

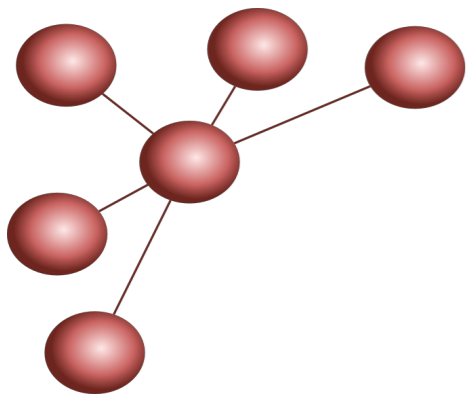
DUBii 2019

# **Introduction to Network Science**

# **Introduction to Network Biology**

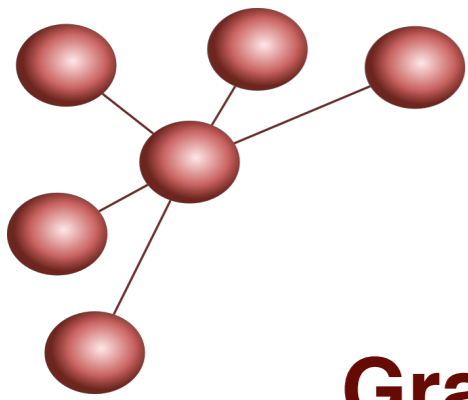
Anaïs Baudot [anais.baudot@univ-amu.fr](mailto:anais.baudot@univ-amu.fr)

Costas Bouyioukos [costas.bouyioukos@univ-paris-diderot.fr](mailto:costas.bouyioukos@univ-paris-diderot.fr)



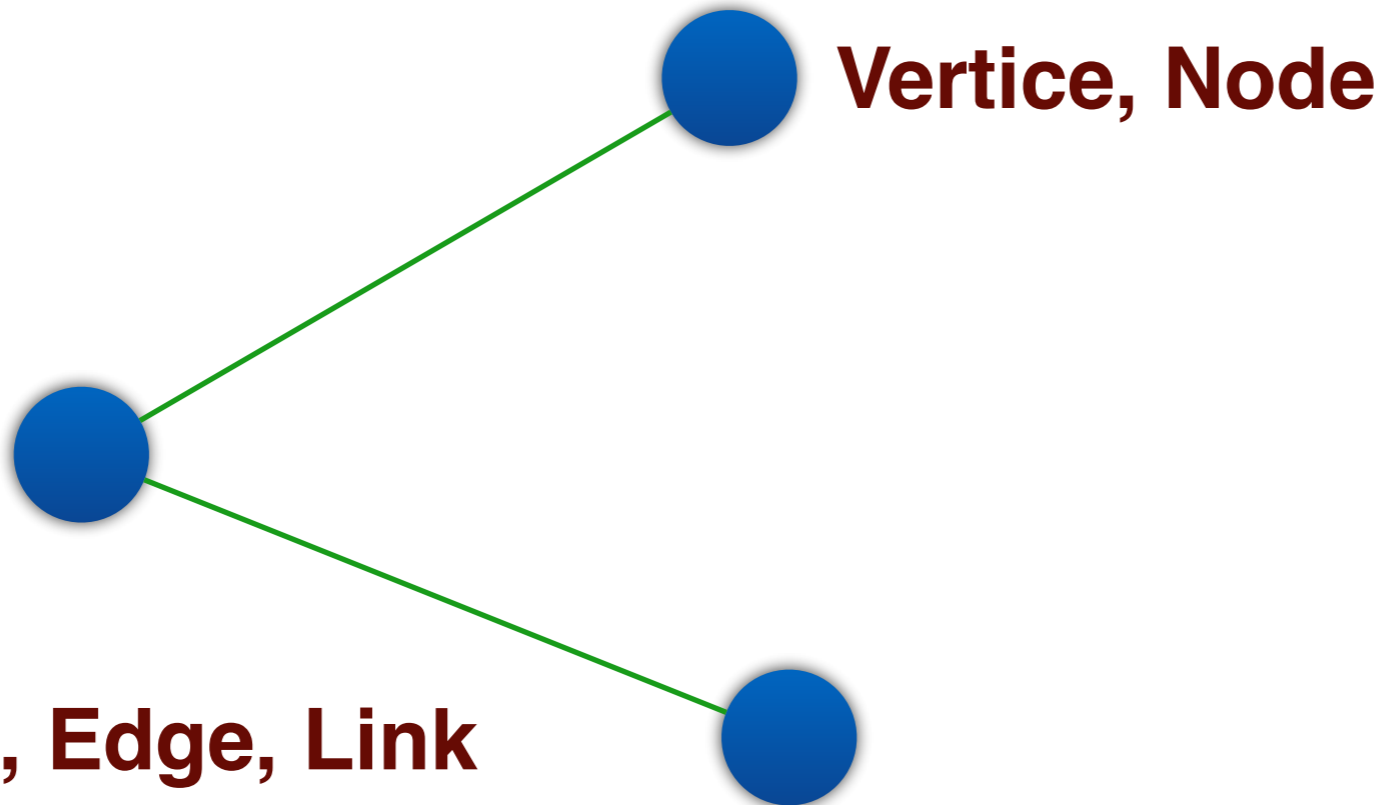
# Network Science

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



**Graph, Network, Web**

# Definition

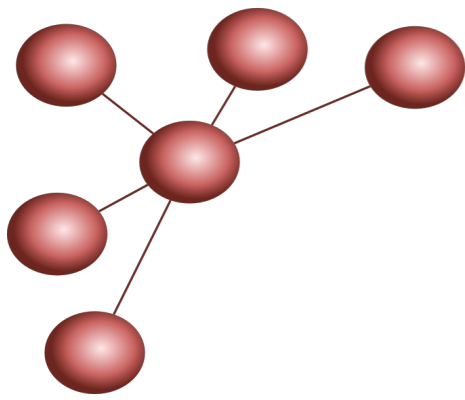


**Vertice, Node**

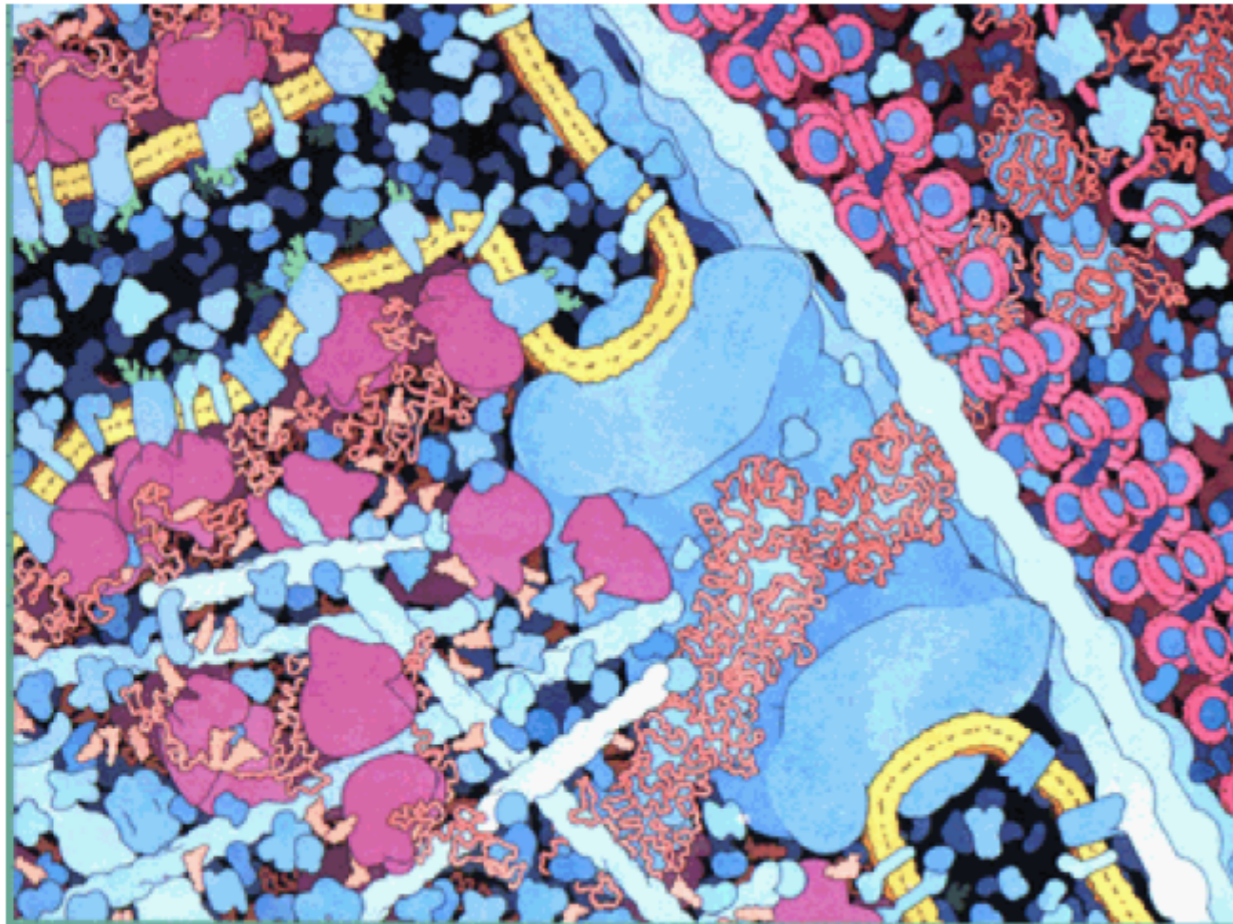
**Interaction, Edge, Link**

$$G = \{V, E\}$$

**Topology, motifs**



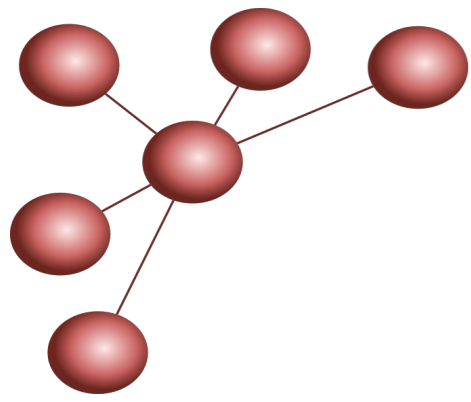
**Proteins do not act isolated but  
interact with each other to  
perform their functions**



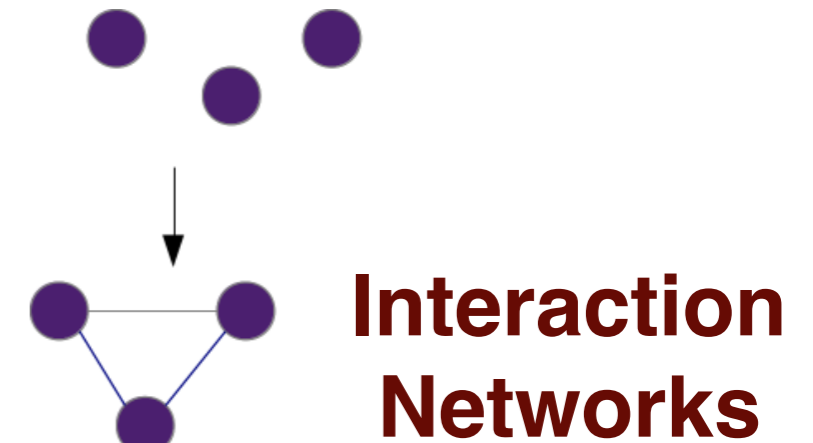
**Molecular interactions**  
**Protein-DNA**  
**Protein-RNA**  
**Protein-protein**  
**Protein-lipid**

...

# Systems Biology because living organisms are complex systems



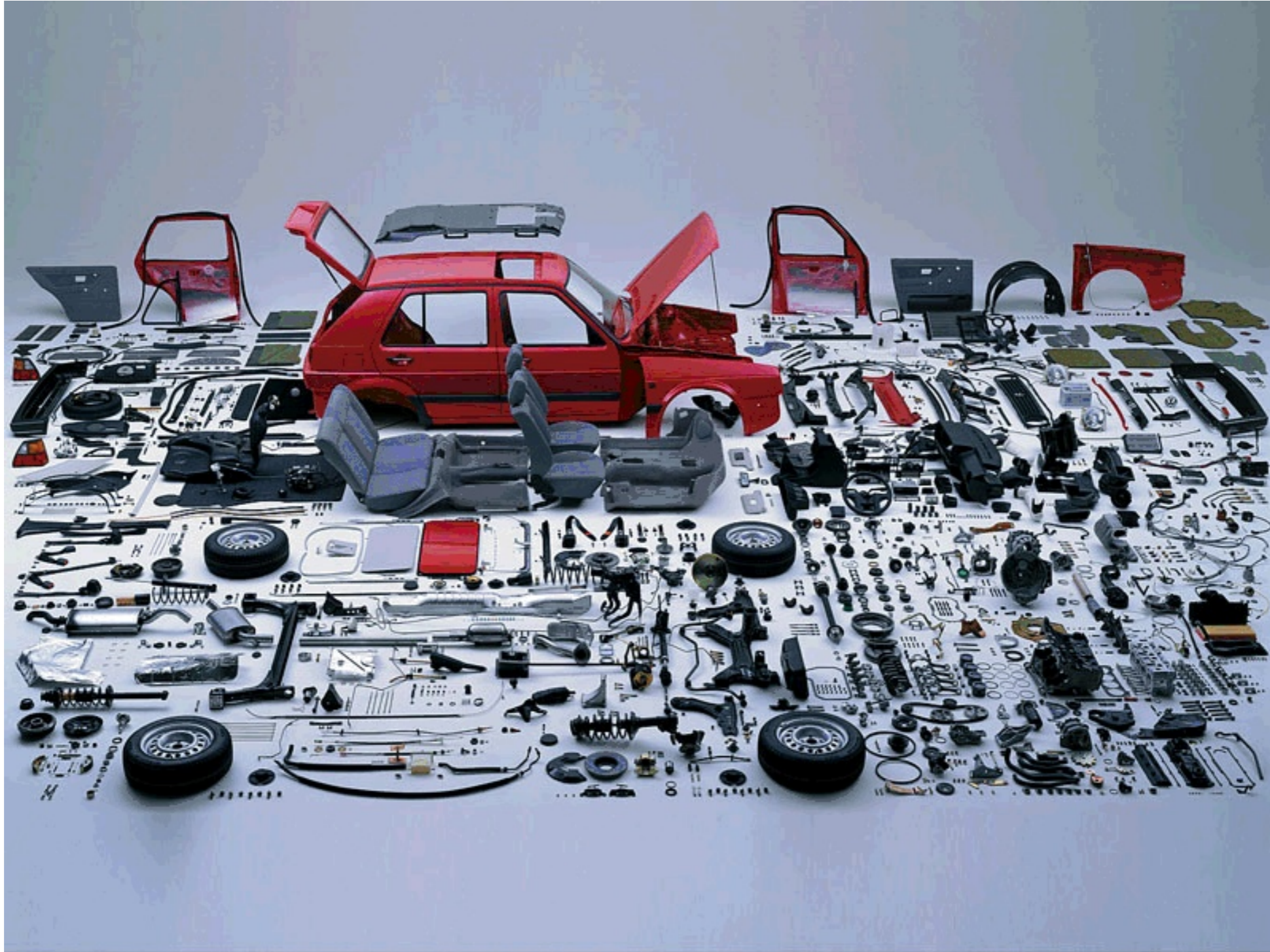
**Systems component : genes/  
proteins**

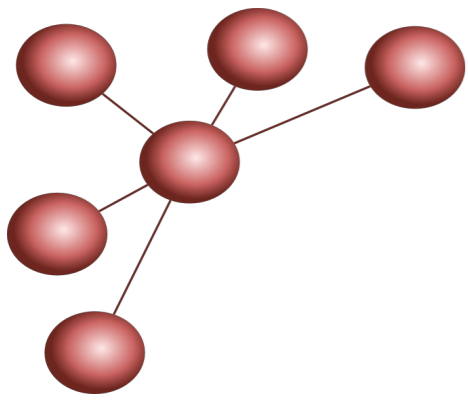


**Emerging properties : phenotypes**

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

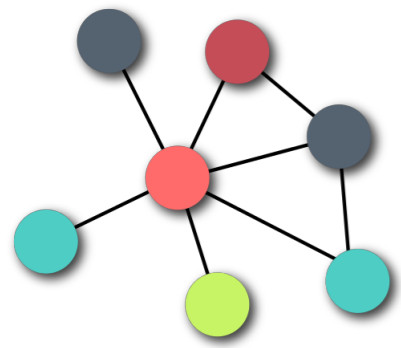
## Exemple :





