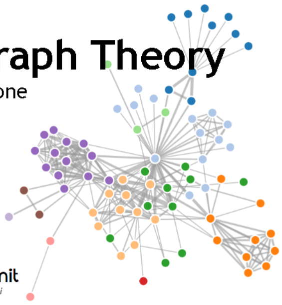
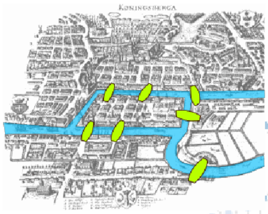
Introduction to Graph Theory

Alberto Calderone



Seven Bridges of Königsberg

Leonhar Euler



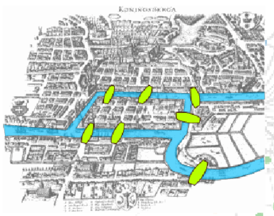
The problem was to devise a walk through the city that would cross each bridge once and only once. Under the following conditions:

1) The islands could only be reached by the bridges
2) Every bridge once accessed must be crossed to its other end
3) The starting and ending points of the walk need not be the same

Seven Bridges of Königsberg

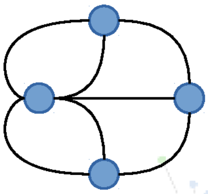
- The year 1736 can be considered the beginning of graph theory when EULER considered the **Seven Bridges of Königsberg** problem.
- 200 years later the first book on graph theory was written:
  - "Theorie der endlichen und unendlichen Graphen", by Kőnig In 1936.
- Since then, graph theory has developed into an extensive branch of mathematics, which has been applied to many problems in science and other areas.

Leonhar Euler



Seven Bridges of Königsberg

Leonhar Euler



The problem was to devise a walk through the city that would cross each bridge once and only once. Under the following conditions:

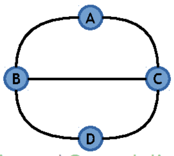
1) The islands could only be reached by the bridges
2) Every bridge once accessed must be crossed to its other end
3) The starting and ending points of the walk need not be the same

Definition of Graph (Network)

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 $v_j \in V$.
These pairs are links between two elements called **Edges**

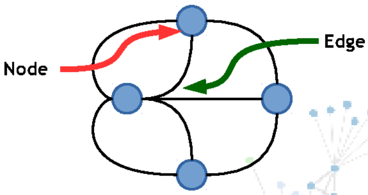


V = {A; B; C; D}

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

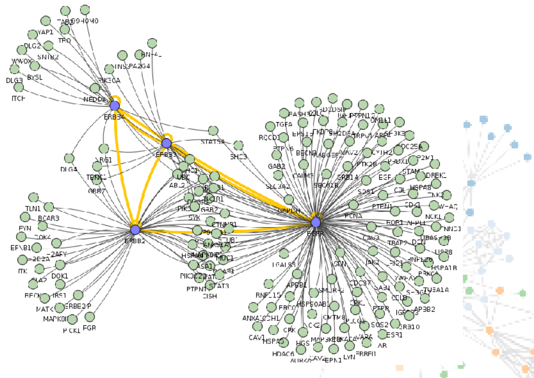
Seven Bridges of Königsberg

Leonhar Euler



Solution: a graph is walkable without passing twice on the same edge if and only if all its nodes have an even number of connections, or if two of them have an odd number of connections; to travel a graph of this second type, it is necessary to start from one of the two nodes with an odd number of connections ending the walk on the other one.

