Graph Theory and Biological Networks

Master-Module Biological Networks
July 21, 2016

Emidio Capriotti
http://biofold.org/

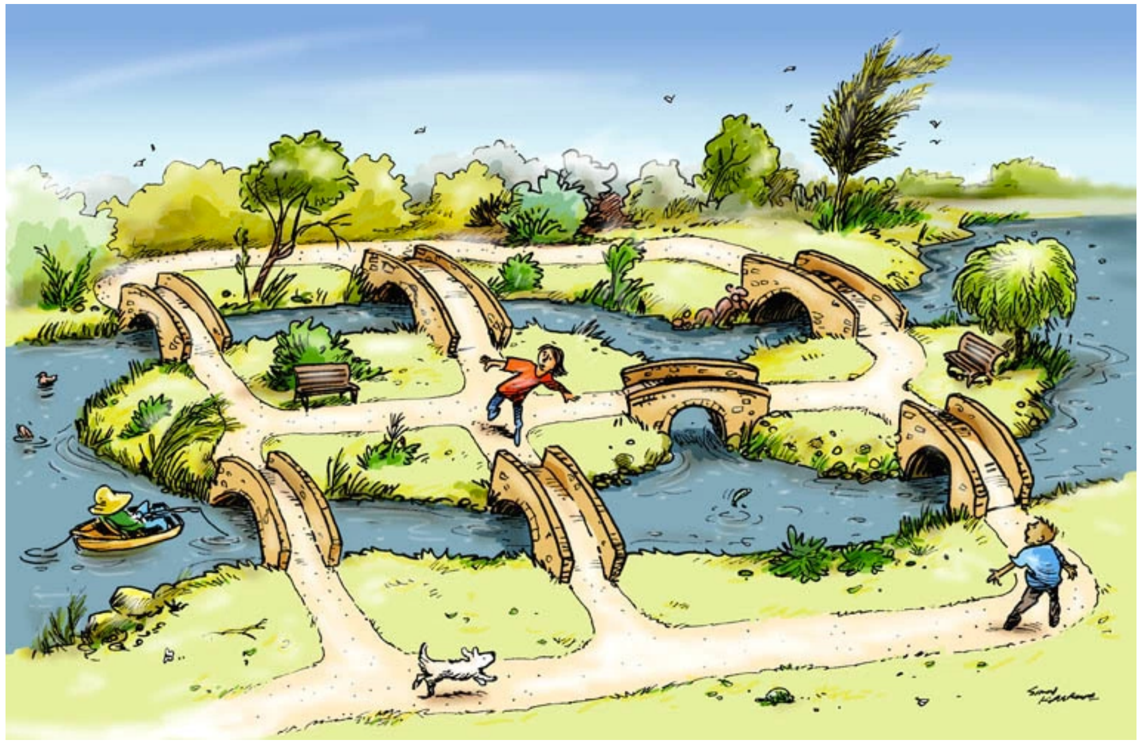
Biomolecules Folding and Disease

Institute for Mathematical Modeling
of Biological Systems
Department of Biology

HEINRICH HEINE
UNIVERSITÄT DÜSSELDORF

Historical Perspective

With the Seven Bridges of Königsberg problem, Euler in 1737 laid the foundations of the graph theory.

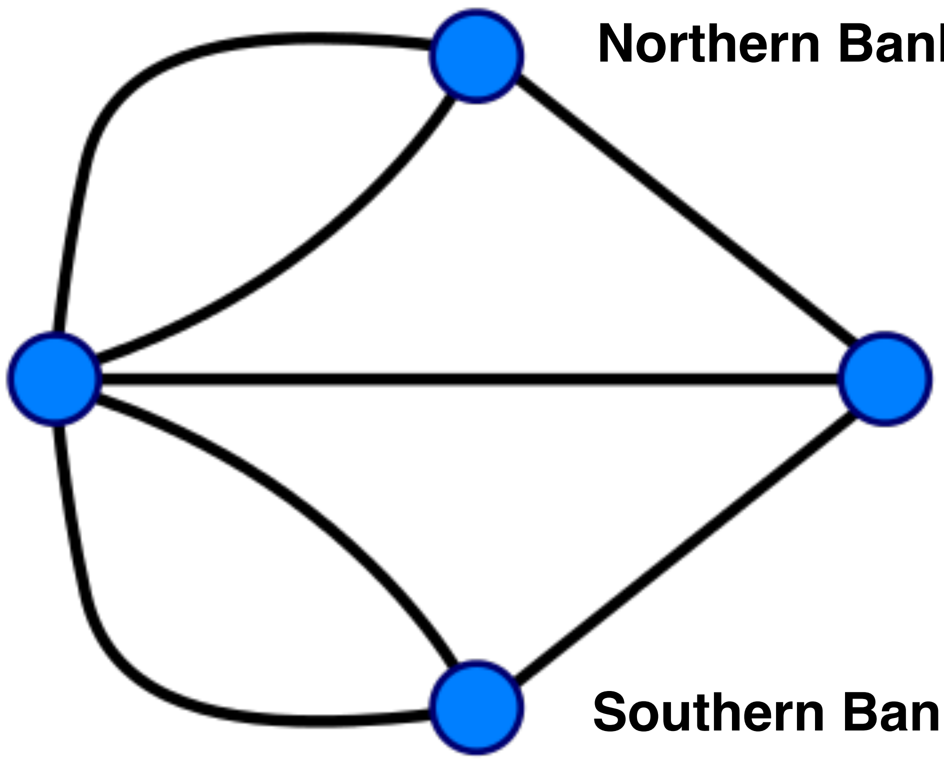


Simon Kneebone – simonkneebone.com

- Find path (Eulerian Path) that traverses all the Pregel's bridges.
- Find walk (Eulerian Circuit) that traverses all the Pregel's bridges and has the same starting and ending point.

Solution

Describe the problem as a graph where the nodes represent the 4 locations and the edges correspond to the bridges



Eulerian path exists only if none or 2 nodes are connected by an odd number of bridges.

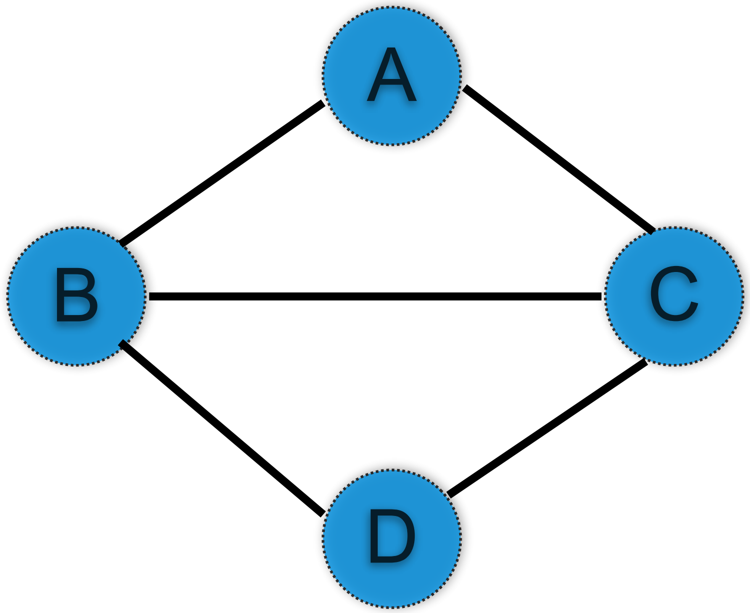
Eulerian circuit exists only if no nodes are connected by an odd number of bridges.

Graph Definition

A graph is a pair $G=(V,E)$ consisting of two sets:

- V is a set of elements called Nodes or Vertices.
- E is a set of pairs (v_i,v_j) where $v_i \in V$ and $v_j \in V$.