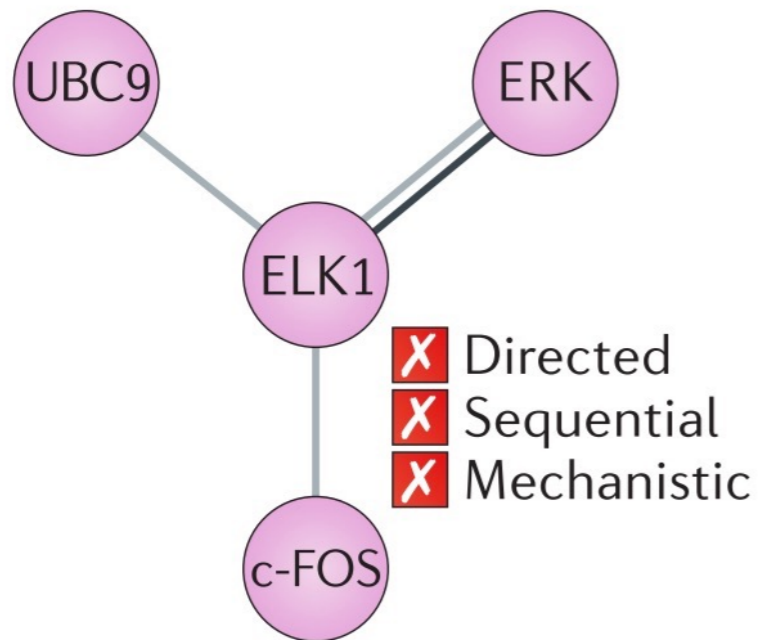
# Biological Networks

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

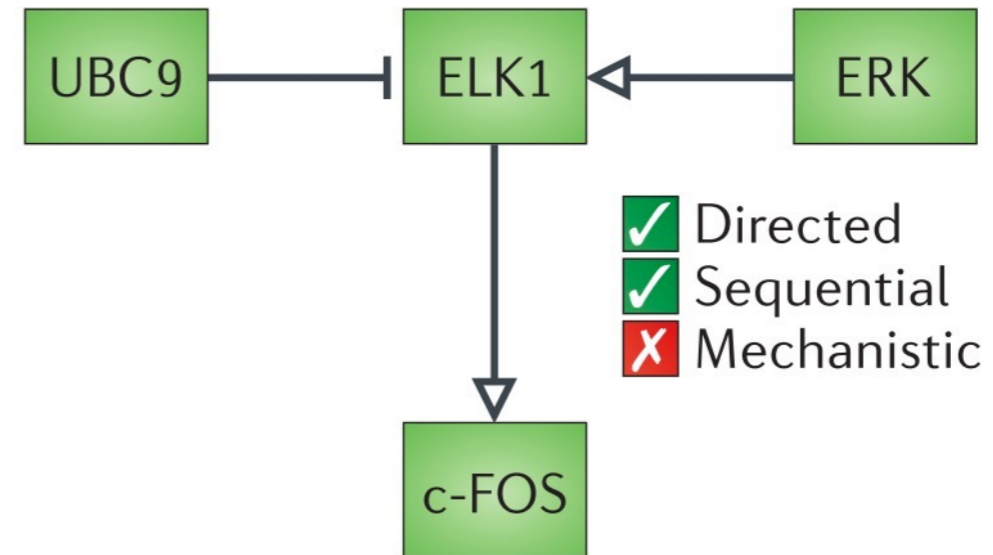


# 2 main types of networks to represent biological information flow

**a** Interaction network

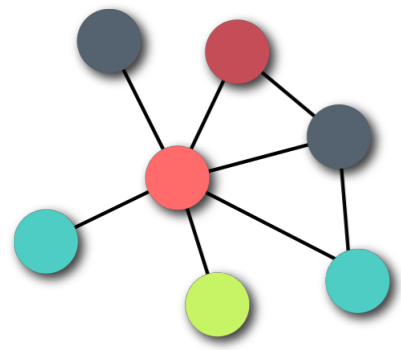


**b** Activity flows



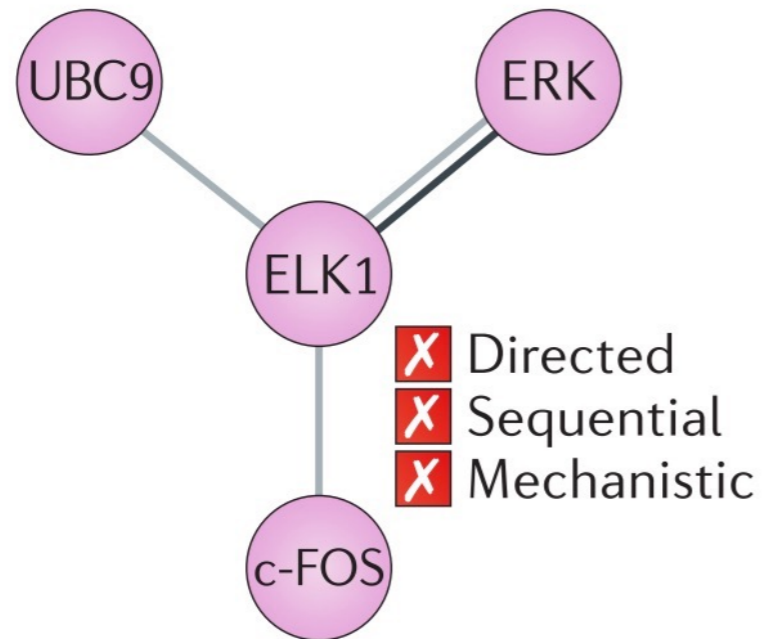
**Le Novère et al. 2015**





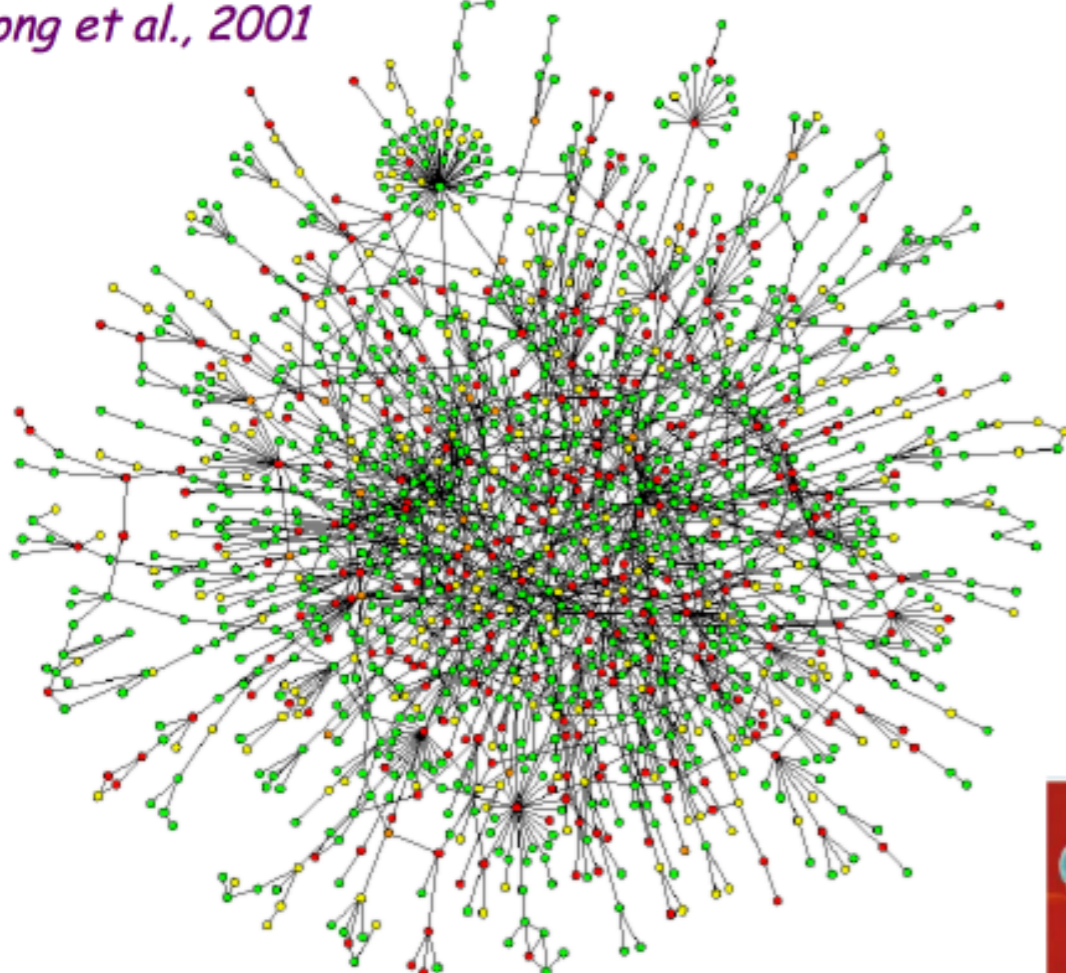
# Interaction networks / Interactome

**a** Interaction network

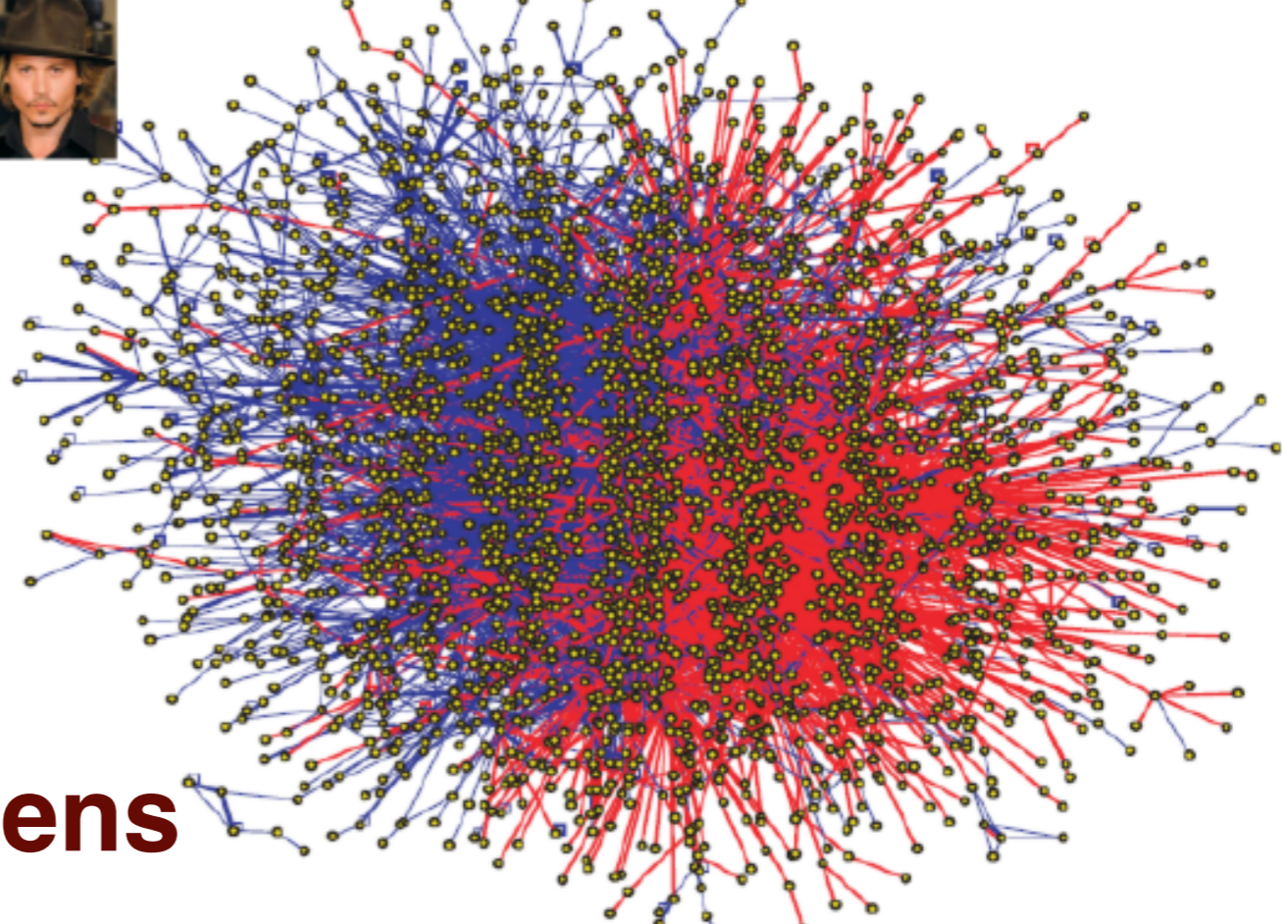
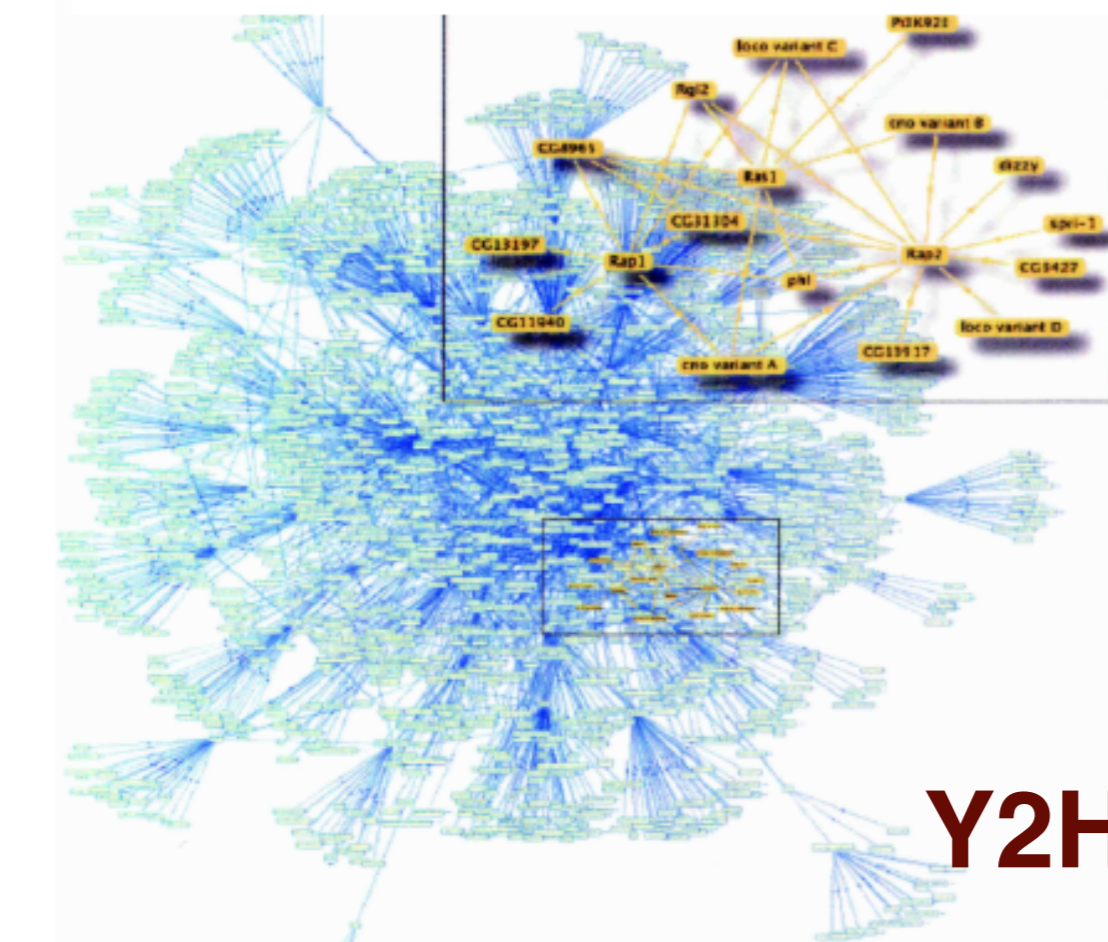
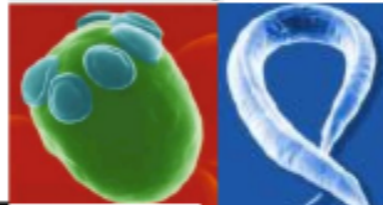
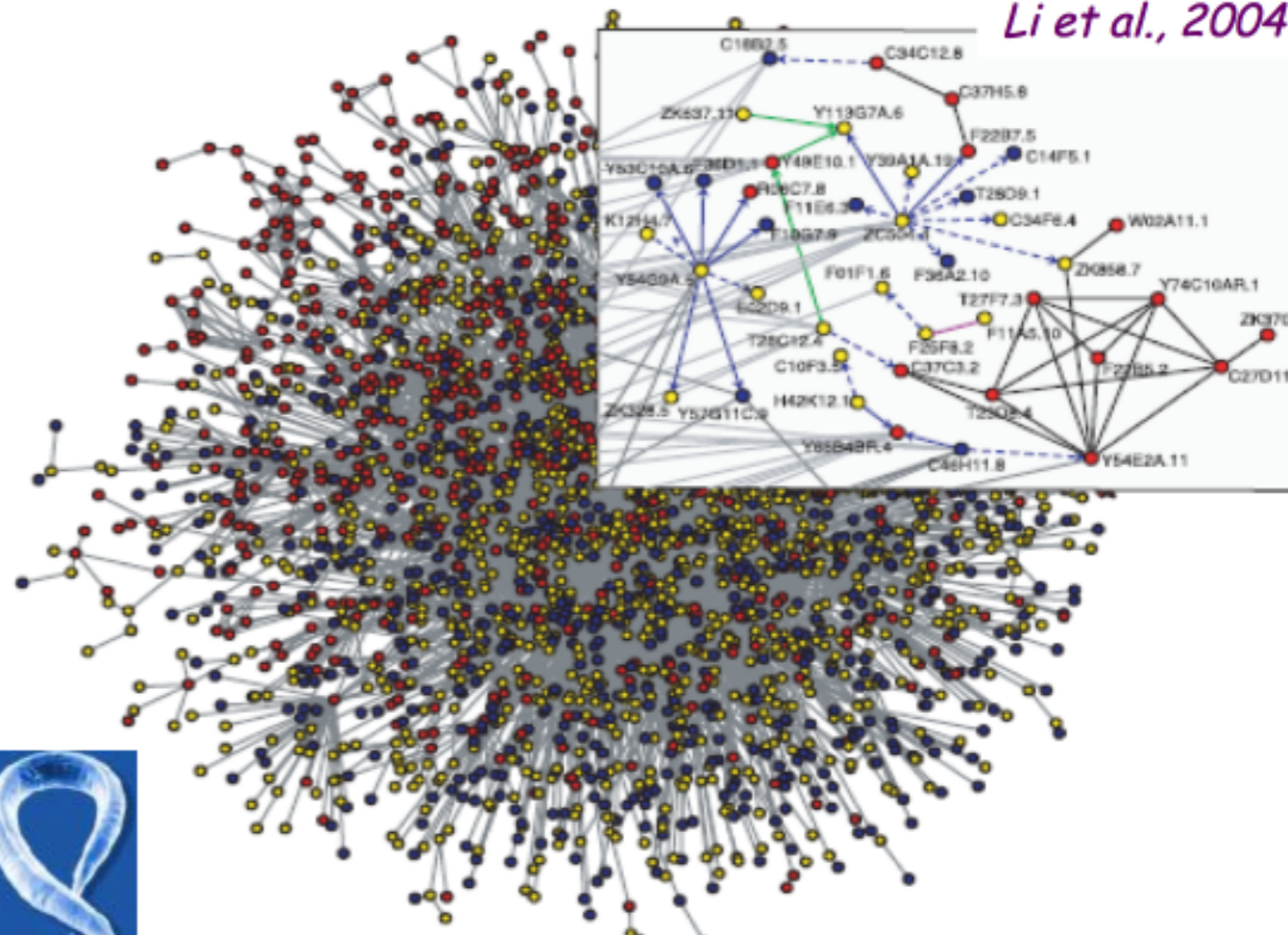


**Le Novère et al. 2015**

Jeong et al., 2001



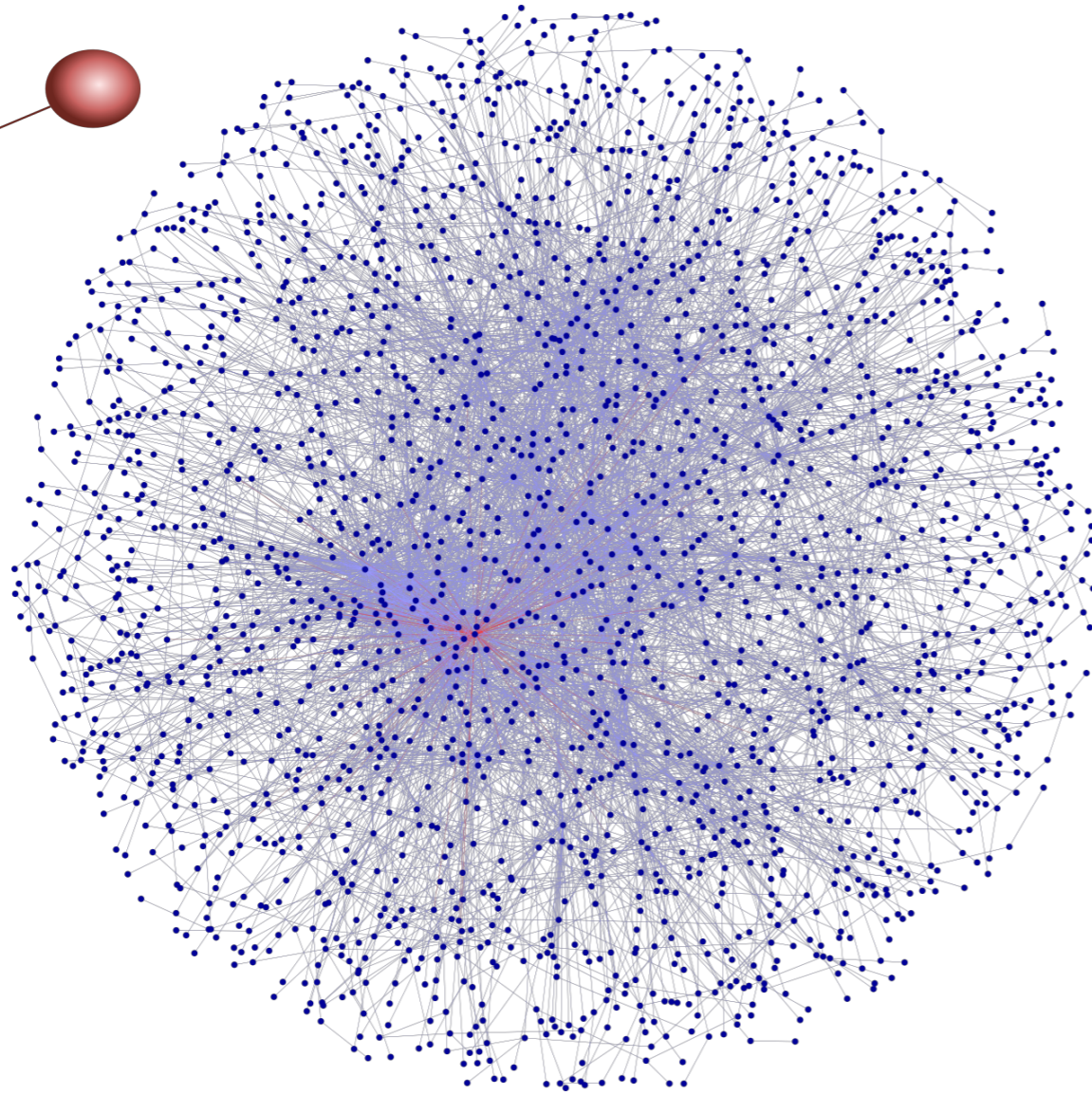
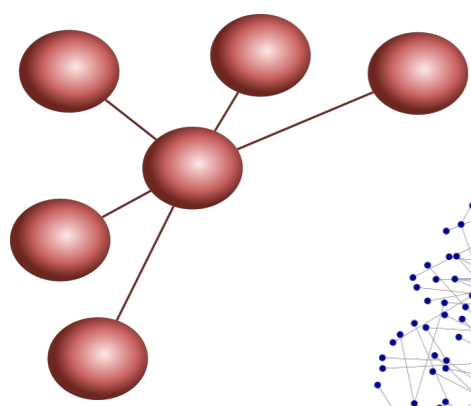
Li et al., 2004



