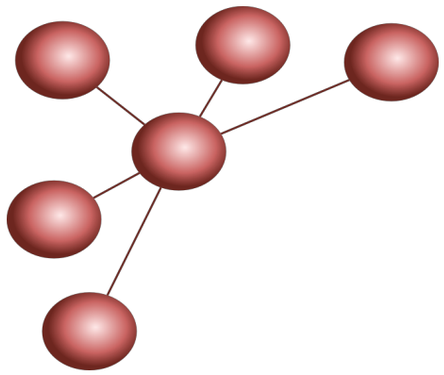
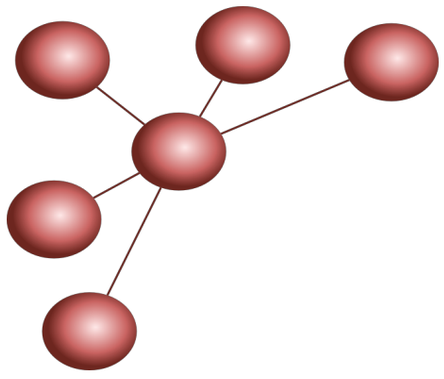


DUBii 2019

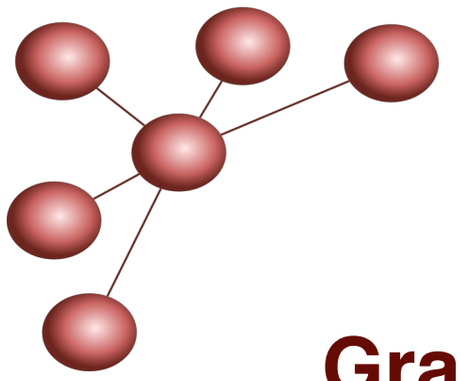
Introduction to Network Science
Introduction to Network Biology

Anaïs Baudot anais.baudot@univ-amu.fr
Costas Bouyioukos costas.bouyioukos@univ-paris-diderot.fr



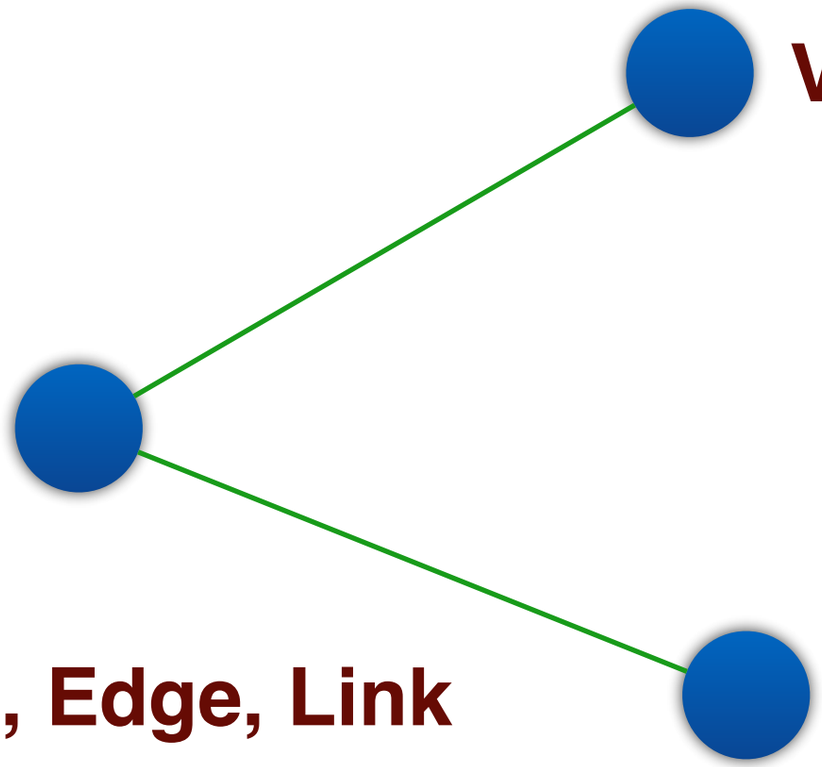
Network Science

- Start 21th century
- Roots on graph theory
- In the context of data production and computer sciences



Definition

Graph, Network, Web

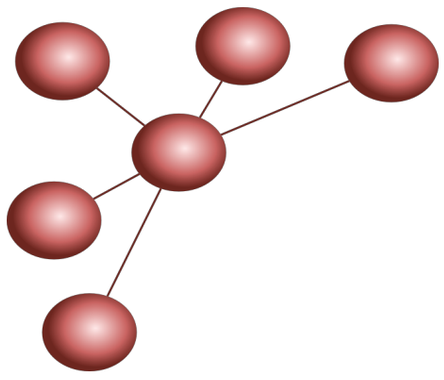


Vertice, Node

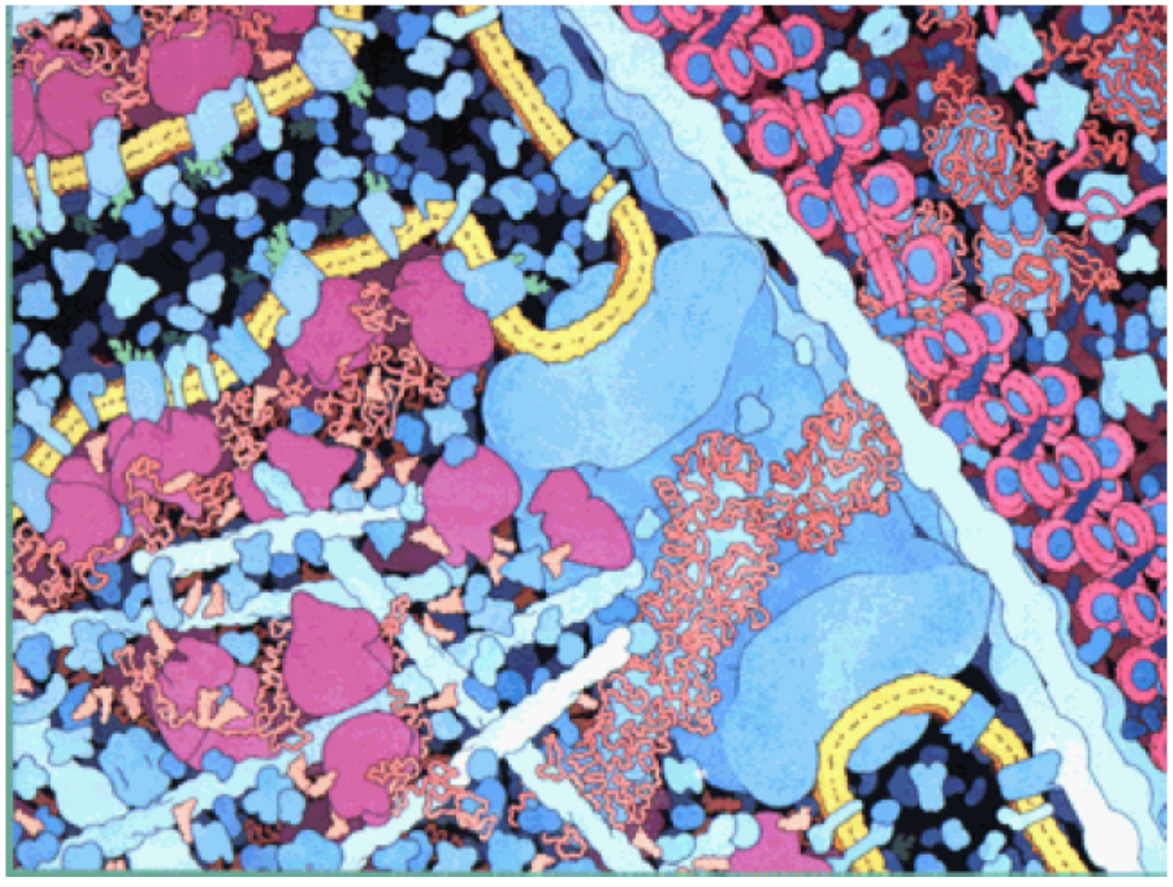
Interaction, Edge, Link

G={V, E}

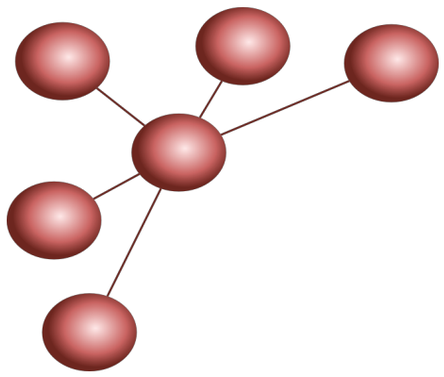
Topology, motifs



Proteins do not act isolated but interact with each other to perform their functions

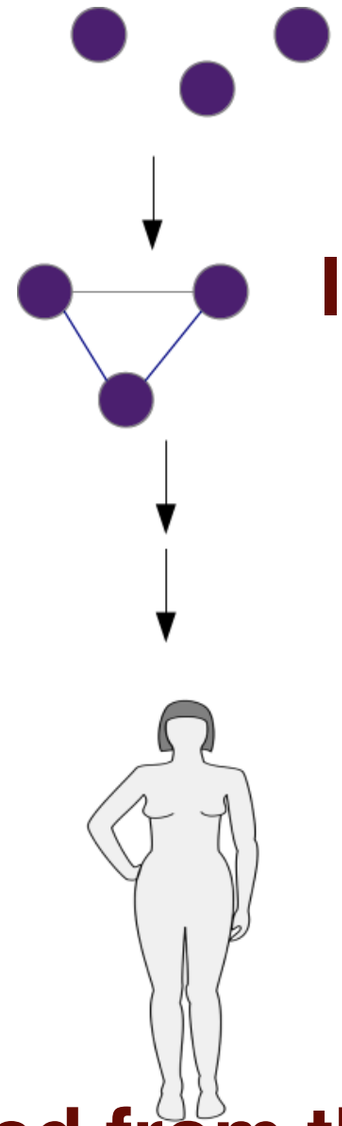


Molecular interactions
Protein-DNA
Protein-RNA
Protein-protein
Protein-lipid
…



Systems Biology because living organisms are complex systems

Systems component : genes/ proteins

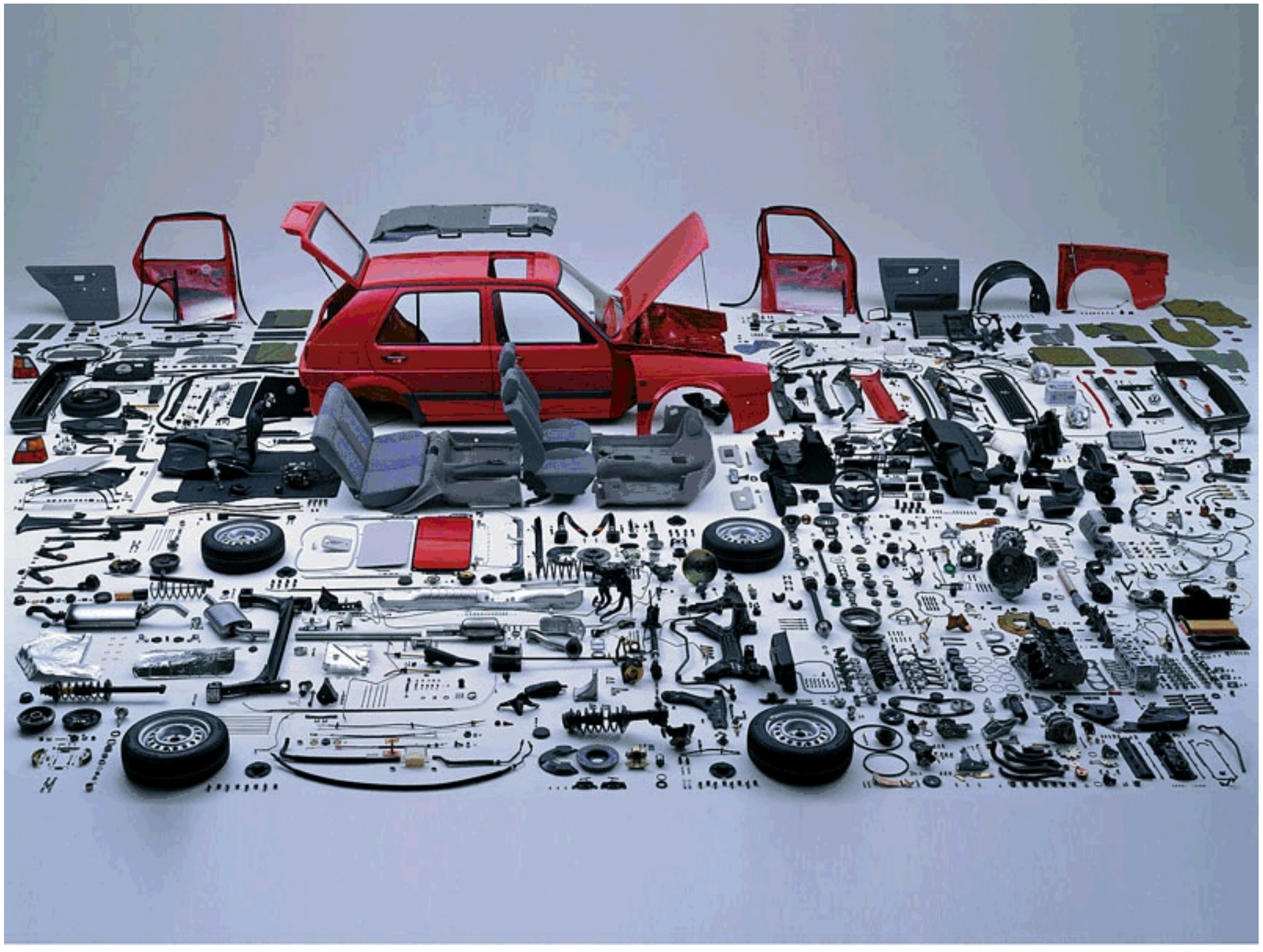


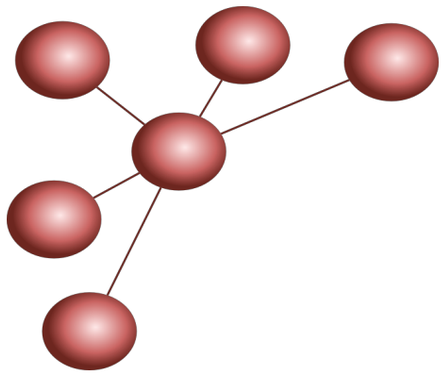
Interaction Networks

Emerging properties : phenotypes

- **Global/collective behavior cannot be deduced from the knowledge on the components**
- **Phenotype does not emerge from isolated biological molecules but from their interactions**

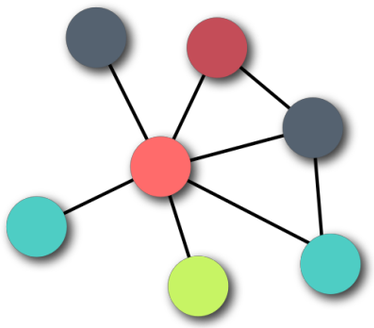
Exemple :



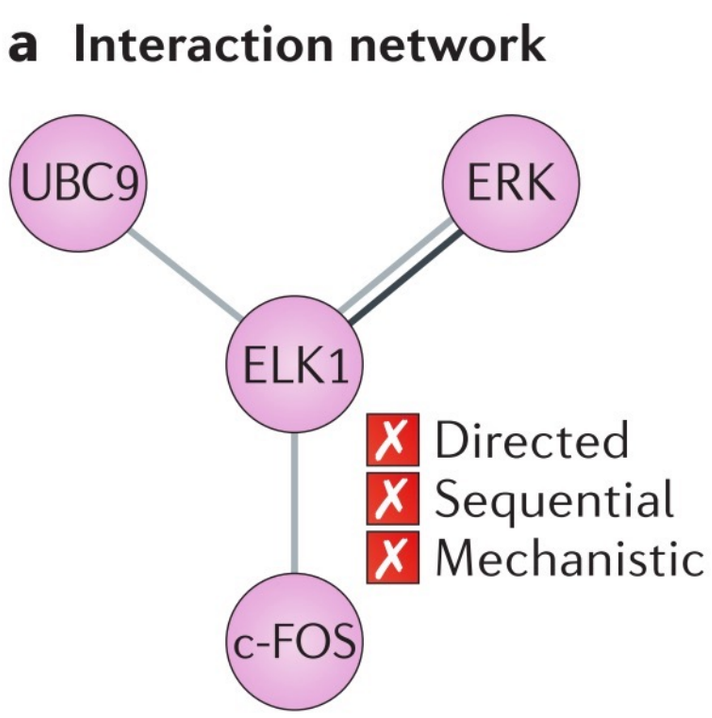


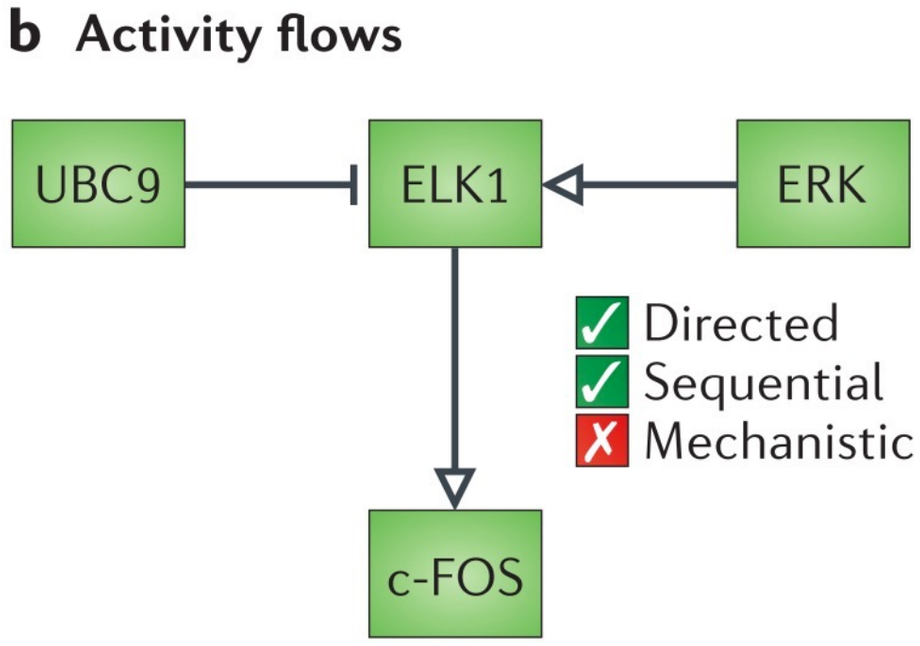
Biological Networks

- From literature, knowledge, curation
- From large-scale interaction experiments
- From inference from large-scale experiments

