计算系统生物学

王 勇

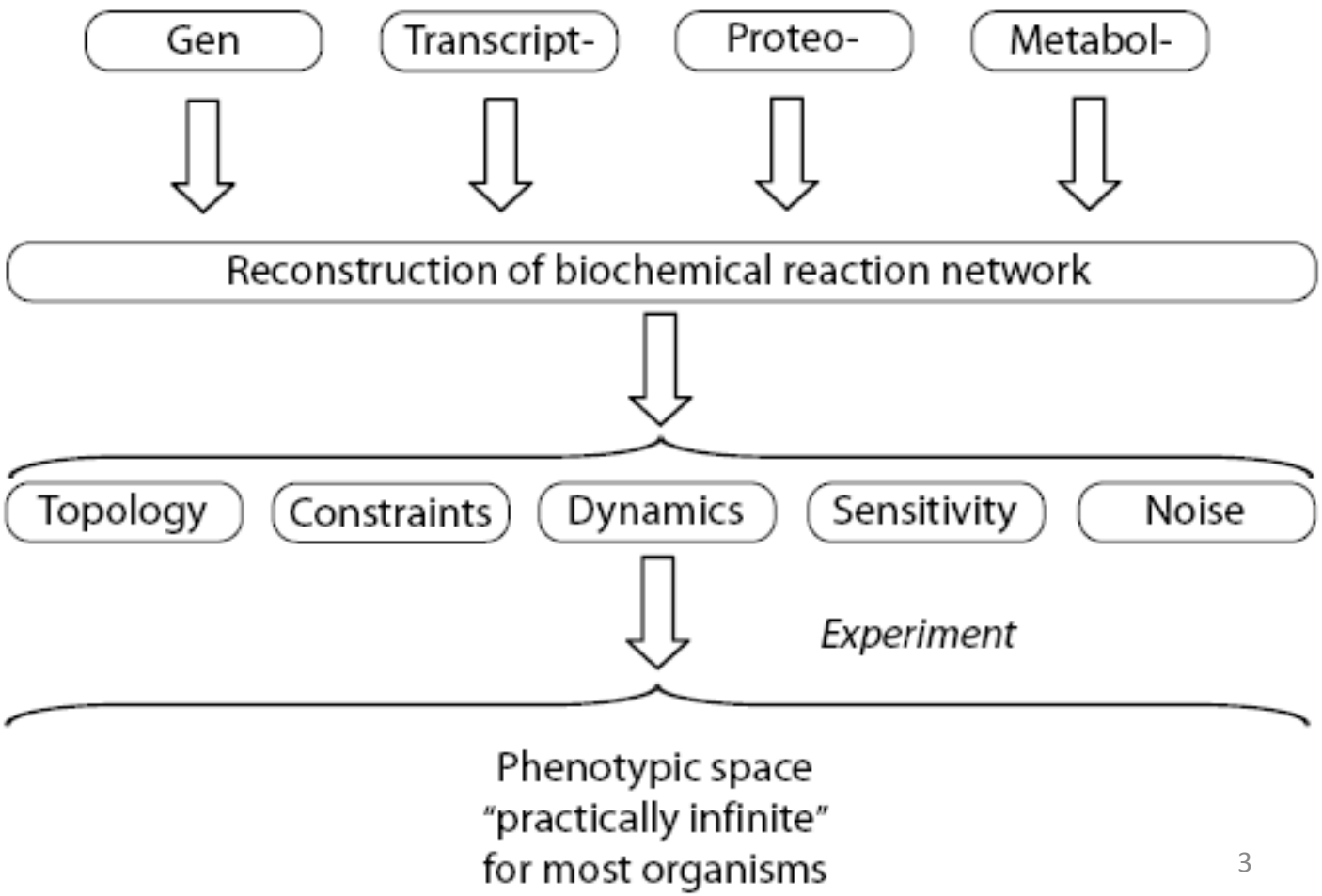
中国科学院数学与系统科学研究院

Bio-molecular network analysis

Yong Wang

Beyond the network reconstruction

1. Components
   Plurality of -omics
2. Reconstruction
   "Systemic annotation"
   one set of reactions
   arising from the
   genome
3. *In silico* modeling
   plurality of methods
4. Hypothesis generation
   and testing
   -CHiP-chip
   -Fluxomics



生物分子网络特性及分析方法

Characteristic	Analysis method
They are complex	Bioinformatics
They are autonomous	Control theory
They are robust	System science
They function to execute a physicochemical process	Transport and kinetic theory
They have "creative functions"	Bifurcation analysis
They are conserved, but can adjust	Evolutionary dynamics

网络分析

- 拓扑分析 (Topology)

Hub and bottleneck

Hierachy structure

Network motif

- 网络动态分析 (Dynamics)

Hubs in different conditions

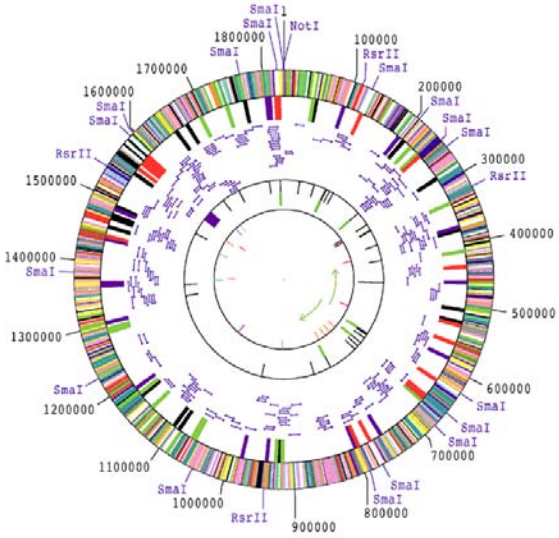
Subnetworks in different conditions

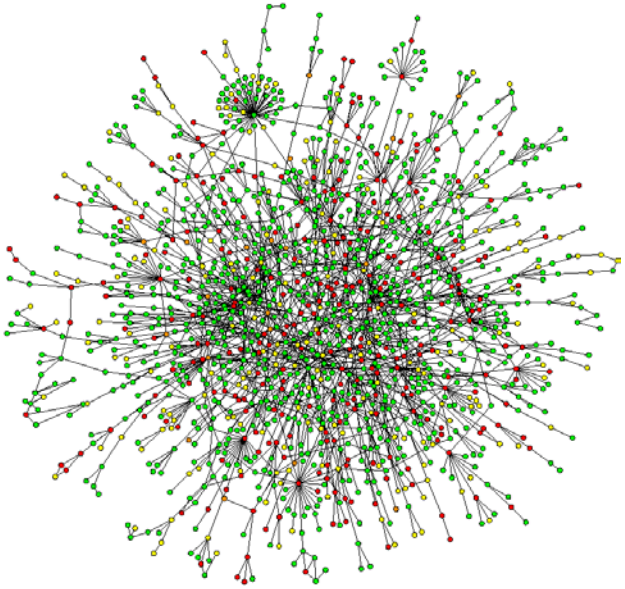
- 子网络分析 (Subnetworks)

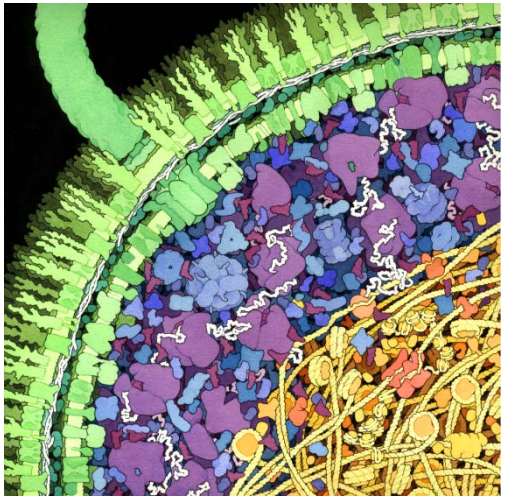
Aging and disease subnetwork

Evolution in TF subnetwork

Networks occupy a midway point in terms of level of understanding







1D: Complete Genetic Partslist

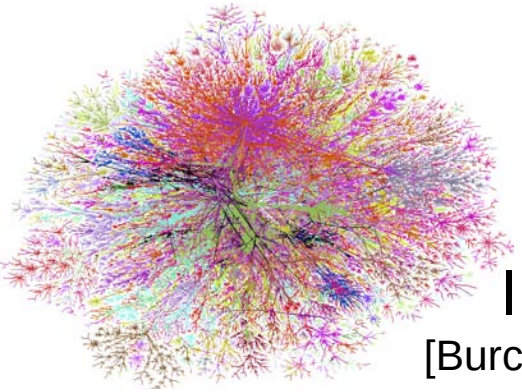
~2D: Bio-molecular Network Wiring Diagram

3D: Detailed structural understanding of cellular machinery

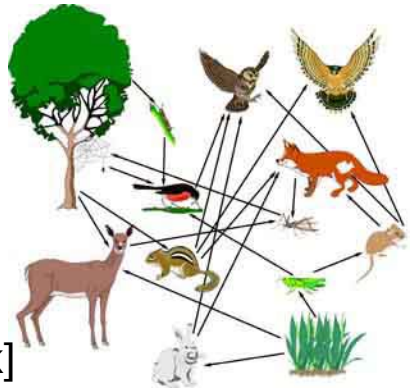
[Fleischmann et al., Science, 269 :496]

[Jeong et al. Nature, 41:411]

Networks as a universal language

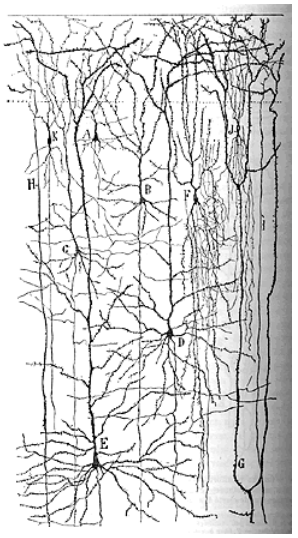


Internet
[Burch & Cheswick]

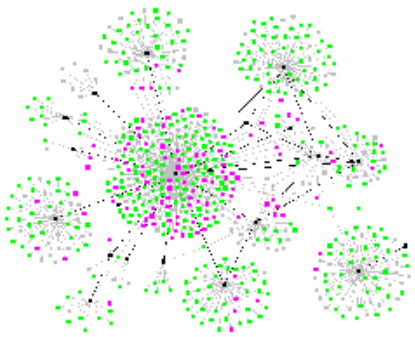


Food Web

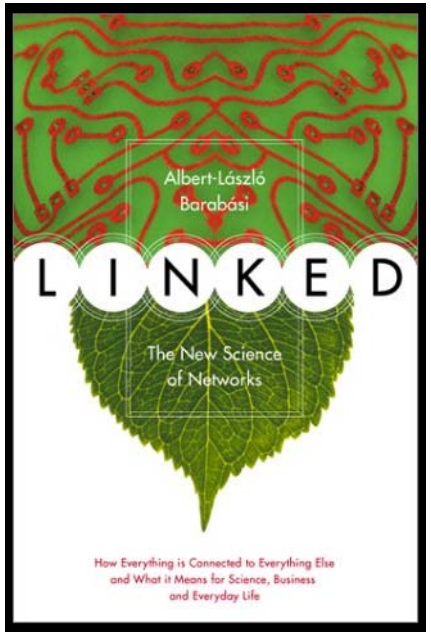
Electronic Circuit



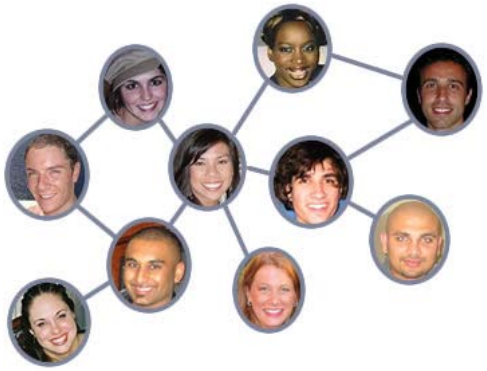
Neural Network
[Cajal]