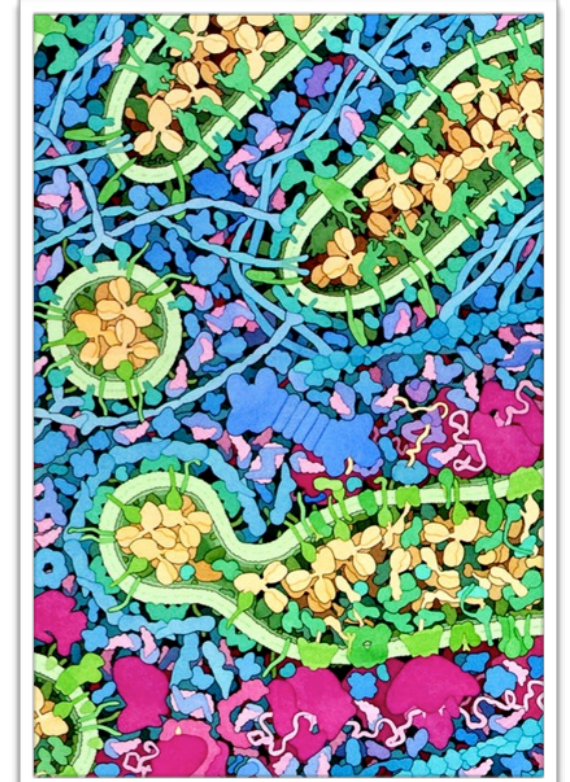
Formstecher et al., 2005

Rual et al., 2005

# Y2H screens



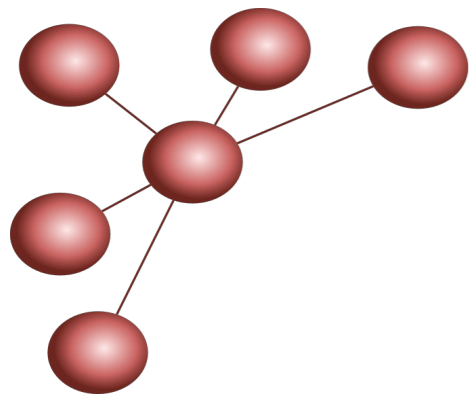
# Interactomes



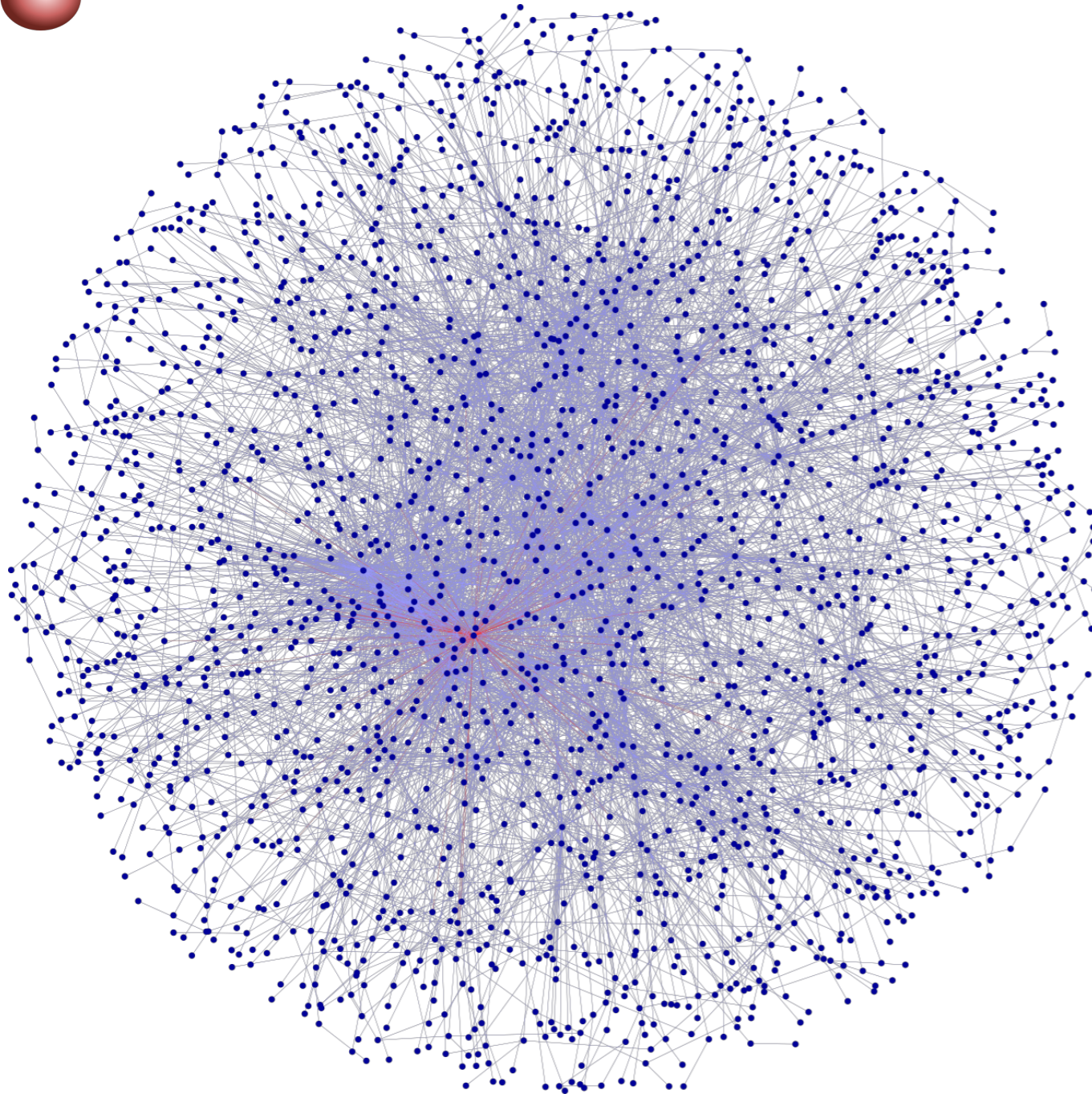
**Set of protein-protein interactions detected in an organism**

Physical interactions, but physiological interactions ?

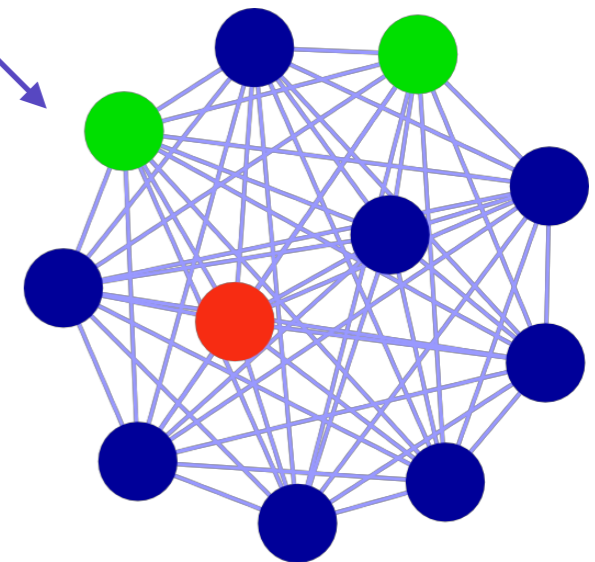
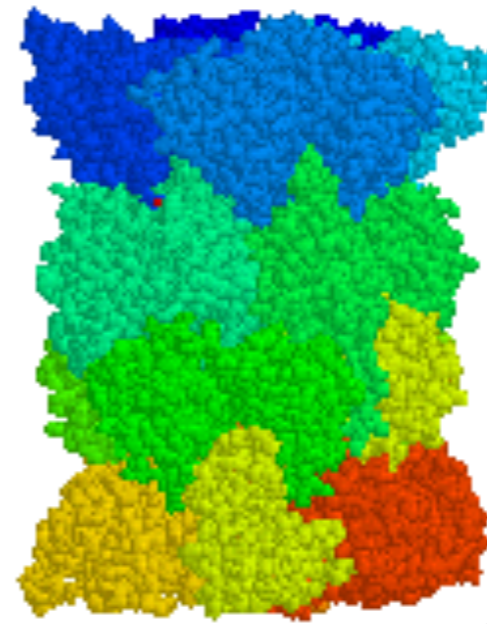
Interactomes are devoid of spatio-temporal information



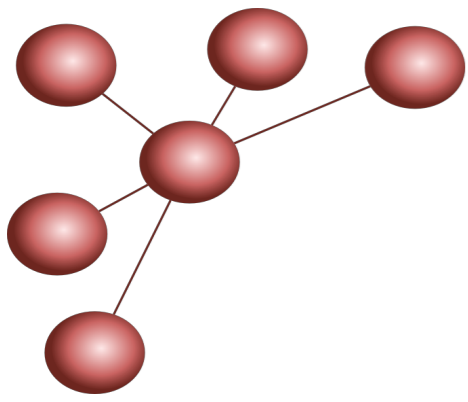
# Interactomes



**Protein-protein interaction networks**



**Protein complexes**



# Interaction databases

## Multi-organisms:

DIP ([dip.doe-mbi.ucla.edu](http://dip.doe-mbi.ucla.edu))

IntAct ([www.ebi.ac.uk/intact](http://www.ebi.ac.uk/intact))

MINT ([mint.bio.uniroma2.it/mint](http://mint.bio.uniroma2.it/mint))

BioGRID ([www.thebiogrid.org](http://www.thebiogrid.org))

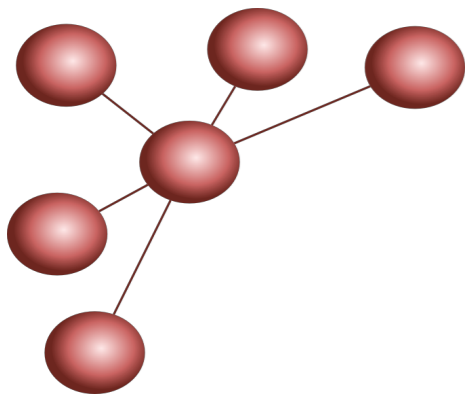
BIND ([www.blueprint.org](http://www.blueprint.org))

## Human


Reactome-FI



**International  
Molecular  
Exchange  
Consortium**



# Psicquic portal

EMBL-EBI  Services Research Training About us

## PSICQUIC View

BRCA2   
Examples: BRCA2, Q06609, dmc1, 10831611

Input Form **Browse** Help

[Input Form](#) > [Browse](#)

### 1,832 binary interactions found for search term *BRCA2*

<input type="checkbox"/> APID Interactomes	<input checked="" type="checkbox"/> BAR-6	<input checked="" type="checkbox"/> bhf-ucl-0	<input type="checkbox"/> BIND
<input checked="" type="checkbox"/> BindingDB-0	<input checked="" type="checkbox"/> BioGrid-322	<input checked="" type="checkbox"/> ChEMBL-0	<input type="checkbox"/> DIP
<input type="checkbox"/> DIP-IMEx	<input type="checkbox"/> DrugBank	<input checked="" type="checkbox"/> EBI-GOA-miRNA-0	<input checked="" type="checkbox"/> EBI-GOA-nonIntAct-65
<input type="checkbox"/> GeneMANIA	<input checked="" type="checkbox"/> HPIDb-0	<input checked="" type="checkbox"/> I2D-0	<input checked="" type="checkbox"/> IMEx-241
<input checked="" type="checkbox"/> InnateDB-0	<input checked="" type="checkbox"/> InnateDB-All-561	<input checked="" type="checkbox"/> IntAct-107	<input type="checkbox"/> Interoprc
<input type="checkbox"/> iRefIndex	<input checked="" type="checkbox"/> MatrixDB-12	<input checked="" type="checkbox"/> MBInfo-0	<input checked="" type="checkbox"/> mentha-380
<input checked="" type="checkbox"/> MINT-84	<input checked="" type="checkbox"/> MPIDB-0	<input checked="" type="checkbox"/> Reactome-0	<input checked="" type="checkbox"/> Reactome-FIs-29
<input type="checkbox"/> Spike	<input type="checkbox"/> TopFind	<input checked="" type="checkbox"/> UniProt-25	<input type="checkbox"/> VirHostNet
<input type="checkbox"/> ZINC			

