

Complex Network Analysis Jan. 15<sup>th</sup>, 2018

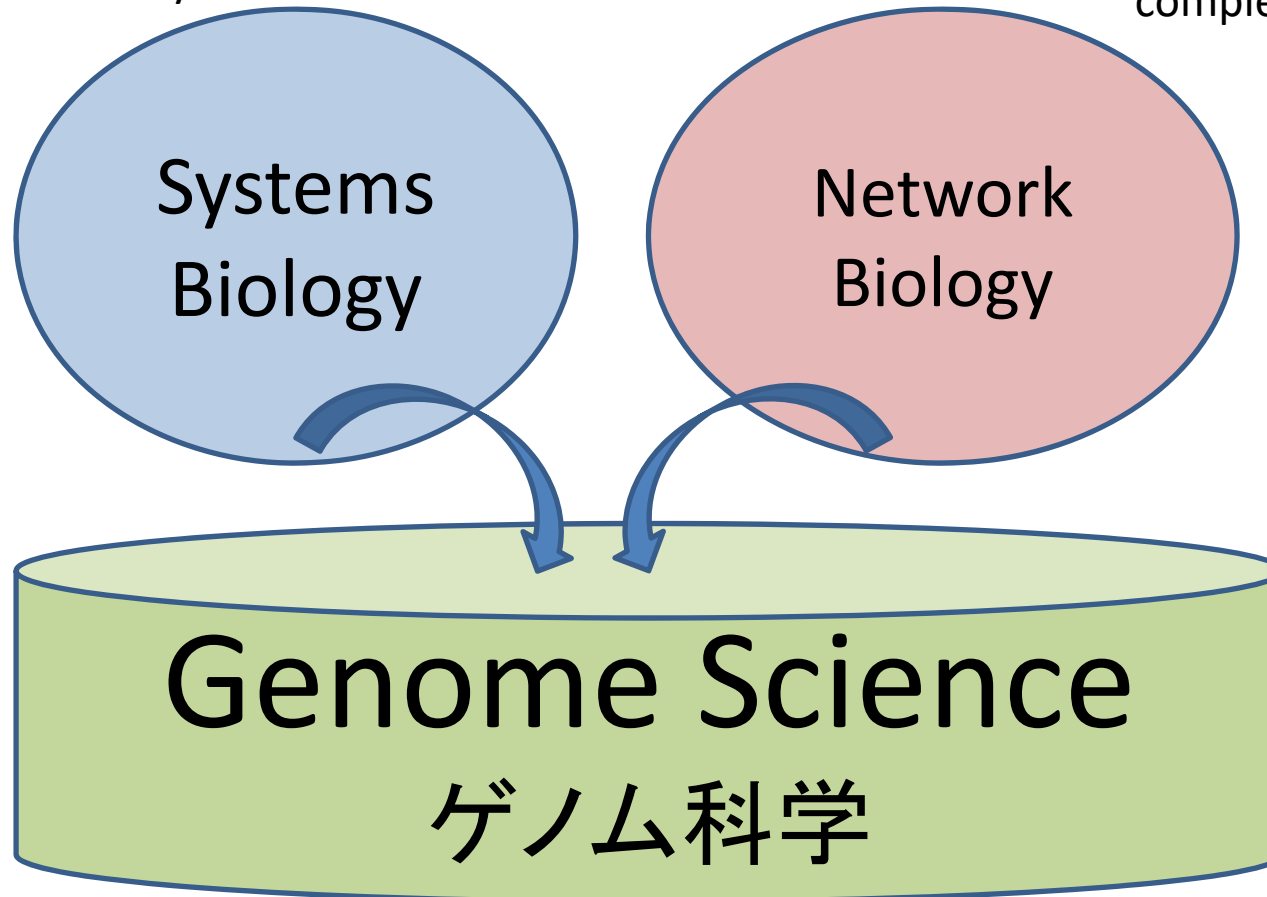
# Protein-Protein Interaction Network Analysis