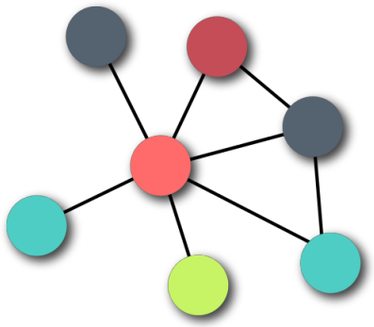


2 main types of networks to represent biological information flow

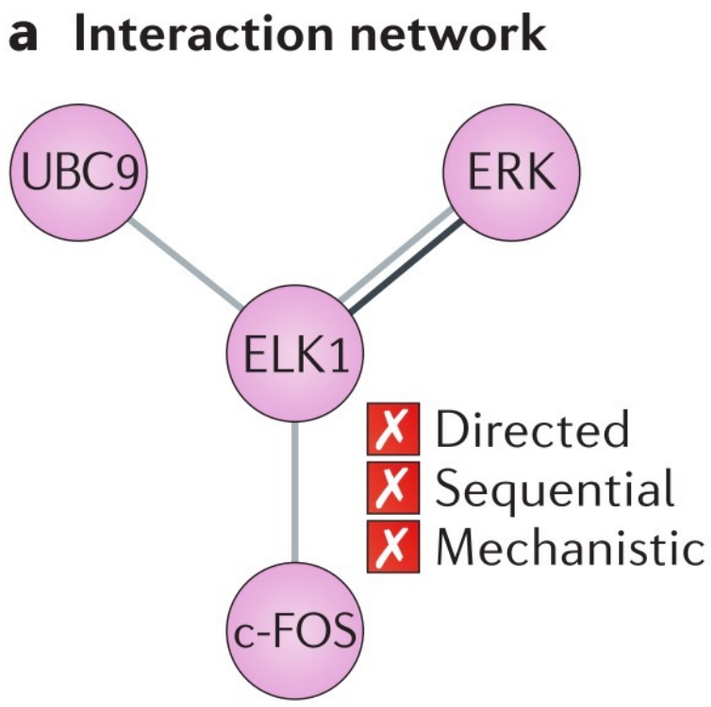




Le Novère et al. 2015

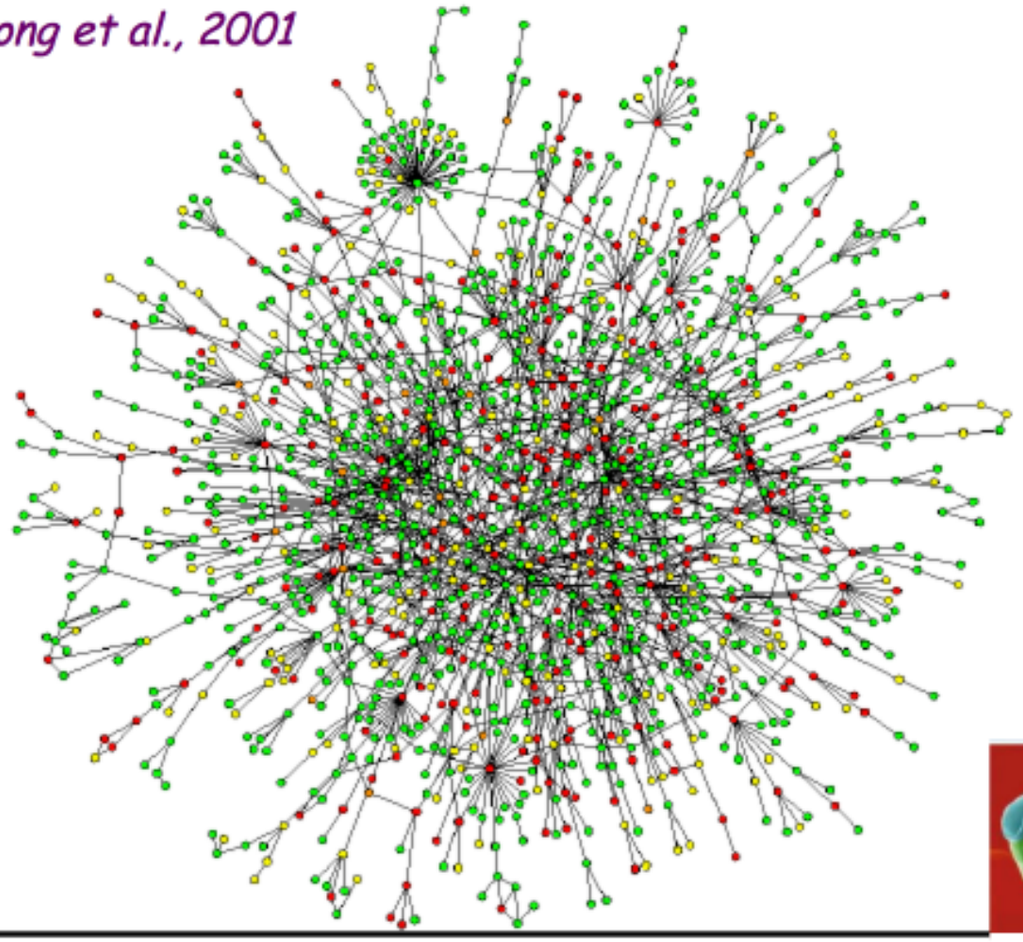


Interaction networks / Interactome

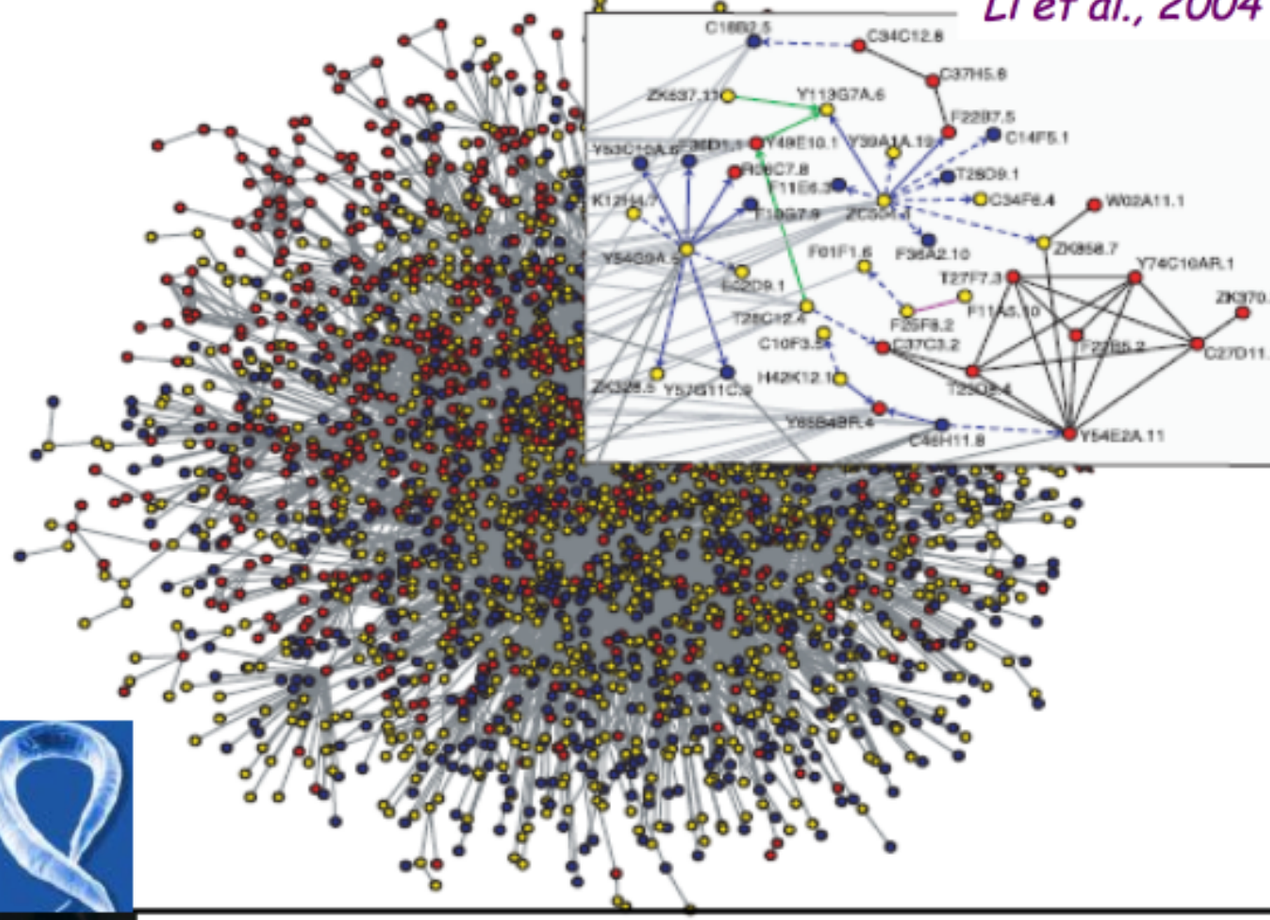


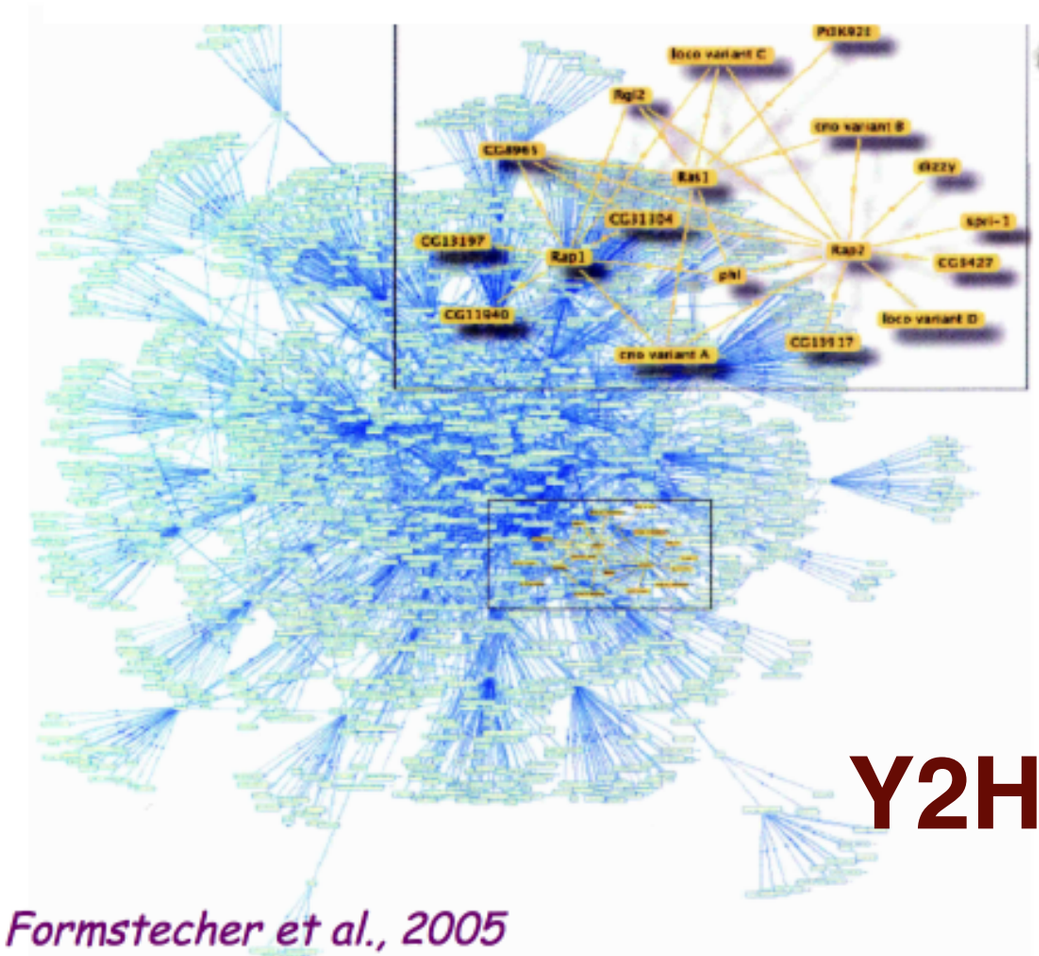
Le Novère et al. 2015

Jeong et al., 2001



Li et al., 2004

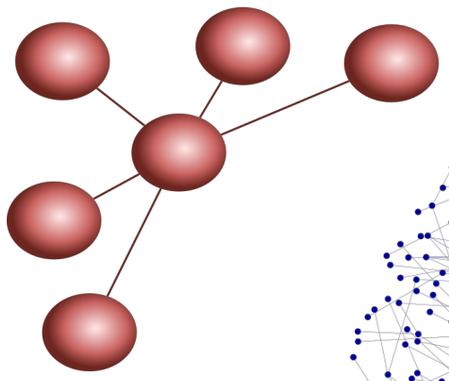


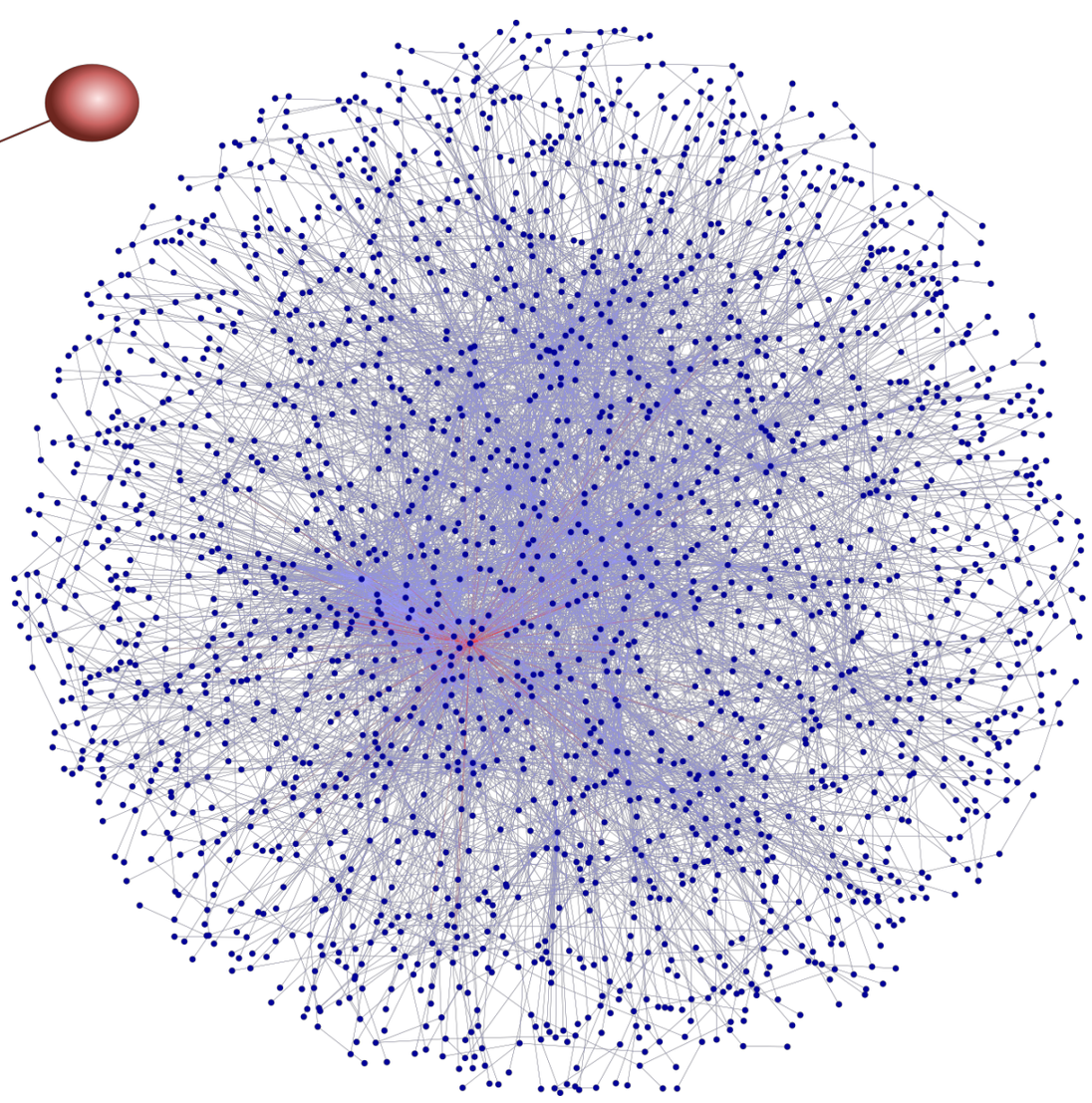


Formstecher et al., 2005

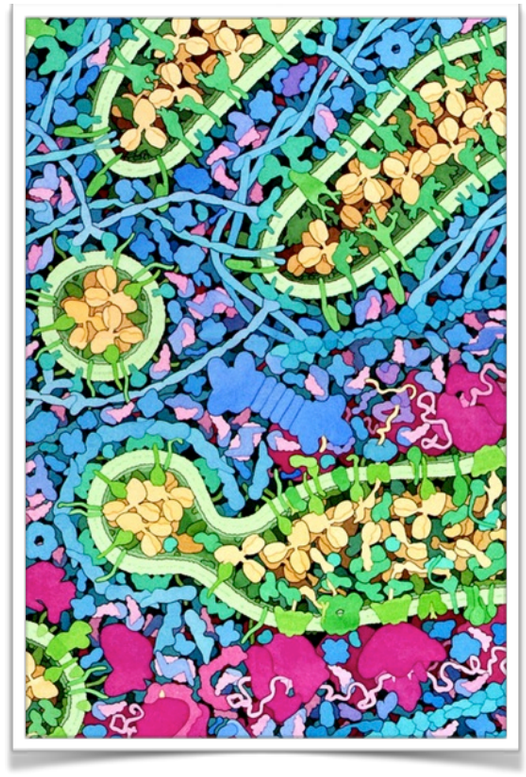
Y2H screens

Rual et al., 2005





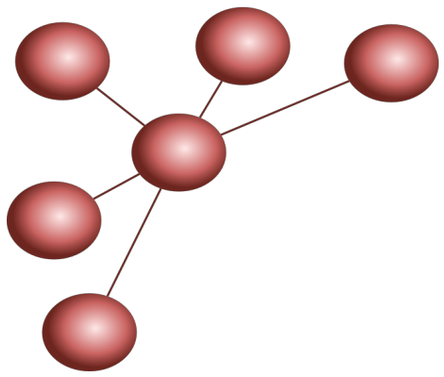
Interactomes



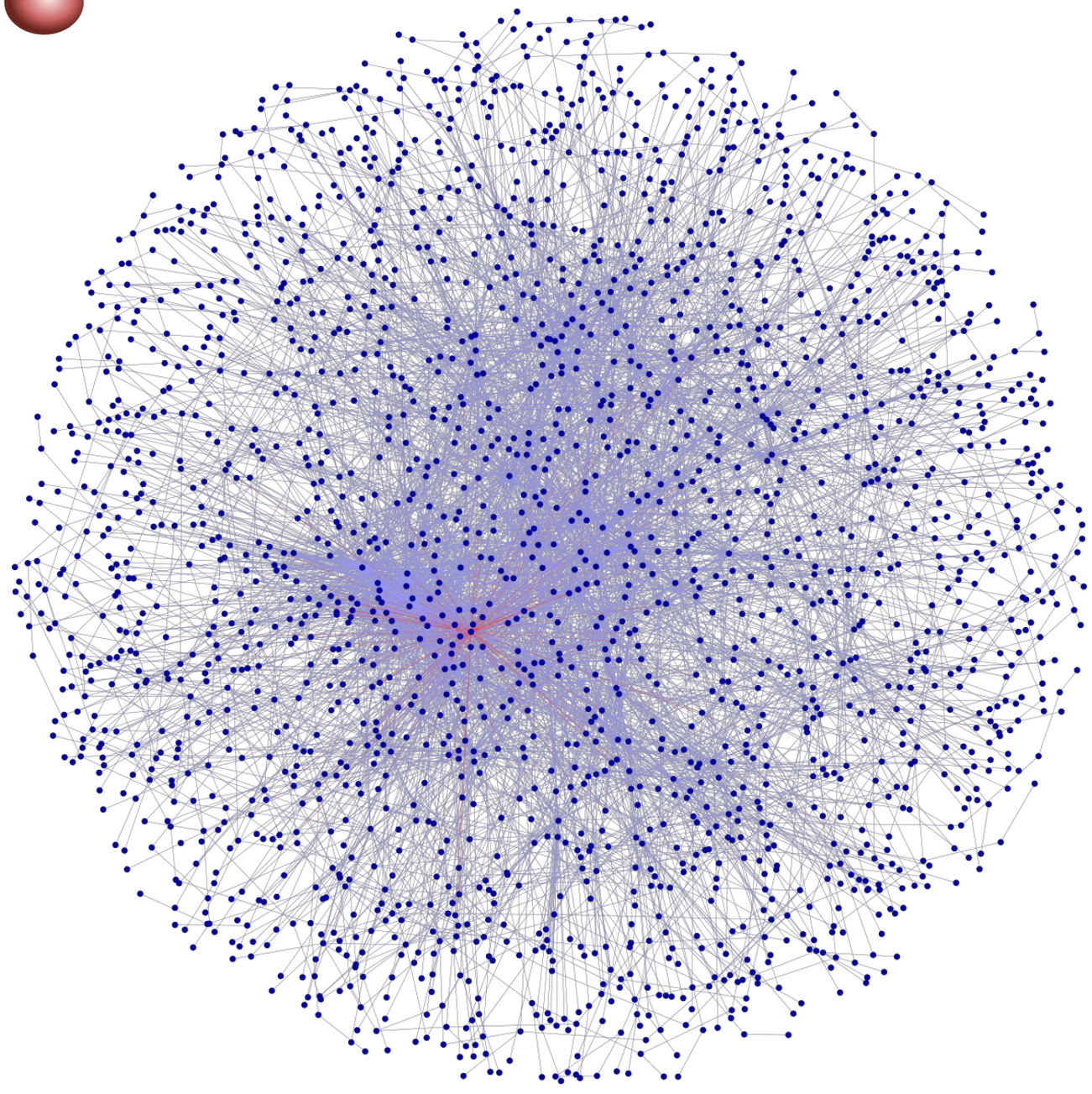
Set of protein-protein interactions detected in an organism

Physical interactions, but physiological interactions ?

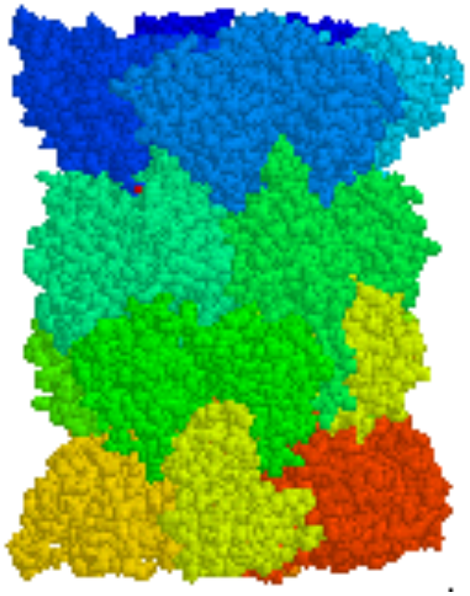
Interactomes are devoid of spatio-temporal information

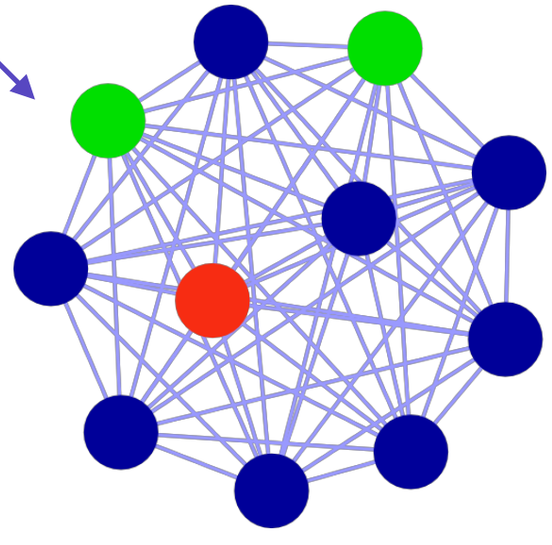


Interactomes



Protein-protein interaction networks





Protein complexes