The pairs E are links between two nodes and are called Edges

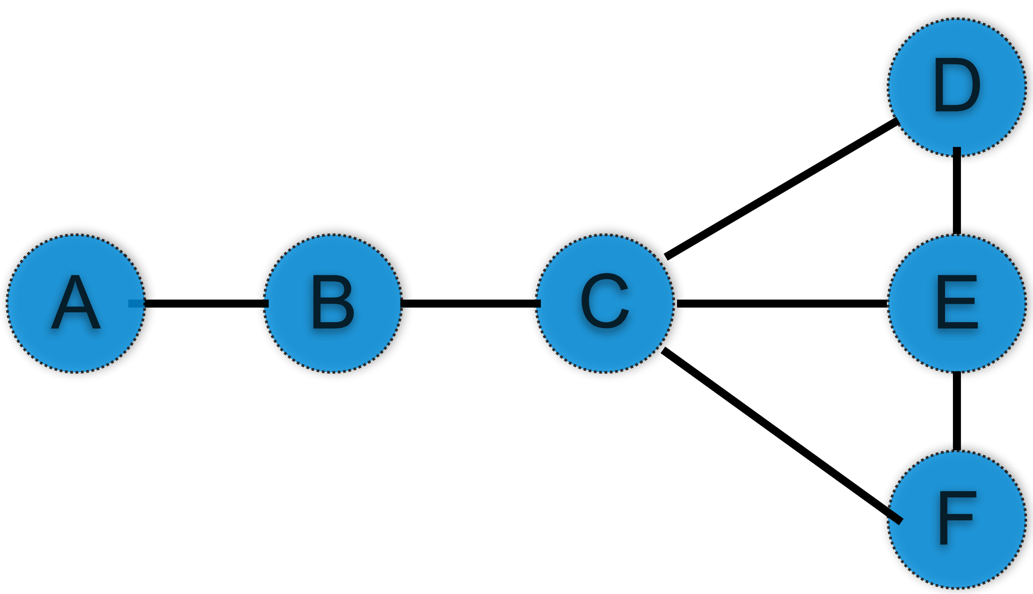


V = {A; B; C; D}

E = {(A,B); (A,C); (B,C); (B,D); (C,D)}

Undirected Graph

Undirected graph is a network where the relationship between nodes are symmetric.

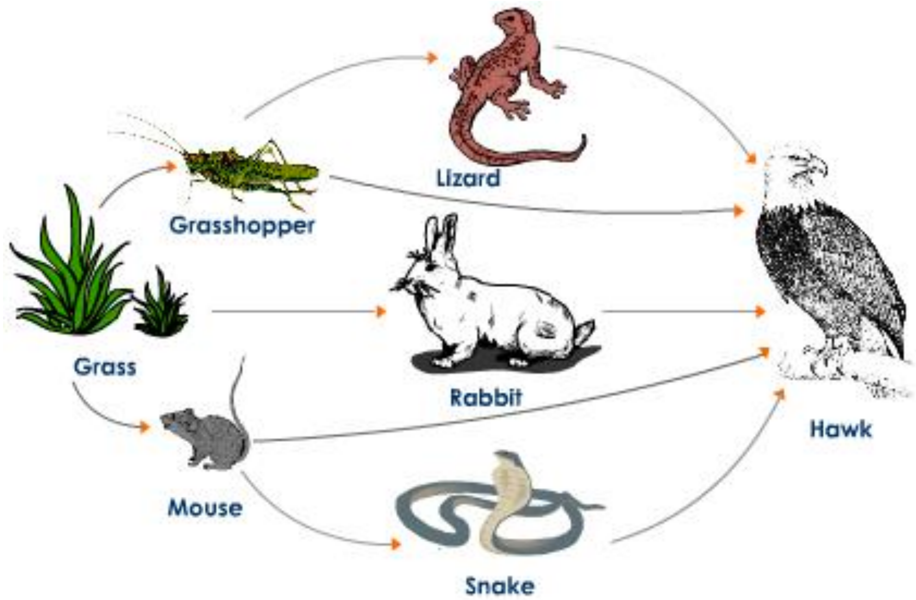


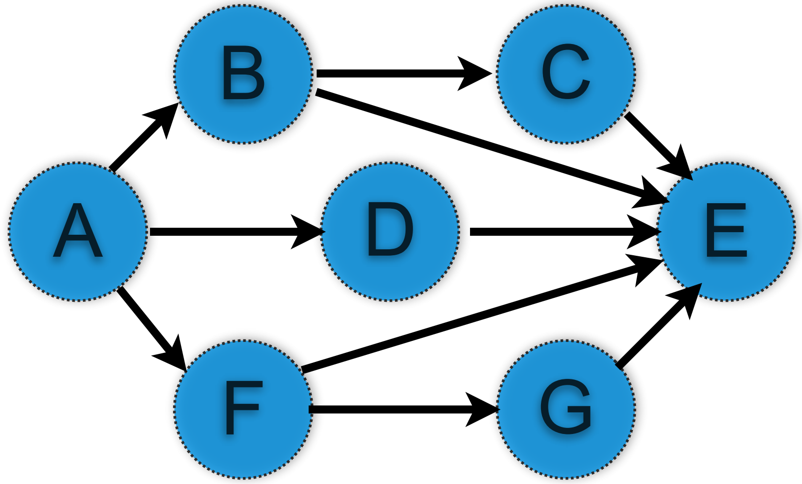
V = {Group of People}

E = {Pairs of Friends}

Directed Graph

Directed graph is a network where the relationship between nodes are asymmetric. In this case the edges are directed lines.



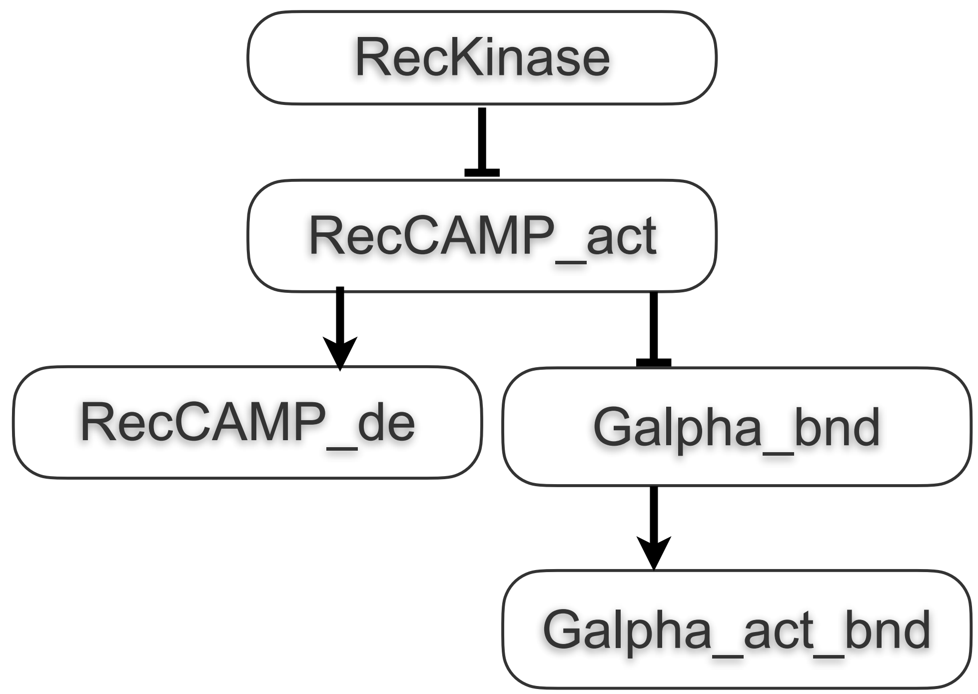


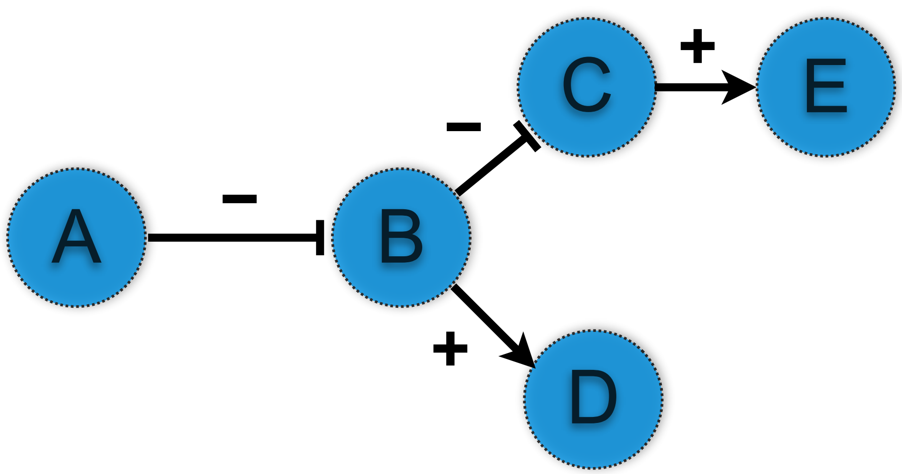
V = {Group of Animals}

E = {Pray/Predator Relationships}

Signed Directed Graph

Signed Directed graph is a network where the relationship between nodes are asymmetric and have positive or negative associated signs





