



# Modular structures of the Protein-protein interaction networks

---

Ka-Lok Ng (吳家樂)

Department of Biotechnology and Bioinformatics (生物科技與生物資訊學系)

Asia University (亞洲大學)

# Content

---

1. From genomes to biological networks
2. Topological features of protein-protein interaction networks
  - Graph theory, topological parameters (node degrees, average path length, clustering coefficient, node degree correlation ..etc.)
  - Random graph, Scale-free network, Hierarchical network
  - Stability of the network
3. Results
4. Summary

# From Genomes to Biological Networks

Over 1000 on-going genomes projects, and more than 200 microbes have been completely sequenced. By now, scientists have a lot genomic and protein data.

*NCBI Genome Database*

<http://www.ncbi.nlm.nih.gov/Genomes/index.html>

*Genomes OnLine Database*

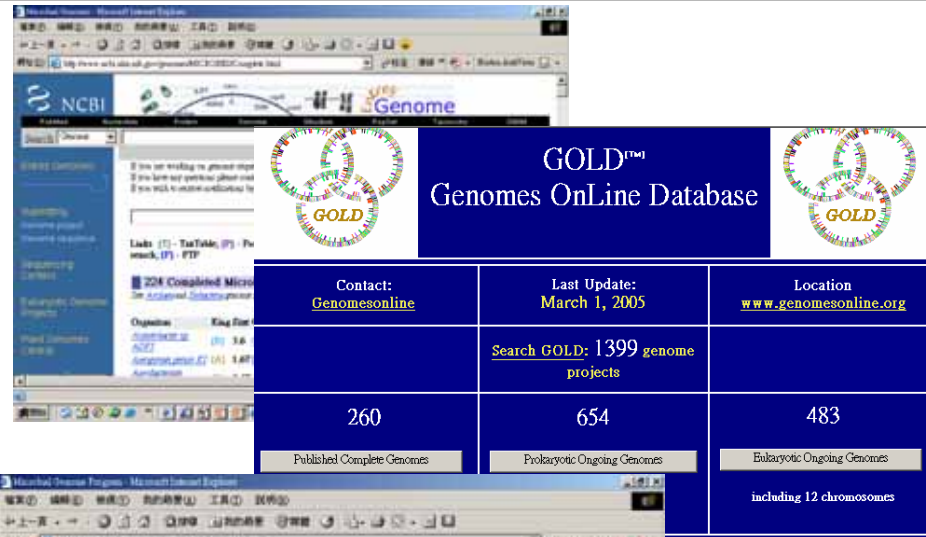
<http://www.genomesonline.org/>

*Microbial Genome Program*

<http://www.er.doe.gov/production/ober/microbial.html>

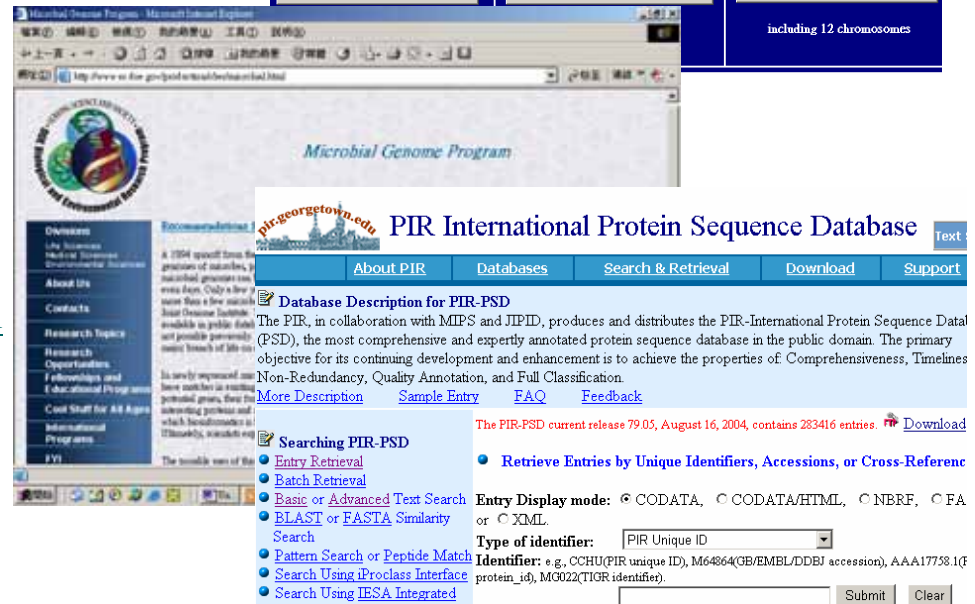
*PIR*

<http://www-nbrf.georgetown.edu/>



The screenshot shows the NCBI Genome Database interface. At the top, there's a navigation bar with 'Genome' and 'GOLD™ Genomes OnLine Database'. Below this, there's a table with the following data:

Contact:	Last Update:	Location
<a href="http://www.genomesonline.org">Genomesonline</a>	March 1, 2005	<a href="http://www.genomesonline.org">www.genomesonline.org</a>
	Search GOLD: 1399 genome projects	
260	654	483
Published Complete Genomes	Prokaryotic Ongoing Genomes	Eukaryotic Ongoing Genomes
		including 12 chromosomes



The screenshot shows the PIR International Protein Sequence Database website. At the top, there's a navigation bar with 'About PIR', 'Databases', 'Search & Retrieval', 'Download', and 'Support'. Below this, there's a section titled 'Database Description for PIR-PSD' with the following text:

The PIR, in collaboration with MIPS and JIPID, produces and distributes the PIR-International Protein Sequence Database (PSD), the most comprehensive and expertly annotated protein sequence database in the public domain. The primary objective for its continuing development and enhancement is to achieve the properties of: Comprehensiveness, Timeliness, Non-Redundancy, Quality Annotation, and Full Classification.

More Description    Sample Entry    FAQ    Feedback

The PIR-PSD current release 79.05, August 16, 2004, contains 283416 entries. Download

Searching PIR-PSD

- Entry Retrieval
- Batch Retrieval
- Basic or Advanced Text Search
- BLAST or FASTA Similarity Search
- Pattern Search or Peptide Match
- Search Using iProClass Interface
- Search Using IESA Integrated

Retrieve Entries by Unique Identifiers, Accessions, or Cross-Reference

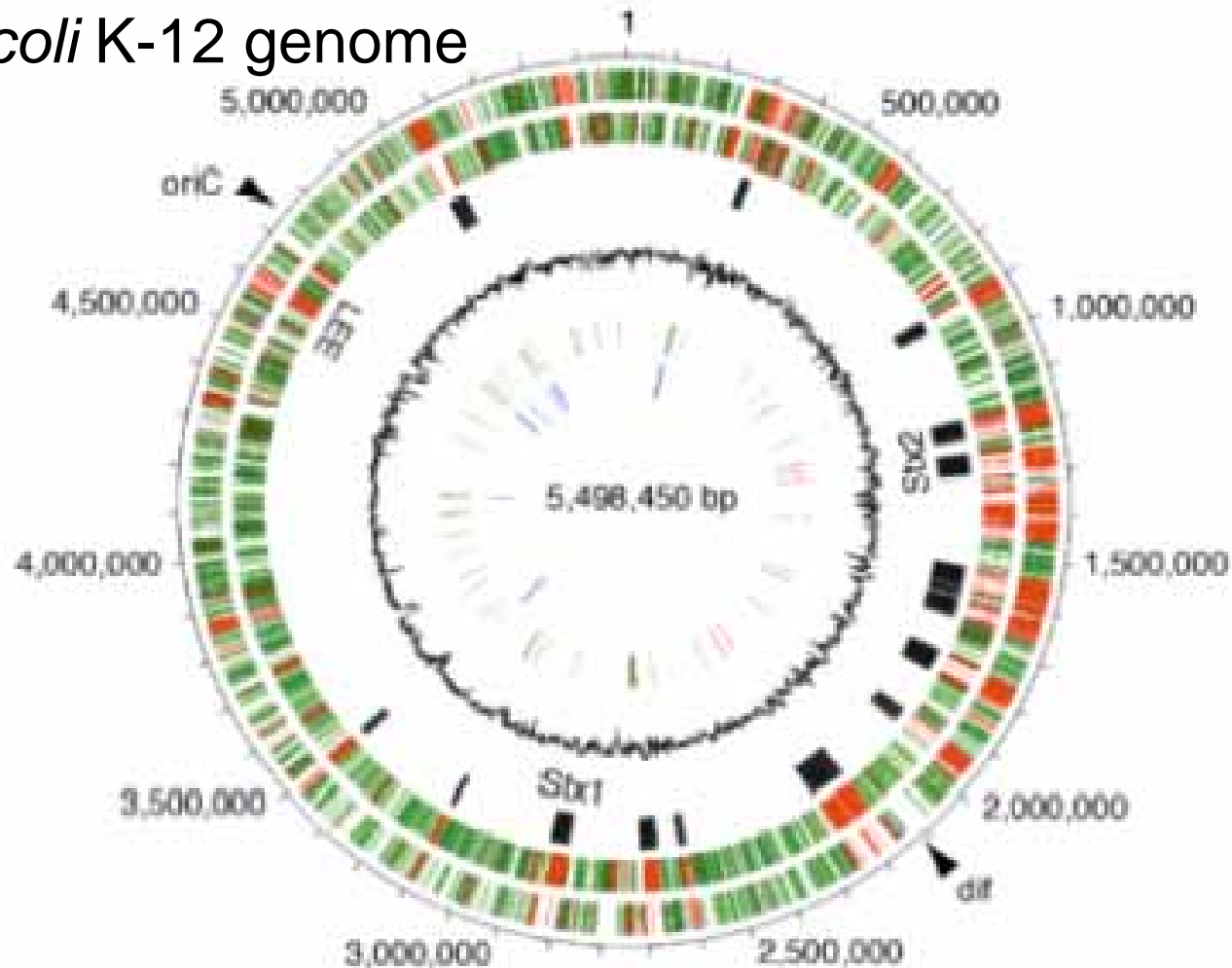
Entry Display mode:  CODATA,  CODATA/HTML,  NBRF,  FA or  XML

Type of identifier:

Identifier: e.g., CCHU(PIR unique ID), M64864(GB/EMBL/DDBJ accession), AAA17758.1( protein\_id), MG022(TIGR identifier).

# From Genomes to Biological Networks

## *E. coli* K-12 genome



# From Genomes to Biological Networks

---

- **Knowing the functionality and interactions is far more important** than just having the genomic or protein sequencing information.
- **Networks of interactions** are fundamental to all biological processes; for example, the **cell can be described as a complex network of chemicals connected by chemical reactions**. Cellular processes are controlled by various types of biochemical networks.

