

# Biological networks

(BIO 390, 08.11.2022)



University of  
Zurich<sup>UZH</sup>

**Session 1 (45 min): Examples of biological networks**

**Session 2 (45 min): Properties of networks**



**Instructor:** Pouria Dasmeh

Postdoctoral Fellow, Wagner Lab., University of Zurich.

Long-term research fellow, Shakhnovich Lab., Harvard University.



<https://github.com/dasmeh>



@PouriaDasmeh

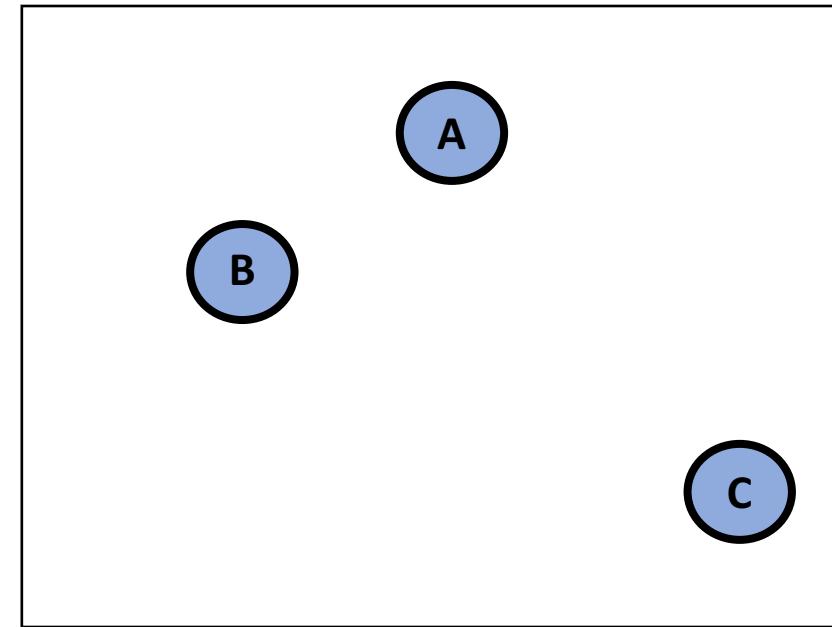
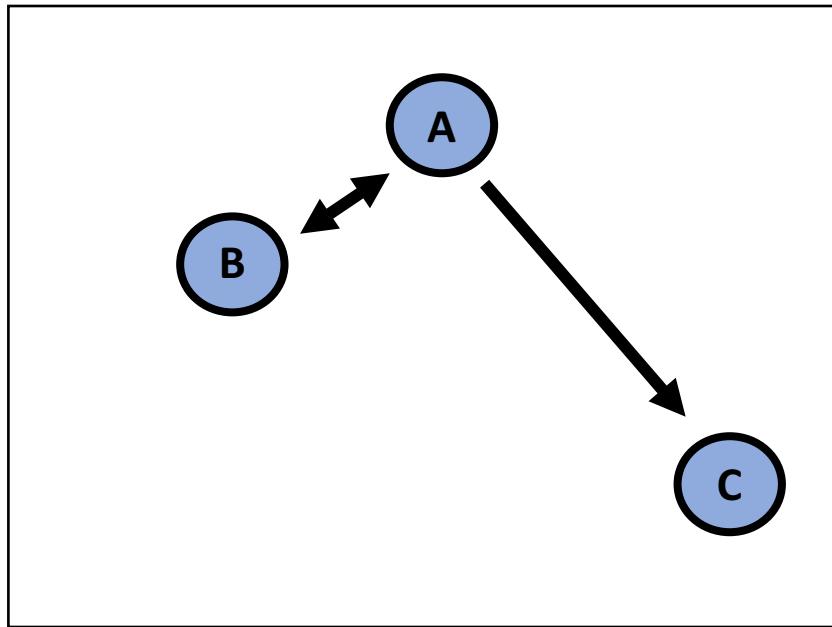
Pouria.dasmeh@uzh.ch

*Session 1*

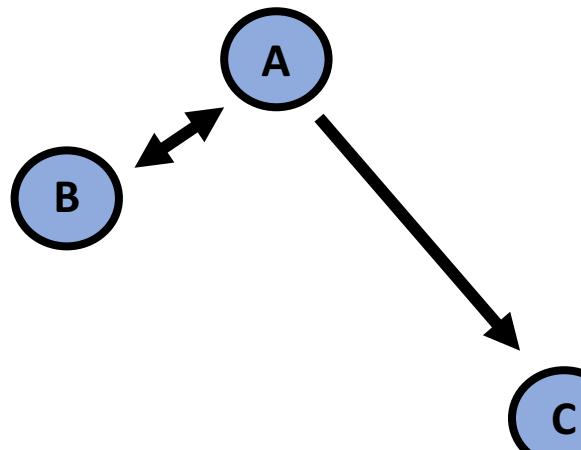
## Examples of biological networks

# What is a network?

Two or more components that can “interact” with each other



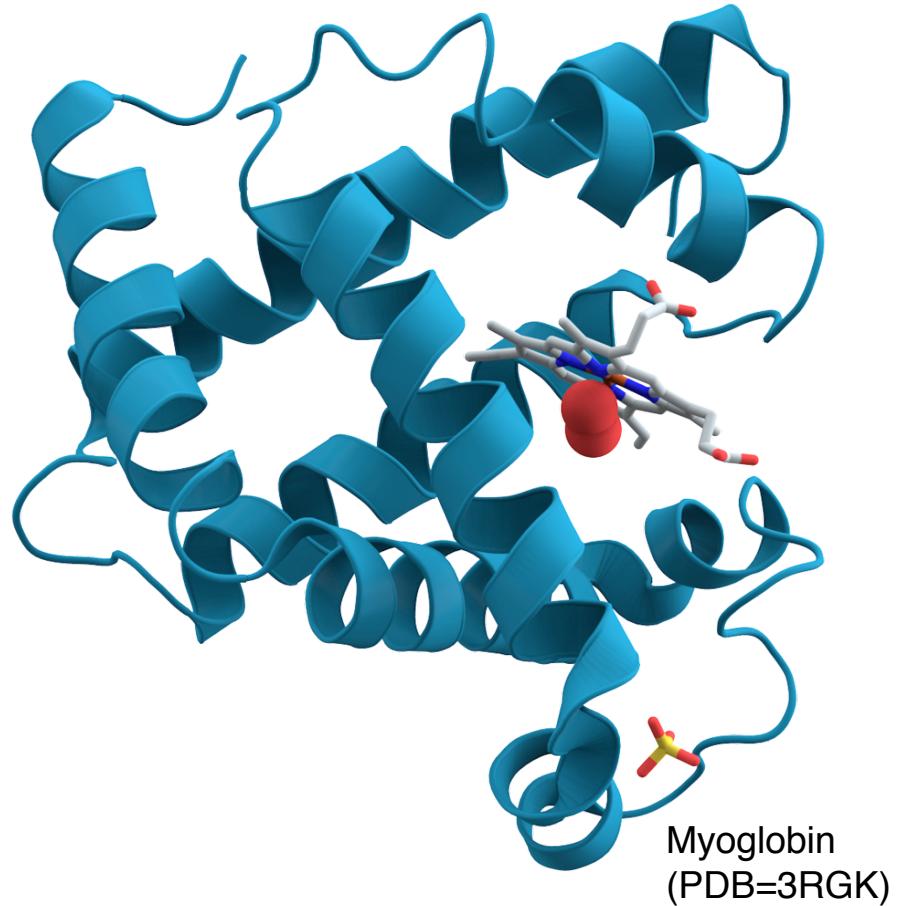
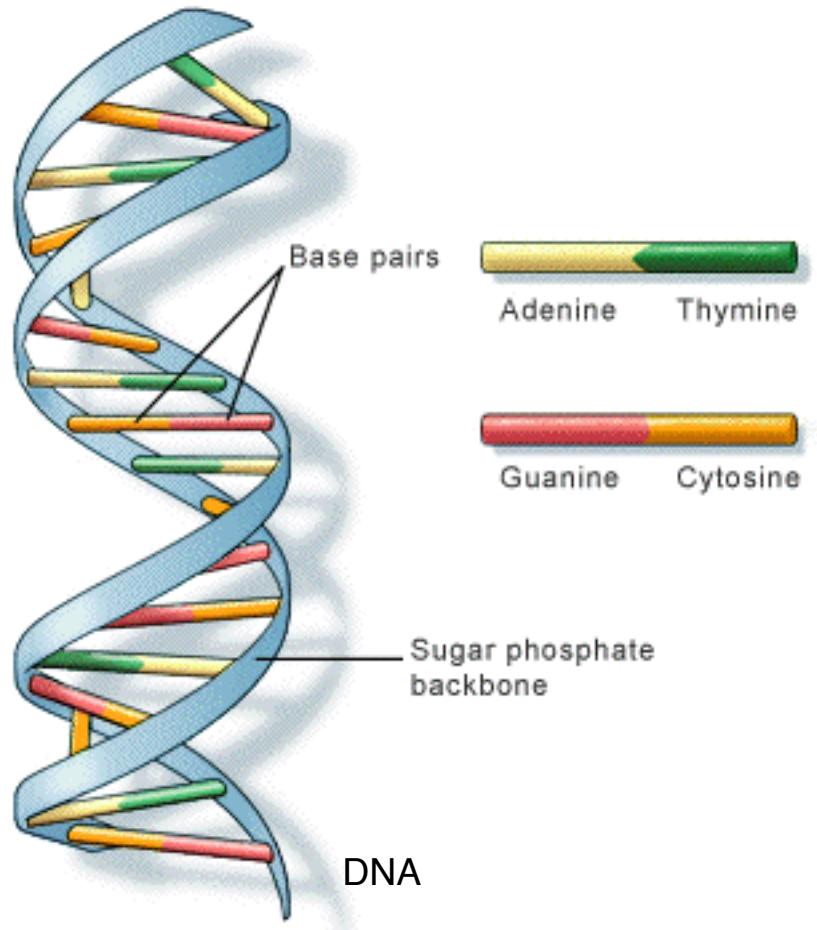
# Biological networks



Networks that apply to biological systems:

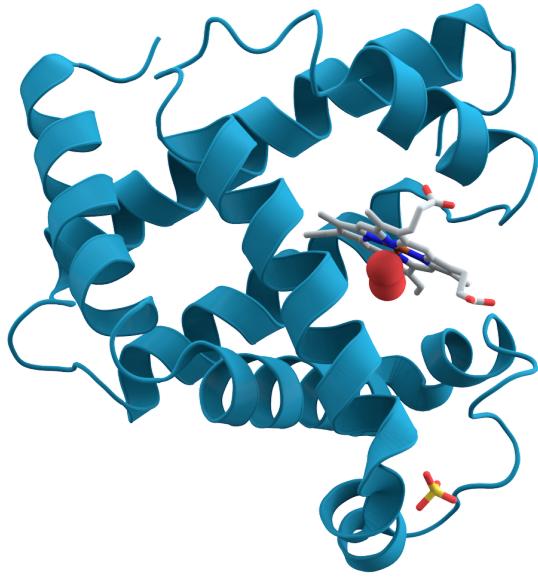
- 1) DNA
- 2) Protein
- 3) Cell
- 4) Individual organisms
- 5) Communities
- 6) ....

# DNA and Proteins

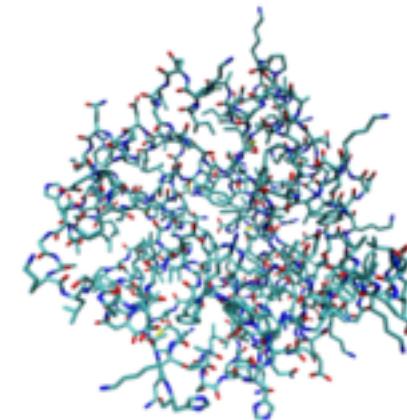
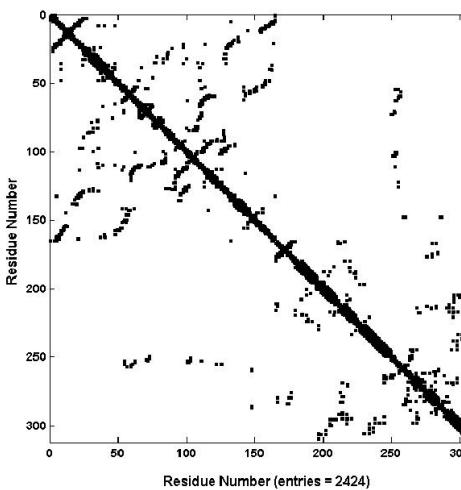


Chemical structures (networks)

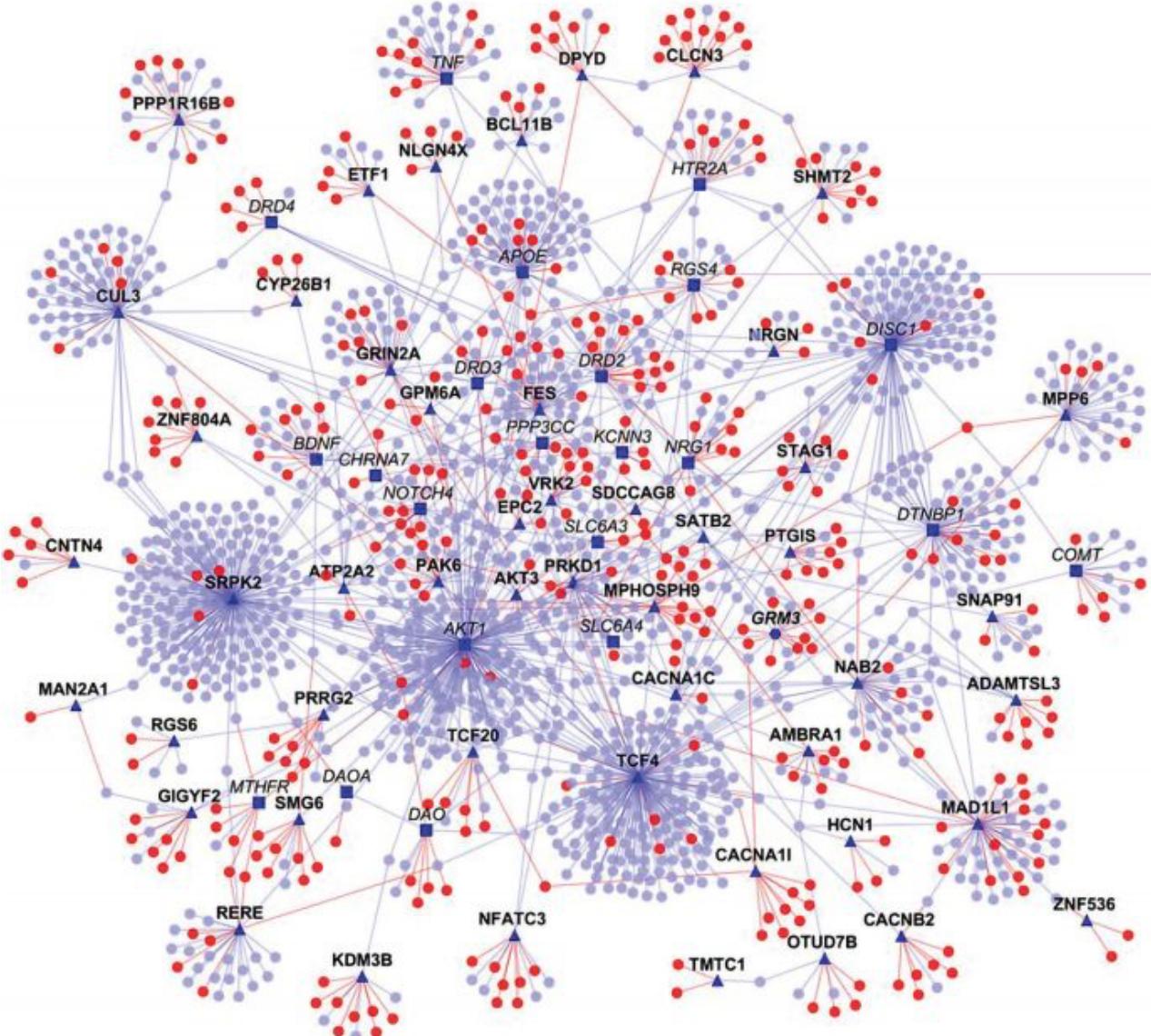
## Chemical vs. Biological views



Chemical structures  
Contact map  
(Biological)

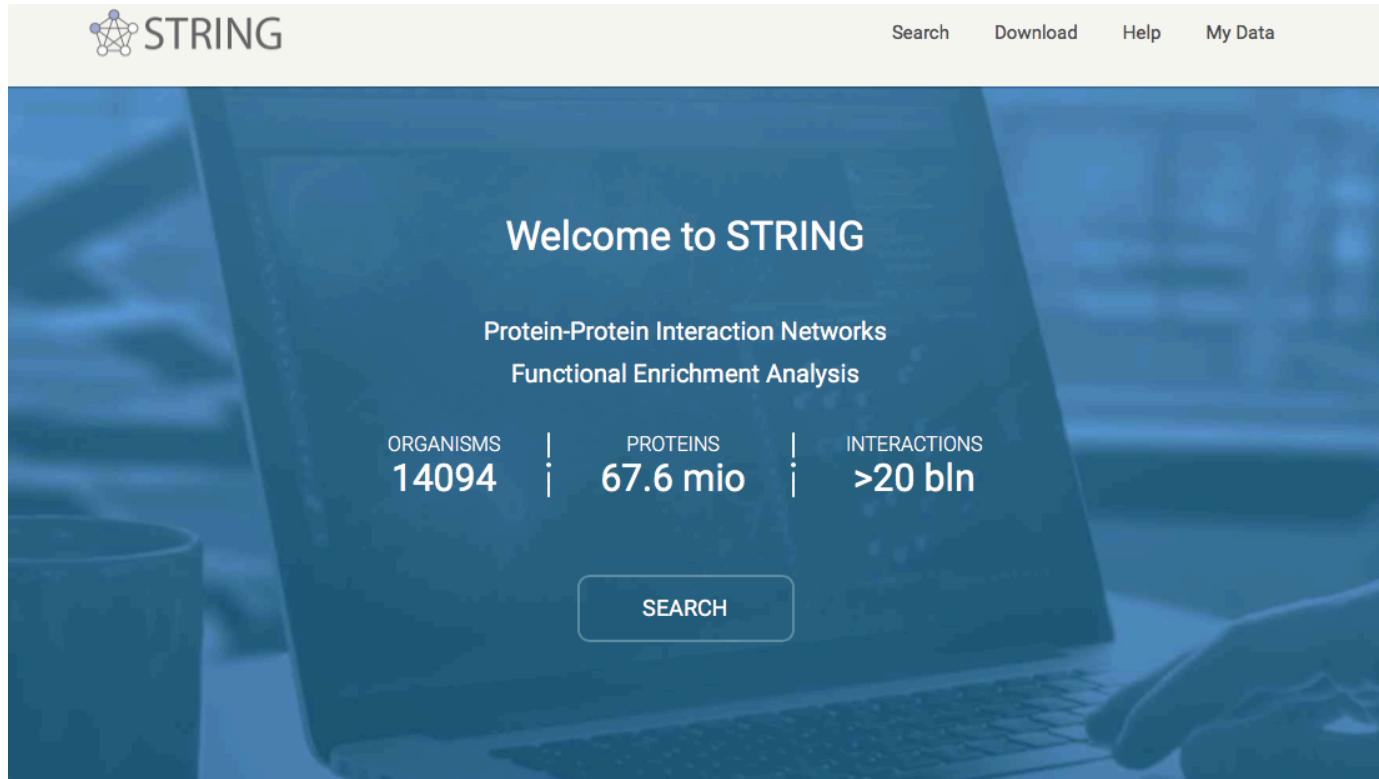


# Protein-protein interaction network



Example:  
Pol II complex  
(PDB=1Y1W)

# Protein-protein interaction network



The screenshot shows the STRING homepage. At the top, there's a navigation bar with links for Search, Download, Help, and My Data. The main header "Welcome to STRING" is followed by a subtitle "Protein-Protein Interaction Networks Functional Enrichment Analysis". Below this, key statistics are displayed: ORGANISMS 14094, PROTEINS 67.6 mio, and INTERACTIONS >20 bln. A large "SEARCH" button is centered below these stats. The footer contains copyright information for the STRING Consortium 2021, logos for SIB, CPR, and EMBL, and links to various sections like About, Info, Access, and Credits.

