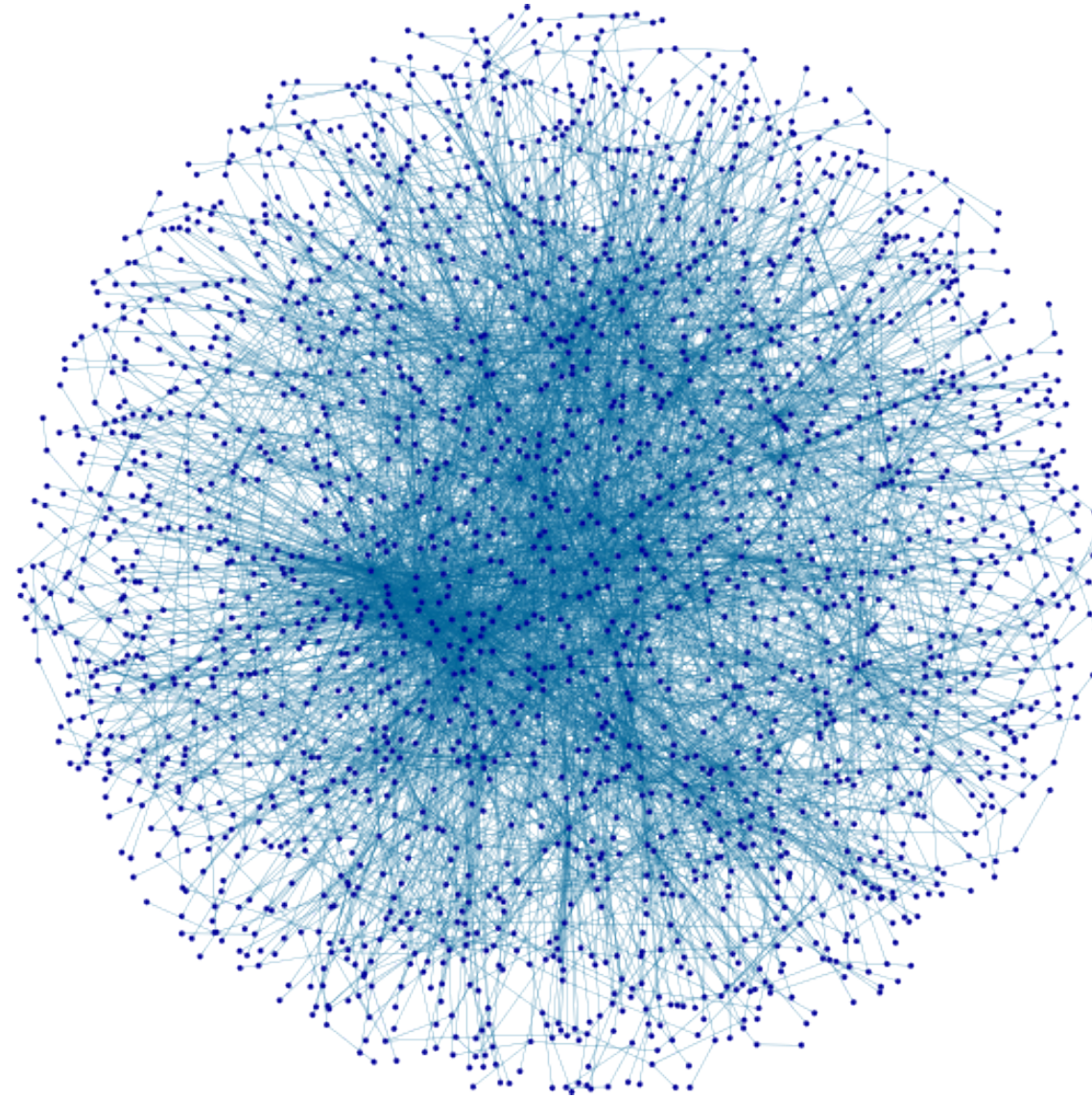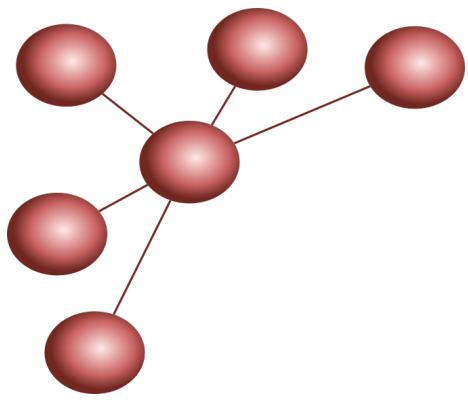# Integrative Bioinformatics

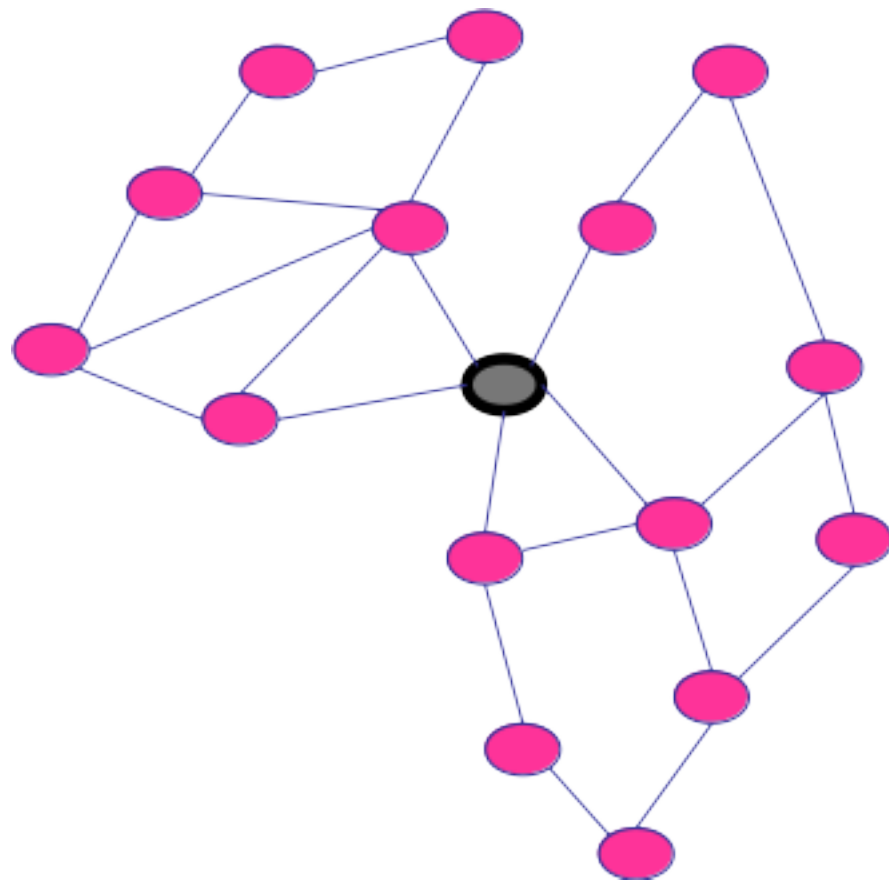## Module 6 - Session 6

What next ?
=> Network Analysis
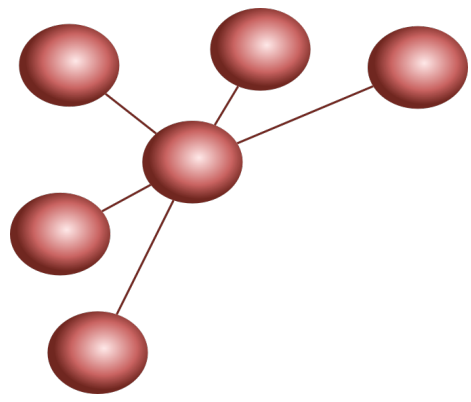
How to use large-scale biological networks ?