Disease Spread
[Krebs]

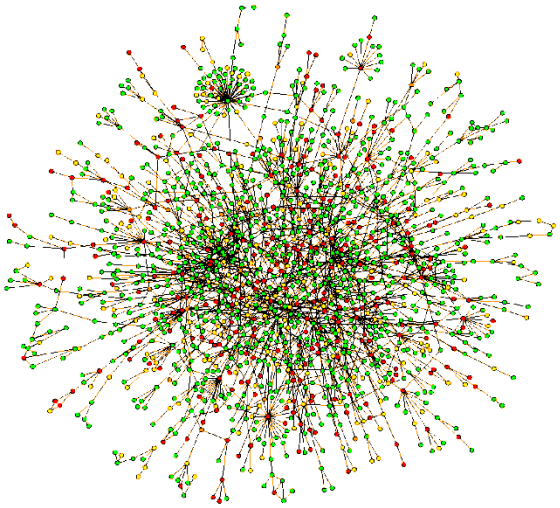


Protein Interactions
[Barabasi]

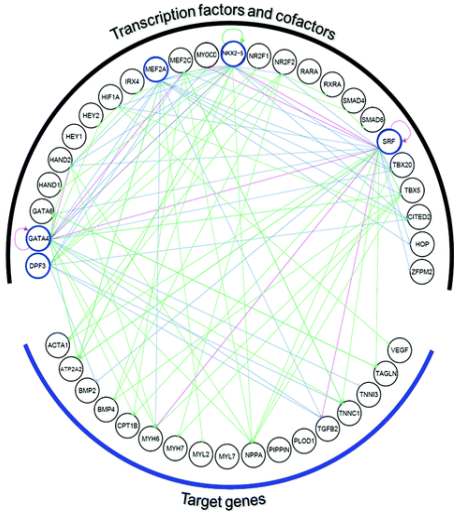


Social Network

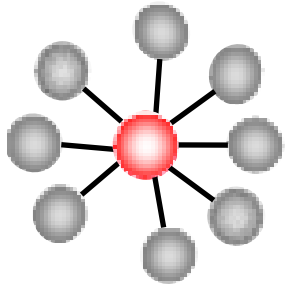
Different Types of Molecular Networks



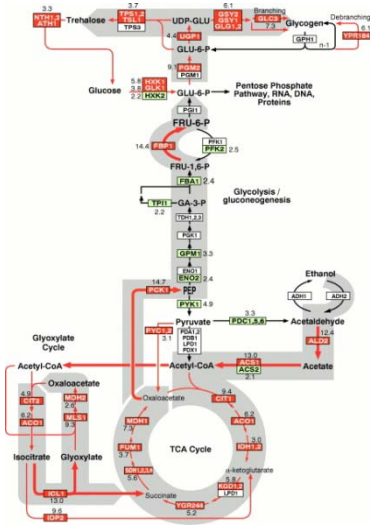
Protein-protein Interaction networks



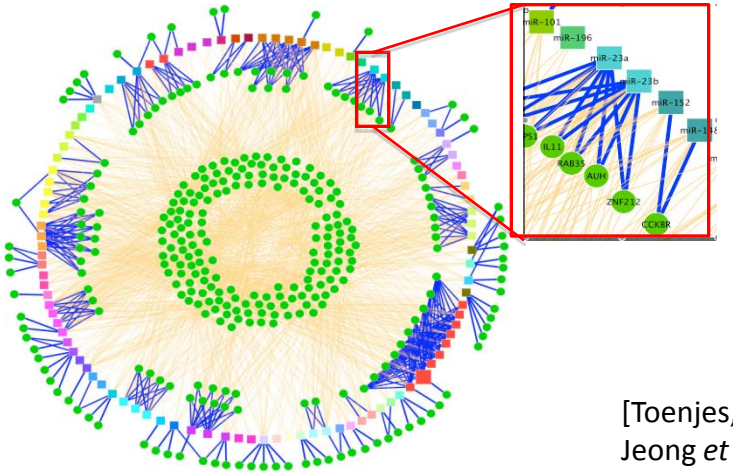
TF-target-gene Regulatory networks



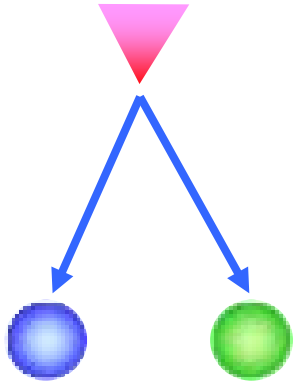
Undirected



Metabolic pathway networks



miRNA-target networks



Directed

[Toenjes, *et al*, *Mol. BioSyst.* (2008); Jeong *et al*, *Nature* (2001); [Horak, et al, Genes & Development, 16:3017-3033; DeRisi, Iyer, and Brown, Science, 278:680-686]

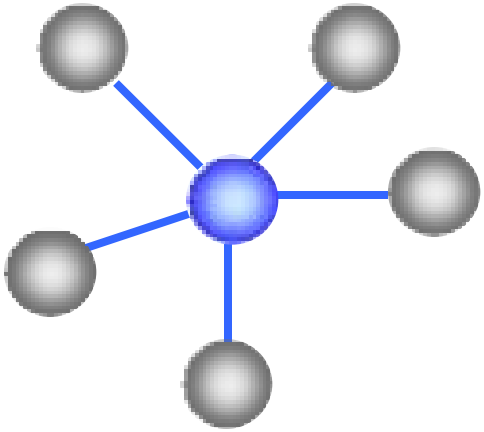
Q1: Finding Central Points in Networks: Hubs & Bottlenecks

Where are key points in networks ? How do we locate them ?

Hub & bottleneck?

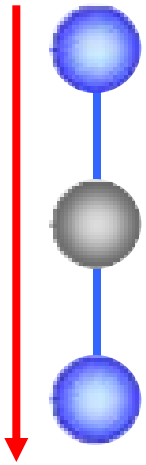
Global topological measures

Indicate the gross topological structure of the network



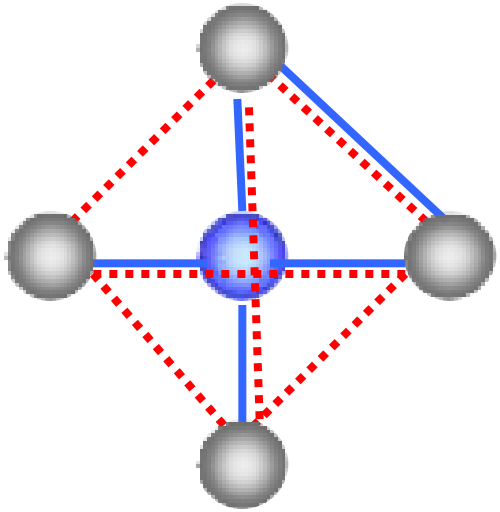
Degree (K)

5



Path length (L)

2

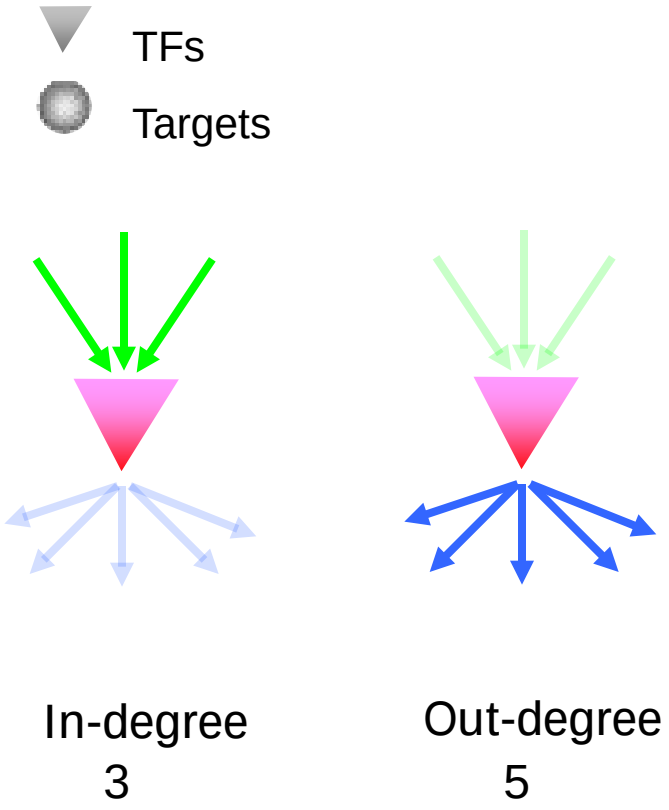


Clustering coefficient (C)

1/6

Interaction and expression networks are *undirected*

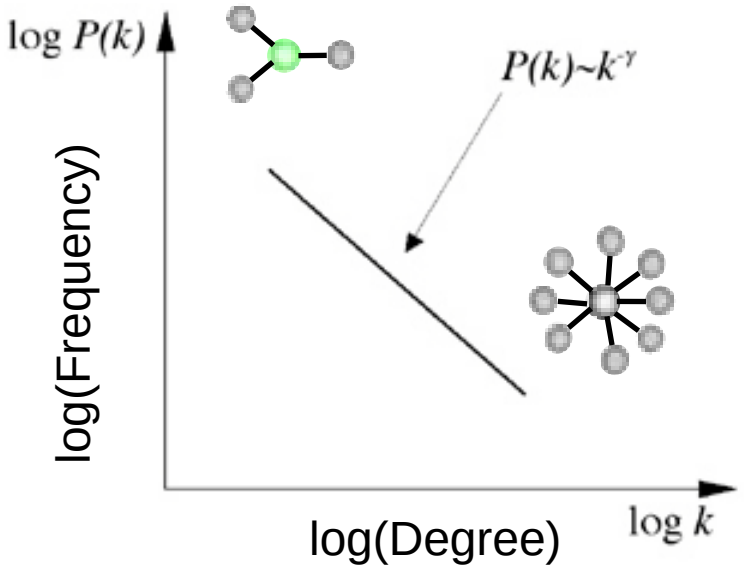
Global topological measures for directed networks



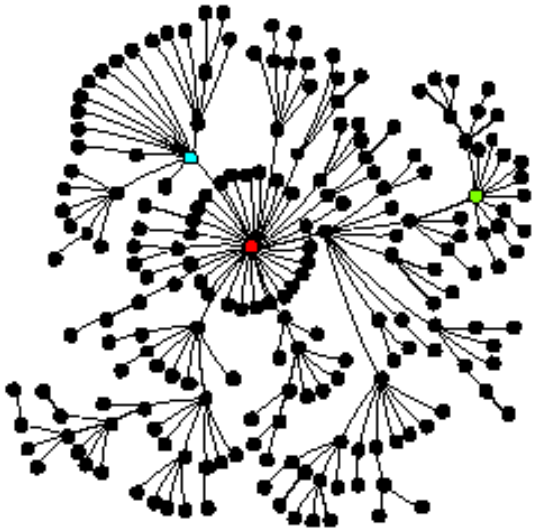
Regulatory and metabolic networks are *directed*

Scale-free networks

Power-law distribution



A *scale-free* network is a network whose degree distribution follows a power law