Types of biochemical networks :

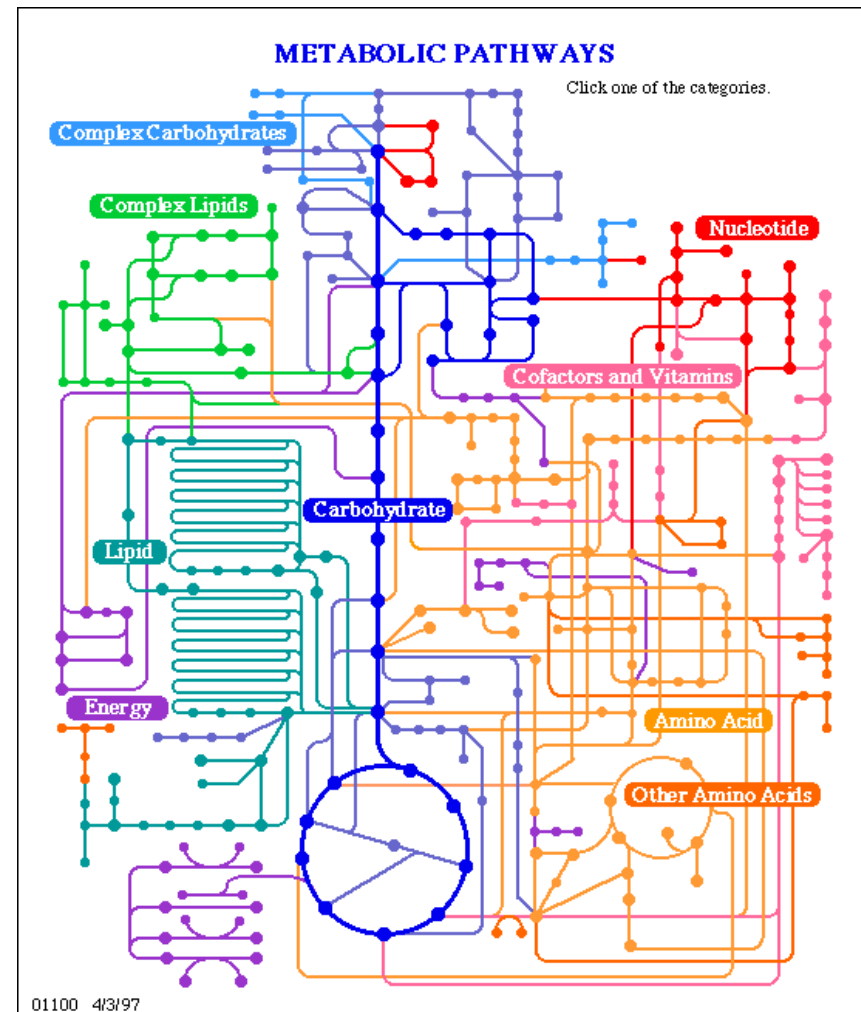
- (i) metabolic networks
- (ii) protein-protein interaction networks
- (iii) gene regulatory networks
- (iv) Cell signaling networks

```
graph LR; A["(i) metabolic networks  
(ii) protein-protein interaction networks  
(iii) gene regulatory networks  
(iv) Cell signaling networks"] --> B["Modular Framework"]
```

Modular Framework

# Biological Networks - metabolic networks

Metabolism is the most basic network of biochemical reactions, which generate energy for driving various cell processes, and degrade and synthesize many different bio-molecules.



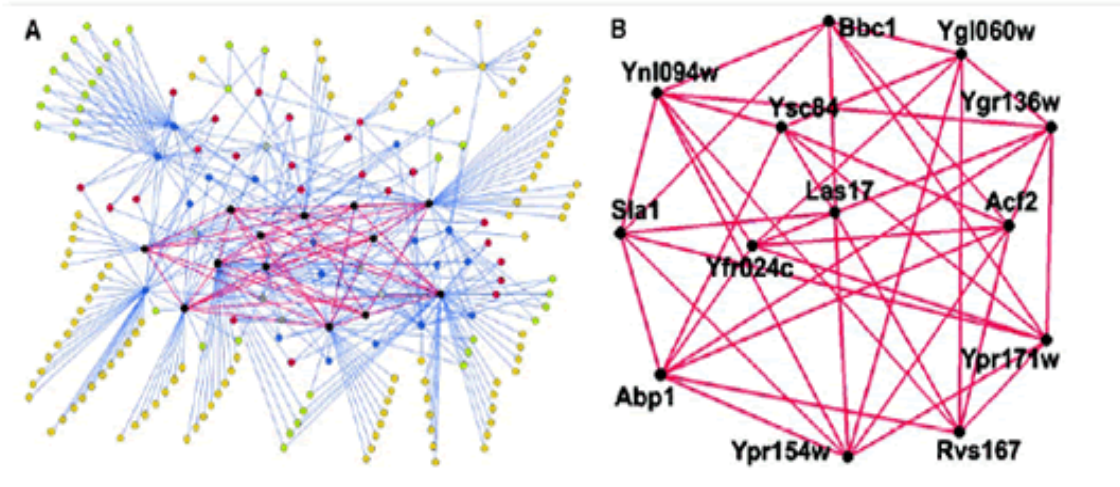
# Biological Networks

## – Protein-protein interaction networks

### Protein-protein interaction network (PIN)

Proteins perform distinct and well-defined functions, but little is known about how interactions among them are structured at the cellular level. Protein-protein interaction network, such as **binding interactions and formation of protein complex**.

- Experiment – Yeast two-hybrid method



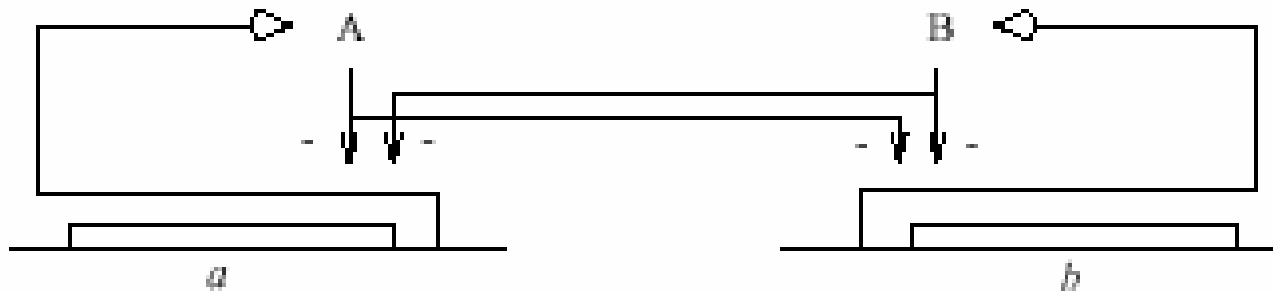
Limitation:  
**No subcellular location, and temporal information.**

# Biological Networks

## – gene regulation networks

---

In a gene regulatory network, the **protein encoded by a gene can regulate the expression of other genes**, for instance, by activating or inhibiting DNA transcription [Lewin 2000].. These genes in turn produce new regulatory proteins that control other genes.

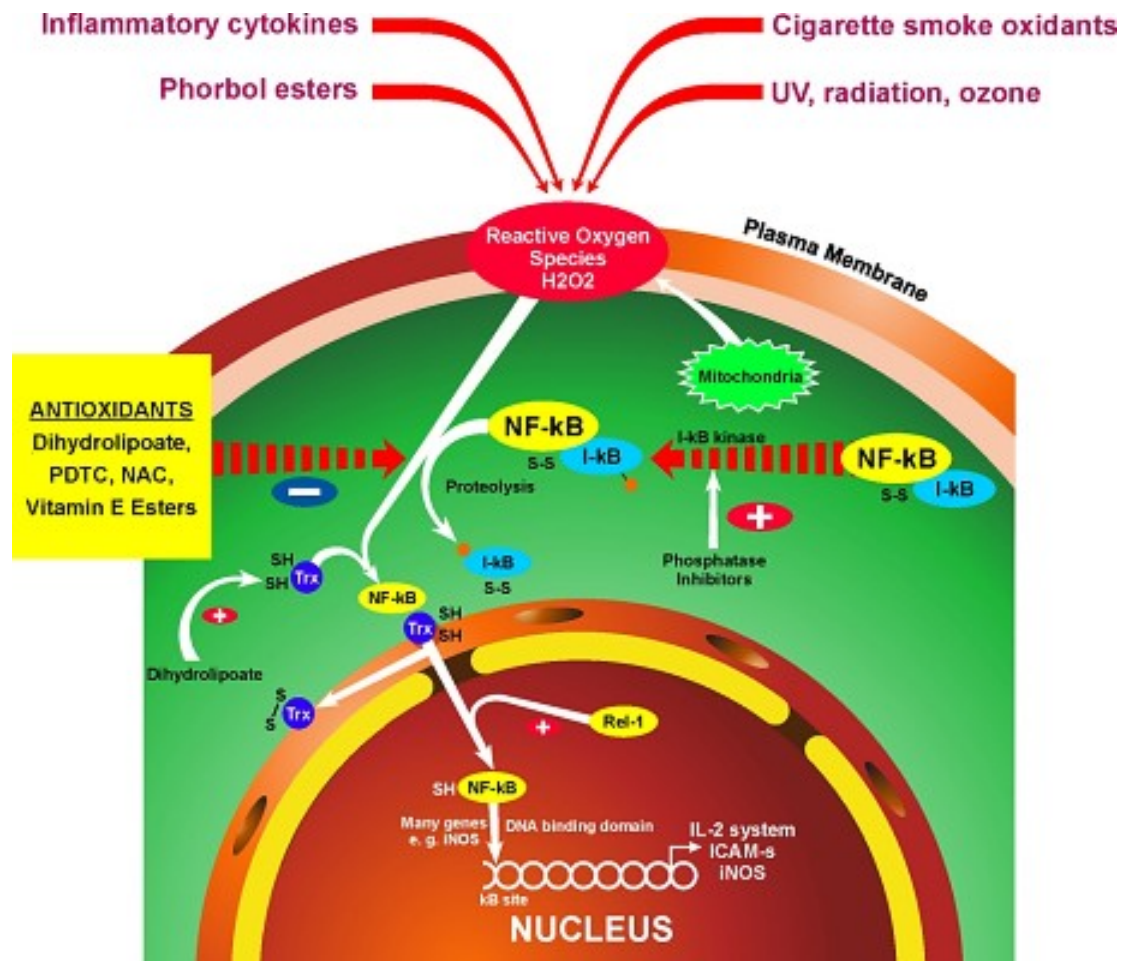


Example of a genetic regulatory network of two genes (*a* and *b*), each coding for a regulatory protein (A and B).