Local approaches

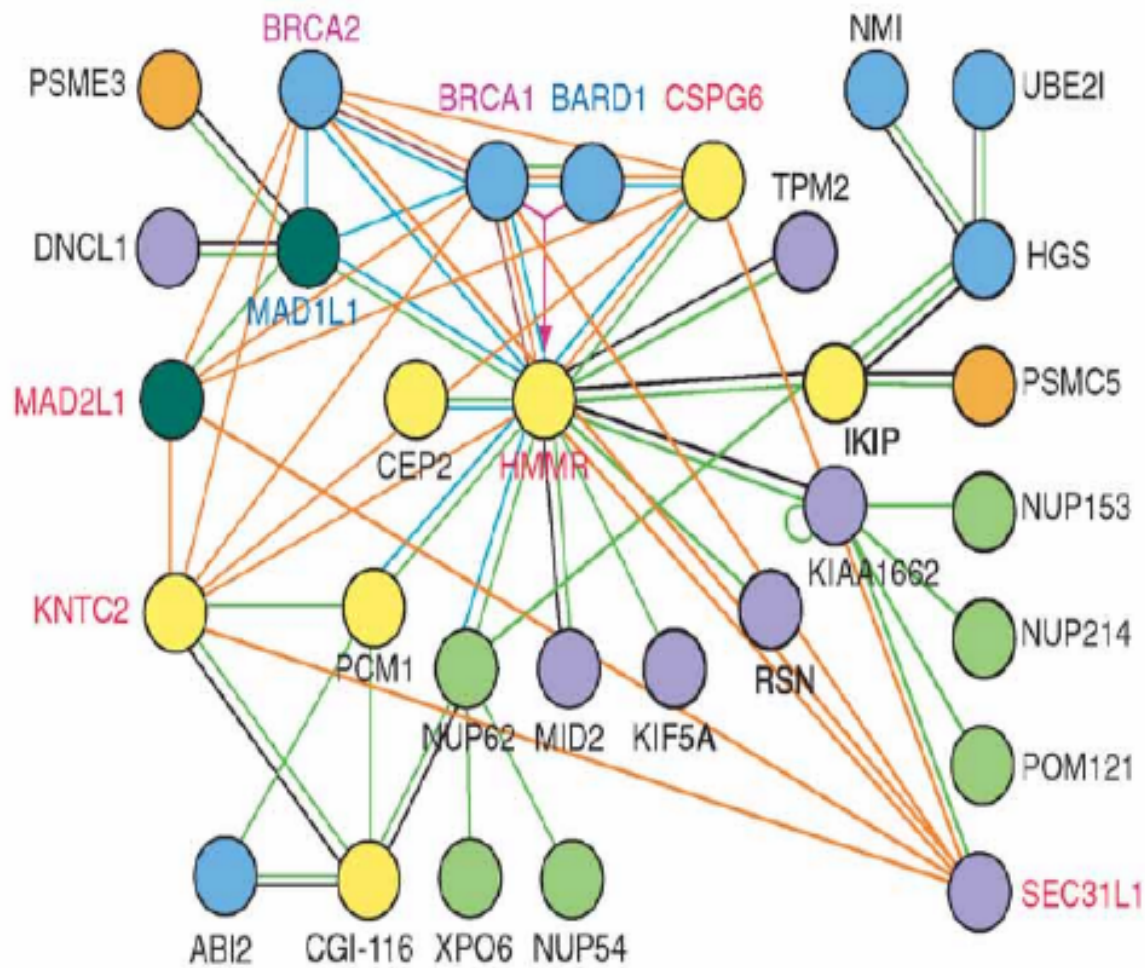**Global approaches**

"guilt by association"
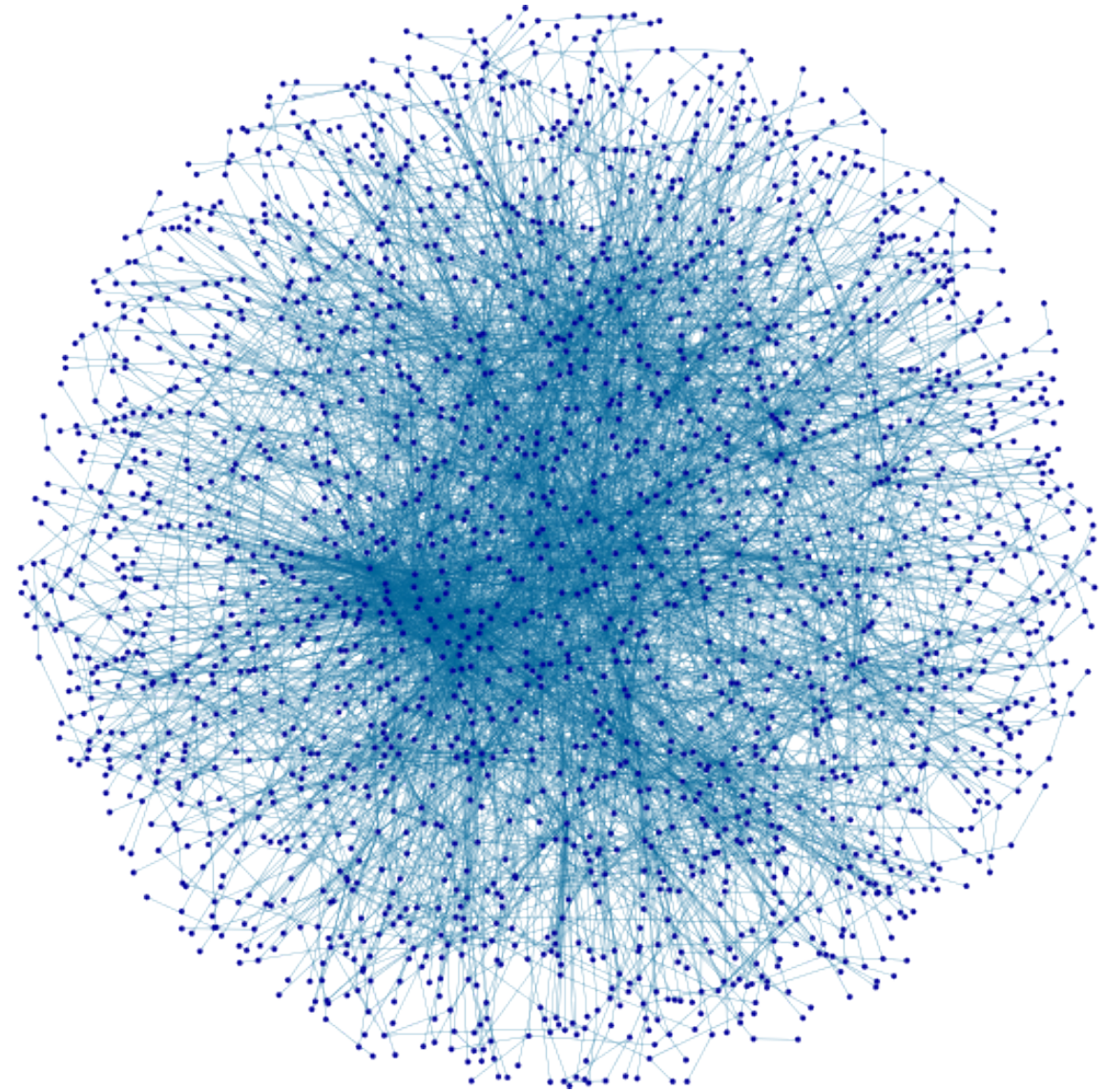
**Topological features
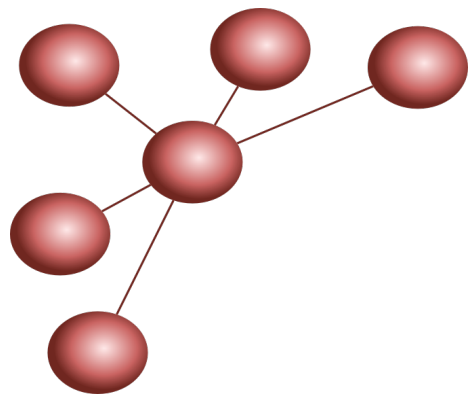Clustering / communities**
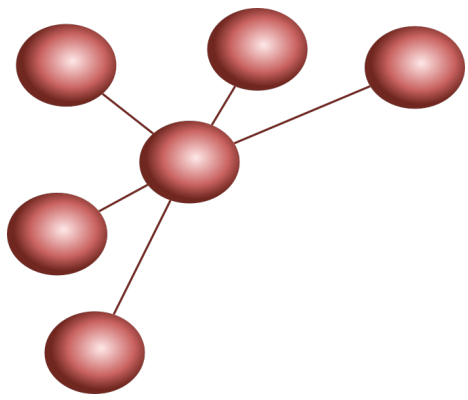
Identification of a new gene involved in breast cancer

