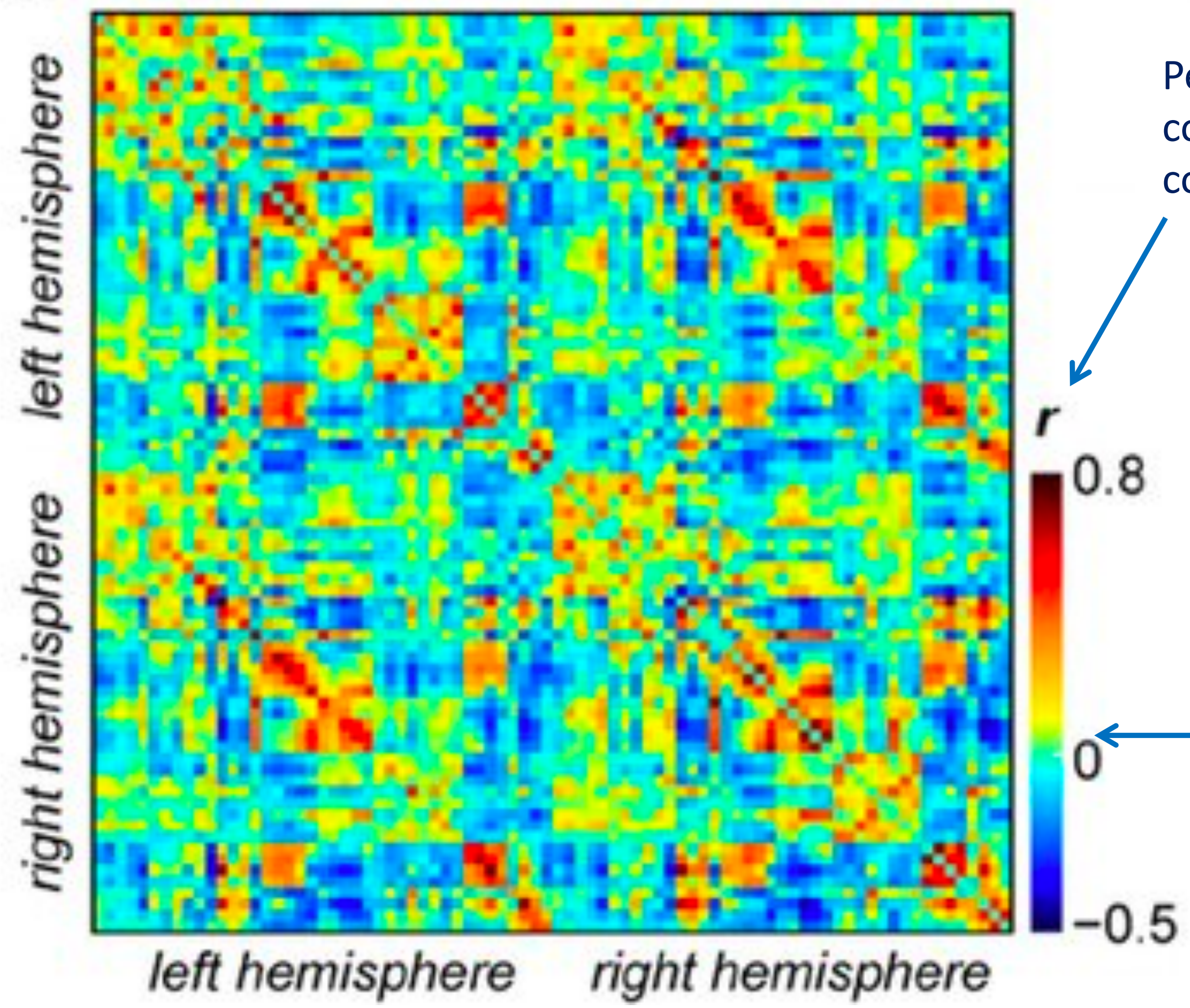
A personality network
(Costantini et al, 2015)



Negative links are displayed in red

Signed graph example (cont’d)

An fMRI adjacency matrix
(fMRI = functional magnetic resonance imaging)