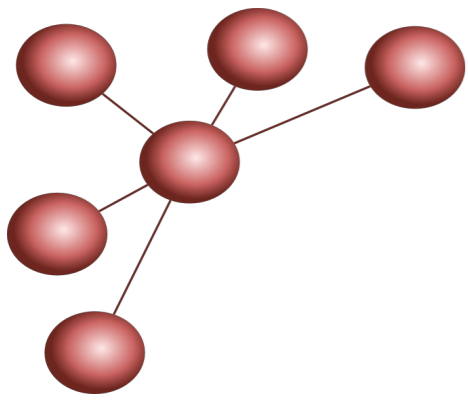
#### Status of the service

- ONLINE
- OFFLINE
- WARNING: Time out
- ERROR: Unexpected Error

1,832 selected interactions

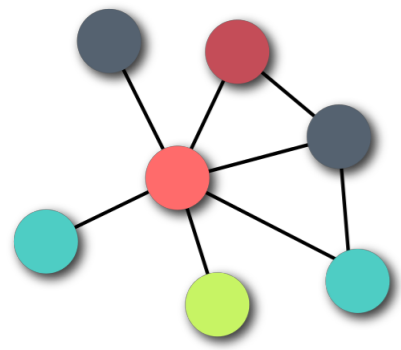
version: 1.4.11

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



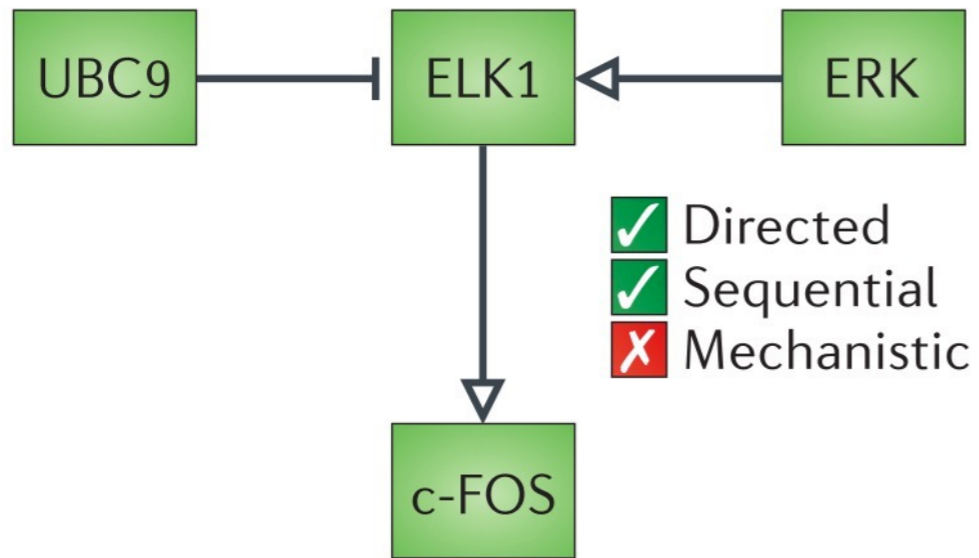
# Network inference from -omics data

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



# Activity flow / Gene Regulatory Networks / Influence Graphs

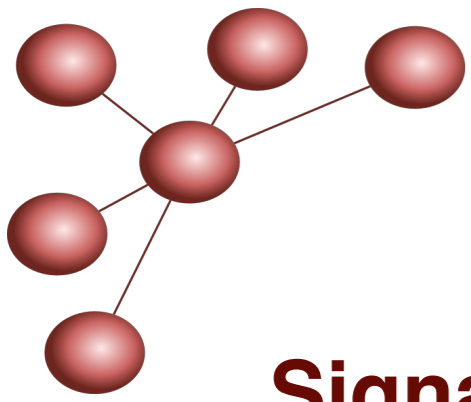
## b Activity flows



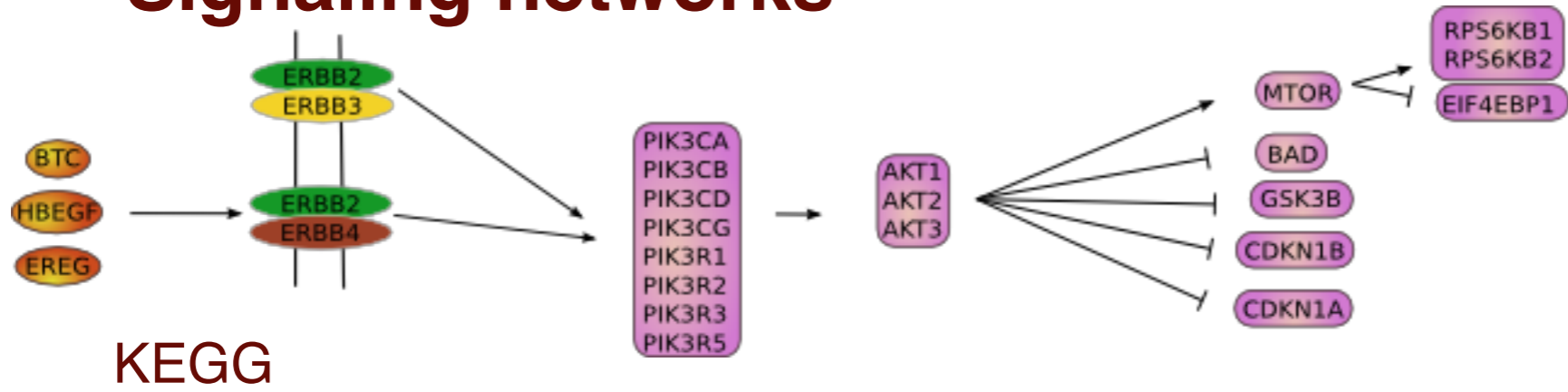
**Le Novère et al. 2015**

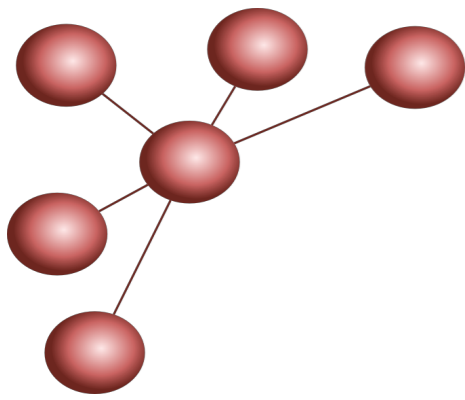


# Signaling Networks



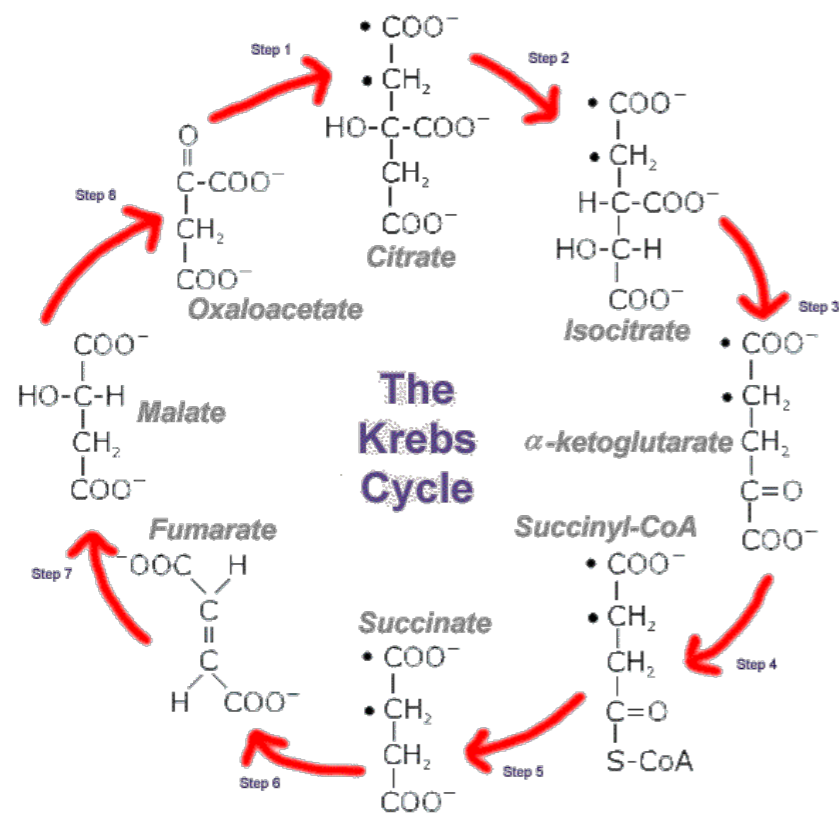
## Signaling networks



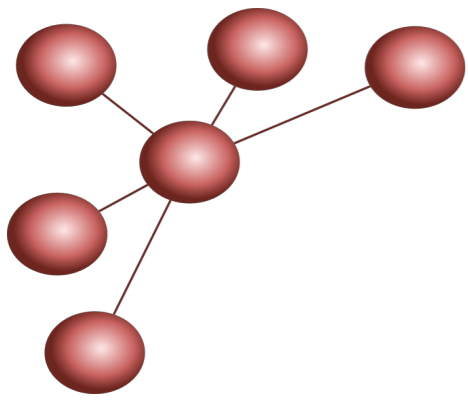


# Metabolic networks

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

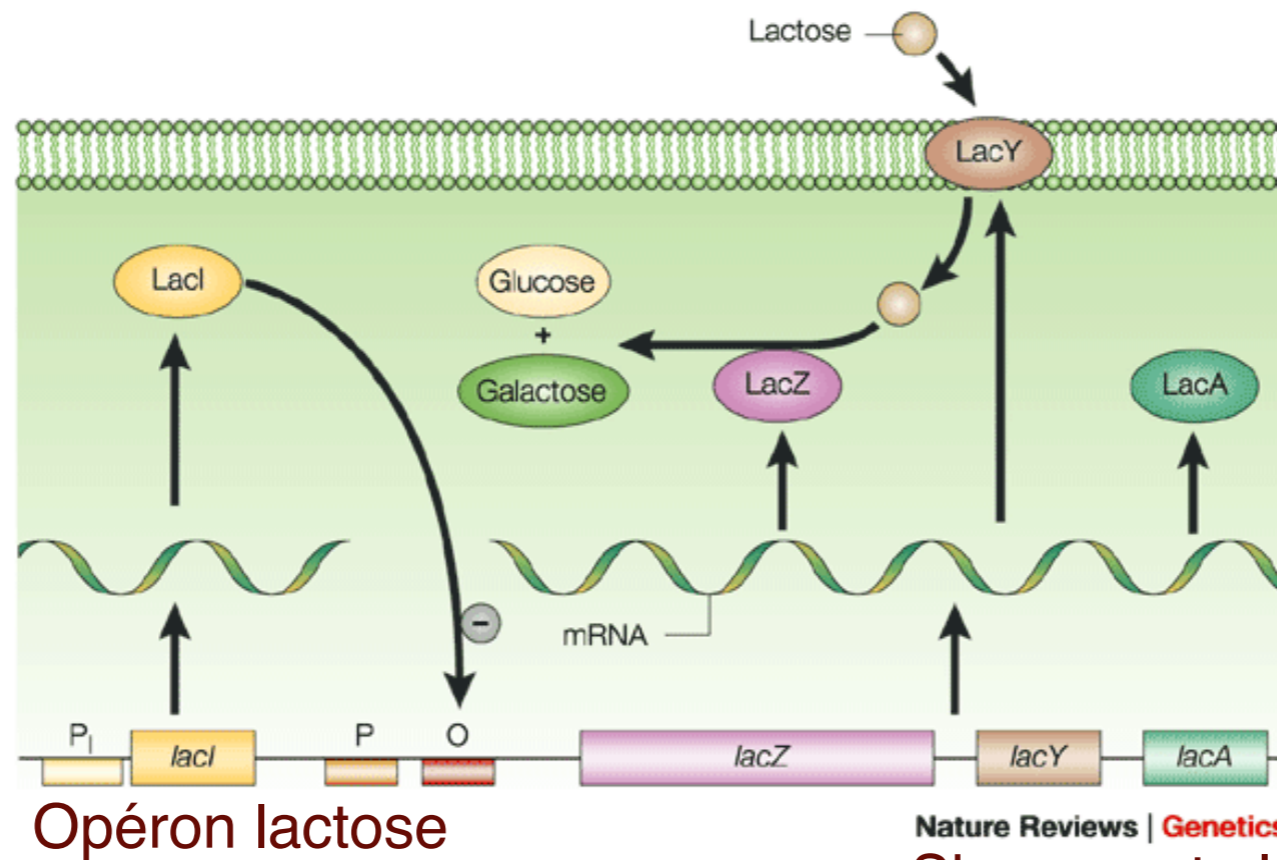


**Metabolic Cycle**



# Gene Regulatory Networks

## Operon / Regulatory networks



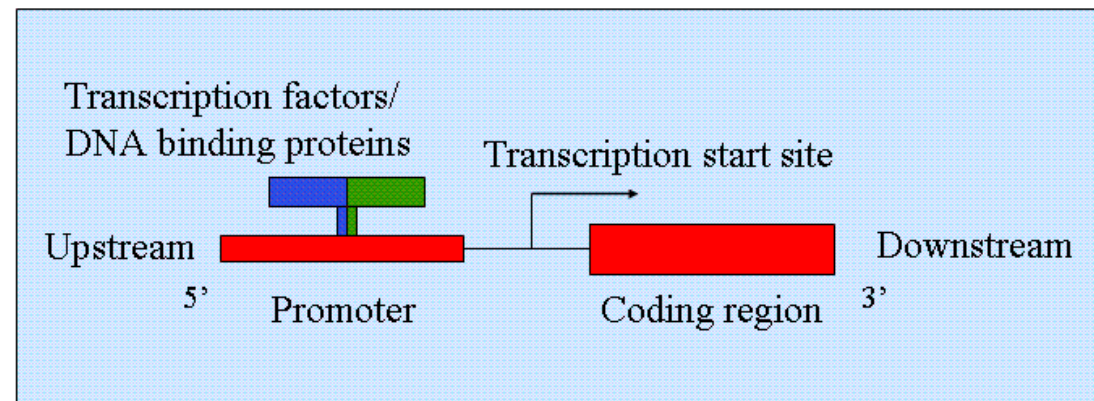
Nature Reviews | Genetics  
Shuman et al.



# Refreshing Gene Regulation

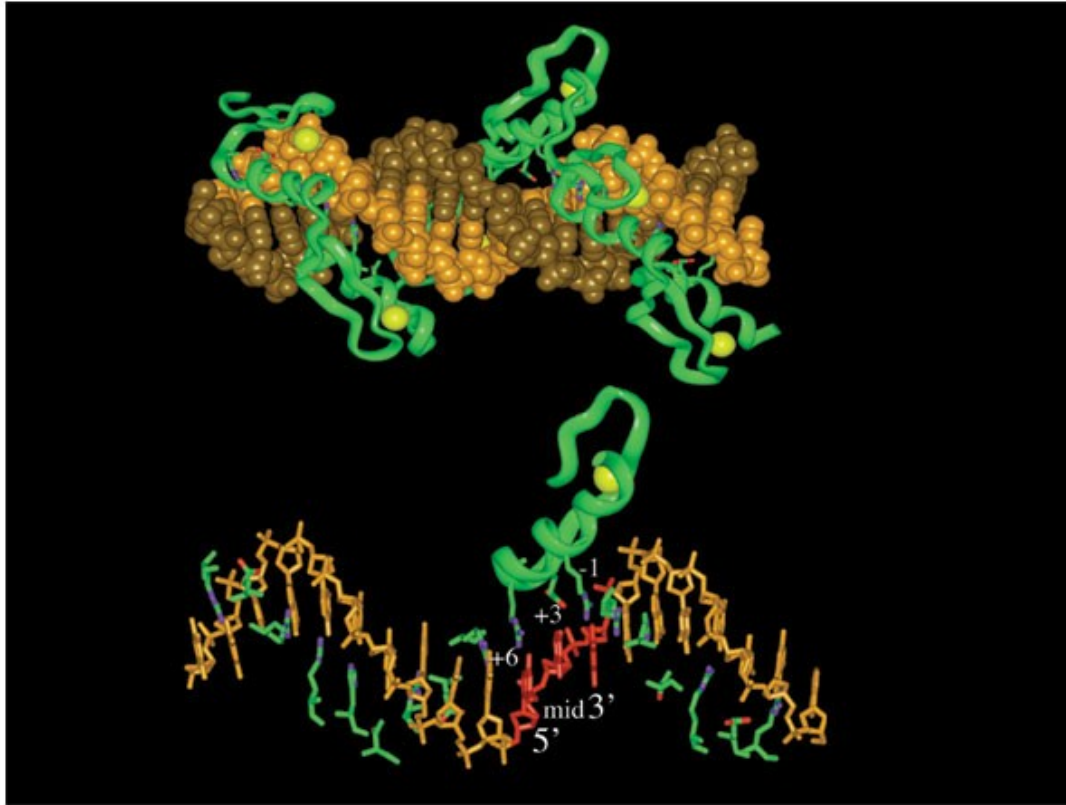
- A gene consists of a coding region AND a regulatory region
- There are special “sites”, called Transcription Factor Binding Sites (TFBS)
- TFBS are short DNA sequences, with specific nucleotide composition, recognised by a TF.

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

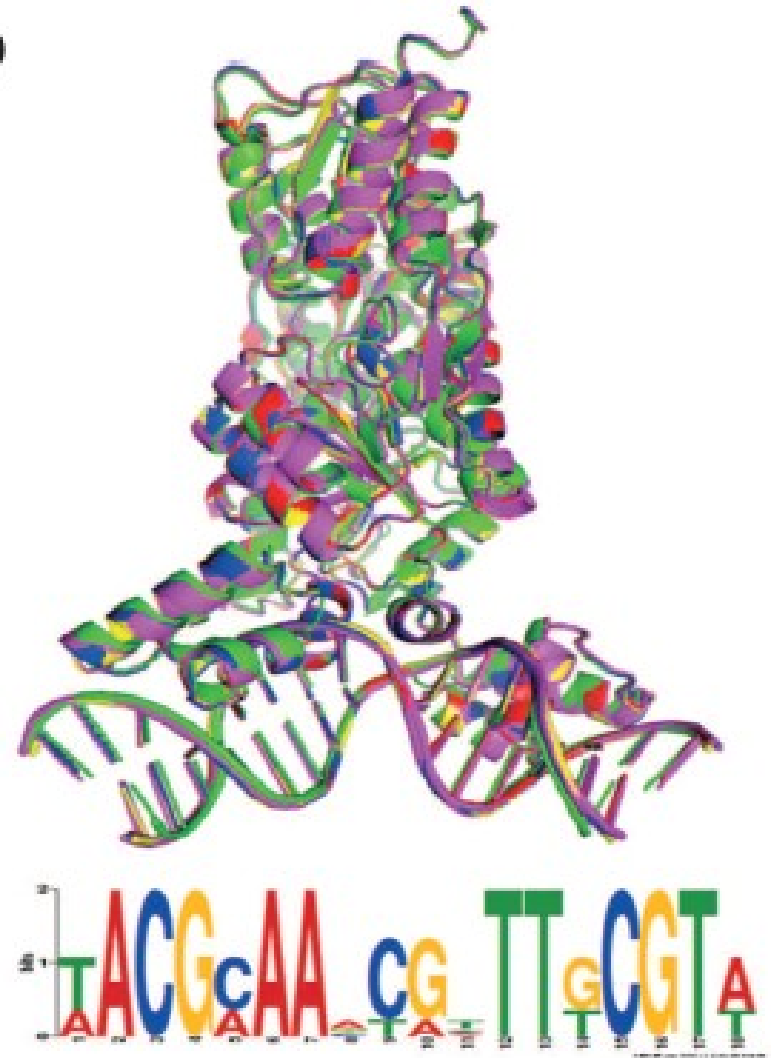


- Transcription Factors (TFs) are special type of proteins which bind to DNA.
- TFs “recognise” via their structure (i.e the amino-acid sequence) only a specific sequence on the DNA major groove.
- The “strength” of the binding (i.e. affinity) depends on the TFBS.

# What are TFBSs (and motifs)

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

(a)



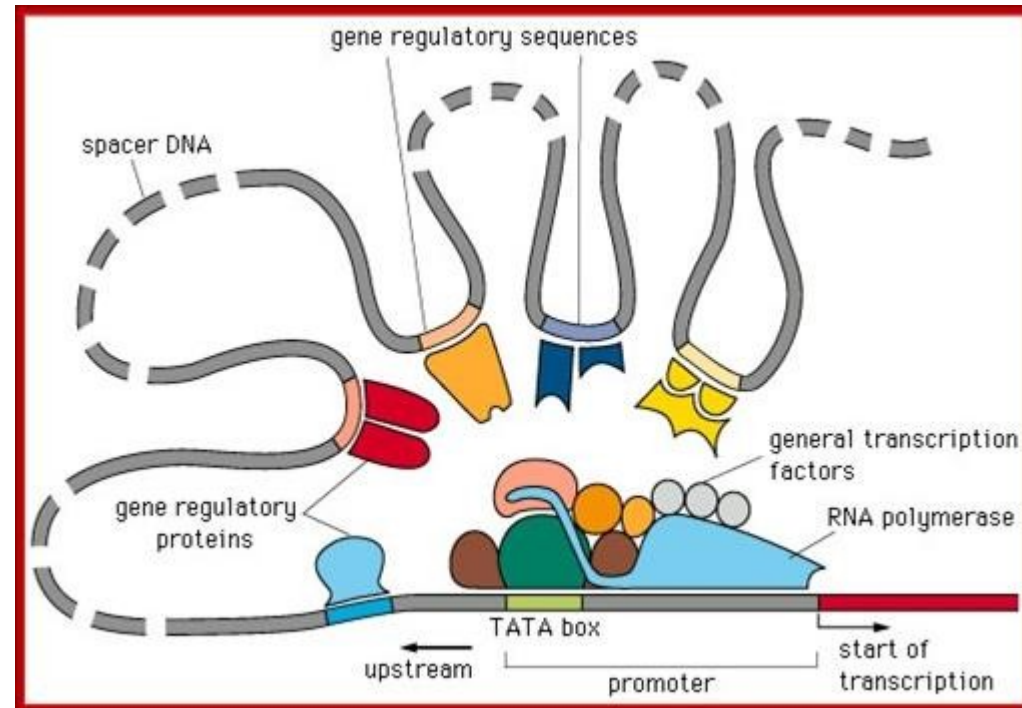
# How all this works together

- 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 proteins, which are encoded by genes which are also regulated by...

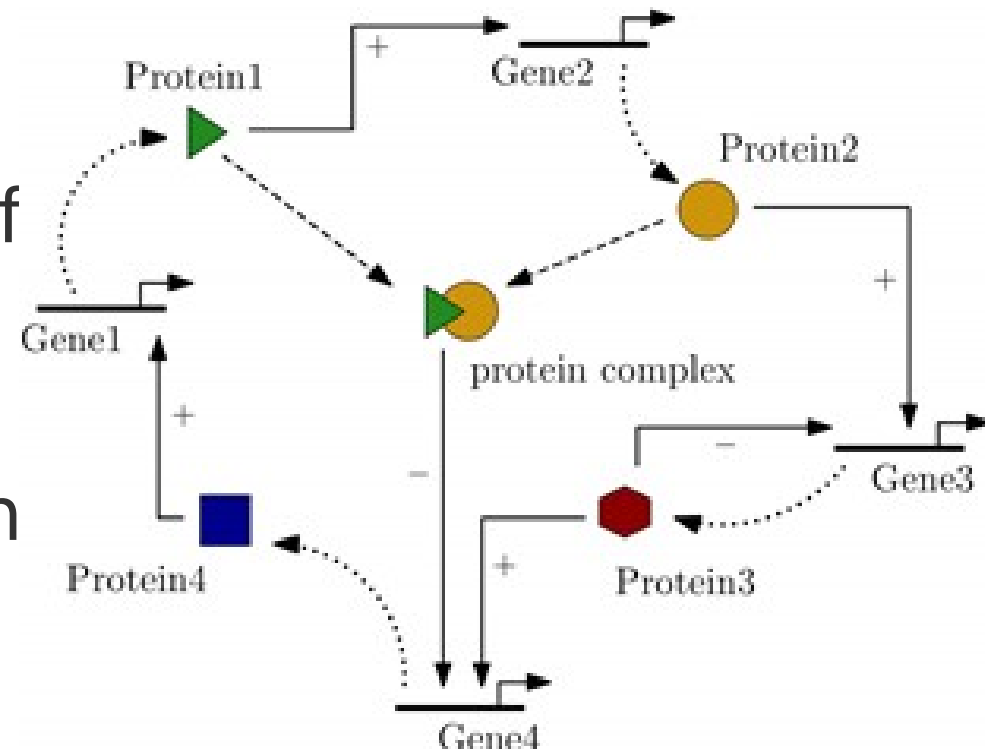




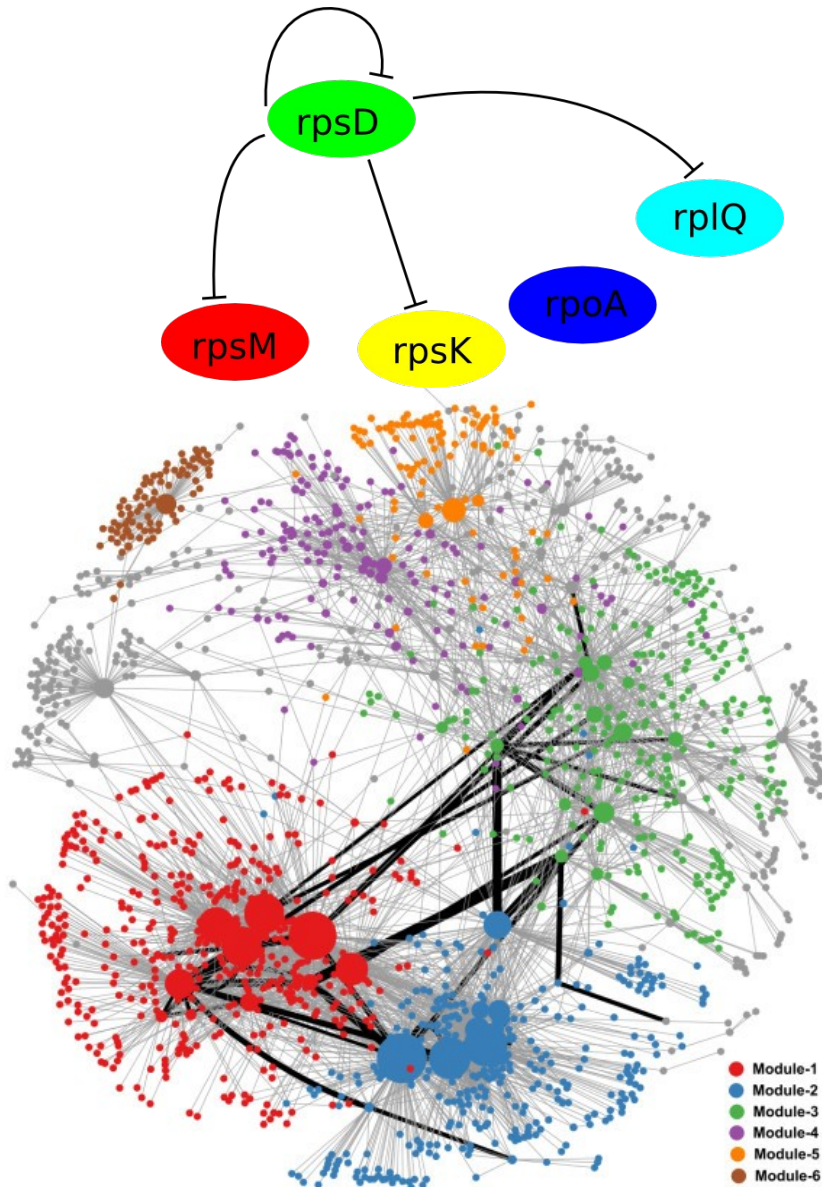
# Gene products regulate the expression of other genes.