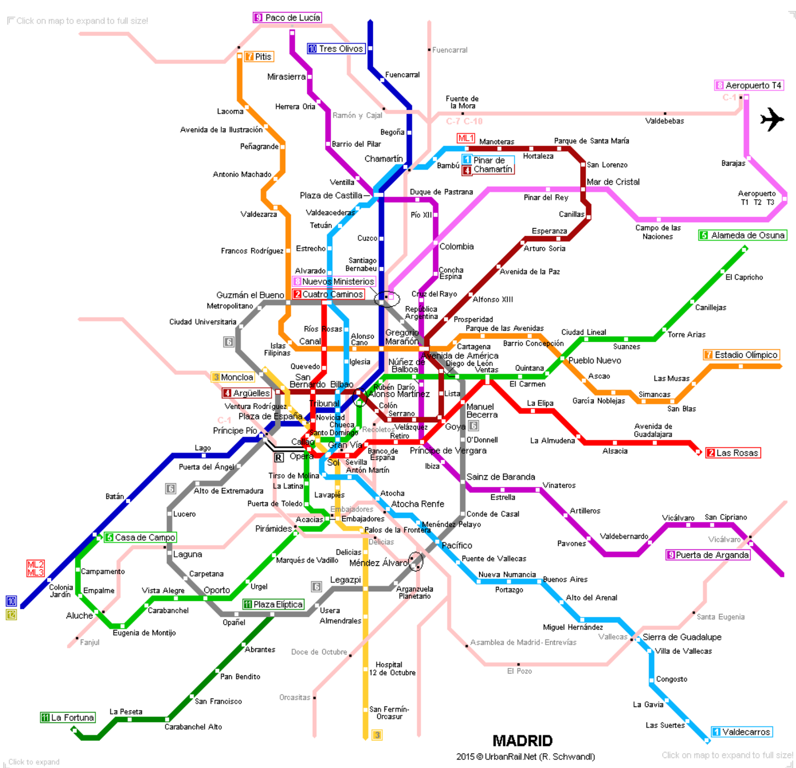
V = {Group of Genes}

E = {Activation/Inhibition Relationships}

Graph and Networks

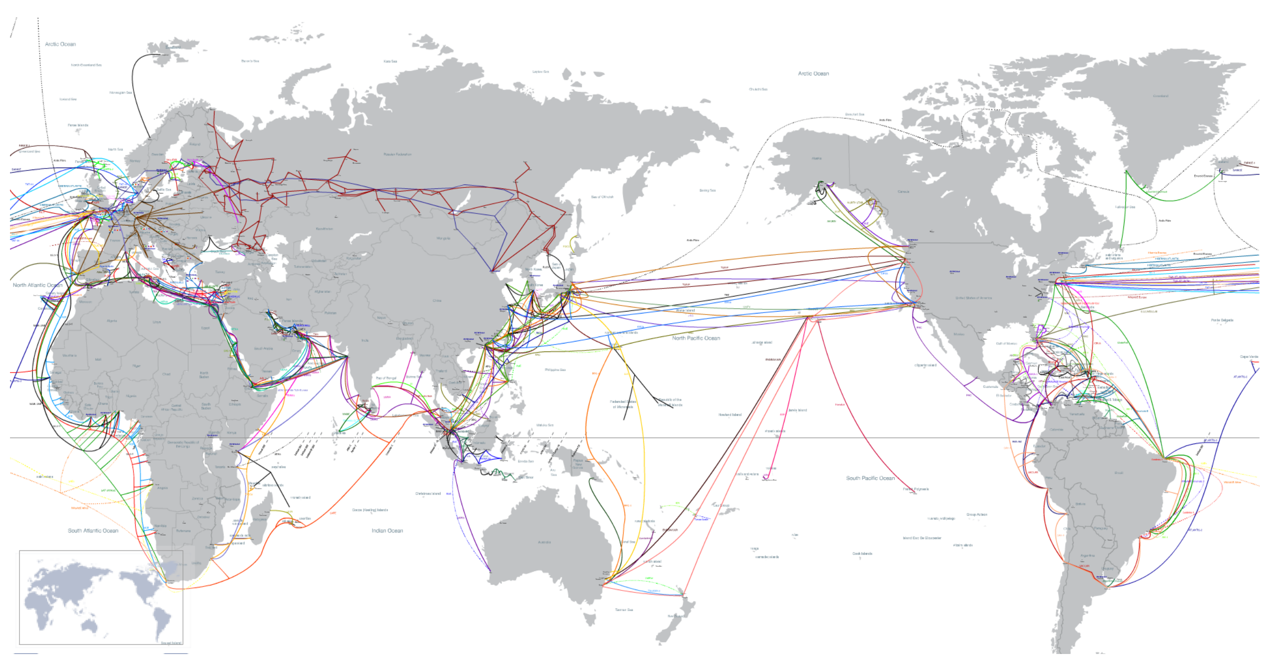
Graphs can be used to represent any observed of network.

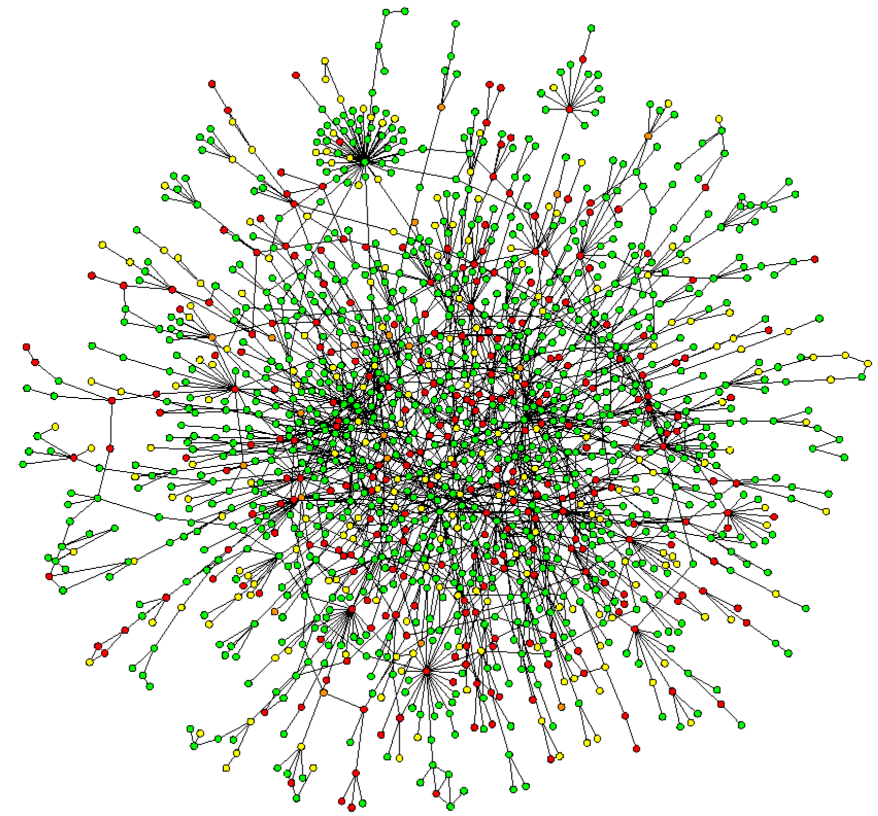
Networks in nature tend to be highly complex