Hubs dictate the structure of the network

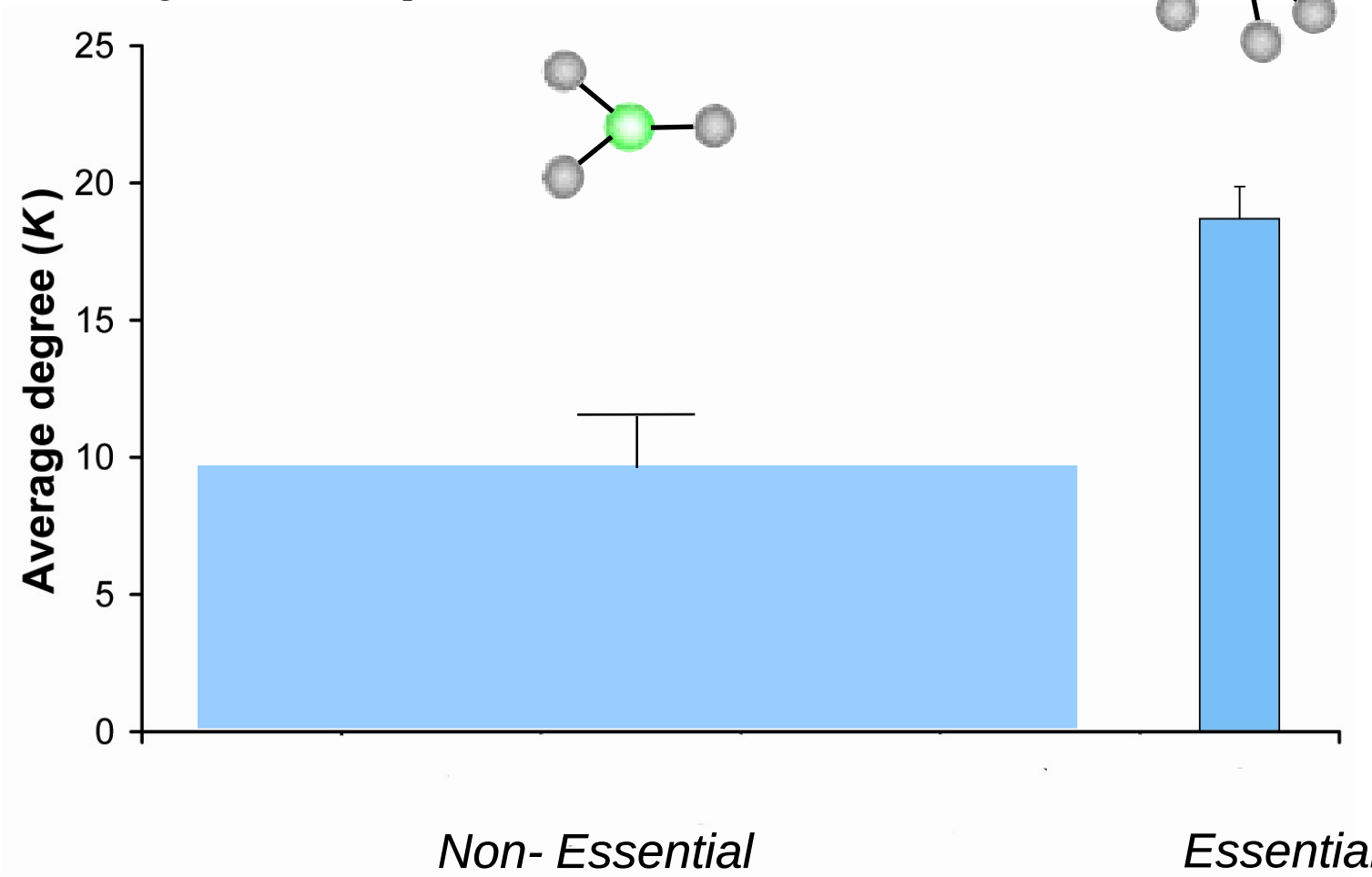
[Barabasi]

Hubs tend to be Essential

Integrate gene essentiality data with protein interaction network. Perhaps hubs represent vulnerable points?

[Lauffenburger, Barabasi]

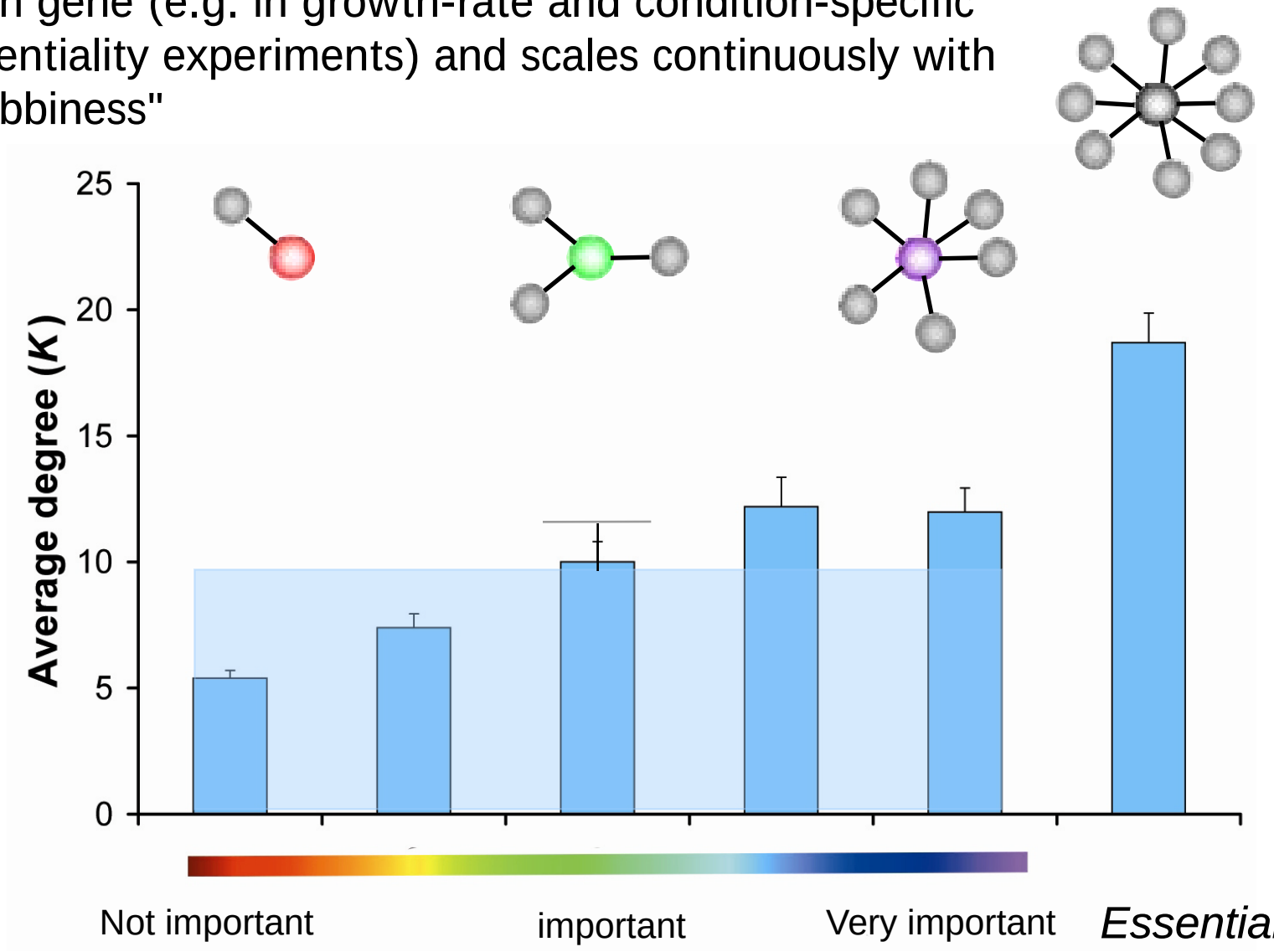
"hubbiness"



[Yu et al., 2003, TIG]

Relationships extends to "Marginal Essentiality"

Marginal essentiality measures relative importance of each gene (e.g. in growth-rate and condition-specific essentiality experiments) and scales continuously with "hubbiness"



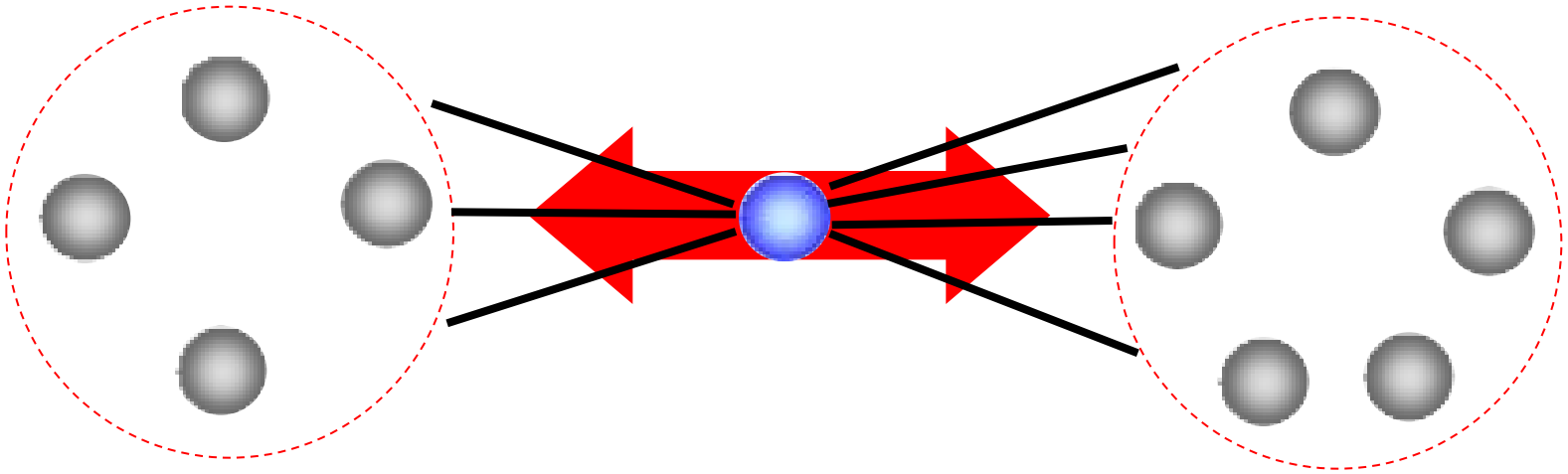
[Yu et al., 2003, TIG]

Another measure of Centrality: Betweenness centrality

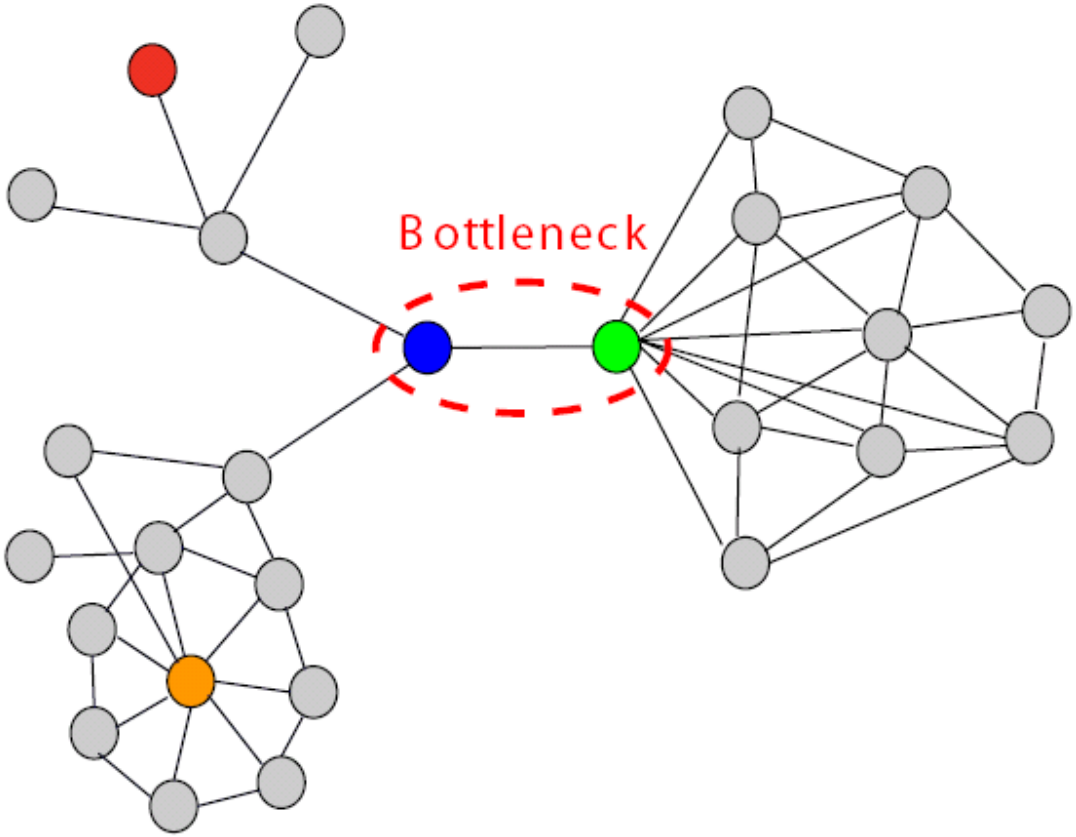
Betweenness of a node is the number of shortest paths of pairs of vertices that run through it -- a measure of information flow.

