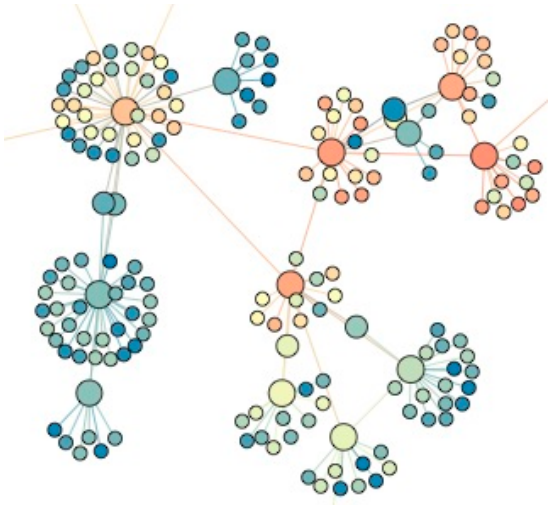


SNA L1C: degree, connected components

Lada Adamic

school of information

Network elements: edges

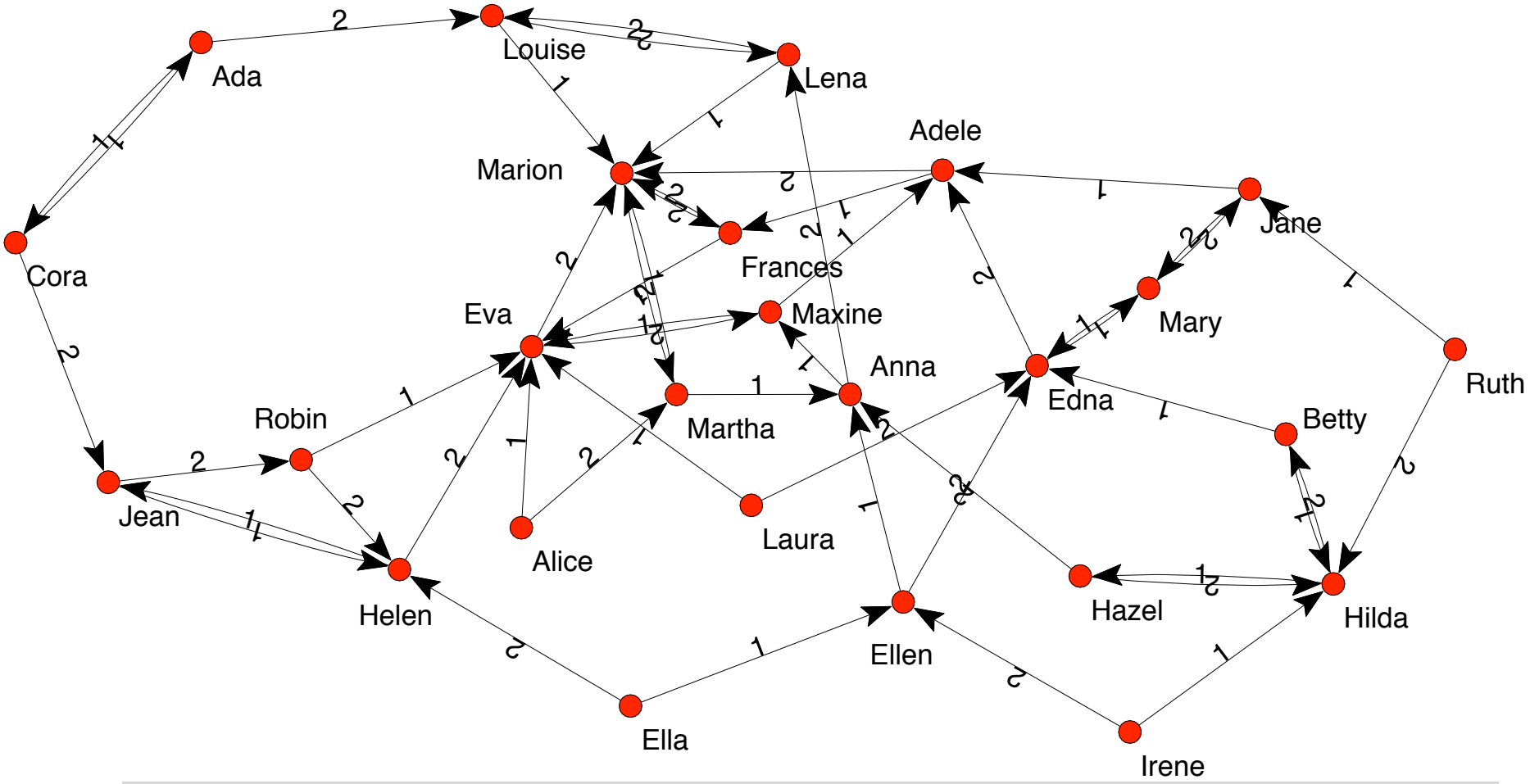
- Directed (also called arcs, links)
  - A -> B
    - A likes B, A gave a gift to B, A is B's child
- Undirected
  - A <-> B or A – B
    - A and B like each other
    - A and B are siblings
    - A and B are co-authors