Interaction databases

Multi-organisms:

- DIP (dip.doe-mbi.ucla.edu)
- IntAct (www.ebi.ac.uk/intact)
- MINT (mint.bio.uniroma2.it/mint)
- BioGRID (www.thebiogrid.org)
- BIND (www.blueprint.org)

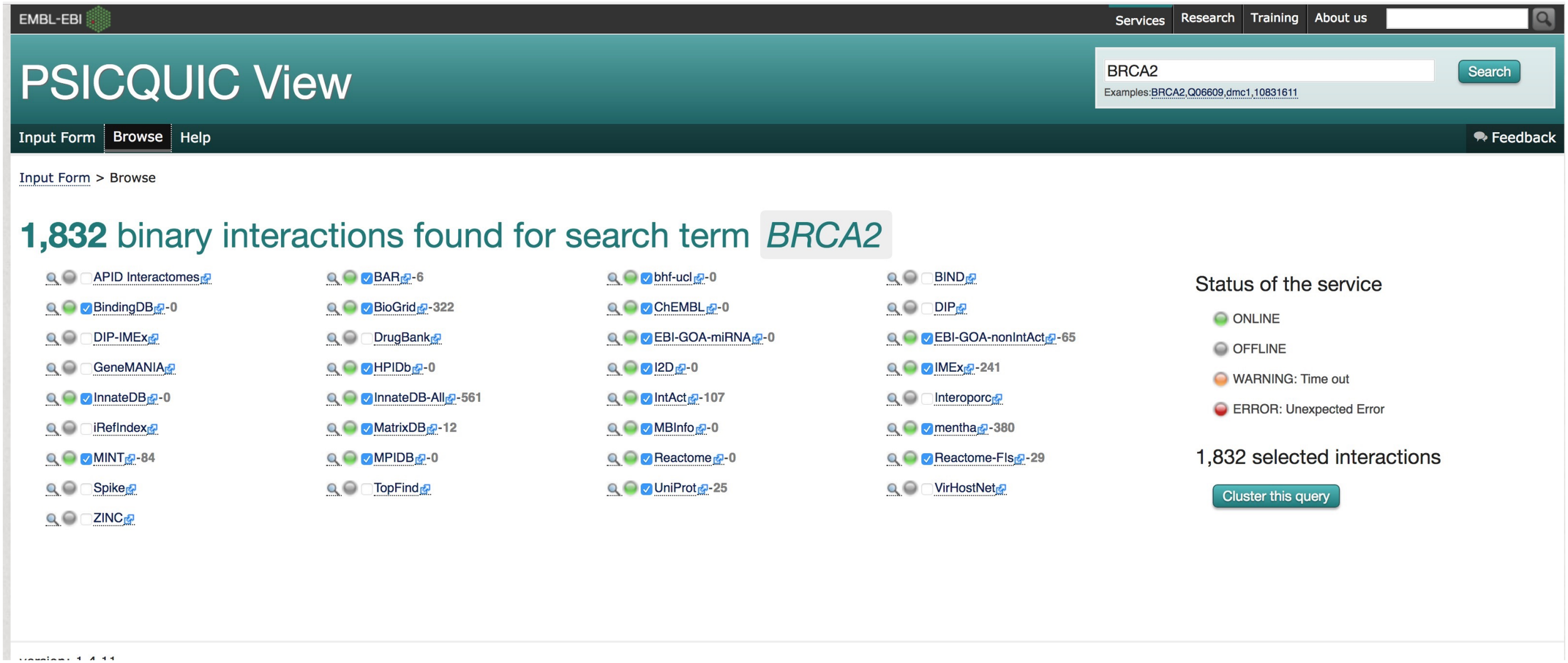
Human

- Reactome-FI

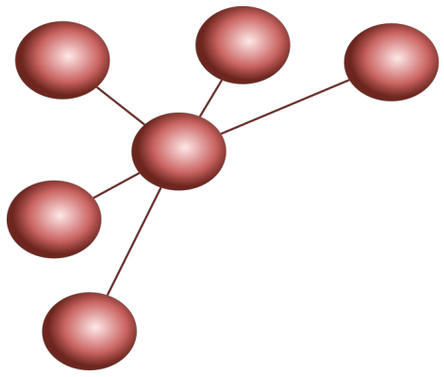
IMEx

International Molecular Exchange Consortium

Psicquic portal

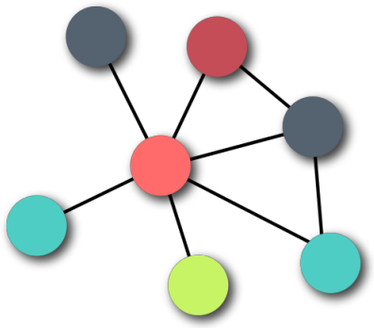


http://www.ebi.ac.uk/Tools/webservices/psicquic/

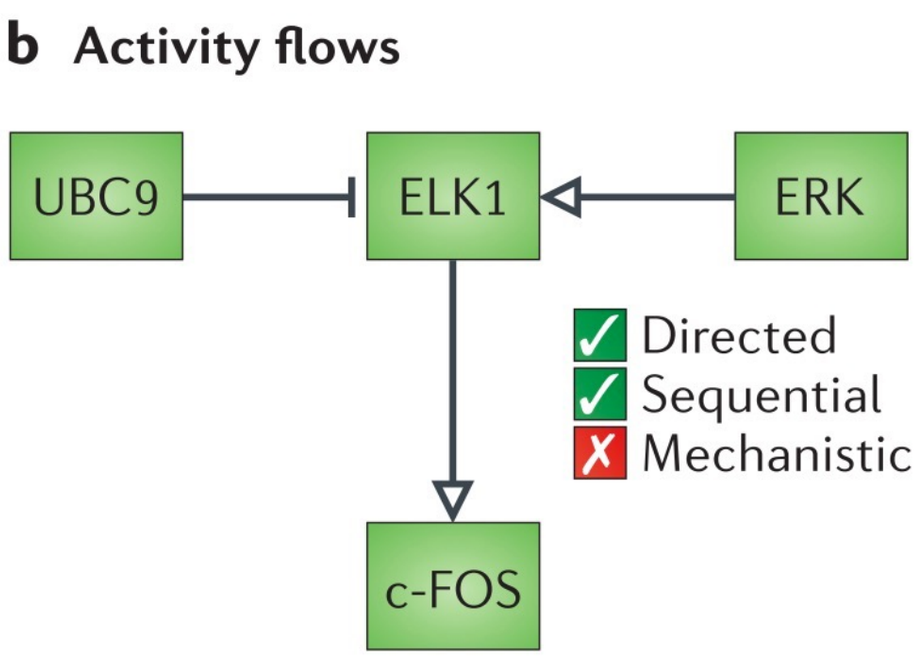


Network inference from -omics data

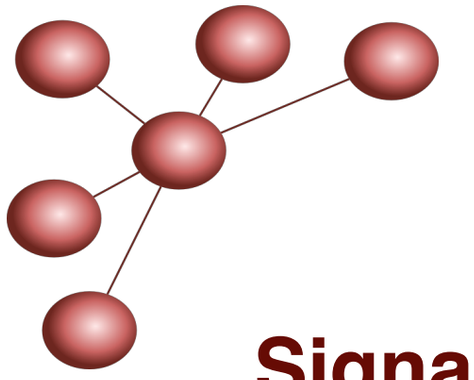
- Co-expression networks from transcriptomics data => Session #2



Activity flow / Gene Regulatory Networks / Influence Graphs

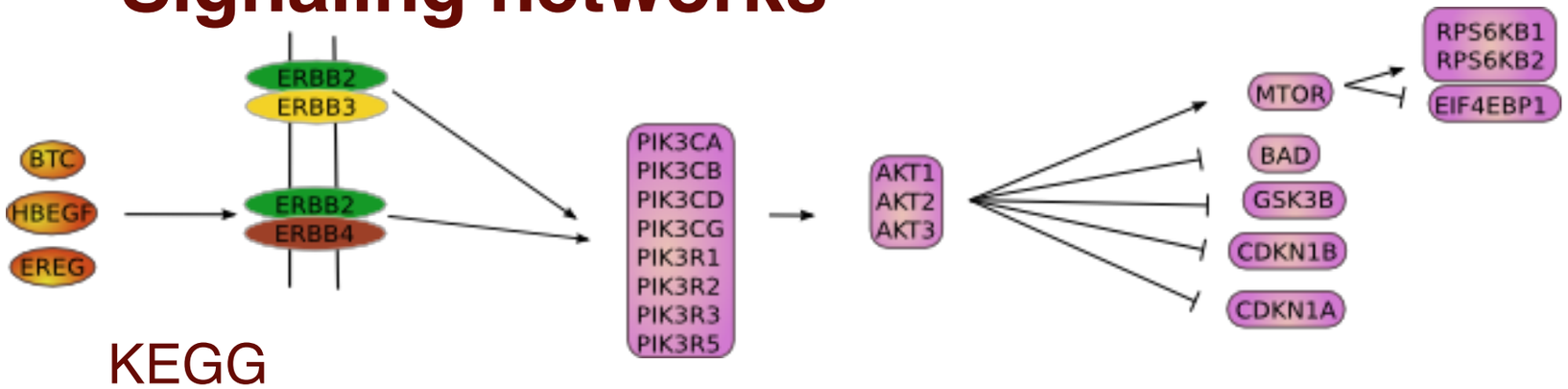


Le Novère et al. 2015



Signaling Networks

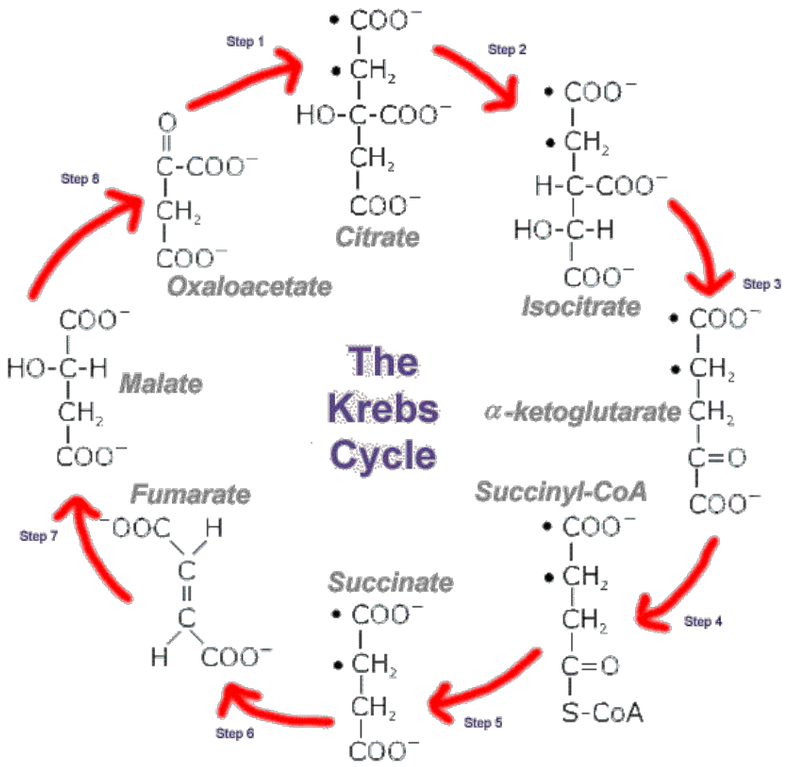
Signaling networks



KEGG

Metabolic networks

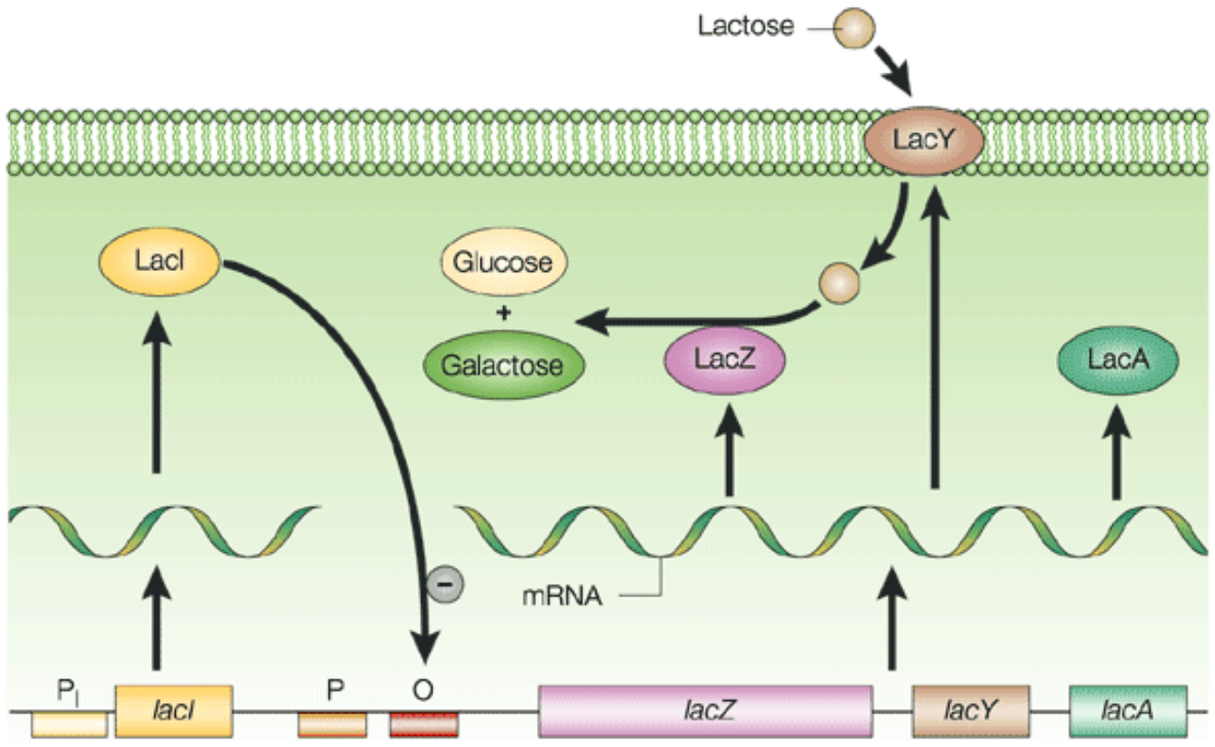
2 types of nodes : enzymes & substrates, reaction directional or bidirectional