Madrid Metro

Internet connections

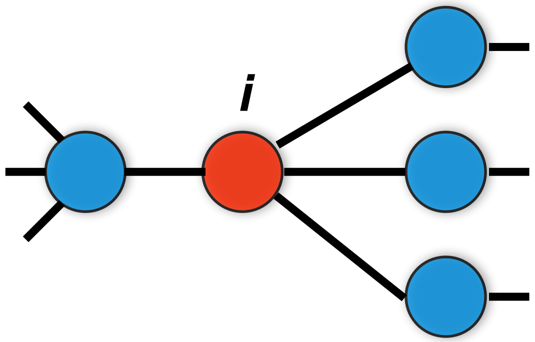




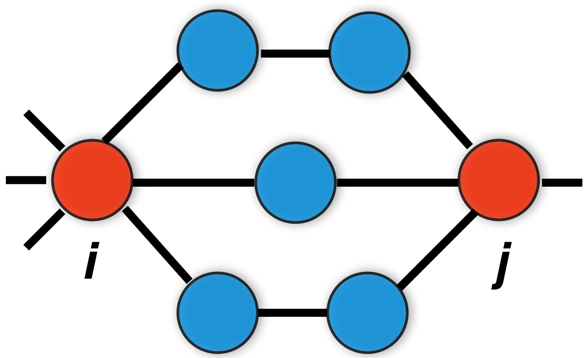
Yeast interactome

Network properties (I)

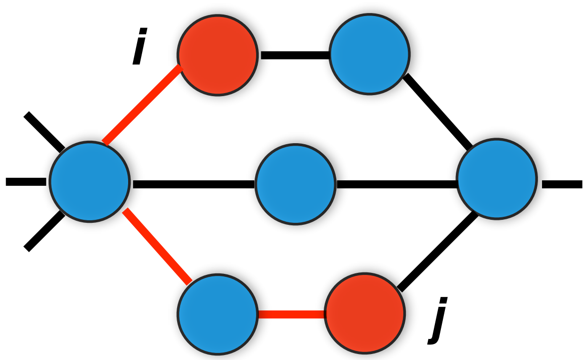
The topology of the network defines its properties. The level of connectivity among the nodes depends on the number of edges.



Degree k_i = number of links connected to node i



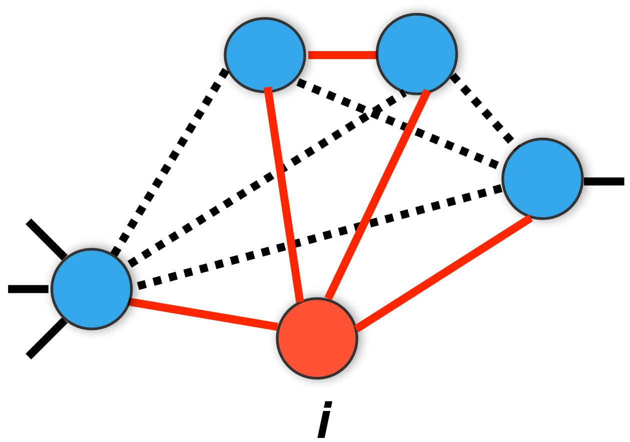
Distance d_{ij} = shortest path between nodes i and j



Diameter D = longest path between all pairs of nodes

Network properties (II)

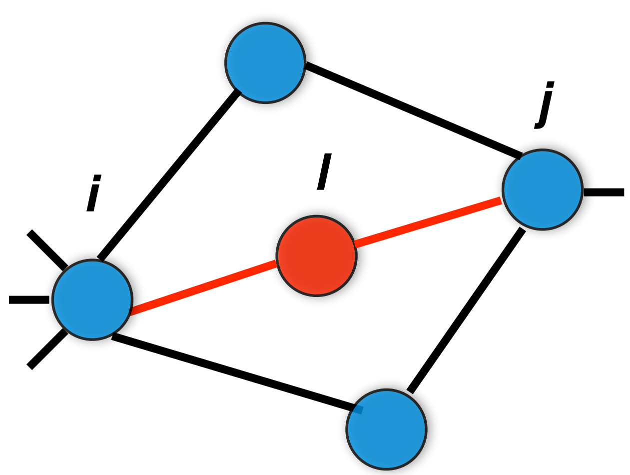
The topology of the network defines its properties. The level of connectivity among the nodes depends on the number of edges.



Transitivity or Clustering Coefficient

$$c_i = \frac{2e_i}{k_i(k_i - 1)}$$

k_i = number of nodes connected to *i*
e_i = number of edges between the k_i nodes



Betweenness

$$g_l = \sum_{i \neq l \neq j} \frac{\sigma_{ij}(l)}{\sigma_{ij}}$$

σ_{in} = number of shortest path between *i and j*
σ_{ij}(l) = number of shortest path passing through node *l*

Types of Network

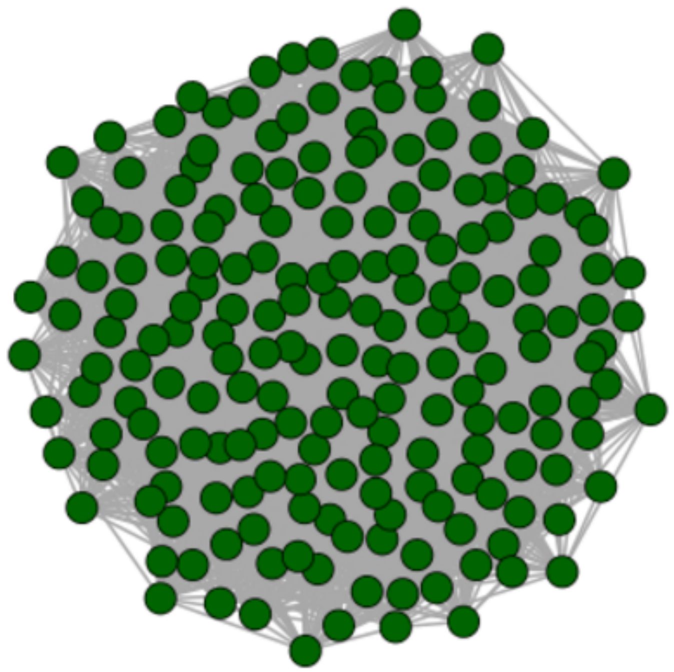
The topology of the network depends on the distribution of the degree for all the nodes.

We can define three types of network:

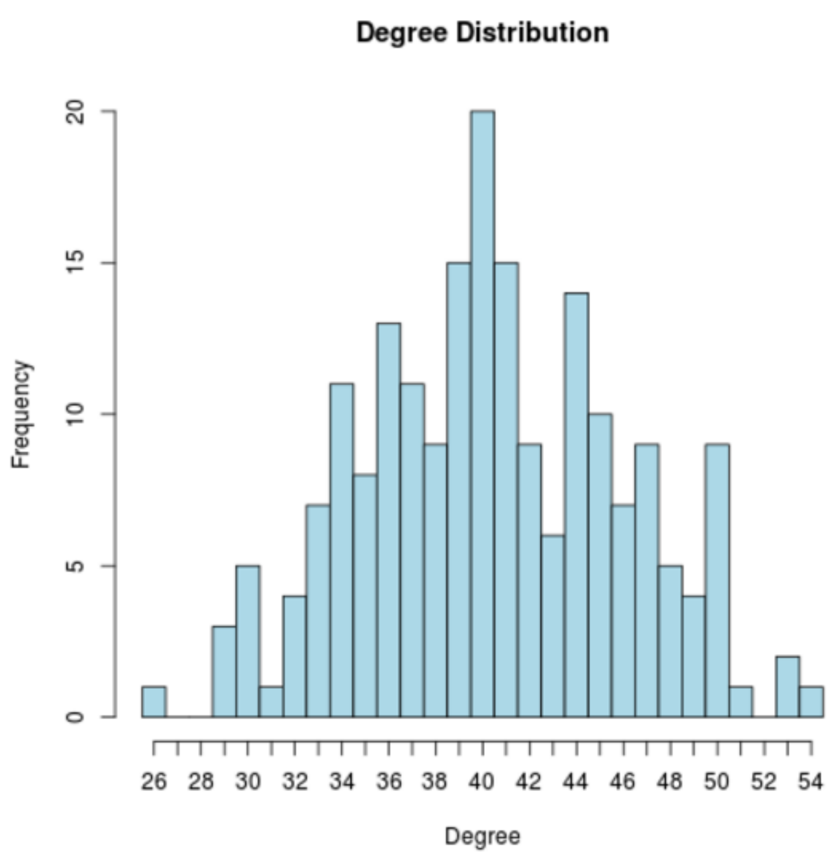
- Random network: generated by a constant probability of having a edge between two nodes.
- Small-world network: when the degrees follow a Poisson distribution
- Scale-Free network: the degrees follow a Power Law distribution

Random Network

Can be generated by Erdős–Rényi model which assume a constant probability of generating edges between nodes.