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

Pujana et al. 2007
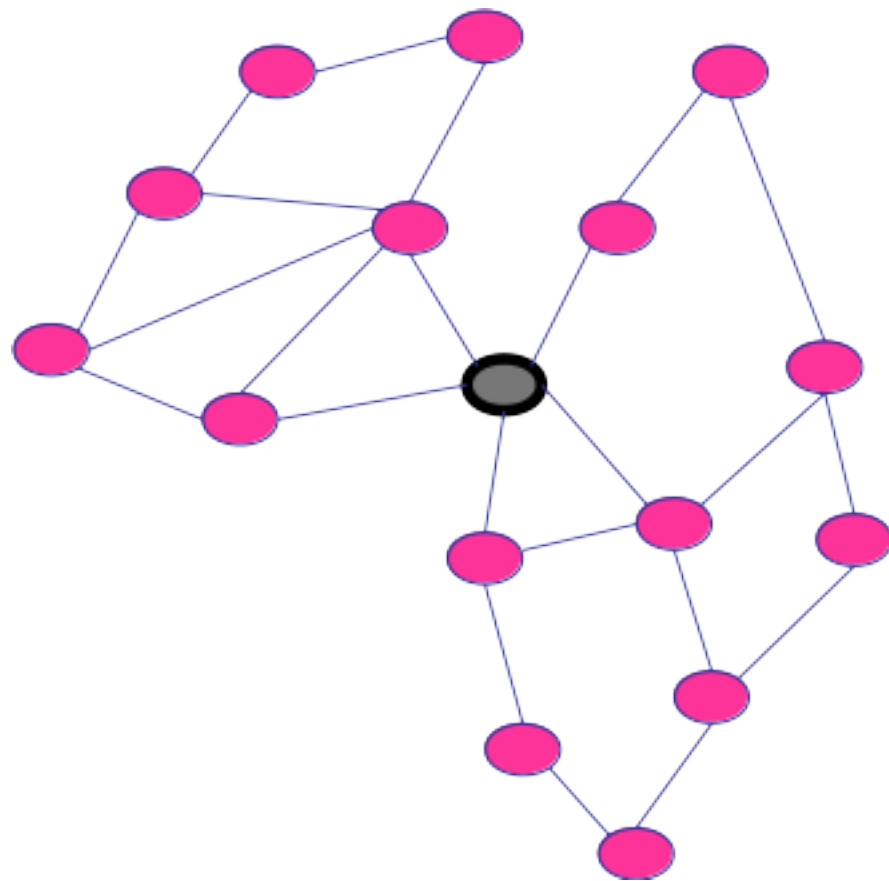
How to use large-scale biological networks ?

Local approaches

Global approaches

"guilt by association"
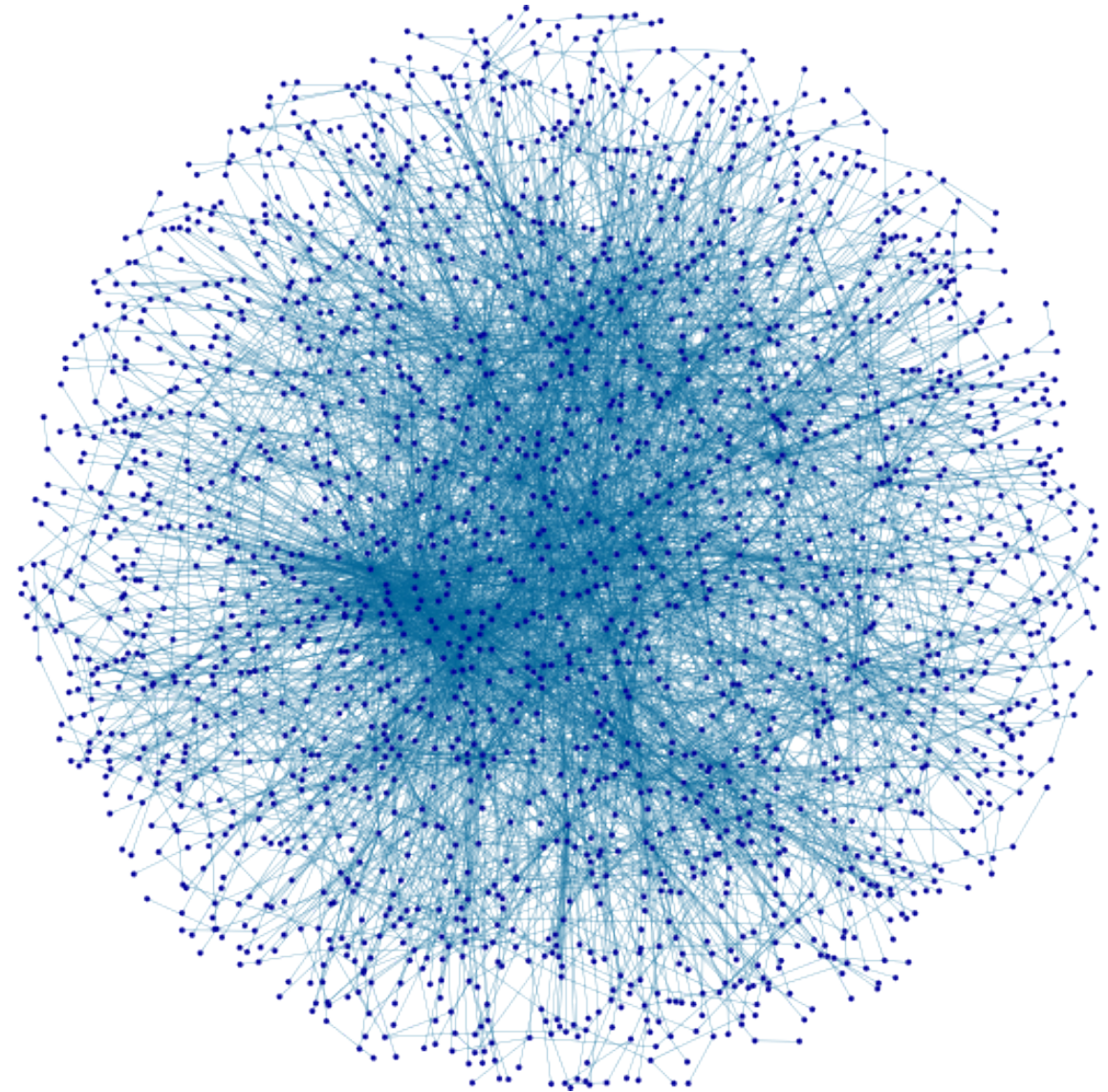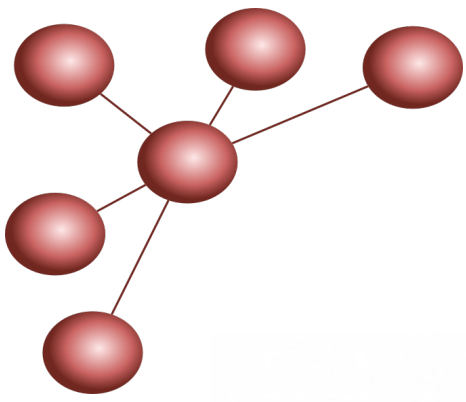
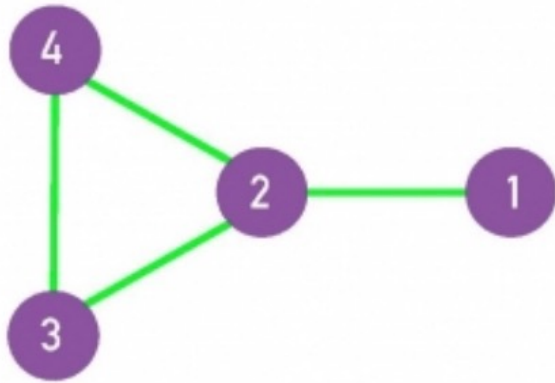**Topological features**
Clustering / communities
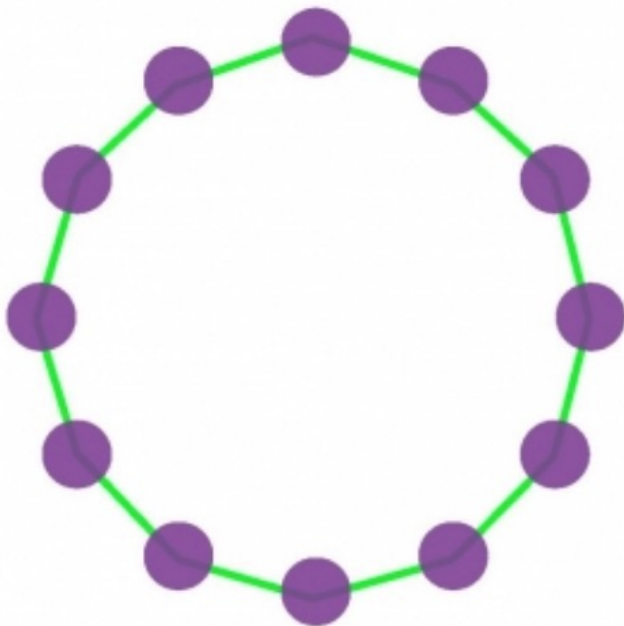
Degree distribution