Akihiko Konagaya  
School of Computing

# Network Biology

To understand life as system

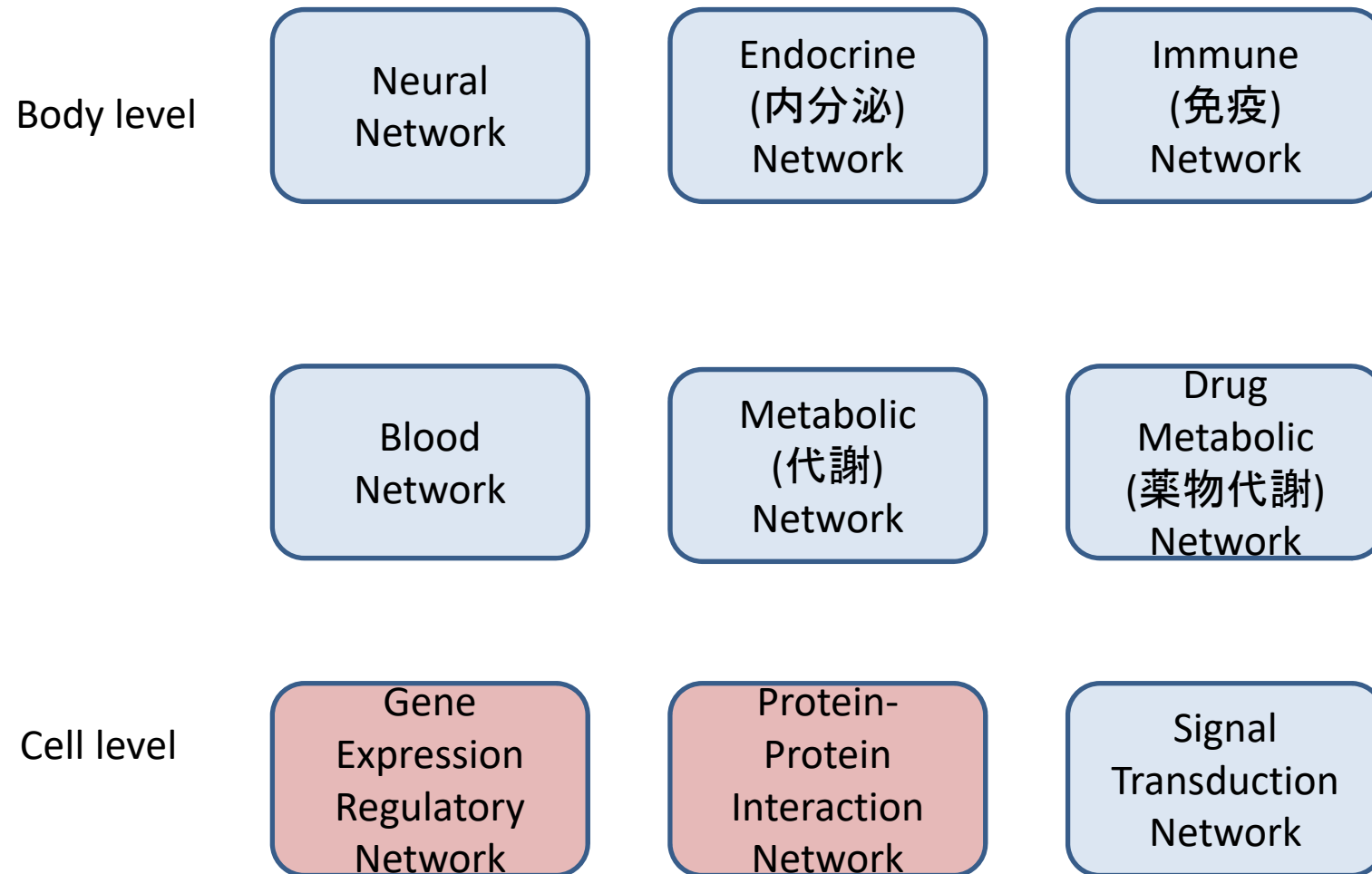
To understand life as complex network



Data driven biology: data science of genome-wide molecular information

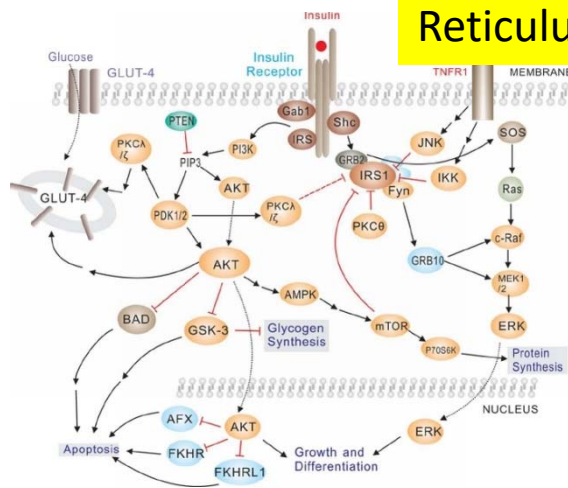
Genome: 生命を構成する一揃いの遺伝子群

# Networks in Life

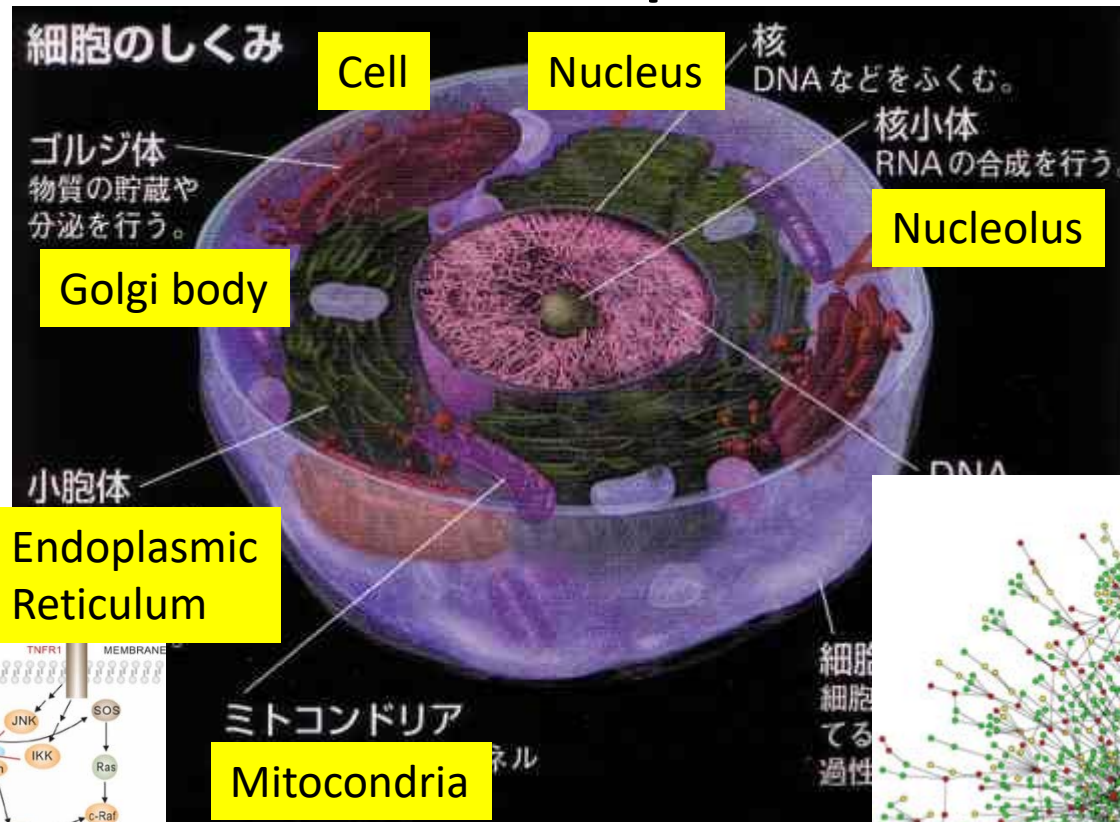


# Understanding Complex Phenomena as Complex

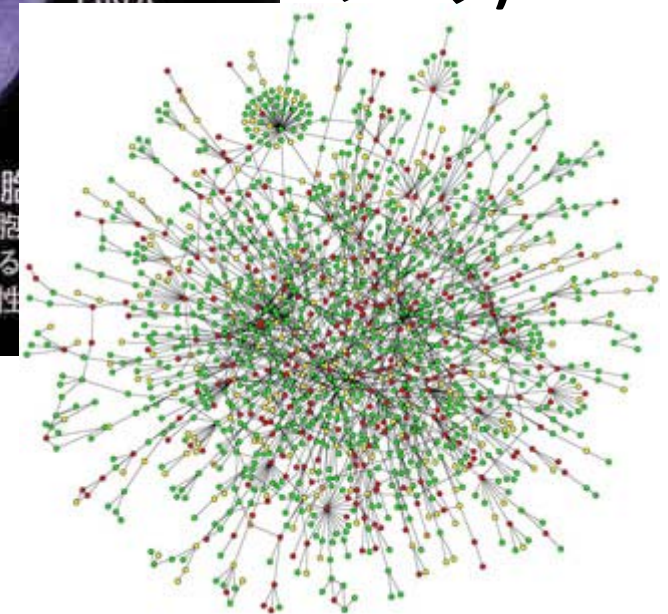
## Insulin (インスリン) Signal Transduction Pathway



2018/1/15



## Complex Network (複雑ネットワーク)



Complex Network Analysis

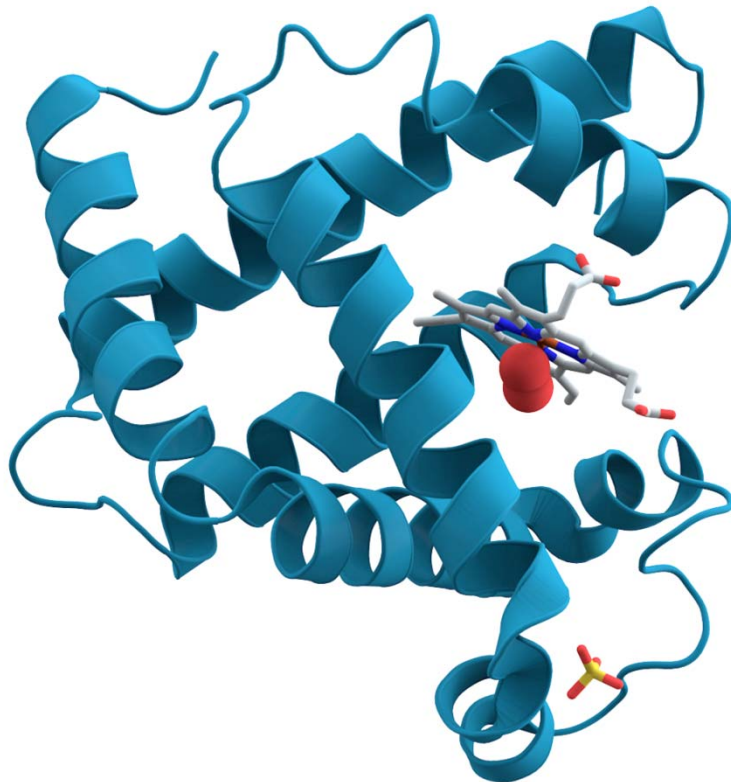
# What is a protein?

What is a protein? (6m57s)

<https://www.youtube.com/watch?v=wwTv8TqWC48>

From DNA to protein-3D (2m41s)

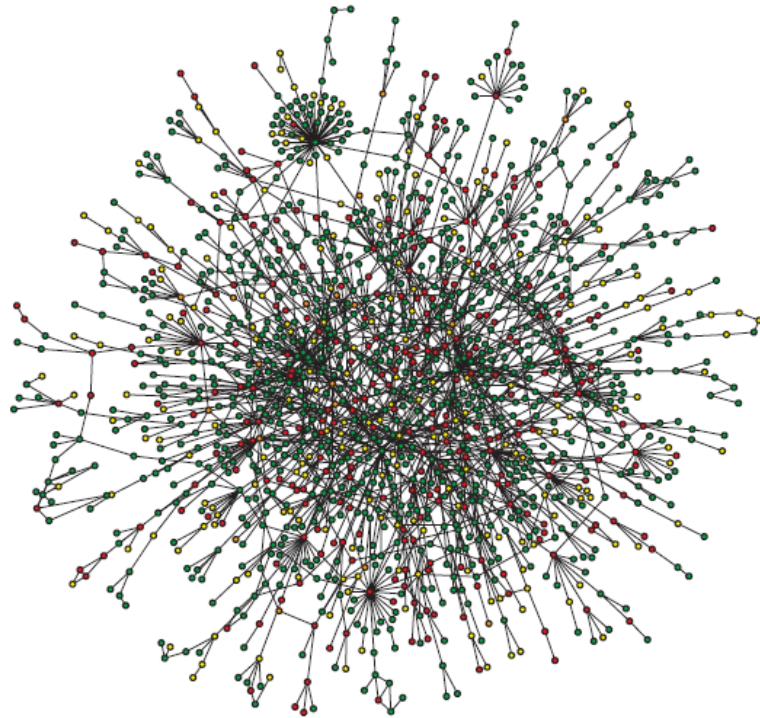
<https://www.youtube.com/watch?v=gG7uCskUOrA>



- A protein consists of a long sequence of amino acids (20 types)
- An amino acid sequence form alpha helixes and beta sheets.
- The sequence forms complex molecular structure by self-folding.
- Some proteins have metabolic reactions (enzymes, 酵素)
- **Proteins interact with each other**
- Some proteins form complex molecular machines.

o o o

# Network Biology (Barabasi et al 2004)



Protein-Protein Network  
By Y2H (Yeast to Hybrid)

Goals: escape from “reductionism”  
represented by molecular biology.

Network biology aims to understand life  
phenomena with complex network which  
keep all information hidden in life.

## Indexes

Degree (# of links) distribution  
Cluster coefficient (ratio of complete graphs)  
distribution

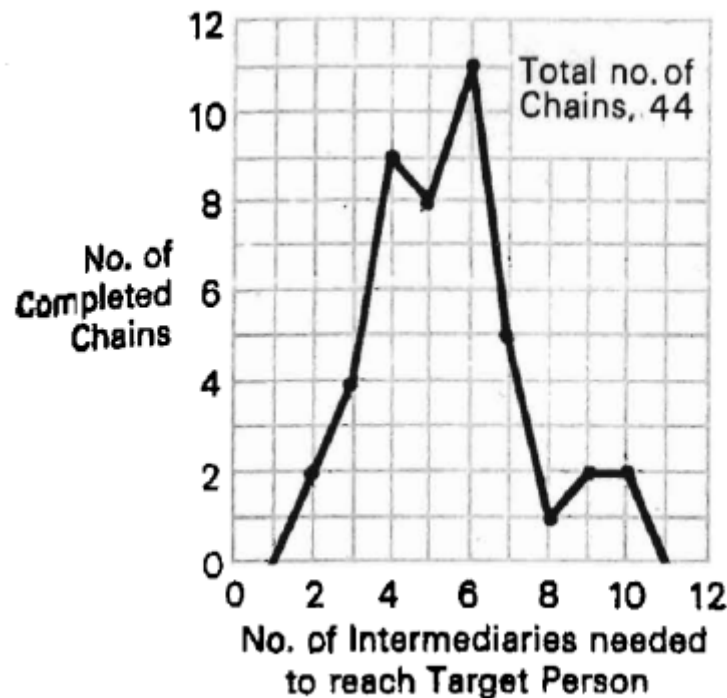
## Network architecture

Random network  
Scale free network  
Hierarchical network

# Small-world

**X - a - b - c - .... y- Z**

How many hops are required to connect given X and Z?



## The Small-World Problem

*By Stanley Milgram*

Fred Jones of Peoria, sitting in a sidewalk cafe in Tunis, and needing a light for his cigarette, asks the man at the next table for a match. They fall into conversation; the stranger is an Englishman who, it turns out, spent several months in Detroit studying the operation of an interchangeable-bottlecap-factory. "I know it's a foolish question," says Jones, "but did you ever by any chance run into a fellow named Ben Arkadian? He's an old friend of mine, manages a chain of supermarkets in Detroit . . ."

"Arkadian, Arkadian," the Englishman mutters. "Why, upon my soul, I believe I do! Small chap, very energetic, raised merry hell with the factory over a shipment of defective bottlecaps."

"No kidding!" Jones exclaims in amazement.

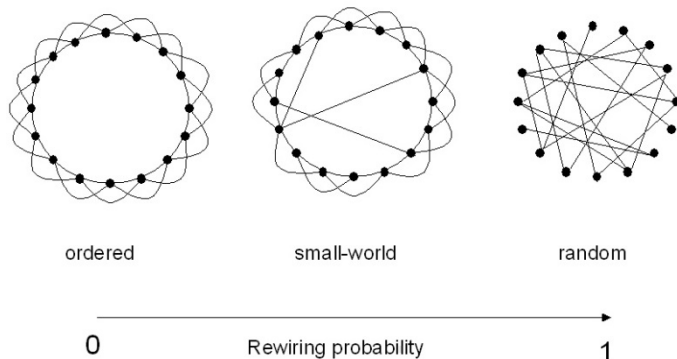
"Good lord, it's a small world, isn't it?"