Recently, it was reported that [Farkas et al. 2003] in the yeast organism, there is **a hierarchical and combinatorial organization of transcriptional activity pattern.**

# Biological Networks

## – Signal transduction networks



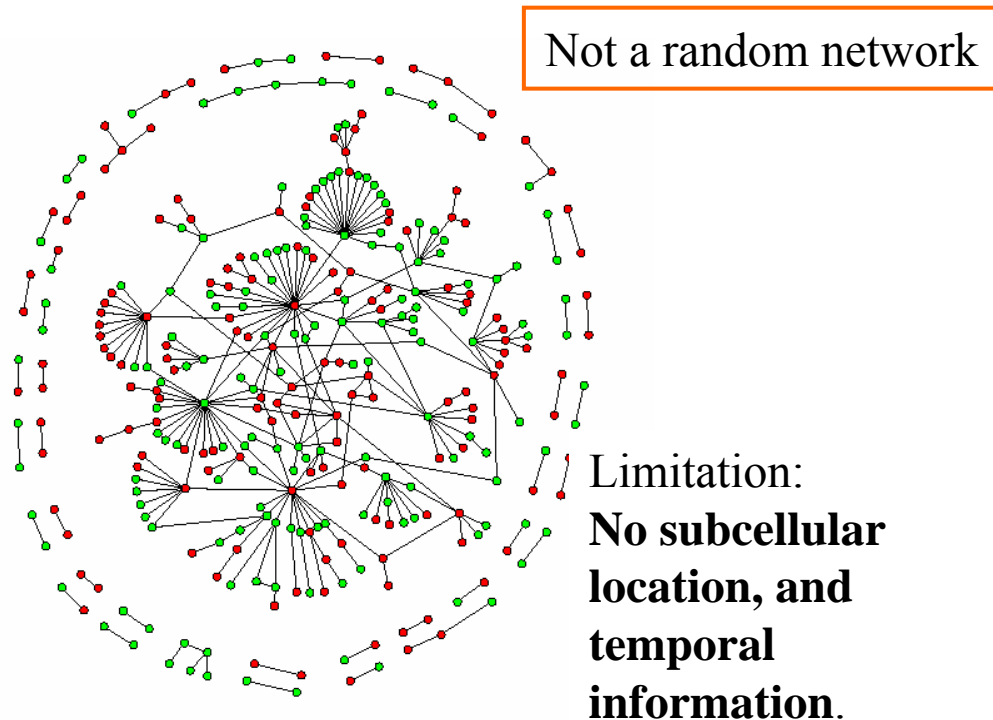
Nuclear transcription factor NF- $\kappa$ B  
 -control of apoptosis (cell suicide),  
 -development of B and T cells,  
 -anti-viral and bacterial responses

Oxidant-induced activation of NF- $\kappa$ B signal transduction

# From Genomes to Biological Networks - PIN

## Yeast Protein-protein interaction network

- Experiment – Yeast two-hybrid method
- protein-protein interactions are *not random*
- *highly connected proteins* are *unlikely to interact* with each other.



- Data from the high-throughput two-hybrid experiment (T. Ito, *et al.* PNAS (2001) )
- The full set containing **4549** interactions among **3278** yeast proteins
- **87%** nodes in the largest component
- $k_{\max} \sim 285$  !
- Figure shows nuclear proteins only

# From Genomes to Biological Networks

---

Biological networks are *not randomly connected*

Underlying architecture → *clustering*

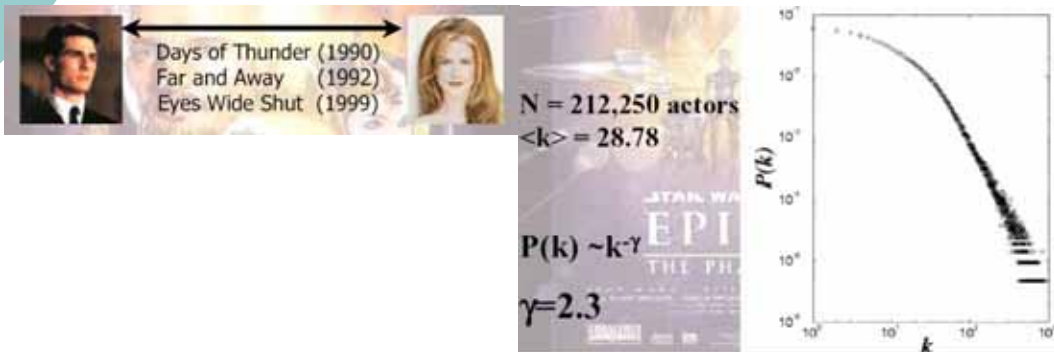
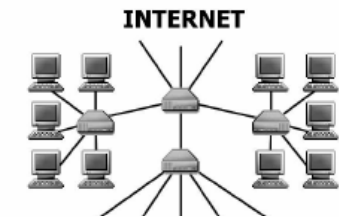
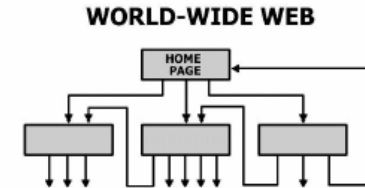
*How to characterize ?*

*Are there any similar features  
across different species ?*

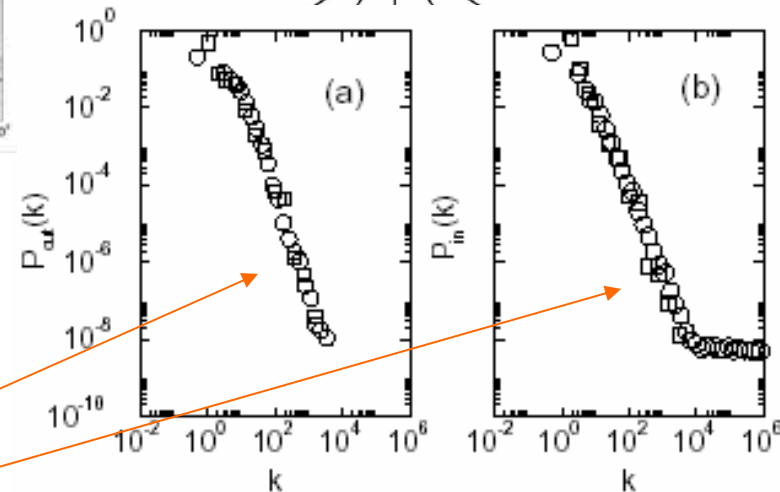
# Real networks

Many networks show scale-free behavior

- World-Wide Web
- Internet
- Ecology network (food web)
- Science collaboration network
- Movie actor collaboration network



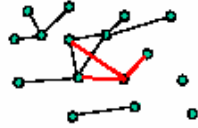
- Cellular network
- Network in linguistic
- Power and neural network
- Sexual contacts within a population (important for disease prevention!)
- etc.



**Power law behavior**

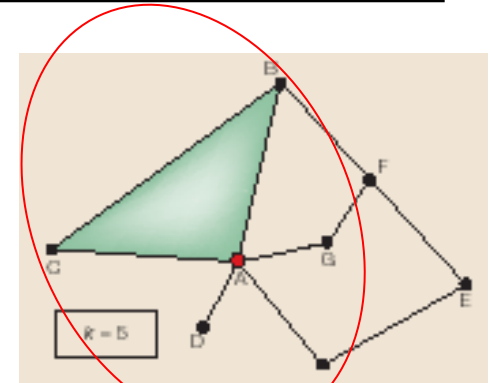
# Graph Theory – Basic concepts

## Node degree



## Clustering coefficient $C_i$

if A-B, B-C, then it is highly probable that A-C



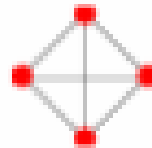
## Components



$$C_i = \frac{2E_i}{k_i(k_i - 1)}$$

$$C_A = \frac{2 \cdot 1}{5(5-1)} = 0.1$$

## Complete graph (Clique)

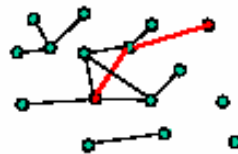


Two ways to compute  $C_i$

-  $E_i$  actual connections out of  $C^k_2$  possible connections

## Shortest path length

- number of triangles that included  $2i/k_i(k_i-1)$



## Average clustering coefficient

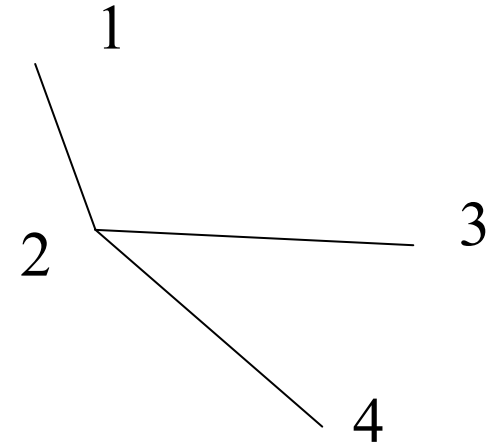
$$C = \frac{1}{N} \sum_{i=1}^N C_i$$

# Graph Theory – Adjacency matrix

---

## Undirected graph

$$M = \begin{pmatrix} 0 & 1 & \infty & \infty \\ 1 & 0 & 1 & 1 \\ \infty & 1 & 0 & \infty \\ \infty & 1 & \infty & 0 \end{pmatrix} \quad \begin{matrix} k_i \\ 1 \\ 3 \\ 1 \\ 1 \end{matrix}$$



- means not directly connected
- node  $i$  connectivity,  $k_i = \text{count}_j(m_{ij} = 1)$

## Bipartite graph

$$M = \begin{pmatrix} 0 & B \\ B^T & 0 \end{pmatrix}$$



# Graph Theory – average network distance

---

Interaction path length or *average network distance*,  $d$

- the **average of the distances between all pairs of nodes**
- frequency of the shortest interaction path length,  $f(L)$
- determined by using the Floyd algorithm

The average network diameter  $d$  is given by

$$d = \frac{\sum_L Lf(L)}{\sum_L f(L)}$$

where  $L$  is the shortest path length between two nodes.