Freeman LC (1977) Set of measures of centrality based on betweenness. Sociometry 40: 35–41.

Girvan & Newman (2002) PNAS 99: 7821.



Bottlenecks & Hubs



Hub-bottleneck **node**

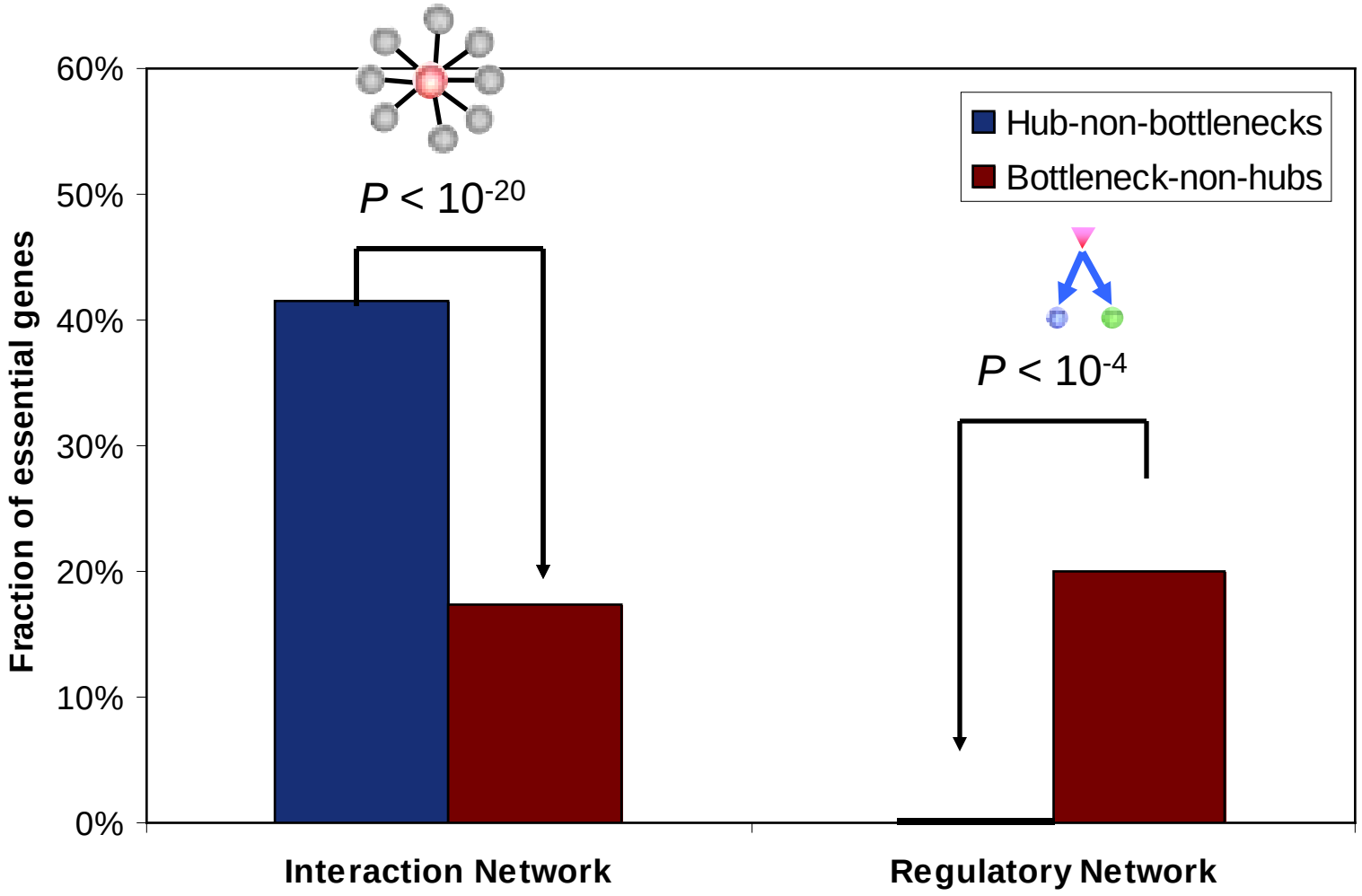
Non-hub-bottleneck **node**

Hub-non-bottleneck **node**

Non-hub-non-bottleneck **node**

[Yu et al., PLOS CB (2007)]

Bottlenecks are what matters in regulatory networks

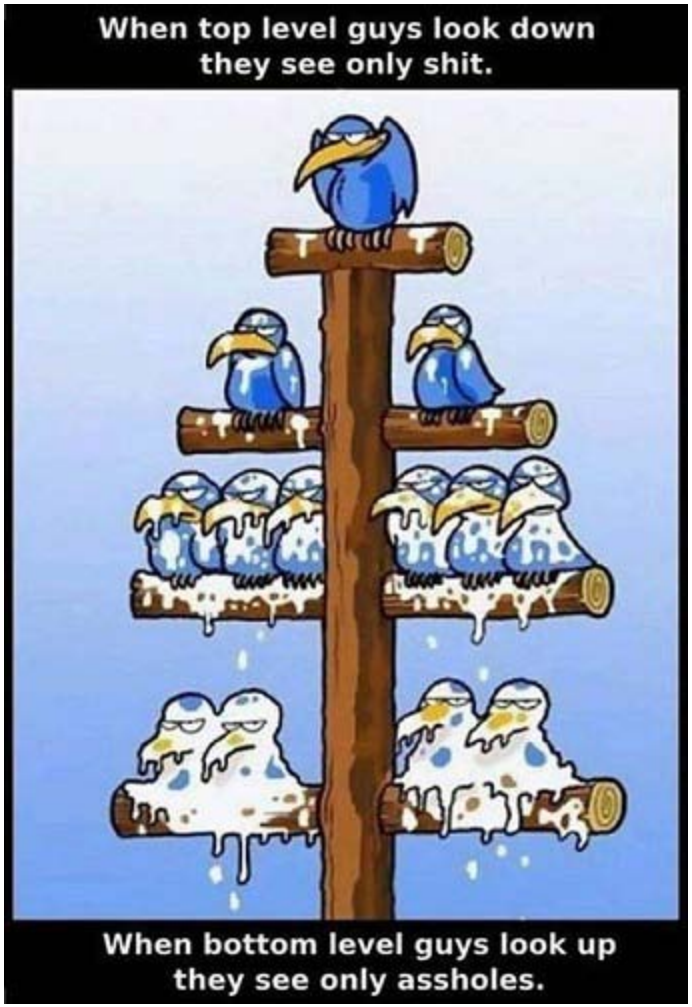


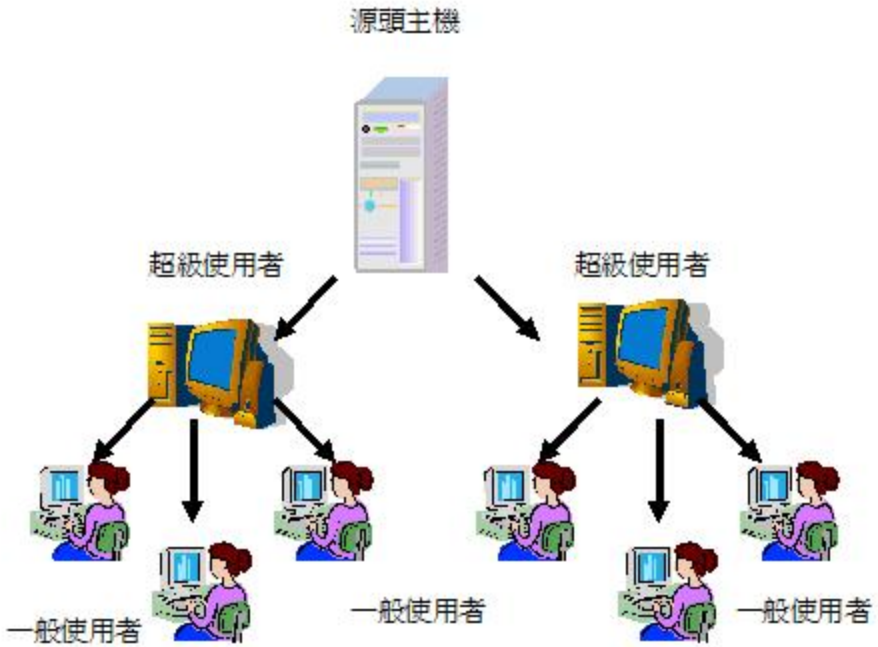
[Yu *et al.*, *PLoS Comput Biol* (2007)]

Q2: Does the Bio-molecular networks posses hierarchy structure

If the network has the hierarchy structure? How do we identify them?
What does it mean?

Hierarchy structure

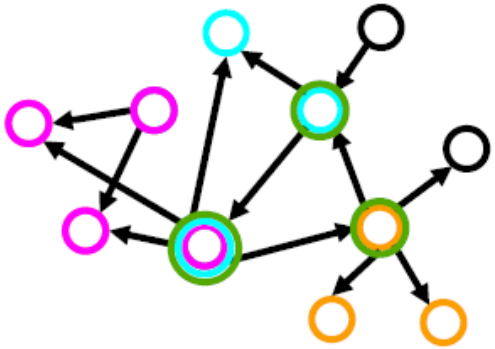




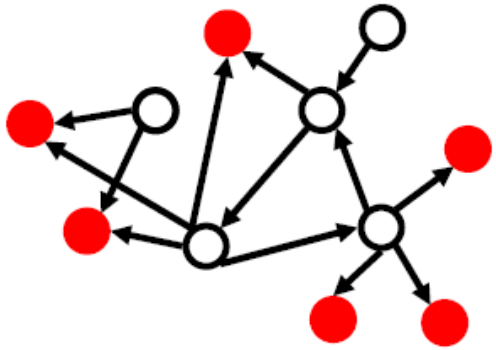
Management Hierachy

Determination of "Level“ in Regulatory Network Hierarchy with Breadth-first Search

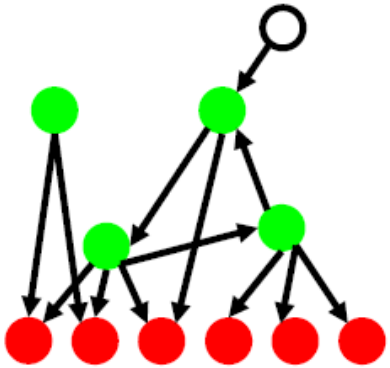
I. Example network with all 4 motifs



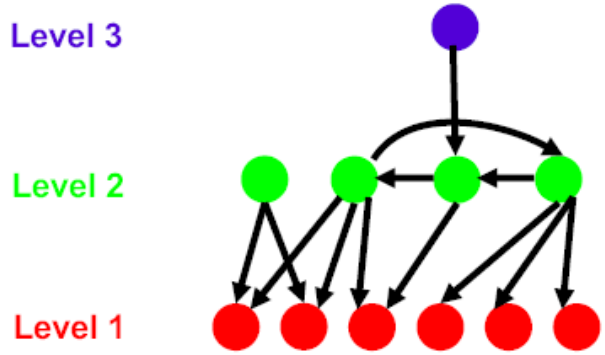
II. Finding terminal nodes (Red)