Edge attributes

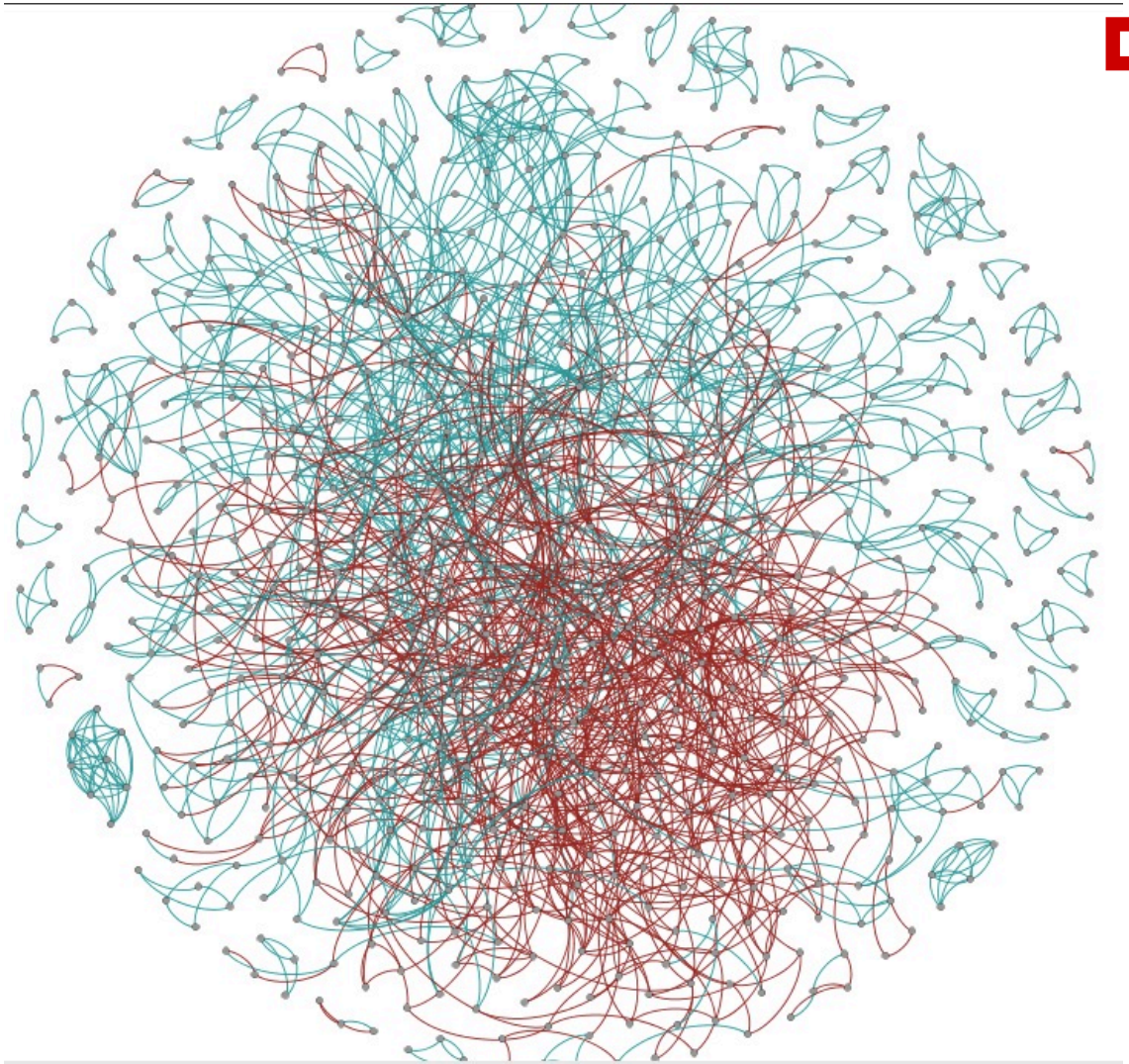
- Examples
  - weight (e.g. frequency of communication)
  - ranking (best friend, second best friend…)
  - type (friend, relative, co-worker)
  - properties depending on the structure of the rest of the graph: e.g. betweenness