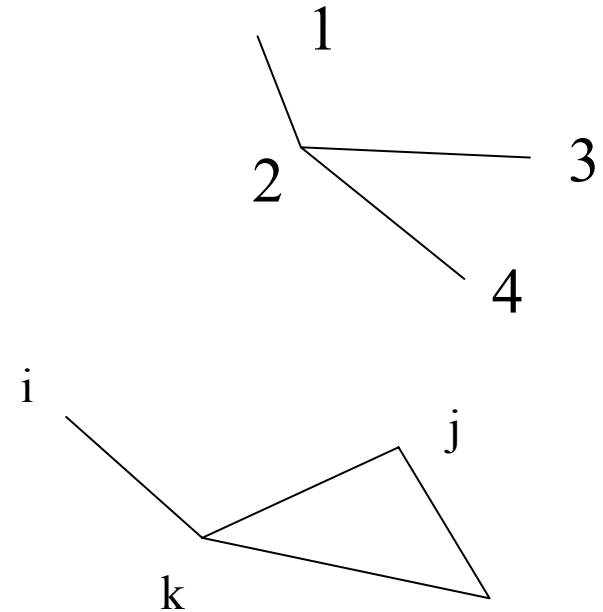
Network diameter (global)  $\neq$  Average network distance (local)

# Graph Theory – the shortest path

---

## The shortest path

- Floyd algorithm, an  $O(N^3)$  algorithm. For iteration  $n$ ,  
$$M^n_{ij} = \min \{ M^{n-1}_{ij}, M^{n-1}_{ik} + M^{n-1}_{kj} \}$$
- search for all possible paths, e.g. 1-2, 1-2-3, 1-2-4, 2-3, 2-4



# Random Graph Theory

= Graph Theory + Probability

---



Paul Erdős and A. Rényi,

On the evolution of random graphs

Magyar Tud. Akad. Mat. Kut. Int. Kozl. 5 (1960) 17-61.

# Random Graph Theory

## = Graph Theory + Probability

---

Random graph (Erdos and Renyi, 1960)

$$N = 4 \rightarrow C_n^6$$

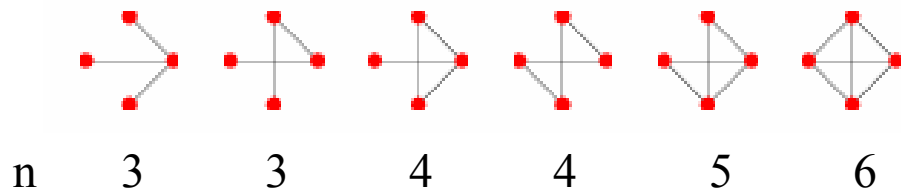
$N$  nodes labeled and connected by  $n$  edges

→  $C_n^N = N(N-1)/2$  possible edges

→  $C_n^{\binom{N}{2}}$  possible graphs with  $N$  nodes and  $n$  edges

n	Number of possible graphs, $C_n^6$
1	6
2	15
3	20
4	15
5	6
6	1

$N = 4$



# Random Graph Theory

## – Random network, Scale free network

### Connectivity distribution $P(k)$

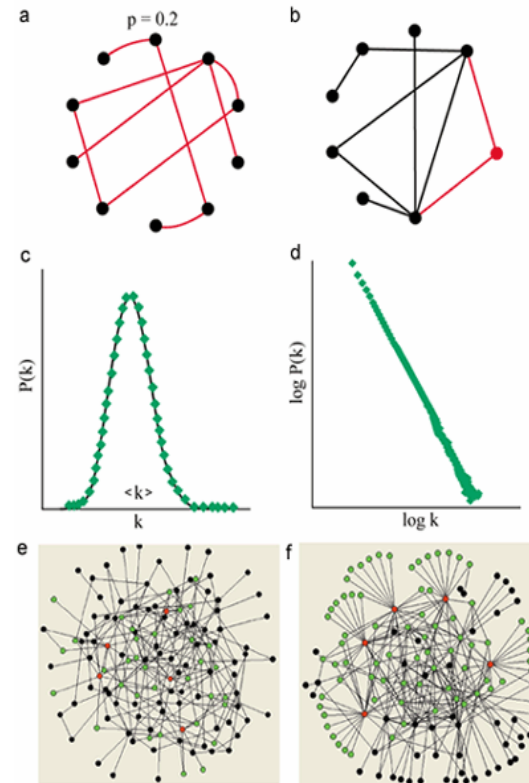
In a **random network**, the links are randomly connected and most of the nodes have degrees close to  $\langle k \rangle = 2E/N$ . The degree distribution  $P(k)$  vs.  $k$  is a **Poisson distribution**, i.e.  $P(k) \sim \langle k \rangle^k e^{-\langle k \rangle} / k!$  for  $k \ll \langle k \rangle$  and  $k \gg \langle k \rangle$ .

**power-law distribution**,  $P(k) \sim k^{-\gamma}$ , where  $\gamma$  is a constant. Such networks are known as **scale-free network**.

Random network  $\rightarrow$   $\text{Log}[P(k)]$  vs  $\text{Log}[k]$  plot has a peak  
 $\rightarrow$  homogenous nodes  
 $\rightarrow d \sim \log N$

Scale-free network  $\rightarrow$   $\text{Log}[P(k)]$  vs  $\text{Log}[k]$  plot is a line with negative slope  
 $\rightarrow$  inhomogenous nodes  
 $\rightarrow d \sim \log(\log N)$

**To reduce fluctuation**, we consider  $P_{cum}(k) \sim k^{-\alpha}$   
 where  $\gamma = \alpha + 1$



Random network      Scale-free network

<http://physicsweb.org/box/world/>

# The Scale-free network model (BA model)

---

The scale-free network model is defined as follows:

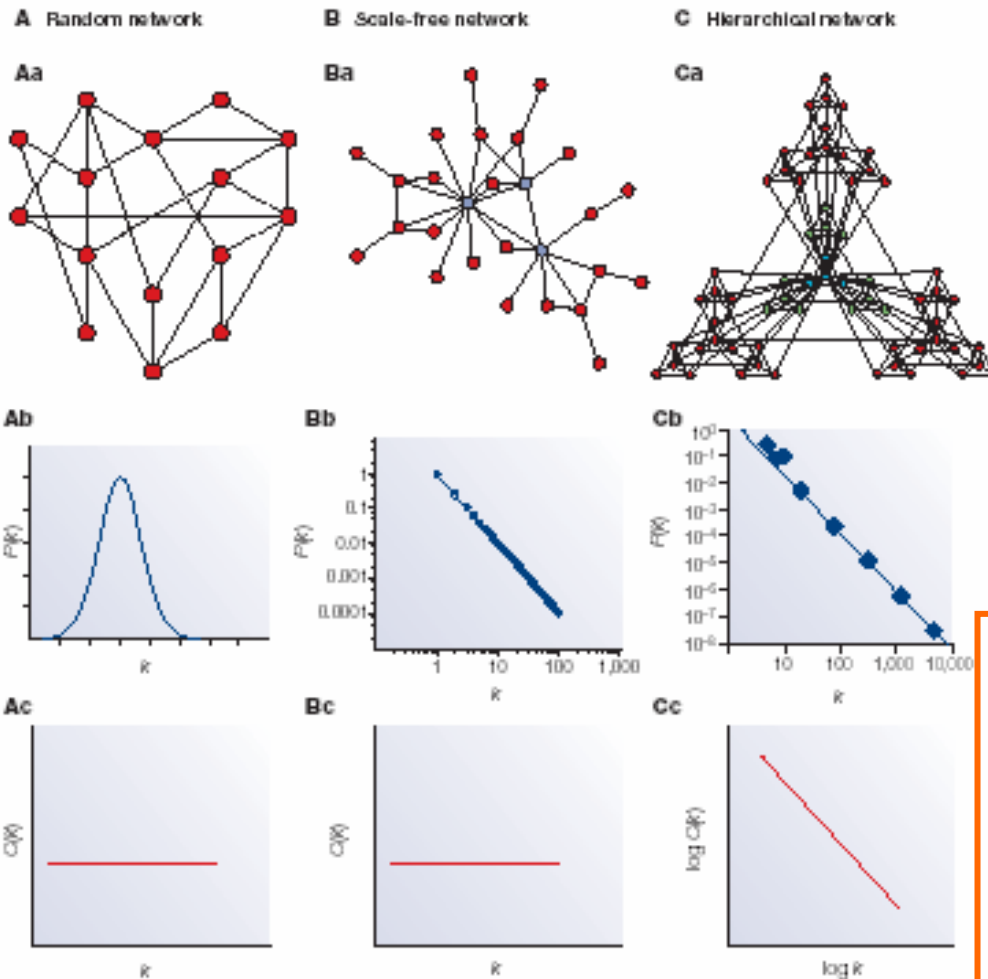
1. **Growth** - add a new vertex with  $m$  ( $\leq m_0$ ) edges at each time-step
2. **Preferential attachment**: A link is added to an old node in proportion to its connectivity

$$\frac{\partial k_i}{\partial t} = m\Pi(k_i)$$

$$\Pi(k_i) = \frac{k_i}{\sum_j k_j}$$

$$\Rightarrow P(k) \sim k^{-3}$$

# Random Graph, Scale-free network, Hierarchical network



*Node degree distribution*

*Clustering coefficient*

Hierarchical network - coexistence of  
 (1) modularity,  
 (2) local clustering, and  
 (3) scale-free behavior

scaling  $C_{ave}(k) \sim k^{-\beta}$   
 $\beta = 1$  or  $> 1$  for  
**Deterministic/Stochastic hierarchical network Model.** See Ravasz and Barabasi, PRE 2003

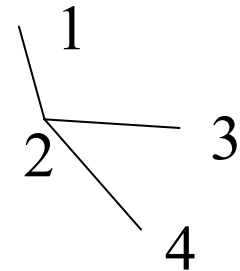
# Graph Theory – Node degree correlation

- *random graph* models → node degrees are *uncorrelated*
- count the frequency  $P(K_1, K_2)$  that two proteins with connectivity  $K_1$  and  $K_2$  connected to each other by a link
- compared it to the same quantity  $P_R(K_1, K_2)$  measured in a randomized version of the same network.
- The average node connectivity *for a fixed  $K_1$*  is given by,

$$\langle K_2 \rangle = \sum K_2 \frac{P(K_1, K_2)}{\langle P_R(K_1, K_2) \rangle}$$

where  $\langle \rangle$  denotes the multiple sampling average, and the summation sums for all  $K_2$  with a fixed  $K_1$ .

- In the randomized version, the *node degrees of each protein are kept the same as in the original network*, whereas their linking partner is totally random.



$$M = \begin{pmatrix} 0 & 1 & \infty & \infty \\ 1 & 0 & 1 & 1 \\ \infty & 1 & 0 & \infty \\ \infty & 1 & \infty & 0 \end{pmatrix}$$

# Network Robustness

To test whether the interaction **network is robust against mutation or experimental errors**, such as false negative or false positive arising from the yeast two-hybrid (Y2H) screening method.