III. Finding mid-level nodes (Green)

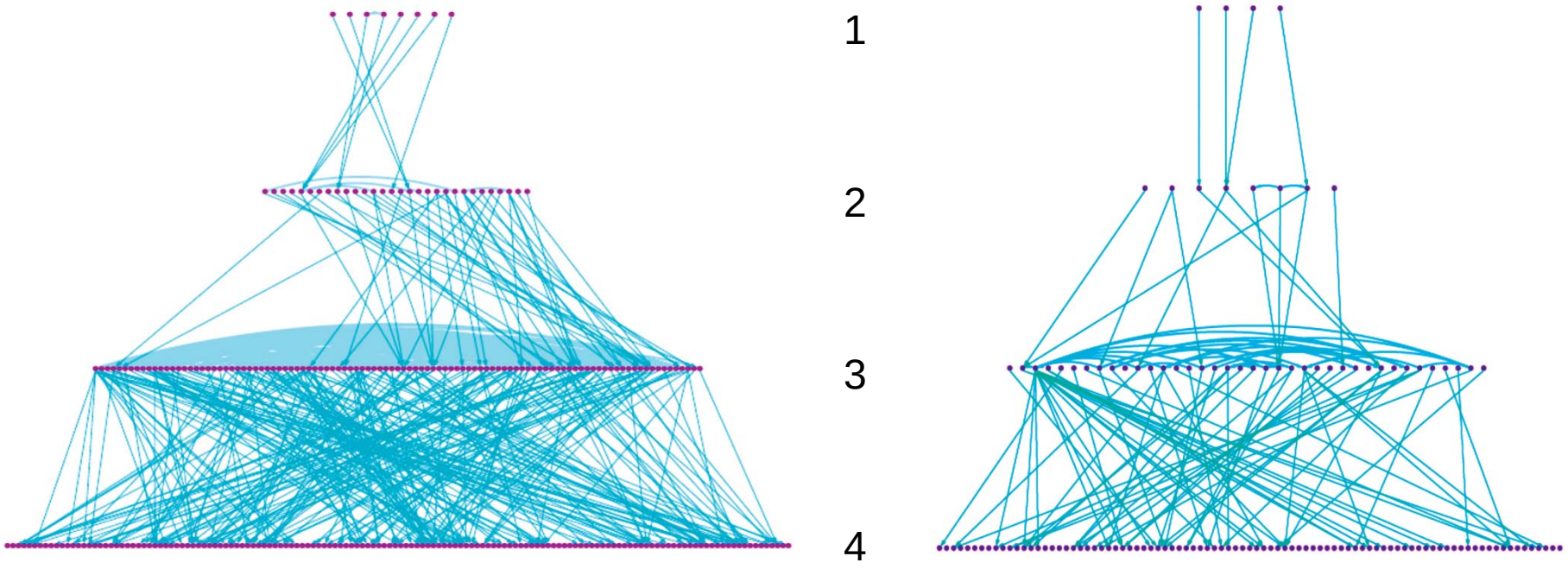


IV. Finding top-most nodes (Blue)



[Yu et al., PNAS (2006)]

Regulatory Networks have similar hierarchical structures

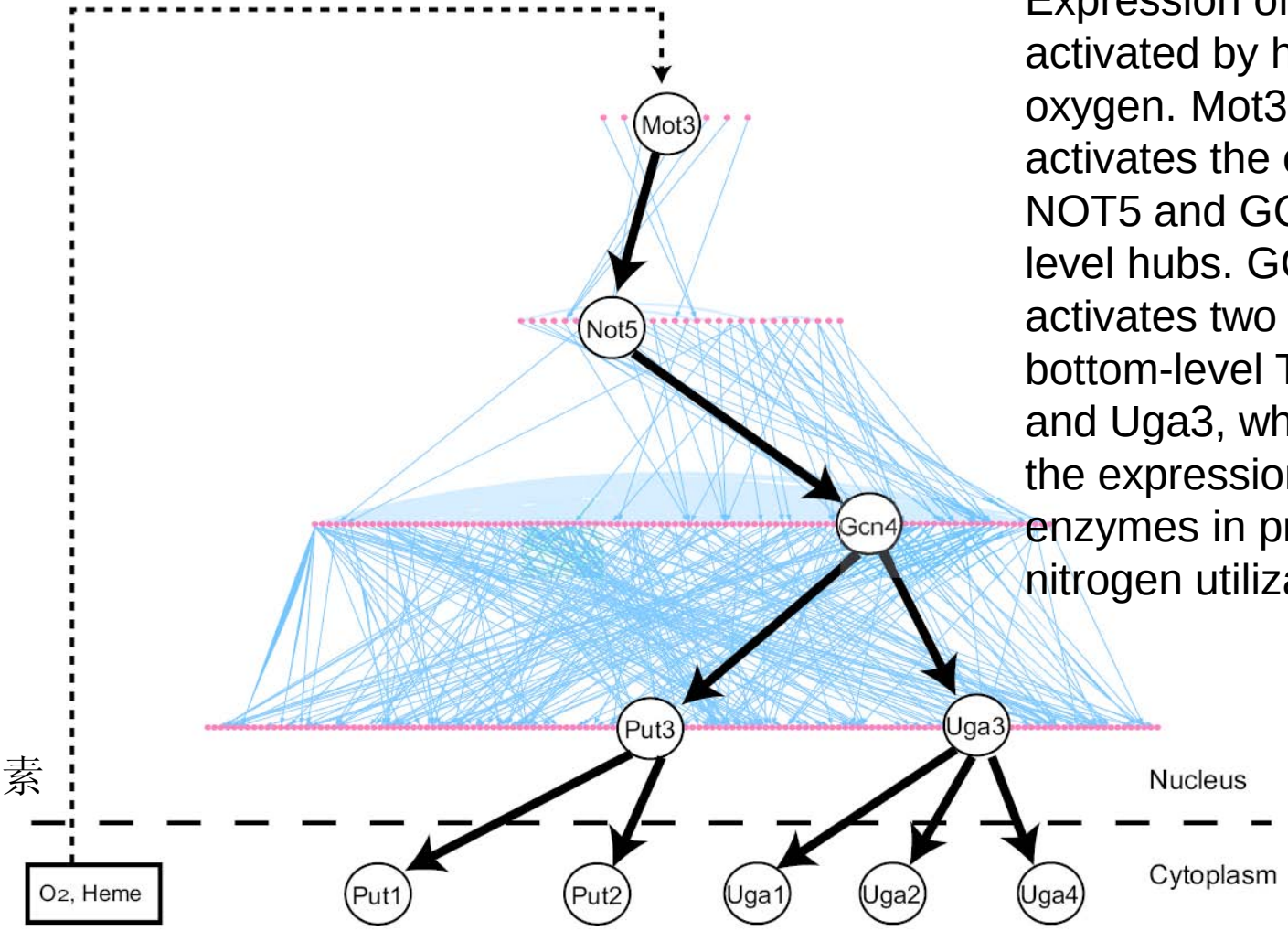


S. cerevisiae

E. coli

[Yu *et al.*, *Proc Natl Acad Sci U S A* (2006)]

Example of Path Through Regulatory Network

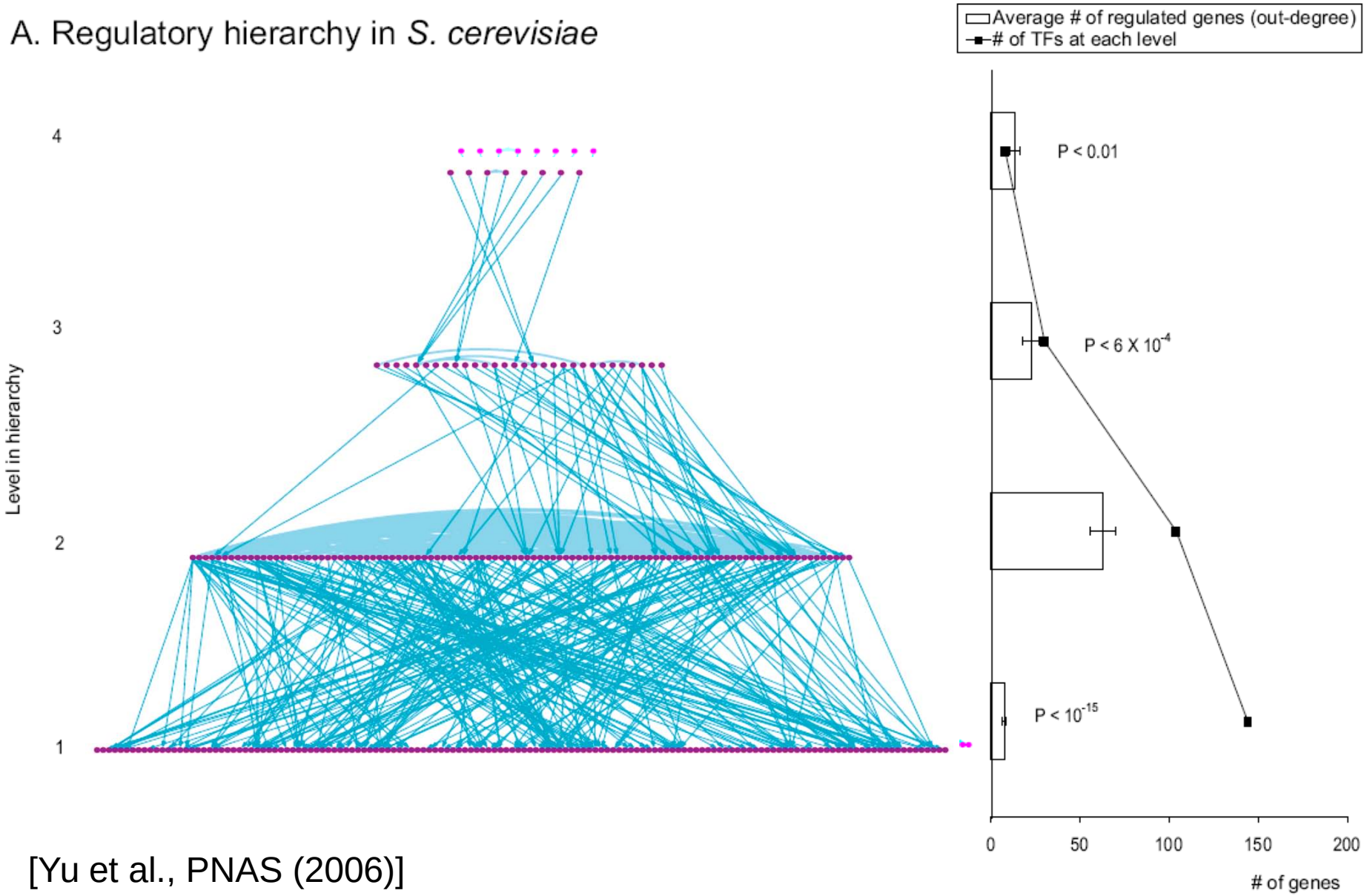


Expression of MOT3 is activated by heme and oxygen. Mot3 in turn activates the expression of NOT5 and GCN4, mid-level hubs. GCN4 activates two specific bottom-level TFs, Put3 and Uga3, which trigger the expression of enzymes in proline and nitrogen utilization.