- High node degree ⇒ low average path length
- Degree distribution tends to be a Gaussian
- High Transitivity
- Small Betweenness



Degree = 40.3
Transitivity = 0.2
Betweenness = 79.3

Small-World Network

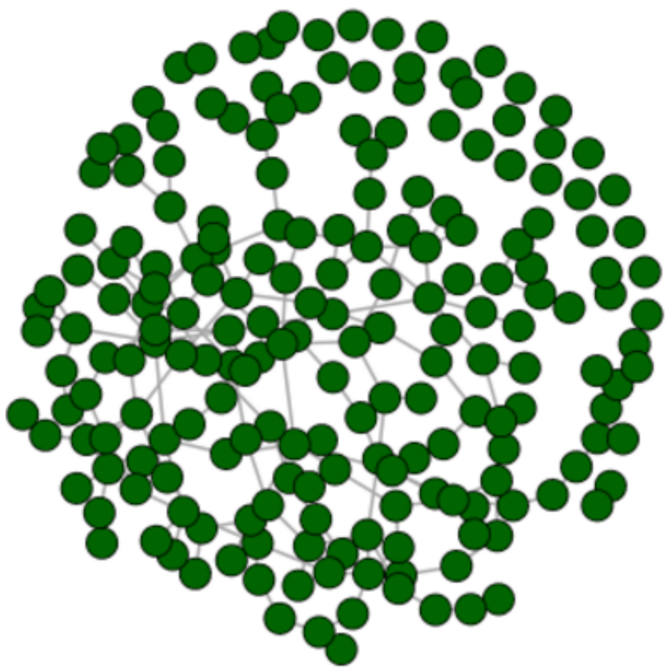
Generated by a Watts–Strogatz model.

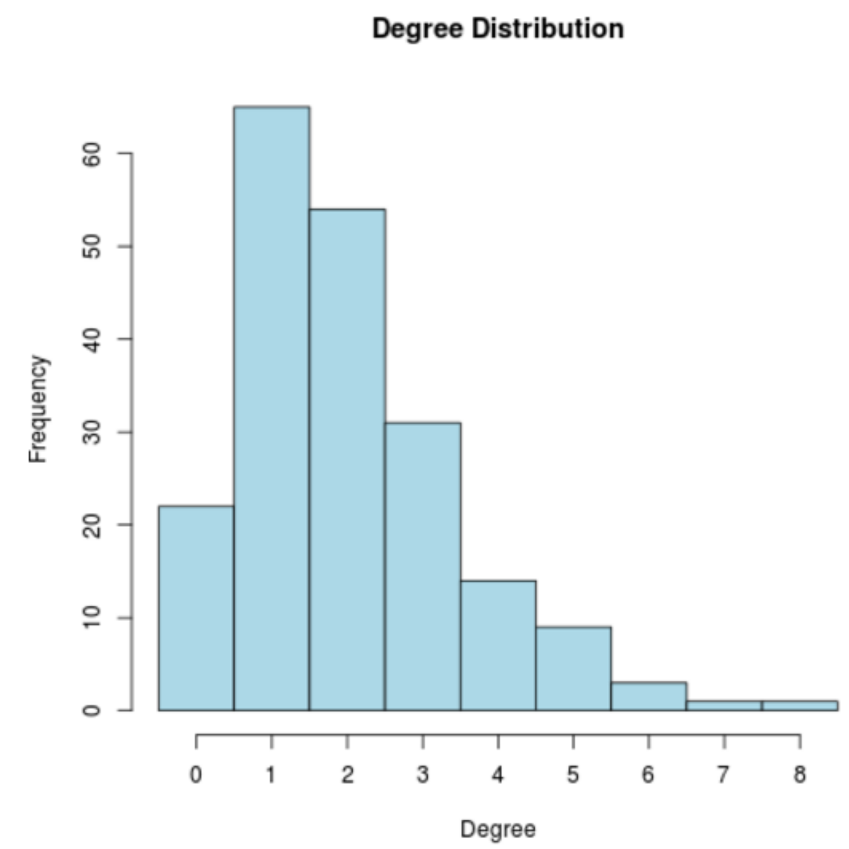
- Low node degree ⇒ “Six degrees of separation”
- Degree follow a Poisson distribution
- Low Transitivity than random
- Higher betweenness than random

$$p(k) = \frac{\lambda^k e^{-\lambda}}{k!}$$

λ = the average value of the distribution
k = number of observed events

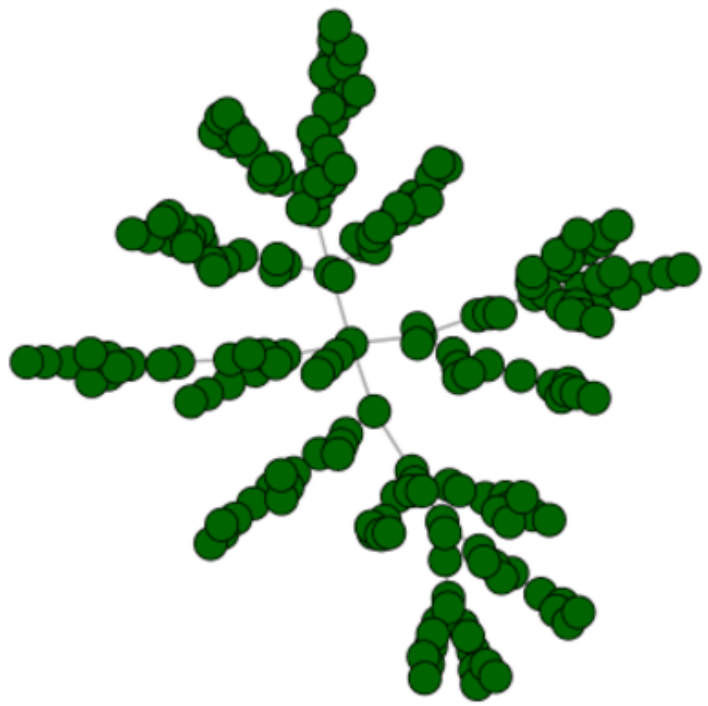
Degree = 2
Transitivity = 0.01
Betweenness = 394.9