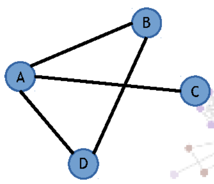
Protein Interactions Networks



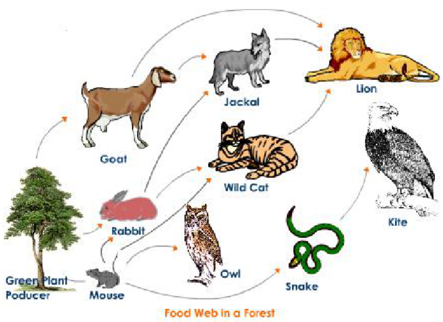
Protein Interactions Networks

Simple Graph

EXAMPLE 1 (**Friends**) :
V = { People living in Italy }
E = { Pairs of friends }



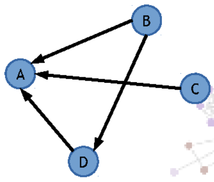
Food Chain Networks



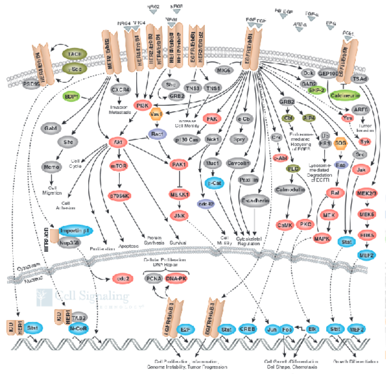
Food Chain Networks

Directed Graph

EXAMPLE 2 (**Mail**) :
V = { People living in Italy }
E = { (x,y) : x sent a mail to y }



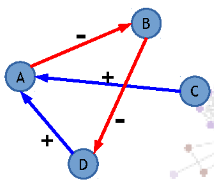
Signalling Networks



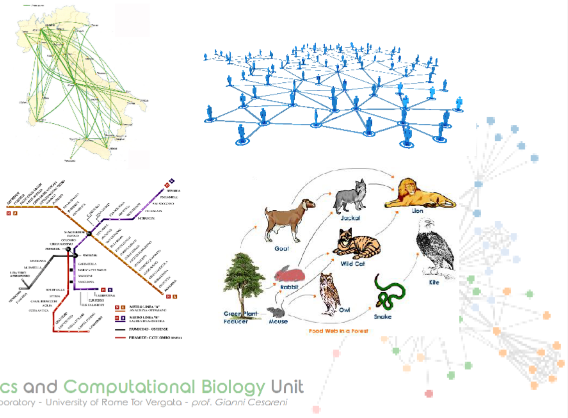
Signalling Networks

Signed Directed Graph

EXAMPLE 3 (**Biological Signalling Networks**) :
V = { Proteins in a cell }
E = { Protein A activates/inactivates protein B }

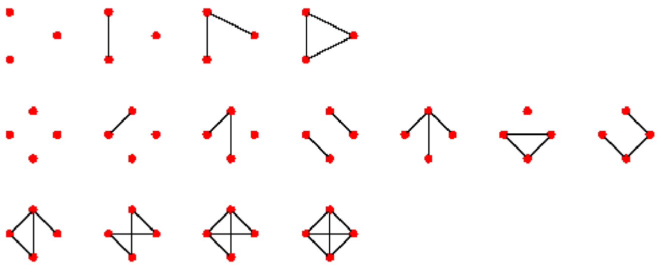


Networks are Everywhere...



Some Important Results

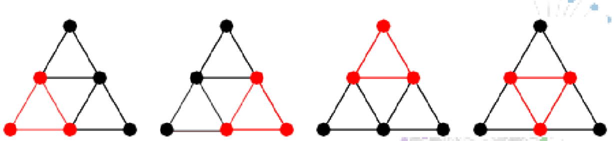
Enumeration: counting graphs meeting specified conditions.



Bioinformatics and Computational Biology Unit
Molecular Genetics Laboratory - University of Rome Tor Vergata - prof. Gianni Cesareni

Some Important Results

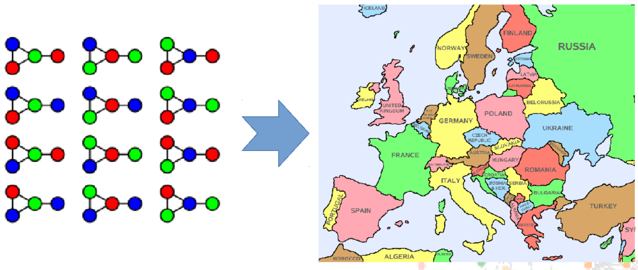
Subgraphs, induced subgraphs: finding the largest complete subgraph, finding the largest edgeless induced subgraph.