Metabolic Cycle

Gene Regulatory Networks

Operon / Regulatory networks



Opéron lactose

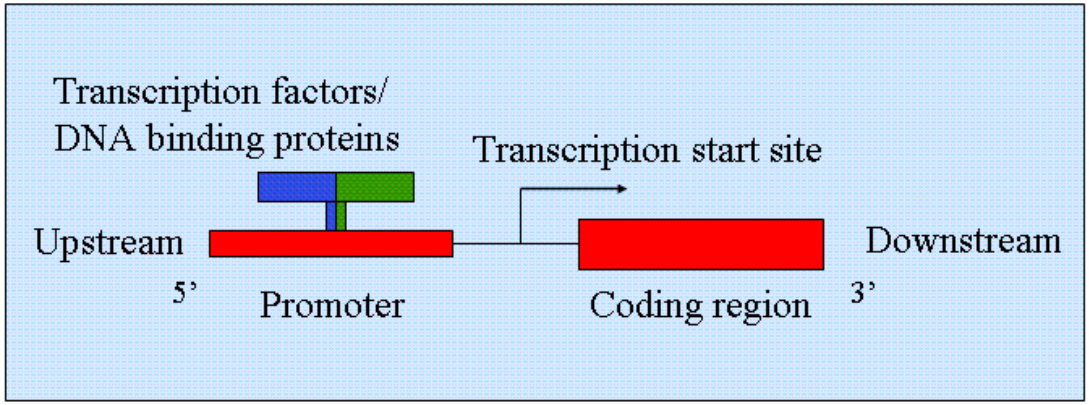
Nature Reviews | Genetics

Shuman et al.

Refreshing Gene Regulation

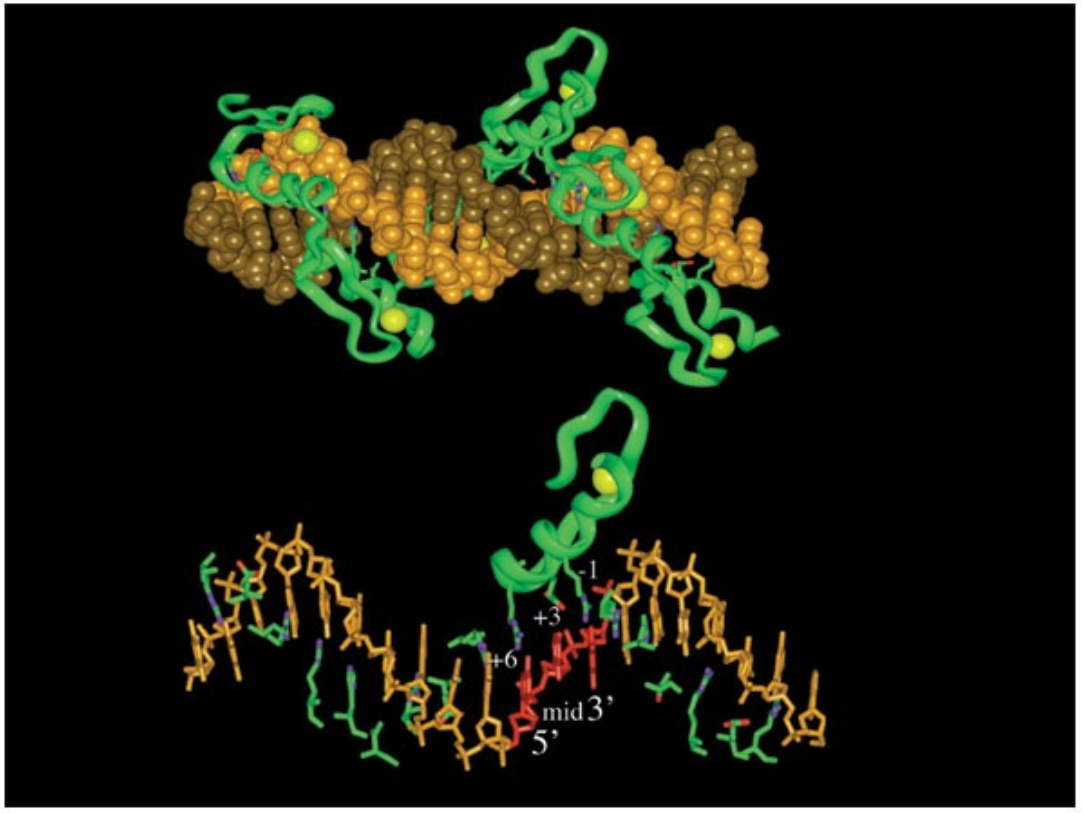
- A gene consists of a coding region AND a regulatory region
- There are special "sites", called Transcription Factor Binding Sites (TFBS)
- TFBS are short DNA sequences, with specific nucleotide composition, recognised by a TF.

Gene Control: Regulatory Regions



- There can be one or more transcription factors (also called DNA binding proteins) that can initiate (or stop) transcription.
- The transcription start site is where RNA polymerase transcribes mRNA from the DNA template.

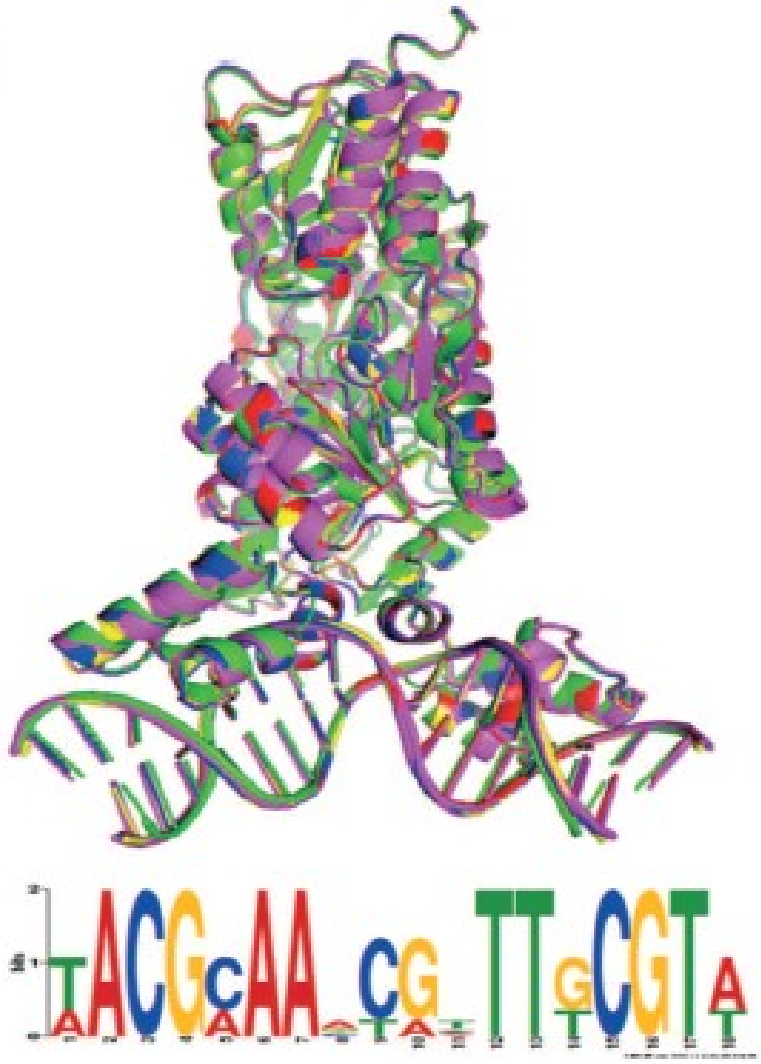
TF binding to a TFBS



- Transcription Factors (TFs) are special type of proteins which bind to DNA.
- TFs “recognise” via their structure (i.e the amino-acid sequence) only a specific sequence on the DNA major grove.
- The “strength” of the binding (i.e. affinity) depends on the TFBS.

What are TFBSs (and motifs)

- Genomic sequences that are found in the <u>regulatory region</u> of genes.
  - a) Short
  - b) Conserved
  - c) Characteristic nucleotide probability (Logo)
- Can be computationally predicted (not see today).

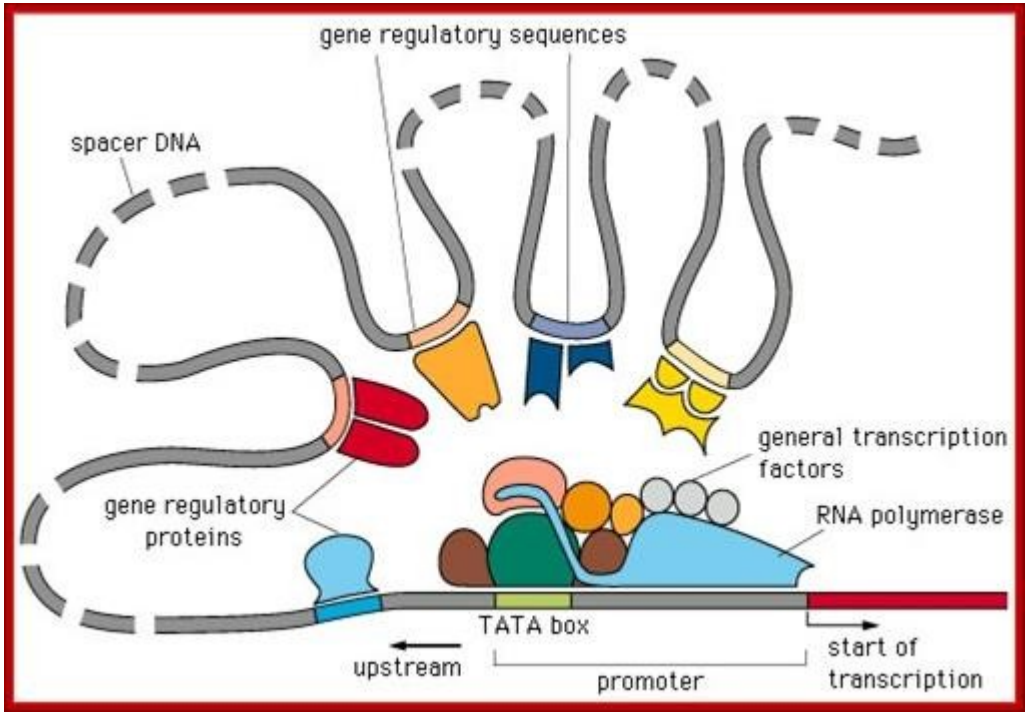


How all this works together

- A whole apparatus is organised in space and regulates transcriptional activity and comprises:

  a) DNA sequences: *cis*-regulatory elements, TFBS

  b) TFs: proteins, regulation in *trans*

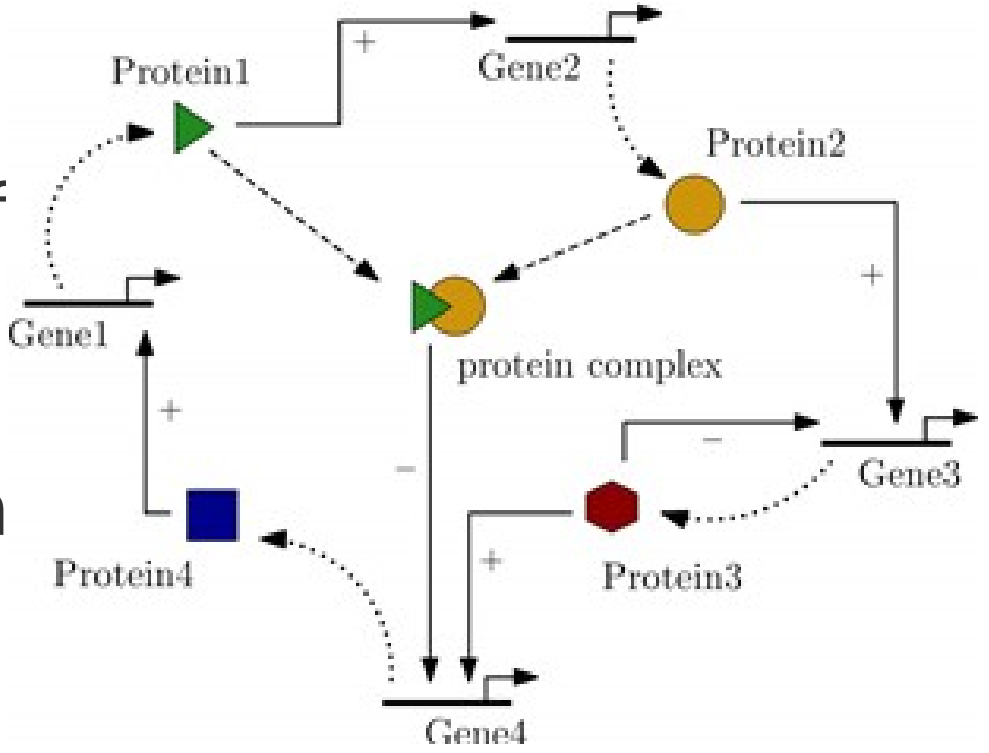


- BUT TFs are also <u>proteins</u>, which are encoded by <u>genes</u> which are also regulated by…

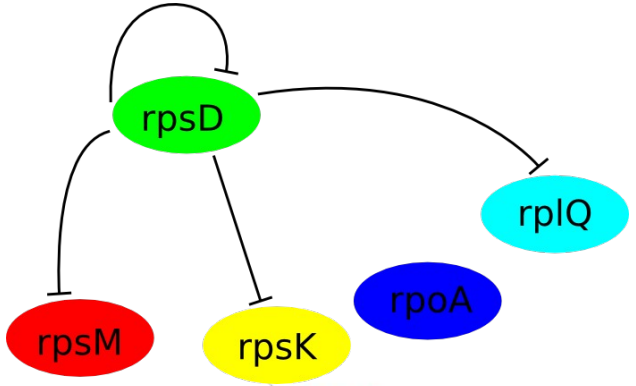
Gene products regulate the expression of other genes.

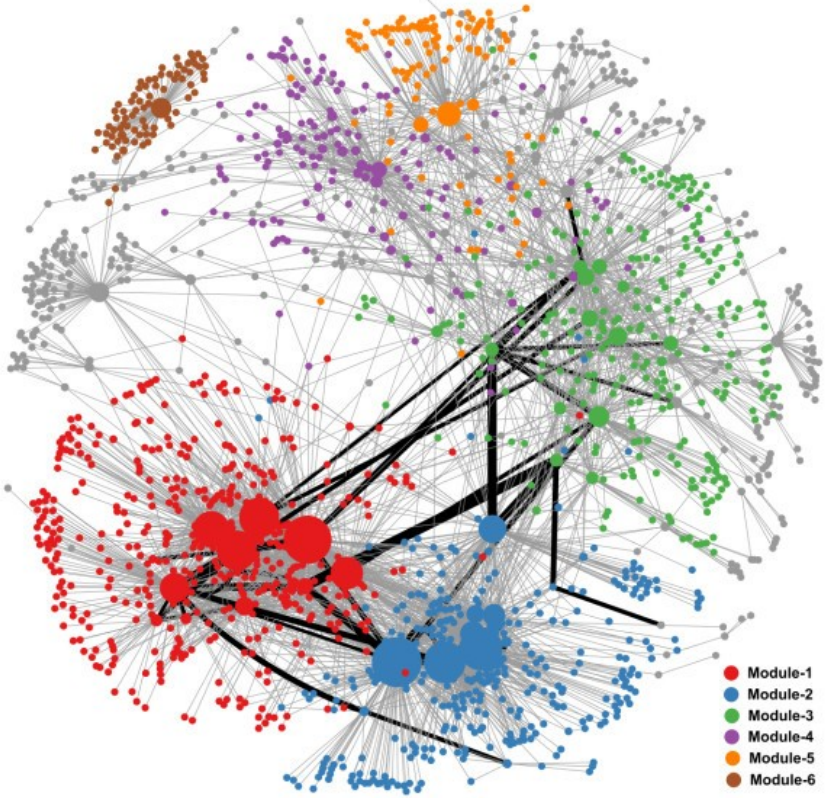
- The product of *Gene1* might form a complex with the product of *Gene2* and together repress the expression of *Gene4.*
- To simplify we just plot the connections between genes.
- Formally a GRN is:

  A network of interactions between genes and their products.