© STRING CONSORTIUM 2021

ABOUT    INFO    ACCESS    CREDITS

SIB - Swiss Institute of Bioinformatics

CPR - Novo Nordisk Foundation Center Protein Research

EMBL - European Molecular Biology Laboratory

Content    Scores    Versions    Funding

References    Use scenarios    APIs    Datasources

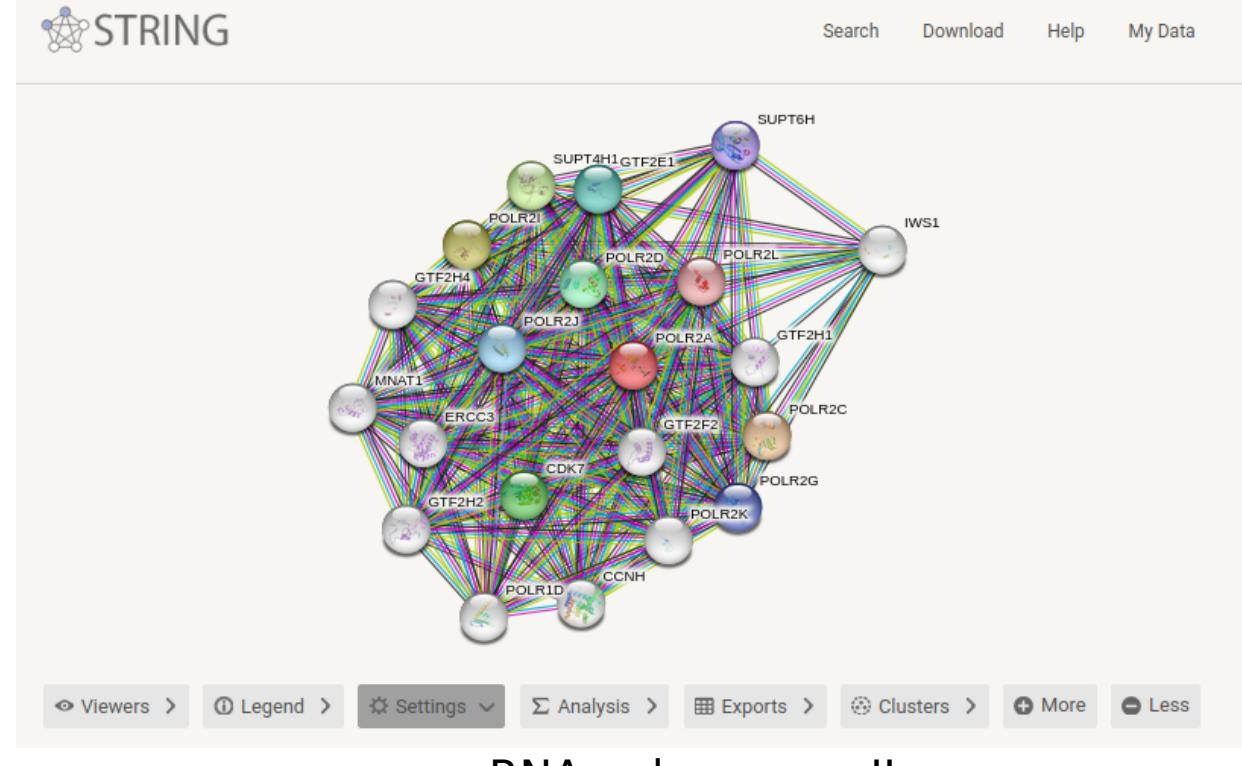
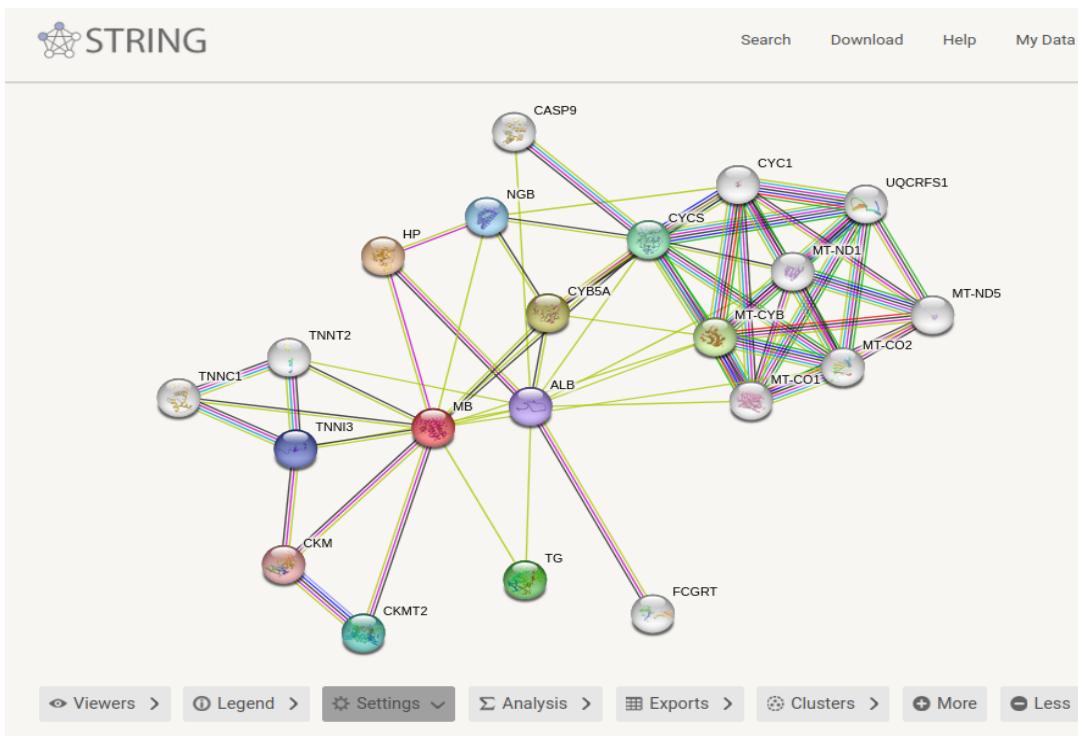
Contributors    FAQs    Licensing    Partners

Statistics    Cookies/Privacy    Usage    Software

Compare the interaction networks of two proteins in human:

- Myoglobin (Mb)
- DNA-directed RNA polymerase II (POLR2A)

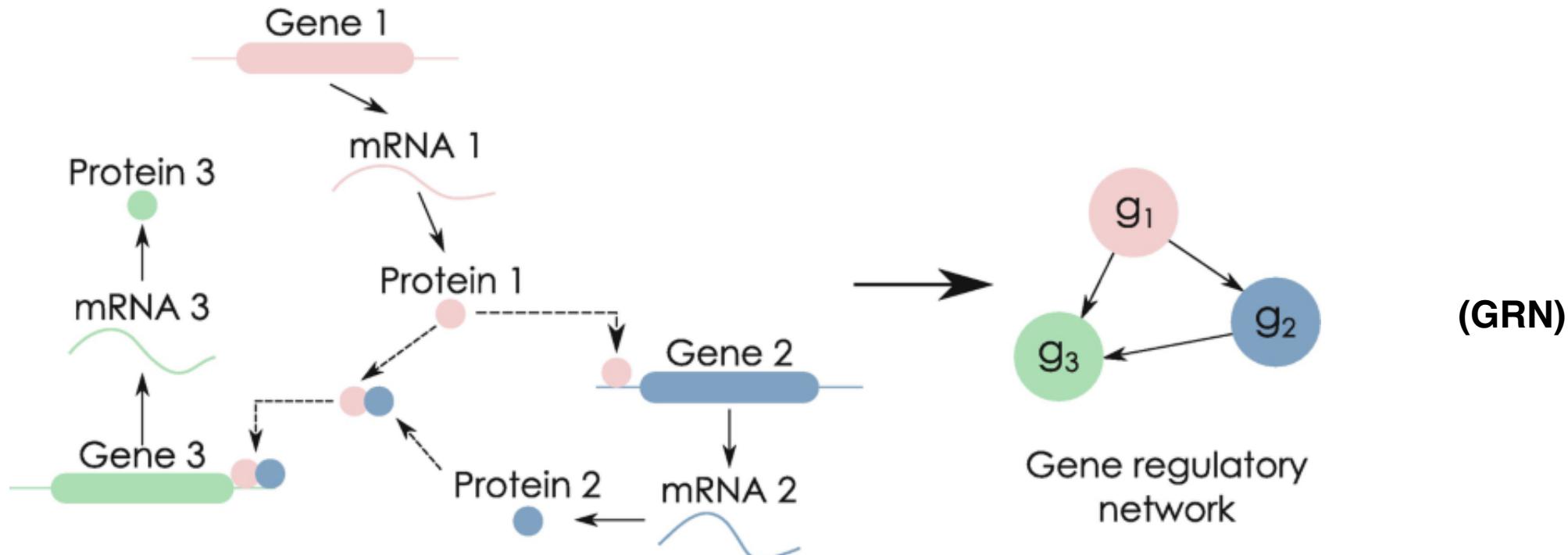
# Protein-protein interaction network



Why do we have different tissues  
and cell types if we have only one DNA  
molecule?



# Gene regulatory networks (DNA-protein interaction networks)

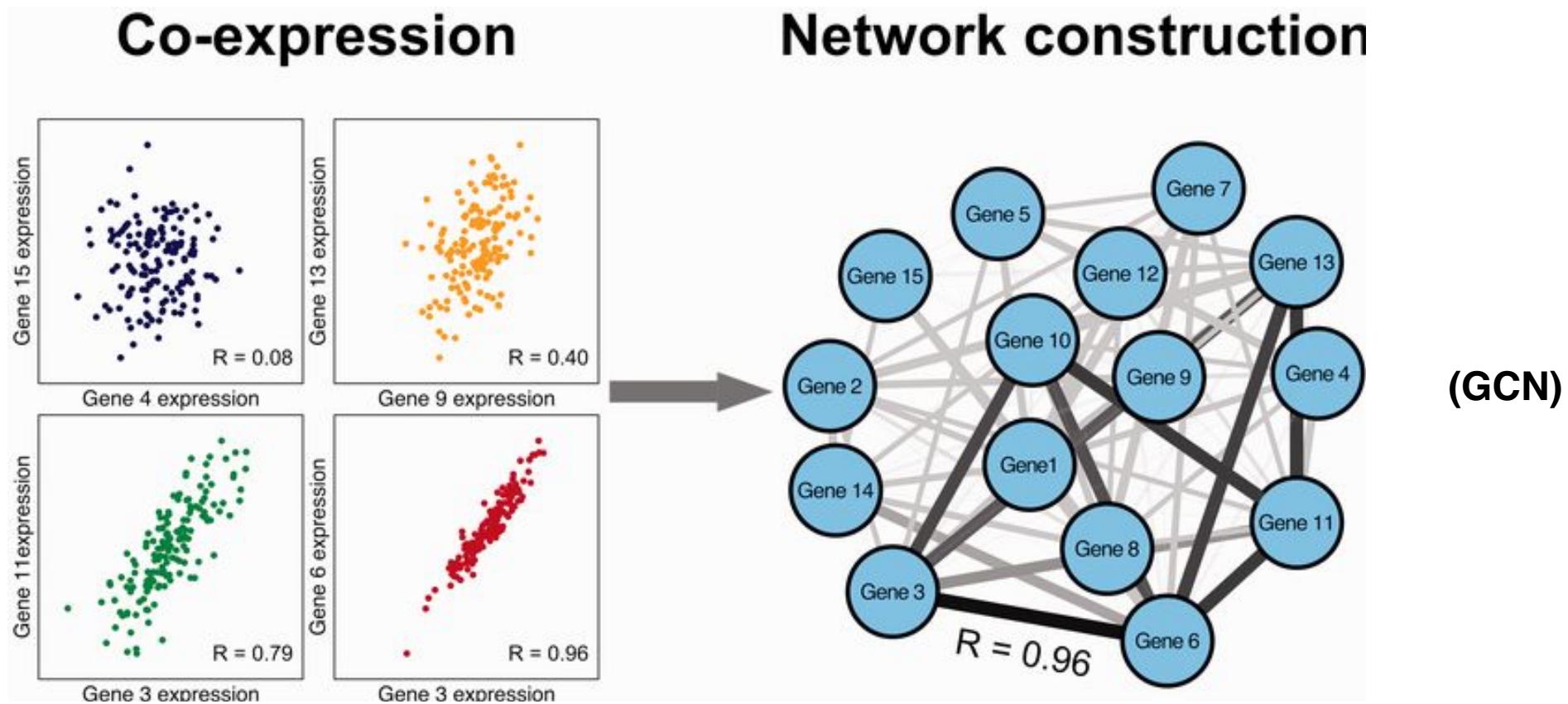


A set of molecular **regulators** that interact with each other to govern the **gene expression** levels of mRNA and proteins >> the **function** of the **cell**.

Responses to stress often involve co-expression of many genes. How do cells achieve this coordinated response?



# Gene co-expression network



A gene co-expression network can be built by finding pairs of genes which show a similar expression pattern across samples.

# Gene co-expression network

RNA-seq  
or  
Microarray  
experiments



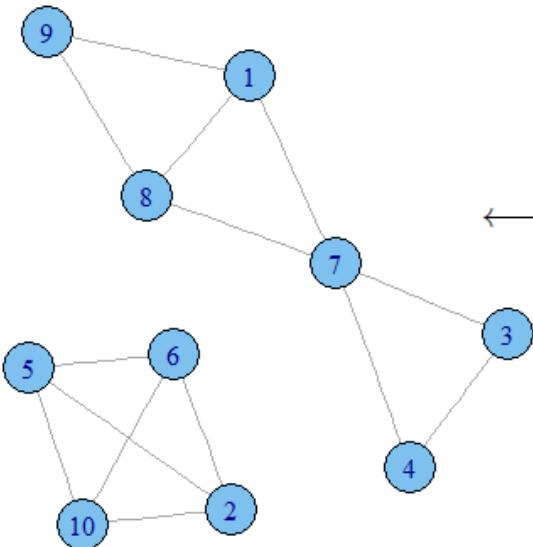
	$S_1$	$S_2$	$S_3$
$G_1$	43.26	40.89	5.05
$G_2$	166.6	41.87	136.65
$G_3$	12.53	39.55	42.09
$G_4$	28.77	191.92	236.56
$G_5$	114.7	79.7	99.76
$G_6$	119.1	80.57	114.59
$G_7$	118.9	156.69	186.95
$G_8$	3.76	2.48	136.78
$G_9$	32.73	11.99	118.8
$G_{10}$	17.46	56.11	21.41

Gene expression values

$|r(G_i, G_j)|$   
 Pearson  
 correlation

	$G_1$	$G_2$	$G_3$	$G_4$	$G_5$	$G_6$	$G_7$	$G_8$	$G_9$	$G_{10}$
$G_1$	1.00	0.23	0.61	0.71	0.03	0.35	<b>0.86</b>	<b>1.00</b>	<b>0.97</b>	0.37
$G_2$	0.23	1.00	0.63	0.52	<b>0.98</b>	<b>0.99</b>	0.29	0.30	0.46	<b>0.99</b>
$G_3$	0.61	0.63	1.00	<b>0.99</b>	0.77	0.53	<b>0.93</b>	0.56	0.41	0.51
$G_4$	0.71	0.52	<b>0.99</b>	1.00	0.69	0.41	<b>0.97</b>	0.66	0.52	0.40
$G_5$	0.03	<b>0.98</b>	0.77	0.69	1.00	<b>0.95</b>	0.48	0.09	0.27	<b>0.94</b>
$G_6$	0.35	<b>0.99</b>	0.53	0.41	<b>0.95</b>	1.00	0.17	0.41	0.57	<b>1.00</b>
$G_7$	0.86	0.29	<b>0.93</b>	<b>0.97</b>	0.48	0.17	1.00	<b>0.83</b>	0.72	0.16
$G_8$	<b>1.00</b>	0.30	0.56	0.66	0.09	0.41	0.83	1.00	<b>0.98</b>	0.42
$G_9$	<b>0.97</b>	0.46	0.41	0.52	0.27	0.57	0.72	<b>0.98</b>	1.00	0.58
$G_{10}$	0.37	<b>0.99</b>	0.51	0.40	<b>0.94</b>	<b>1.00</b>	0.16	0.42	0.58	1.00