Directed networks

- girls' school dormitory dining-table partners, 1st and 2nd choices
  (Moreno, *The sociometry reader*, 1960)



Positive and negative weights



- e.g. one person trusting/distrusting another
  - Research challenge: How does one 'propagate' negative feelings in a social network? Is my enemy's enemy my friend?

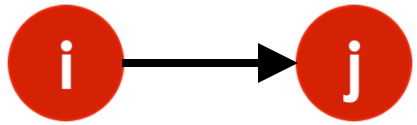
sample of positive & negative ratings from Epinions network

Data representation

- adjacency matrix
- edgelist
- adjacency list

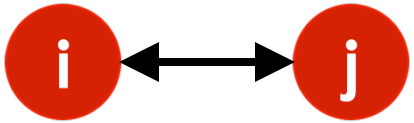
Adjacency matrices

- Representing edges (who is adjacent to whom) as a matrix
  - A_{ij} = 1 if node *j* has an edge to node *i*
    = 0 if node *j* does not have an edge to i

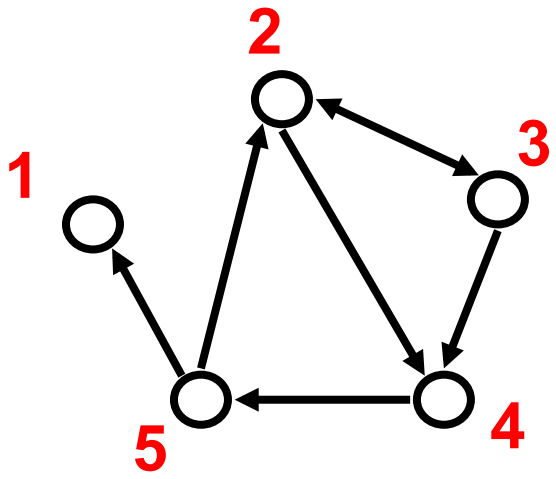


  - A_{ii} = 0 unless the network has self-loops

  - $A_{ij} = A_{ji}$ if the network is undirected, or if *i* and *j* share a reciprocated edge