Bioinformatics and Computational Biology Unit
Molecular Genetics Laboratory - University of Rome Tor Vergata - prof. Gianni Cesareni

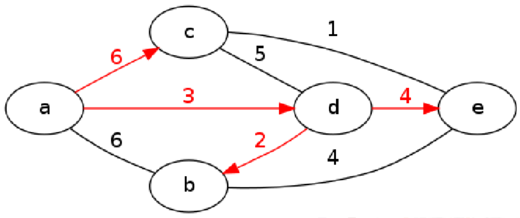
Some Important Results

Graph colouring: it is an assignment of labels traditionally called "colours" to elements of a graph subject to certain constraints.



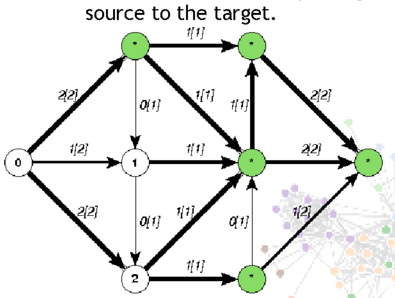
Some Important Results

Route problems: weighted or unweighted path discovery, spanning trees.



Some Important Results

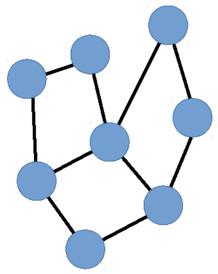
Network flow: maximum amount of flow passing from the source to the target.

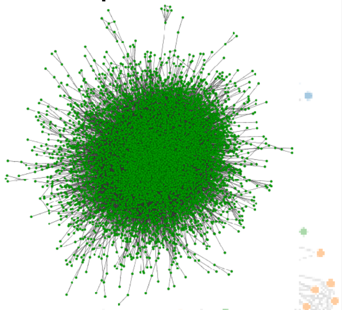


Complex Networks

Network

Complex Network

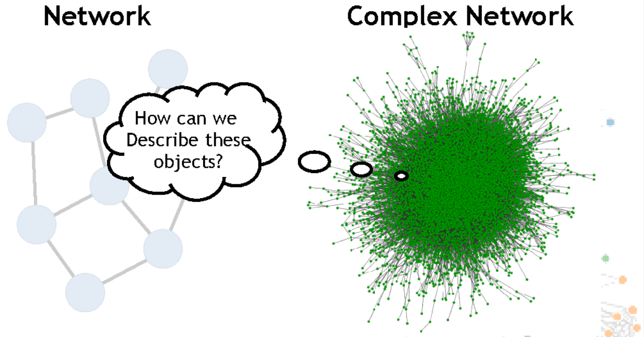




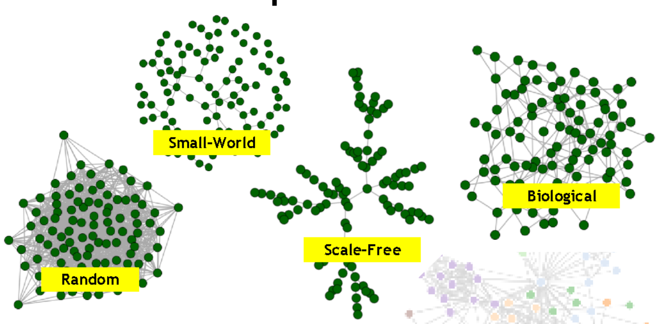
Complex Networks

Network

Complex Network



Graph Models



http://160.80.34.9/elixir2015/

Activity: Describing Graphs

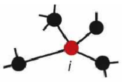
Nodes and edges Characteristics

- Definition of “Degree”
- Definition of “Betweenness” (*)
- Definition of “Transitivity” (*)