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

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

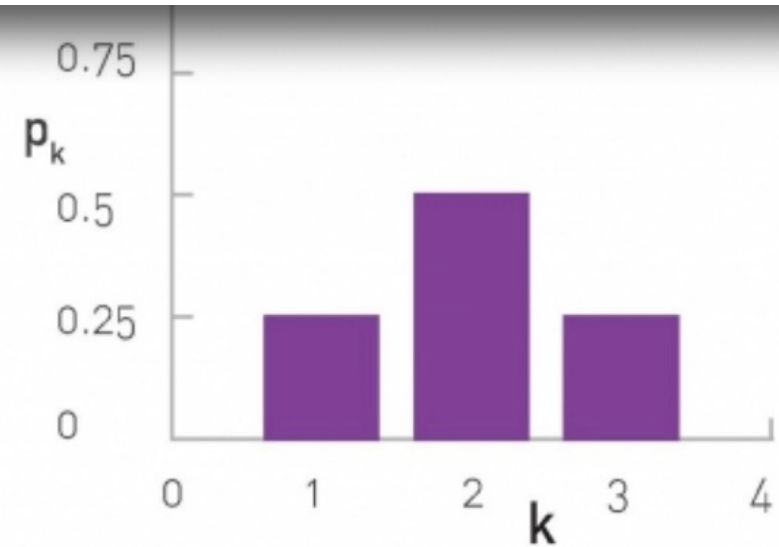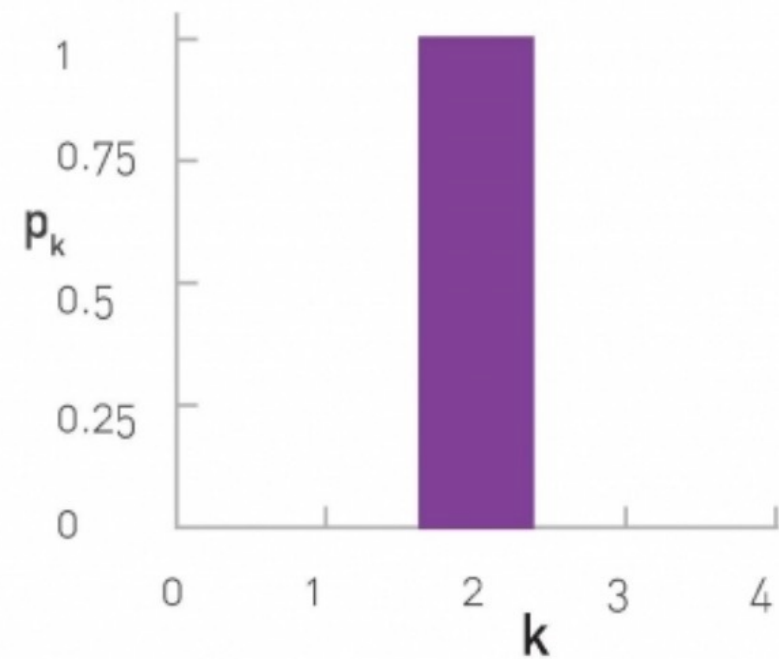**NetworkAnalyzer**
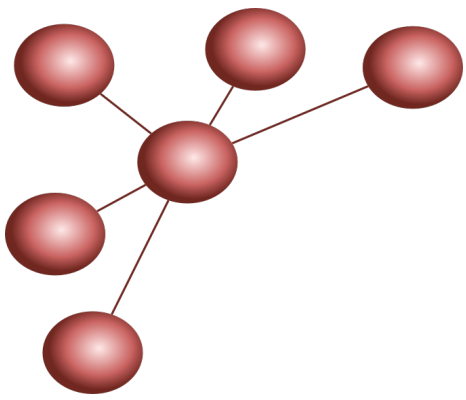Core App: Computes basic properties of whole network (degree distribution, clustering coefficients, centrality, etc.)