- 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



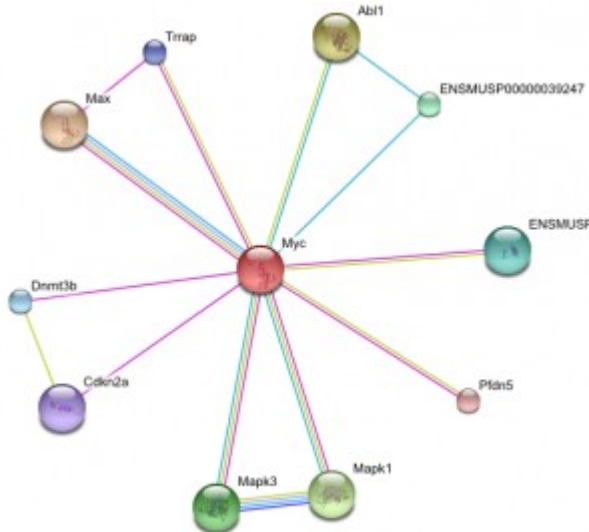
- It can be something simple like the first...
- ... but in fact they look more like the second.
- In any case, they contain arrows that represent the interactions between genes, the direction and the nature (activation, repression).
- For that we need specialised tools to view and analyse.

# A few words on *graphs*!

## Graphs consists of:

- ***Nodes or Vertices***

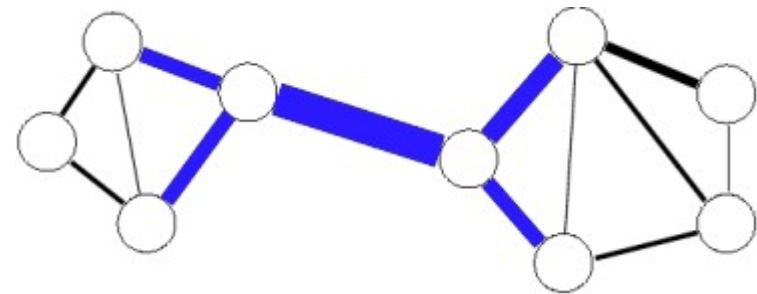
The nodes represent genes (or gene products) in a GRN.



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.

The top screenshot displays the plantgrn.noble.org interface for the protein GIK. The page shows the following details:

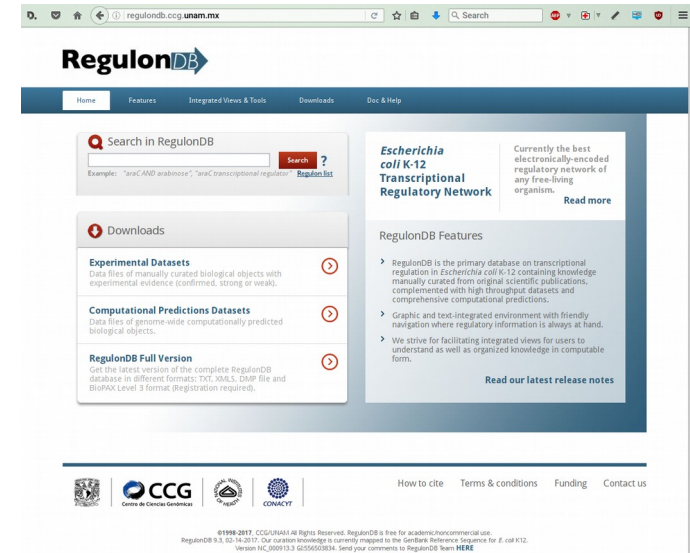
- Type: Protein
- GRN ID: np21957
- Name: GIK
- Annotation: GIK | Predicted AT-hook DNA-binding family protein
- Gene Locus: AT2G35270
- Gene Expression (Affy ATH1 probe(s)): 20854\_4\_at

The main content is a network graph with GIK as the central node. It is connected to several other genes: AT60936, AT62240, AT202480, AT619300, AT593320, AT631840, AT631930, AT5620550, and ABE13. The graph is displayed in a 'Concentric' layout. On the right side, there is a 'Graph Customization' panel with options for highlighting, displaying edges, and a 'Data Filter' panel with various interaction types checked.

The bottom screenshot shows the yeastgenome.org interface for the DOG1 locus. The page title is 'Interaction Network'. The main content is a network graph with DOG1 as the central node. It is connected to several other genes: BAS2, HHT1, ABR1, and SHI1. The graph is displayed in a 'Concentric' layout. On the left side, there is a sidebar with navigation options: DOG1 / YHR044C, Interactions, Overview, Annotations, Interaction Network, and Resources. On the right side, there is a 'Filter Interaction by Type' section with options for 'All', 'Physical', and 'Genetic', and a 'Filter Interactions by # of Experiments' section with a value of 1.

# 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/>

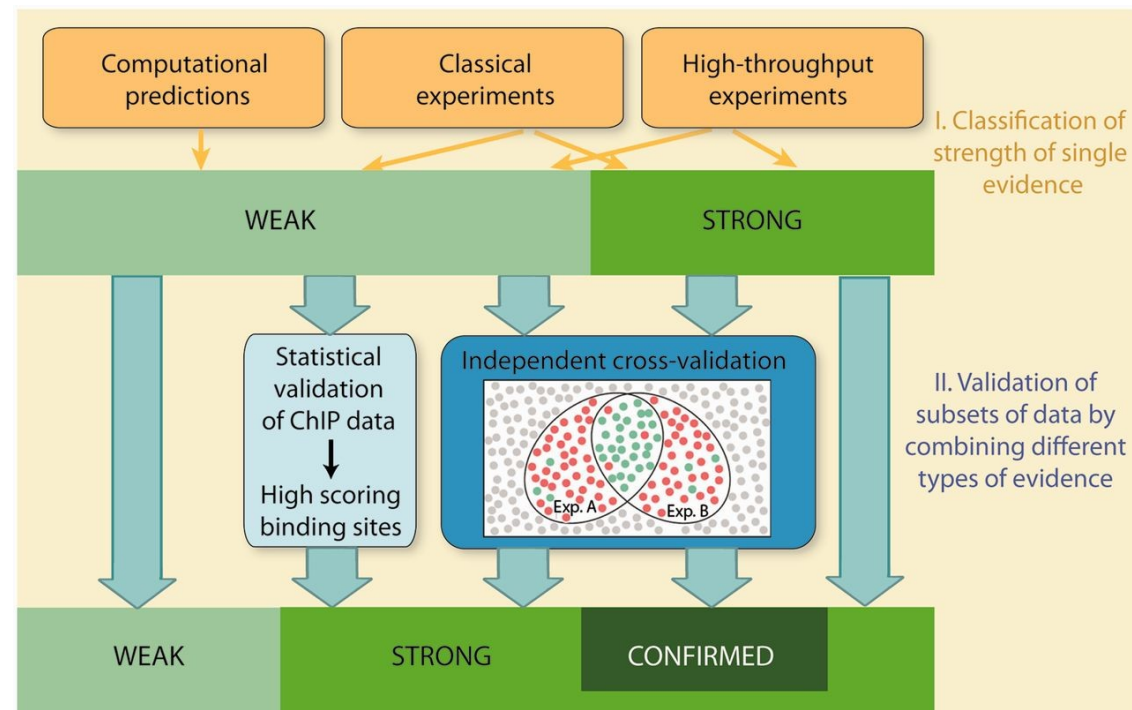


# How regulatory interactions have been deciphered

- A combination of expert knowledge and high-throughput datasets.
  - CHIP-Seq, CHIP-Chip, CHIP-exo
  - RNA-Seq, microarrays
  - GSelex

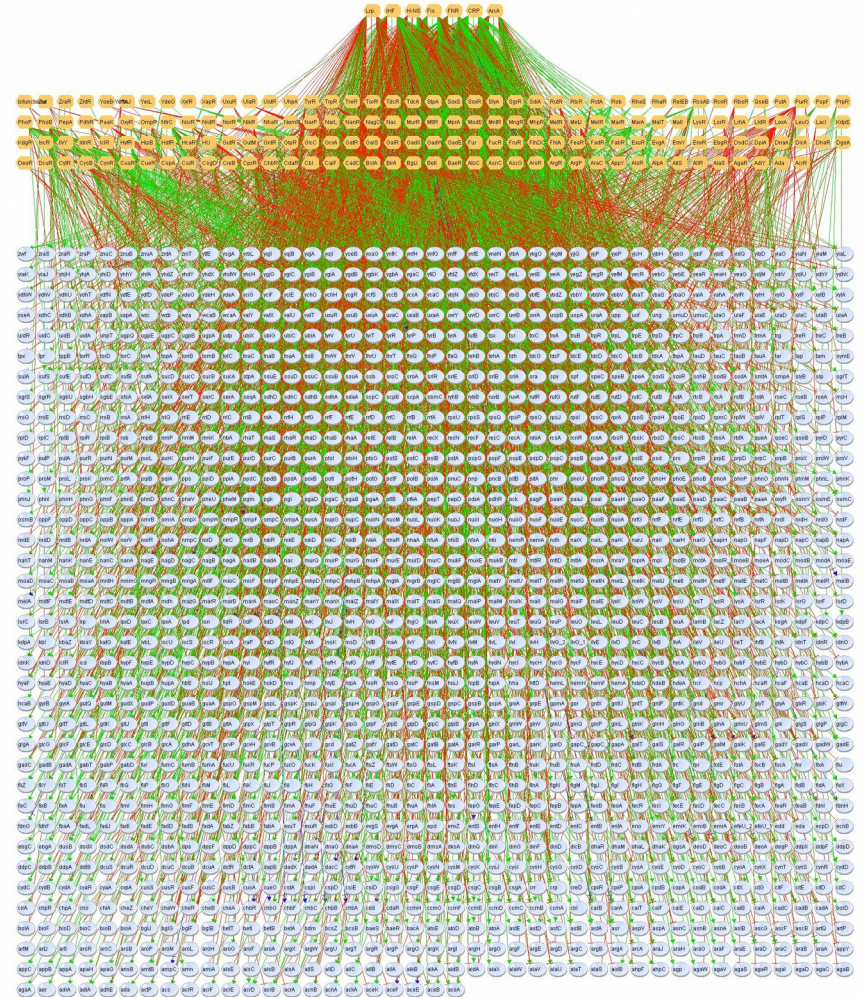
- Two classifications:

- STRONG evidence (CHIP-SV)
- WEAK evidence (CHIP, GEA, GSELEX)

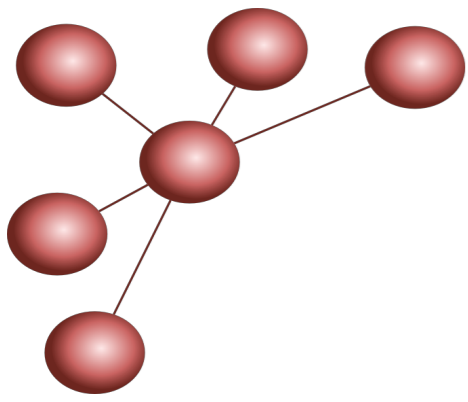


# Tables of regulatory interactions

- 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.

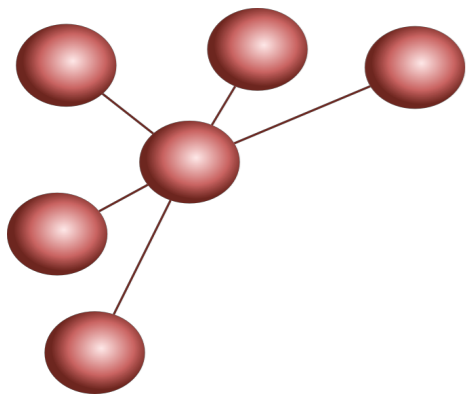


<http://regulondb.ccg.unam.mx/menu/download/datasets/index.jsp>



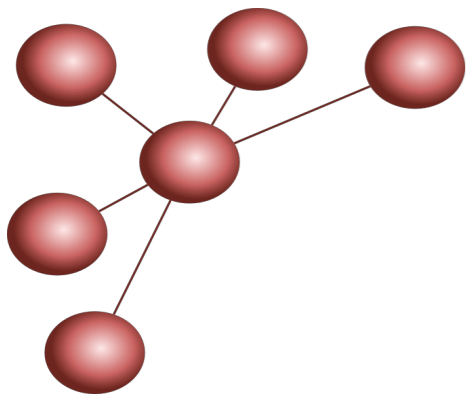
**What next ?**  
**=> Network Analysis**



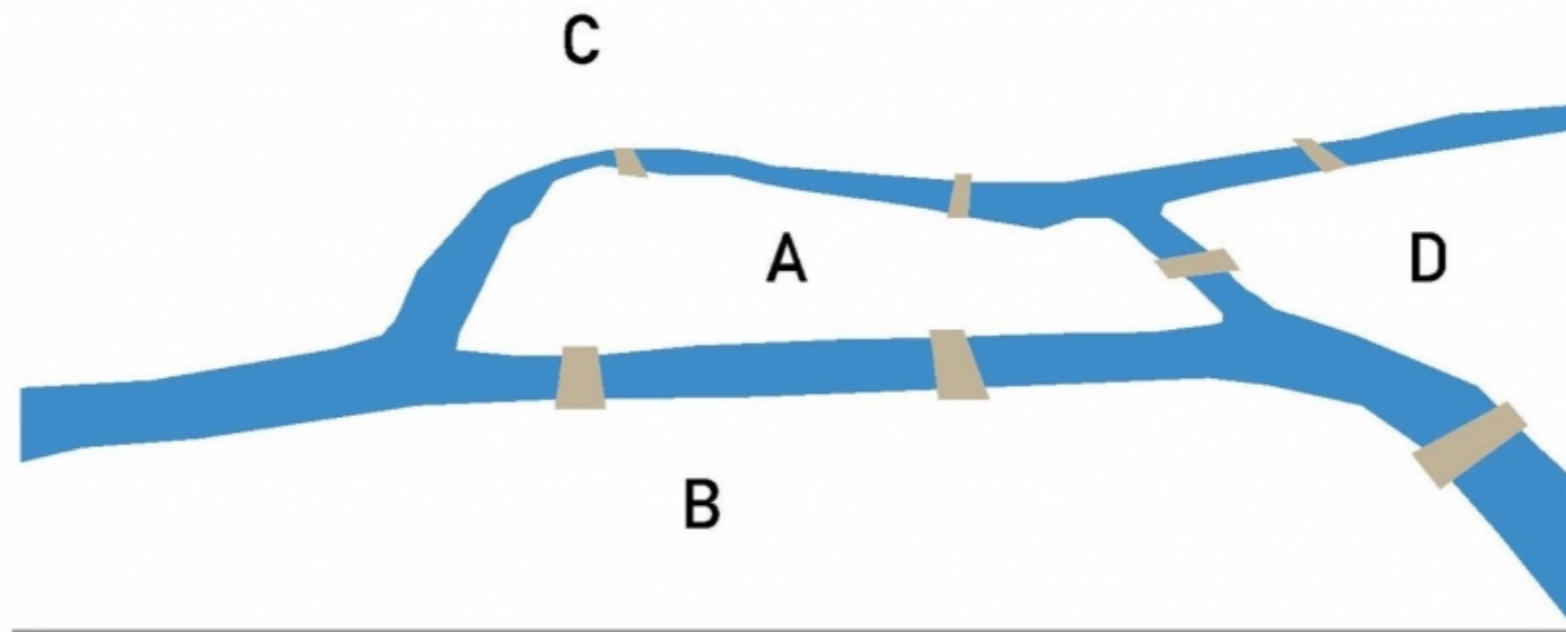
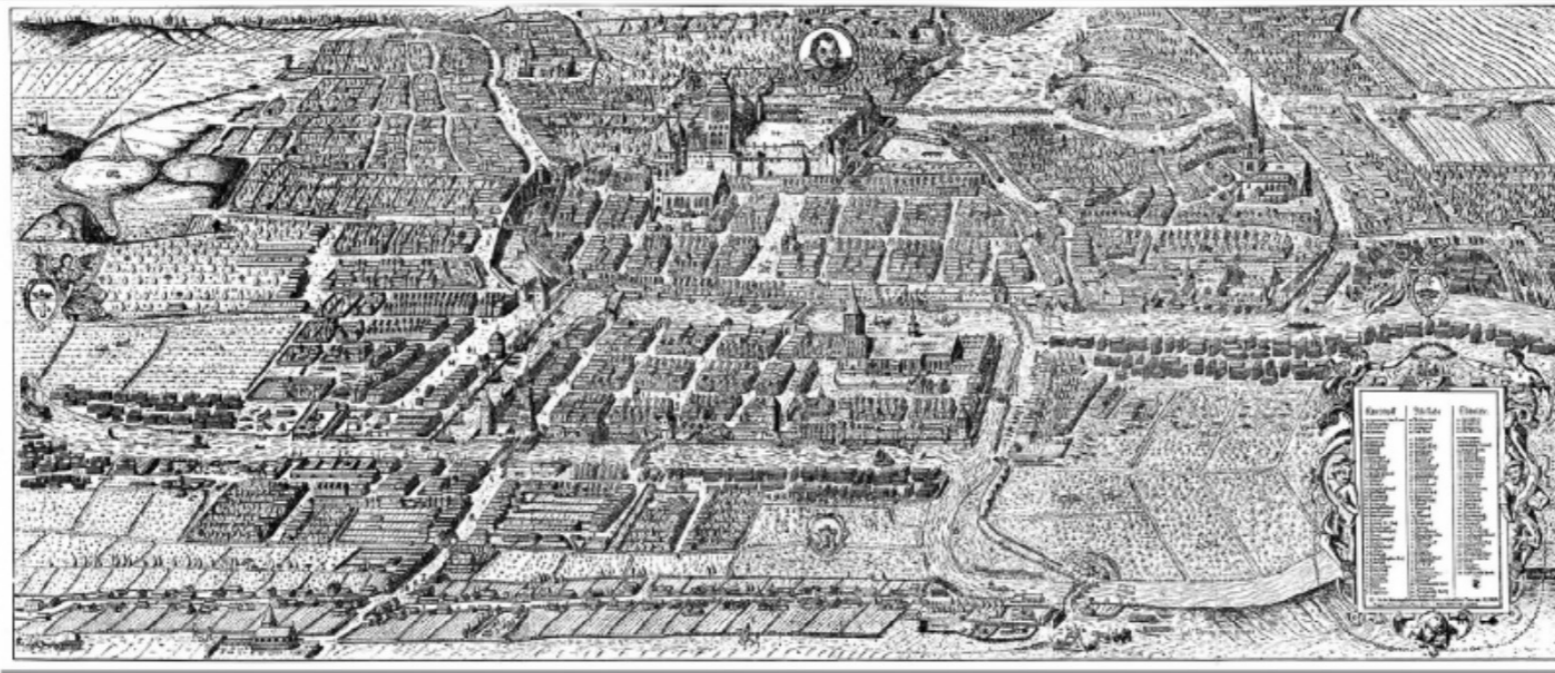