Consider four types of perturbations,

(i) remove nodes randomly (failure) → protein malfunction (FN)

(ii) remove nodes with highest connectivity in descending order (attack)

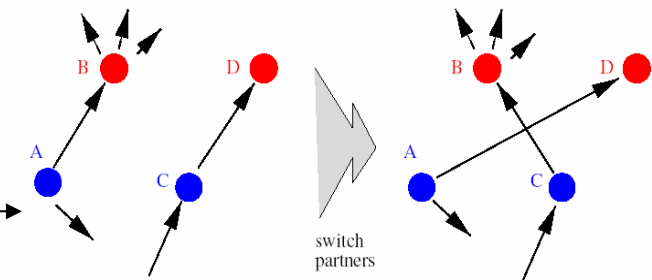
(iii) rewire nodes randomly (introduce FP and FN)

(iv) remove links randomly → protein malfunction (FN)

Multiple sampling of the randomized networks

→  $\langle d_{pert} \rangle / d$

Nodes rewiring

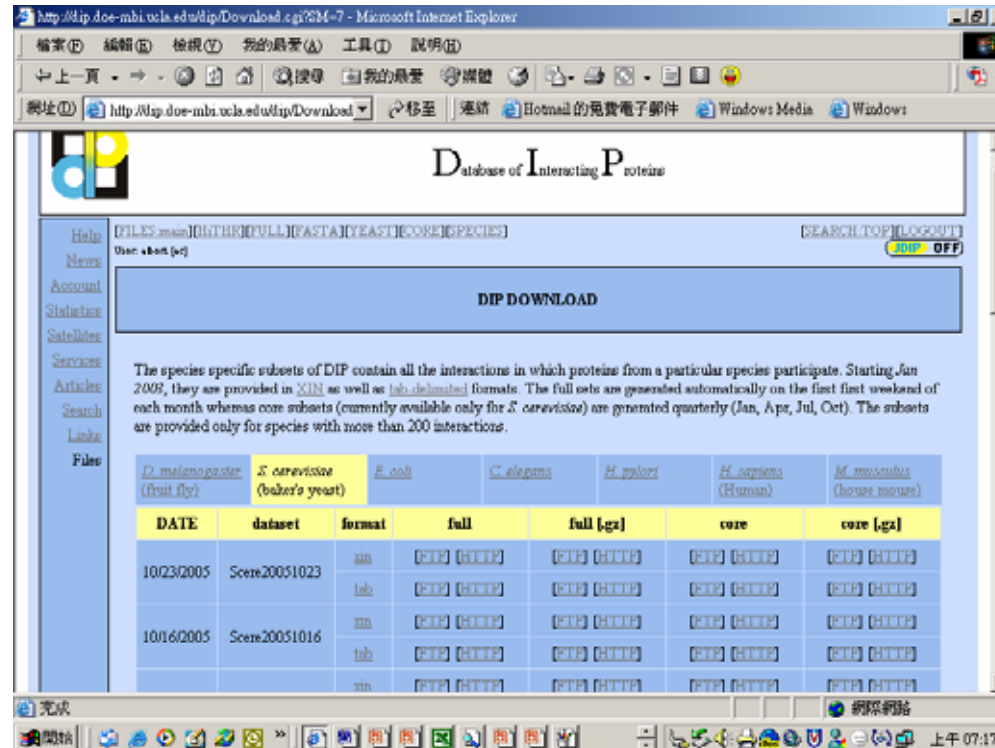


*this process will not change the degree of connectivity of each node*

# Input - Database of Interacting Proteins (DIP)

DIP <http://dip.doe-mbi.ucla.edu>

DIP is a database that documents experimentally determined protein-protein interactions. We analyze the protein-protein interaction for seven different species, *S. cerevisiae*, *E.coli*, *H.pylori*, *C. elegans*, *D. melanogaster*, *H. sapiens* and *M. musculus*.



The screenshot shows the DIP website interface in a Microsoft Internet Explorer browser. The page title is "Database of Interacting Proteins". The main content area is titled "DIP DOWNLOAD" and contains a table of species-specific datasets. The table has columns for species, date, dataset ID, format, and download links for full and core datasets. The species listed are *D. melanogaster*, *S. cerevisiae*, *E. coli*, *C. elegans*, *H. pylori*, *H. sapiens*, and *M. musculus*.

Species	DATE	dataset	format	full	full [gz]	core	core [gz]
<i>D. melanogaster</i> (fruit fly)	10/23/2005	Scere20051023	sh	<a href="#">FTP</a> <a href="#">HTTP</a>	<a href="#">FTP</a> <a href="#">HTTP</a>	<a href="#">FTP</a> <a href="#">HTTP</a>	<a href="#">FTP</a> <a href="#">HTTP</a>
			tab	<a href="#">FTP</a> <a href="#">HTTP</a>	<a href="#">FTP</a> <a href="#">HTTP</a>	<a href="#">FTP</a> <a href="#">HTTP</a>	<a href="#">FTP</a> <a href="#">HTTP</a>
			sh	<a href="#">FTP</a> <a href="#">HTTP</a>	<a href="#">FTP</a> <a href="#">HTTP</a>	<a href="#">FTP</a> <a href="#">HTTP</a>	<a href="#">FTP</a> <a href="#">HTTP</a>
			tab	<a href="#">FTP</a> <a href="#">HTTP</a>	<a href="#">FTP</a> <a href="#">HTTP</a>	<a href="#">FTP</a> <a href="#">HTTP</a>	<a href="#">FTP</a> <a href="#">HTTP</a>
<i>S. cerevisiae</i> (baker's yeast)	10/16/2005	Scere20051016	sh	<a href="#">FTP</a> <a href="#">HTTP</a>	<a href="#">FTP</a> <a href="#">HTTP</a>	<a href="#">FTP</a> <a href="#">HTTP</a>	<a href="#">FTP</a> <a href="#">HTTP</a>
			tab	<a href="#">FTP</a> <a href="#">HTTP</a>	<a href="#">FTP</a> <a href="#">HTTP</a>	<a href="#">FTP</a> <a href="#">HTTP</a>	<a href="#">FTP</a> <a href="#">HTTP</a>
			sh	<a href="#">FTP</a> <a href="#">HTTP</a>	<a href="#">FTP</a> <a href="#">HTTP</a>	<a href="#">FTP</a> <a href="#">HTTP</a>	<a href="#">FTP</a> <a href="#">HTTP</a>
			tab	<a href="#">FTP</a> <a href="#">HTTP</a>	<a href="#">FTP</a> <a href="#">HTTP</a>	<a href="#">FTP</a> <a href="#">HTTP</a>	<a href="#">FTP</a> <a href="#">HTTP</a>

- Look for general and different features of PIN for different species

# The seven species

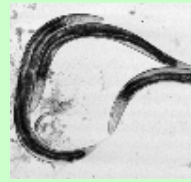
---



*E. coli*



*H. pylori*



*C. elegans*



*D. melanogaster*



*H. sapiens*



*M. musculus*



*S. cerevisiae*

# Results – PPI statistics

Table 1. DIP database statistics for *E. coli*, *H. pylori*, *C. elegans*, *D. melanogaster*, *H. sapiens*, *M. musculus* and *S. cerevisiae*, the total number of proteins  $N_p$ , total number of interactions  $N_E$ , the estimated gene numbers, total number of proteins in the largest cluster  $N_{largest}$ , and the ratio of the size of the largest cluster to the total number of proteins  $N_{largest}/N_p$ .

Organism	$N_p$	$N_E$	Estimate gene numbers	$N_{largest}$	$N_{largest}/N_p$
<i>E. coli</i>	336	611	~4000	145	0.43
<i>H. pylori</i>	702	1420	~1500	686	0.98
<i>C. elegans</i>	2629	4030	~19000	2386	0.91
<i>D. melanogaster</i>	7057	20988	~13000	6926	0.98
<i>H. sapiens</i>	1059	1369	~30000	563	0.53
<i>M. musculus</i>	327	286	~37000	49	0.15
<i>S. cerevisiae (CORE)</i>	2609	6574	~6300	2440	0.94

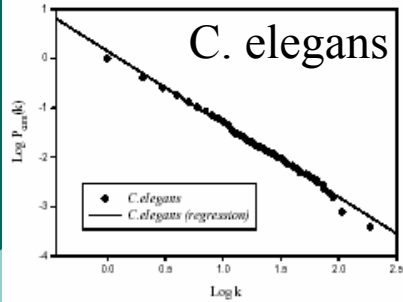
- The largest connected clusters (>90% of the nodes)
- minimize experimental uncertainty → CORE subset of budding yeast, *S. cerevisiae*
- according to the criteria described in Deane CM, Salwinski L, Xenarios I, Eisenberg D. (2002). Protein Interactions. *Mol Cell Prot* 1, 349.

# Results – scale-free network study

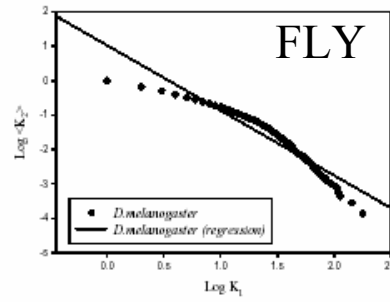
Table II. The average node degree  $\langle k \rangle$ , standard deviation of node degree,  $\sigma_k$ , maximum connectivity  $k_{max}$ , average network diameter  $d$ , and the regression and correlation results of the  $Ln P_{cum}(k)$  vs  $Ln k$  plots ( $\alpha$ ,  $r^2$  and  $p$ -value) for the seven species.

$P_{cum}(k)$  vs  $Ln k$  plots ( $\alpha$ ,  $r^2$  and  $p$ -value) for the seven species.