Psychology Today, vol.1, no.1,  
May 1967, pp61-67



# Small-world Network

Three basic network types



**Watts and Strogatz :**  
**Collective dynamics of**  
**‘small-world’**  
**networks,**  
**Nature 393, 440-442**  
**(4 June 1998)**

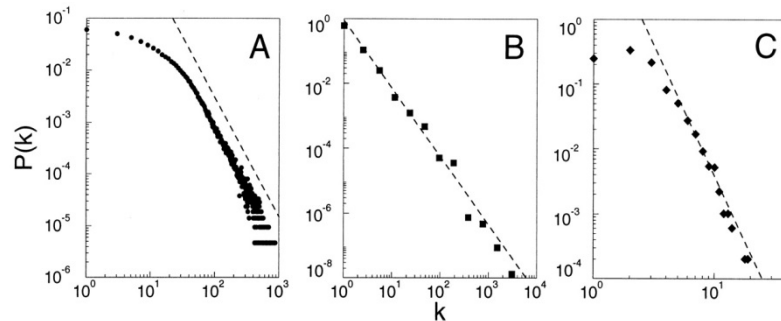
2018/1/15

Networks of coupled dynamical systems have been used to model biological oscillators, Josephson junction arrays, excitable media, neural networks, spatial games, genetic control networks and many other self-organizing systems. Ordinarily, the connection topology is assumed to be either completely regular or completely random. But many biological, technological and social networks lie somewhere between these two extremes. Here we explore simple models of networks that can be tuned through this middle ground: regular networks 'rewired' to introduce increasing amounts of disorder. We find that these systems can be highly clustered, like regular lattices, yet have small characteristic path lengths, like random graphs. We call them 'small-world' networks, by analogy with the small-world phenomenon (popularly known as six degrees of separation). The neural network of the worm *Caenorhabditis elegans*, the power grid of the western United States, and the collaboration graph of film actors are shown to be small-world networks. Models of dynamical systems with small-world coupling display enhanced signal-propagation speed, computational power, and synchronizability. In particular, infectious diseases spread more easily in small-world networks than in regular lattices.

Complex Network Analysis



# Power Law



**Figure 1** The distribution function of connectivities for various large networks. (A) Actor collaboration graph with  $N = 212,250$  vertices and average connectivity  $\langle k \rangle = 28.78$ . (B) WWW,  $N = 325,729$ ,  $\langle k \rangle = 5.46$  (6). (C) Power grid data,  $N = 4941$ ,  $\langle k \rangle = 2.67$ . The dashed lines have slopes (A)  $\gamma_{\text{actor}} = 2.3$ , (B)  $\gamma_{\text{www}} = 2.1$  and (C)  $\gamma_{\text{power}} = 4$ .

Systems as diverse as genetic networks or the World Wide Web are best described as networks with complex topology. A common property of many large networks is that the vertex connectivities follow a scale-free power-law distribution. This feature was found to be a consequence of two generic mechanisms: (i) networks expand continuously by the addition of new vertices, and (ii) new vertices attach preferentially to sites that are already well connected. A model based on these two ingredients reproduces the observed stationary scale-free distributions, which indicates that the development of large networks is governed by robust self-organizing phenomena that go beyond the particulars of the individual systems.

**Albert-László Barabási\* and Réka Albert: Emergence of Scaling in Random Networks**

2018/1/15

Science. 1999 Oct 15;286(5439):509-12.

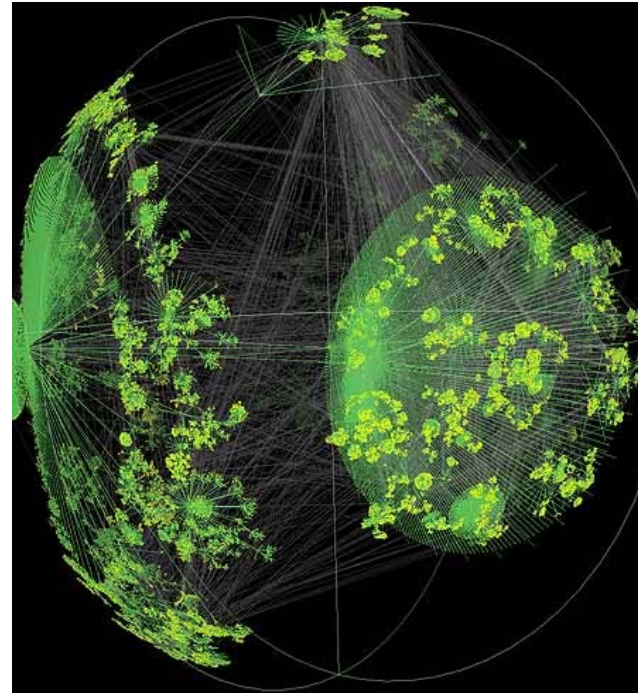
Complex Network Analysis



# The Internet Topology



[http://www.topology-zoo.org/eu\\_nren.html](http://www.topology-zoo.org/eu_nren.html)



[http://mappa.mundi.net/maps/maps\\_020/march2001\\_sm\\_lo.jpg](http://mappa.mundi.net/maps/maps_020/march2001_sm_lo.jpg)

# Degree

Degree: # of links connected to some node

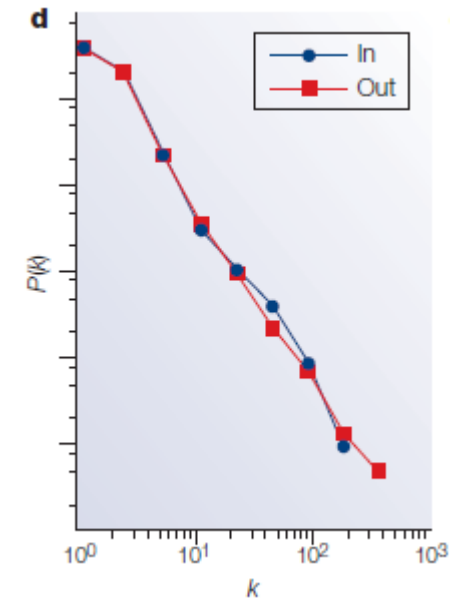
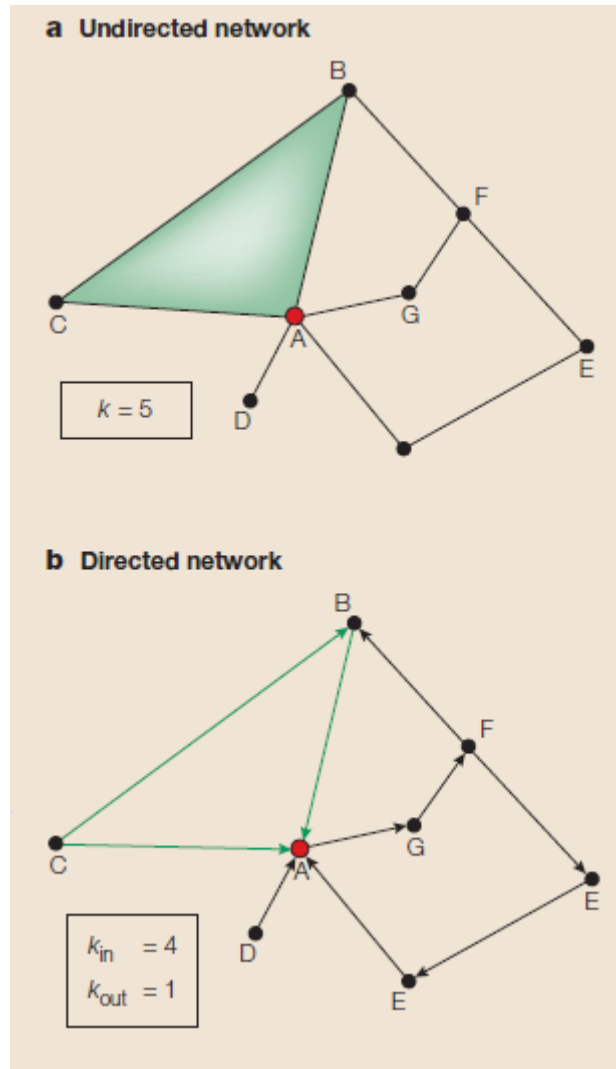
$$\text{Degree}(A) = 5$$

In case of directed links,  
incoming degree: # of links incoming  
outgoing degree: # of links outgoing

$$K_{\text{in}}(A) = 4$$
$$k_{\text{out}}(A) = 1$$

Example:

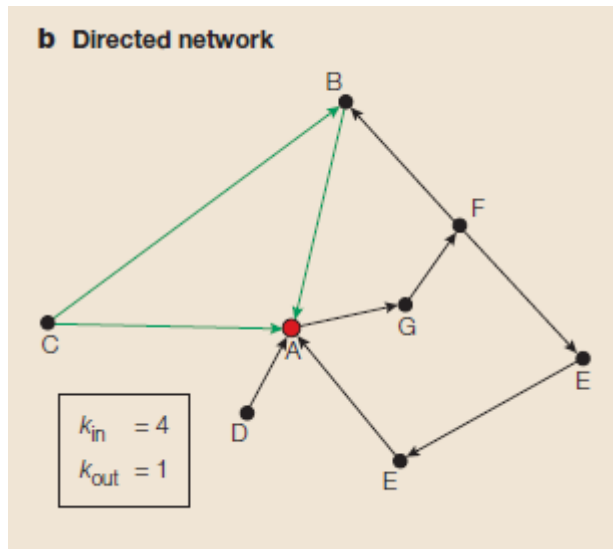
$P(k)$  : Distribution  
of degrees in  
metabolic pathway  
network



# Cluster Coefficient

$$C_i = 2n_i / k(k-1)$$

The ratio of  
# of triangles divided by  
# of possible triangles in neighbor nodes within k hops;  
**Density of graphs**