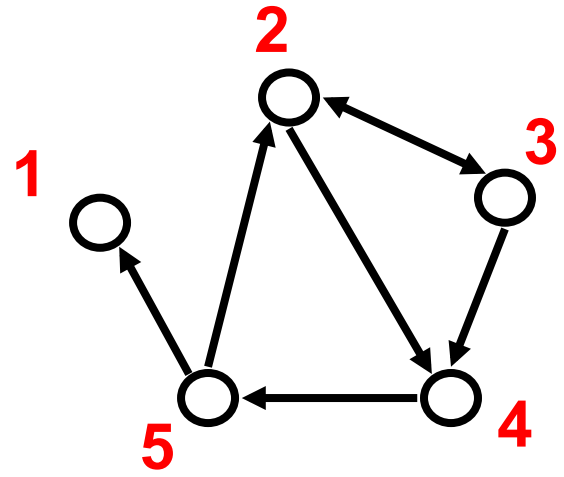
Example adjacency matrix



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

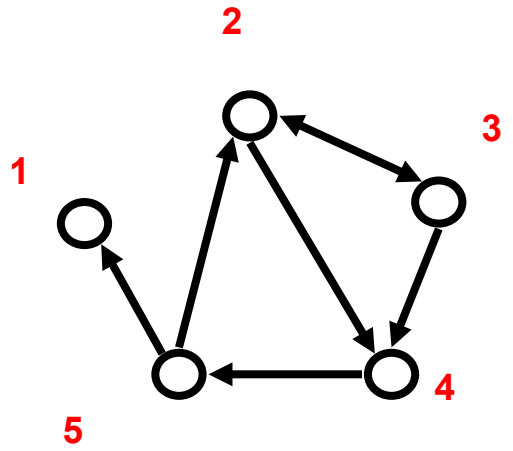
Edge list

- Edge list
  - 2, 3
  - 2, 4
  - 3, 2
  - 3, 4
  - 4, 5
  - 5, 2
  - 5, 1



Adjacency lists

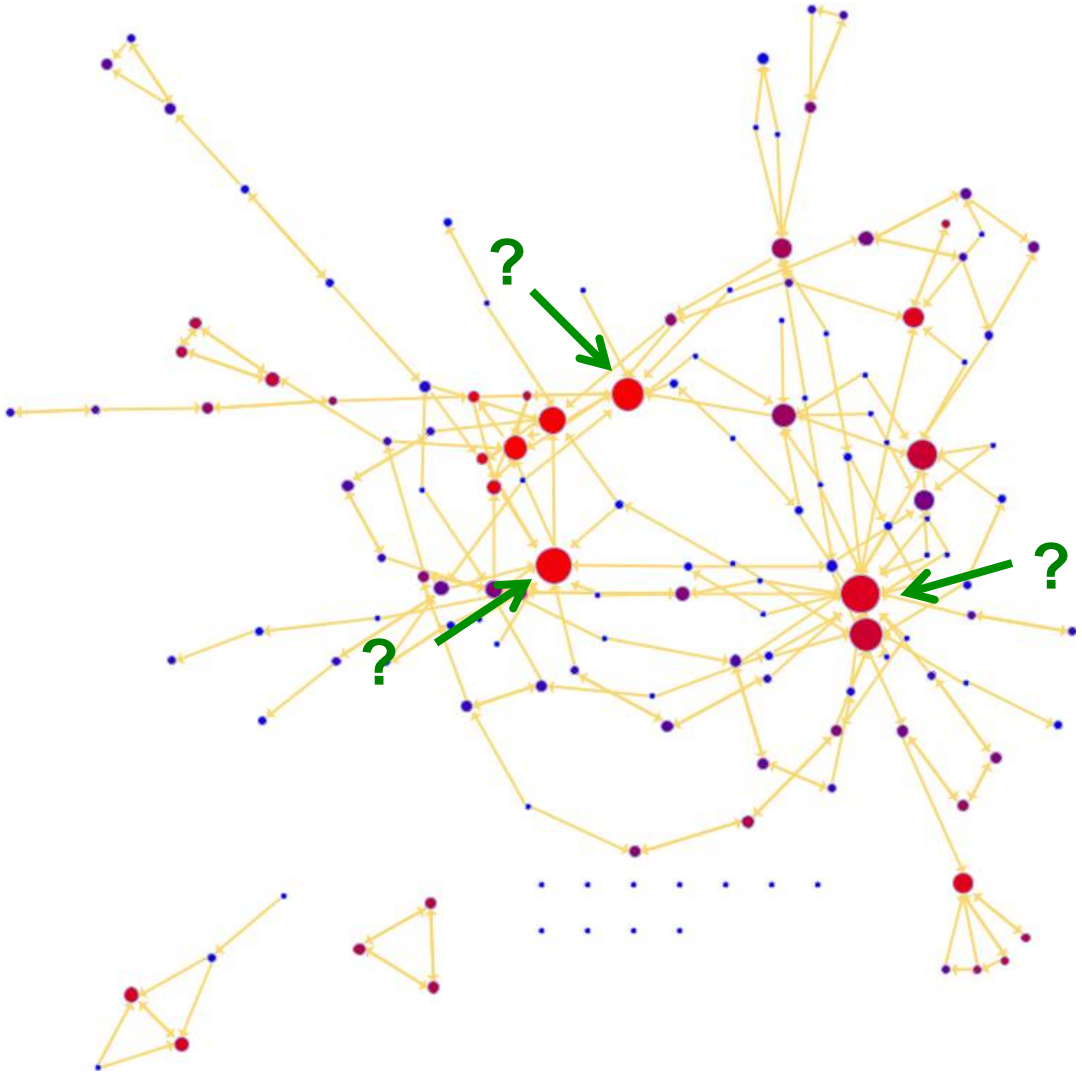
- Adjacency list
  - is easier to work with if network is
    - large
    - sparse
  - quickly retrieve all neighbors for a node
    - 1:
    - 2: 3 4
    - 3: 2 4
    - 4: 5
    - 5: 1 2



Computing metrics

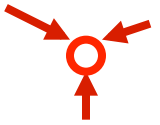
- degree & degree distribution
- connected components

Degree: which node has the most edges?