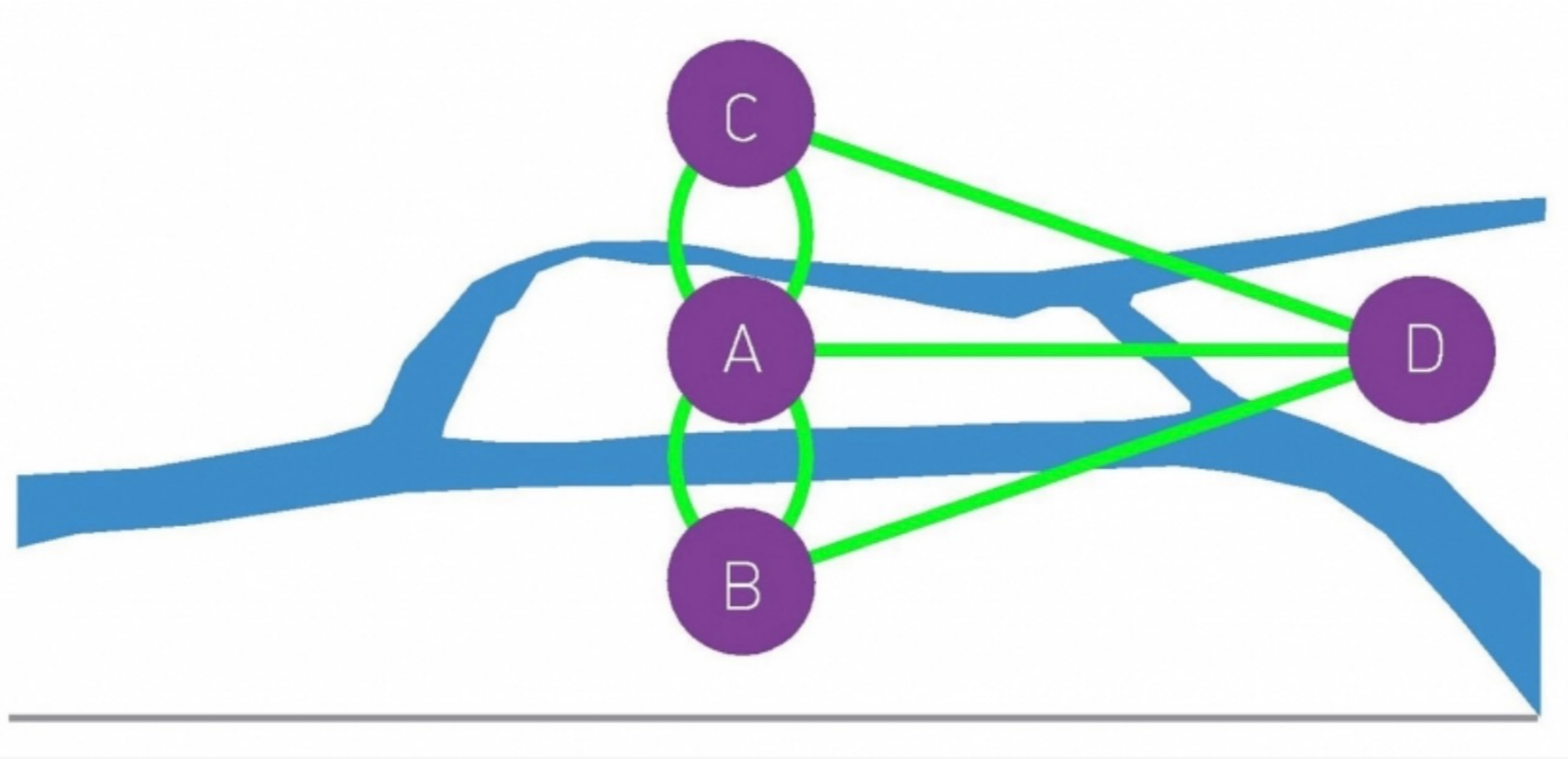
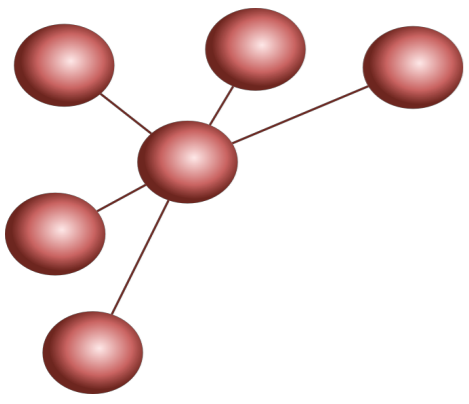
# Network analysis / graph theory

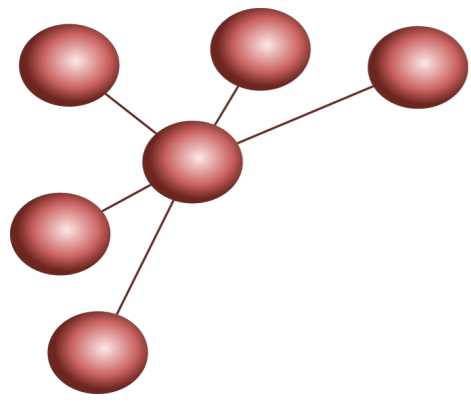




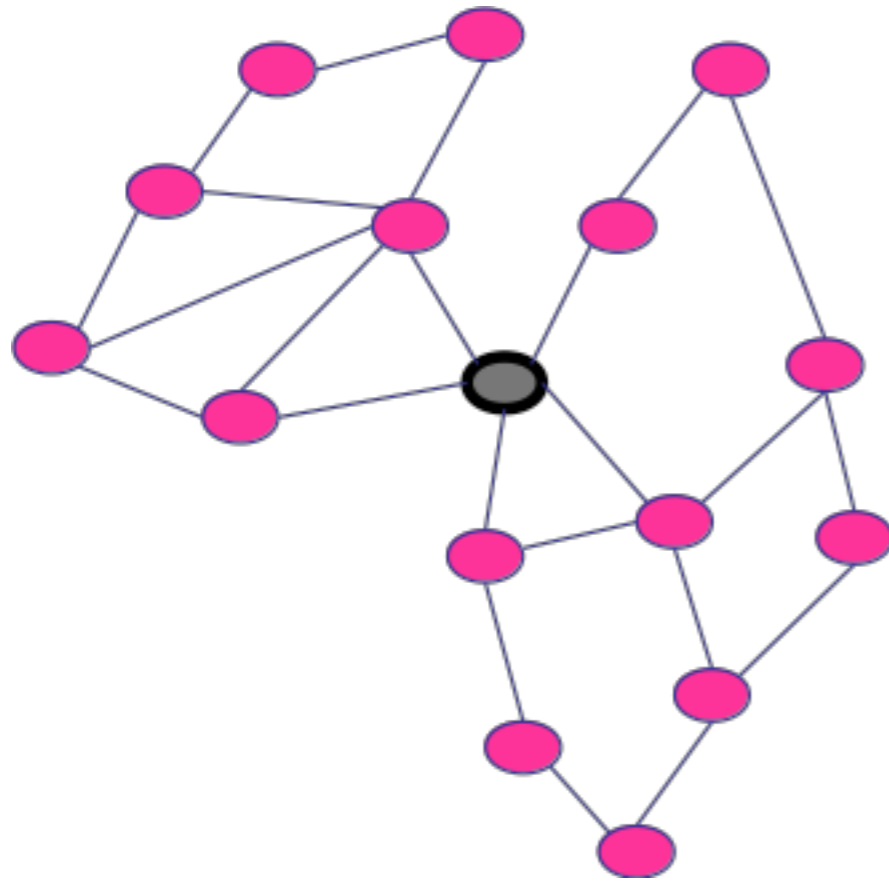
# Königsberg







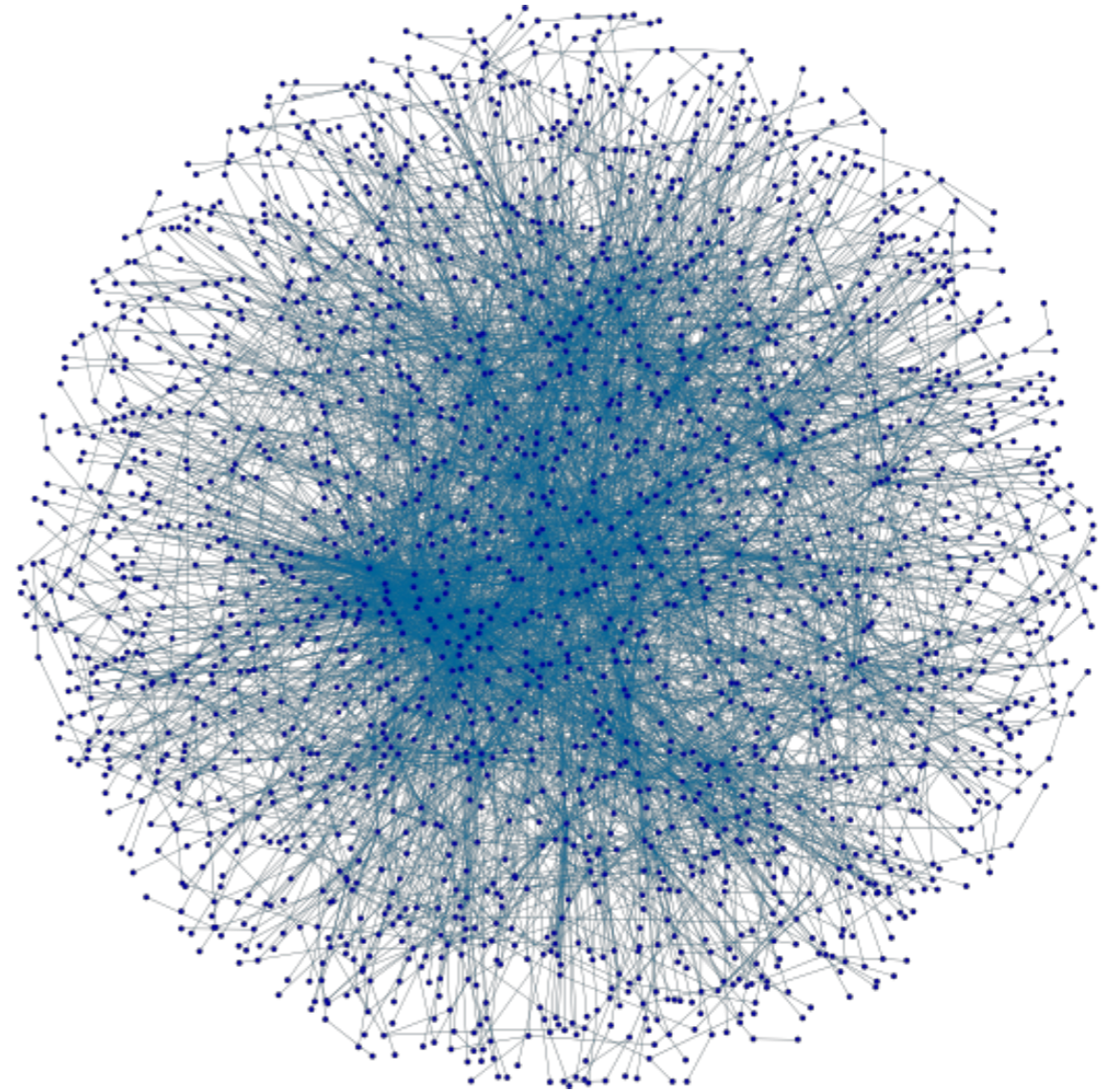
**Local approaches**



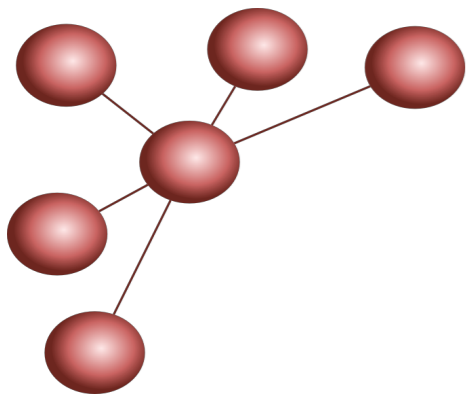
**“guilt by association”**

# How to use large-scale biological networks ?

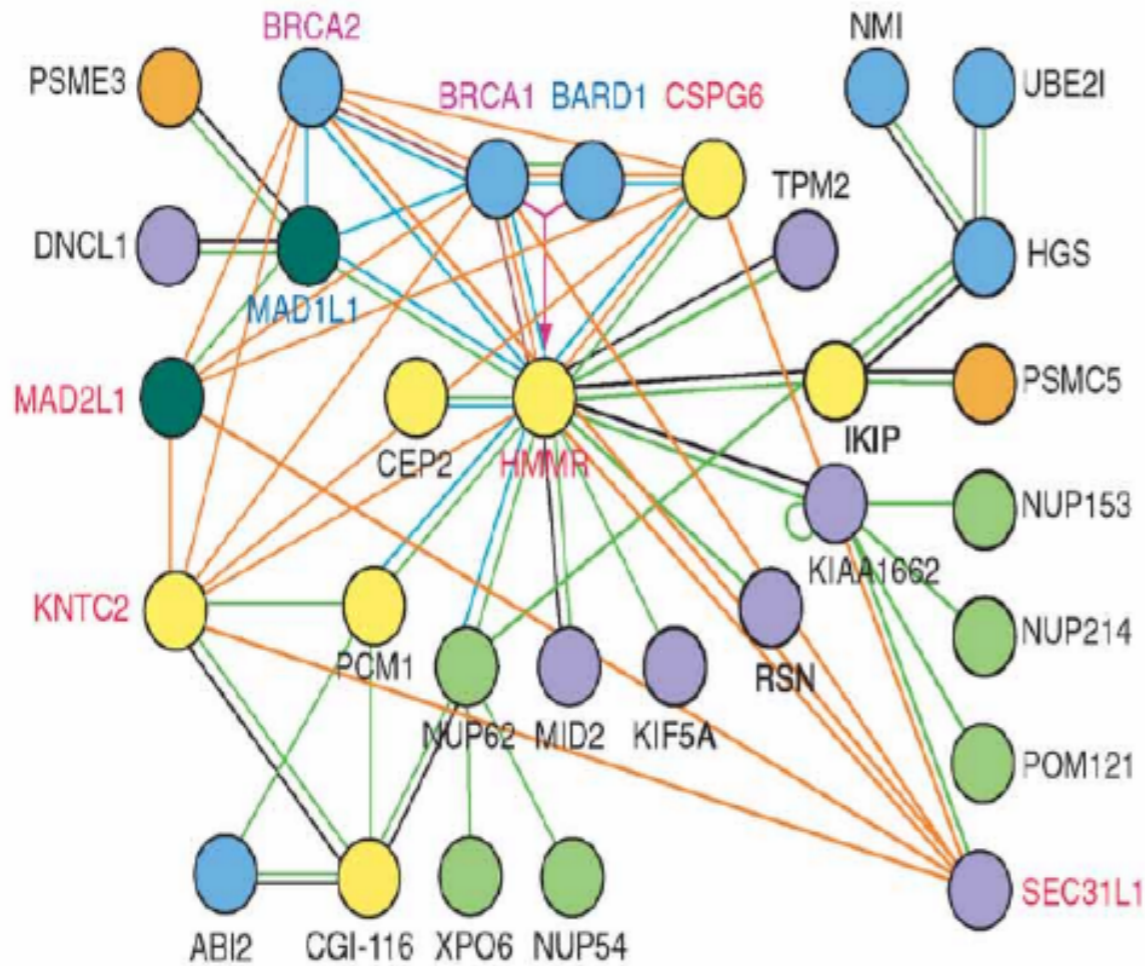
**Global approaches**



**Topological features  
Clustering / communities**



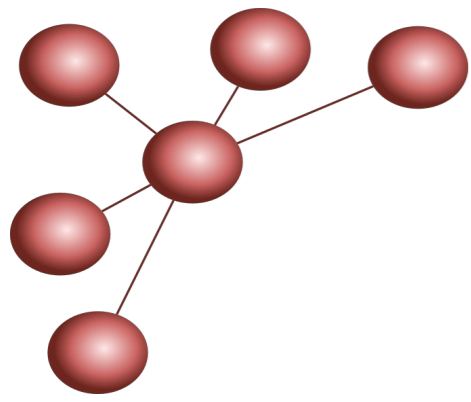
# Identification of a new gene involved in breast cancer



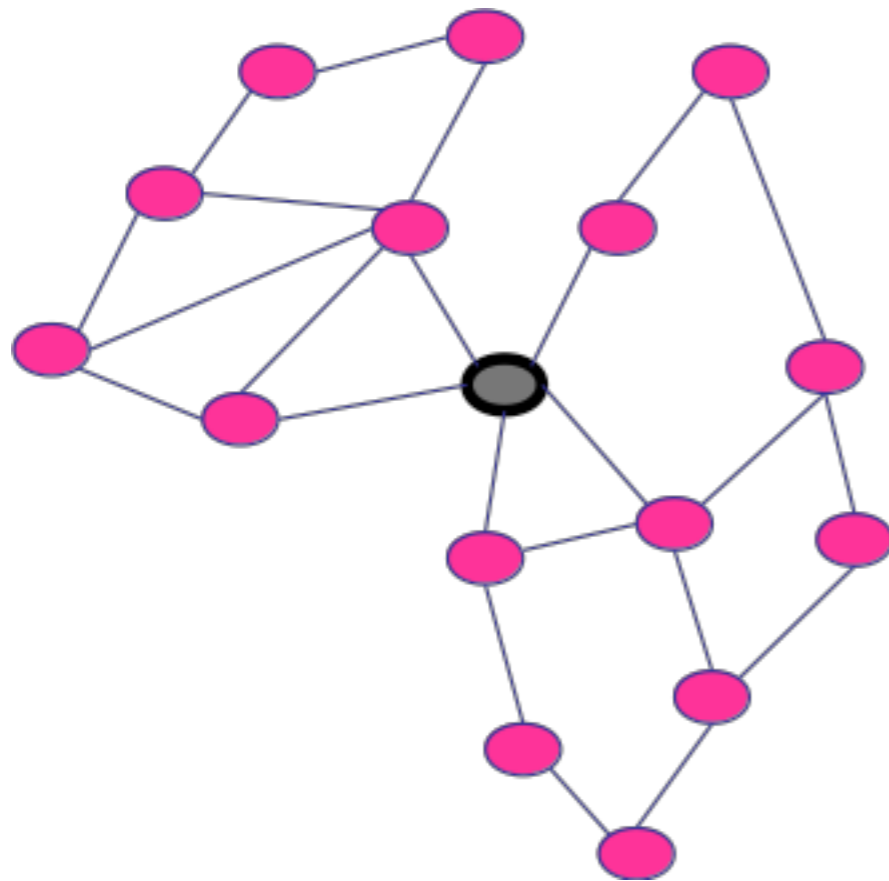
**Nodes correspond to proteins,  
edges to interactions  
identified by different  
experimental techniques**