Similarity (Co-expression) score



	$G_1$	$G_2$	$G_3$	$G_4$	$G_5$	$G_6$	$G_7$	$G_8$	$G_9$	$G_{10}$
$G_1$	0	0	0	0	0	0	1	1	1	0
$G_2$	0	0	0	0	1	1	0	0	0	1
$G_3$	0	0	0	1	0	0	1	0	0	0
$G_4$	0	0	1	0	0	0	1	0	0	0
$G_5$	0	1	0	0	0	1	0	0	0	1
$G_6$	0	1	0	0	1	0	0	0	0	1
$G_7$	1	0	1	1	0	0	0	1	0	0
$G_8$	1	0	0	0	0	0	1	0	1	0
$G_9$	1	0	0	0	0	0	0	1	0	0
$G_{10}$	0	1	0	0	1	1	0	0	0	0

$|r(G_i, G_j)| \geq 0.8$   
 Significance threshold

Network adjacency matrix

## Metabolic networks

A set of chemical reactions that produces:

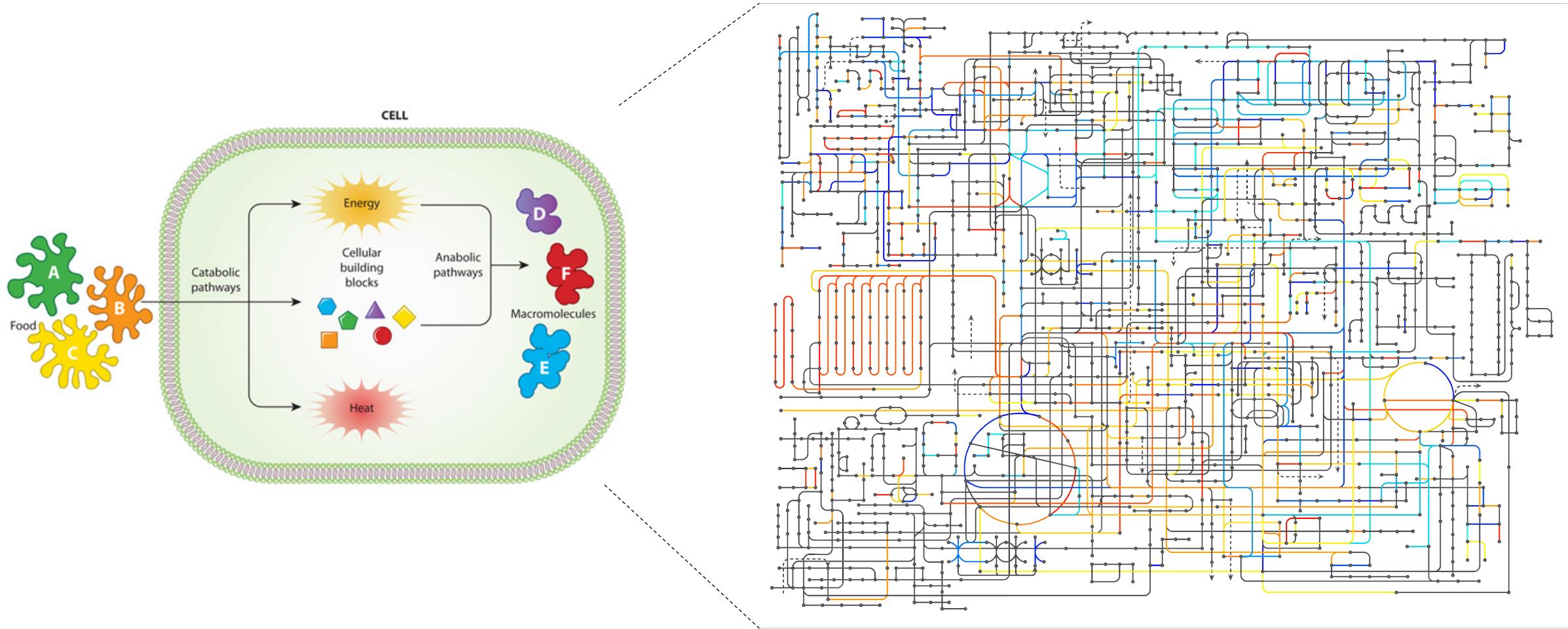
1) energy

(for maintenance of cell functions and for biosyntheses)

2) molecular building blocks for biosyntheses

These reactions are catalyzed by enzymes that are encoded by genes.

# Metabolic networks



# Metabolic networks



**KEGG PATHWAY Database**  
Wiring diagrams of molecular interactions, reactions and relations

**KEGG2 PATHWAY BRITE MODULE KO GENES COMPOUND DISEASE DRUG**

Select prefix  Enter keywords  Go Help

[ New pathway maps | Update history ]

### Pathway Maps

KEGG PATHWAY is a collection of manually drawn pathway maps representing our knowledge of the molecular interaction, reaction and relation networks for:

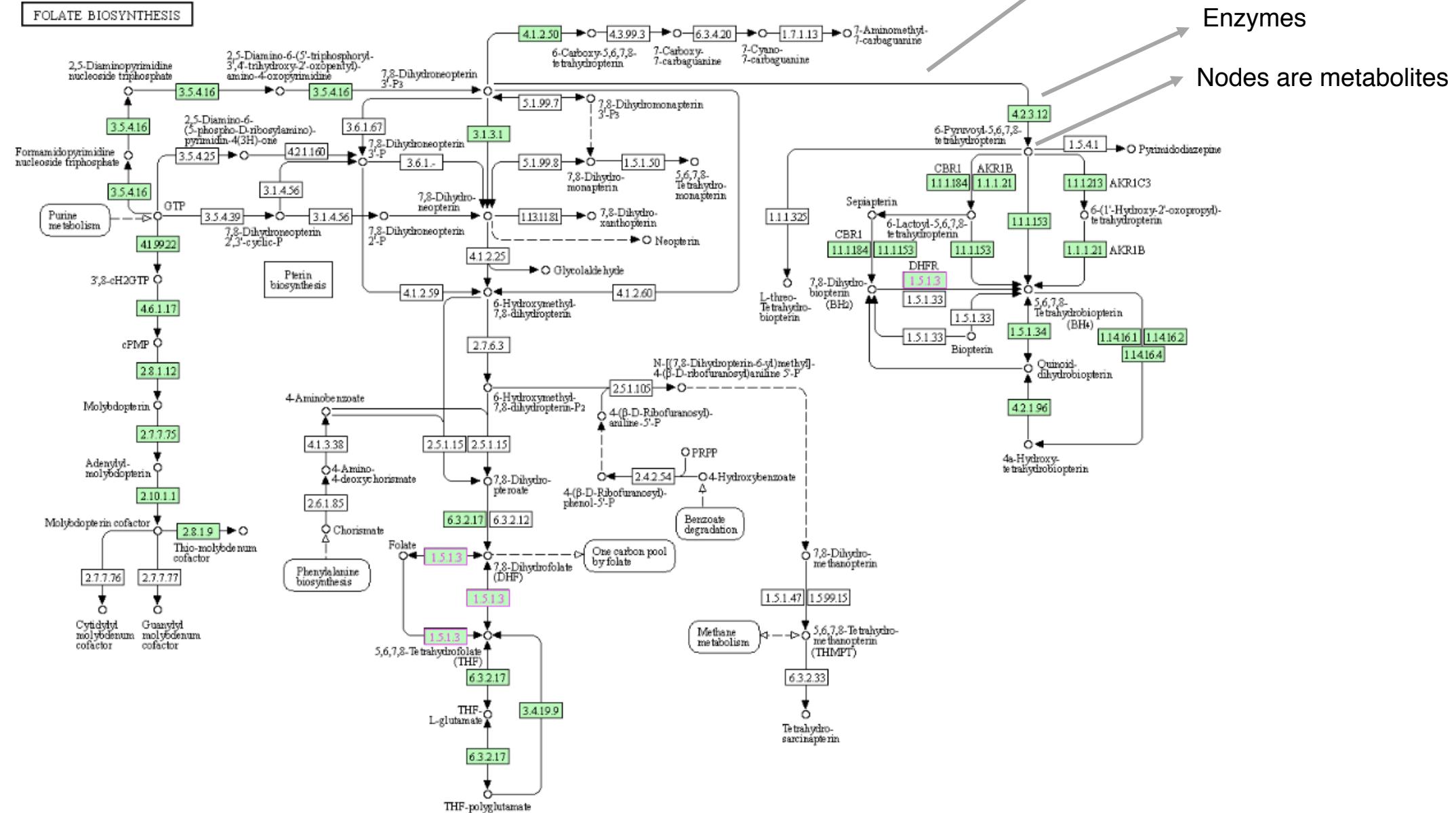
- 1. Metabolism**  
Global/overview Carbohydrate Energy Lipid Nucleotide Amino acid Other amino Glycan Cofactor/vitamin Terpenoid/PK Other secondary metabolite Xenobiotics Chemical structure
- 2. Genetic Information Processing**
- 3. Environmental Information Processing**
- 4. Cellular Processes**
- 5. Organismal Systems**
- 6. Human Diseases**
- 7. Drug Development**

KEGG PATHWAY is the reference database for pathway mapping in **KEGG Mapper**.

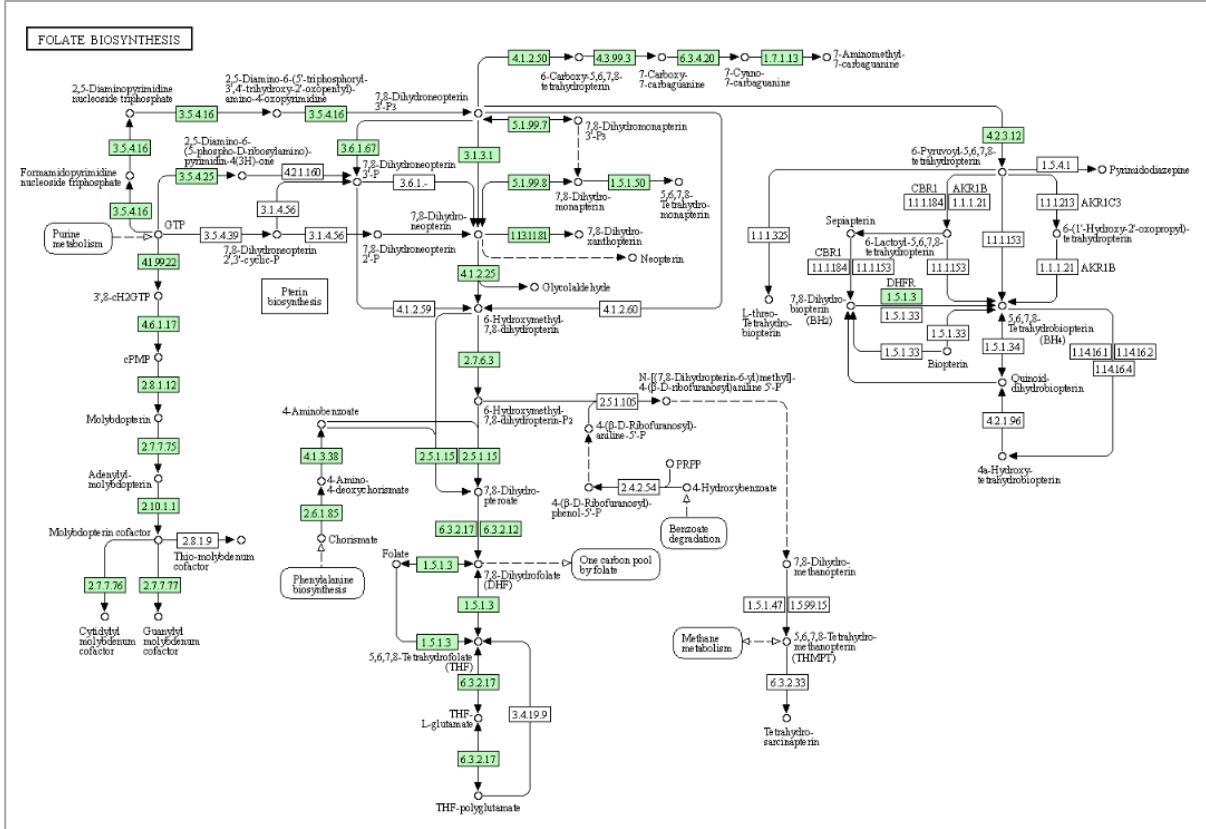
Find the folate biosynthetic pathway  
in Human and *E.coli*