k = 4

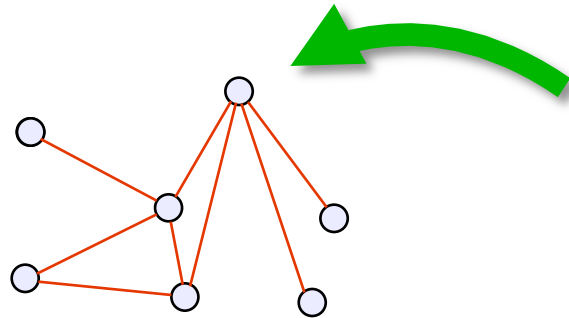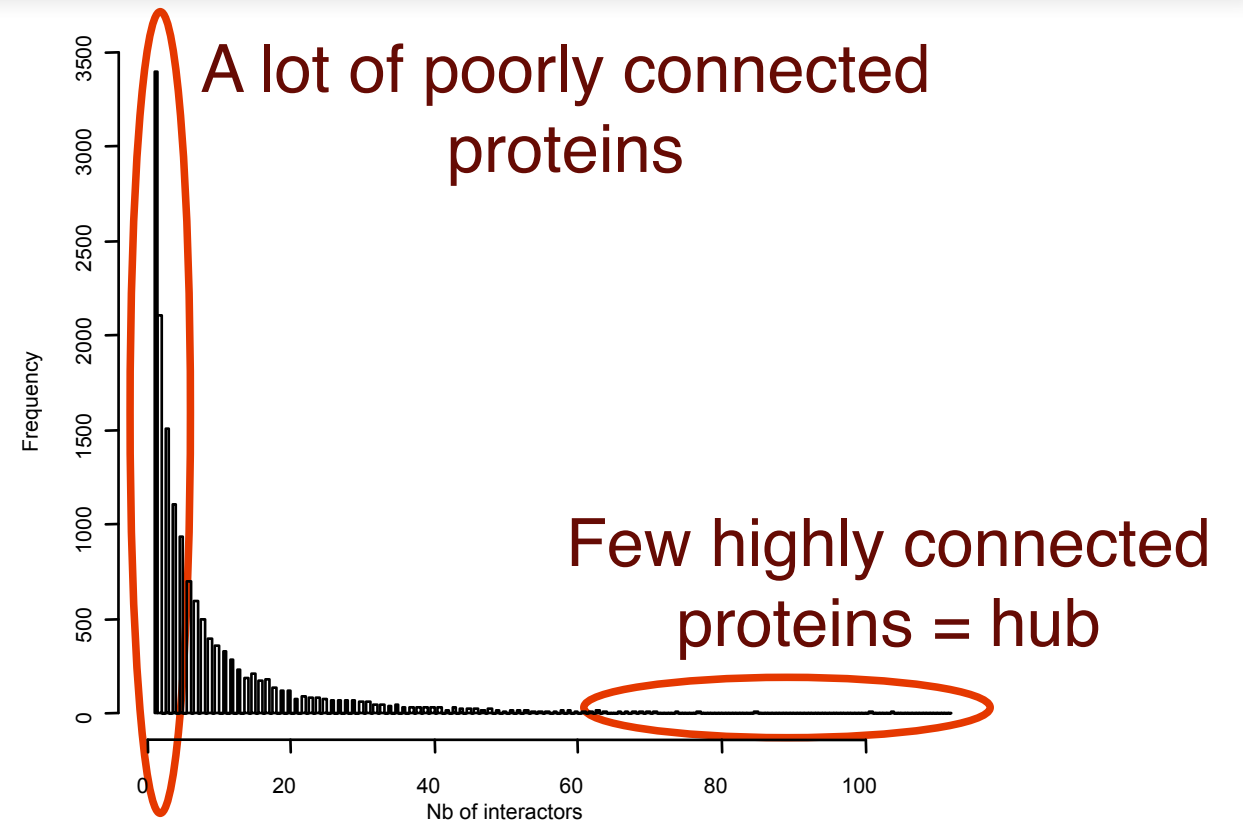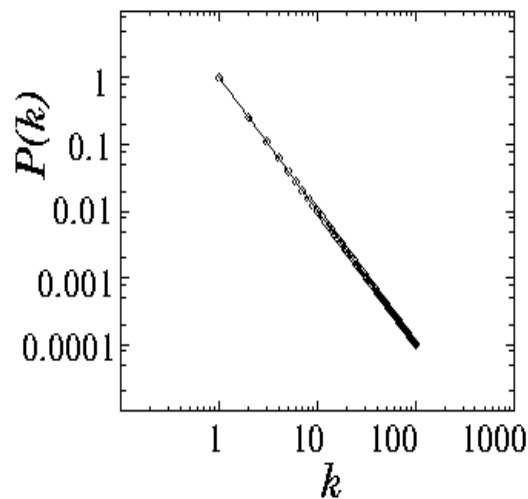A lot of poorly connected proteins

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

NetworkAnalyzer
Core App: Computes basi... ...bution,
clustering coefficie...

www.rsc.org/molecularbiosystems | Molecular BioSystems

k = 4

REVIEW

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

Gipsi Lima-Mendez* and Jacques van Helden*

Few highly connected proteins = hub

Nb of interactors

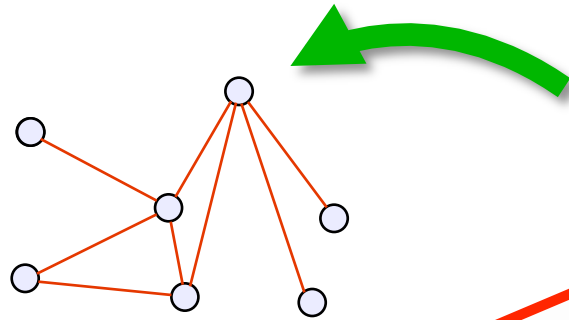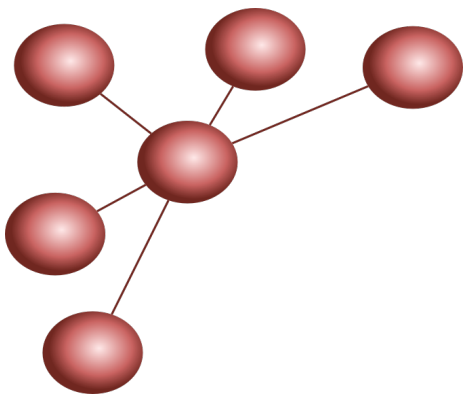0   20   40   60   80   100
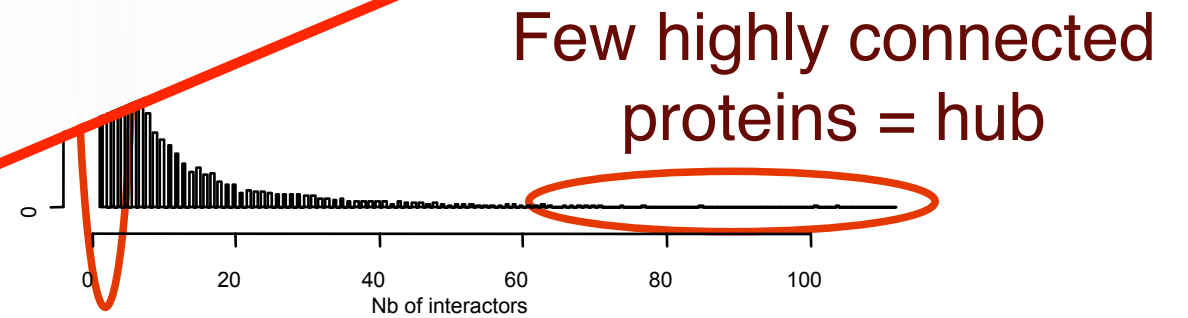
Biological interpretation?
Robust to random attack, sensitive to targeted attacks
Growth with preferential attachement ("rich get richer") => create "hubs"