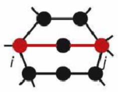
(*) Not easy

http://160.80.34.9/elixir2015/

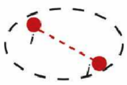
Definitions

A 

Degree — k_i = number of links connected to node i

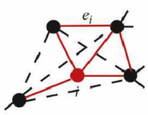
B 

Distance — d_{ij} = shortest path length between node i and j

C 

Diameter — $D = \max\{d_{ij} \mid i, j \in N\}$ N : all nodes in the network

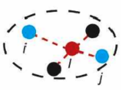
Definitions

D 

Transitivity or *Clustering Coefficient*

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

e_i : number of existing links (labeled in red) among the k_i nodes that connect to node i

E 

Betweenness

Defined for Nodes and Edges

$$b_l = \sum_{ij} p_{ij}(l) / p_{ij}$$

p_{ij} : number of shortest paths between i and j

$p_{ij}(l)$: number of shortest paths between i and j going through node l

Activity: Describing Graphs

Nodes and edges Characteristics

- Definition of “Degree”
- Definition of “Betweenness” (*)
- Definition of “Transitivity” (*)

Global characteristics

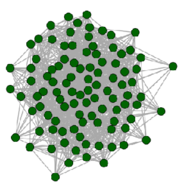
- Compare different graphs
- Find characteristics

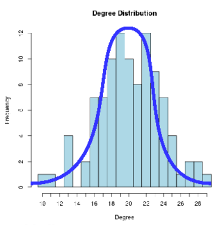
(*) Not easy

http://160.80.34.9/elixir2015/

Random Networks Model
(Erdős-Rényi Model)

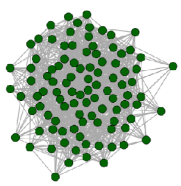
1) Starting from a set of n vertices
2) Give two random vertices there is a probability P that they are linked together

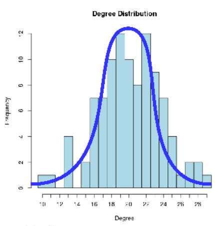




Random Networks Characteristics
(Erdős-Rényi Model)

1) High node degree => Short Average Path Length
2) Degree distribution follows Normal Distribution
3) Small Betweenness
4) High Transitivity

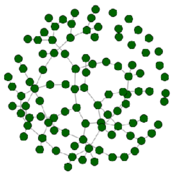


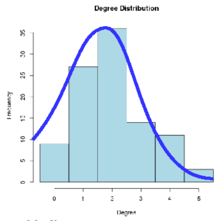


Small-World Networks Model

(Watts and Strogatz Model)

1) Generate a lattice
2) Nodes are initially linked to k closest neighbours
3) Apply a rewiring probability

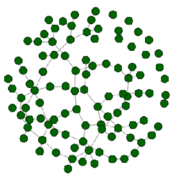


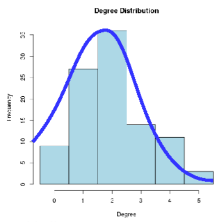


Small-World Networks Characteristics

(Watts and Strogatz Model)

1) Small node degree => “Six-Degrees of Separation”
2) Degree distribution follows Poisson Distribution
3) Betweenness smaller than Scale-Free
4) Transitivity: smaller than Random, Higher than Scale-Free





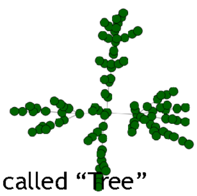
Scale-Free Networks

(Barabási-Albert Model)

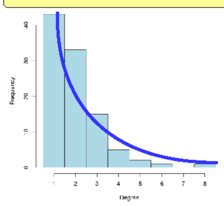
1) Smallest node degree
2) Degree distribution follow Power-Law
3) Highest Betweenness
4) Transitivity: smallest

Preferential Attachment: The rich gets richer

Scale-free refers to any function f(x) That remains unchanged to within a multiplicative factor under a rescaling of x. This means power-law since these are the only solutions to f(ax)=bf(x).