Scale-Free Network

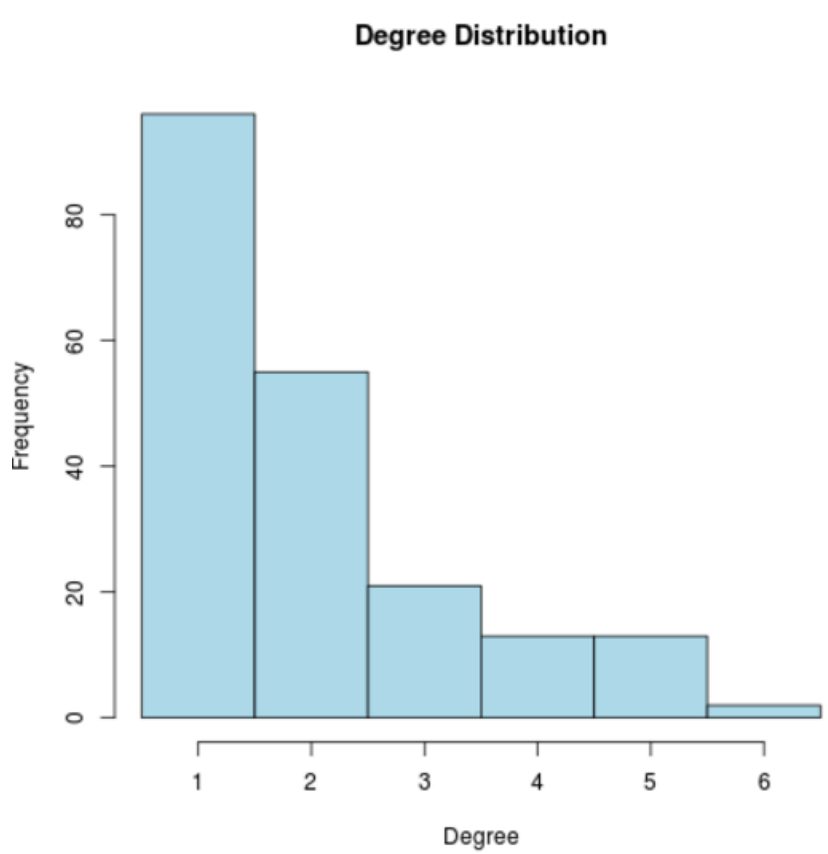
Generated by the Barabasi-Albert model.



- Smallest degree
- Degree follow a Power Law distribution
- Lowest Transitivity
- Highest Betweenness

$$p(k) = Ax^{-k}$$

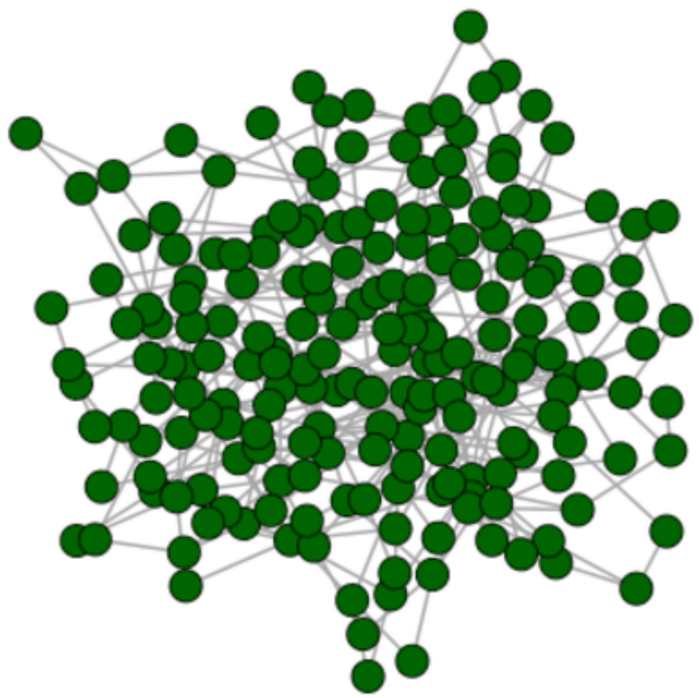
x = is a constant
k = number of observed events



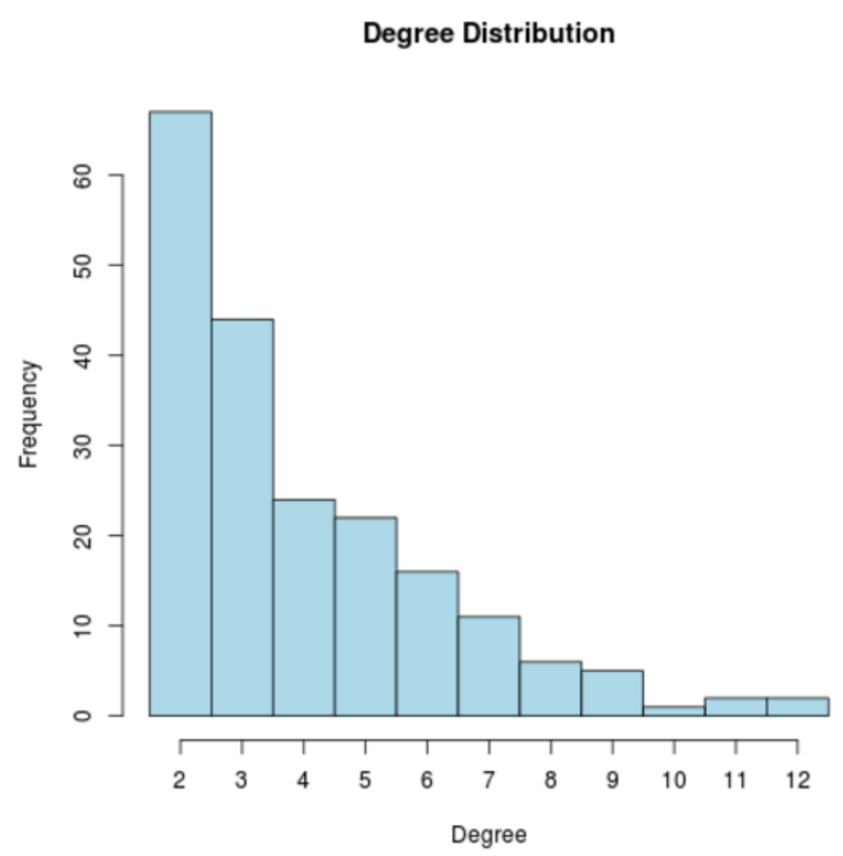
Degree = 2
Transitivity = 0
Betweenness = 753.4

Biological Network

Similar to Small-World and Scale-Free networks



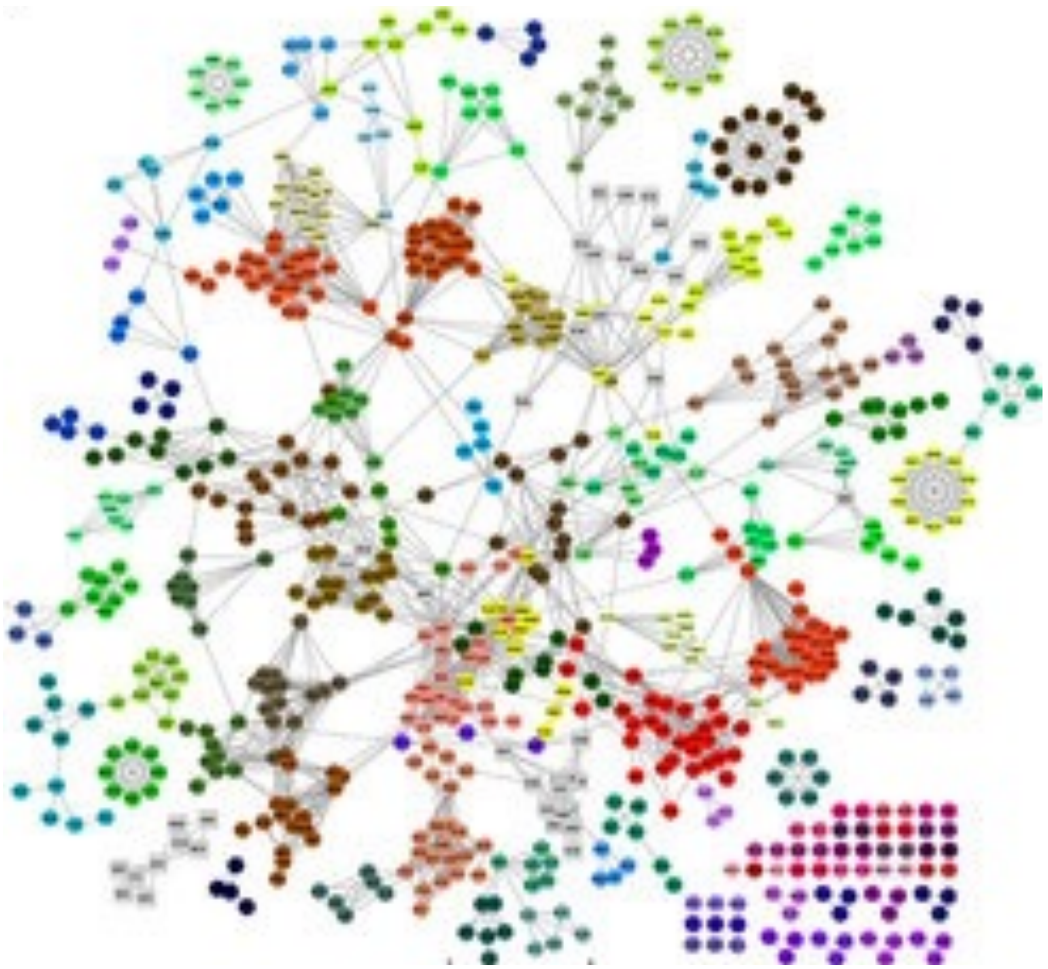
- Small degree
- Average path length proportional to *ln(ln(#nodes))*
- Transitivity high than Small-World and Scale Free
- Betweenness lower than Small-World and Scale Free