Network topological structure : Small-world property

(a) Random network

(b) Scale-free network

- Milgram, 6 degrees of separation

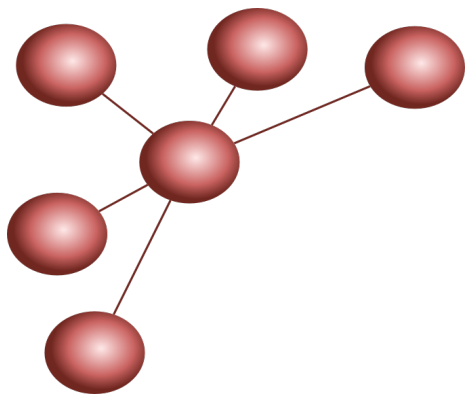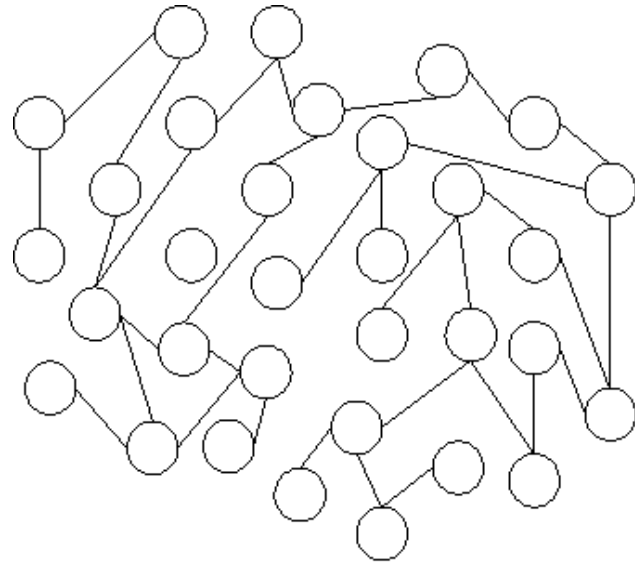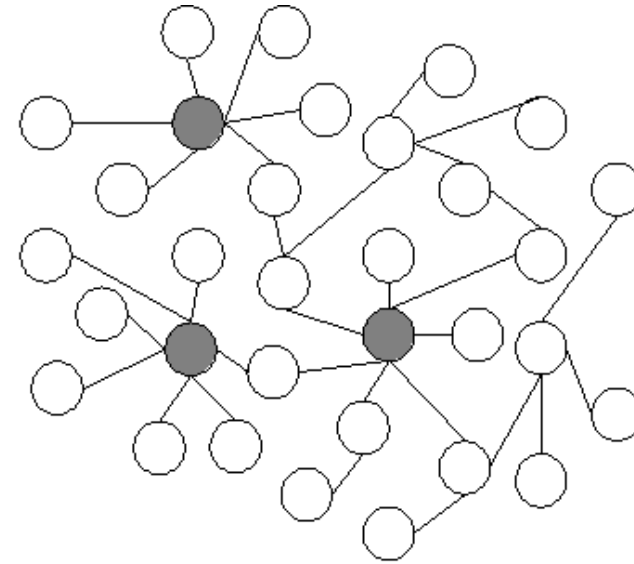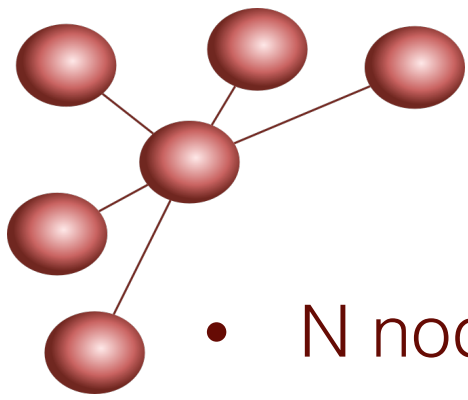# Metrics on graphs

- N nodes, V edges

- Network size

- Adjacency matrix

- Degree, degree distribution

- Path, shortest path, distances

- Connectivity, clustering coefficient
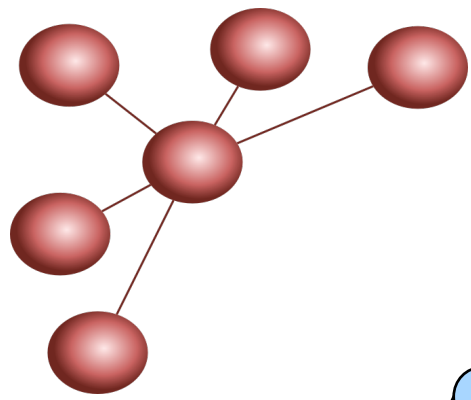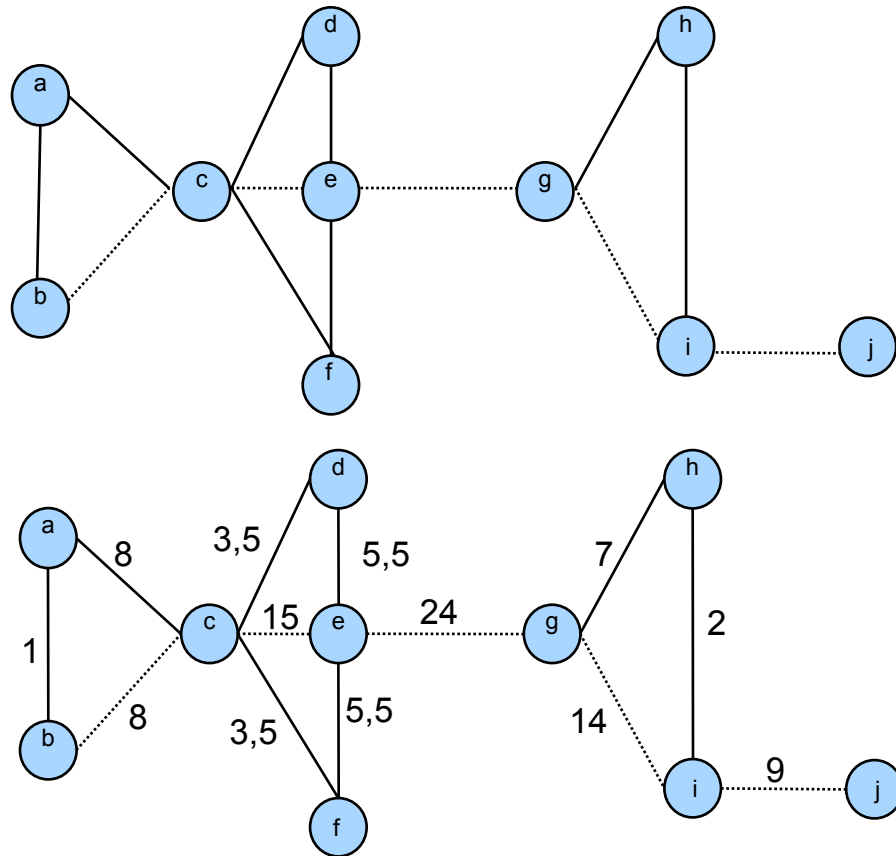
- Betweenness

- Motifs

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)

$C_A = \dfrac{\text{Actual links between A's neighbors (black)}}{\text{Possible links between A's neighbors (orange)}}$

k_A=Nb of edges through A=5

$C_A = n_A/[k_A(k_A-1)/2]$
$=2/[4\times(4-1)/2]=0.333$

$NC_A=(k_B+k_C+k_D+k_E+k_J)/5$
$=(5+2+2+3+1)/5=2.6$

$SP_{FH}=(F,D,A,B,H)=4$

B_4=Fraction of SPs passing through A
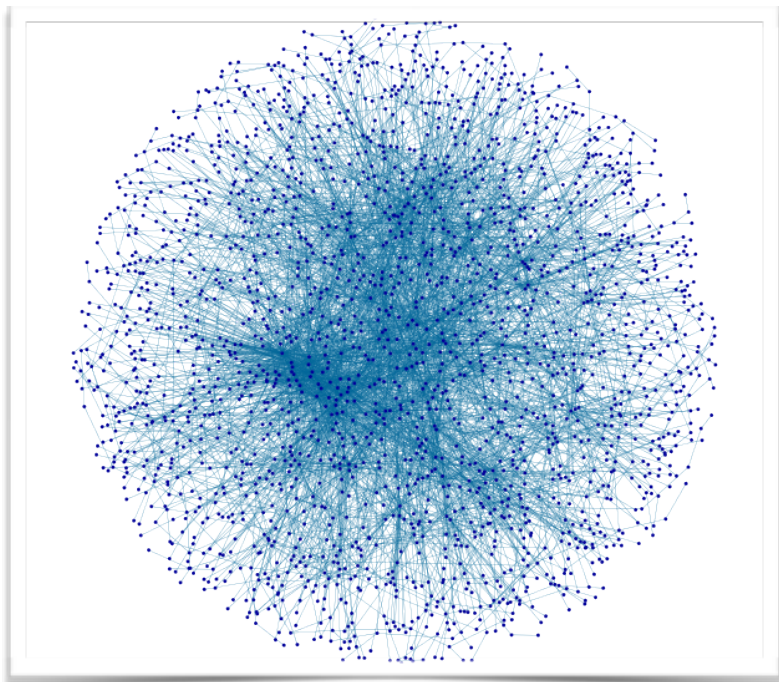$=0.090$