## Database 2: KEGG

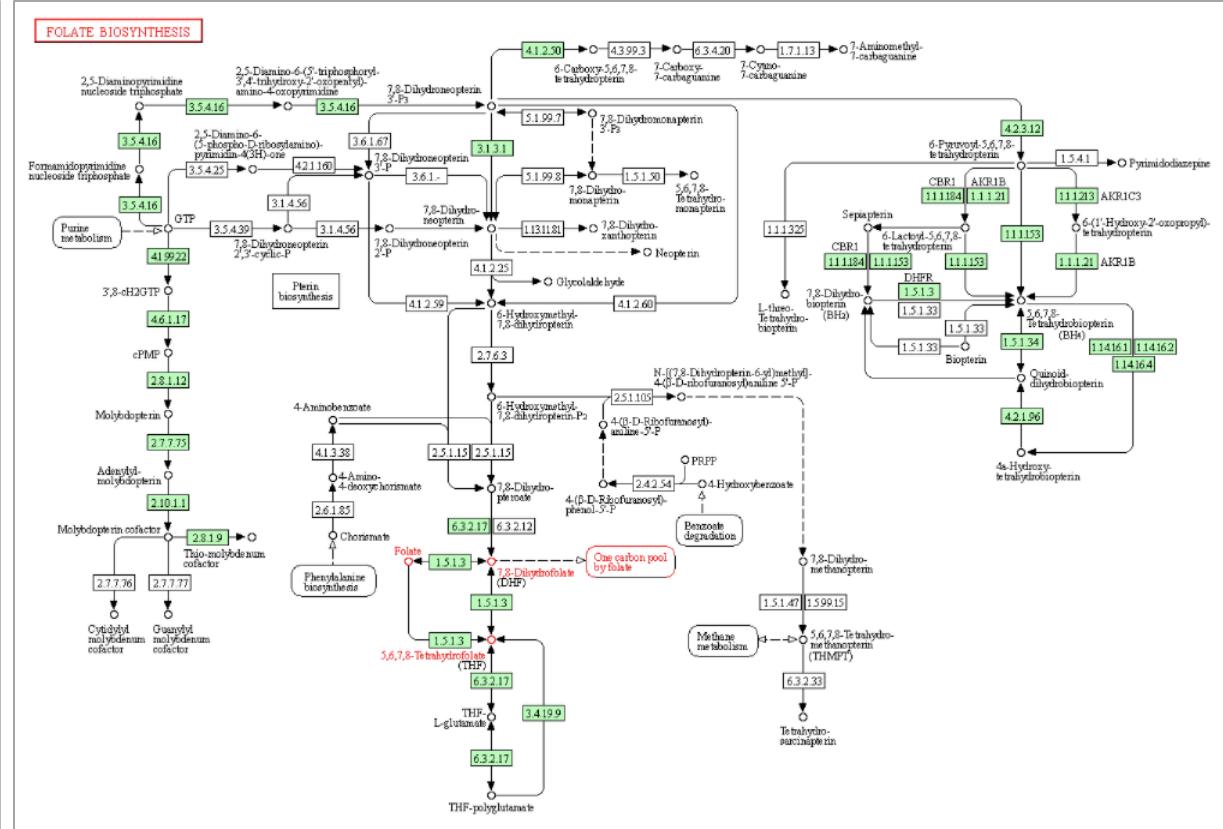
# Folate pathway



# Folate pathway in E.coli and Human

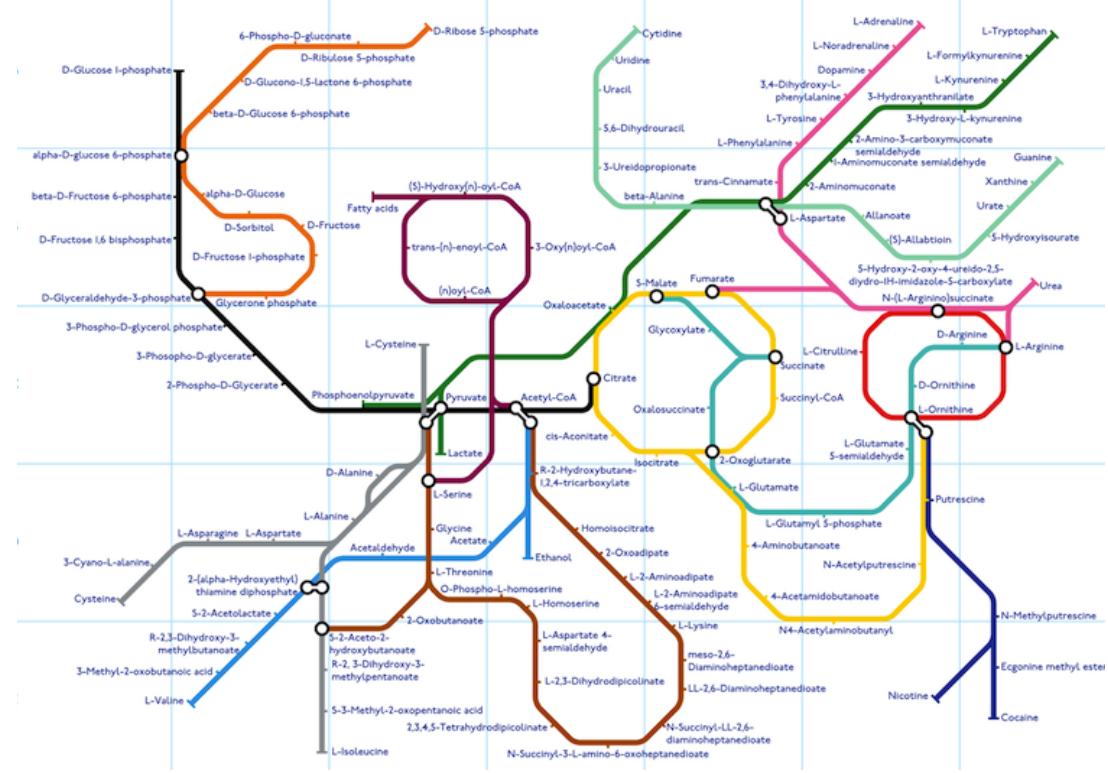


*E.coli*



*Human*

# Biological and other complex networks have similar structures



Metabolic map

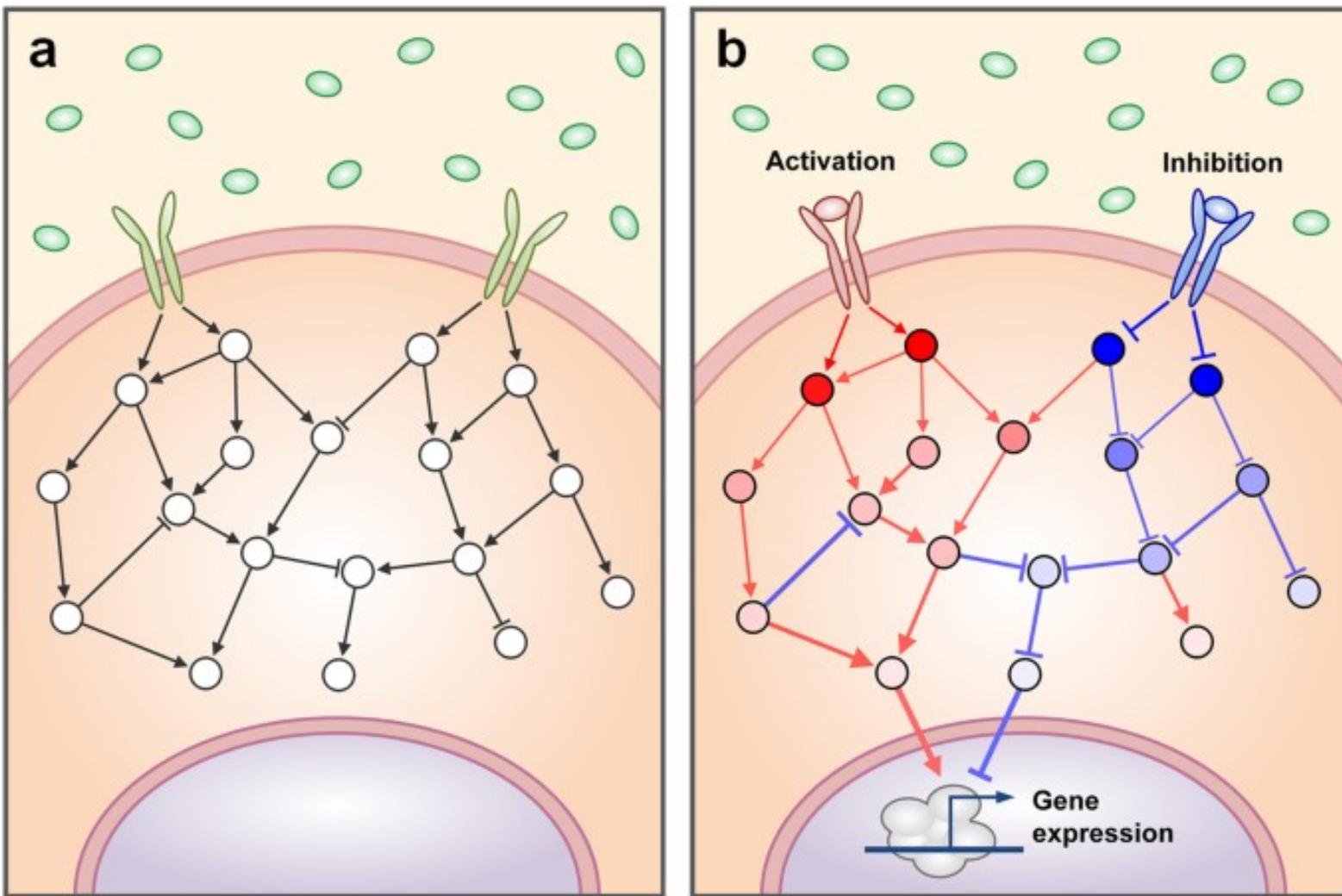


New York subway map

(Obviously!) cells should communicate.  
How do they achieve this?



# Signaling networks



## Find Reactions, Proteins and Pathways

e.g. O95631, NTN1, signaling by EGFR, glucose

**Go!****Pathway Browser**

Visualize and interact with Reactome biological pathways

**Analysis Tools**

Merges pathway identifier mapping, over-representation, and expression analysis

**ReactomeFIViz**

Designed to find pathways and network patterns related to cancer and other types of diseases

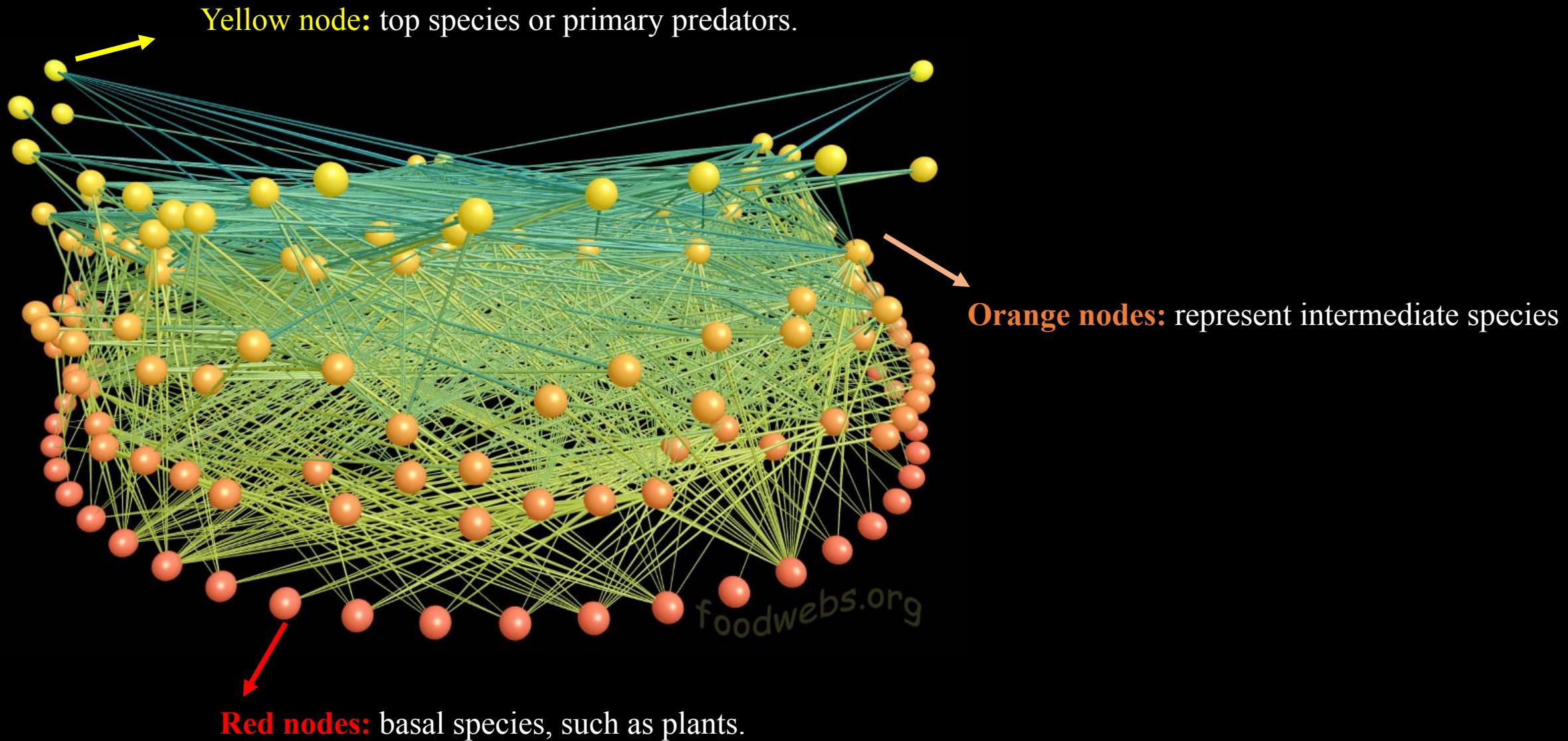
**Documentation**

Information to browse the database and use its principal tools for data analysis

Any higher-level network to talk about?



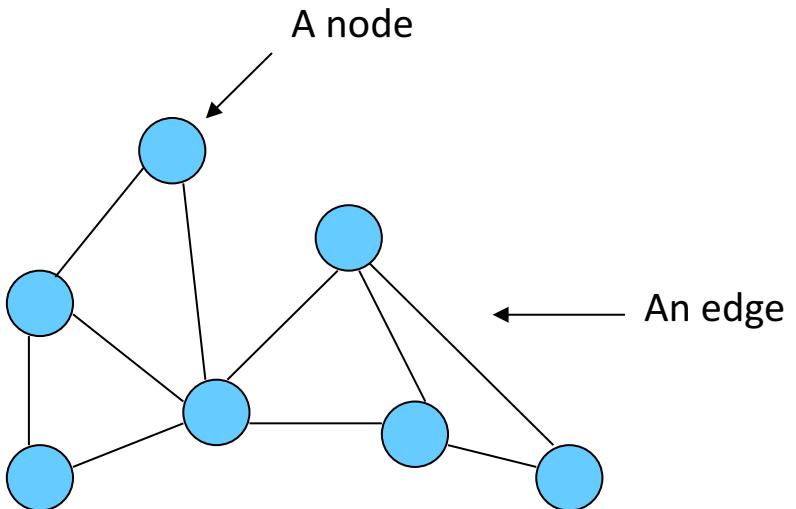
## Food web



*Session 2*

## Properties of networks

# Graphs

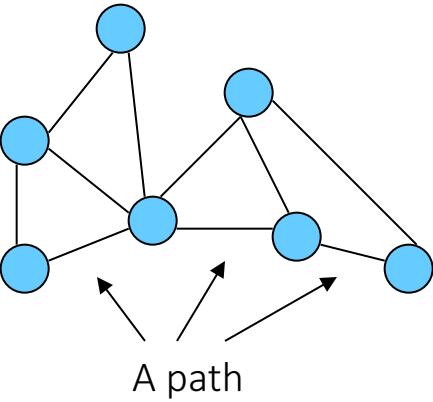


A graph  $G=(V,E)$  comprises  
a set  $V$  of nodes (vertices)  
a set  $E$  of edges

$$V = \{V_1, \dots, V_n\}$$
$$E = \{(V_i, V_j), \dots, (V_k, V_l)\}$$

Protein interaction networks are undirected graphs  
(Individual node pairs in  $E$  are unordered.)

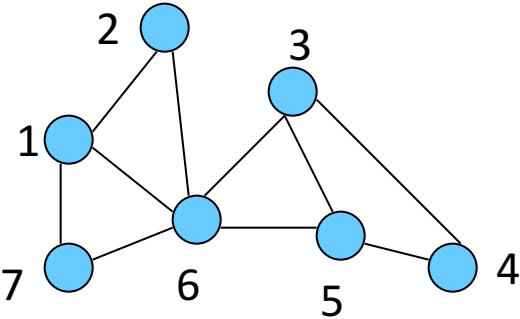
# Graphs



A path is a sequence of alternating nodes and edges in which no node is visited more than once

A geodesic is the shortest path between two nodes.

## Graphs can be represented by matrices

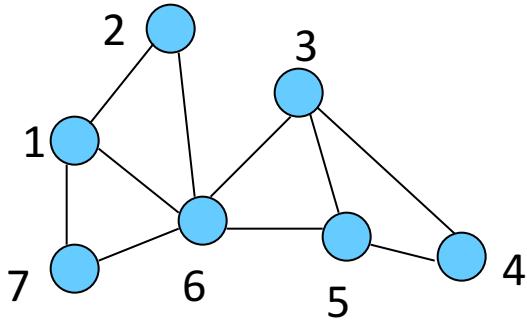


Adjacency matrix  $A=(a_{ij})$

$$\begin{array}{ll} a_{ij}=1 & (V_i, V_j) \in E \\ a_{ij}=0 & \text{otherwise} \end{array}$$

$$A = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

## Degree distribution of graphs

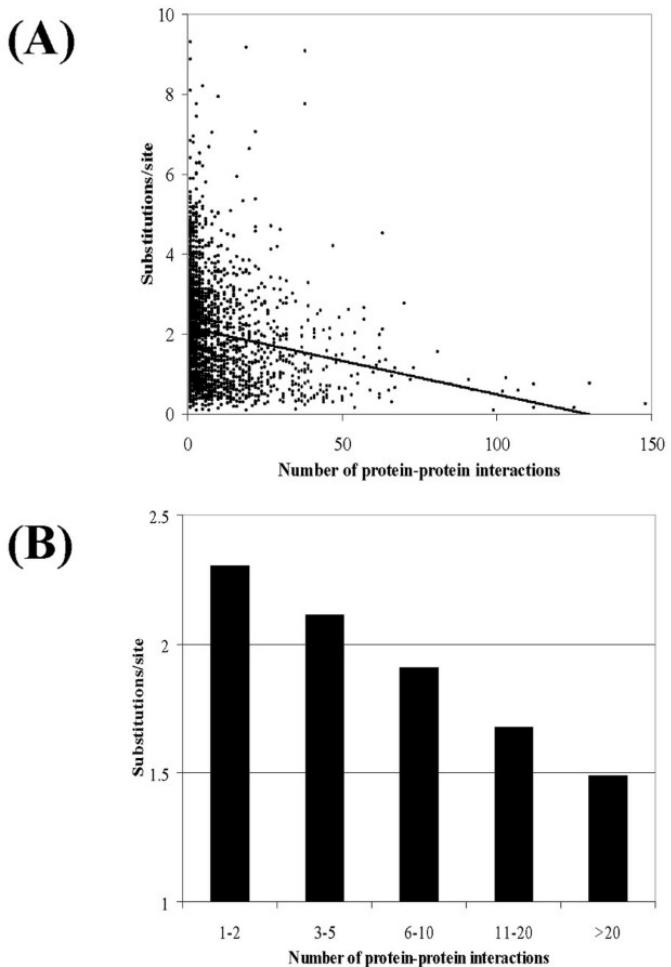


The degree (connectivity)  $k_i$  of a node  $V_i$  is the number of edges incident with the node (e.g.,  $k_1=3$ ,  $k_6=5$ ).

$$k_i = \sum_j a_{ij}$$

Graphs can be characterized according to their degree distribution  $P(k)$ , the fraction of nodes having degree  $k$ .

# Connectivity can affect the evolution of proteins!



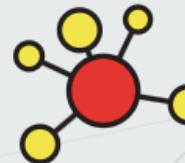
Highly connected proteins tolerate  
fewer amino acid substitutions in their evolution.

BMC evolutionary biology 3.1 (2003): 1-6

Class activity:

Determine the degree distribution of your network of friends ( $n=5$ )





# igraph – The network analysis package

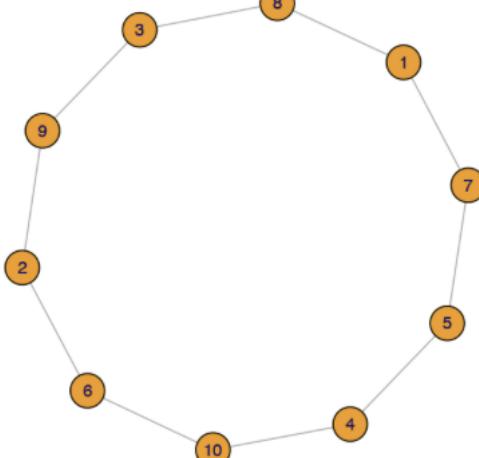
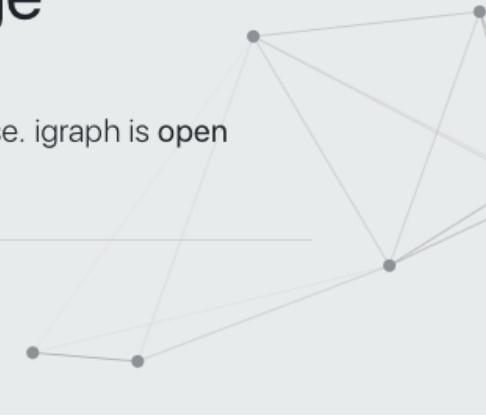
igraph is a collection of network analysis tools with the emphasis on efficiency, portability and ease of use. igraph is open source and free. igraph can be programmed in R, Python, Mathematica and C/C++.