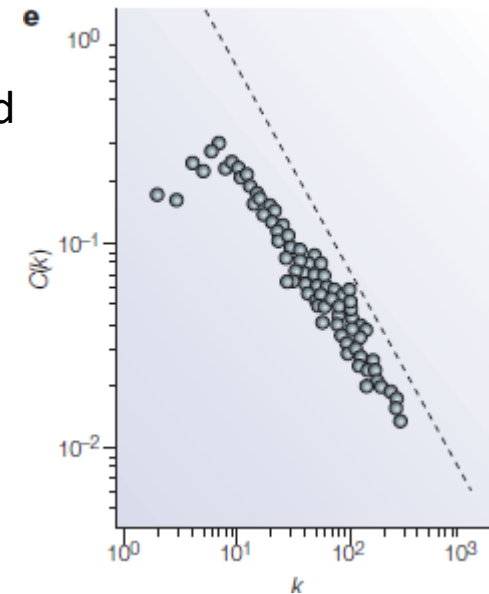


# of neighbor nodes around  
the node A within 1 hop:  
5 nodes (B, C, D, E and G)

# of triangle = 1

# of possible triangles  
=  $5(5-1)/2$   
= 10

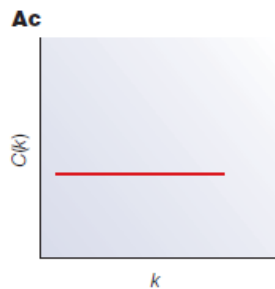
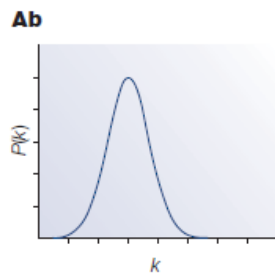
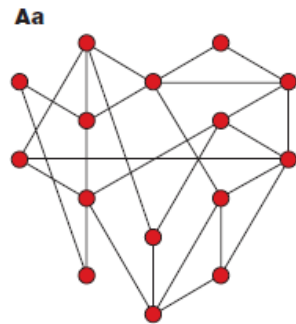
So, cluster coefficient  
 $C = 1/10$



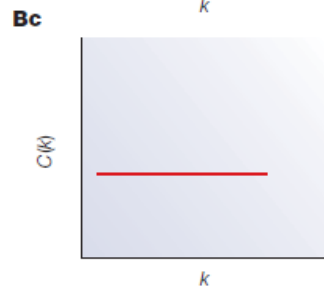
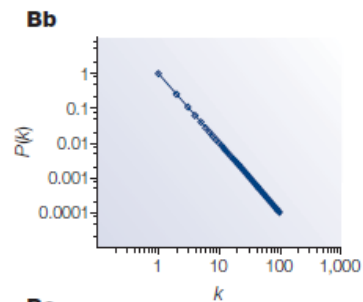
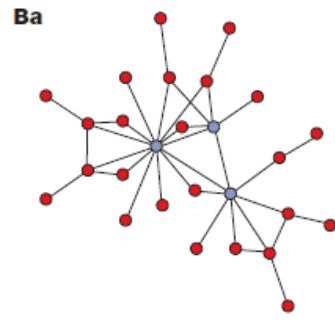
Distribution of cluster  
coefficient in  
metabolic pathway  
network

# Network Classification

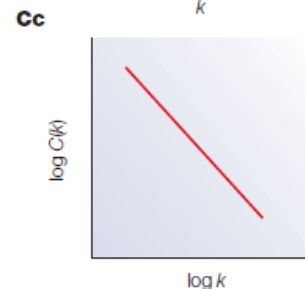
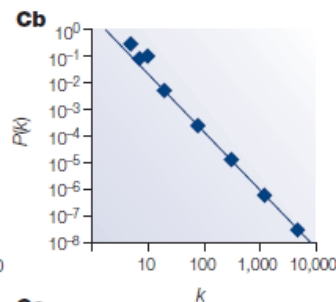
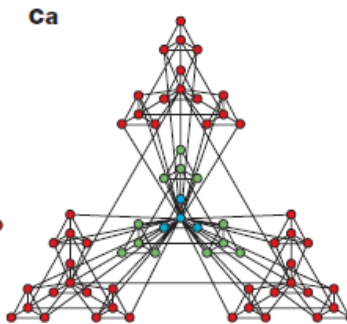
**A Random network**



**B Scale-free network**



**C Hierarchical network**



## Random Network

$P(K)$ : normal distribution

Peak at the average in # of edges (Degree)

$C(K)$ : uniform distribution

Uniform in cluster coefficient

## Scale Free Network

$P(K)$ : power law distribution

Decreasing # of edges constantly

$C(K)$ : uniform distribution

Uniform in network density

## Hierarchical Network

(Module Hierarchy)

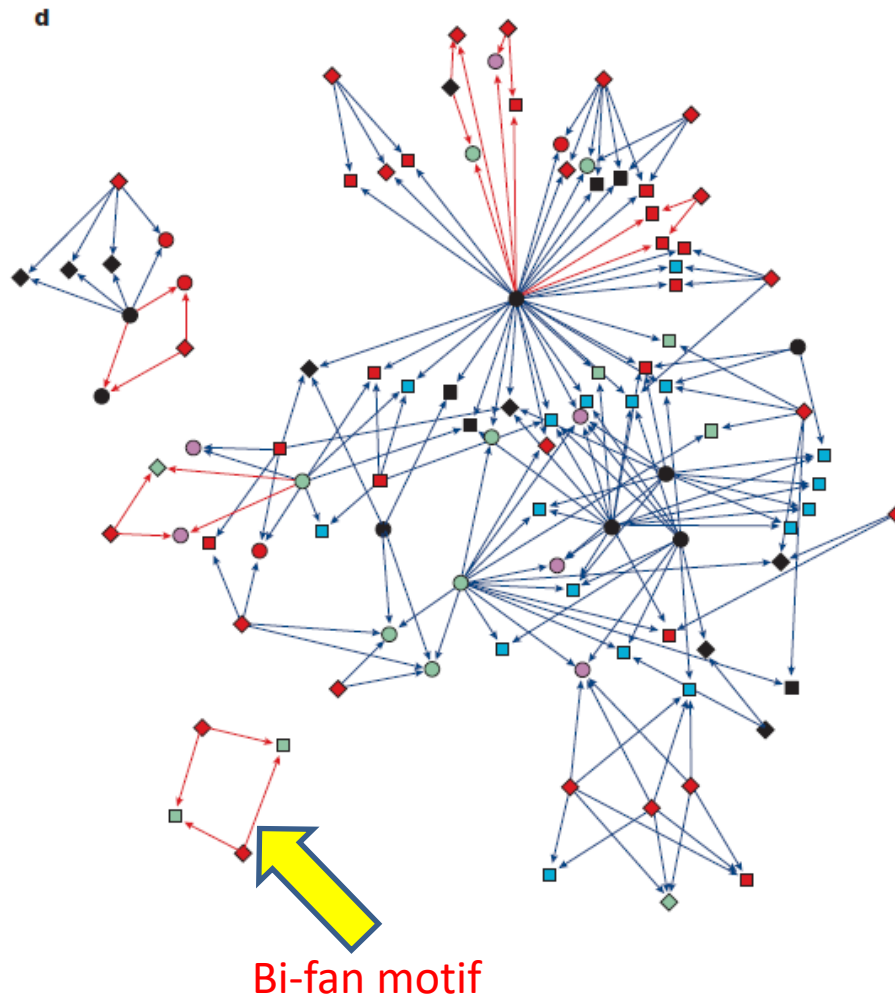
$P(K)$ : power law distribution

Decreasing # of edges constantly

$C(K)$ : power law distribution

Decreasing graph density constantly

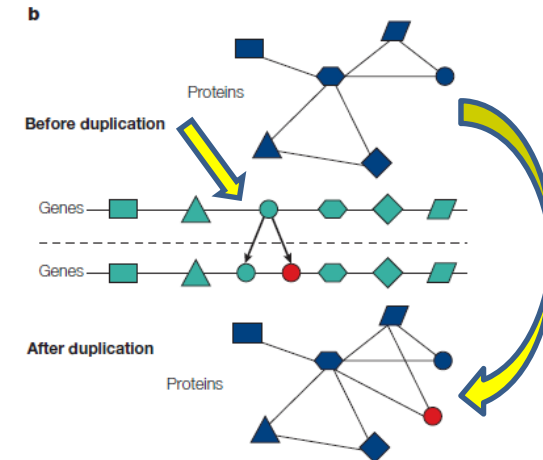
# Network Motif



Common subgraphs found in complex network

Network motifs may indicate important discovery or interpretation.

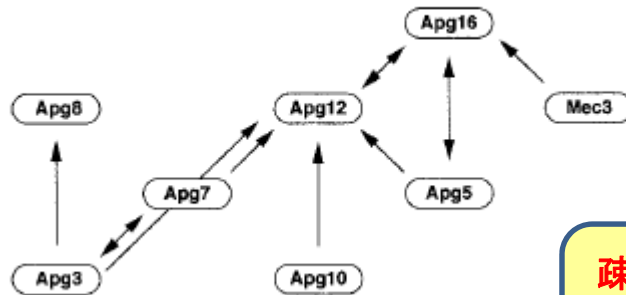
Ex) Bi-fan motif in gene expression regulatory network might be result from gene duplication.





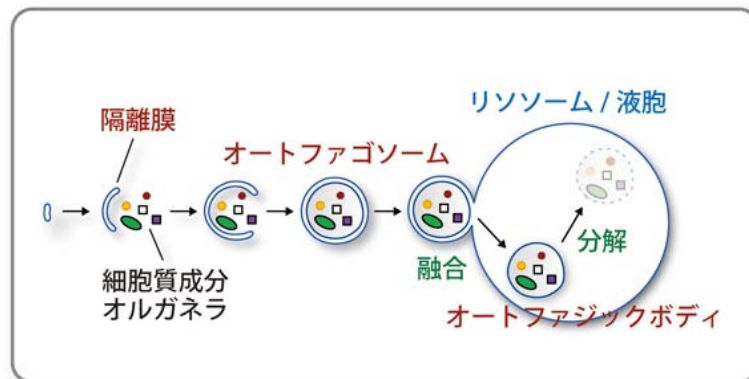
# タンパク-タンパク相互作用 (PPI) とは？

A



Autophagy

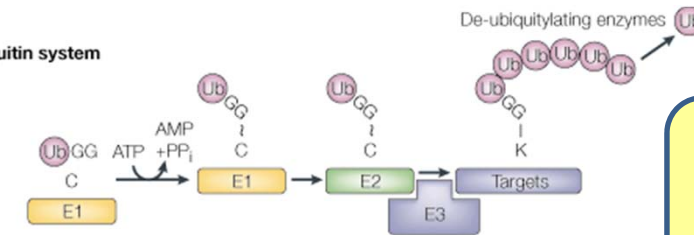
(自食作用) 不要なオルガネラ,  
たんぱく質のリソソームによる  
分解作用