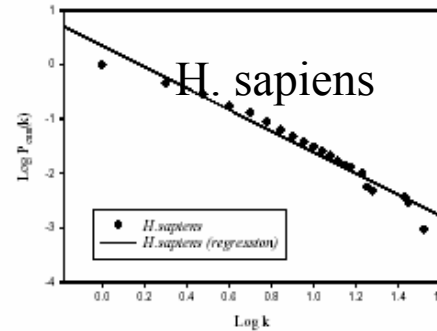
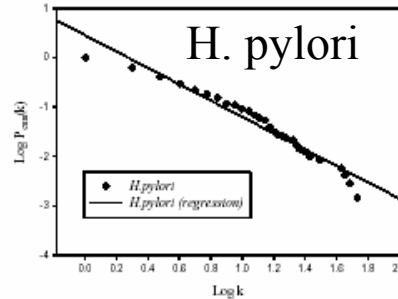
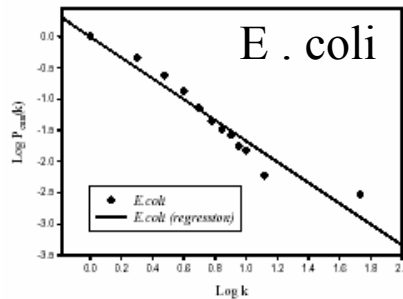
Species	$\langle k \rangle \pm \sigma_k$	$k_{max}$	$d$	$\alpha$	$r^2$	$p$ -value
<i>E. coli</i>	$2.17 \pm 3.31$	54	5.63	$1.67 \pm 0.14$	0.93	$\ll 10^{-6}$
<i>H. pylori</i>	$3.87 \pm 5.42$	54	4.14	$1.65 \pm 0.07$	0.95	$\ll 10^{-6}$
<i>C.elegans</i>	$3.02 \pm 7.19$	187	4.81	$1.48 \pm 0.02$	0.99	$\ll 10^{-6}$
<i>D. melanogaster</i>	$5.90 \pm 9.37$	178	4.02	$1.90 \pm 0.02$	0.93	$\ll 10^{-6}$
<i>M. musculus</i>	$1.68 \pm 1.40$	12	3.57	$1.95 \pm 0.09$	0.96	$\ll 10^{-6}$
<i>H. sapiens</i>	$2.49 \pm 2.99$	33	6.80	$2.34 \pm 0.16$	0.96	$\ll 10^{-6}$
<i>S. cerevisiae(CORE)</i>	$4.87 \pm 6.07$	111	5.01	$1.97 \pm 0.09$	0.91	$\ll 10^{-6}$



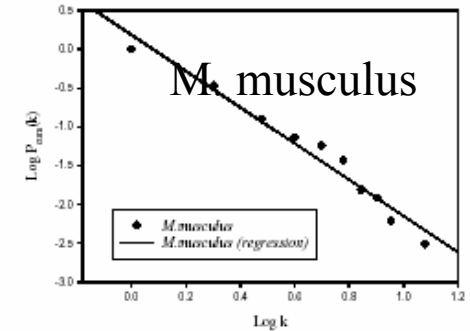
(a)



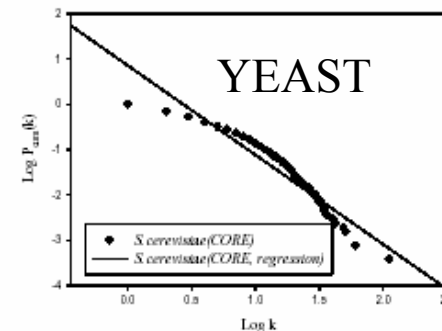
(b)



(e)



(f)



- Large standard deviation of  $k$
  - Coefficient of determination,  $r^2 = SSR/SST > 0.90$
  - To account for the flat plateau and long tail behaviors, assume a short-length scale correction  $k_0$  and an exponential cut-off tail at  $k_c$
- $$P(k) \propto (k + k_0)^{-\gamma} e^{-k/k_c}$$

# Results – Hierarchical structures

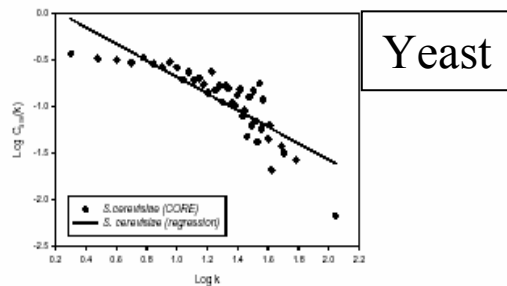


Fig. 2. The  $\text{Log } C_{ave}(k)$  vs  $\text{Log } k$  plot for *S. cerevisiae*, and the regression line.

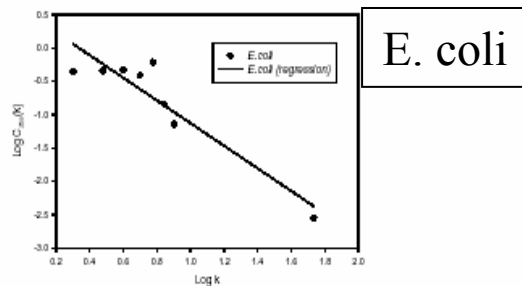
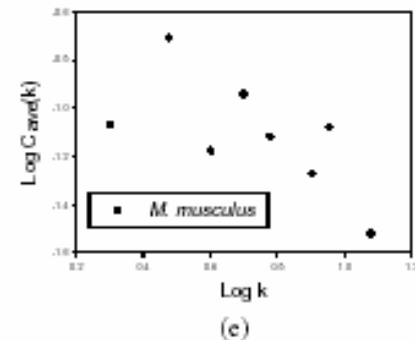
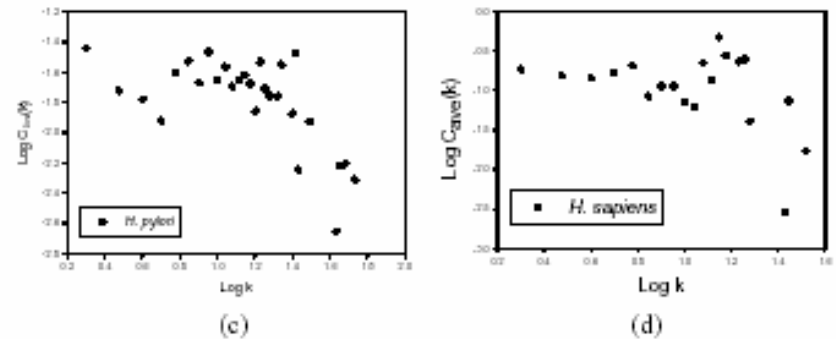
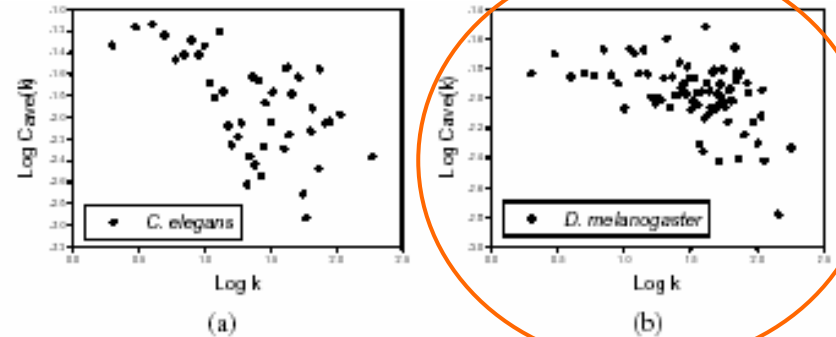


Fig. 3. The  $\text{Log } C_{ave}(k)$  vs  $\text{Log } k$  plot for *E. coli* and the regression line.

$$C_{ave}(k) \sim k^{-\beta}$$

The  $\text{Log } C_{ave}(k)$  vs  $\text{Log } k$  plots for the seven species.



# Results – Hierarchical structures

Table III. Regression and correlation results of the  $\text{Log } C_{ave}(k)$  vs. the  $\text{Log } k$  plots ( $\beta$ ,  $r^2$  and  $p$ -value), and the cluster coefficients  $C$  for the seven species.

Species	$\beta$	$r^2$	$p$ -value	$C$	$C$ [22]
<i>E. coli</i>	1.70±0.28	0.86	8.3*10 <sup>-4</sup>	0.43	
<i>H. pylori</i>	0.49±0.13	0.37	5.6*10 <sup>-4</sup>	0.025	
<i>C. elegans</i>	0.63±0.13	0.36	2.0*10 <sup>-5</sup>	0.049	0.047
<i>D. melanogaster</i>	0.26±0.05	0.24	8.2*10 <sup>-6</sup>	0.015	0.014
<i>H. sapiens</i>	0.66±0.32	0.19	0.055	0.17	
<i>M. musculus</i>	0.60±0.28	0.43	0.077	0.10	
<i>S. cerevisiae</i> (CORE)	0.89±0.09	0.72	<<10 <sup>-6</sup>	0.30	
<i>S. cerevisiae</i>				0.13	0.12

Stochastic hierarchical network,  $\beta > 1$

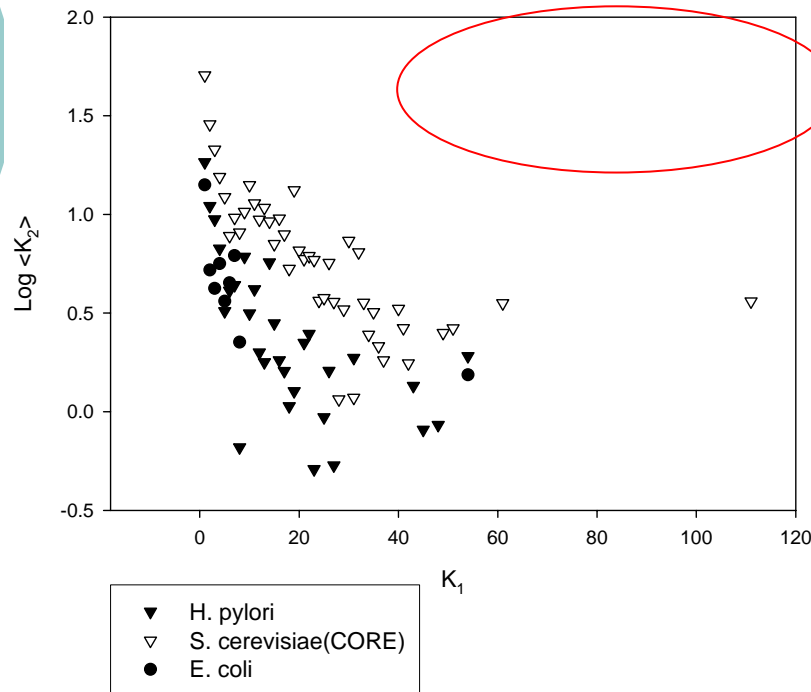
More like a scale-free network  $\beta = 0.26$  close to 0 rather 1, it exhibits a flat plateau for small  $k$

Hierarchical network model is a better description for *E.coli* and *S.cerevisiae* PINs, and it is probably **not a universal feature** across different species.