Pearson’s correlation coefficient

No correlation

Negative correlation

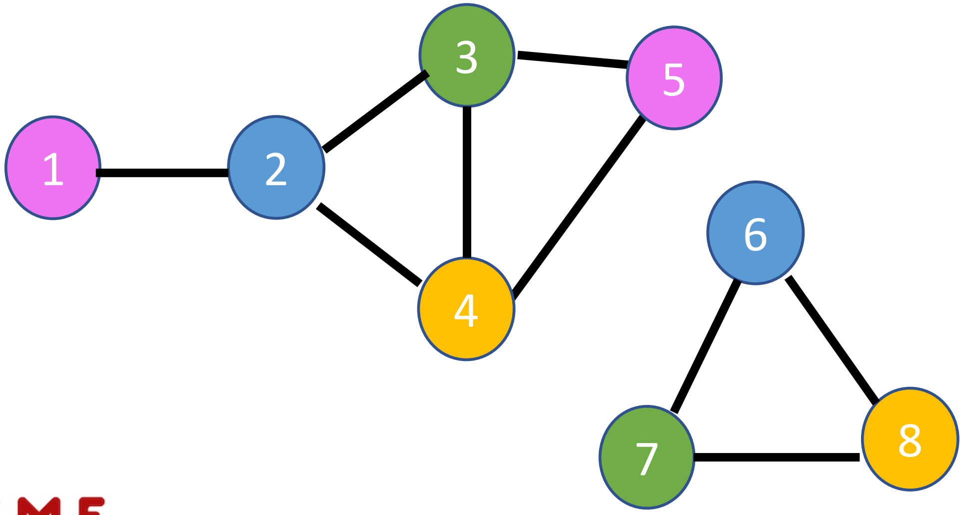
Connectivity

- 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)



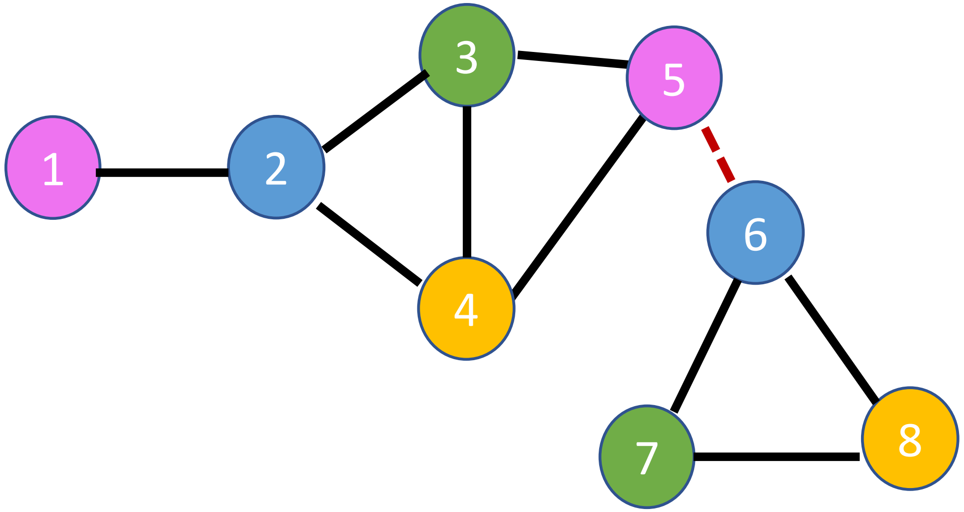
$$A=\begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \end{bmatrix}$$

block-diagonal matrix

Connectivity

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

its removal would make the network disconnected



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

Connectivity in directed nets

For directed networks we distinguish between

- ☐ Strongly connected components

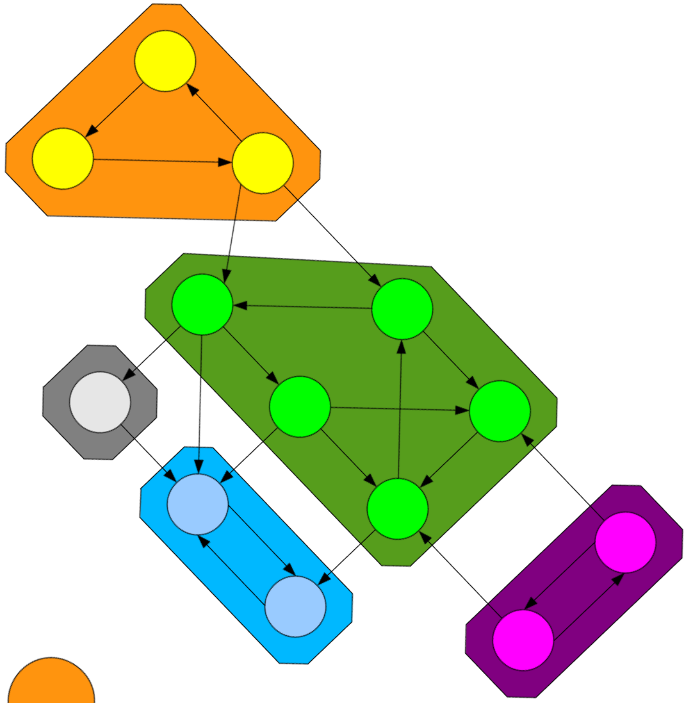
  where *i*→*j* and *j*→*i* for all choices of (*i*,*j*) in the component

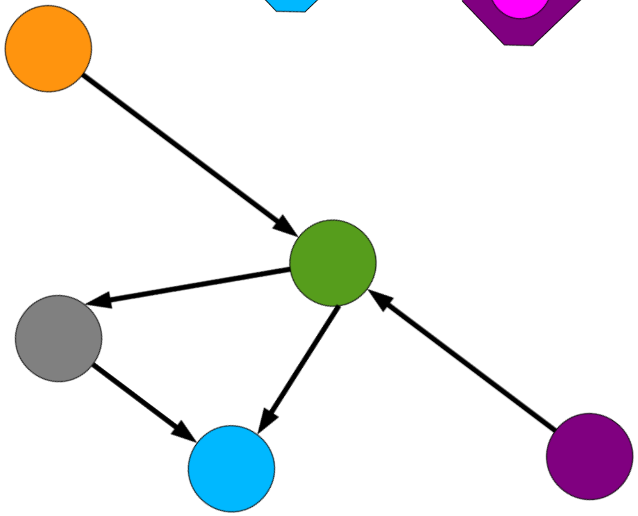
- ☐ Weakly connected components

  connected in the undirected sense (i.e., disregard link directions)

