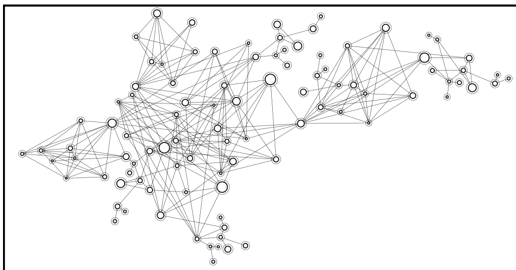


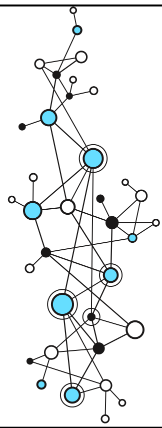
Network science in biology

measuring, visualizing and modelling real world complex networks

Petra Vertes

A quick show of hands

1. Experience with networks:
   A. None
   B. From popular media or general talk
   C. Read textbook/papers or did a course
2. Experience coding
   A. Yes (matlab?)
   B. No
3. Experience reading scientific papers
   A. Yes
   B. No

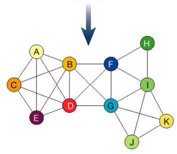


Overview of topics

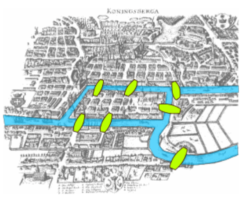
1. What is a network? – examples from social and biological sciences.
2. Constructing and representing complex networks.
3. Topological properties of networks – how to measure them and why they matter?
4. Network analysis in biological sciences – six examples
5. Generative modelling of networks – why and how?
6. Getting hold of data and code – tools and resources for network analysis

1.

What is a Network?

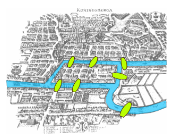


1736 - The birth of graph theory (Euler)



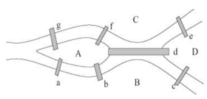
Seven bridges of Königsberg – can we cross all bridges exactly once?

1736 - The birth of graph theory (Euler)

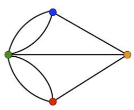


Seven bridges of Königsberg – can we cross all bridges exactly once?

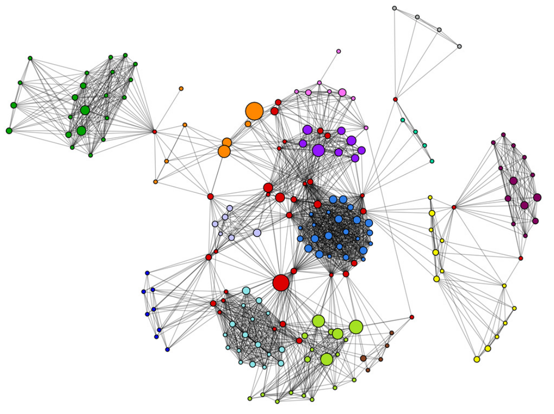
Euler was able to reformulate this question by representing each land mass by an abstract node and each bridge by a link.



This reveals that the desired path can only exist if the number of nodes with odd degree is zero (if start/end are the same) or two (if start/end are not the same), because for all other nodes need to arrive/leave through different edge.



Not the case here, so answer is: NO



Example 1: Social Network

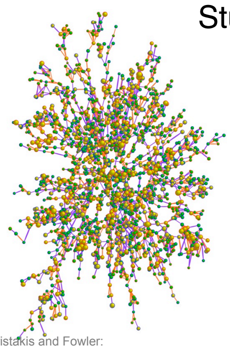
Nodes: People

Links: Social tie (defined as…)

Node Attributes:
- Location
- Gender
- Age
- Religion
- Political views
- Hobbies

Edge Attributes:
- Type of communication
- Direction of link
- Weight of link

Example 1: Farmingham Heart Study



Nodes: 12,067 subjects in a long-running federal study in the USA.

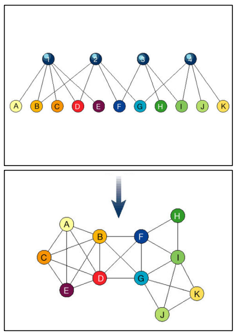
Links:
- friendship
- marriage
- kinship

Node Attributes:
32 years of medical records, including:
- body weight
- smoking habits
- mood
- genetics

Have data over time (32 years)

Christakis and Fowler:
Are friends genetically close? Are obesity, happiness and smoking 'contagious'?

Example 2: Actors' Network



Projection of bipartite network.