亚铁血红素

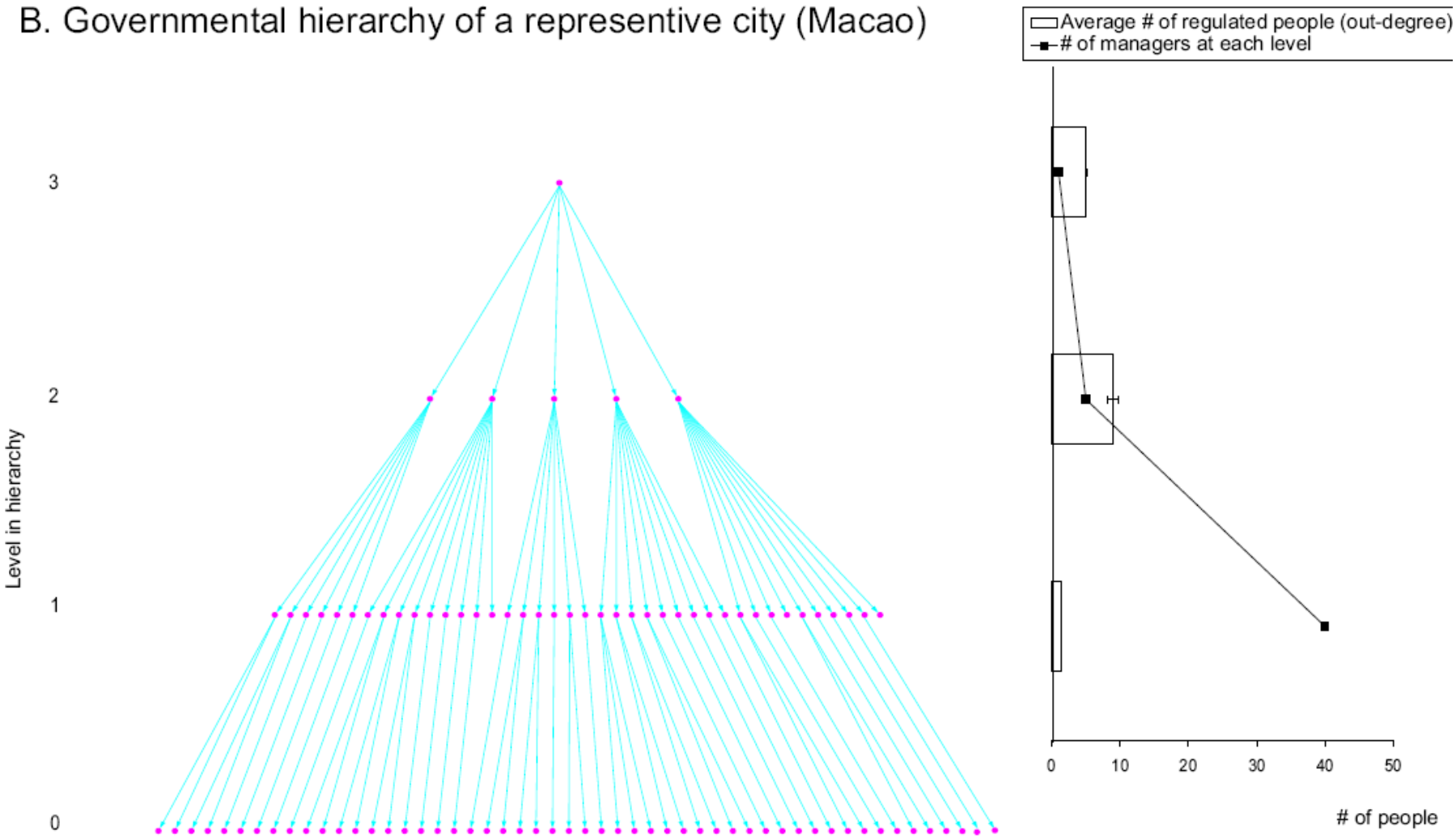
[Yu et al., PNAS (2006)]

Yeast Regulatory Hierarchy

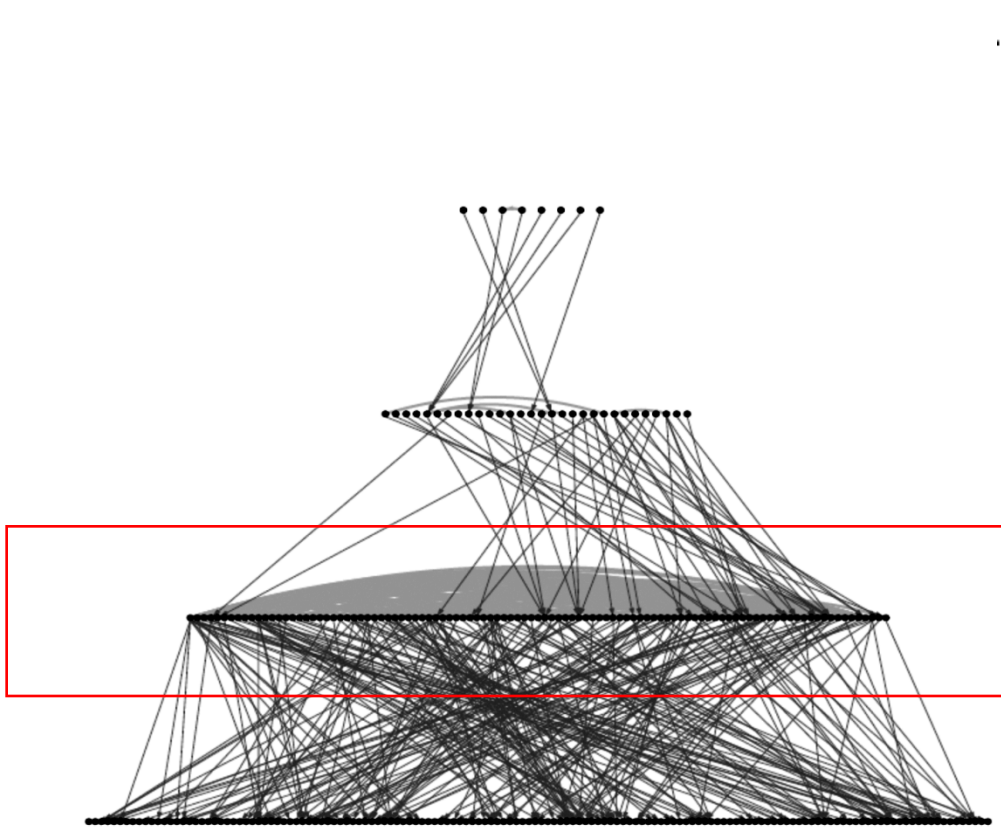


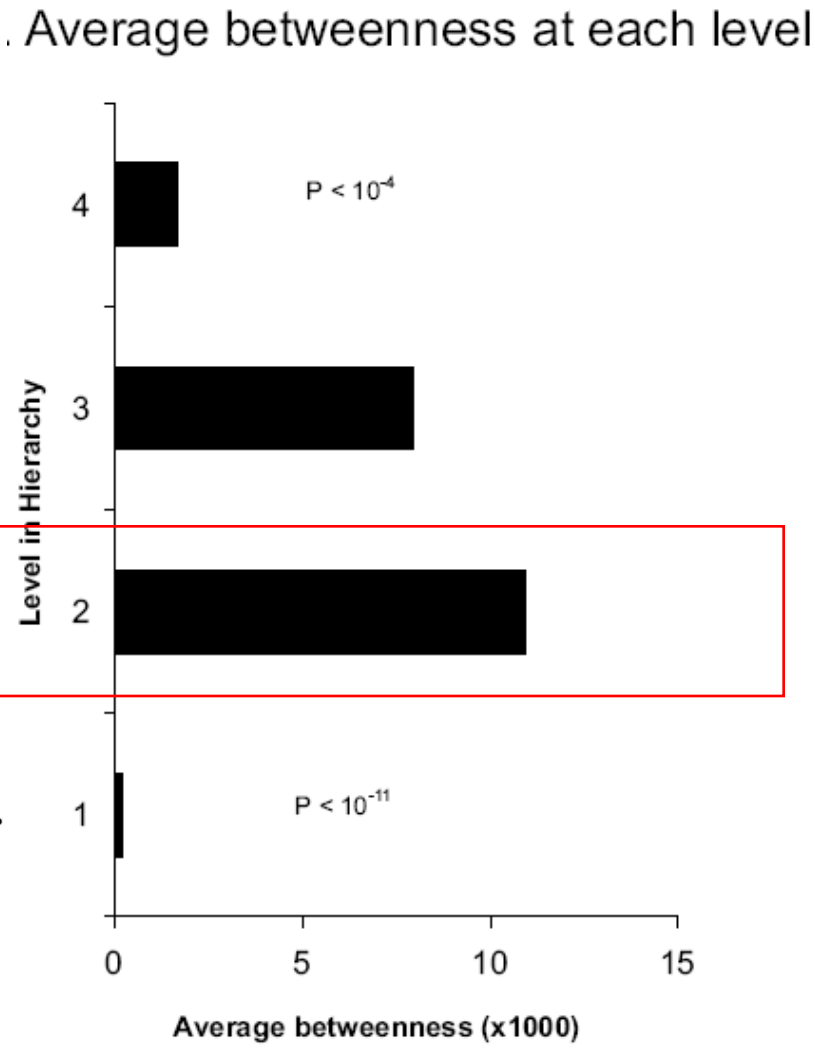
[Yu et al., PNAS (2006)]

Yeast Network Similar in Structure to Government Hierarchy with Respect to Middle-managers



Characteristics of Regulatory Hierarchy: Middle Managers are Information Flow Bottlenecks

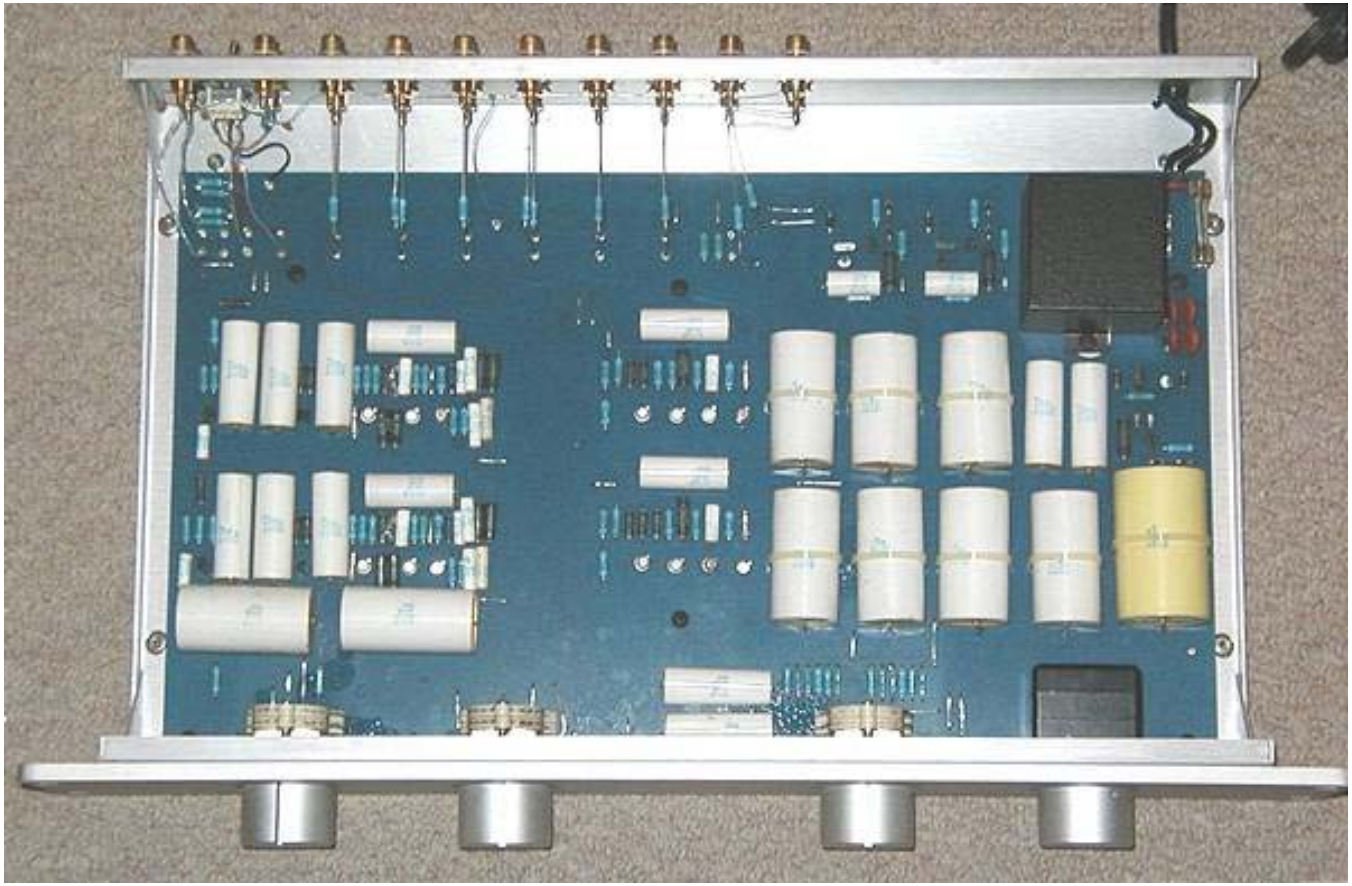




[Yu et al., PNAS (2006)]

Q3: Are there some building blocks in the Bio-molecular networks?