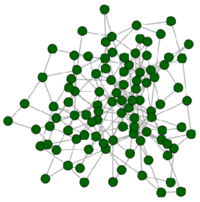
This is called “Tree”

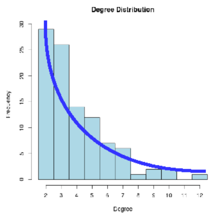


Natural Networks

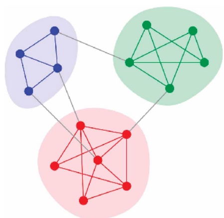
(Biological Networks)

1) Small-World & Scale-Free
2) Small Node Degree, Short Path Length, Power-Law
3) Betweenness between Small-World and Scale-Free
4) Transitivity higher than Small-World and Scale Free
5) Average path length *ln(ln(n))*





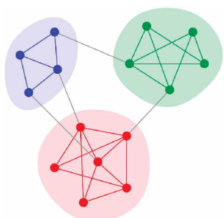
Communities (or Clusters)



The presence of communities in a graph is one of the most important features.

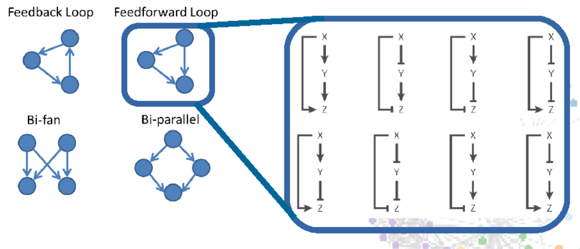
- For biologists to detect proteins with the same function
- For physicians to detect related diseases
- Communities are important for Amazon, to run their businesses

Communities Discovery Algorithms



- Modularity - Edge-Betweenness Algorithm
  modularity is the fraction of edges that fall within a given group minus the expected fraction if edges were distributed at random (*it is unable to detect small communities*)
  Biological networks, including animal brains, exhibit a high degree of modularity
- Other Algorithms - e.g. Information Theory

Network Motifs



Motifs can be seen as the smallest version of communities

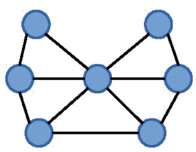
Network Robustness

Complex systems tend to have a surprising degree of tolerance to errors (Robustness):

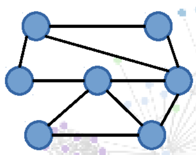
- Biological networks persist despite environmental noise, mutations, attacks ...
- Communication networks resist to malfunctions, hackers, nodes failure

Network Robustness

Preferential Attachment

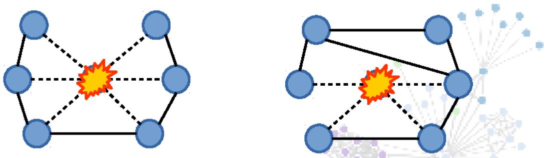


Random



Network Robustness

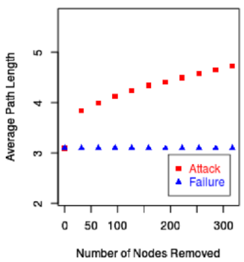
Preferential Attachment | Random



Bioinformatics and Computational Biology Unit
Molecular Genetics Laboratory - University of Rome Tor Vergata - prof. Gianni Cesareni

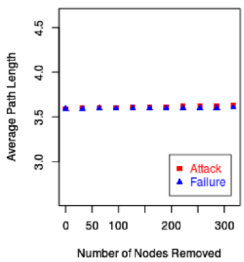
Network Robustness

Homo Sapiens



Scale-Free
Preferential Attachment

Random Network



Bioinformatics and Computational Biology Unit
Molecular Genetics Laboratory - University of Rome Tor Vergata - prof. Gianni Cesareni

Conclusions

- What is a graph/network
- Type of graphs
  - Simple, Directed, Signed
- Graph Algorithms Overview
- Graph Models
  - Random, Scale-Free, Small-World, Preferential attachment
- How to describe graphs: specific characteristics
  - Degree, Betweenness, Transitivity, Distance
- Complex Networks: Global Characteristics
- Networks Structure
  - Communities, Motifs
- Networks Robustness

Bioinformatics and Computational Biology Unit
Molecular Genetics Laboratory - University of Rome Tor Vergata - prof. Gianni Cesareni