## Functional associations (*n*)

- Expression profiling similarity (20)
- Similar gene deficiency phenotype (2)
- Y2H binary protein interaction (32)
- Protein co-AP (13)
- Protein co-IP (11)
- Biochemical interaction (1)



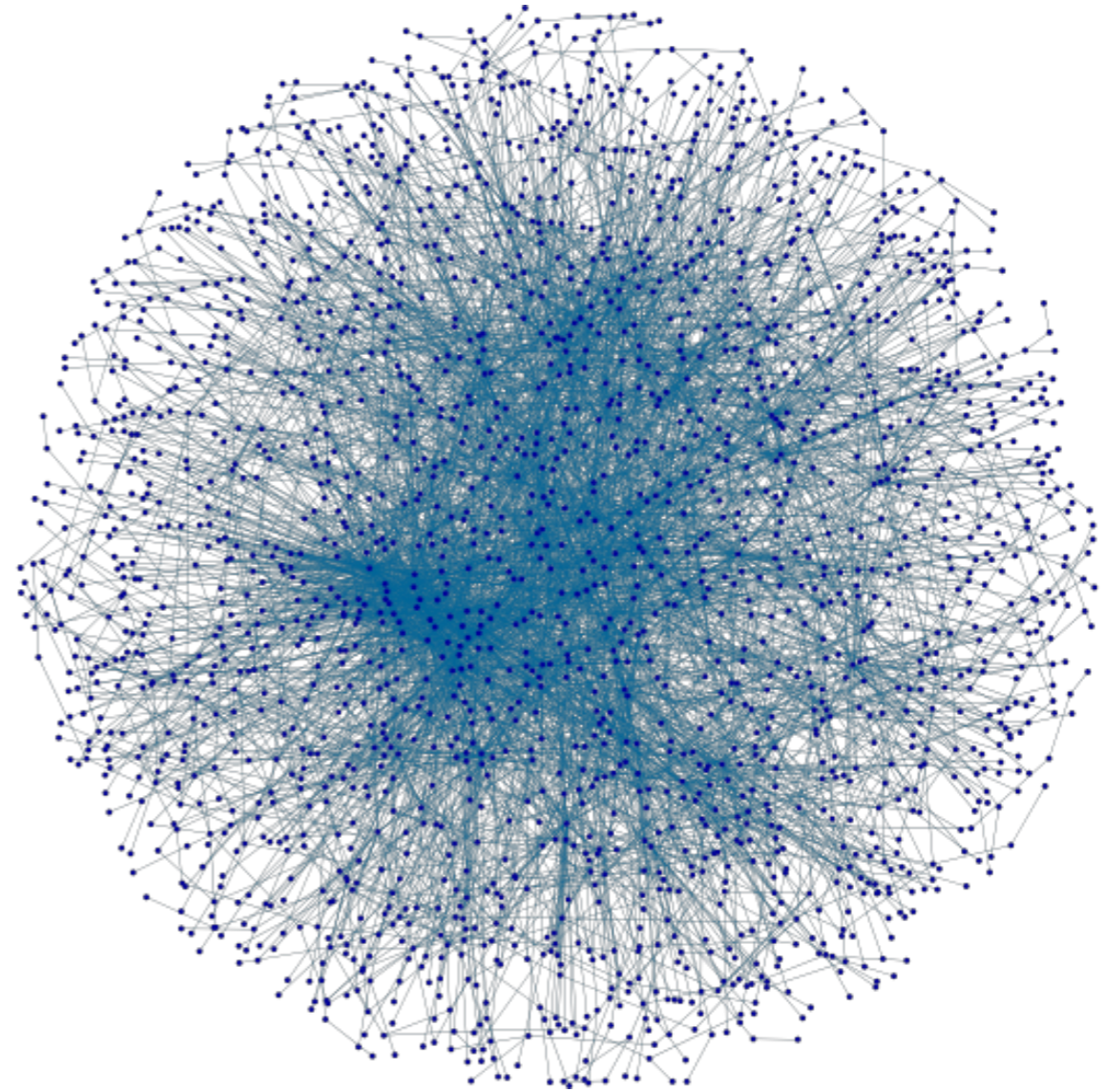
**Local approaches**



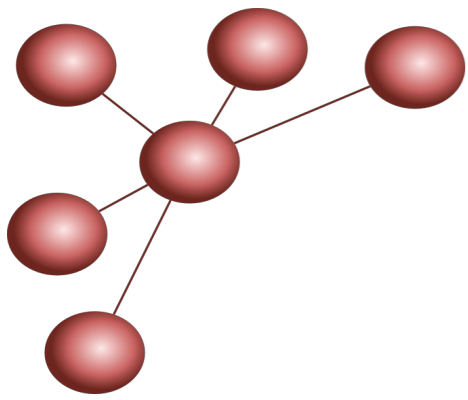
**“guilt by association”**

# How to use large-scale biological networks ?

**Global approaches**



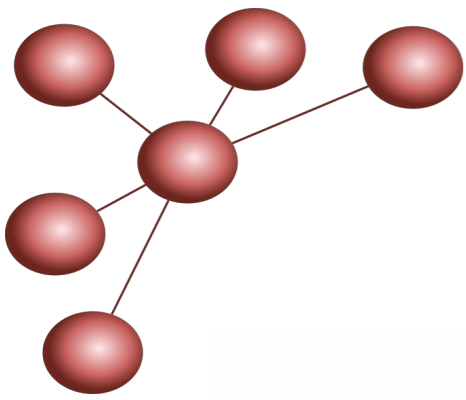
**Topological features  
Clustering / communities**



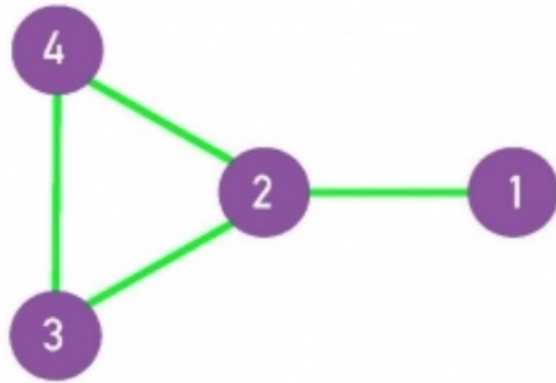
# Global approaches - topological measures

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

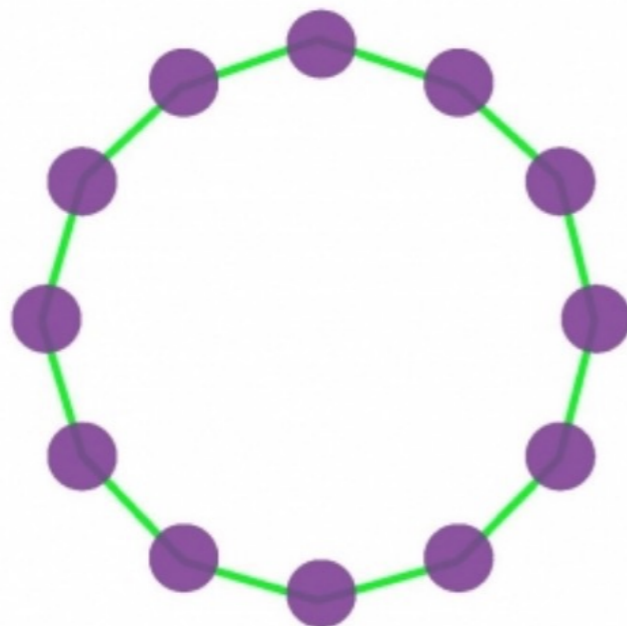
# Degree distribution



a.

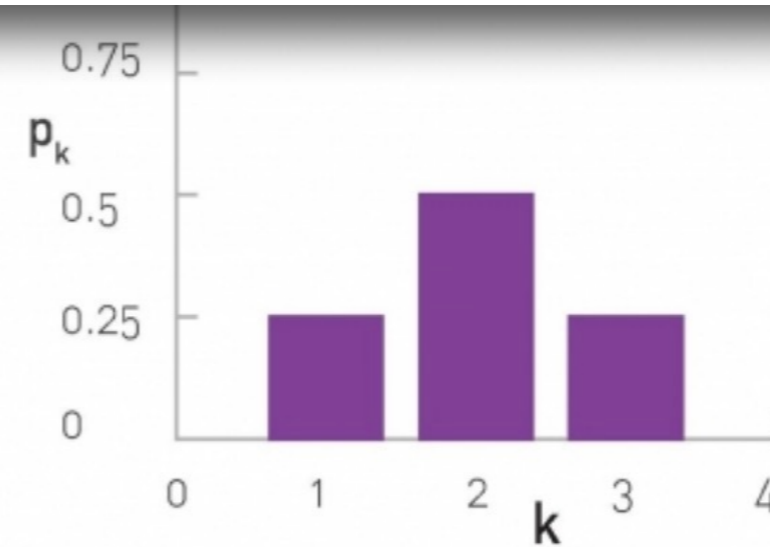


c.

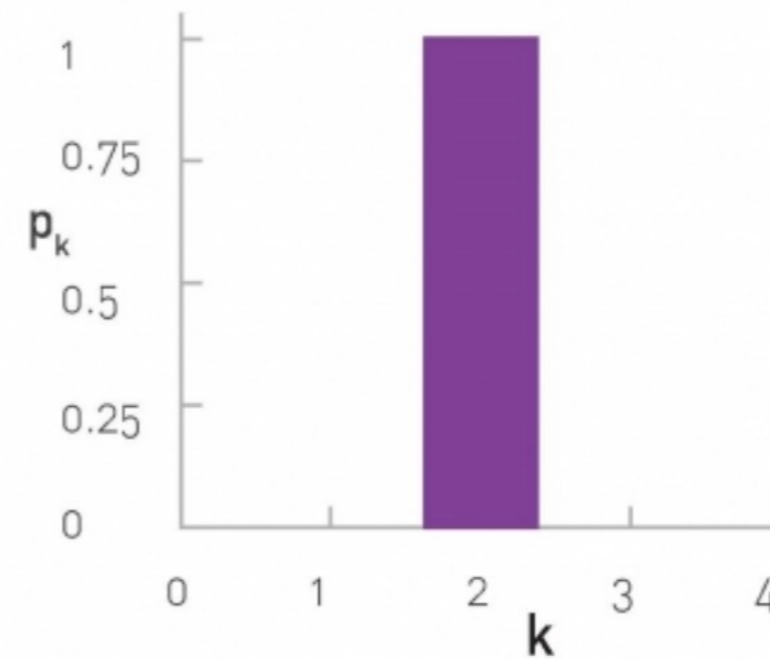


## NetworkAnalyzer

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

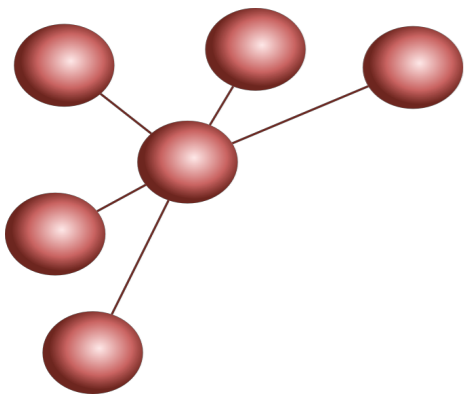


d.



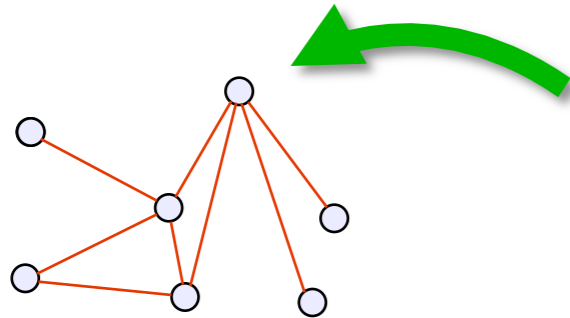


# Protein degree distribution : interactomes are scale-free and small-world



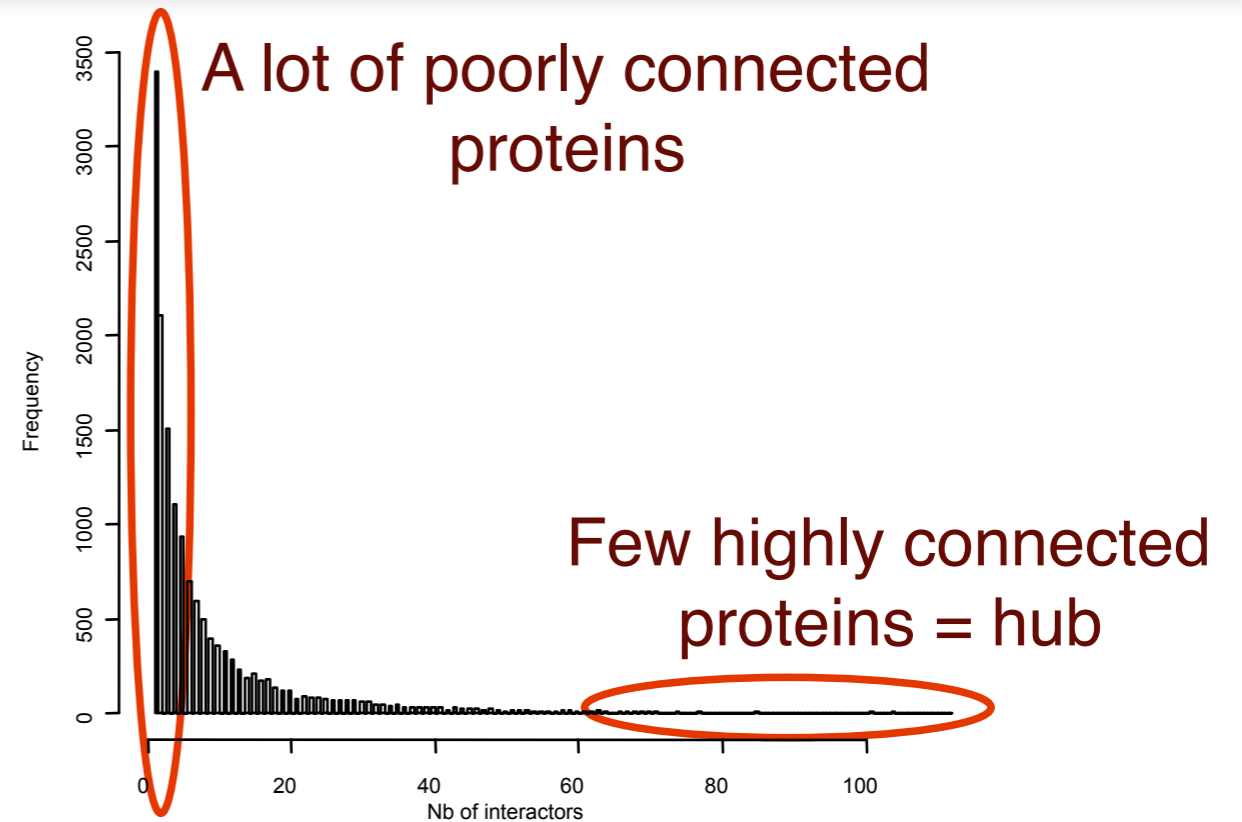
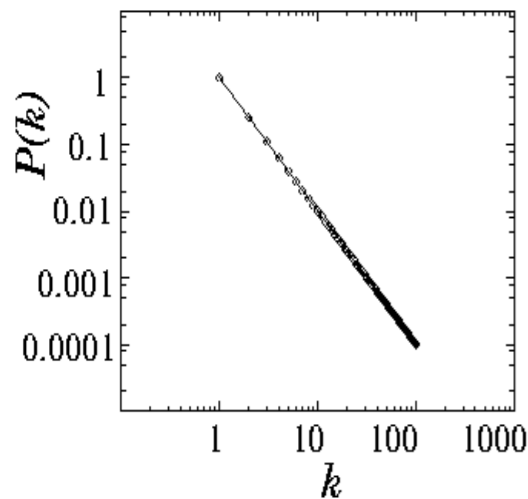
## NetworkAnalyzer

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



$k = 4$

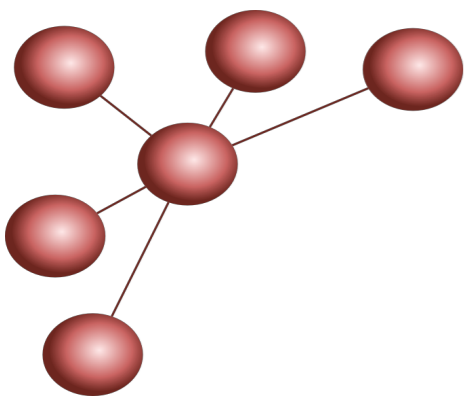
Power-law distribution




**Biological interpretation?**

**Robust to random attack, sensitive to targeted attacks**

**Growth with preferential attachment (“rich get richer”) => create “hubs”**



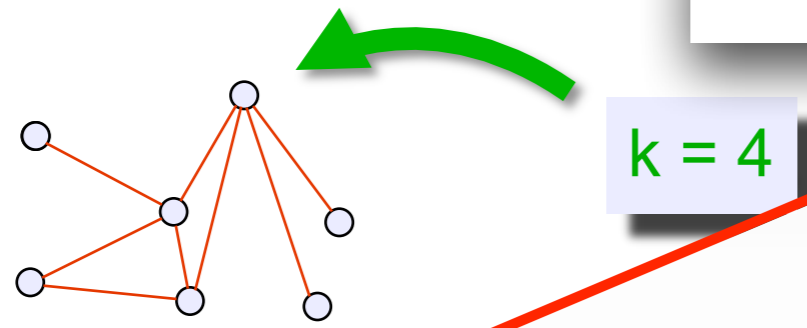
# Protein degree distribution : interactomes are scale-free and small world



## NetworkAnalyzer

Core App: Computes basic network statistics, clustering coefficient, etc.

www.rsc.org/molecularbiosystems | Molecular BioSystems

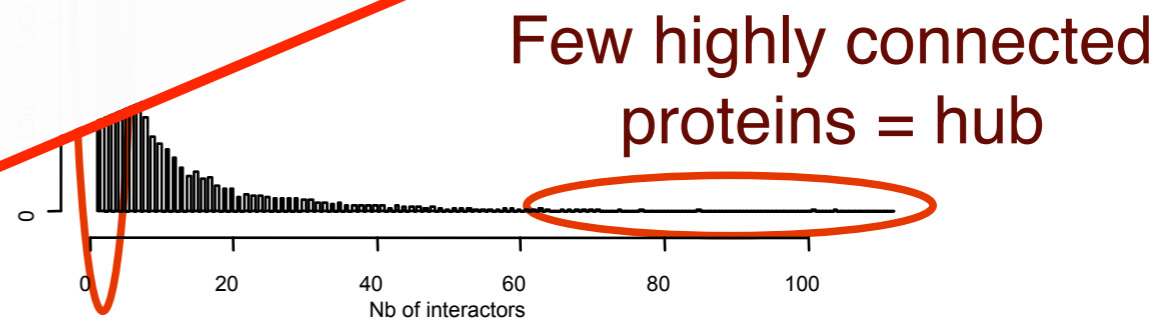


The powerful law of the power law and other myths in network biology†

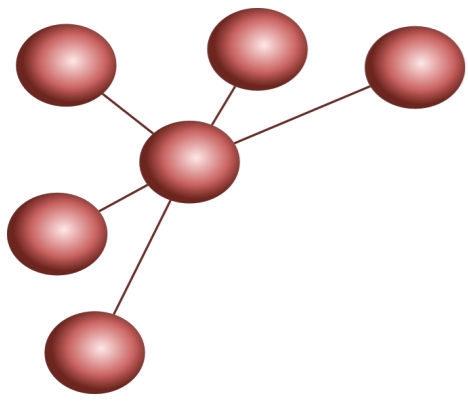
Gipsi Lima-Mendez\* and Jacques van Helden\*