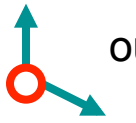
Nodes

- Node network properties
  - from immediate connections
    - indegree
      how many directed edges (arcs) are incident on a node



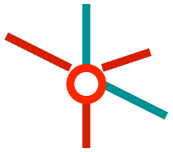
indegree=3

    - outdegree
      how many directed edges (arcs) originate at a node



outdegree=2

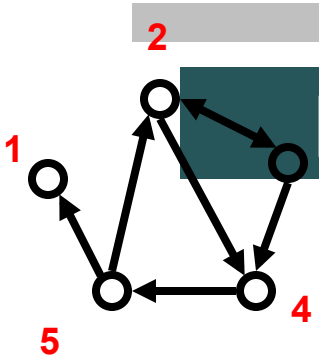
    - degree (in or out)
      number of edges incident on a node



degree=5

  - from the entire graph
    - centrality (betweenness, closeness)

Node degree from matrix values



- Indegree = $\sum_{j=1}^{n} A_{ij}$

A =

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

example: indegree for node 3 is 2, which we obtain by summing the number of non-zero entries in the 3rd *row*

$$\sum_{j=1}^{n} A_{3j}$$

- Outdegree = $\sum_{i=1}^{n} A_{ij}$

A =

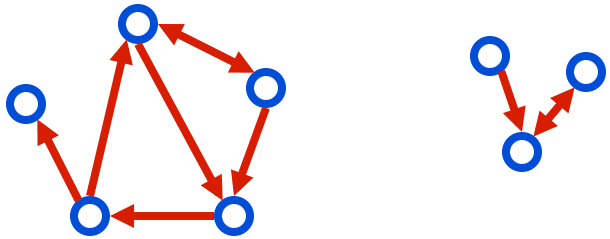
$$\begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 \end{pmatrix}$$

example: the indegree for node 3 is 1, which we obtain by summing the number of non-zero entries in the 3rd *column*