GRNs are represented by graphs





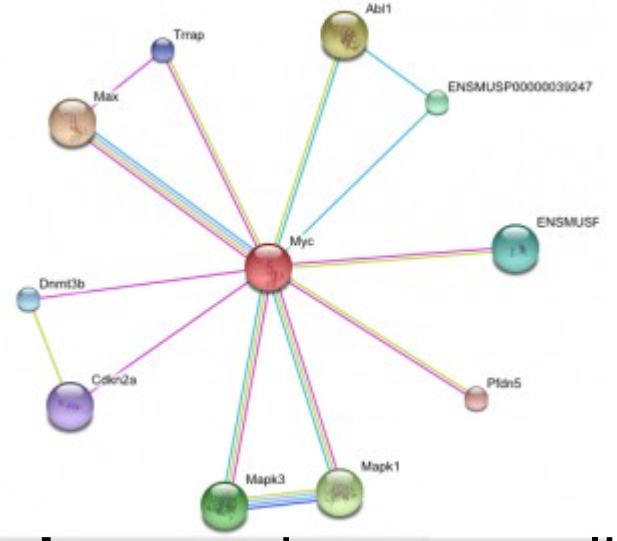
- It can be something simple like the first...
- ... but in fact they look more like the second.
- In any case, they contain arrows that represent the interactions between genes, the direction and the nature (activation, repression).
- For that we need specialised tools to view and analyse.

A few words on *graphs!*

Graphs consists of:

- ***Nodes or Vertices***

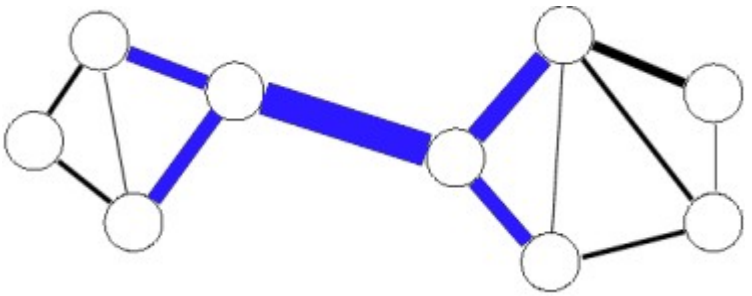
  The nodes represent genes (or gene products) in a GRN.



The **degree** –how many links, is the most important property of a node.

- ***Links or Edges***

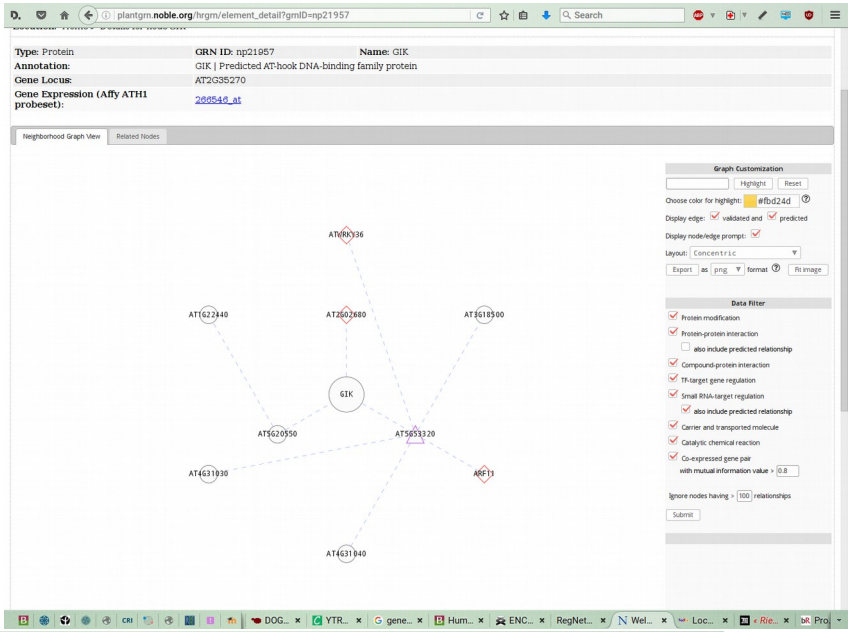
  The links represent interactions in a GRN

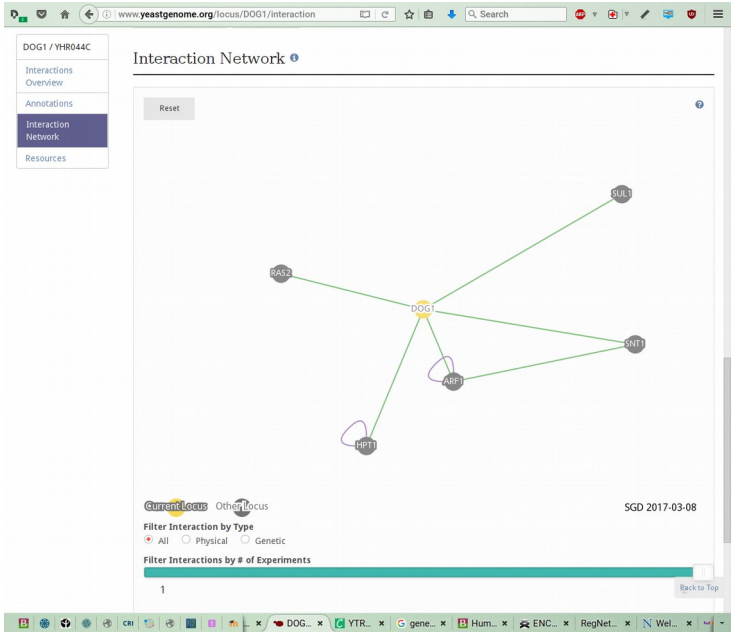


The **betweenness** –how many paths, is the most important property of edges.

GRN online resources.

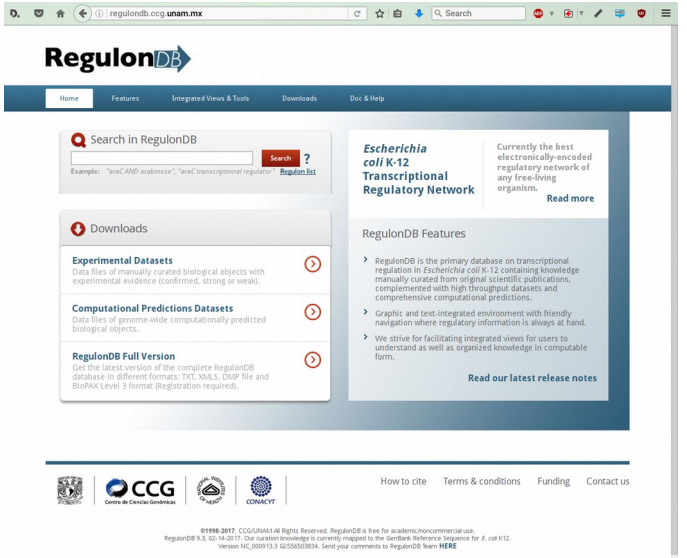
- There is an unknown number of databases that holds gene regulation information.
- For the record keep some: RegulonDB, Jaspar, PlantRegMap, Plantgrn, Yeastract, SGCB and *of course* ENCODE (too controversial).
- Today we will see a simple example with RegulonDB.





From online resources to analysis Cytoscape/RegulonDB

- **RegulonDB:** a highly curated database for transcription regulation in *E.coli*
  - *http://regulondb.ccg.unam.mx/*



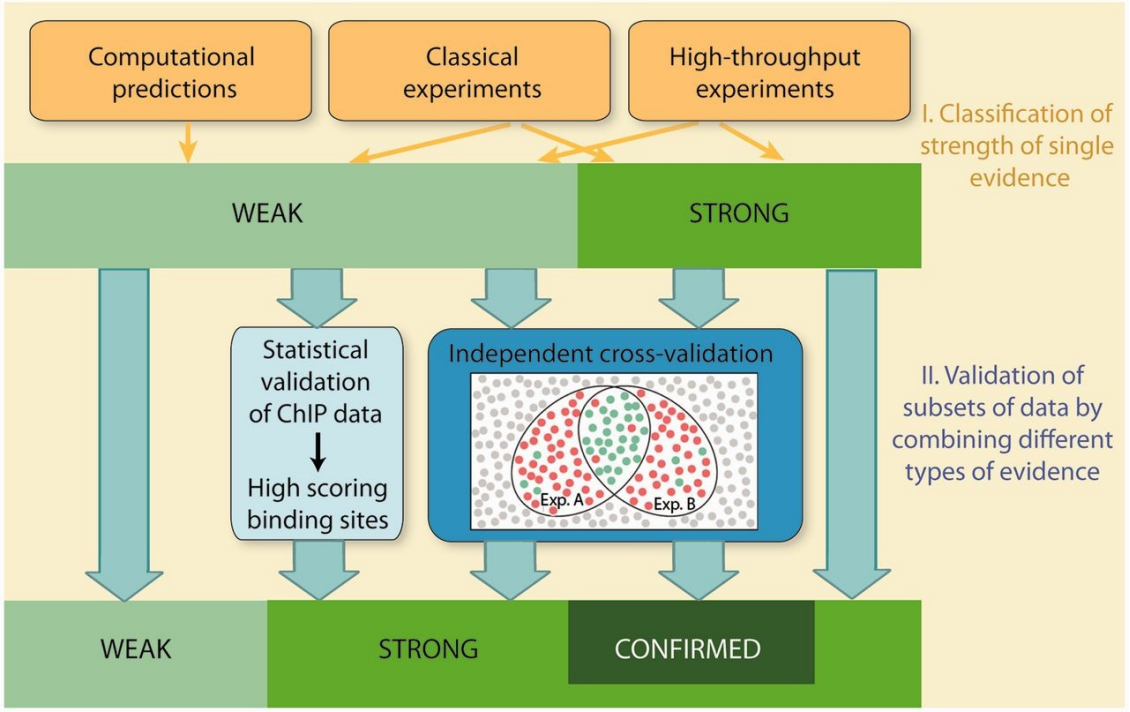
- **Cytoscape:** a versatile and easy to use analysis tool that can do (almost) everything on networks.
  - http://cytoscape.org/

How regulatory interactions have been deciphered

- A combination of expert knowledge and high-throuput datasets.
  - ChIP-Seq, ChIP-Chip, ChIP-exo
  - RNA-Seq, microarrays
  - GSelex
- Two classifications:

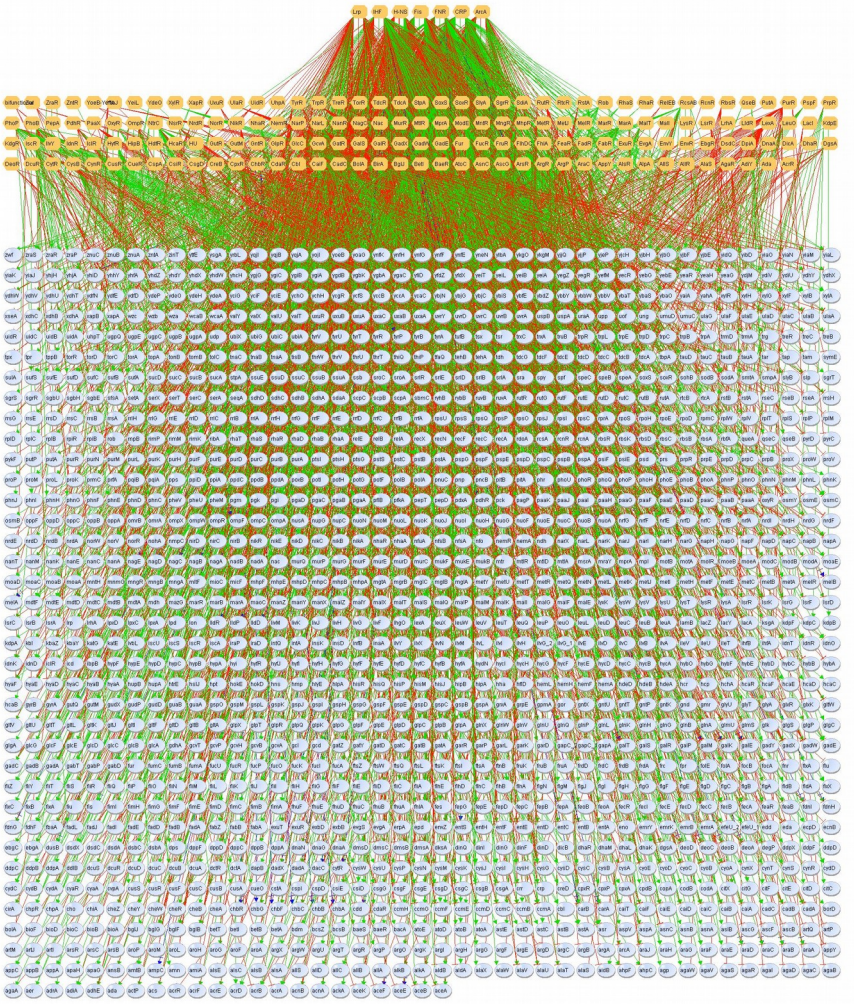
  - STRONG evidence (CHIP-SV)

  - WEAK evidence (CHIP, GEA, GSELEX)

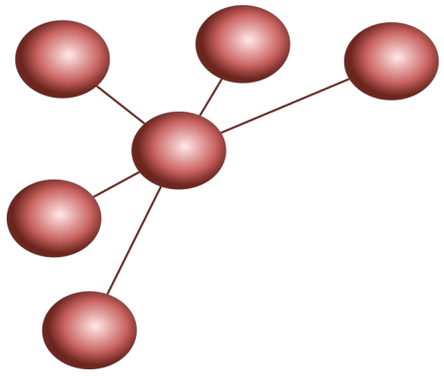


Tables of regulatory interactions

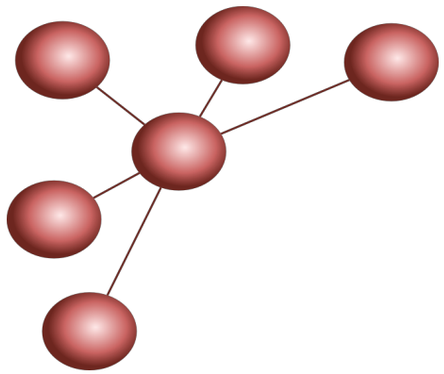
- Start by downloading the GRN of *E.coli* from RegulonDB. We need the gene-TF file and the TF-TF file.
- In RegulonDB looks something like the figure…
- … let‘s see if we can visualise it better with Cytoscape.