Where are they? How do we identify them? What does it mean?



Circuit network

Building blocks: Switch, feed-back loop, oscillator…

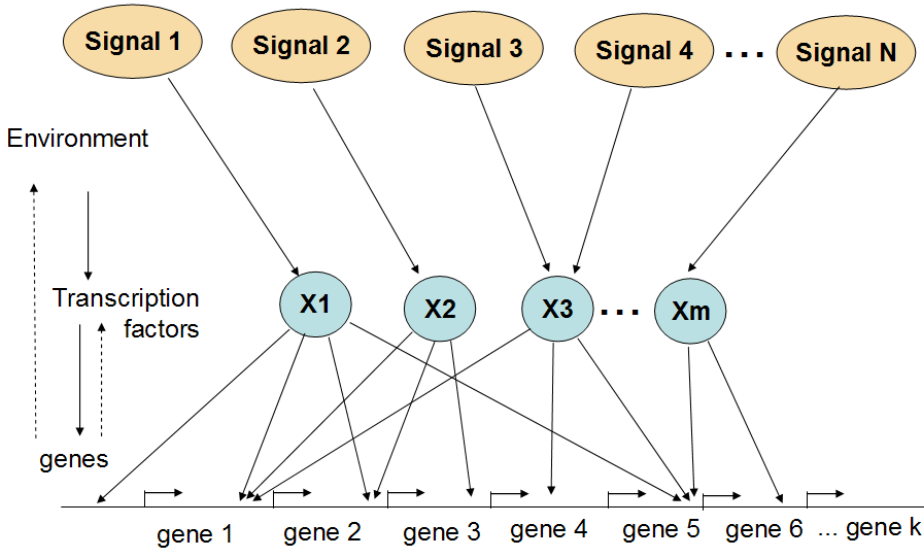
Network Motifs: simple Building Blocks of Complex Networks

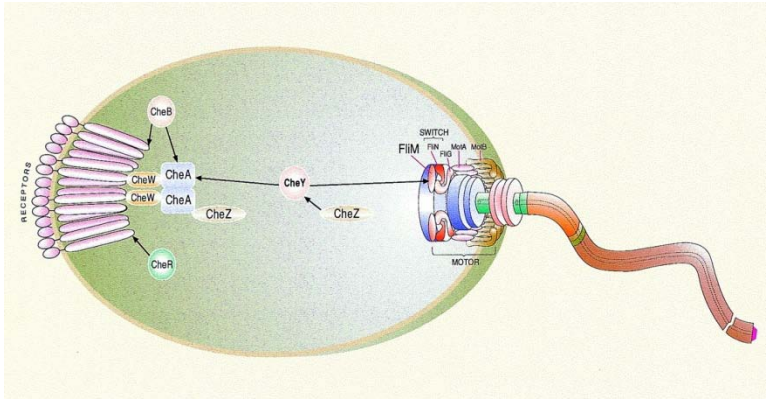
- R. Milo *et. al. Science* ***298****, 824 (2002)*

- the design principles of this network
- “Evolution preserves modules that define specific function.”
- Motifs are those subgraphs which occur in higher frequencies than in random graphs.

The cell and the environment

- Cells need to react to their environment
- Reaction is by synthesizing task-specific proteins, on demand.
- The solution – regulated transcription network

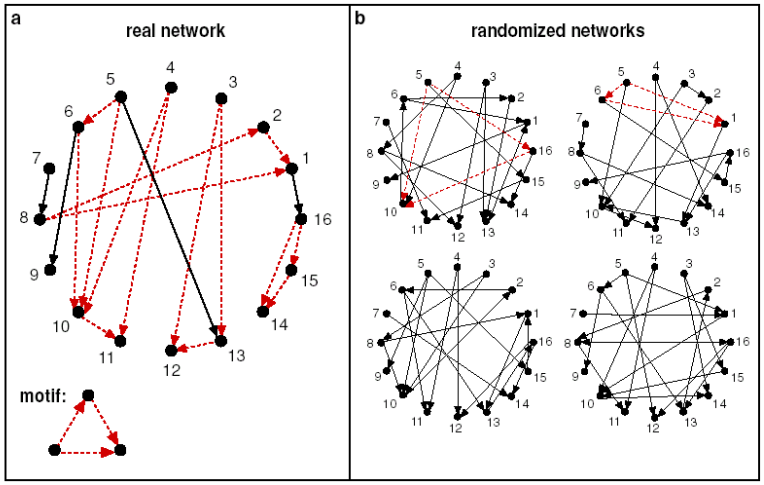




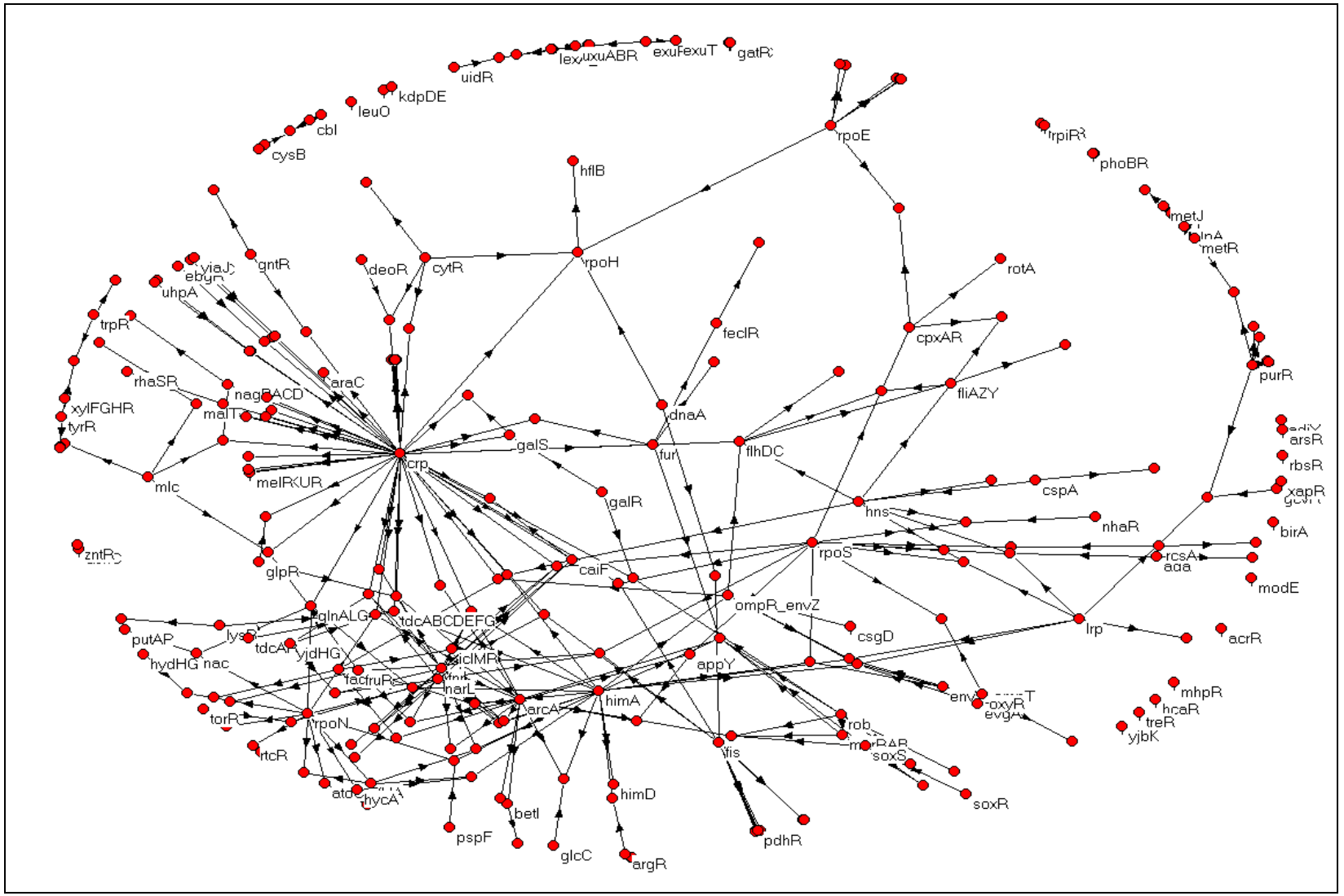
- E. Coli – 1000 protein types at any given moment >4000 genes (or possible protein types) – need regulatory mechanism to select the active set
- We are interested in the design principles of this network