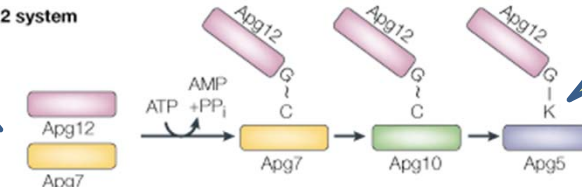
2018/1/15

疎水性  
相互作用  
による結合

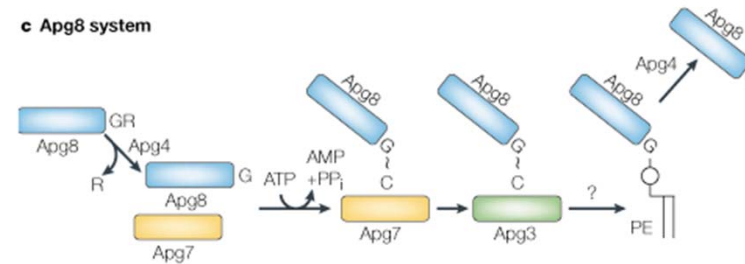
a Ubiquitin system



b Apg12 system



c Apg8 system



Apg12の  
C末端の  
GとApg5  
のKとの  
共有結合

Nature Reviews | Molecular Cell Biology

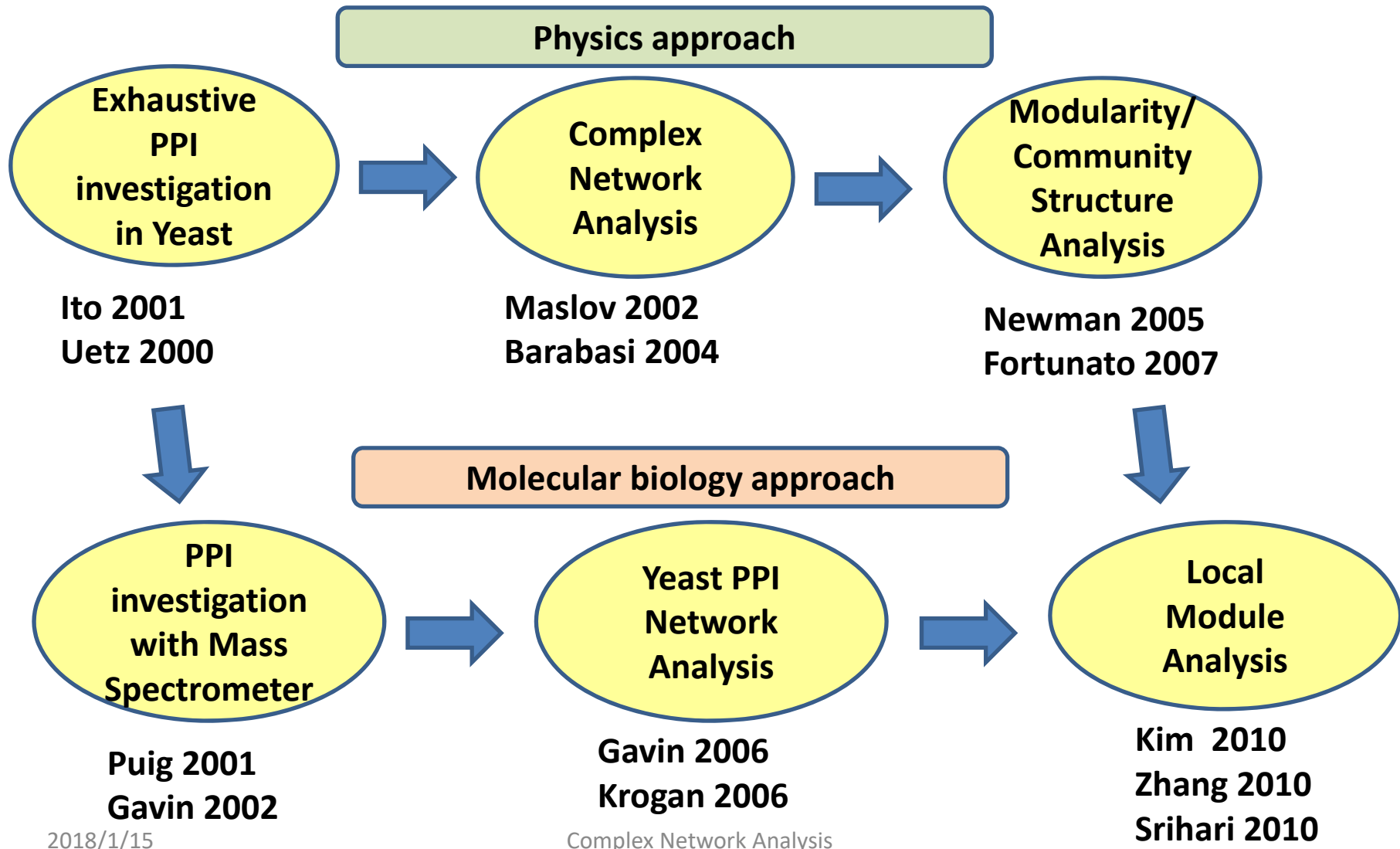
Ito T. et al (2001) PNAS

Ohsumi Y. et al (2001) Nature Reviews

Complex Network Analysis



# Trend of Protein-Protein Interaction (PPI) Research



# The difference of physics and molecular biology approaches

## Physicist's view

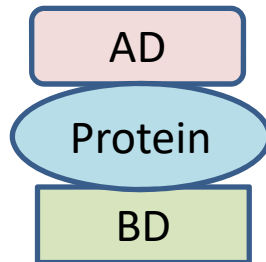
- Understanding complex system as complex
- All natural events including life must be based on a law
- Preferring precision experiments without noise

## Molecular biologist's view

- All facts must be proved by experiments
- Diversity plays an important role in life
- Preferring exhausting experiments even if noise are included

# Y2H(Yeast Two Hybrid, Fields et al 1991)

GAL4 protein

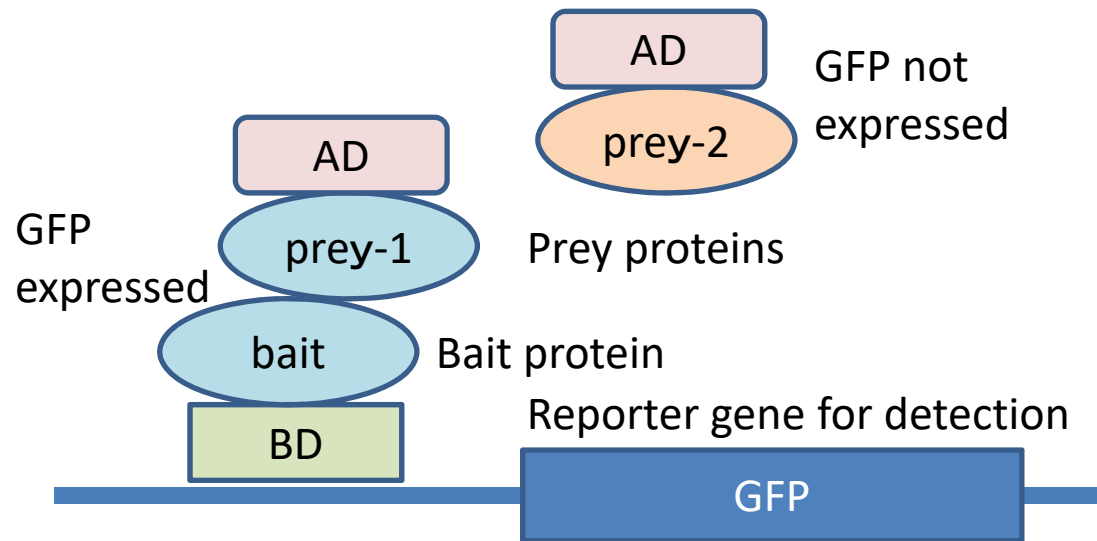


GAL4 protein can hold a protein between its two components AD and BD.

Active Domain (活性化部位)

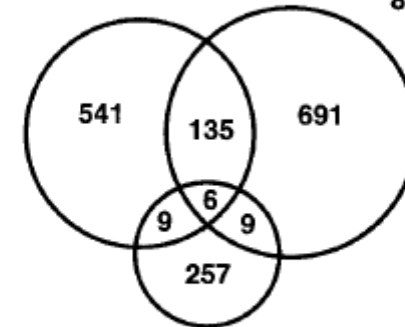
Binding Domain (DNA結合部位)

So, GAL4 can be used for PPI detection by means of providing  
Prey protein (捕捉タンパク質) attached with AD  
Bait protein (餌タンパク質) attached with BD