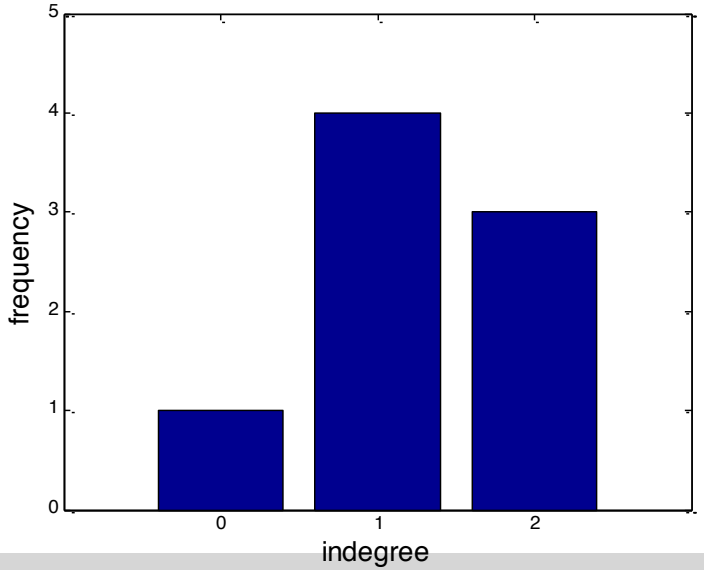
$$\sum_{i=1}^{n} A_{i3}$$

Network metrics: degree sequence and degree distribution

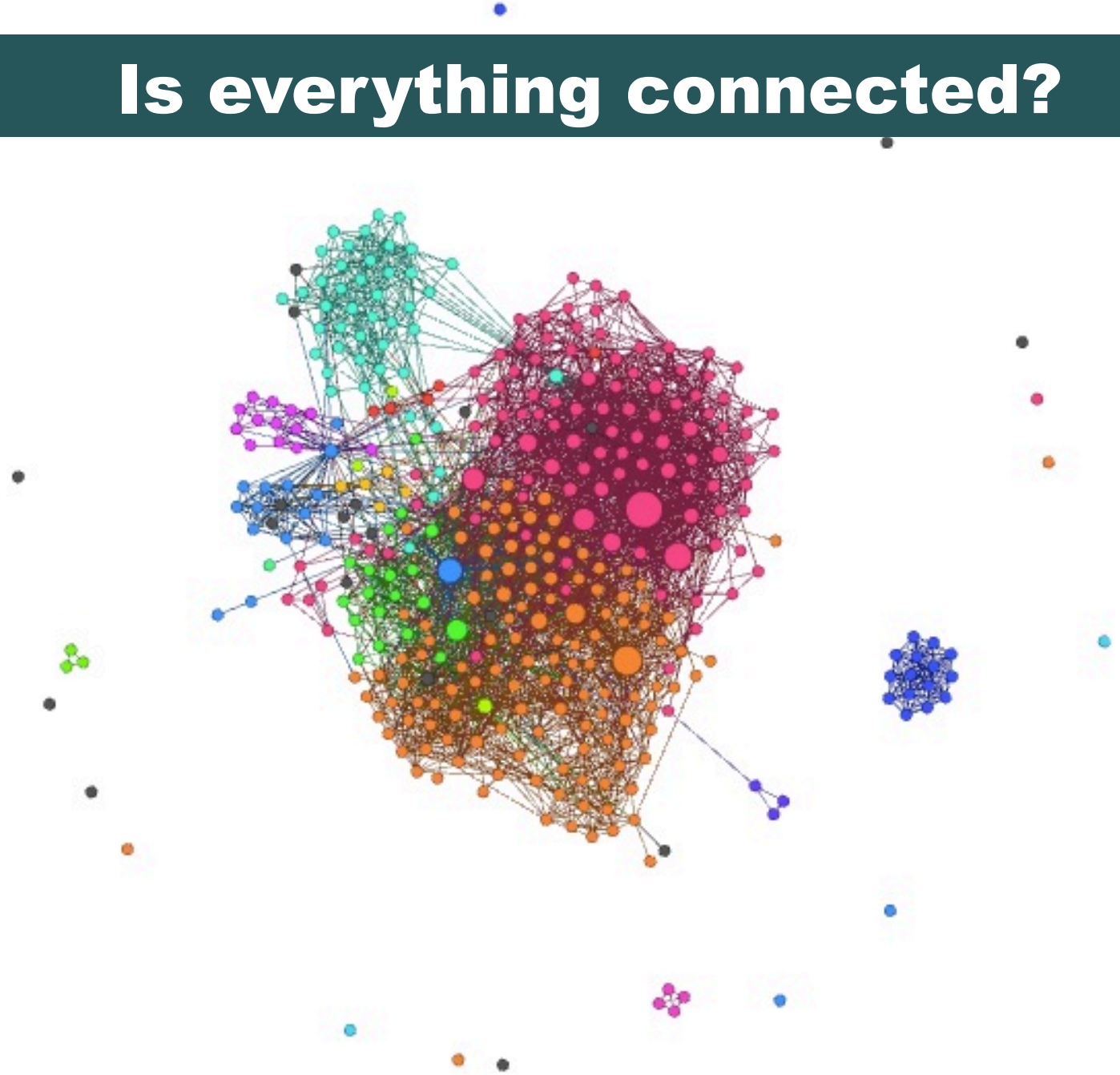
- Degree sequence: An ordered list of the (in,out) degree of each node
  - In-degree sequence:
    - [2, 2, 2, 1, 1, 1, 1, 0]
  - Out-degree sequence:
    - [2, 2, 2, 2, 1, 1, 1, 0]
  - (undirected) degree sequence:
    - [3, 3, 3, 2, 2, 1, 1, 1]



- Degree distribution: A frequency count of the occurrence of each degree
  - In-degree distribution:
    - [(2,3) (1,4) (0,1)]
  - Out-degree distribution:
    - [(2,4) (1,3) (0,1)]
  - (undirected) distribution:
    - [(3,3) (2,2) (1,3)]



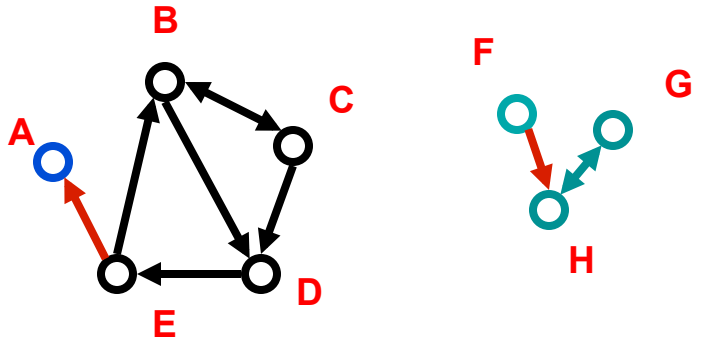
Is everything connected?