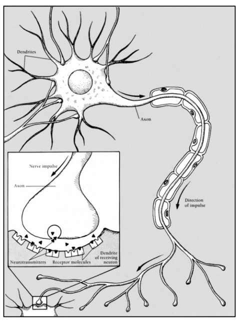
Nodes: actors

Edges: acted in same film

Node Attributes:
- Age
- Gender
- Individual awards

Edge Attributes:
- Film awards
- Success of film
- Language of film

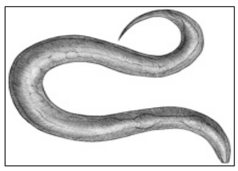
Example 3: Neural Network (I)

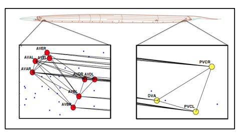


Nodes: Neurons

Edges: Synaptic connections (neurotransmitter release)

Example 4: Neural Network (II)





In C.elegans (only):

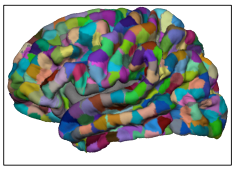
Nodes: known

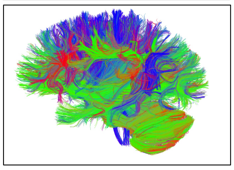
Edges: known

Node Attributes:
- Location
- Type (motor/sensory/interneuron)
- Birth time

Edge Attributes:
- Synapse or Gap junction
- Strength & direction

Example 5: Structural Brain Networks





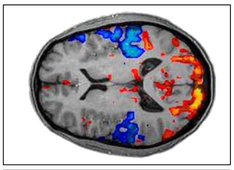
Nodes: Brain regions (in grey matter)

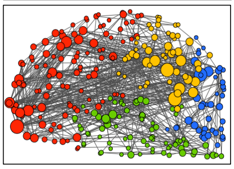
Links:
- White matter connections

Node Attributes:
- Location
- Function
- Metabolic cost

Edge Attributes:
- Weight of connections

Example 6: Functional Brain Networks





Nodes: Brain regions

Links:
- Functional connections (correlation in activity)

Node Attributes:
- Location
- Function
- Metabolic cost

Edge Attributes:
- Weight of connections

Can you think of other examples?

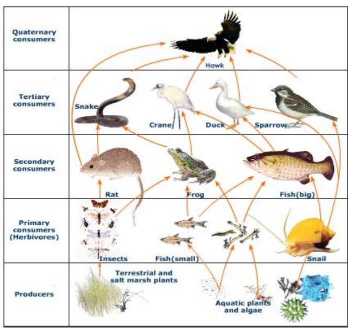
v

(we will cover several other biological networks later in the course)

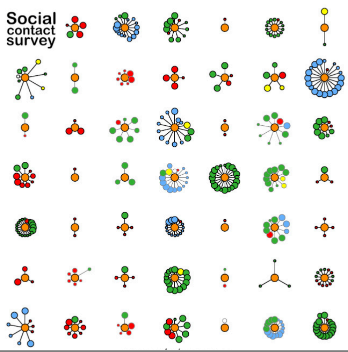
2.

Constructing and Representing Complex Networks

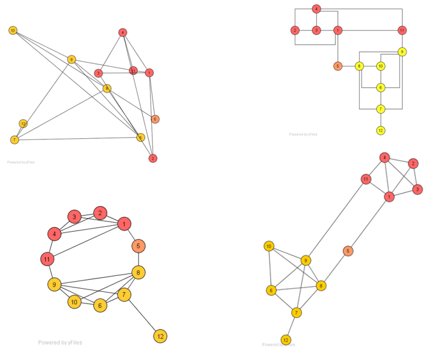
Data Collection



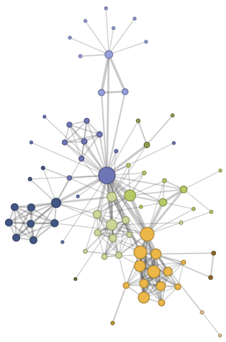
Data Collection 2.0



Representing a network

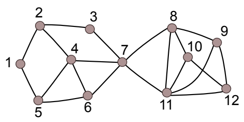


Force-directed graph drawing

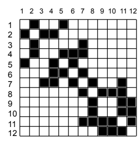


- Class of algorithms for positioning the nodes of a graph in 2D or 3D space to produce clear network visualizations.
- Spring-like attractive forces based on Hooke's law are used to attract pairs of endpoints of the graph's edges towards each other.
- Repulsive forces like those of electrically charged particles based on Coulomb's law are used to separate all pairs of nodes.
- Forces may be defined using functions that are not based on the physical behavior of springs and particles.
- The algorithms simulate the motion of nodes until equilibrium is reached under these forces (state of minimum energy). Other methods search more directly for energy minima (e.g.:simulated annealing).