http://regulondb.ccg.unam.mx/menu/download/datasets/index.jsp

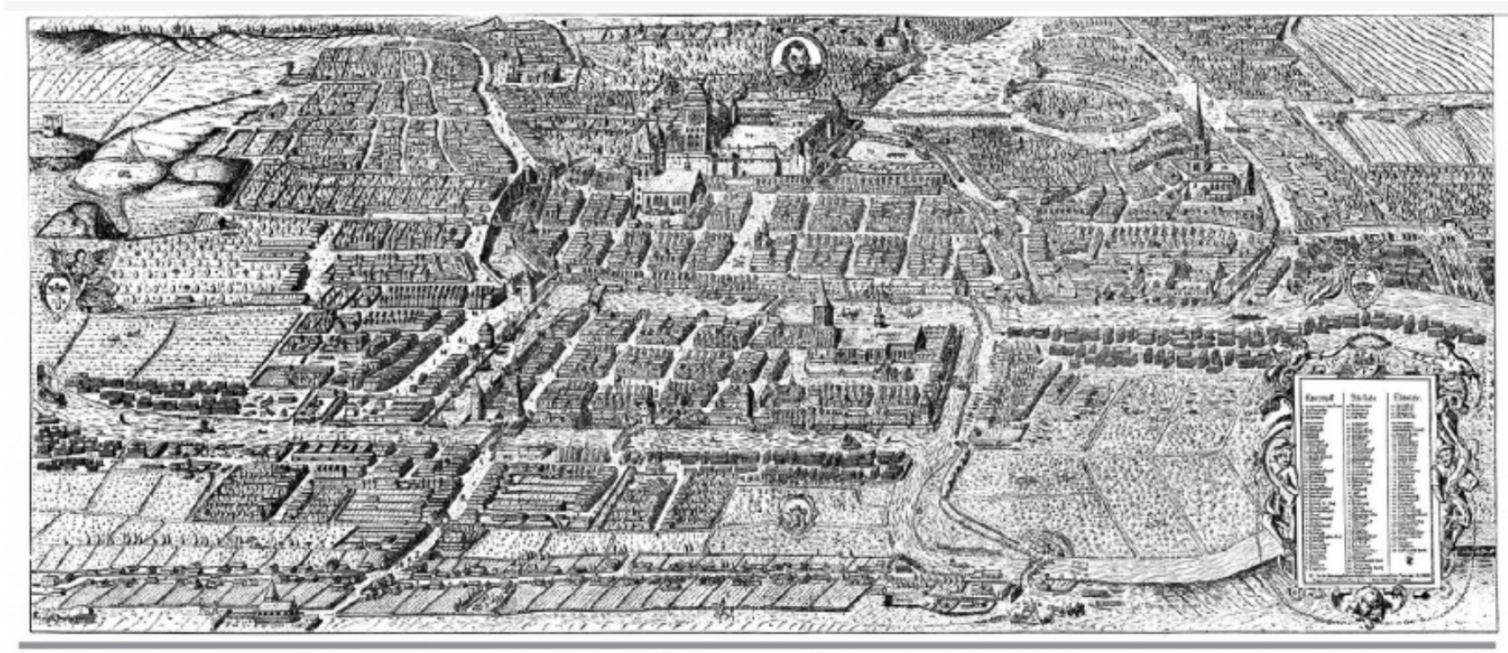


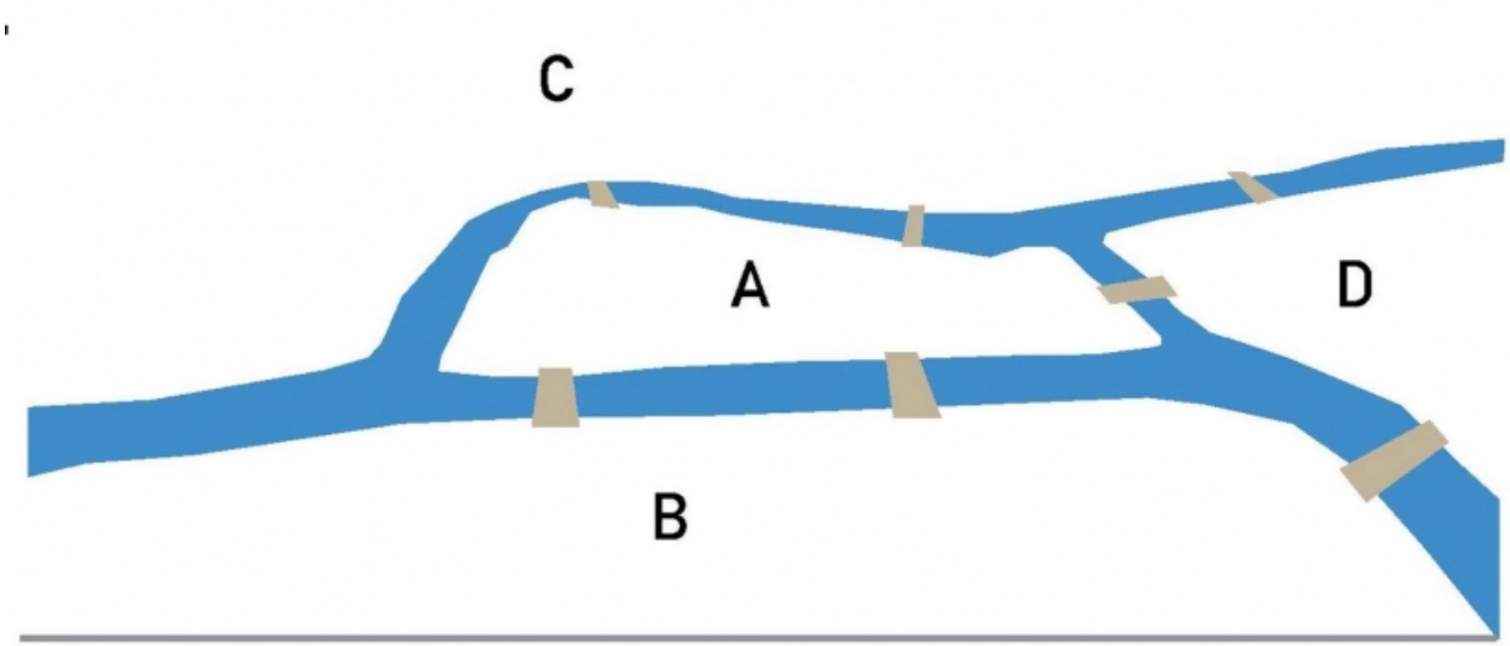
What next ?
=> Network Analysis

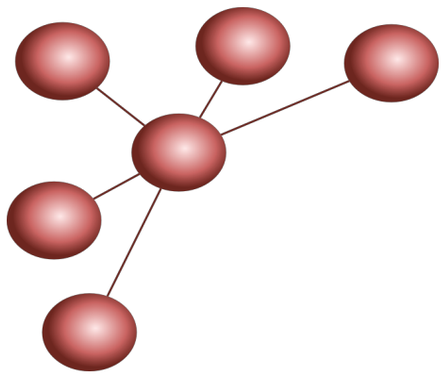


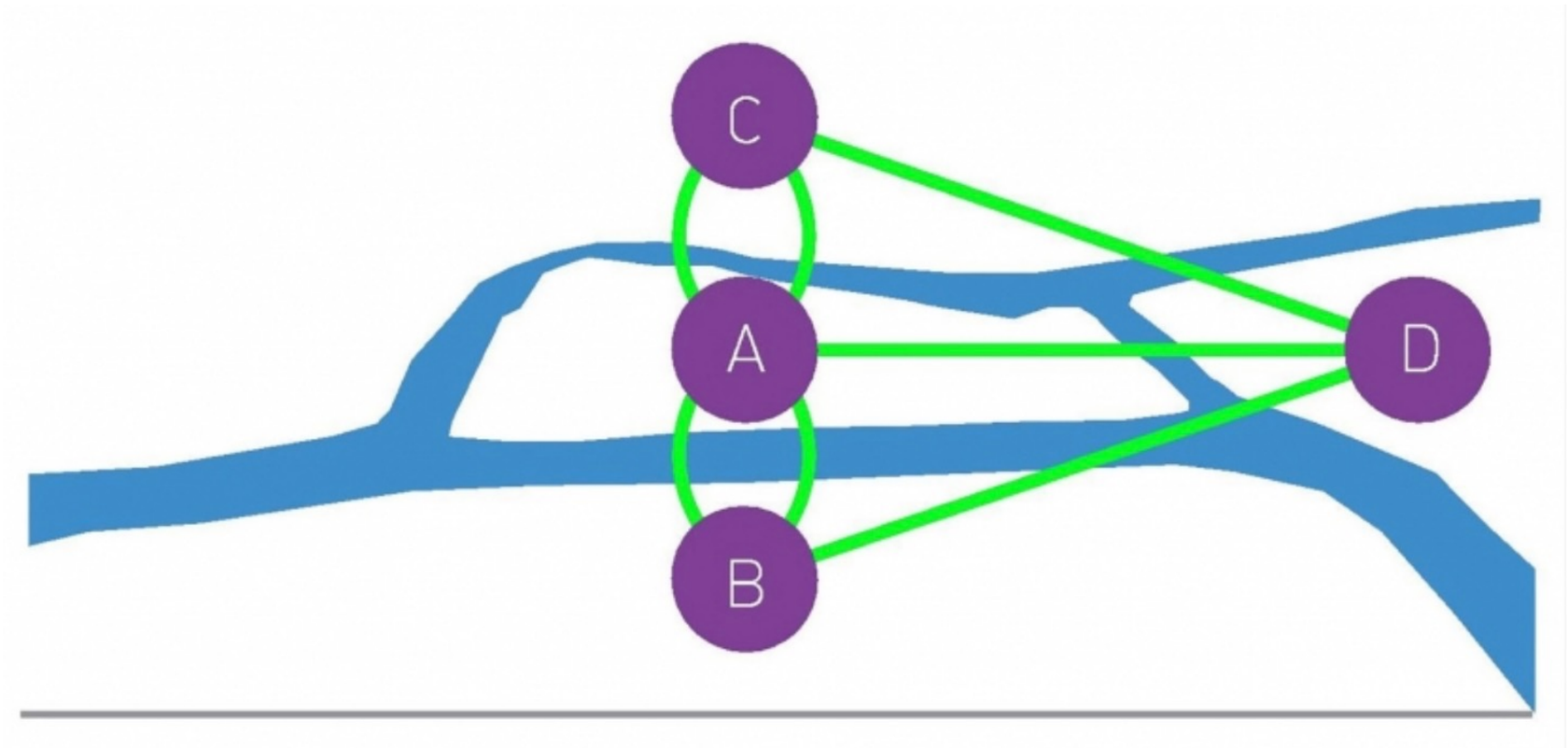
Network analysis / graph theory

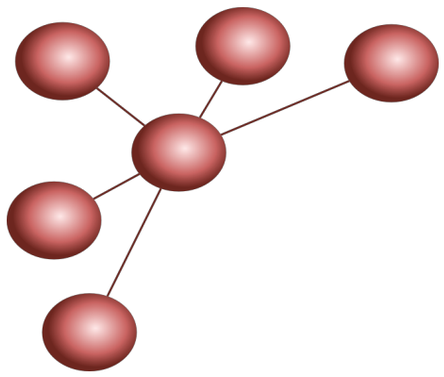
Königsberg





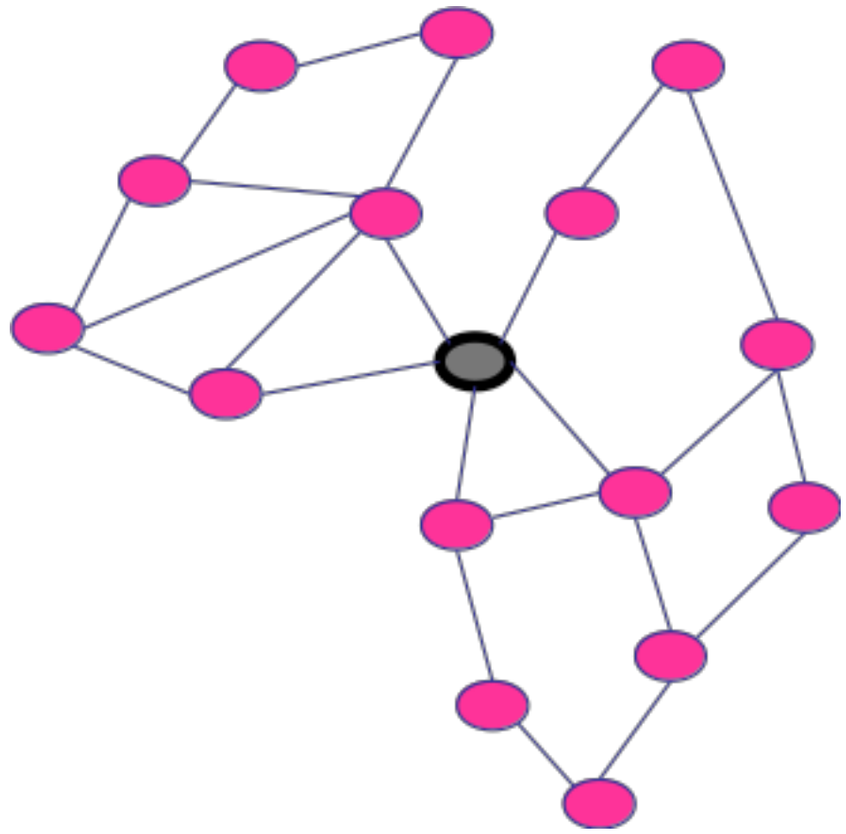



C
A
D
B



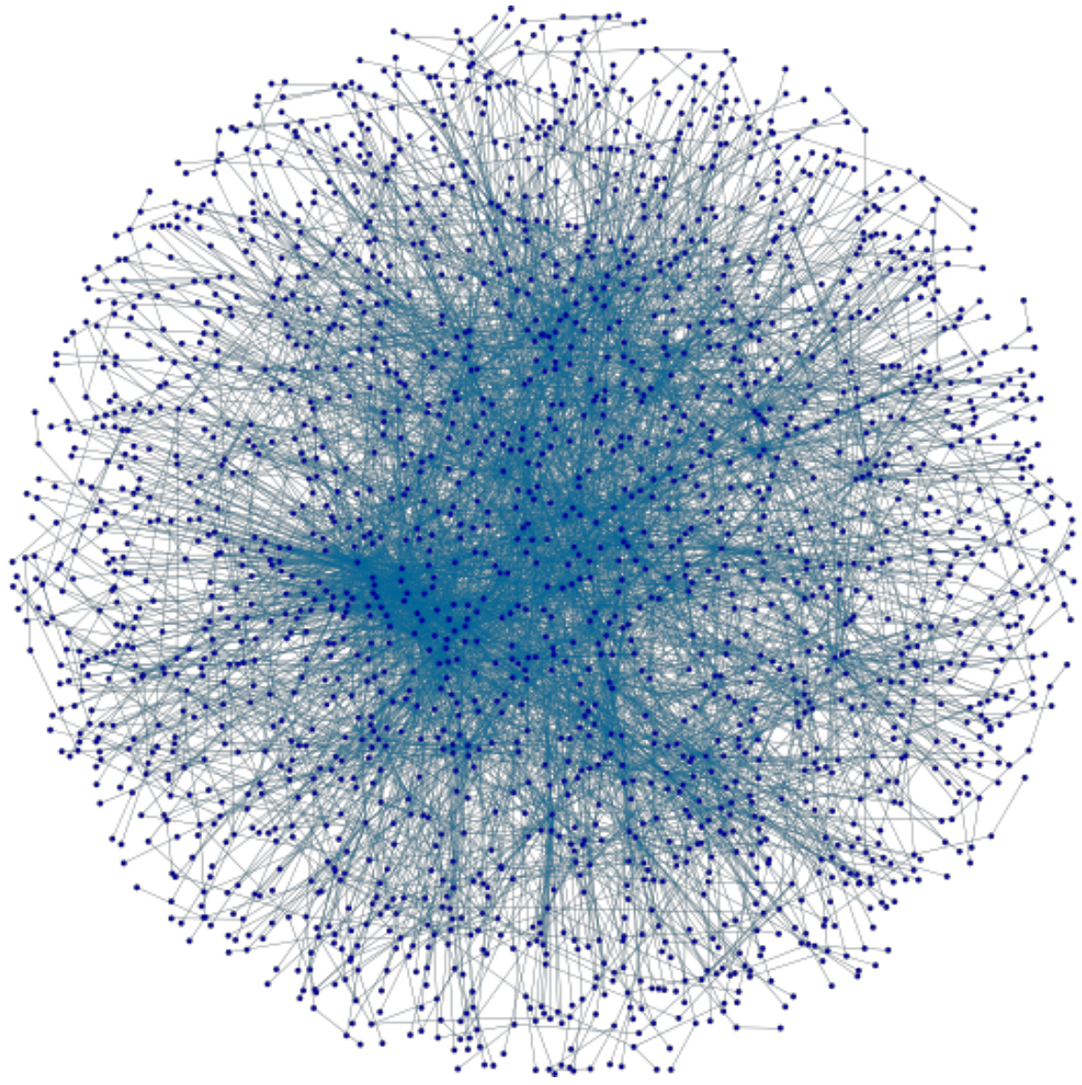
How to use large-scale biological networks ?

Local approaches



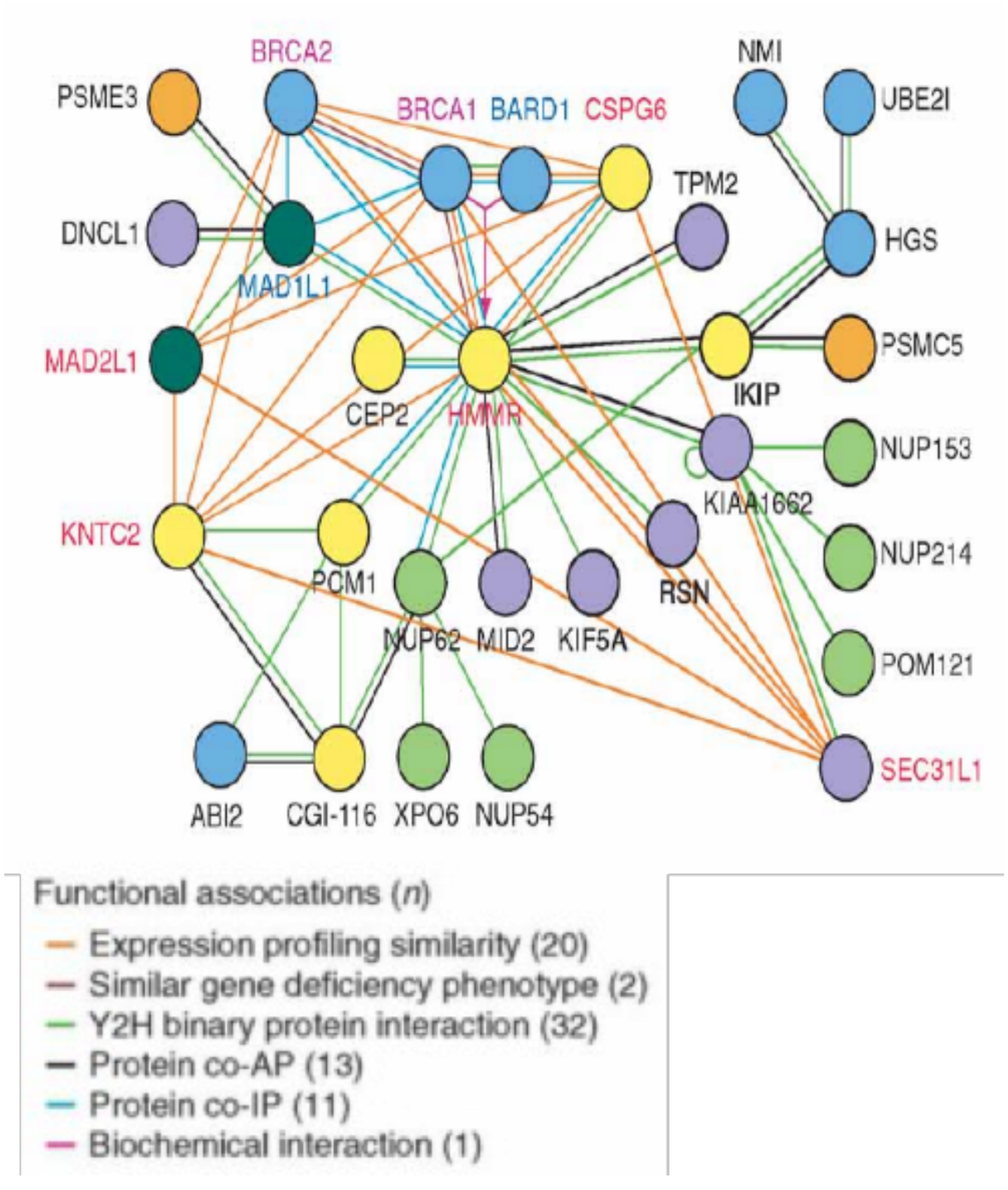
“guilt by association”

Global approaches



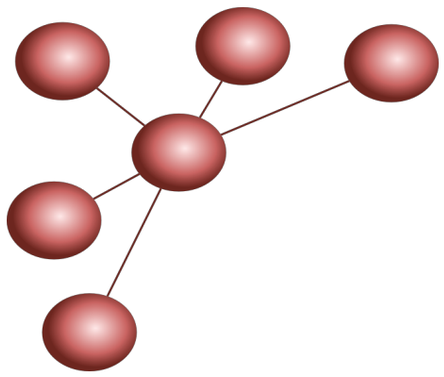
Topological features
Clustering / communities

Identification of a new gene involved in breast cancer



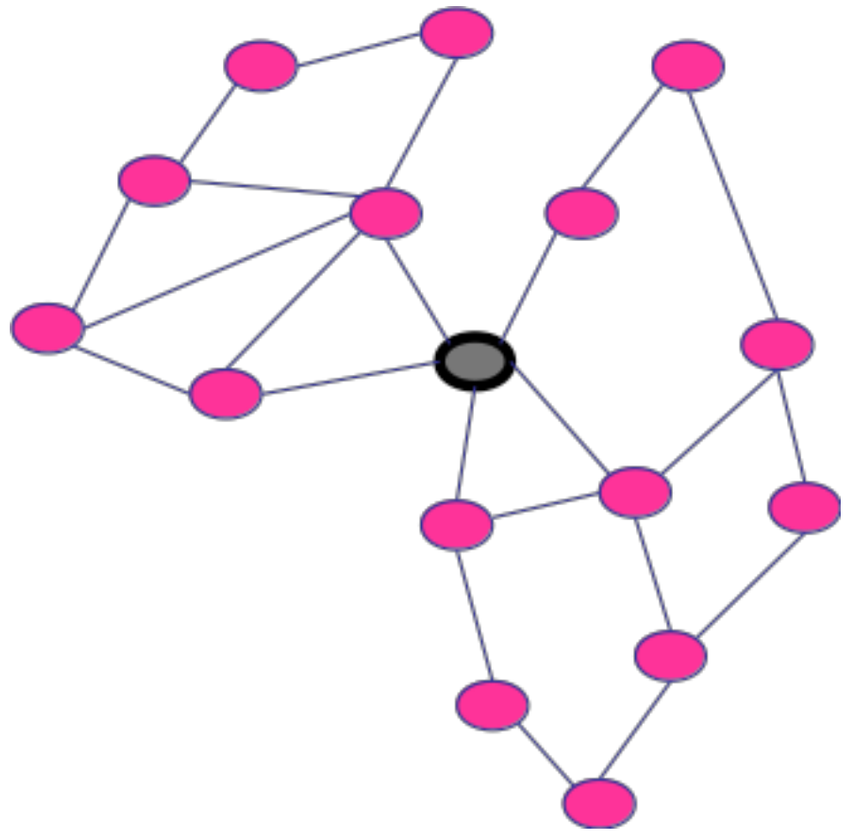
Nodes correspond to proteins, edges to interactions identified by different experimental techniques