Connected components

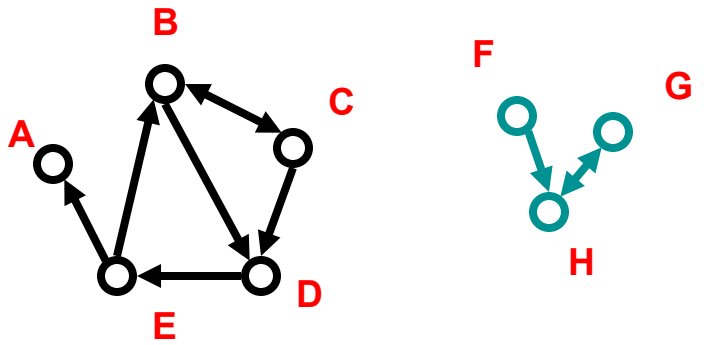
- Strongly connected components
  - Each node within the component can be reached from every other node in the component by following directed links

  - Strongly connected components
    - B C D E
    - A
    - G H
    - F



- Weakly connected components: every node can be reached from every other node by following links in either direction

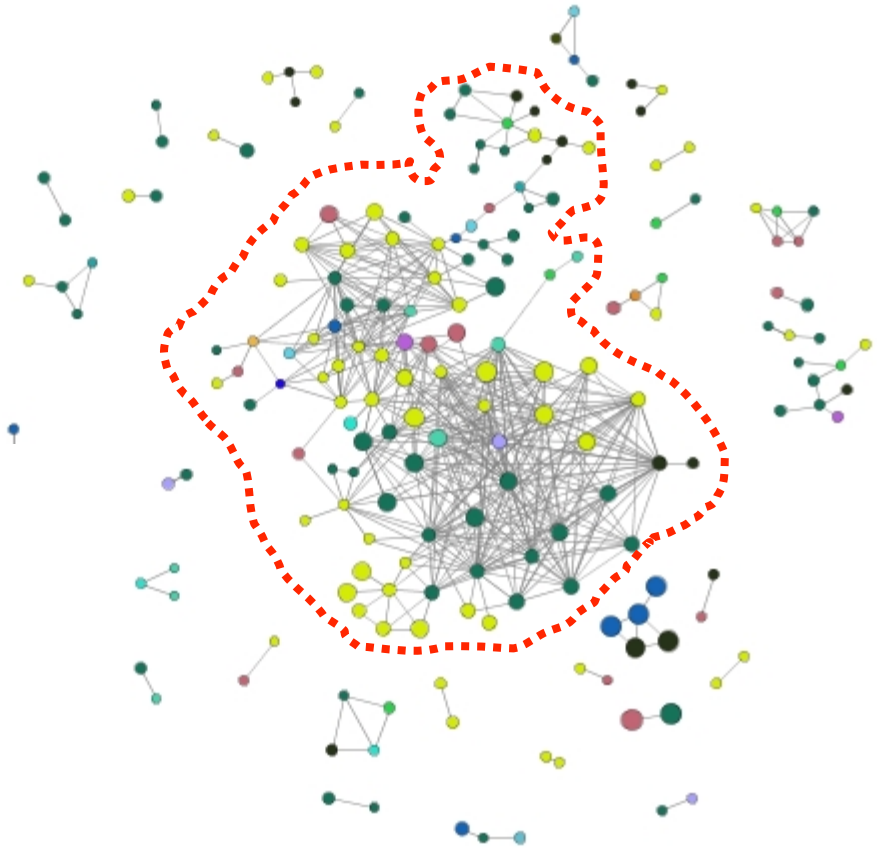
  - Weakly connected components
    - A B C D E
    - G H F



- In undirected networks one talks simply about 'connected components'

Giant component

- if the largest component encompasses a significant fraction of the graph, it is called the **giant component**



Recap

- Networks can be represented as matrices
- Usefull network metrics:
  - degree and degree distribution
  - connected components
    - strong
    - weak
    - giant