Representing a network



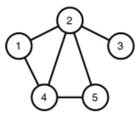
Spatial embedding



Adjacency matrix

Adjacency matrix elements a_{ij}

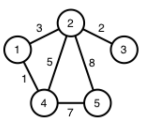
Undirected, unwieghted networks have a *symmetric* adjacency matrix with binary values $a_{ij} = a_{ji}$.



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

Weighted networks

In a *weighted* network a real number is attached to each edge, so that we obtain an adjacency matrix, usually denoted as w_{ij}.

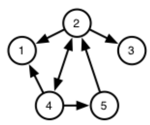


$$\begin{pmatrix} 0 & 3 & 0 & 1 & 0 \\ 3 & 0 & 2 & 5 & 8 \\ 0 & 2 & 0 & 0 & 0 \\ 1 & 5 & 0 & 0 & 7 \\ 0 & 8 & 0 & 7 & 0 \end{pmatrix}$$

Directed Networks

Directed networks in general have *asymmetric* a_{ij}

(which may be binary or weighted) .

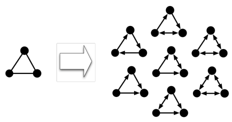


$$\begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 \end{pmatrix}$$

Weights and directionality increase complexity

Number of nodes N, number of edges E

Undirected networks: $E \leq N(N\text{-}1)/2$
Directed networks: $E \leq N(N\text{-}1)$ (double!)



Much more complex topology.

Thresholding and Binarizing

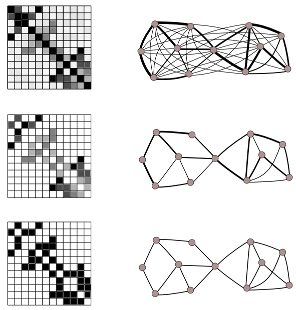
Weighted, Undirected

threshold

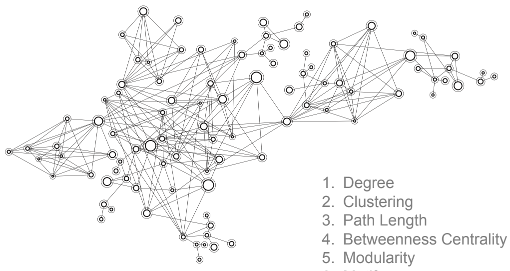
Weighted, Undirected

binarize

Unweighted, Undirected

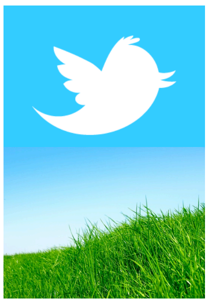


3. Network Measures (and why they matter)



1. Degree
2. Clustering
3. Path Length
4. Betweenness Centrality
5. Modularity
6. Motifs

Campaign Tactics

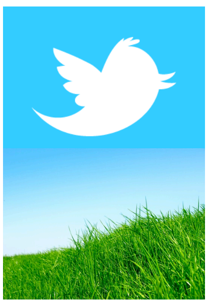


The pattern of Twitter activity generated by an artificial 'movement' differs from more spontaneous Twitter traffic:

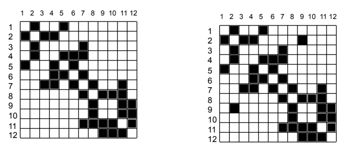
"Just looking at the structure of the network and how the information propagates, we have information about the nature of the message"

Menczer 2010, Election season.

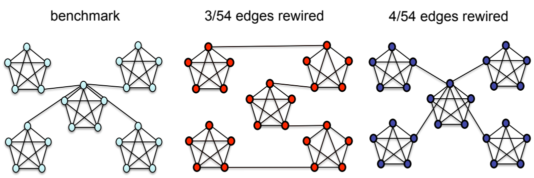
Grassroots or Astroturf?



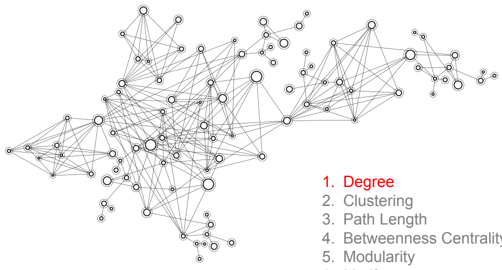
Comparing Networks



Comparing Networks

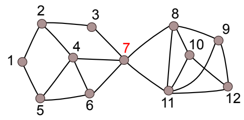


3. Network Measures (and why they matter)



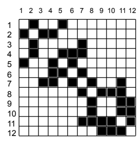
1. Degree
2. Clustering
3. Path Length
4. Betweenness Centrality
5. Modularity
6. Motifs

Degree (k) of a node

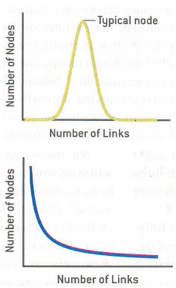


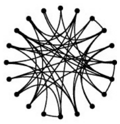
$K_7=5$

How do we find the degree of each node from the adjacency matrix?