Pujana et al. 2007



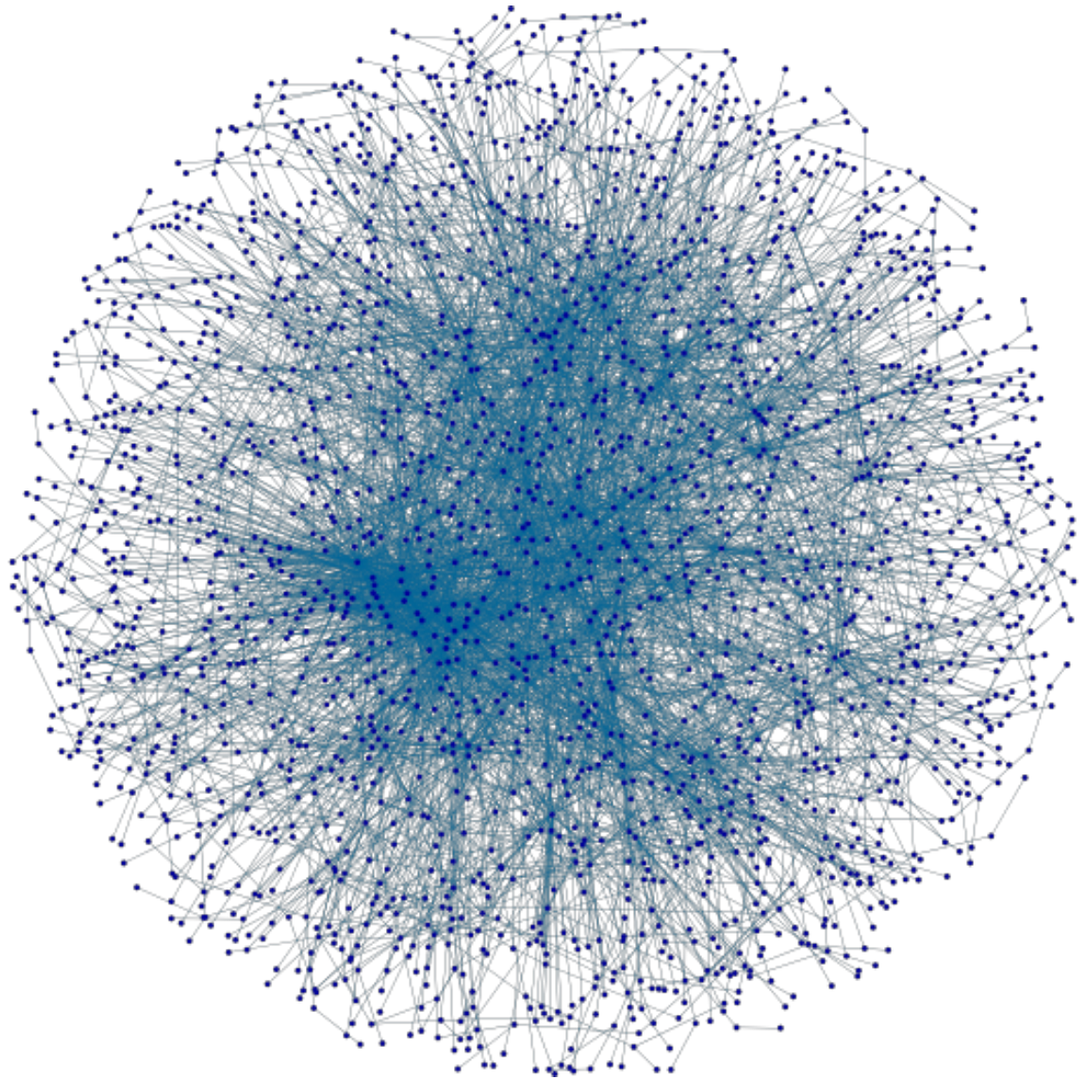
How to use large-scale biological networks ?

Local approaches

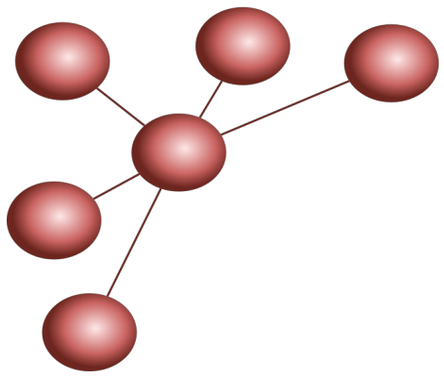


“guilt by association”

Global approaches



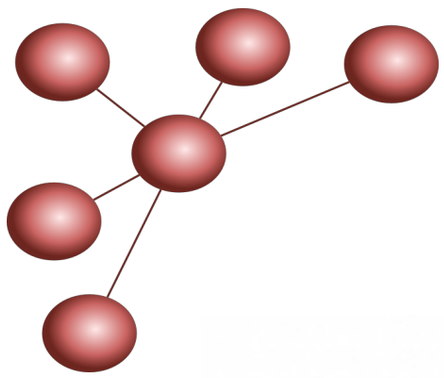
Topological features
Clustering / communities



Global approaches - topological measures

- Degree / degree distribution
- Size / diameter
- Clustering coefficient

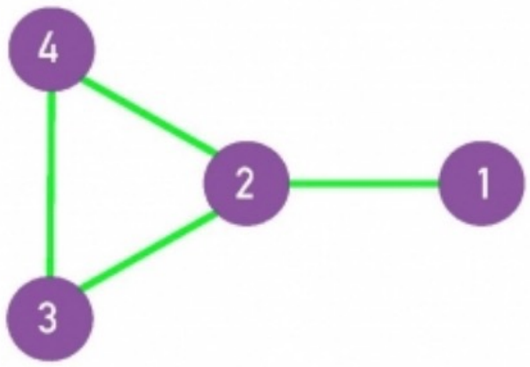
Degree distribution

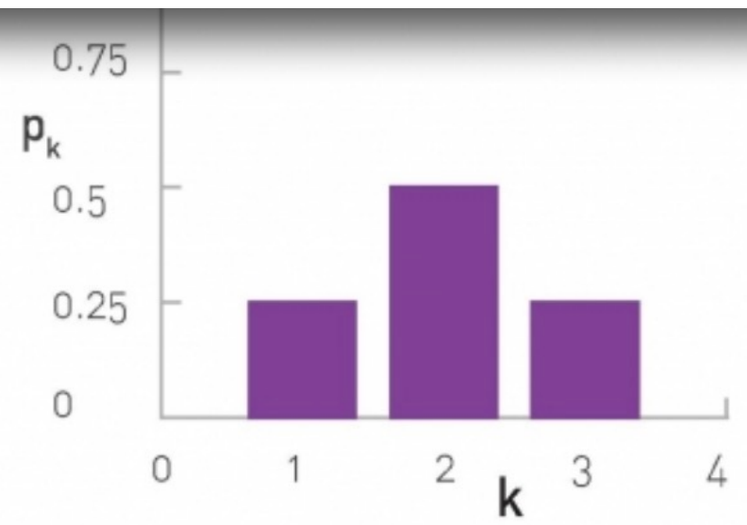


NetworkAnalyzer

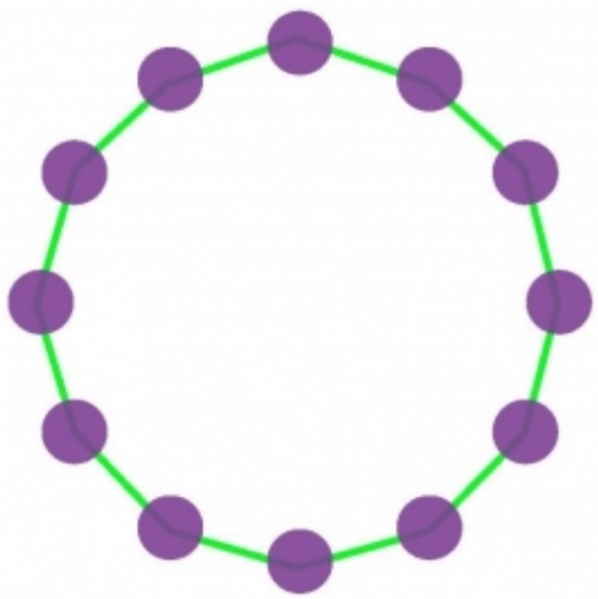
Core App: Computes basic properties of whole network (degree distribution, clustering coefficients, centrality, etc.)

a.

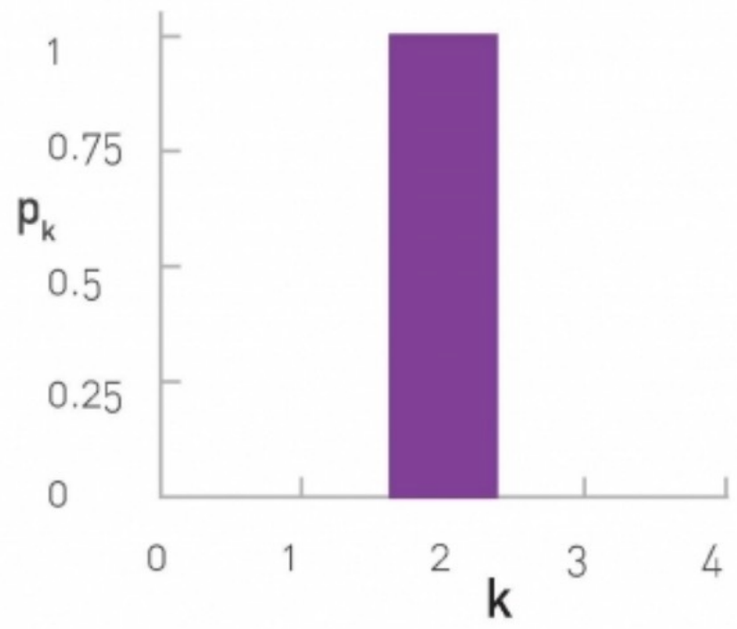


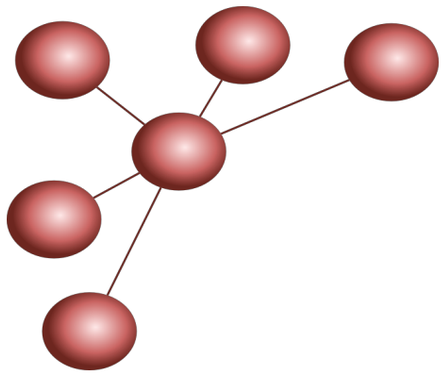


c.



d.

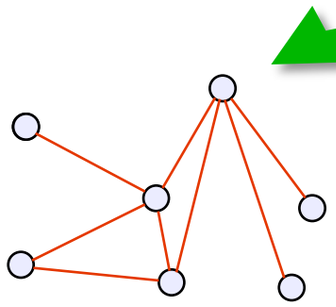




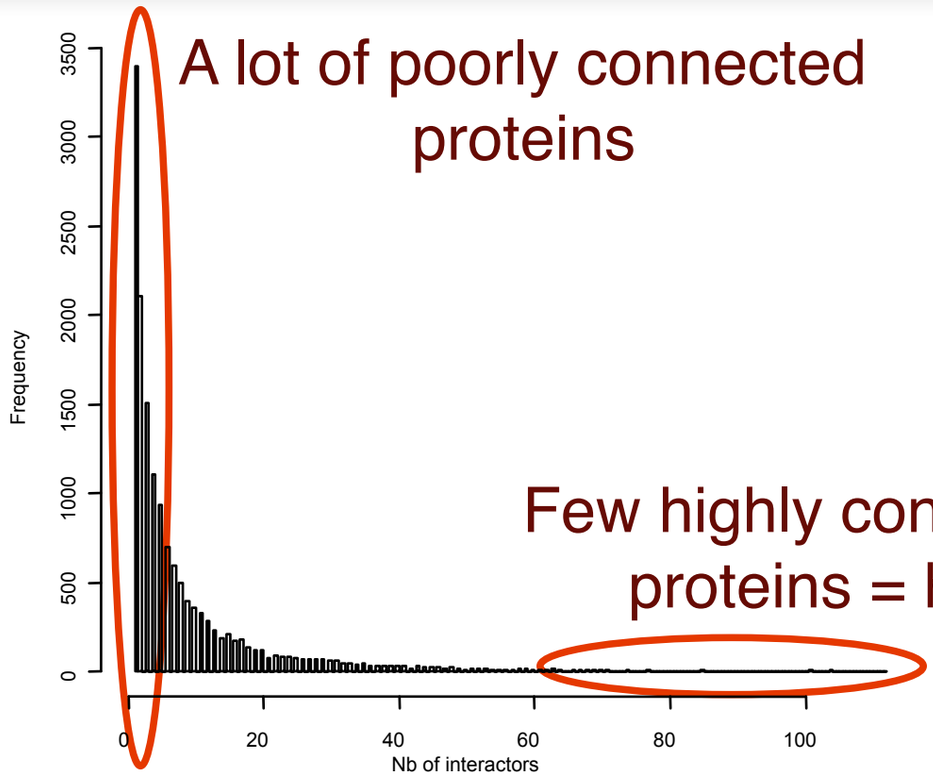
Protein degree distribution : interactomes are scale-free and small-world

NetworkAnalyzer

Core App: Computes basic properties of whole network (degree distribution, clustering coefficients, centrality, etc.)



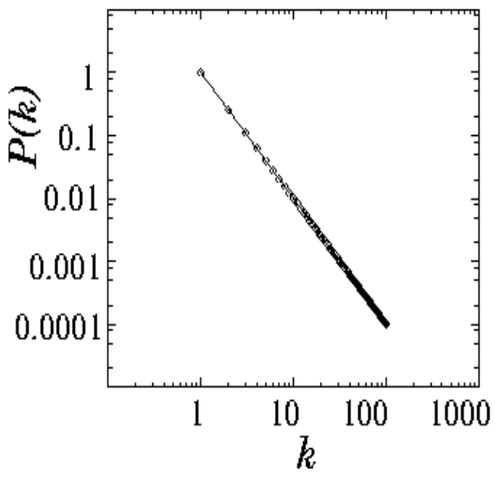
k = 4



A lot of poorly connected proteins

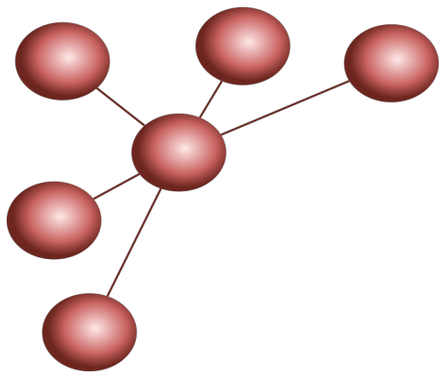
Few highly connected proteins = hub

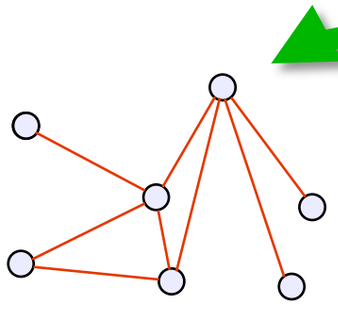
Power-law distribution



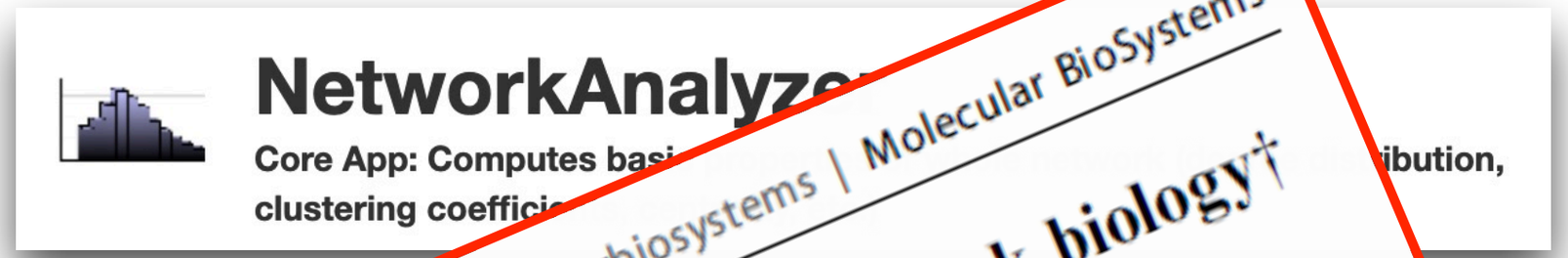
Biological interpretation?
Robust to random attack, sensitive to targeted attacks
Growth with preferential attachement (“rich get richer”) => create “hubs”