Degree = 4.0
Transitivity = 0.04
Betweenness = 290.4

Community or Cluster

One of the main feature of the biological network is the presence of communities or clusters.

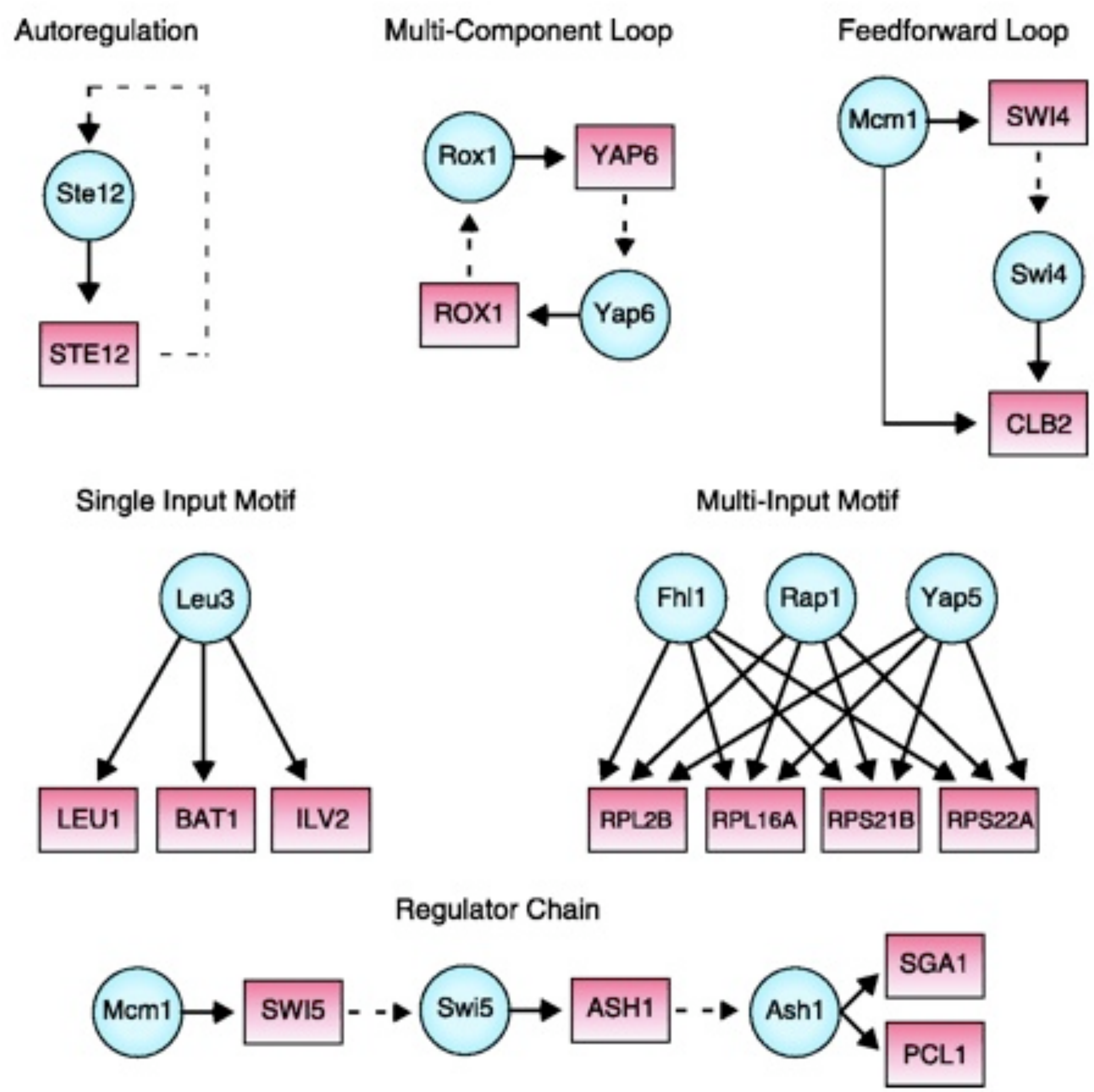


Gaiter, Scientific Reports 2015

Cluster are important to detect similarity between nodes (genes, diseases, etc) in the same cluster.

Network Motifs

Network analysis is important for detecting network motifs, which are recurrent and statistically significant sub-graphs or patterns.



http://compbio.pbworks.com/

Network Robustness

Robustness, the ability to withstand failures and perturbations. It is a critical attribute of many complex systems including biological networks.

Robustness is tested removing nodes and checking if connections between the remaining nodes are conserved. This is possible because may exist alternative paths between two distinct nodes.

Biological networks persists despite the environmental noise, mutations etc.

Telecommunication networks resit to the attach of hackers and hardware failure

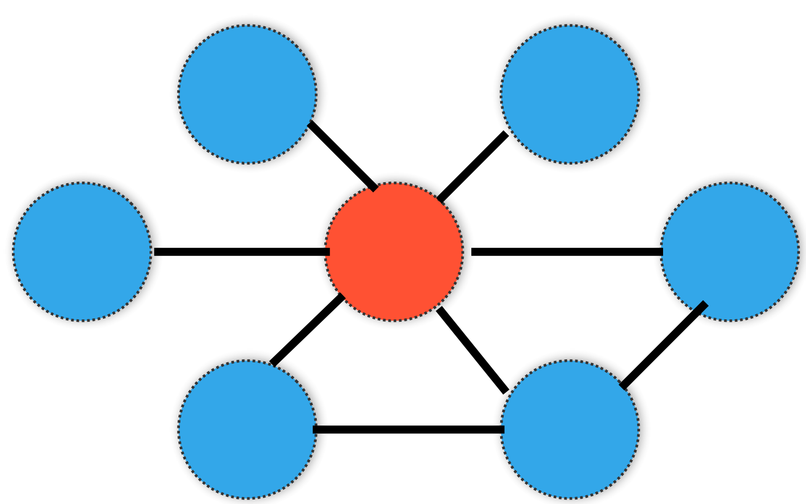
Network Attack

For random networks the effect of removing a single node is on average the same.

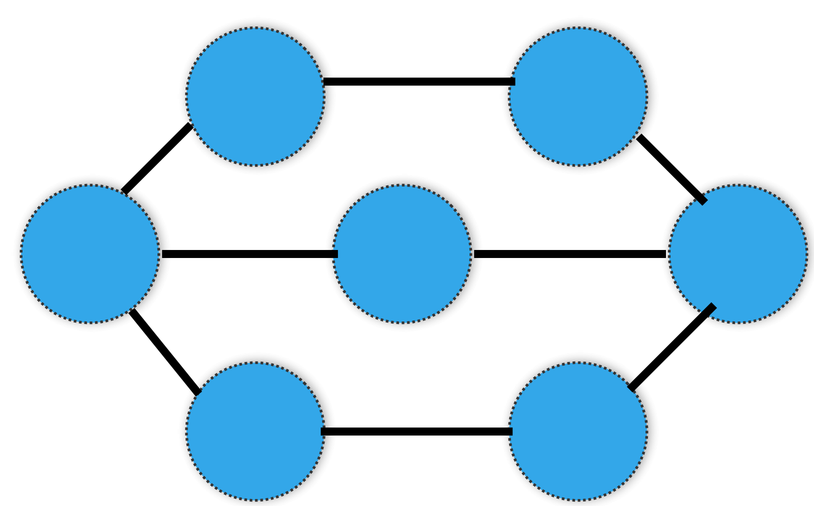
Biological networks are characterized by a small fraction of nodes with high degree (hubs)

An attack that aims to an hub has strong effect on the connectivity of the network.