Received 5th May 2009, Accepted 12th August 2009  
First published as an Advance Article on the web  
DOI: 10.1039/b908681a

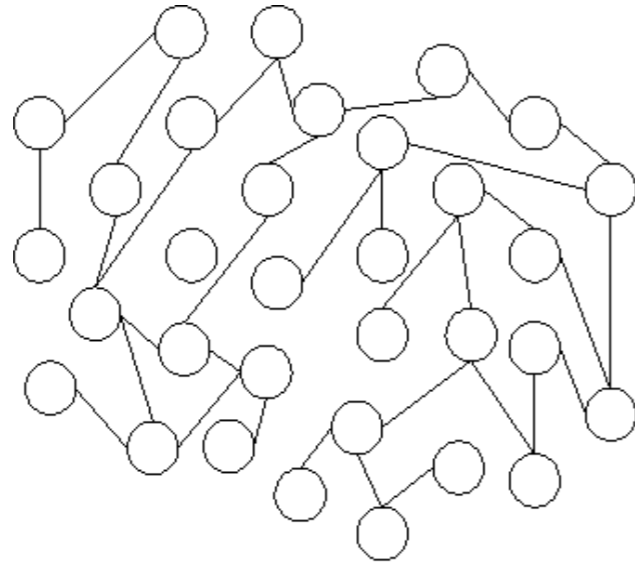
REVIEW



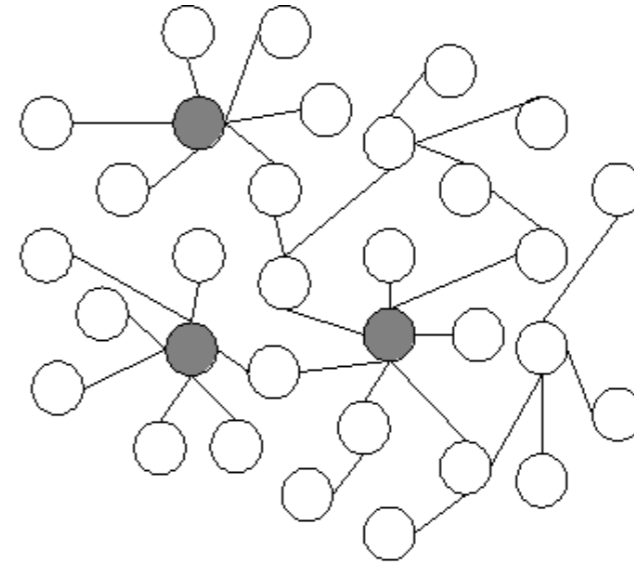
**Biological interpretation?**  
 Robust to random attack, sensitive to targeted attacks  
 Growth with preferential attachment (“rich get richer”) => create “hubs”



# Network topological structure : Small-world property

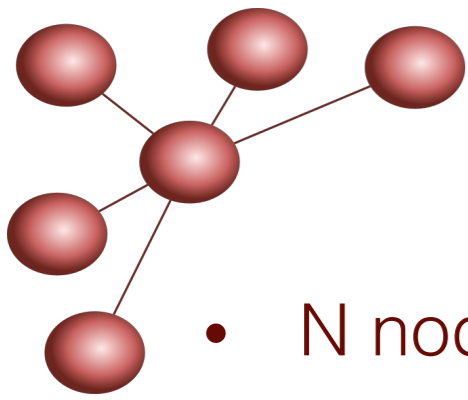


(a) Random network



(b) Scale-free network

- Milgram, 6 degrees of separation



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

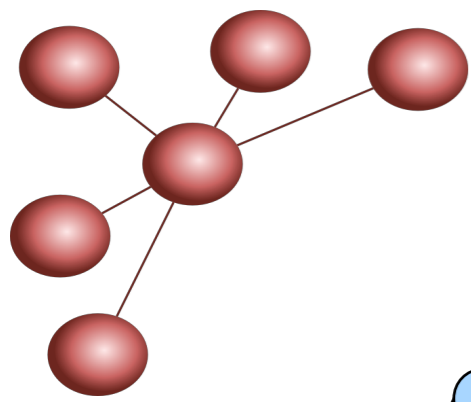
# Metrics on graphs



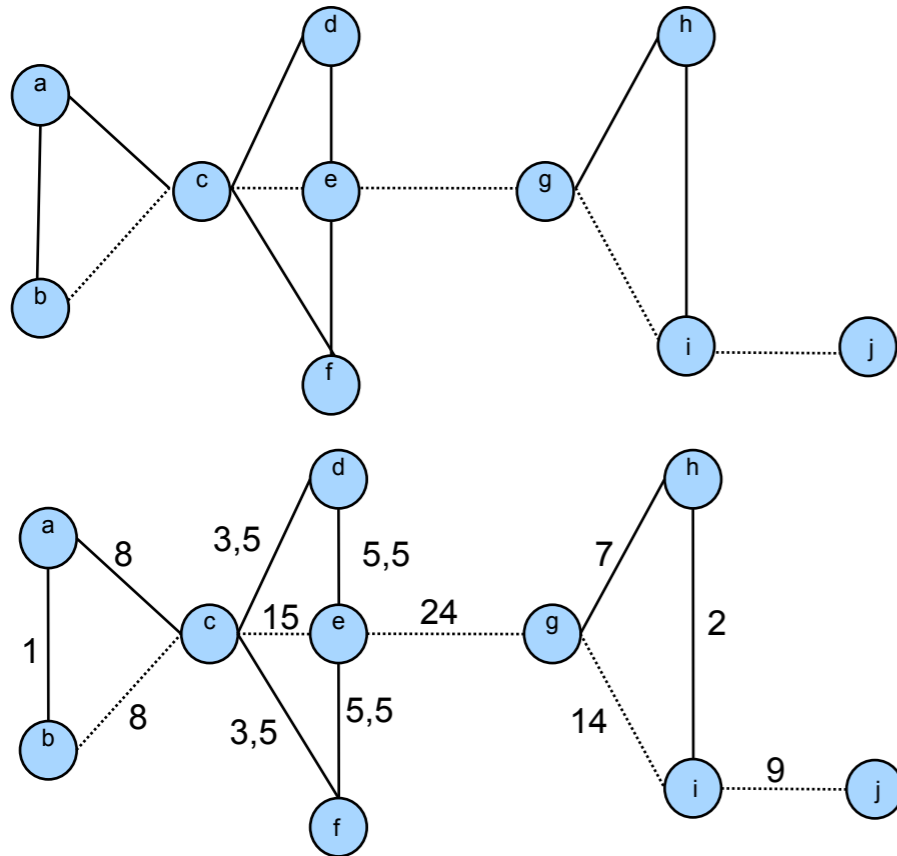
## NetworkAnalyzer

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

NETWORK MEASURES				
Degree/ connectivity (k)	Clustering coefficient/ interconnectivity (C)	Assortativity/average nearest neighbor's connectivity (NC)	Shortest path (SP) between two nodes	Betweenness/ centrality (B)
<p><math>k_A = \text{Nb of edges through } A = 5</math></p>	$C_A = \frac{\text{Actual links between A's neighbors (black)}}{\text{Possible links between A's neighbors (orange)}}$ $C_A = n_A / [k_A(k_A - 1) / 2]$ $= 2 / [4 \times (4 - 1) / 2] = 0.333$	$NC_A = (k_B + k_C + k_D + k_E + k_J) / 5$ $= (5 + 2 + 2 + 3 + 1) / 5 = 2.6$	$SP_{FH} = (F, D, A, B, H) = 4$	$B_4 = \text{Fraction of SPs passing through } A$ $= 0.090$

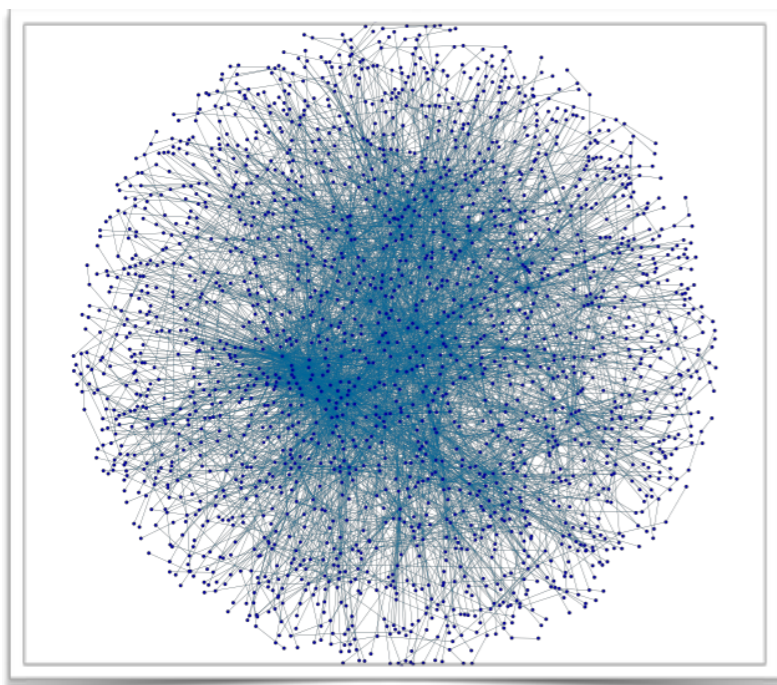


# “Betweenness”



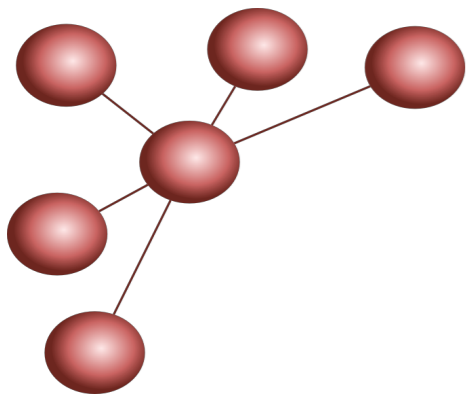
Number of shortest paths  
running through an edge  
=  
“bottleneck”

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

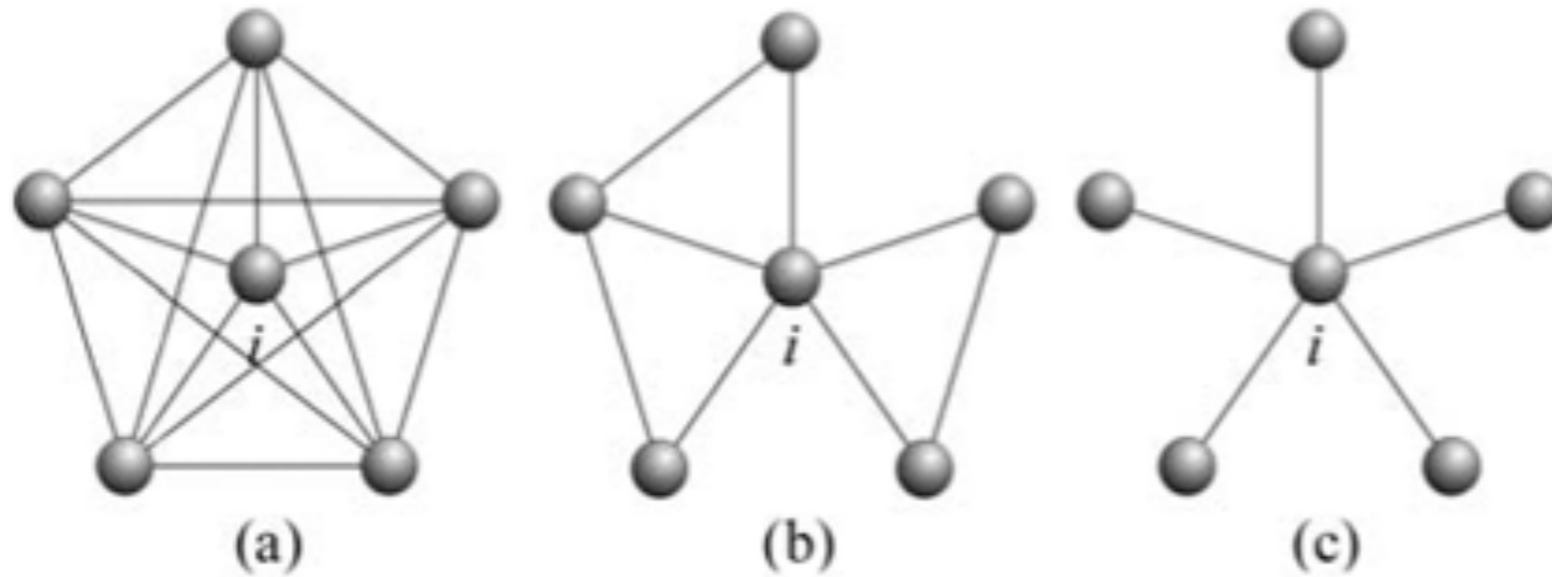


## NetworkAnalyzer

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



# Clustering coefficient / modularity

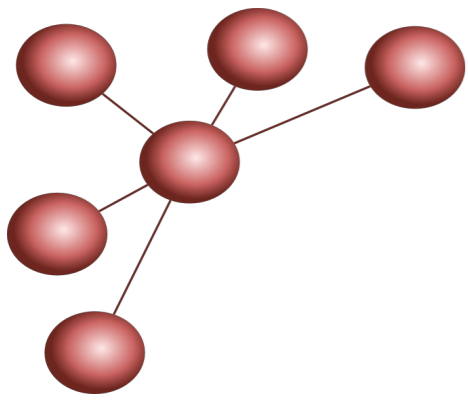


**Actual links between neighbours / Possible links between neighbours**



## NetworkAnalyzer

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

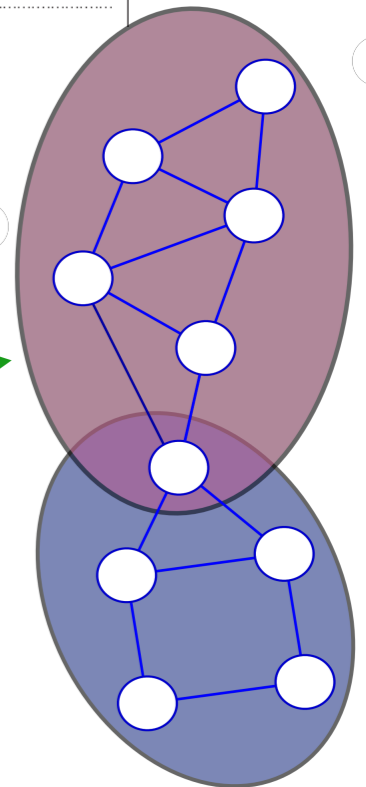
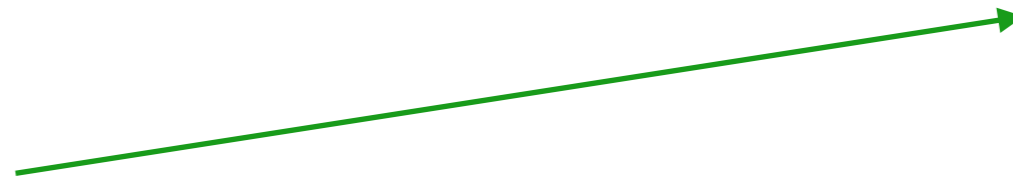
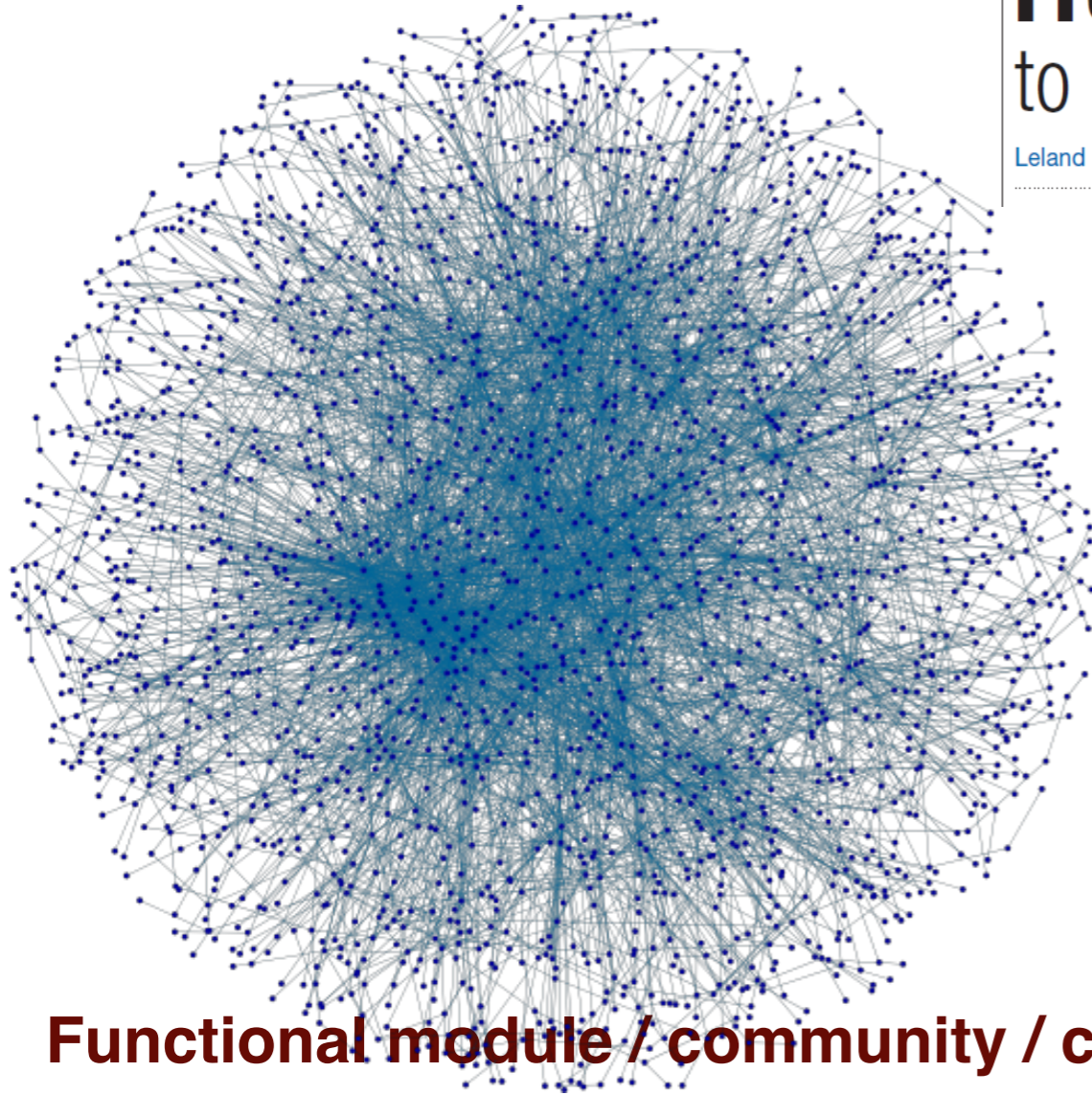


# Global approach - Clustering

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.