Protein degree distribution : interactomes are scale-free and small world

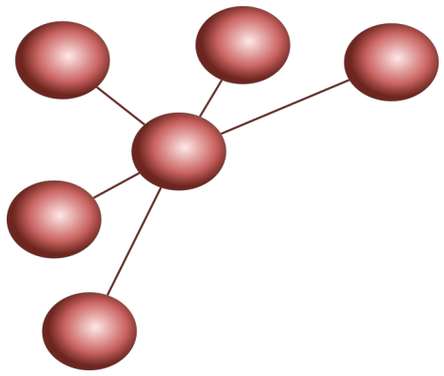




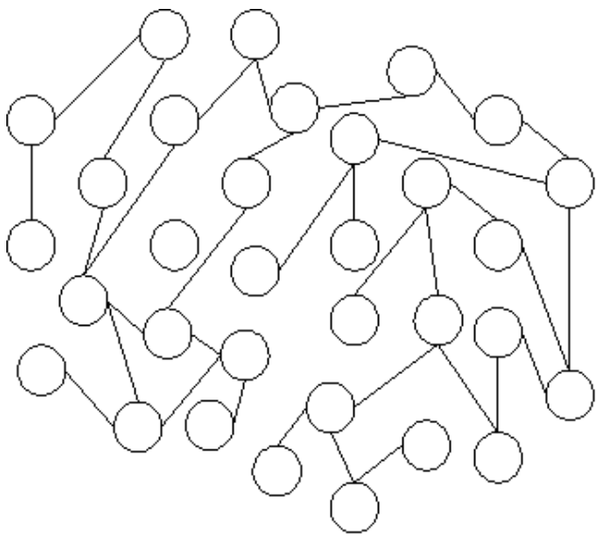
k = 4



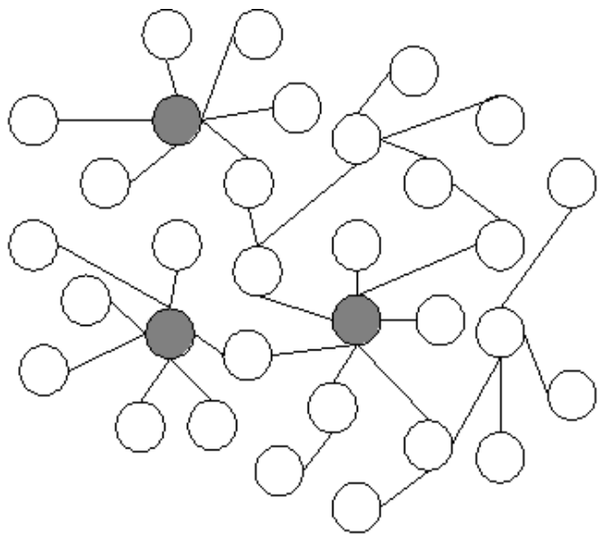
Biological interpretation?
Robust to random attack, sensitive to targeted attacks
Growth with preferential attachement (“rich get richer”) => create “hubs”



Network topological structure : Small-world property

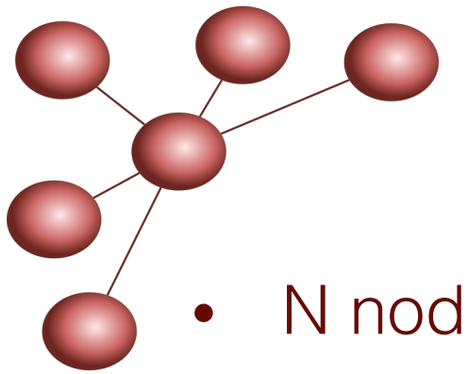


(a) Random network



(b) Scale-free network

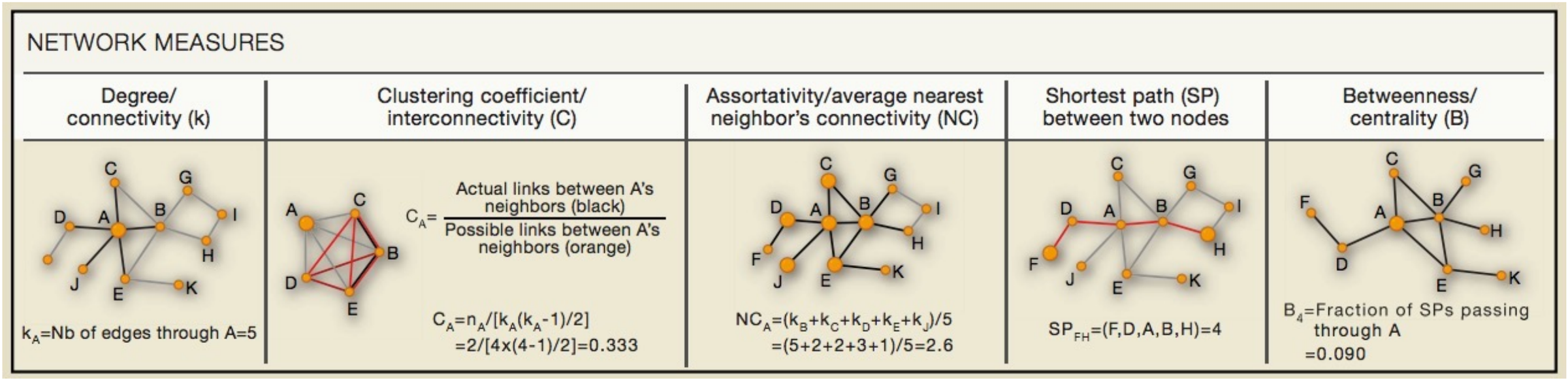
- Milgram, 6 degrees of separation



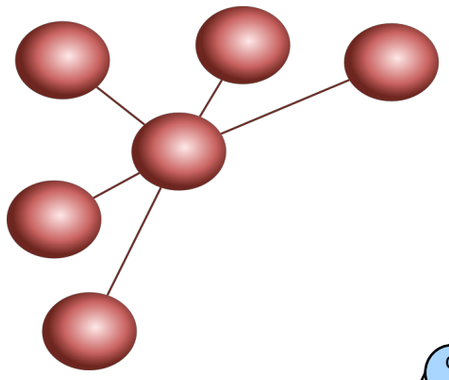
Metrics on graphs

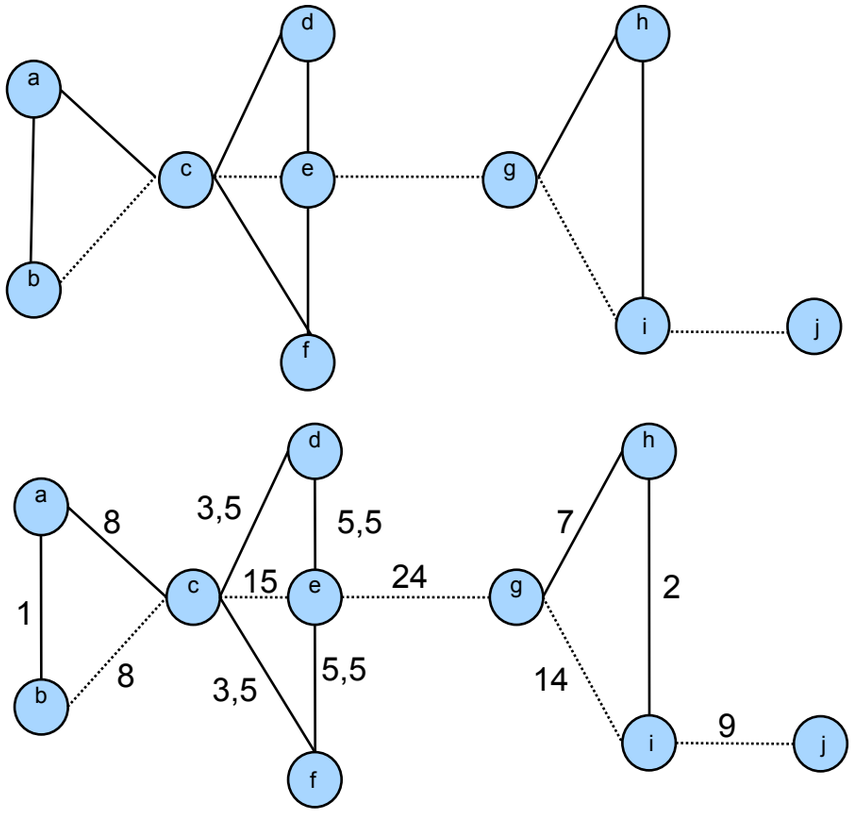
- N nodes, V edges
- Network size
- Adjacency matrix
- Degree, degree distribution
- Path, shortest path, distances
- Connectivity, clustering coefficient

- Betweenness
- Motifs



Gavin, Cell



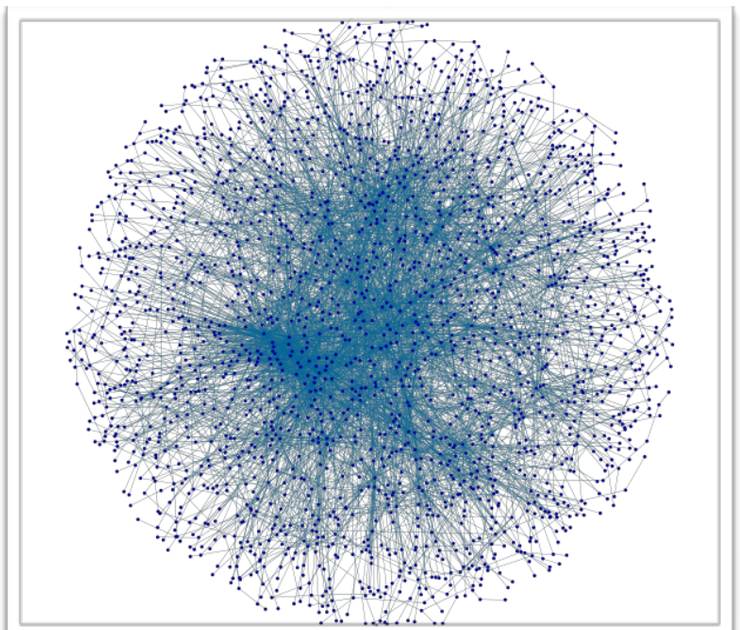


“Betweenness”

Number of shortest paths running through an edge

=

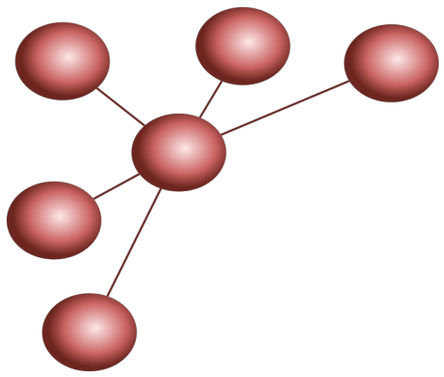
“bootleneck”



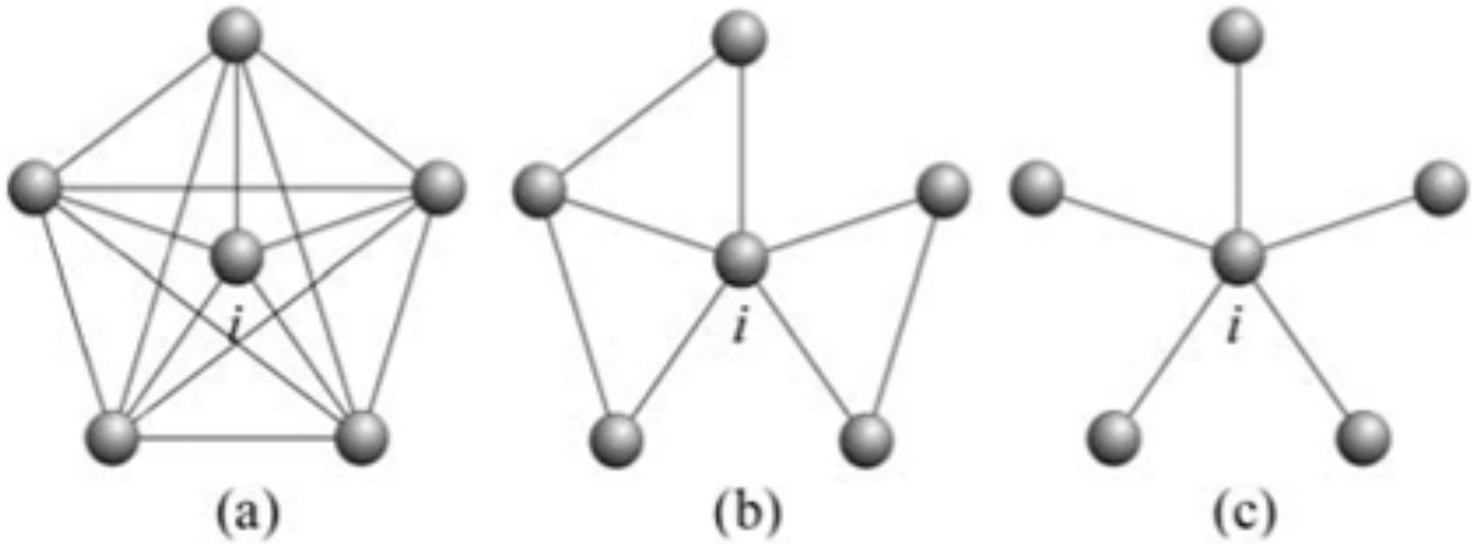
Biological interpretation ?
Correlation with gene essentiality, gene involvement in diseases, importance in flux transmission …

NetworkAnalyzer

Core App: Computes basic properties of whole network (degree distribution, clustering coefficients, centrality, etc.)



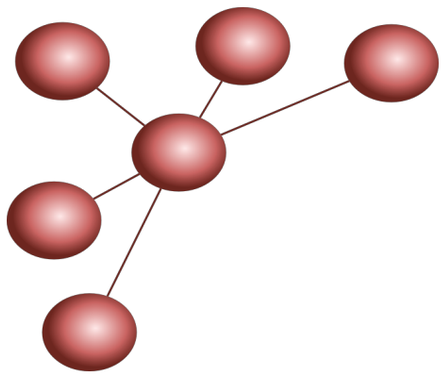
Clustering coefficient / modularity



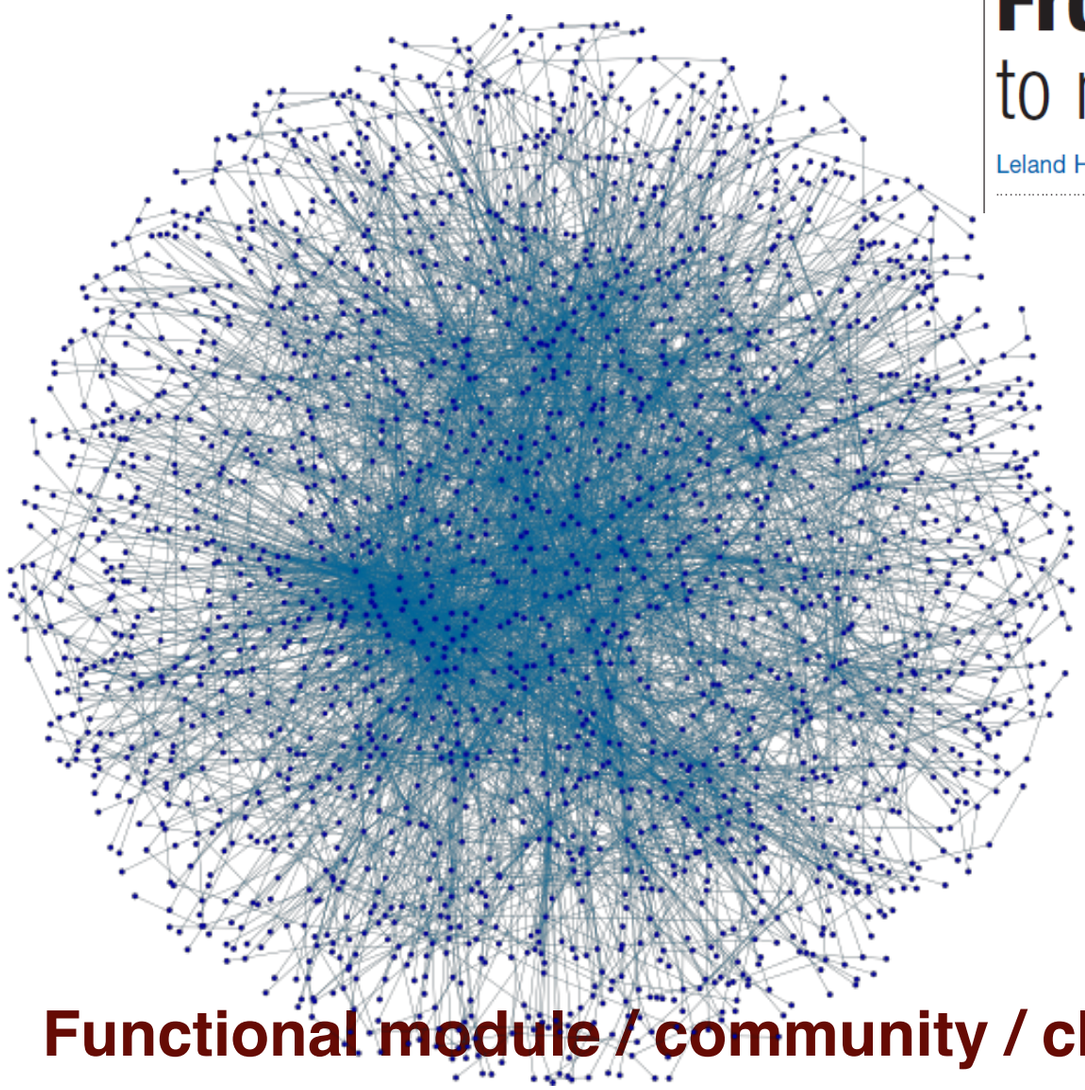
Actual links between neighbours / Possible links between neighbours

NetworkAnalyzer

Core App: Computes basic properties of whole network (degree distribution, clustering coefficients, centrality, etc.)



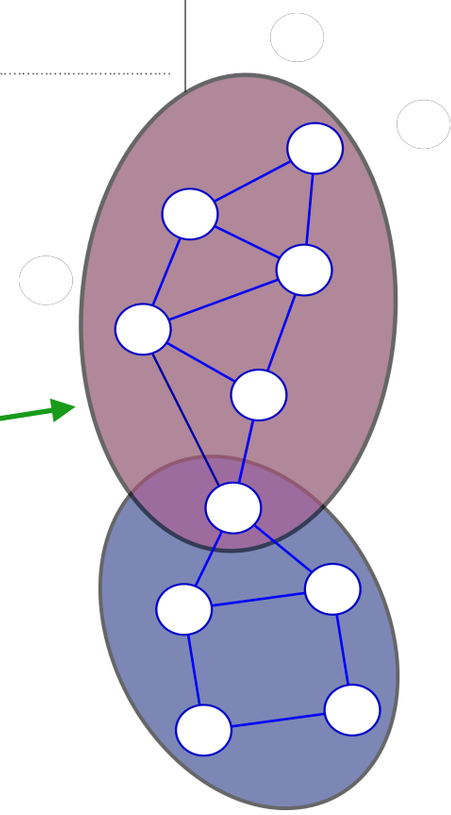
Global approach - Clustering



impacts

From molecular
to modular cell biology

Leland H. Hartwell, John J. Hopfield, Stanislas Leibler and Andrew W. Murray



Functional module / community / cluster / class : discrete function

Modules can be isolated or connected

Groups of proteins involved in a common cellular function

MCODE

Clusters a given network based on topology to find densely connected regions.