Gavin, Cell

"Betweenness"

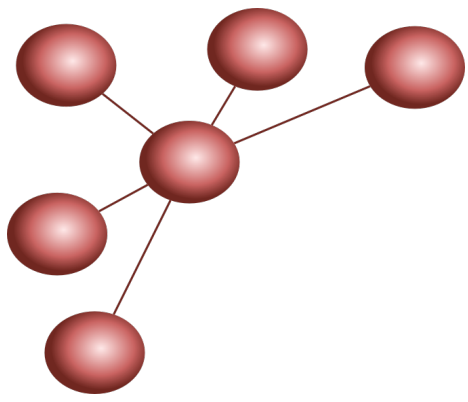## Number of shortest paths running through an edge
=
"bootleneck"

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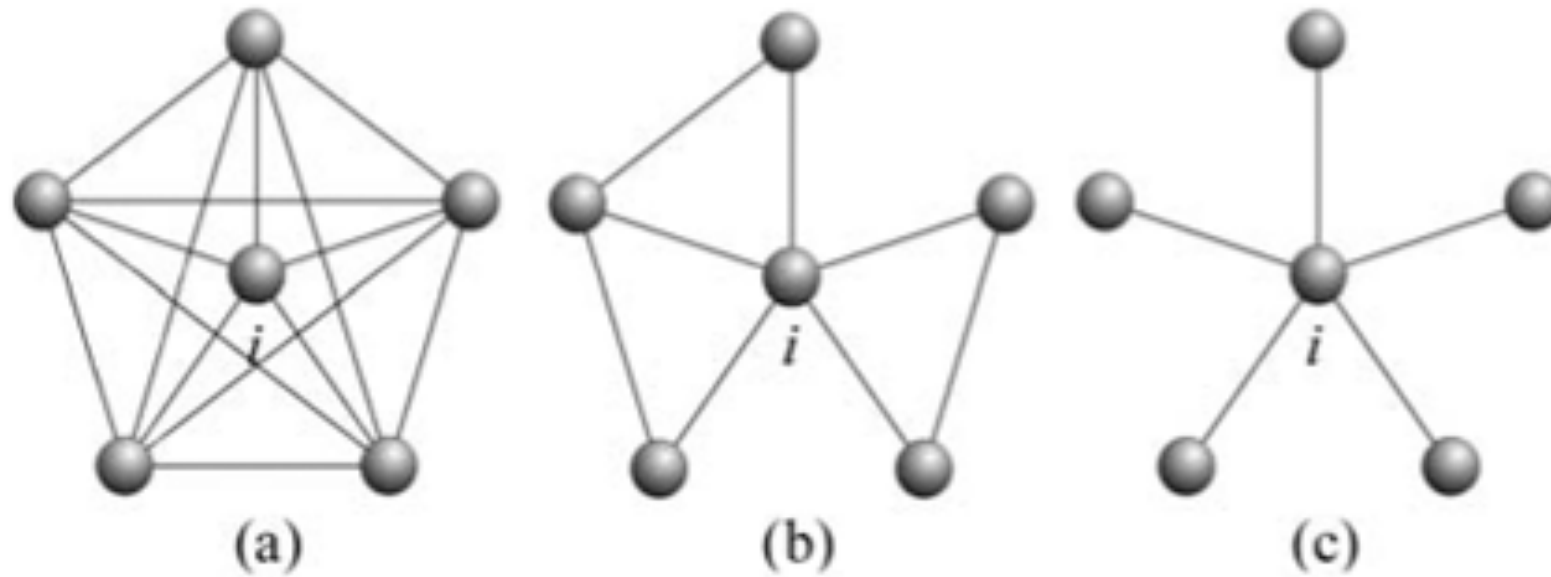**Correlation with gene essentiality, gene involvement in diseases, importance in flux transmission …**

NetworkAnalyzer

Core App: Computes basic properties of whole network (degree distribution, clustering coefficients, centrality, etc.)
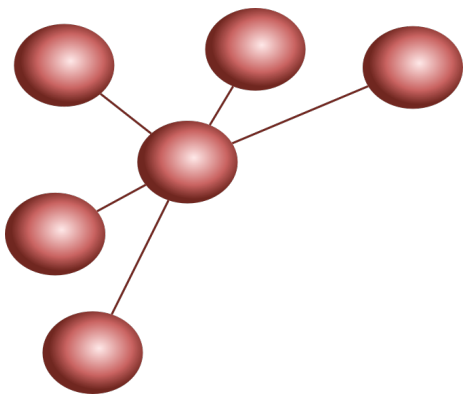
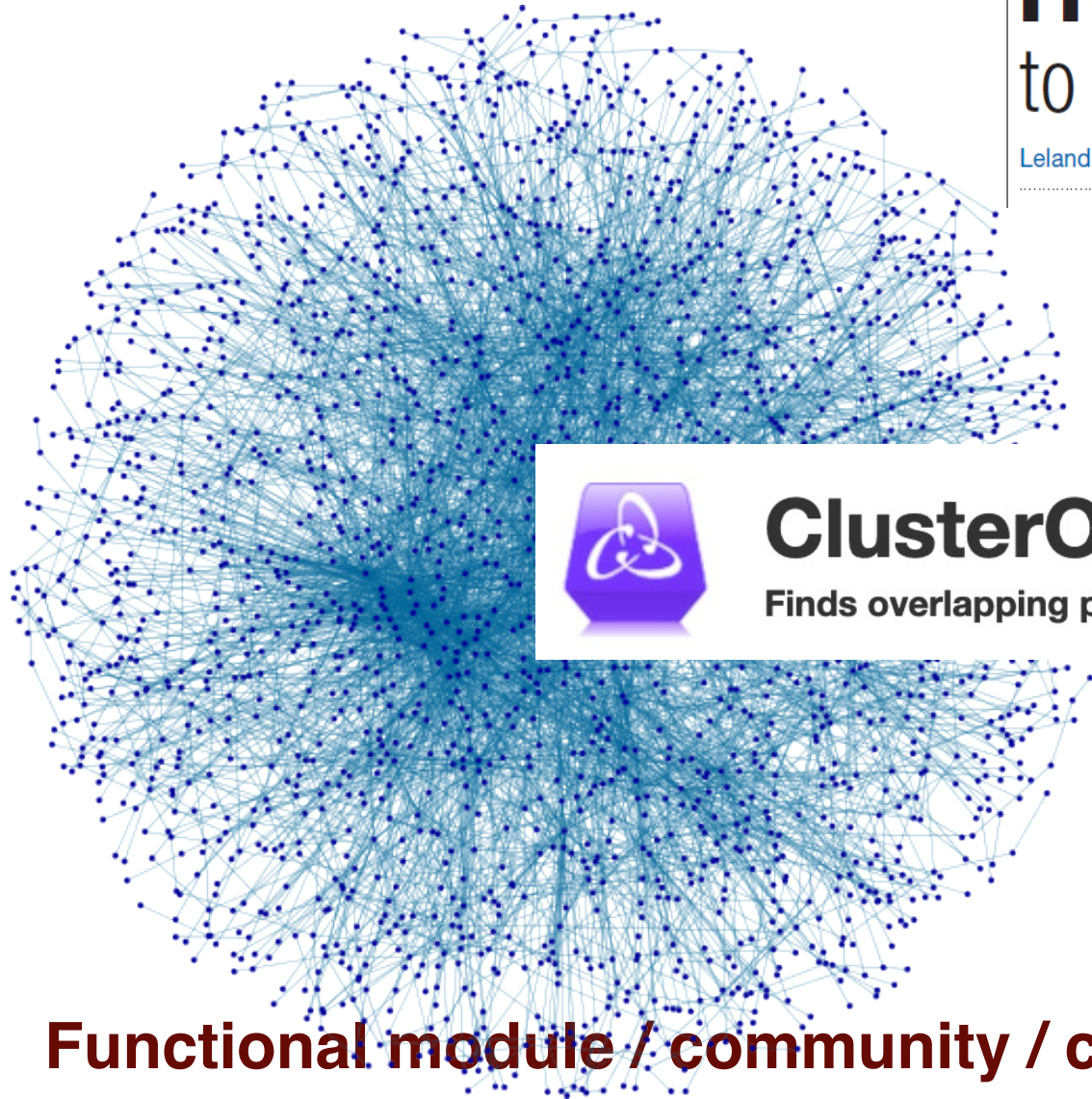# Clustering coefficient / modularity