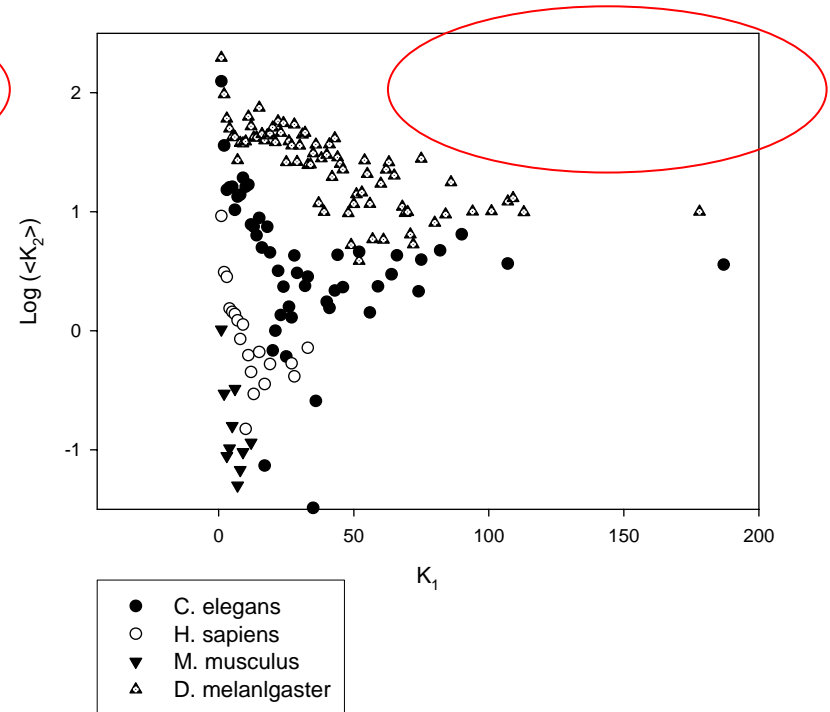
Modular structure → similar functions, involve in the same biological pathway

[22] Goh K.I., Kahng B. and Kim B. (2005). LANL archive, q-bio.MN/0502018

# Results – degree-degree correlations



Average node degree correlation profile of protein-protein interaction network for the single cellular organisms, *H. pylori*, *S. cerevisiae*(CORE) and *E. coli*.



Average node degree correlation profile of protein-protein interaction network for multiple cellular organisms, *C. elegans*, *D. melanogaster*, *H. sapiens* and *M. musculus*.

**Highly connected proteins are unlikely to interact with each other.**

# Robustness of Random and SF Networks

- blue symbols → failure
- diameter of the exponential (triangles)
- scale-free (squares)
- $f$  → a fraction of the nodes are removed randomly

- Red symbols → attack
- exponential (diamonds)
- scale-free (circles)
- the most connected nodes are removed first

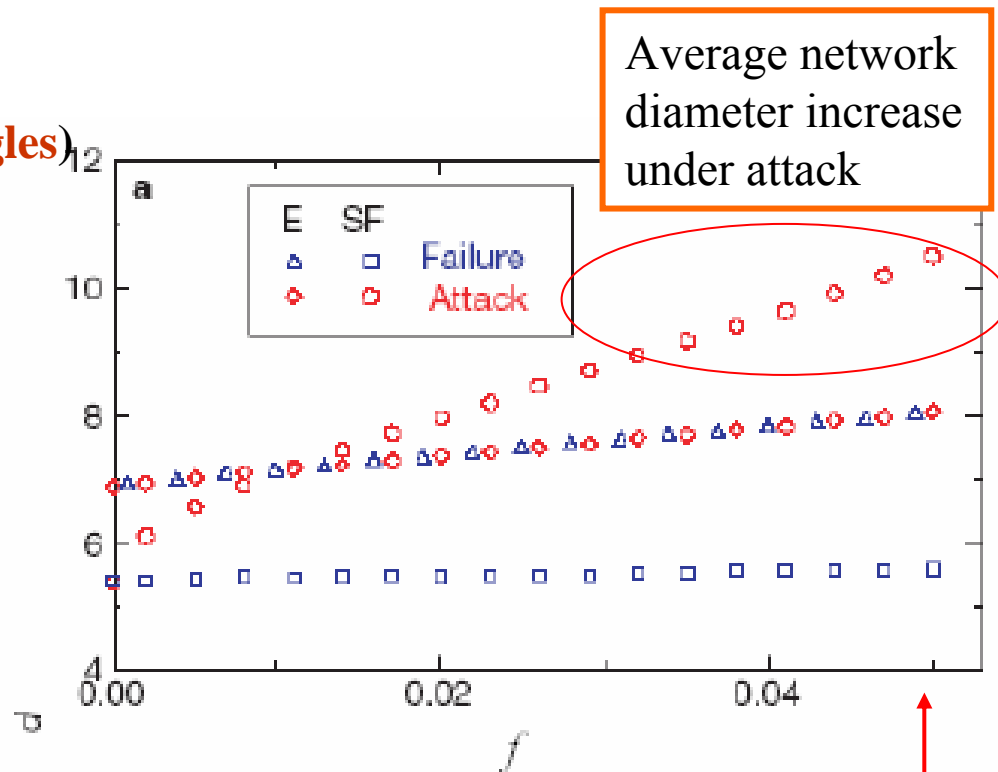
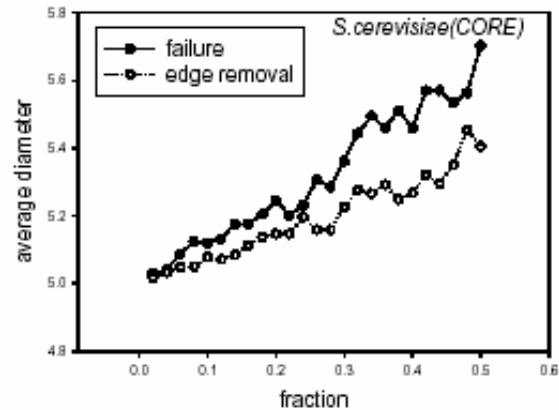
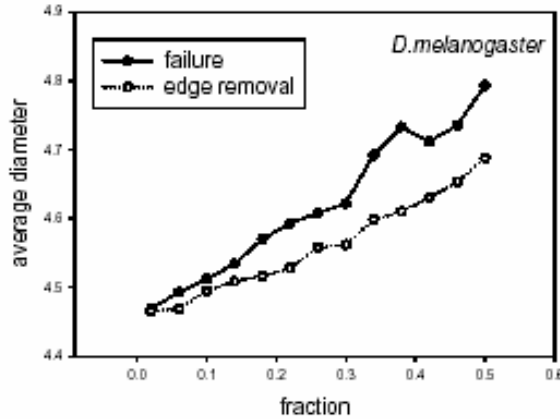
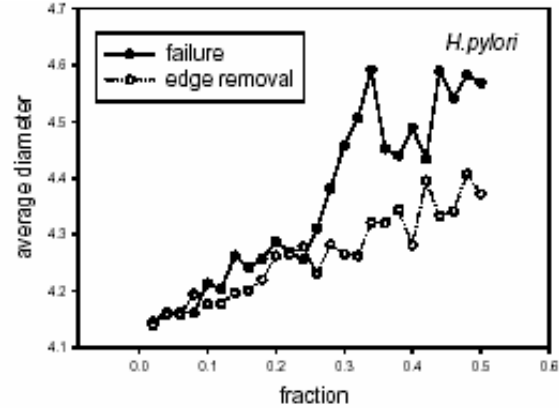
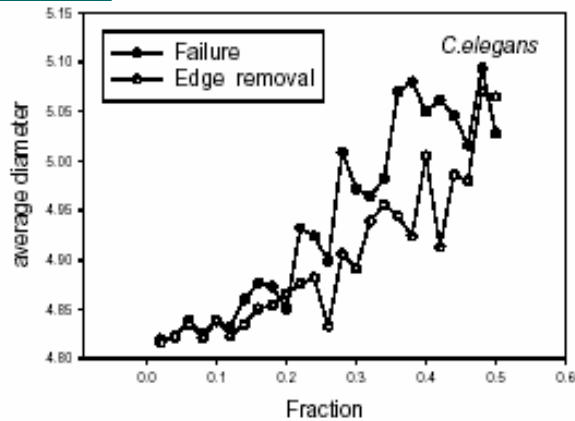


Fig. 2 diameter of network against a fraction of nodes are removed randomly.

scale-free networks display a surprisingly high degree of tolerance against random failures

# Results – remove nodes or edges randomly



Species	$N_p$	Interactions	$N_{cluster}$	$N_{largest}$	$N_{largest}/N_p$
<i>E. coli</i>	336	611	191	145	0.43
<i>H. pylori</i>	702	1420	17	686	0.98
<i>C. elegans</i>	2629	4030	108	2386	0.91
<i>D. melanogaster</i>	7057	20988	73	6926	0.98
<i>M. musculus</i>	327	1369	81	49	0.15
<i>H. sapiens</i>	1059	286	125	563	0.53
<i>S. cerevisiae(CORE)</i>	2609	6574	92	2440	0.94

Pick a network with the largest connected component contains >91% proteins.

Failure has a larger effect than edge removal.