Uetz et al.  
IST approach  
691

this study  
core data  
841



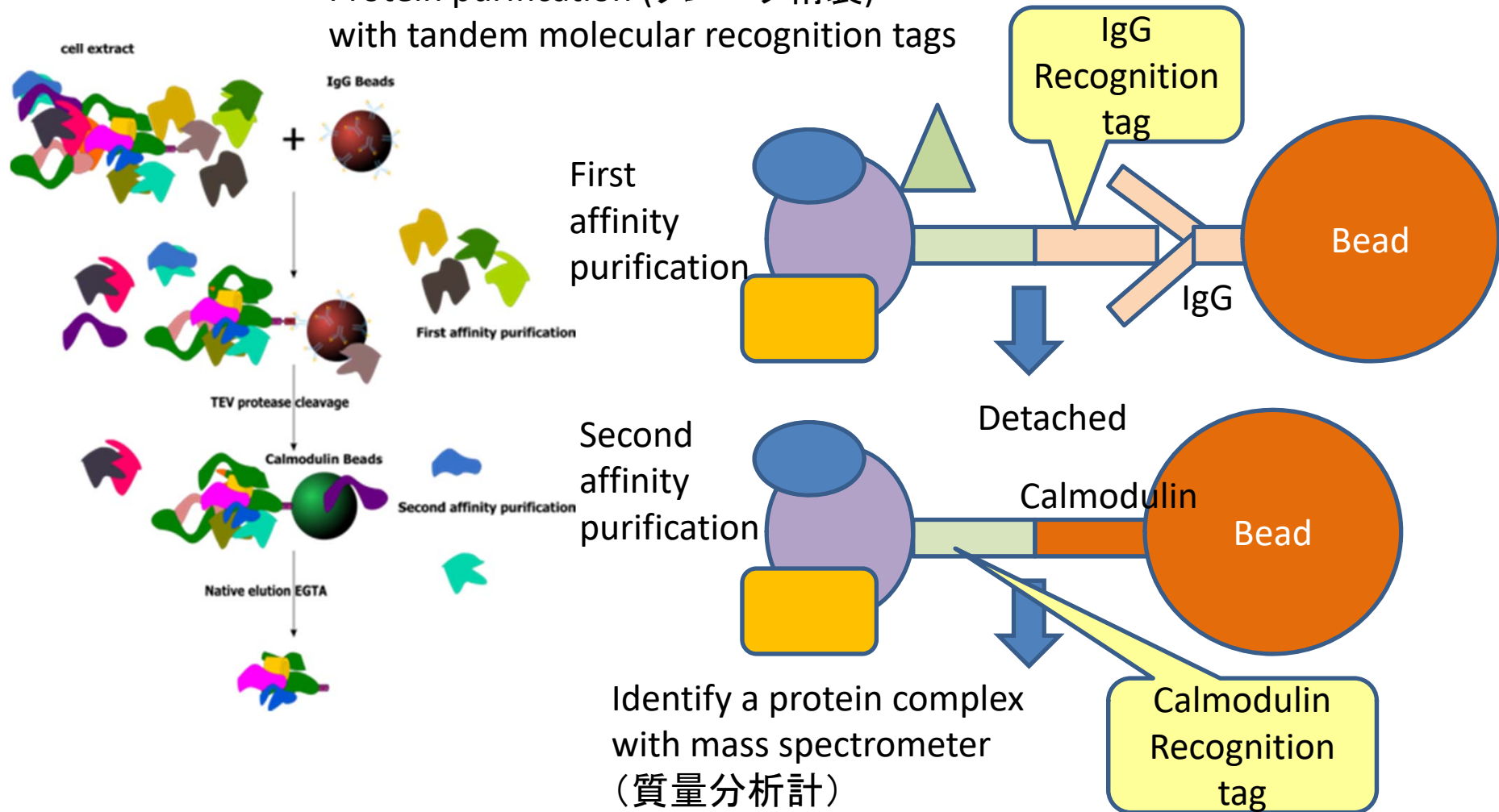
Uetz et al.  
protein array  
281

Only 141 PPIs could be reproduced among the two exhaustive experiments with Y2H. Few common PPIs with protein array result.

Expression of the reporter gene  
Indicates the occurrence of protein-protein interaction between bait and prey proteins.

# TAP (Tandem Affinity Purification)

Protein purification (タンパク精製)  
with tandem molecular recognition tags



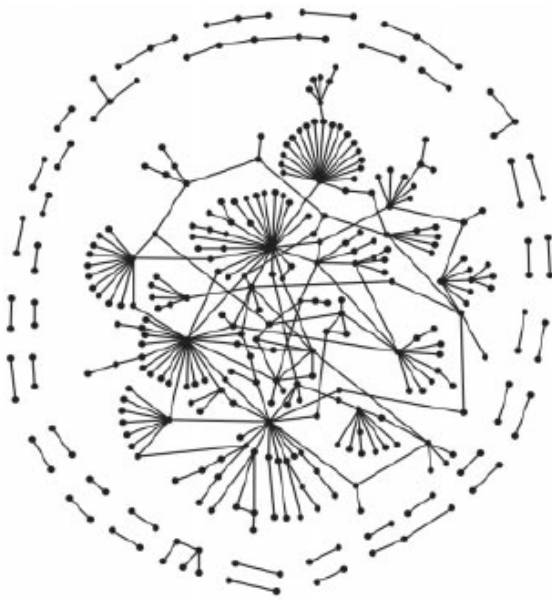
[http://en.wikipedia.org/wiki/Tandem\\_Affinity\\_Purification](http://en.wikipedia.org/wiki/Tandem_Affinity_Purification)

2018/1/15

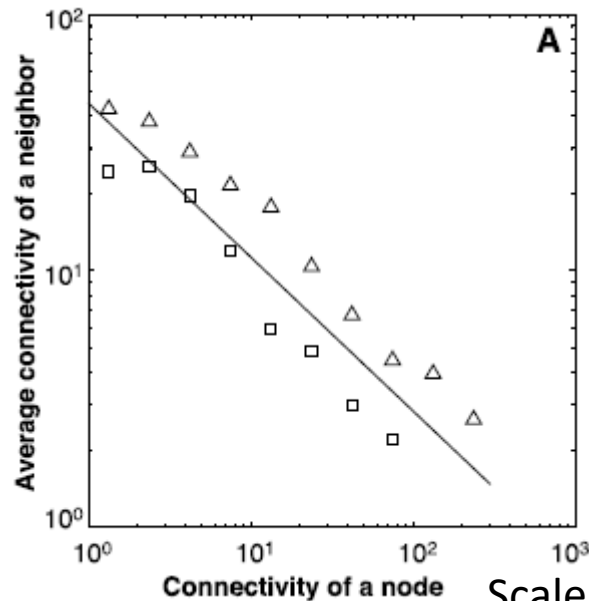
Complex Network Analysis

# Topology of Protein Networks (Maslov 2002)

Investigation of Y2H protein-protein interaction network from the viewpoint of complex network analysis; suggesting important proteins in hub nodes.

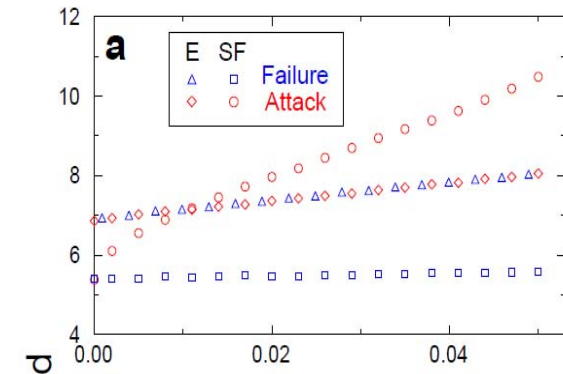


PPI network topology of 329 proteins in Y2H experiment. The network consists of several local hub nodes with many links and many edge nodes with a few links or none.



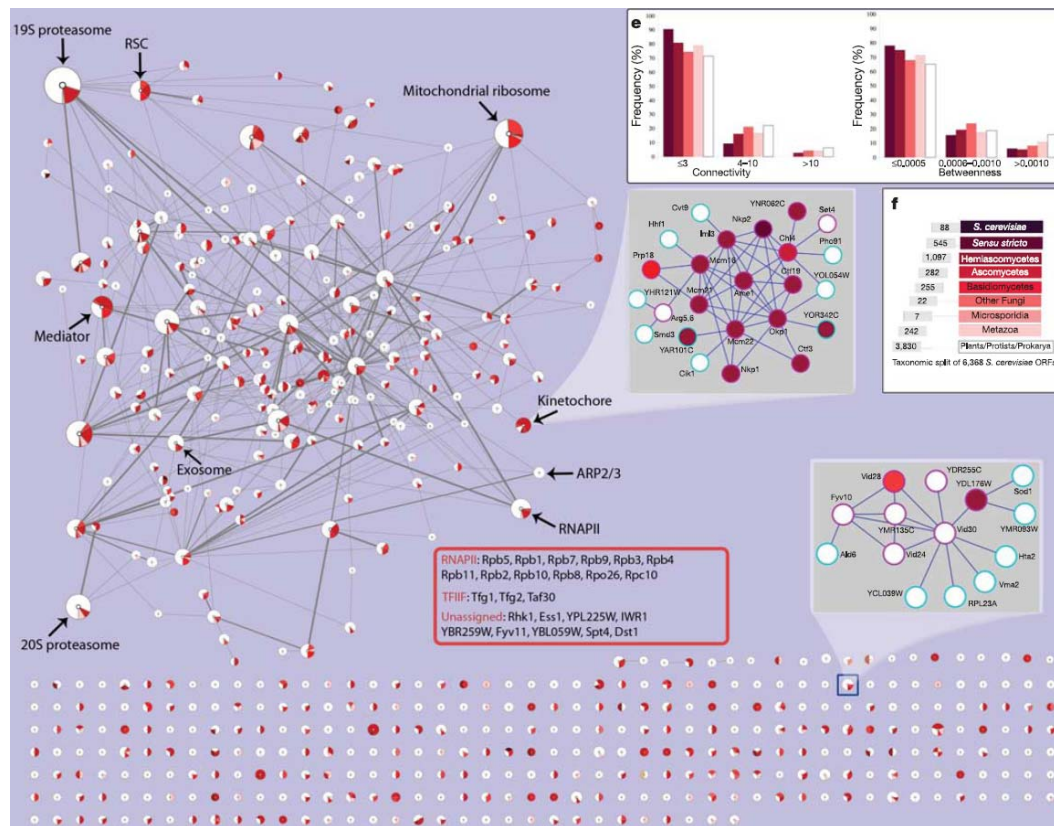
Scale free in the ratio of  
Connectivity (周辺ノード数)  
and  
Average connectivity of a neighbor  
(近傍の平均周辺ノード数)  
△ : PPI network  
□ : Gene expression regulatory network

Albert 2000, Exponential vs Scale Free Networks

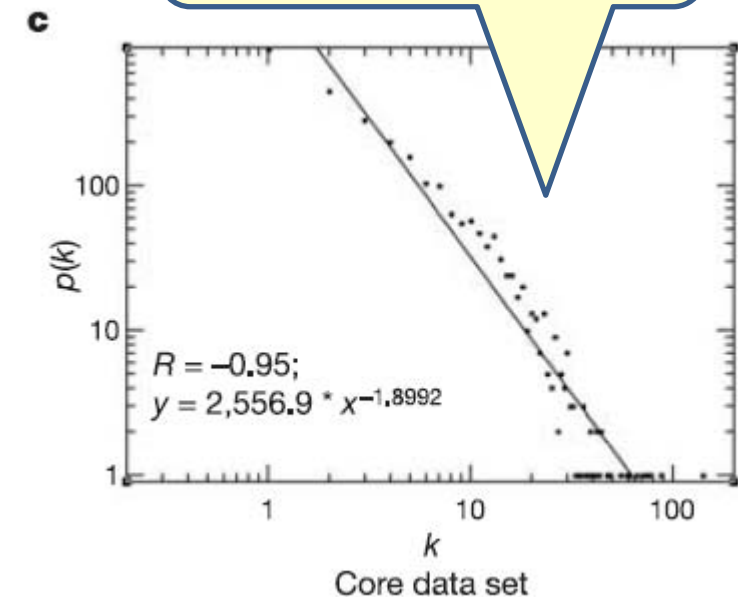


# Analysis of Yeast Protein Complex (Krogan2006)

High throughput high precision protein complex analysis with TAP+LC/MS/MS;  
 Identified (547 complex, 2708 protein, 7123 PPI; 5.26 interactions/protein)  
 However, less precision with regards to locality and Gene Ontology functional analysis  
 If compared with MIPS complex database (knowledge base of PPI extracted from literatures)