(a) (b) (c)

Actual links between neighbours / Possible links between neighbours

Global approach - Clustering
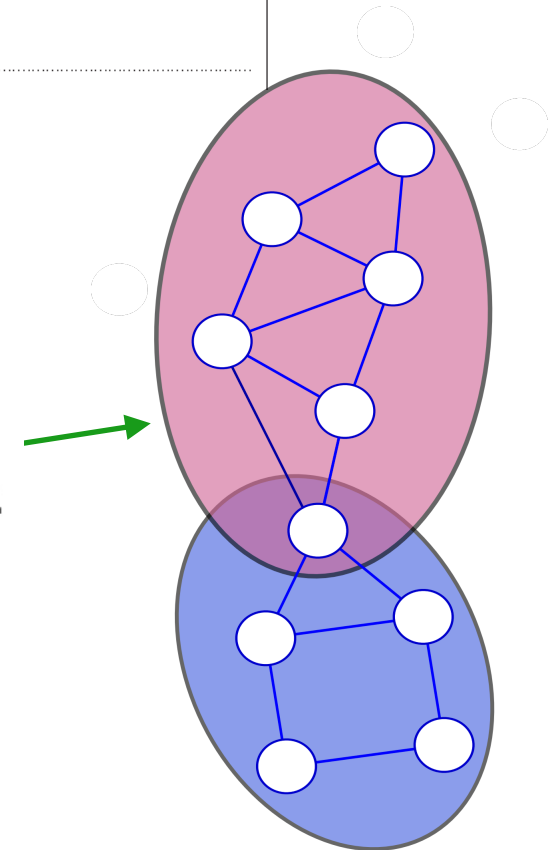
impacts

From molecular
to modular cell biology

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

ClusterONE
Finds overlapping protein complexes in a protein interaction network.

Functional module / community / cluster / class : discrete function

Modules can be isolated or connected

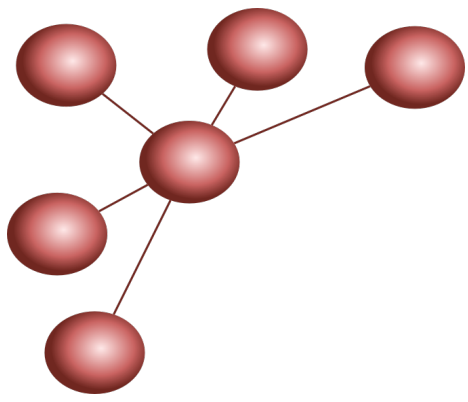**Groups of proteins involved in a common cellular function**

clusterMaker2
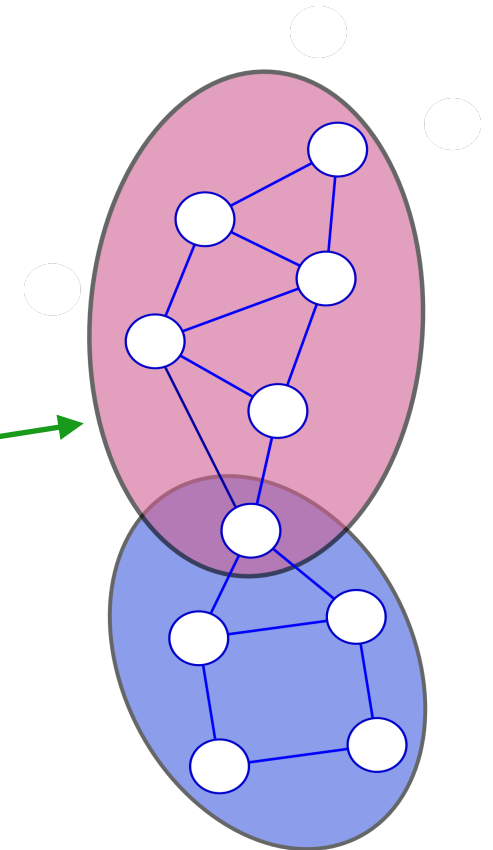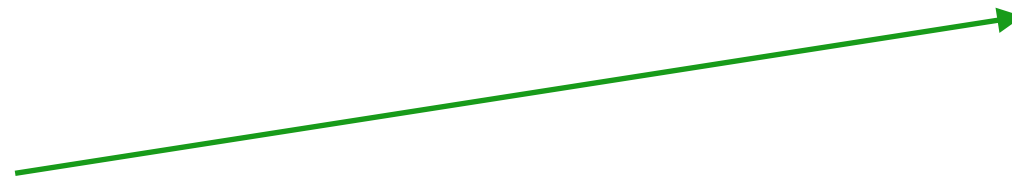Multi-algorithm clustering app for Cytoscape

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

Global approach - Clustering

ClusterONE
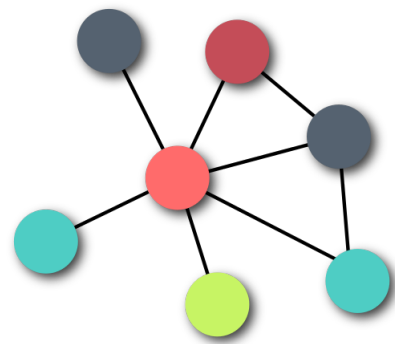Finds overlapping protein complexes in a protein interaction network.

clusterMaker2
Multi-algorithm clustering app for Cytoscape

MCODE
Clusters a given network based on topology to find densely connected regions.
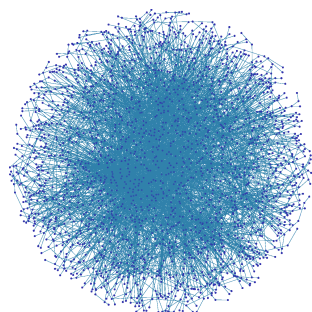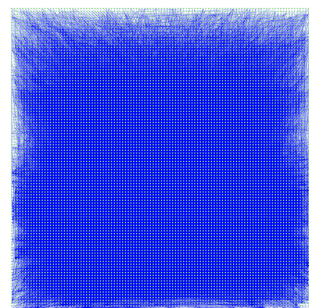
Integration of networks

- Diverse interaction sources
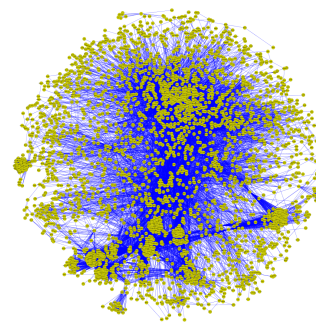
- Each own features, topology, bias

PPI	Co-expr	Pathways	Complexes
~60 000 edges	~1 400 000 edges	~250 000 edges	~40 000 edges

How do we combine many networks / interaction sources?