igraph R package

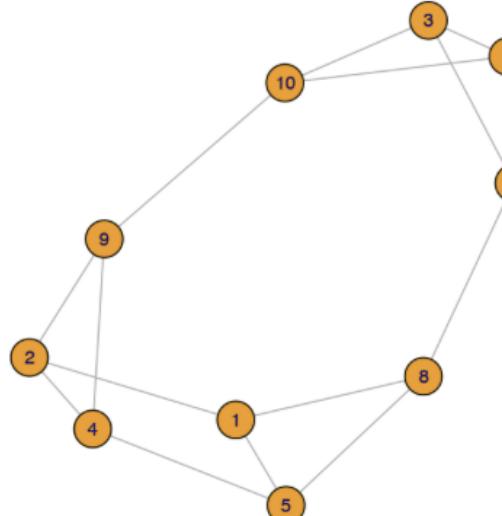
python-igraph

IGraph/M

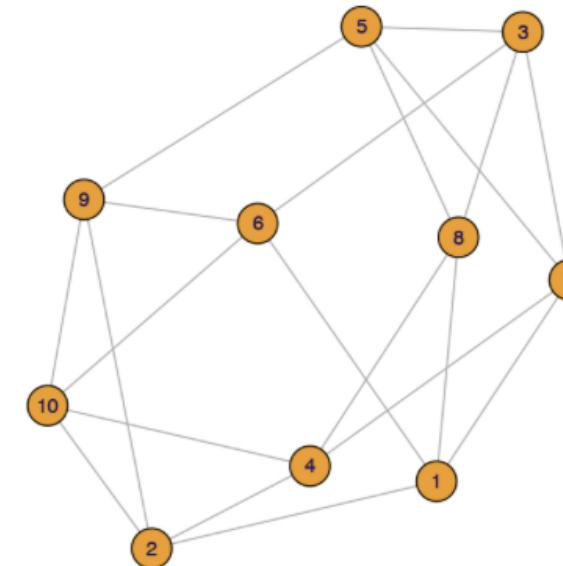
igraph C library



`plot(sample_k_regular(10, 2))`

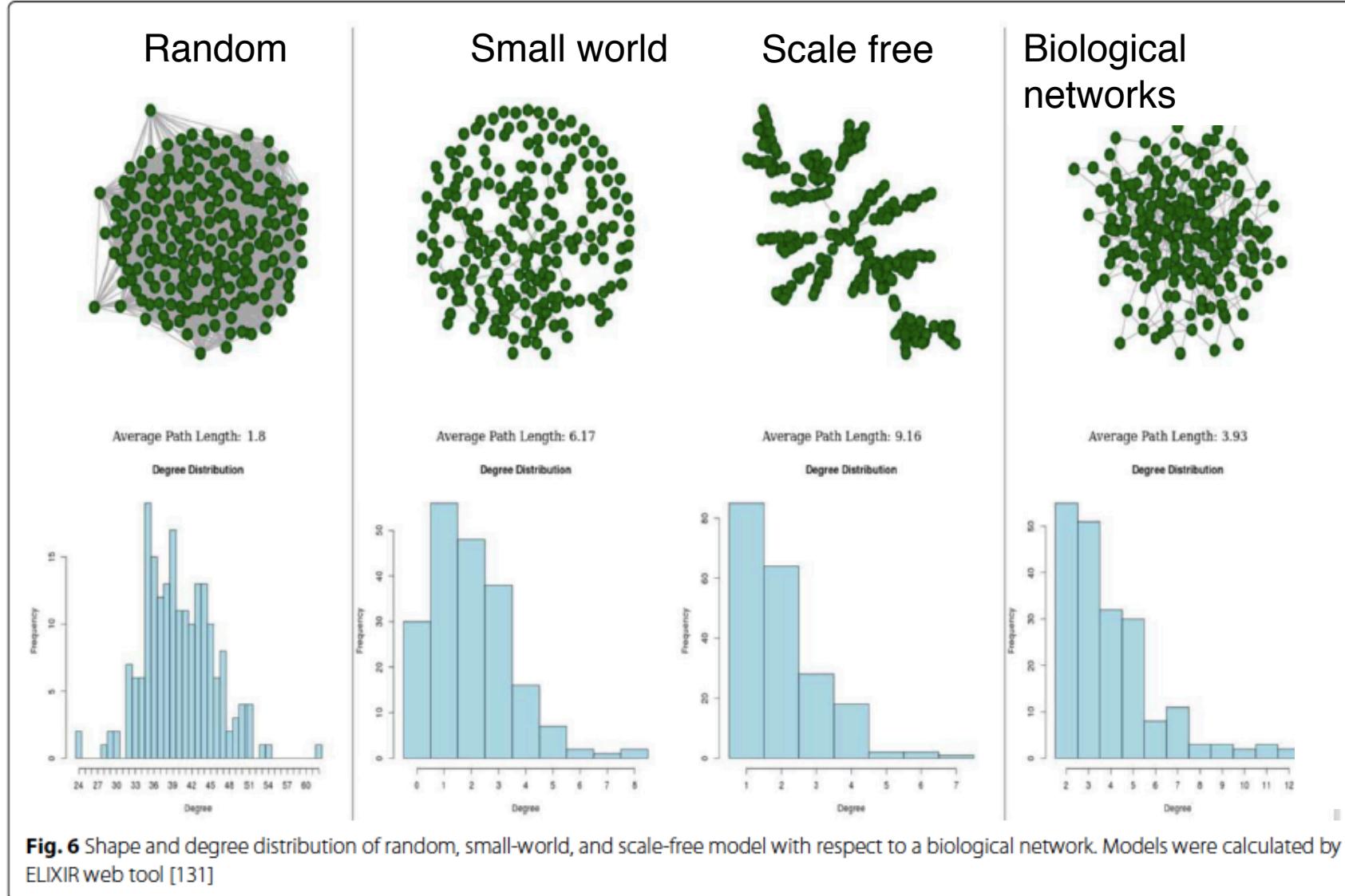


`plot(sample_k_regular(10, 3))`



`plot(sample_k_regular(10, 4))`

# Degree distribution can help us determine the type of networks



EURASIP Journal on  
Bioinformatics and Systems  
Biology 2017.1 (2017): 1-16.

Class activity:

How do you build a random graph?

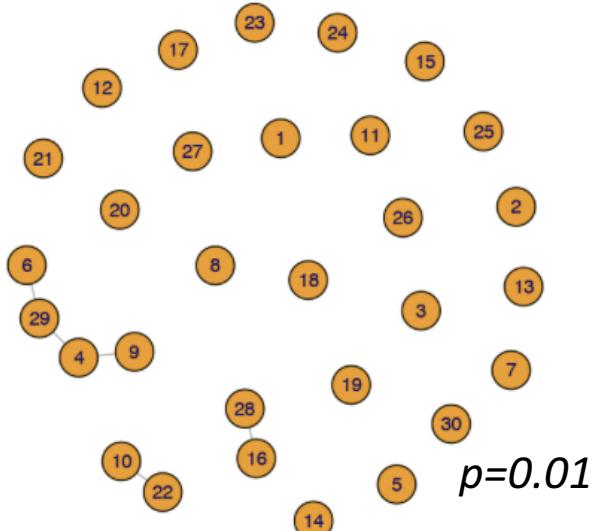


# The random graph

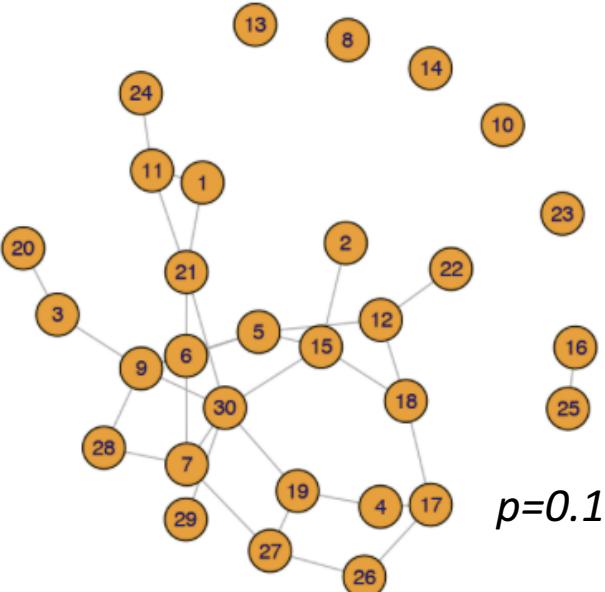
Erdős–Rényi model

$$P(k) = \binom{n-1}{k} p^k (1-p)^{n-1-k}$$

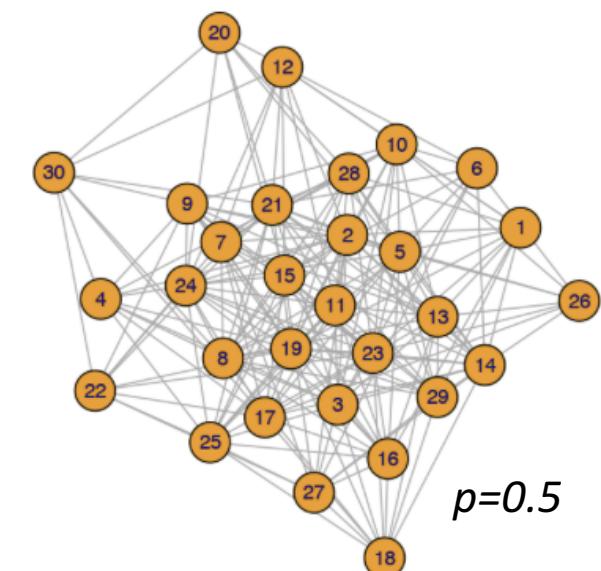
$$P(k) = e^{-(np)} \frac{(np)^k}{k!}$$



```
plot(sample_gnp(30, 0.01))
```

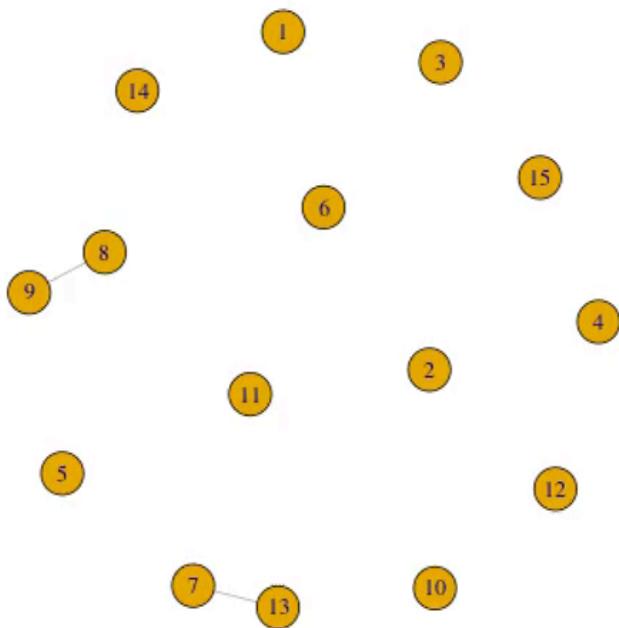


```
plot(sample_gnp(30, 0.1))
```

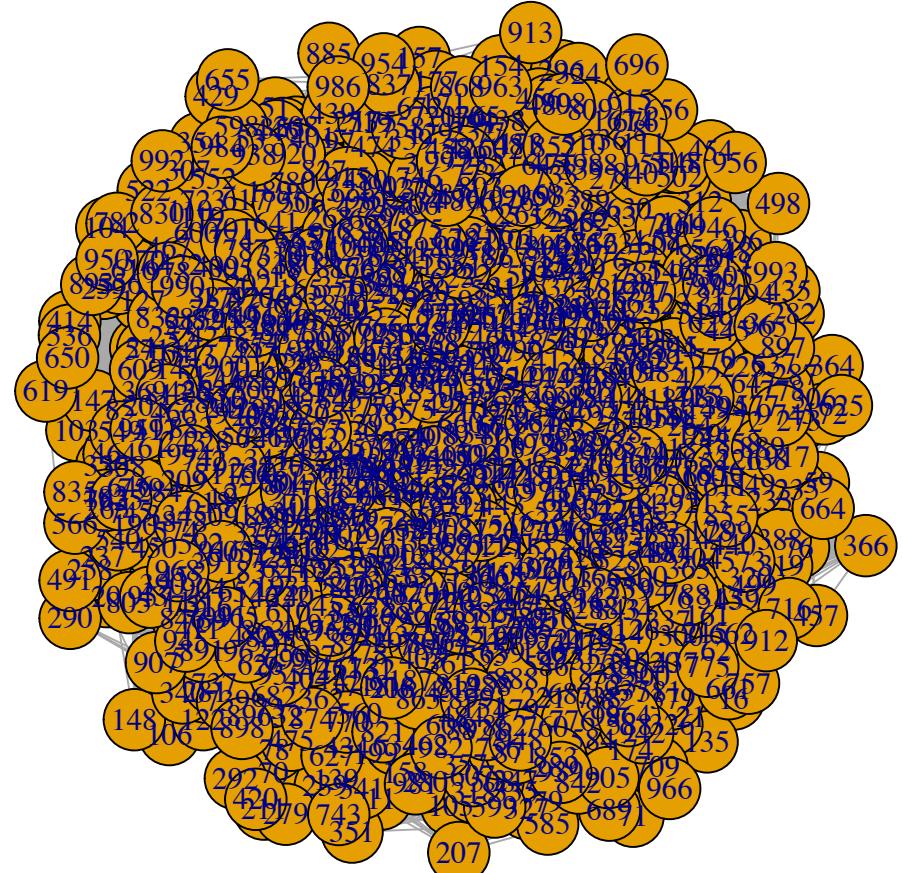
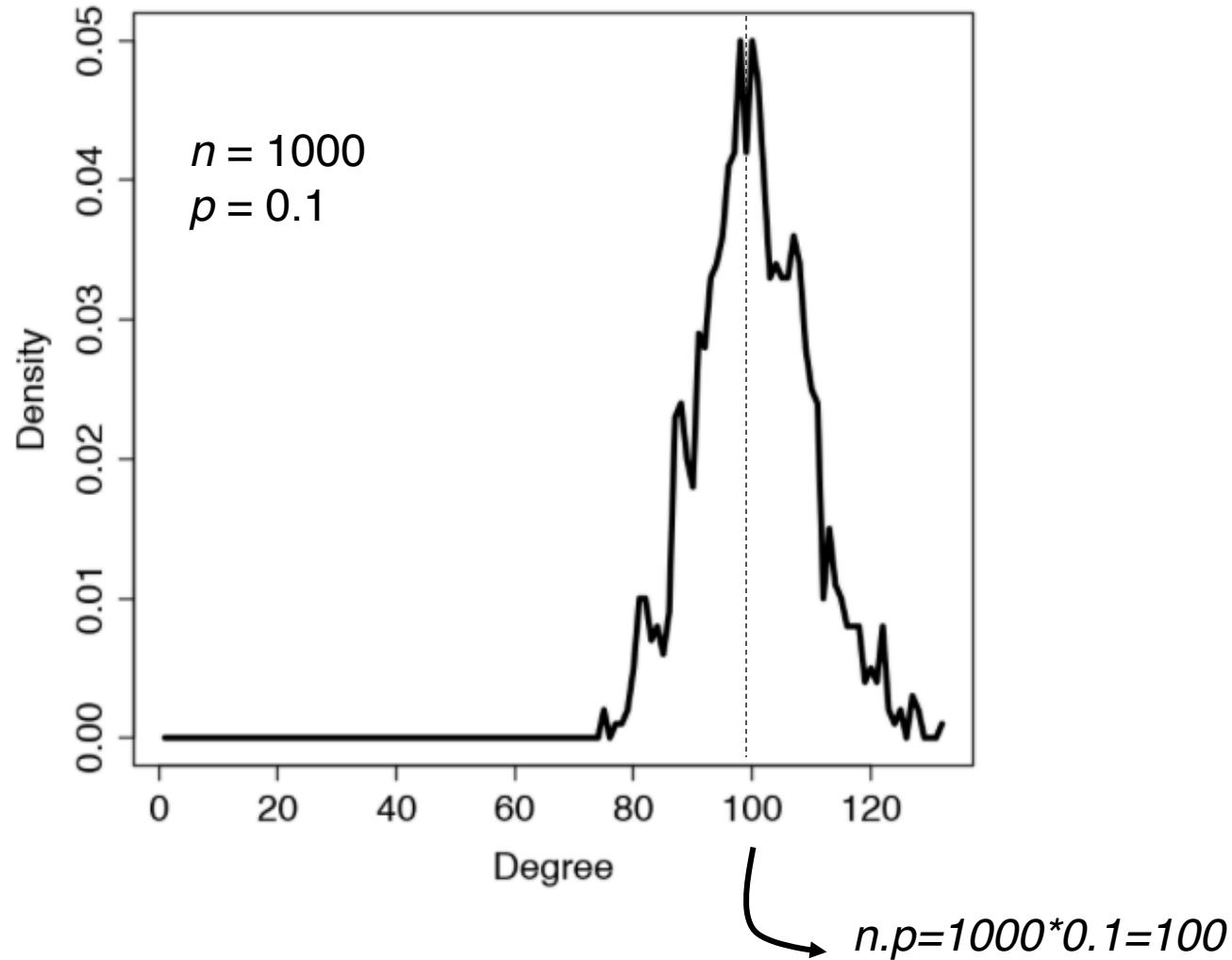


```
plot(sample_gnp(30, 0.5))
```