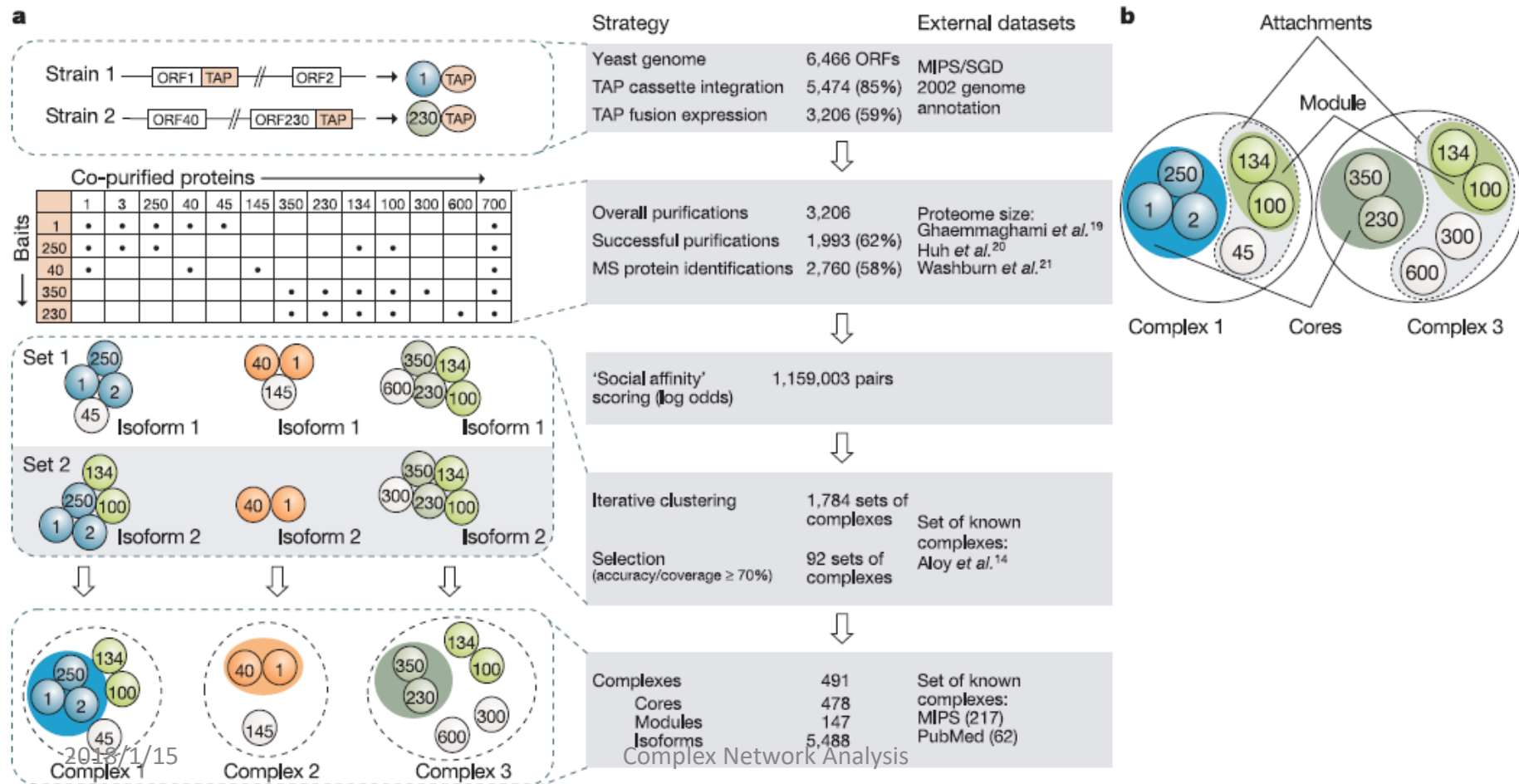
Does protein-protein interaction follow power-law?





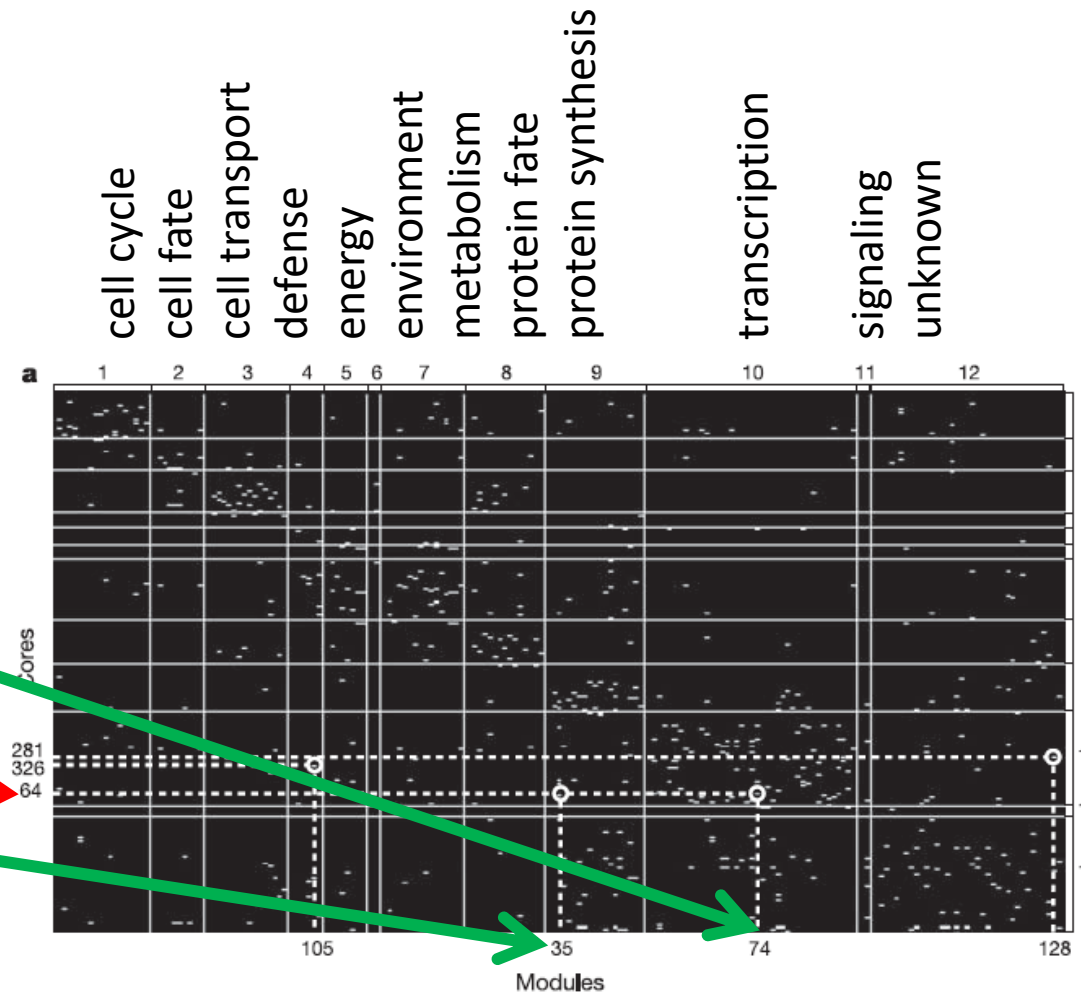
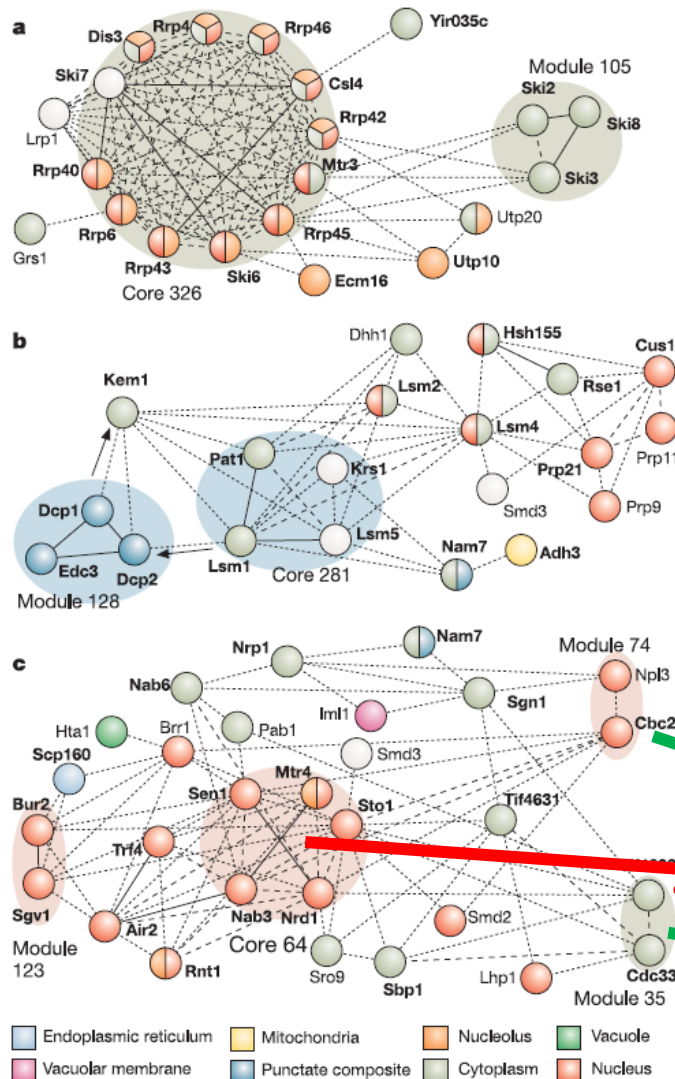
# Analysis of Yeast Protein Complex (Gavin2006)

Identify 491 protein complex for 6466 open reading frame (i.e. genes) with TAP+MS method. Found 257 novel protein complex. There are special protein groups named “Core” expressed in many protein complexes; 478 cores are associated with protein functions and 147 cores are on conditional.





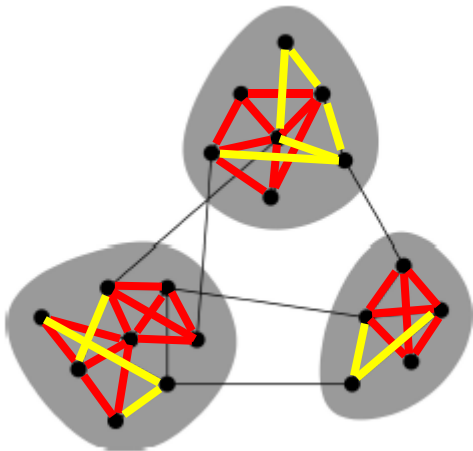
# Relations between cores and modules



Complex Network Analysis

2018/1/15

# Modularity and Community Structure (Newman –Girvan 2004)



Biologically meaningful decomposition cannot be obtained from mathematical algorithms such as counting links and complete cliques.