In normal situation we assume that attacks are random. Thus, on average, an attach should have smaller effect on Biological Network.



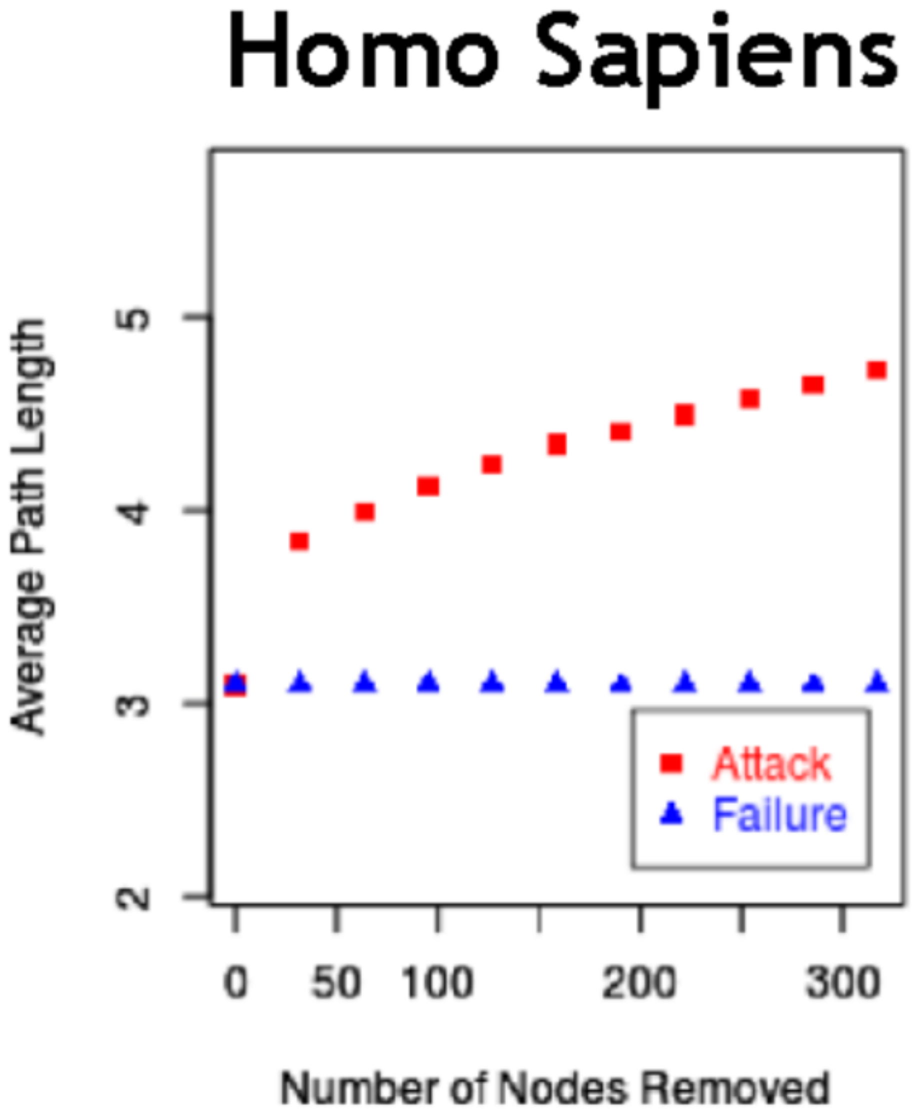
Biological network

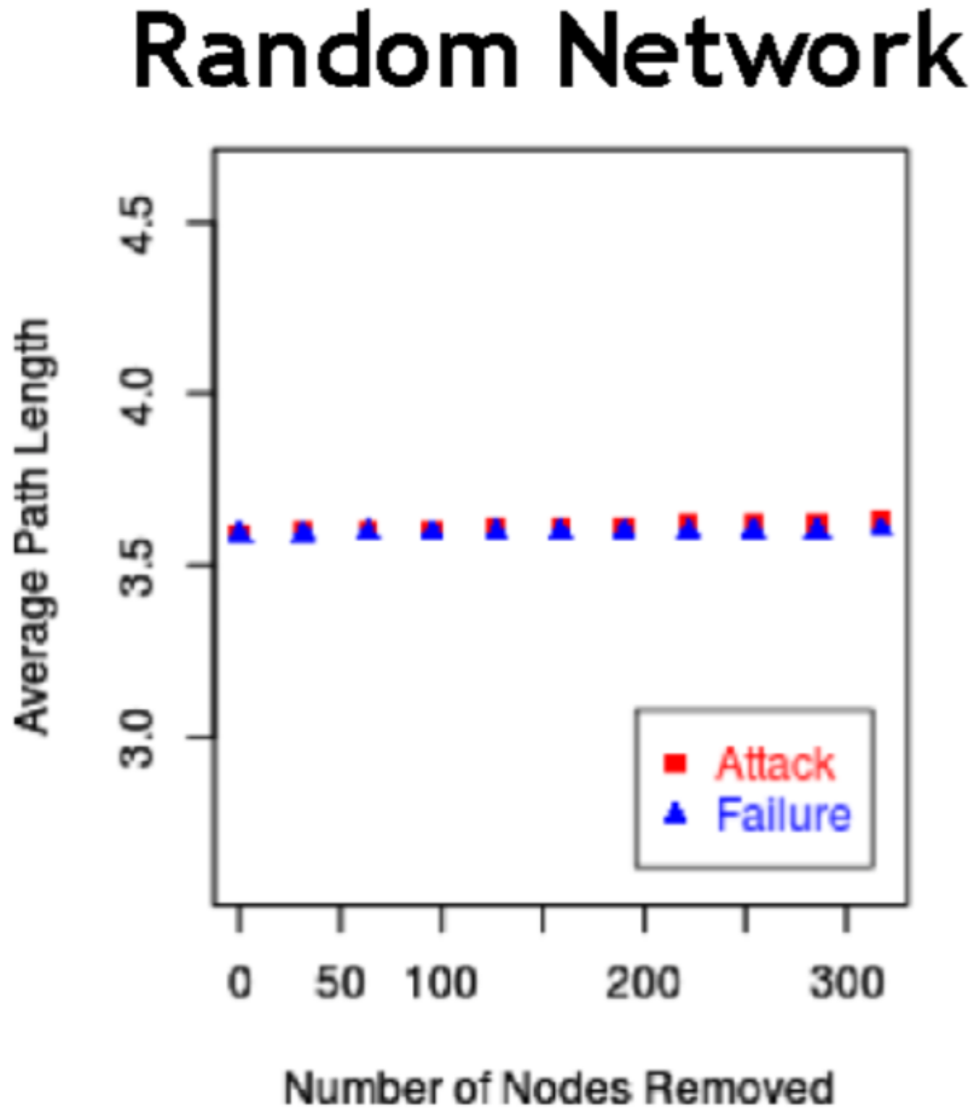


Random network

Multiple attacks

Removing small number of nodes has less impact on biological network with respect to random network. Stronger effect is shown when the number of affected nodes increases.





Exercise

Download the Yeast interactome from BIOGRID

Download the bundle file generate a comma separated value with the interactions

Import the network in R using the following command for edges

data <- read.csv(filename, header=F, as.is=T)

Transform your dat in a graph object using the following command:

net <- graph_from_data_frame(d=links,, directed=F)

Calculate the degree of all the nodes:

deg <- degree(net, mode=“all”)

Plot the distribution.

More command to test can be found at http://kateto.net/networks-r-igraph