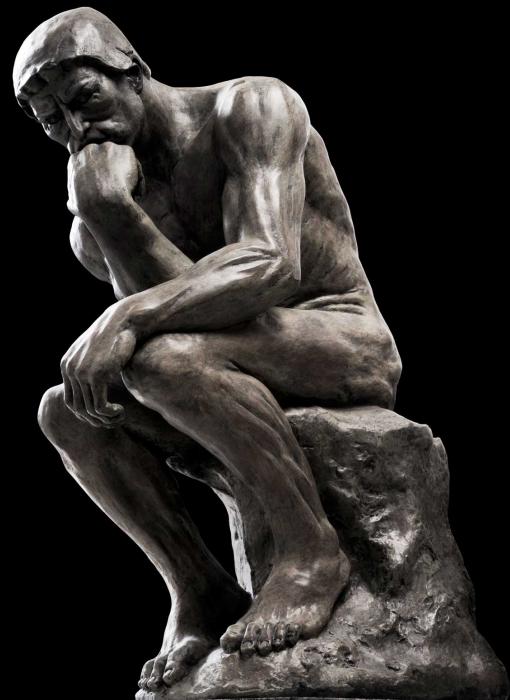
**Probability= 0.03**



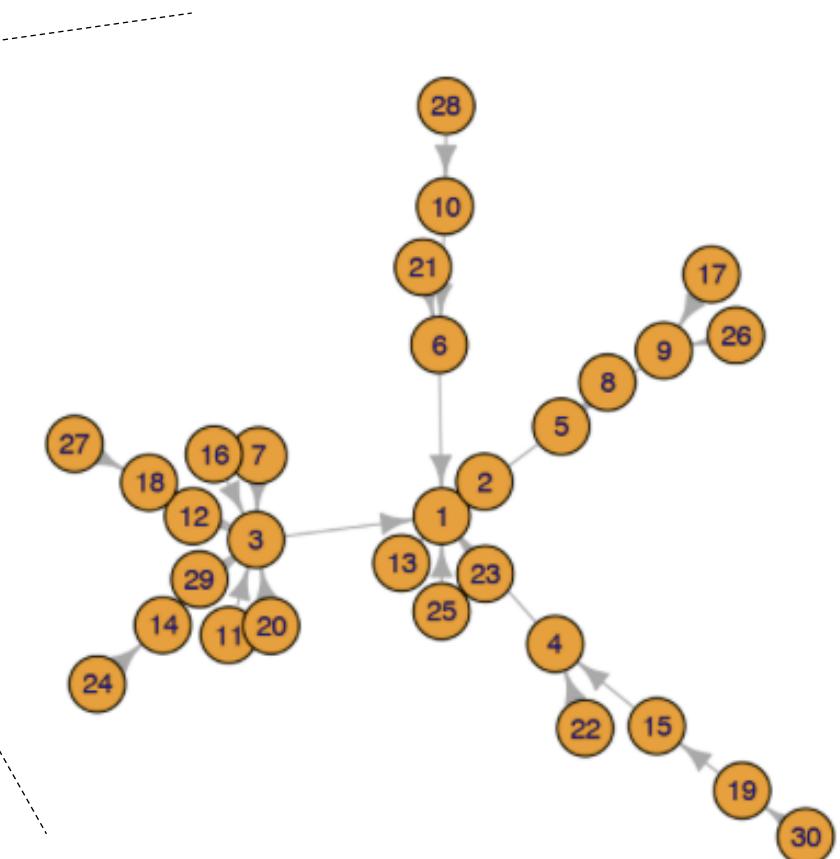
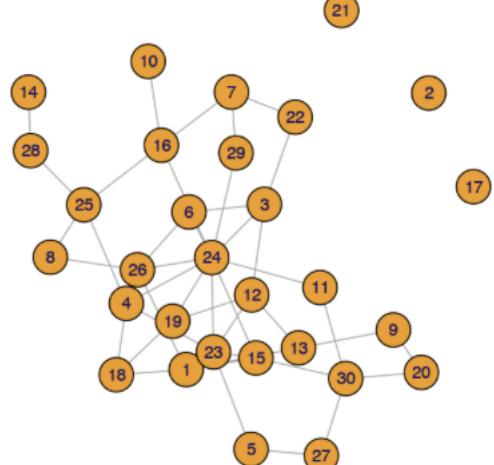
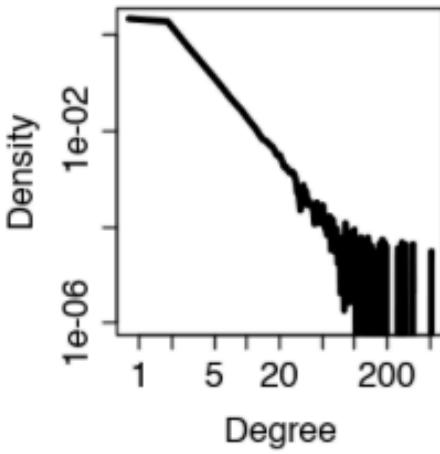
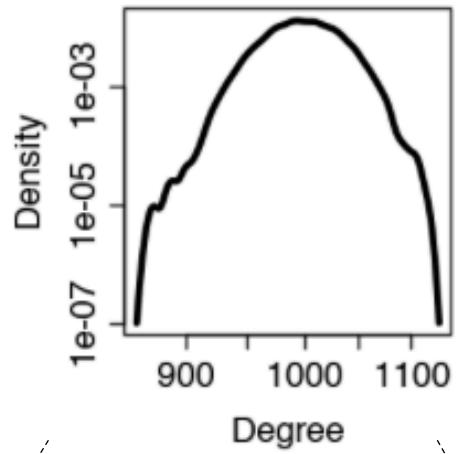
# The random graph



# Are biological networks random?



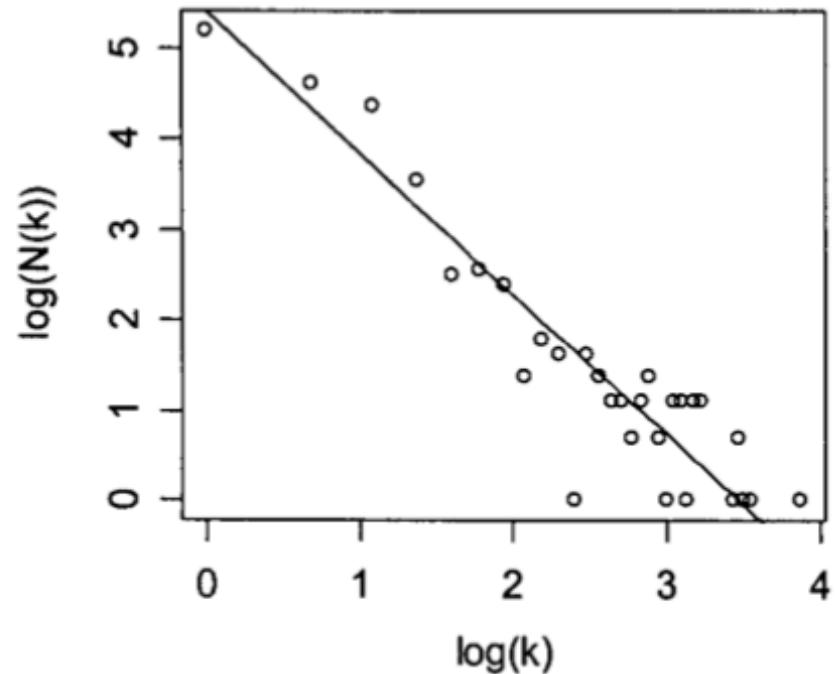
# The random graph vs. scale-free graph



Log-log scale

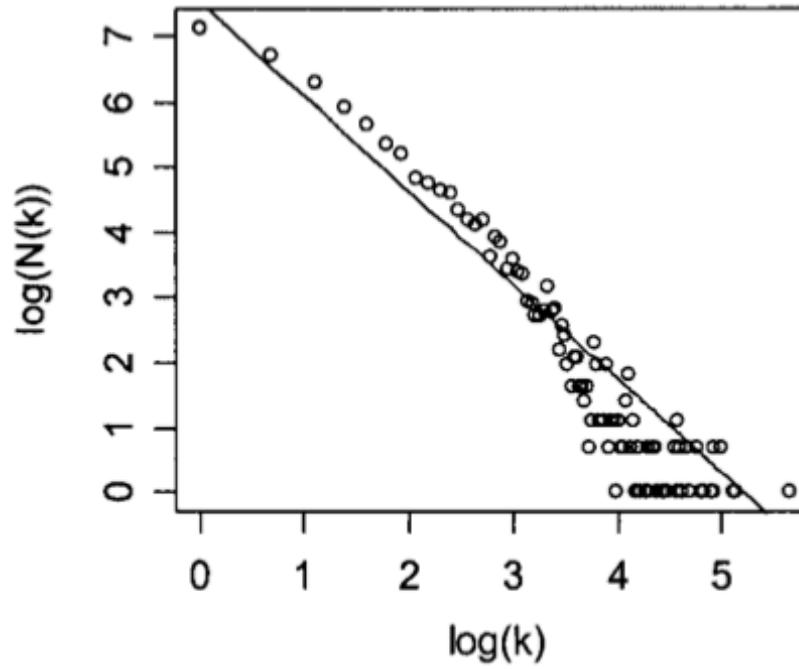
# Biological network are non-random

Guelzim *et al.*, 2002



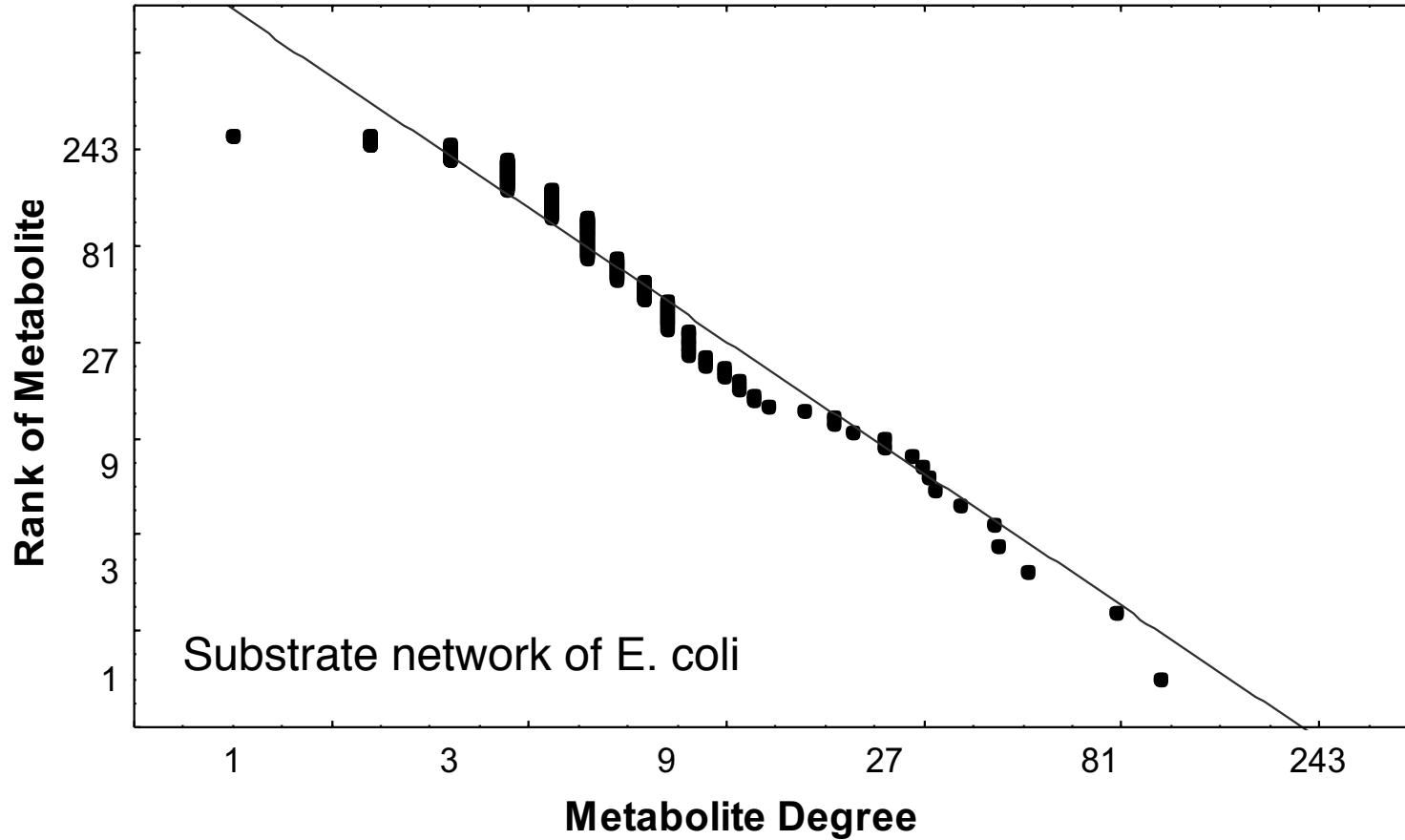
Transcriptional  
regulatory network

Tong *et al.*, 2004

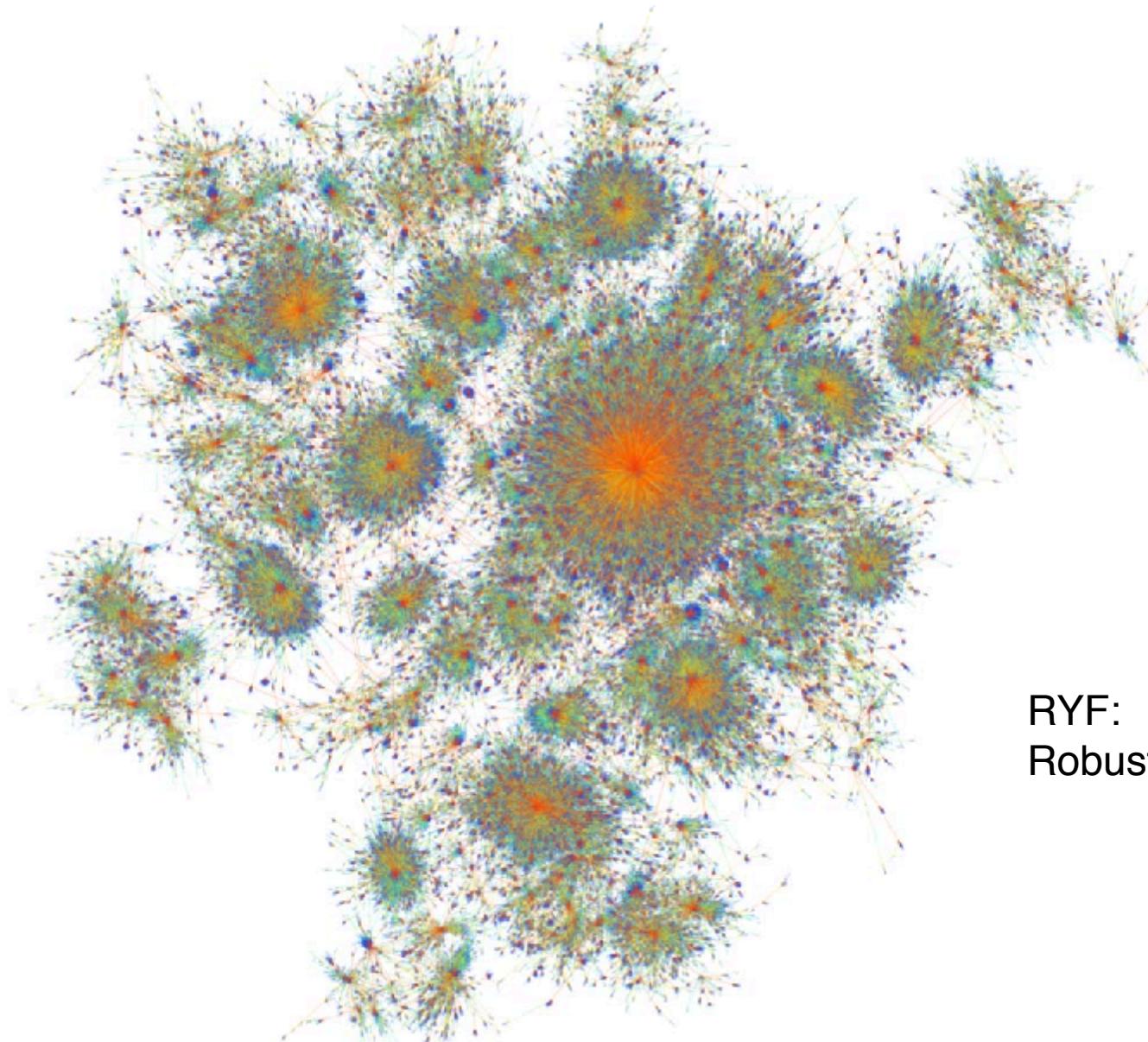


Genetic interaction  
network

# Metabolic networks have a broad-tailed degree distribution



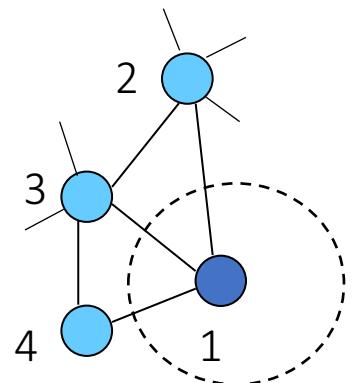
## What is so special about scale-free networks?



**RYF:**  
Robust yet Fragile!

## Nearest-neighbor degree of nodes

Average nearest neighbor degree of a node



$$k_1=3$$

$$k_2=5$$

$$k_3=5$$

$$k_4=2$$

$$k_{nn,1} = (1/3)(5+5+2) = 4$$

$$k_{nn,i} = \frac{1}{k_i} \sum_{j, \text{ nearest neighbors of } i} k_j$$

# Assortative vs. Disassortative networks

Average nearest neighbor degree of all nodes with degree  $k$

$$k_{nn,i} = \frac{1}{k_i} \sum_{j, \text{ nearest neighbors of } i} k_j$$

A graph is assortative if  $k_{nn}(k)$  increases with  $k$   
(You are well-connected; Your neighbors are well-connected)

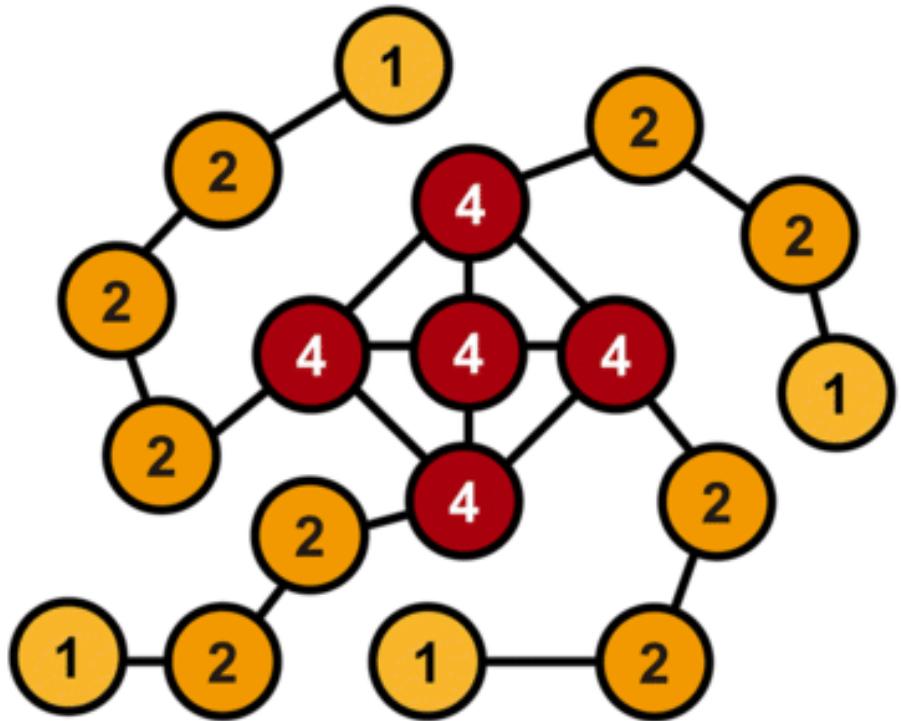
$N_k$ ...number of nodes with degree  $k$

$$k_{nn}(k) = \frac{1}{N_k} \left( \sum_{\text{nodes with degree } k} k_{nn,k} \right)$$

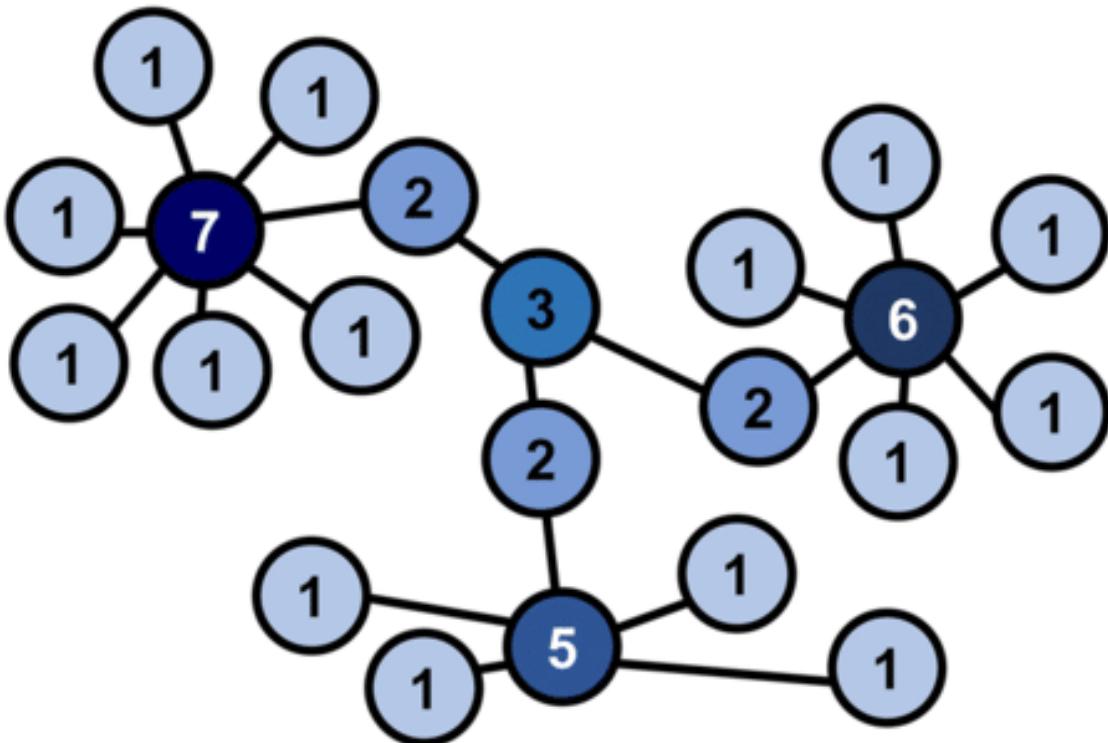
A graph is disassortative if  $k_{nn}(k)$  decreases with  $k$   
(You are well-connected; Your neighbors are not!)