Modularity  $Q$ : indicator of Community Structure available from eigenvalues of module matrix

$$Q = \sum_i (e_{ii} - a_i^2) = \text{Tr } e - \|e^2\|,$$

# of links in a module

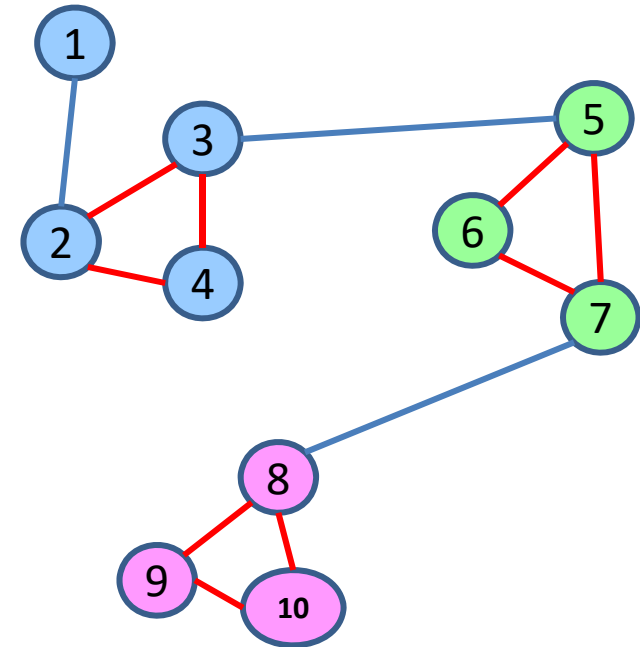
# of link nodes expected in random network

$Q=0$  Random network  
 $Q \approx 1$  Independent decomposition  
 $0.3 < Q < 0.7$

Newman M.E.J , Girvan M. : Finding and evaluating community structure in network  
Physical Review E, 69, 026113 (2004)

Adjacency  
Matrix A<sub>ij</sub>

		1								
1			1	1						
	1			1	1					
	1	1								
			1							
				1		1				
				1	1		1			
						1		1	1	
							1		1	
							1	1		



8/24	1/24	0/24
1/24	6/24	1/24
0/24	1/24	6/24

9/24 8/24 7/24

9/24

8/24

7/24

$$Q = \sum_i (e_{ii} - a_i^2) = \text{Tr } \mathbf{e} - \|\mathbf{e}^2\|,$$

$$= (8/24) + (6/24) + (6/24) - (9/24)^2 - (8/24)^2 - (7/24)^2 = 0.43$$

# Resolution Limit (Fortunato 2007)

Newman's modularity is fast but the size of minimum modules becomes larger and larger if the size of network becomes large. Therefore, the method cannot identify the difference of module structure and a cluster of small module structures in protein-protein interaction network. (**resolution limit**).

Table 1. Results of the modularity analysis on real networks

Network	No. of modules ( $Q_{\max}$ )	Total no. of modules ( $Q$ )
Yeast	9 (0.740)	57 (0.677)
<i>E. coli</i>	27 (0.752)	76 (0.661)
Elect. circuit	11 (0.670)	70 (0.640)
Social	10 (0.608)	21 (0.532)
<i>C. elegans</i>	4 (0.408)	20 (0.319)

# Modularity Density (Zhang 2010)

Modularity (MQ)

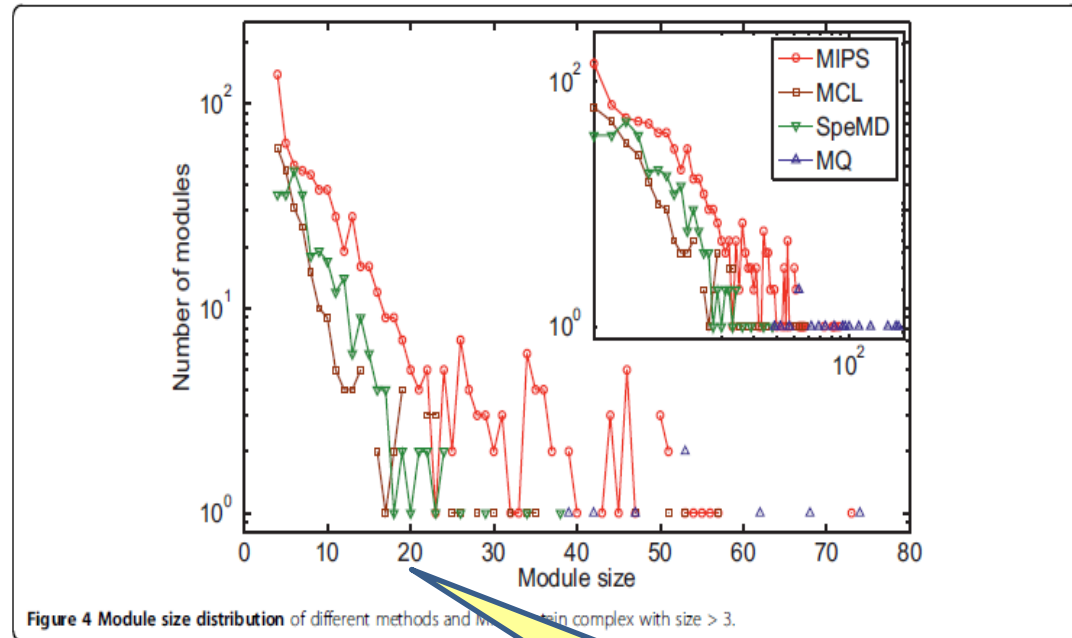
$$Q = \sum_{s=1}^m \left[ \frac{l_i}{L} - \left( \frac{d_i}{2L} \right)^2 \right],$$

Modularity Density (SpeMD)

$$D = \sum_{i=1}^m ad(G_i) = \sum_{i=1}^m \frac{2l_i - \bar{l}_i}{n_i}.$$

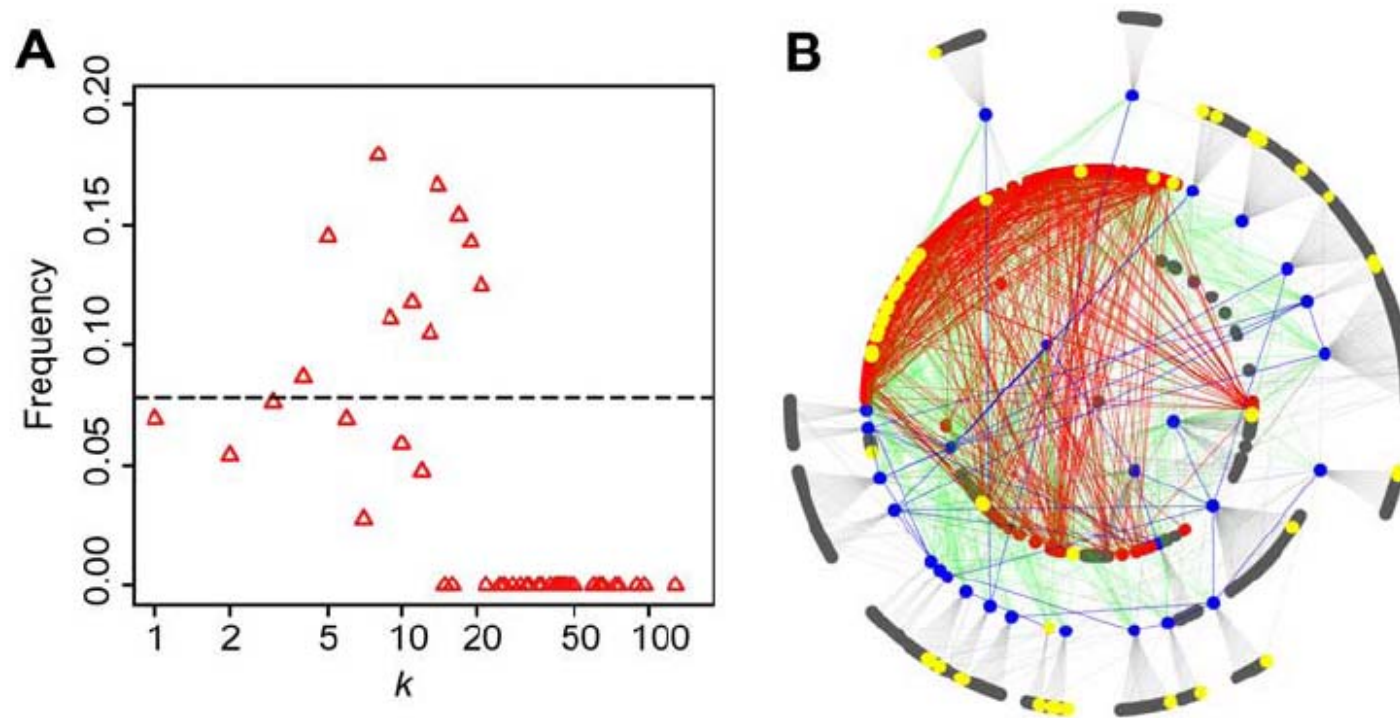
Sum of the difference of double of link # in a module ( $l_i$ ) and the number of links outgoing from the module ( $\bar{l}_i$ ) divided by node # in a module ( $n_i$ )

Zhang S : Determining modular organization of protein interaction networks by maximizing modularity density, BMC System Biology, 4 (Suppl 2): S10 (2010)



**Efficient to find  
modules with small  
nodes**

# Drug target molecule discovery with protein-protein interactions



Drug target molecules are often found in a zone with medium PPI frequency.

Hase Takeshi. et al. : Structure of Protein Interaction Networks and Their Implications on Drug Design, PLOS Computational Biology, vol.5, no.10 (2009)

# Summary

- Emergence of network biology to understand life phenomena from the view point of complex network analysis
- Scale free network in protein-protein interaction
- Drug target proteins often have medium PPI frequency



# Report Issues

- Due date Jan 22<sup>nd</sup> (Mon) Noon
- Submission through OCW
- Subject Protein-protein interaction network
- Style a PDF file A4 1 or 2 pages

## Issue

- Impression or investigation report on the issue of protein-protein interaction network held on Jan. 15<sup>th</sup>.