# Results – edges rewiring

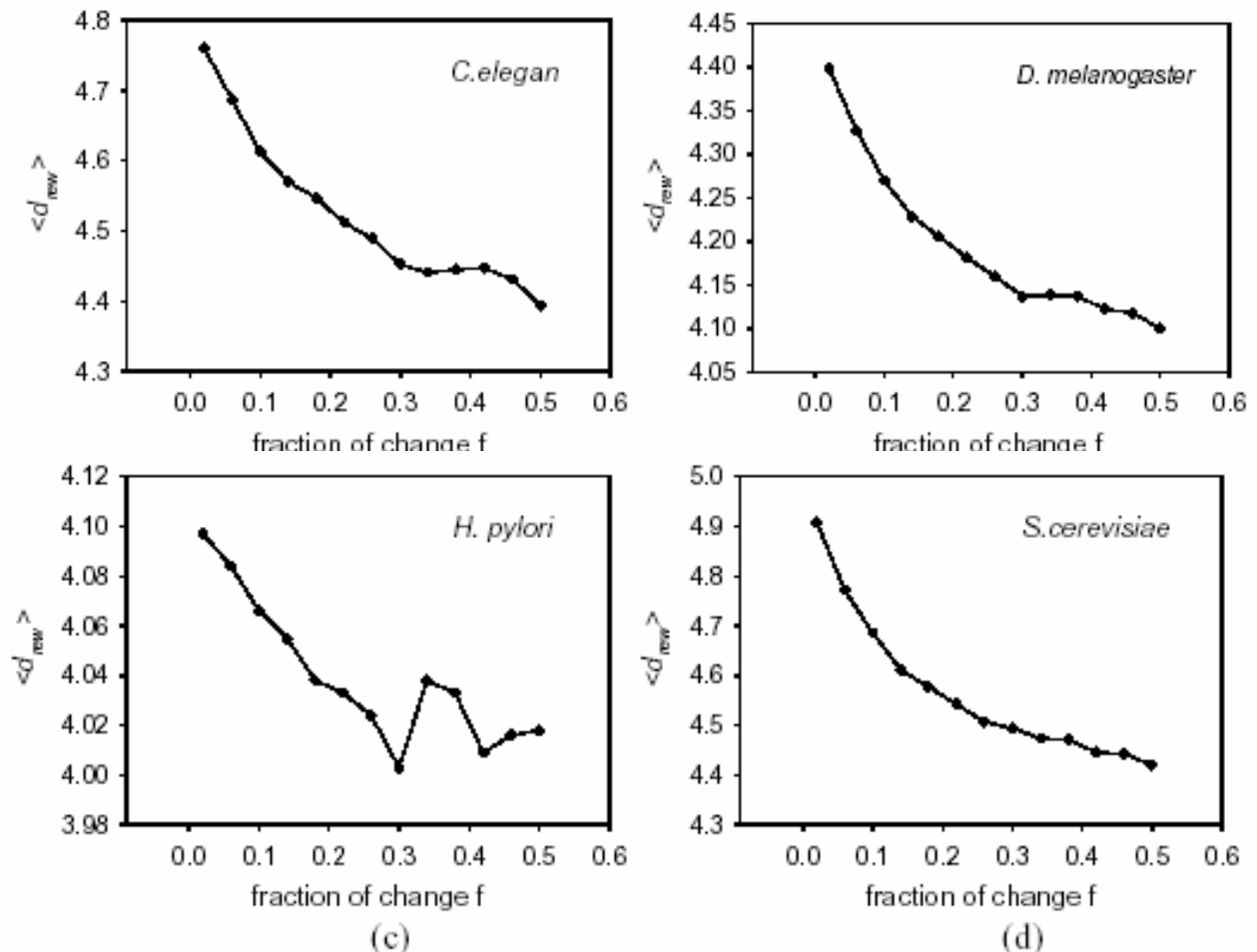


Figure 3. The average network diameter  $\langle d_{rew} \rangle$  vs. the fraction of change of rewired edges for the four species' PINs, (a) *C. elegans* (b) *D. melanogaster* (c) *H. pylori* and (d) *S. cerevisiae*.

# Results – comparison of failure and attack

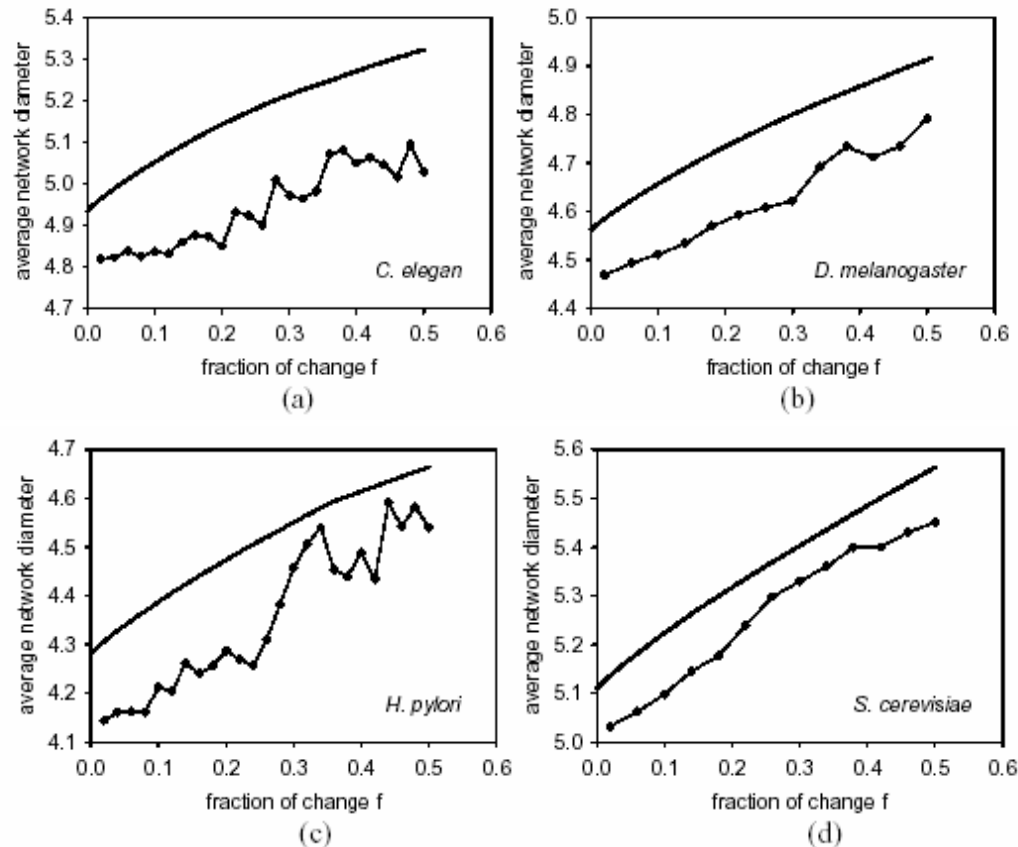


Figure 2. The average network diameter  $\langle d_{fail} \rangle$  (the see-saw line) and  $\langle d_{att} \rangle$  (the solid line) vs. the fraction of change  $f$  for the four species' PINs, (a) *C. elegans* (b) *D. melanogaster* (c) *H. pylori* and (d) *S. cerevisiae*.

**Hubs removal could cause a larger effects than failure.**

# Summary

## – failure, attack, rewire and edge removal

---

Table 3. A summary of the percentage changes for average failure diameter  $\Delta_{fail}$ , average attack diameter  $\Delta_{att}$ , average rewire diameter  $\Delta_{rew}$ , and average edge deletion diameter  $\Delta_{edge}$  results for the four species at the 50% level of changes.

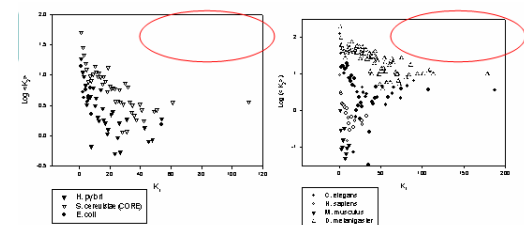
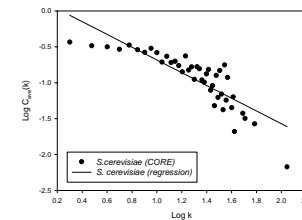
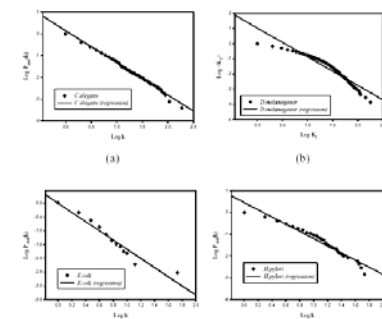
Species	$\Delta_{fail}$	$\Delta_{att}$	$\Delta_{rew}$	$\Delta_{edge}$
<i>H. pylori</i>	9.7	12.8	-2.9	5.6
<i>C.elegans</i>	5.8	10.6	-8.7	5.4
<i>D. melanogaster</i>	7.4	10.1	-10.1	5.2
<i>S. cerevisiae(CORE)</i>	8.8	11.0	-11.8	8.0

The average network diameters for the perturbed cases,  $\langle d_{fail} \rangle$ ,  $\langle d_{att} \rangle$ ,  $\langle d_{rew} \rangle$  and  $\langle d_{edge} \rangle$  differed from unperturbed cases with  $|\Delta_x| < 13\%$  at a 50% level of changes. Albert and Barabasi found that diameters of WWW and the Internet doubled its original value if 5% of the most connected nodes were removed. In contrast, the change in diameter is less drastic for PINs with respect to attack. Nevertheless, attack perturbation induced a larger effect on average network diameter relative to failure or edge deletion perturbation. The situation is somewhat inconclusive for rewiring perturbation because it could also induced comparable percentage change as attack perturbation.

# Summary

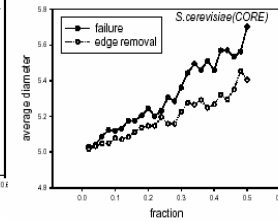
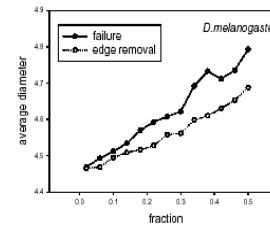
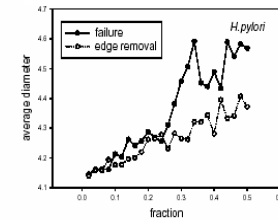
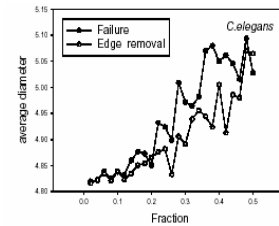
## Protein-protein interaction Network

- PINs are not random networks
- The  $\text{Log}[P_{cum}(k)]$  vs  $\text{Log}[k]$  study indicates that PINs *are well approximate by scale-free networks*
- *Hierarchical network model is a better description for certain species' PINs*
- Node degree correlation study supports the earlier result that *nodes of low connectivity are correlated*, whereas *nodes of high connectivity are not directly linked*



# Summary

PINs are topologically robust with respect to failure, attack, nodes rewiring and edges removal and robust against experimental artifacts, i.e. the false negative and false positive errors.



# Acknowledgement

---

- 1. Professor Chien-Hung Huang (黃建宏)**  
**Department of Computer Science and Information Engineering**  
**National Formosa University**  
**資訊工程系 國立虎尾科技大學**
- 2. Dr. Po-Han Lee (李柏翰)**  
**The Affiliated Senior High School of National Taiwan Normal University**  
**師大附中**
- 3. Professor Jywe-Fei Fang (方覺非 )**  
**Department of Computer Science and Information Engineering**  
**Department of Digital Content and Technology, National Taichung Univ.**  
**數位內容科技系 國立台中大學**
- 4. Professor Jeffery J.F. Tsai (蔡進發)**  
**Department of Computer Science, University of Illinois, USA.**
- 5. National Science Council (國科會)**



*Thank you.*