## Assortative vs. Disassortative networks

**A** Assortative network

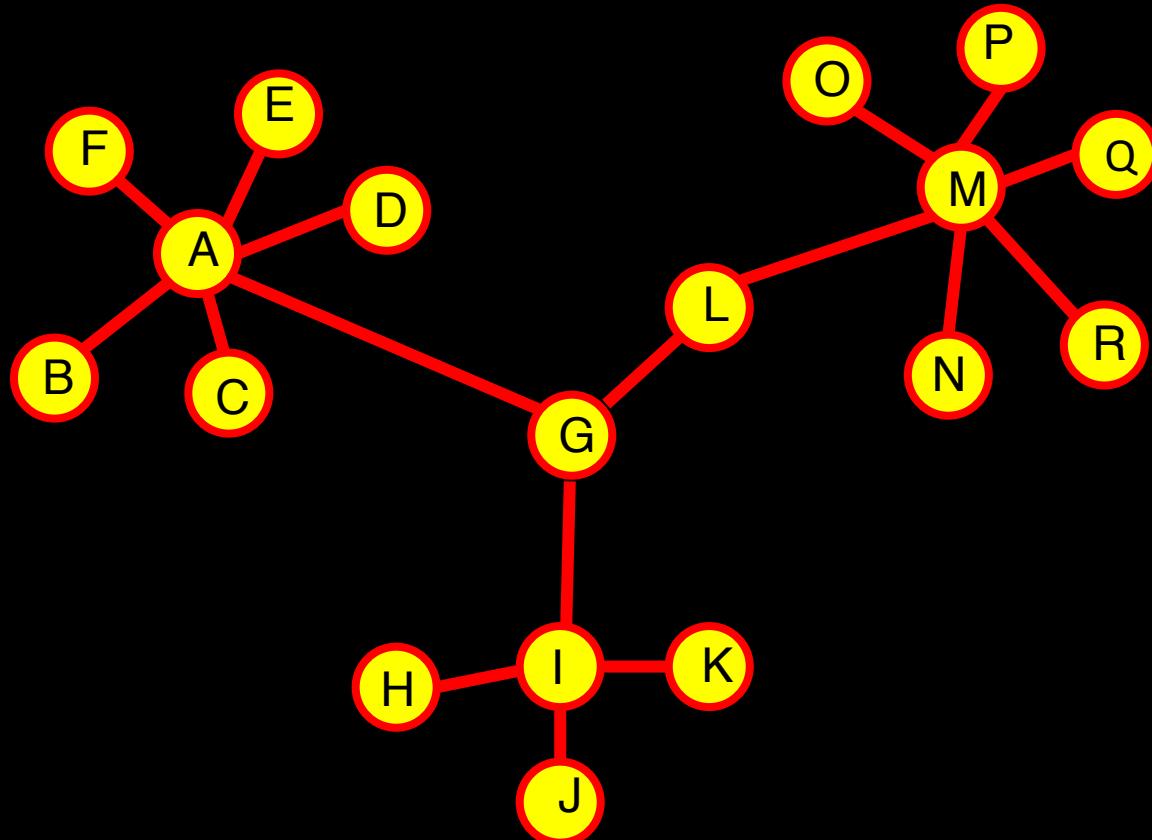


**B** Disassortative network

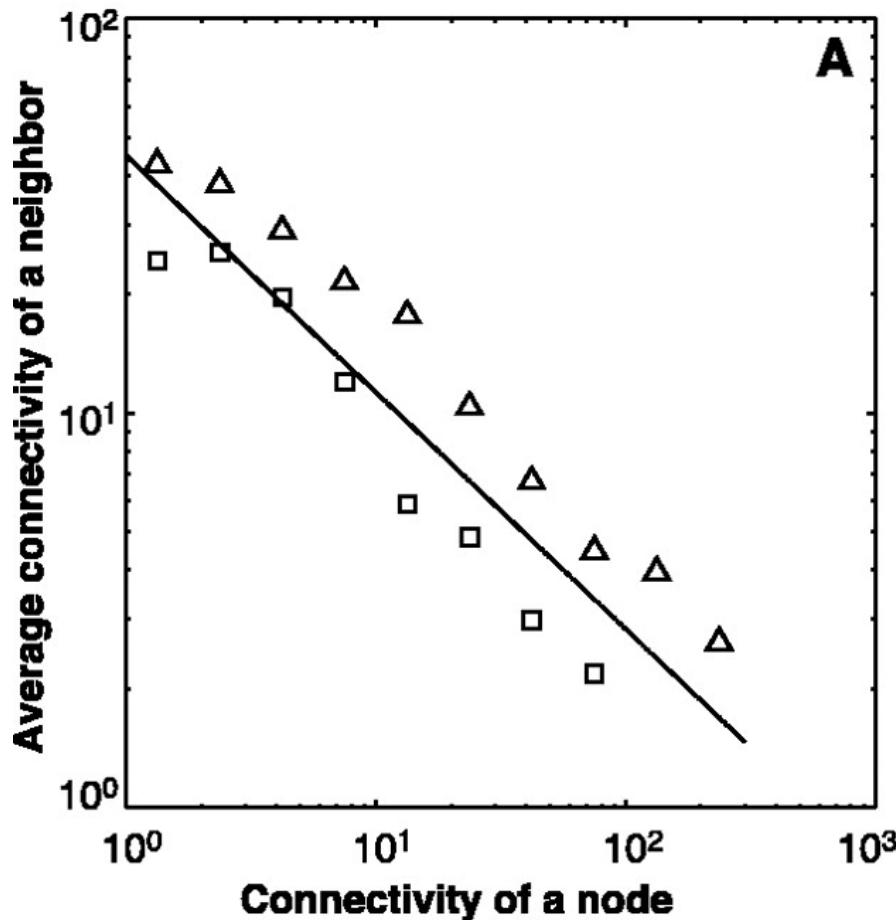


## Class activity:

How would you make the following graph i) more or ii) less disassortative.



# Protein interaction networks are disassortative



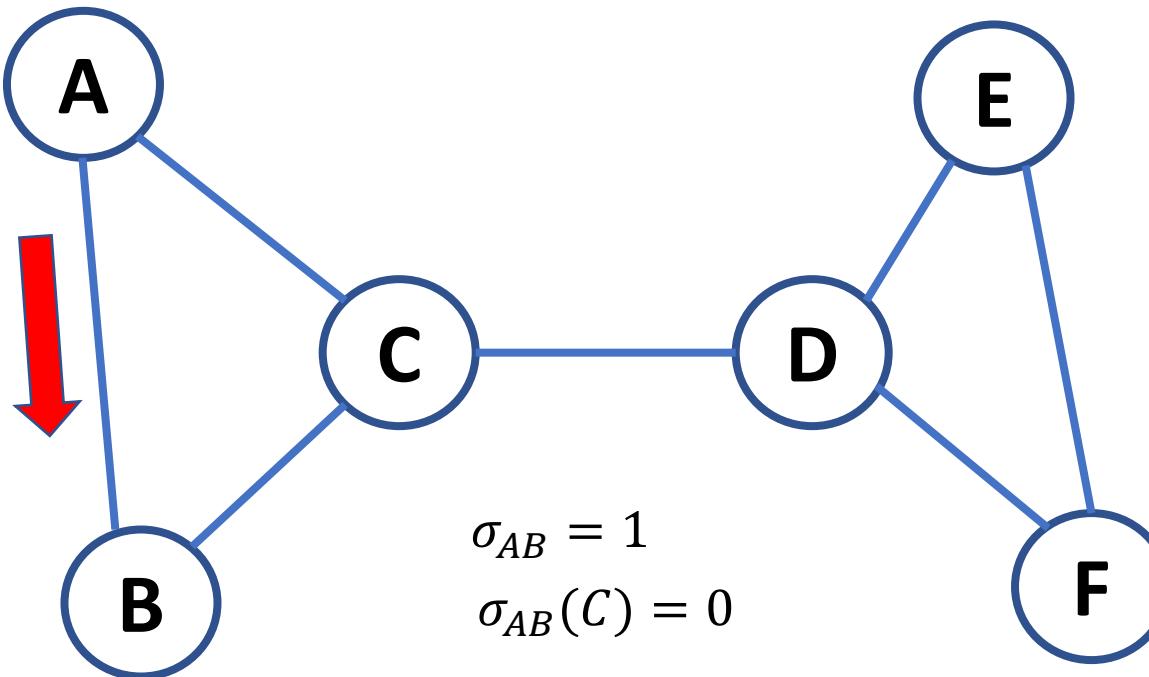
- 1) Few interactions between hubs.
- 2) Many interactions between hubs and neighbors with low degrees.

## Measures of graph compactness: Betweenness centrality

Measure of centrality of node  $v$  in a graph

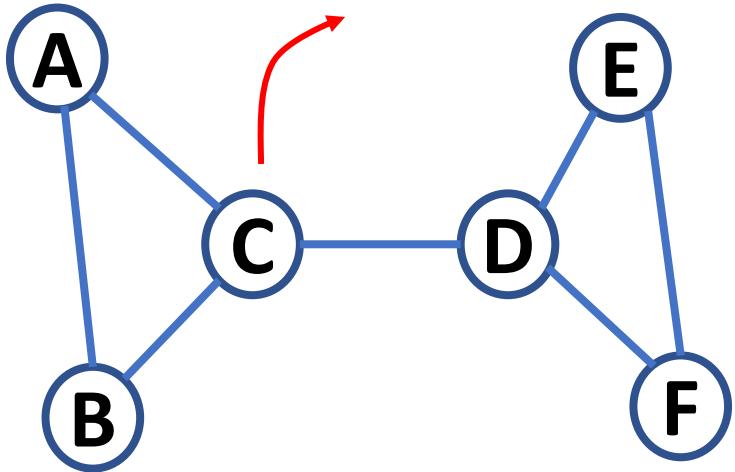
$\sigma_{uw}$       # of shortest paths between nodes  $u$  and  $w$

$\sigma_{uw}(v)$     # of shortest path between nodes  $u$  and  $w$  that involve  $v$



## Betweenness centrality

Betweenness centrality  
for node C

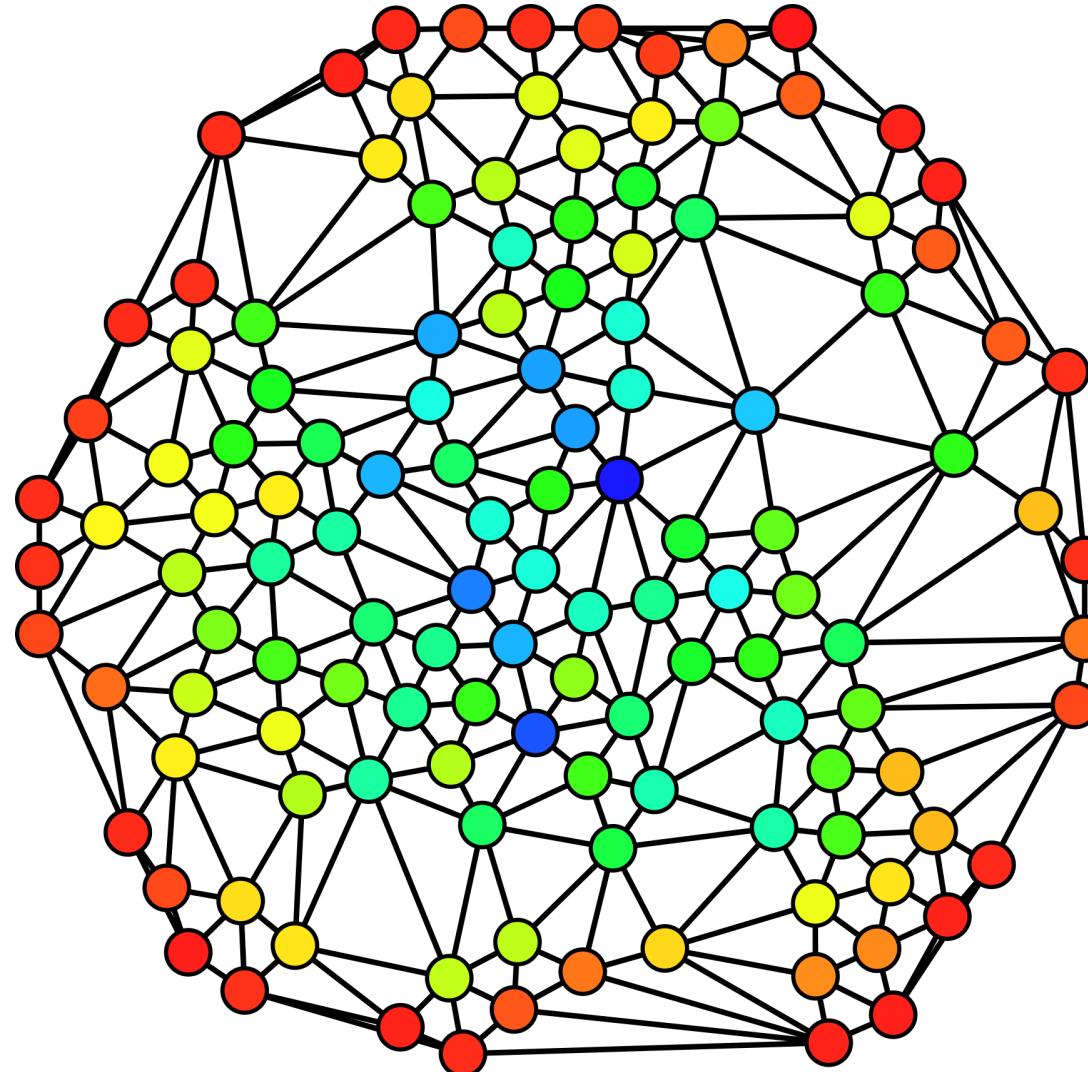


	$\sigma_{uw}$	$\sigma_{uw}(v)$	$\sigma_{uw}(v)/\sigma_{uw}$
(A,B)	1	0	0
(A,D)	1	1	1
(A,E)	1	1	1
(A,F)	1	1	1
(B,D)	1	1	1
(B,E)	1	1	1
(B,F)	1	1	1
(D,E)	1	0	0
(D,F)	1	0	0
(E,F)	1	0	0