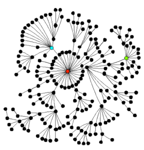


Degree Distribution





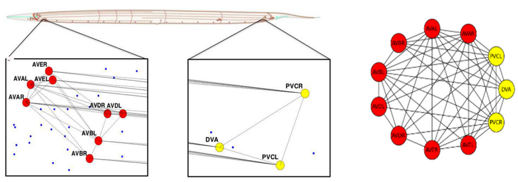
Random Network



Scale-free Network

Barabasi & Albert (1999) Science

Hubs in the C elegans nervous system



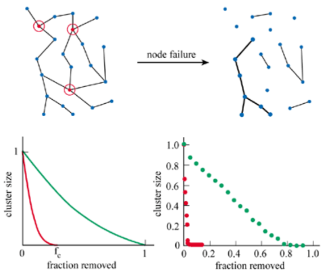
- Most nodes have few connections, a few nodes have >80 links
- These hubs are in the head and tail and are heavily linked to each other, forming a rich club
- This group of neurons controls backward and forward movement (command interneurons)
- During evolution, the worm only begins to twitch once the last of these elite neurons is born

Towlson et al., J. Neurosci. 33, 6380 (2013).

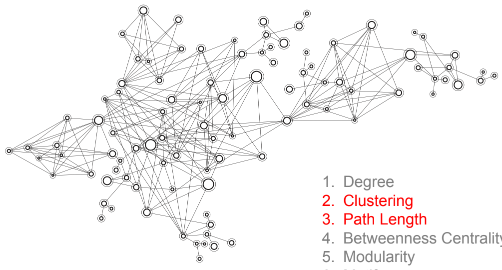
Hubs and robustness to error or attack

As many as 80% of randomly selected internet routers can fail and the remaining ones will still form a compact cluster where a path exists between any two nodes.

Preferential removal of hubs, however, quickly disrupts the network.



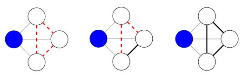
3. Network Measures (and why they matter)



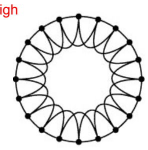
1. Degree
2. Clustering
3. Path Length
4. Betweenness Centrality
5. Modularity
6. Motifs

Clustering Coefficient of a Node

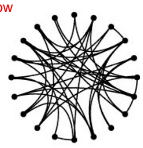
$$C_i = \frac{\textit{Links between 'friends'}}{\textit{Total possible number}}$$



High C

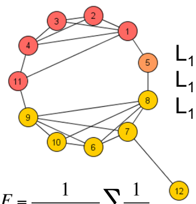


Low C



Path Length (or Efficiency)

$L_{i,j}$ = *Shortest path between nodes i and j*



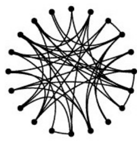
$L_{1,2} = 1$

$L_{1,4} = 1$

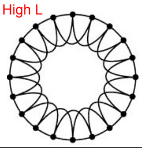
$L_{1,12} = 4$

$$E = \frac{1}{N(N-1)} \sum_{i \neq j} \frac{1}{L_{i,j}}$$

Low L

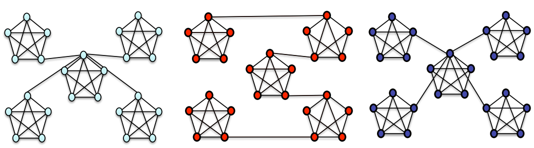


High L



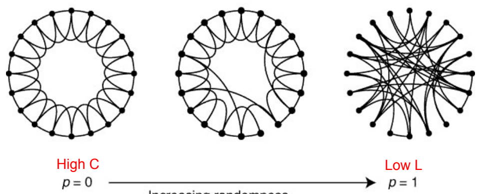
Comparing Networks

benchmark	3/54 edges rewired	4/54 edges rewired



L=2.83 C=0.90	L=4.00 C=0.87	L=2.83 C=0.90

Small- World Networks: Why it is Fun to Surf the Web



Watts & Strogatz (1998) Nature