Condensation graph

- ☐ Strong connectivity induces a partition in disjoint strongly connected sets $\mathcal{V}_1, \mathcal{V}_2, \ldots, \mathcal{V}_K$
- ☐ By reinterpreting the sets as nodes we obtain a condensation graph $\mathcal{G}^*$ where $i \rightarrow j$ is an edge if a connection exists between sets $\mathcal{V}_i \rightarrow \mathcal{V}_j$





Properties of $\mathcal{G}^*$

- $\mathcal{G}^*$ does not contain cycles

  otherwise the sets in the cycle would be strongly connected

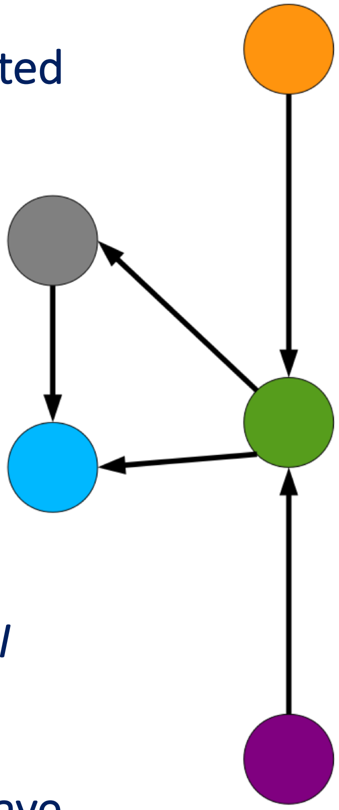
- $\mathcal{G}^*$ has at least one root and one leaf

  and every node in the graph can be reached from one of the roots

- $\mathcal{G}^*$ allows a particular reordering

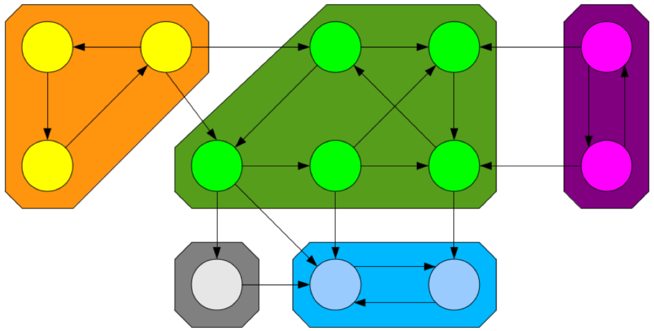
  where node n_i does not reach any of the nodes n_j with $j<l$

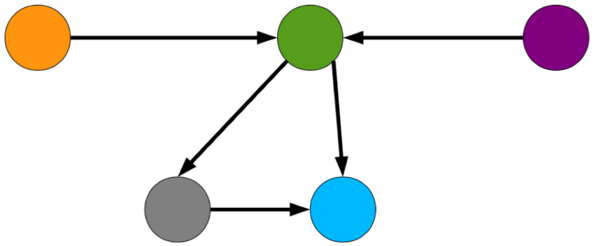
  procedure: identify a root n_1 and remove it from the network, then identify a new root; cycle until all nodes have been selected



Condensation graph

- The condensation graph ordering induces a block-lower-triangular matrix structure on the adjacency matrix





$$A=\left[\begin{array}{ccc|cc|ccccc|c|cc}
 & & 3 & & 2 & & & & & 5 & 1 & & 2 \\
 & \frac{1}{3} & & & & & & & & & & & \\
 & & 1 & & & & & & & & & & \\
1 & & & & & & & & & & & & \\
\hline
 & & & & \frac{1}{2} & & & & & & & & \\
 & & & \frac{1}{2} & & & & & & & & & \\
\hline
 & \frac{1}{3} & & & & & & & & \frac{1}{2} & & & \\
 & & & \frac{1}{2} & & \frac{1}{2} & & & \frac{1}{3} & & & & \\
 & \frac{1}{3} & & & & \frac{1}{2} & & & & & & & \\
 & & & & & & & \frac{1}{3} & & & & & \\
 & & & & \frac{1}{2} & & 1 & & \frac{1}{3} & & & & \\
\hline
 & & & & & & & \frac{1}{3} & & & 0 & & \\
\hline
 & & & & & & & \frac{1}{3} & \frac{1}{3} & & 1 & & 1 \\
 & & & & & & & & & \frac{1}{2} & & 1 &
\end{array}\right]$$

blocks in the diagonal are irreducible = no block-diagonal form !

Readings

- ☐ A.L. Barabási, Network science

  http://barabasi.com/networksciencebook

  Ch.2 “Graph theory”