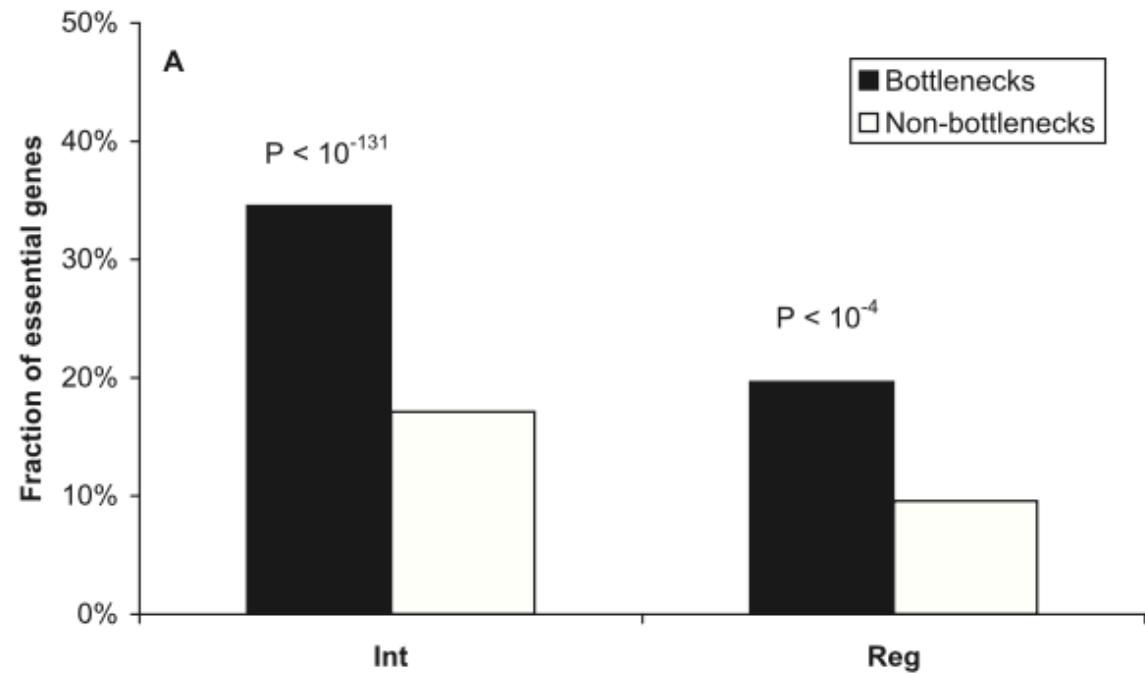
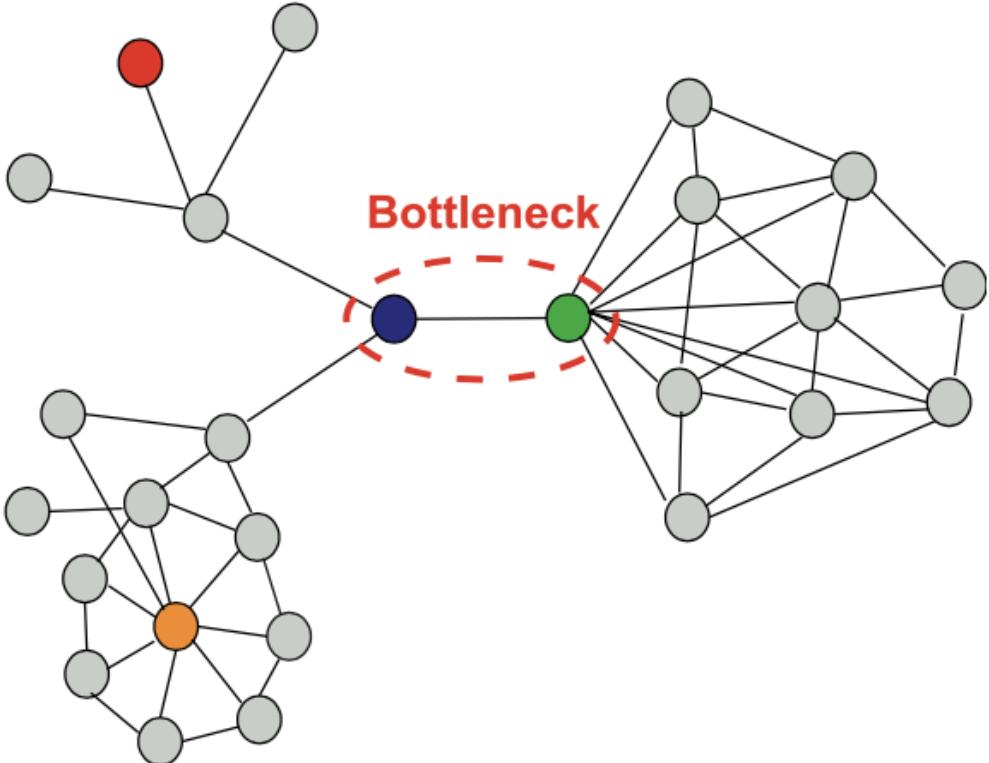
Sum=6

Class activity:

Which nodes have the highest betweenness centrality?

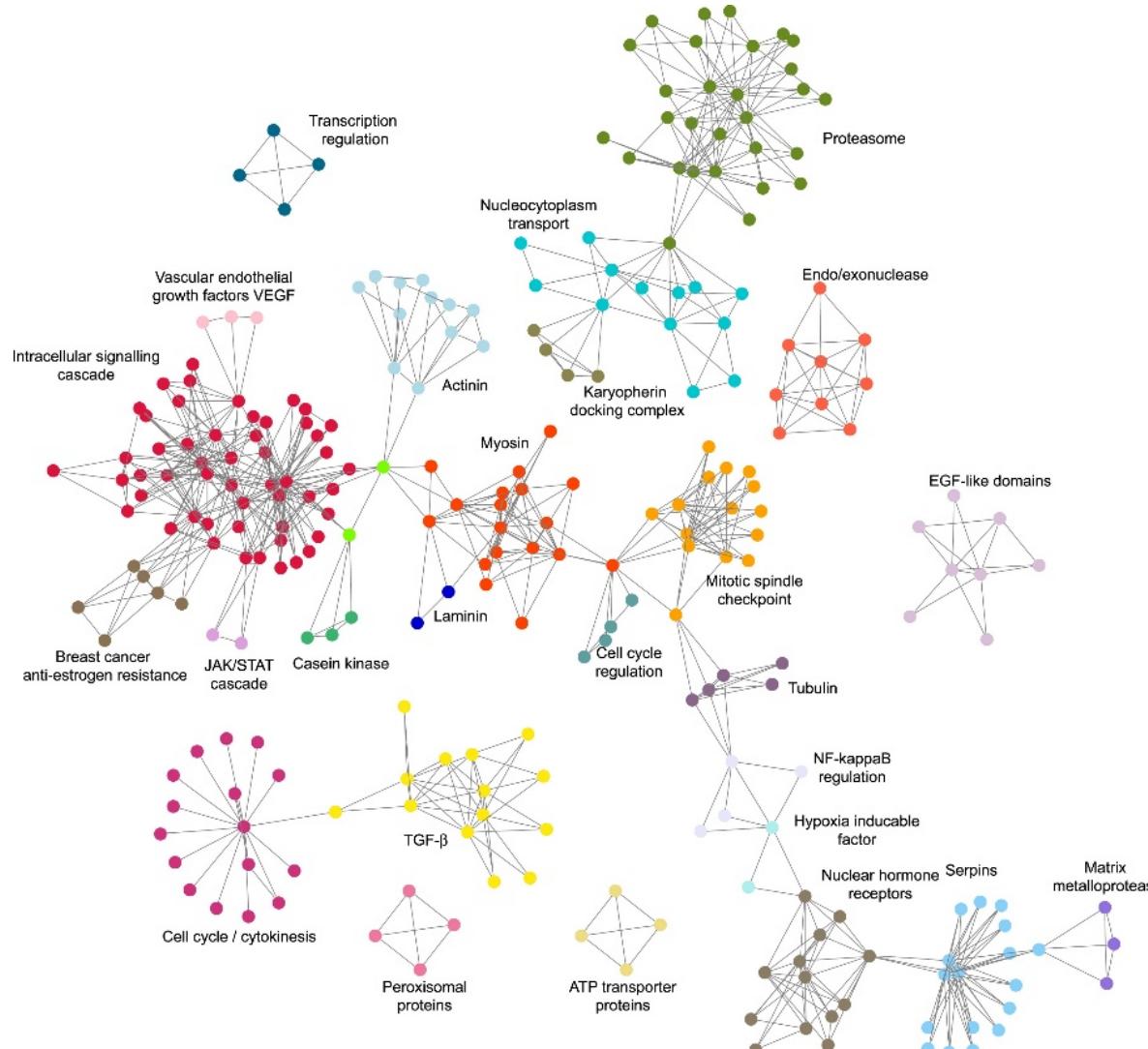


# Bottlenecks have high betweenness centrality scores



Proteins with high betweenness scores  
tend to be more essential

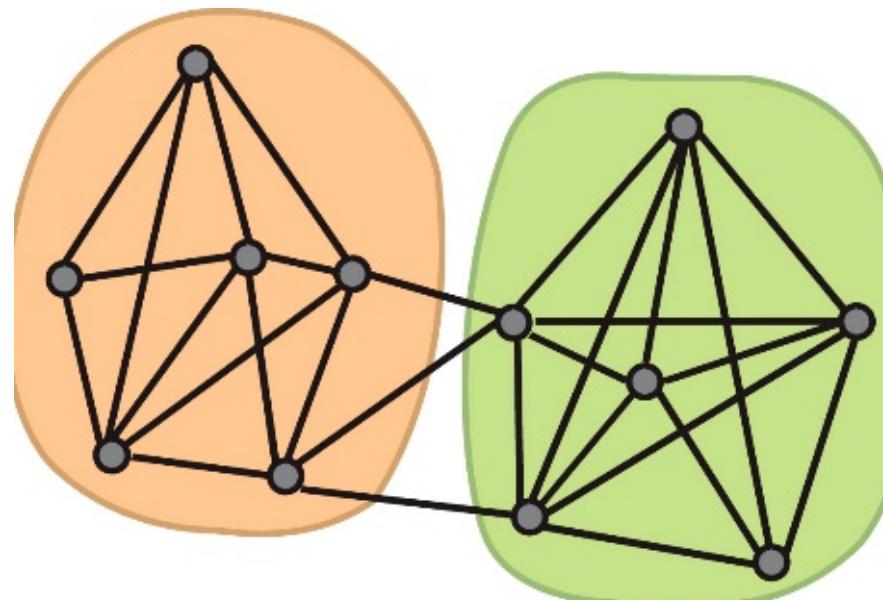
# Advanced topics: Graphs can be subdivided into “communities”



Community structure of a protein interaction network

## *Advanced topics:* Graphs can be subdivided into “communities”

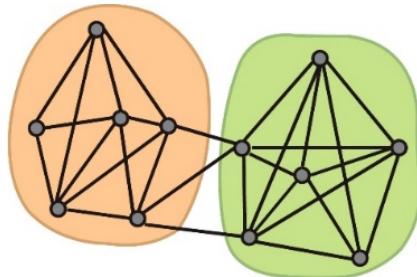
In a graph that can be subdivided into communities (clusters, modules) nodes fall into groups that share more edges with each other than with nodes outside the community



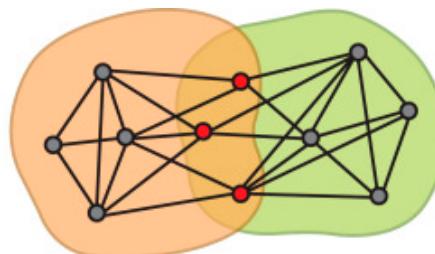
## *Advanced topics:* Graphs can be subdivided into “communities”

Some require information about the total number of communities (easier), others don't (more difficult).

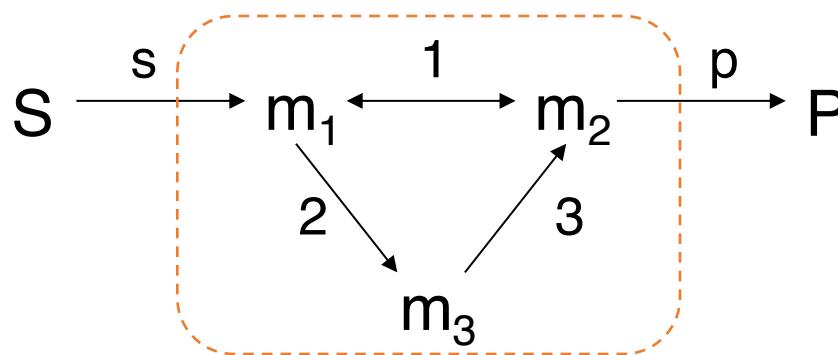
**Hard-clustering** methods generate non-overlapping communities (easier)



**Soft-clustering** methods allow overlapping communities (more difficult)



## Advanced topics: We can do calculations with graphs



A chemical reaction network

Metabolite concentrations  $m_i$  change according to the equations

$$\frac{dm_1}{dt} = v_s - v_1 - v_2$$

$$\frac{dm_2}{dt} = v_1 + v_3 - v_p$$

$$\frac{dm_3}{dt} = v_2 - v_3$$

$$\frac{d\vec{m}}{dt} = \mathbf{S}\vec{v}$$

$$\mathbf{S} = \begin{pmatrix} 1 & -1 & -1 & 0 & 0 \\ 0 & 1 & 0 & 1 & -1 \\ 0 & 0 & 1 & -1 & 0 \end{pmatrix}$$

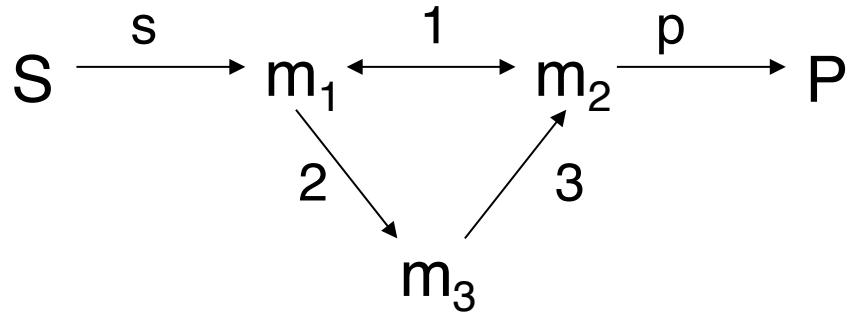
Stoichiometry matrix

$v_i$  metabolic flux through reaction i

$$\vec{v} = (v_s, v_1, v_2, v_3, v_p)^\top$$

Rows: metabolites  
Columns: reactions

## *Advanced topics:* We can do calculations with graphs



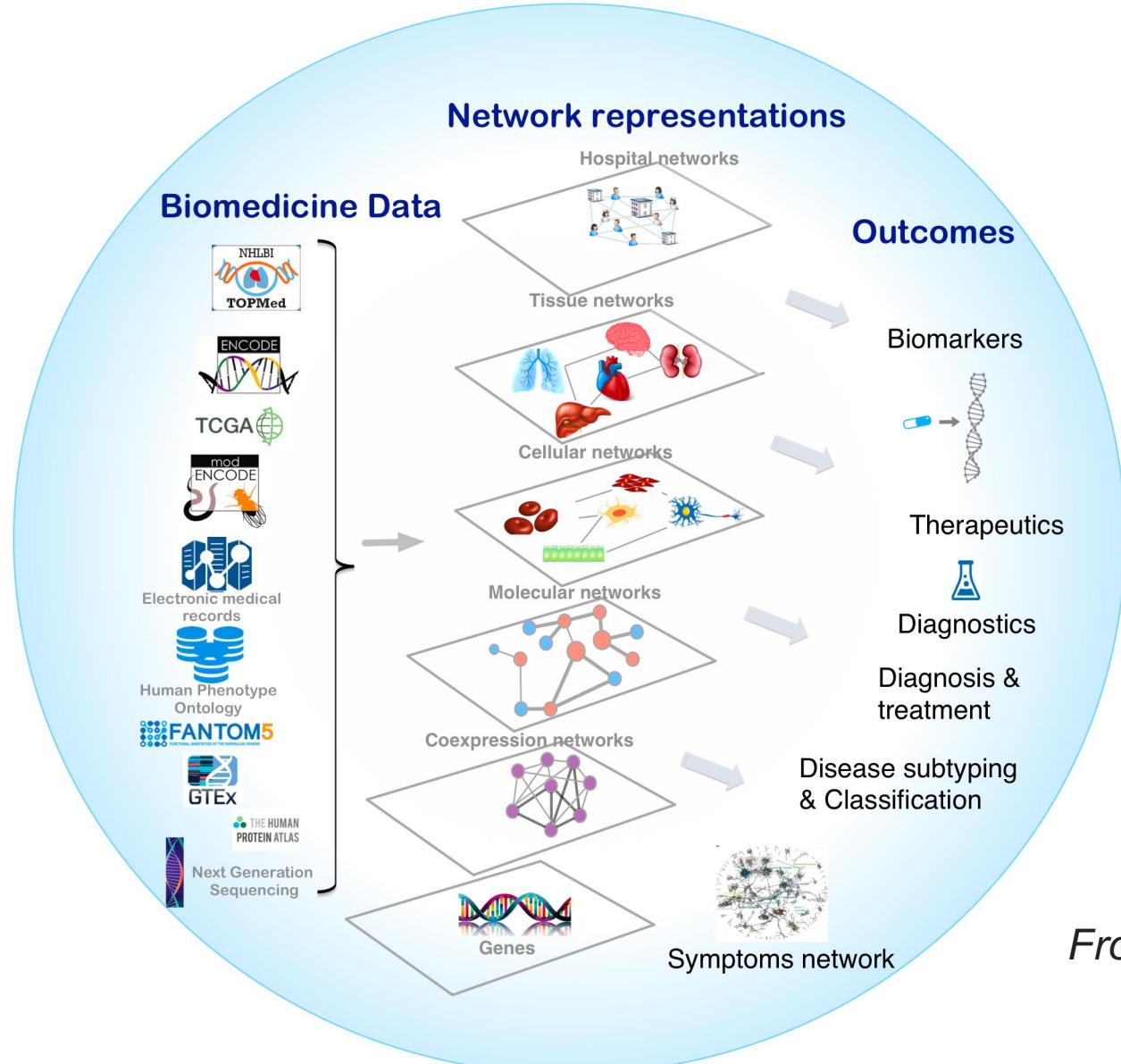
FBA assumes that metabolism is in a steady state where the concentrations of metabolites no longer change

$$\frac{d\vec{m}}{dt} = 0$$

$$\mathbf{S}\vec{v} = 0$$

The solutions of these equations are the allowable metabolic fluxes. They form the so-called null space of S

# Advanced topics: Network medicine



*Frontiers in Genetics* 10 (2019): 294.

## Concluding remarks

- Networks are everywhere! Biological systems also behave as interconnected networks.
- Several mathematical approaches and disciplines such as graph theory help us model networks.
- We don't just visualize networks. We also use networks for:
  - Quantitative analyses
  - Predictions (Bayesian graph networks)
- Don't overlook the experimental approaches to build networks!

# Further reading

## Complex networks in general

Newman, MEJ. The structure and function of complex networks. *SIAM Review* **45**, 167-256, 2003.

Fortunato, S., Hric, D. Community detection in networks: A user guide. *Physics Reports* **659**, 1-44. 2016.

## Protein interaction networks

Xia et al. Analyzing cellular biochemistry in terms of molecular networks. *Annu. Rev. Biochem.* **73**:1051–87, 2004

Rajagopala et a., The binary protein-protein interaction landscape of *Escherichia coli*. *Nature Biotechnology* **32**, 285-290, 2014

## Metabolic networks

Price et al. *Nature Reviews Microbiology* **2**, 886-897, 2004

## Exam questions