Analyzing networks

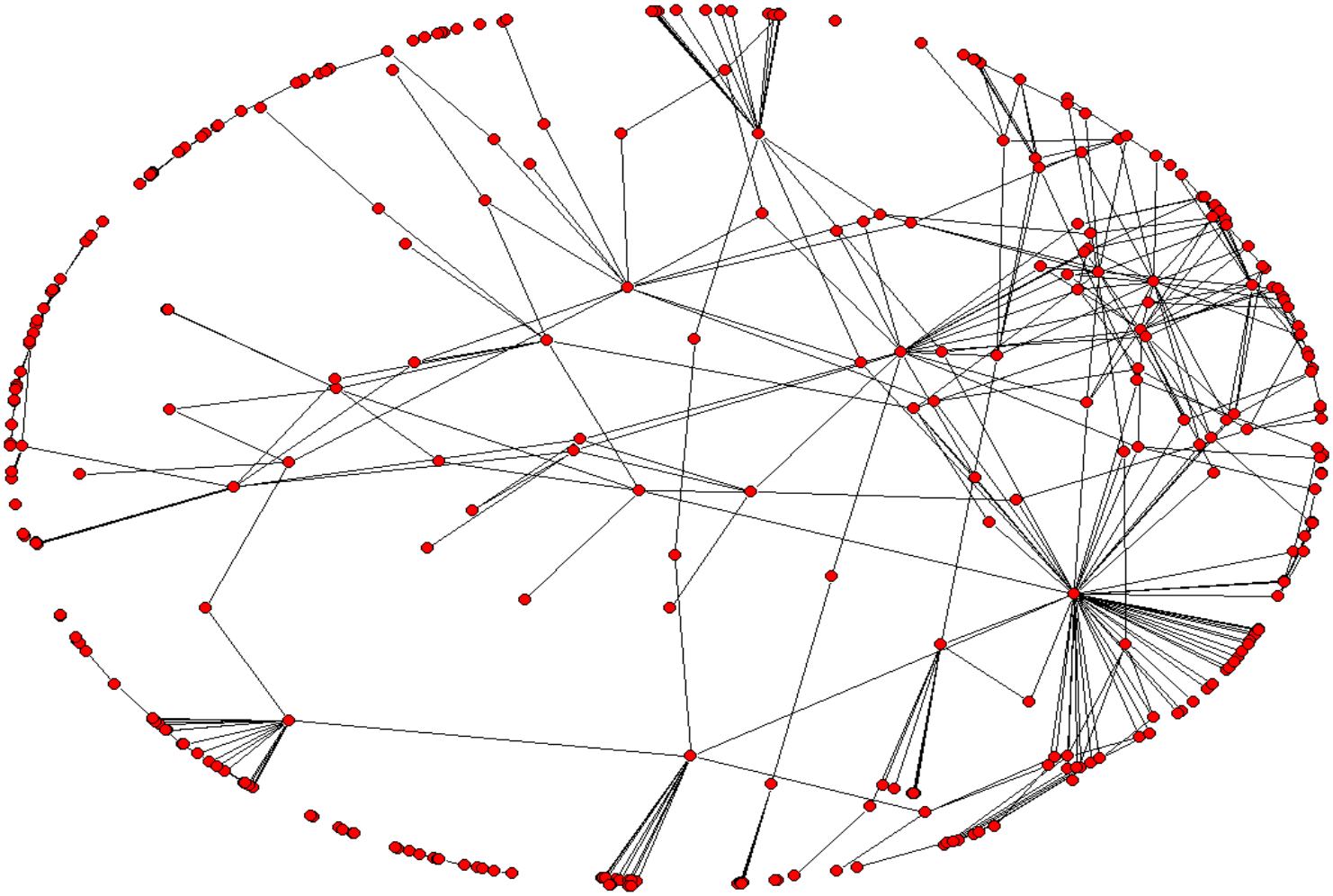
- The idea- patterns that occur in the real network much more then in a randomized network, must have functional significance.
- The randomized networks share the same number of edges and number of nodes, but edges are assigned at random



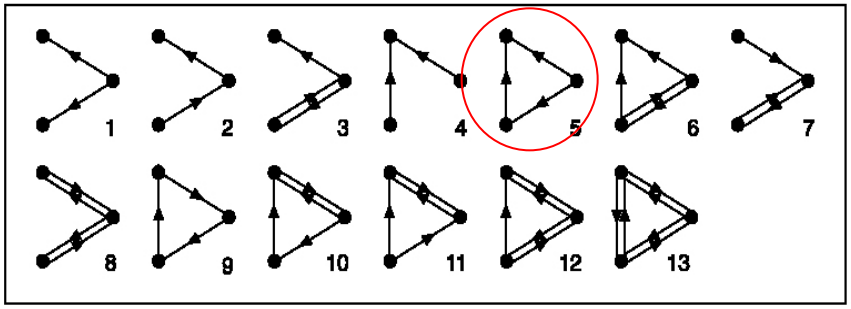
The known E. Coli transcription network

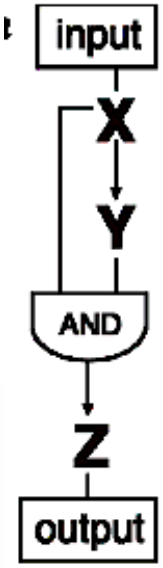


A random graph based on the same node statistics



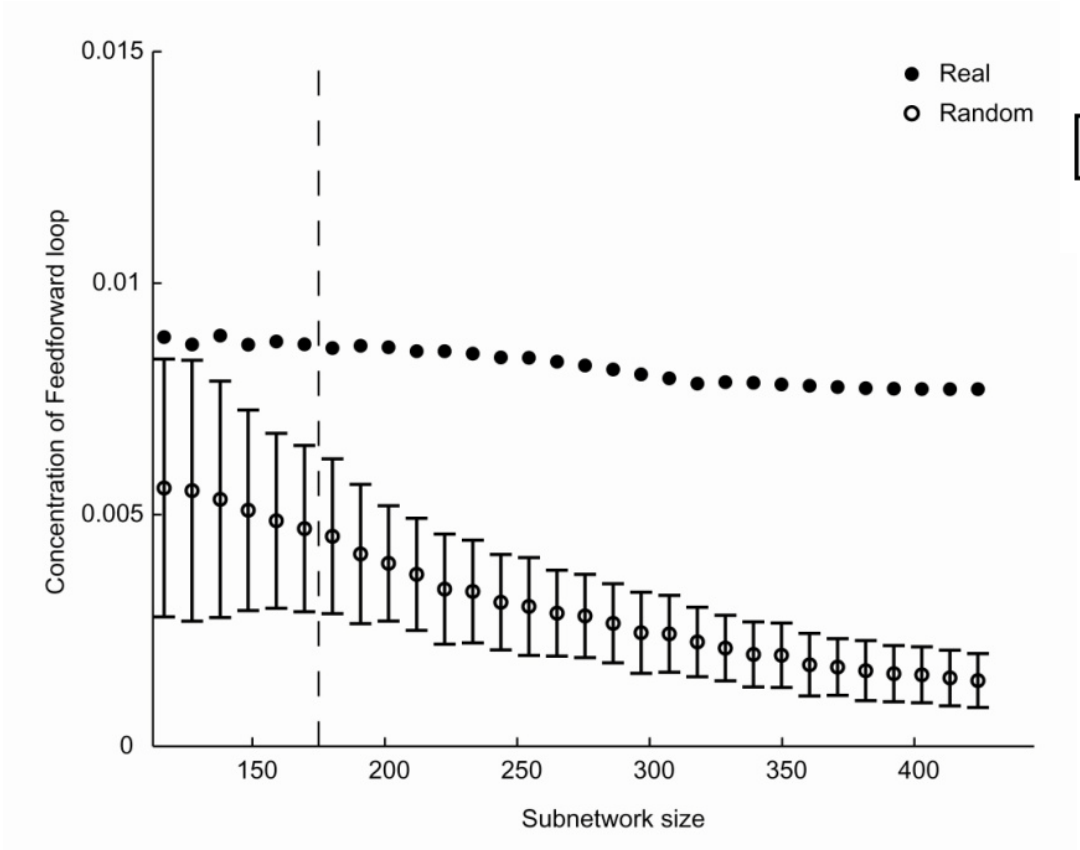
3-node network motif – the feedforward loop



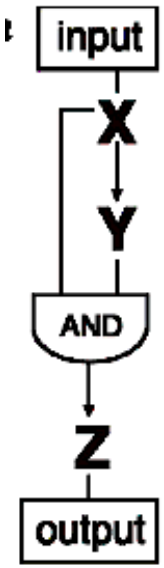


Nreal=40

Nrand=7±3



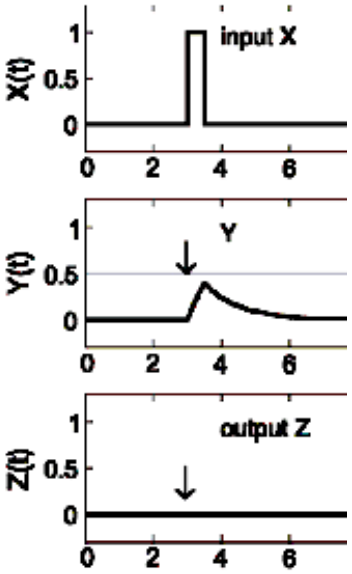
The feedforward loop : a sign sensitive filter



$$X = X(t)$$

$$\frac{dY}{dt} = \theta(X - k_{XY}) - Y$$

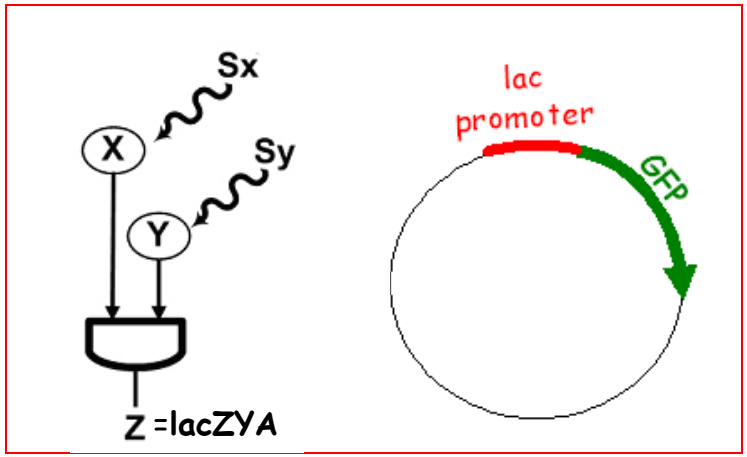
$$\frac{dZ}{dt} = \theta(X - K_{XZ})\theta(Y - K_{YZ}) - Z$$



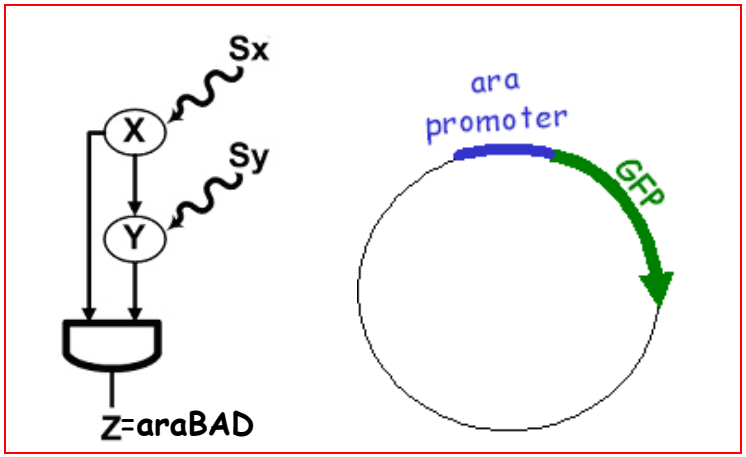
The feedforward loop is a filter for transient signals while allowing fast shutdown

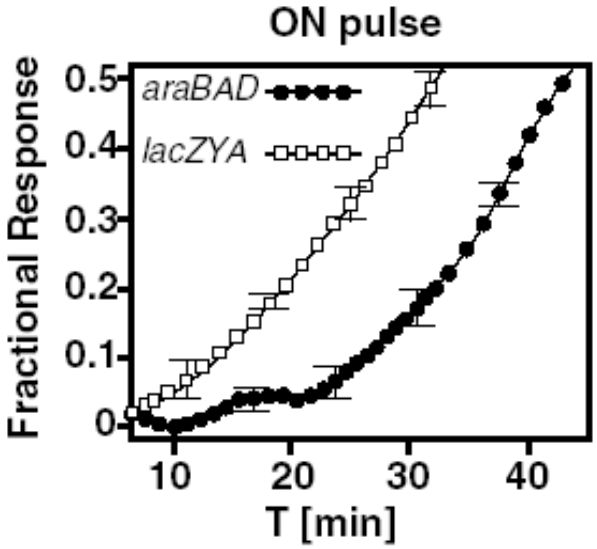
Mangan, Alon, PNAS, JMB, 2003

The Feedforward loop : a sign sensitive filter



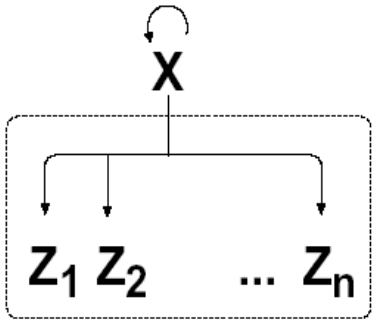
Vs.

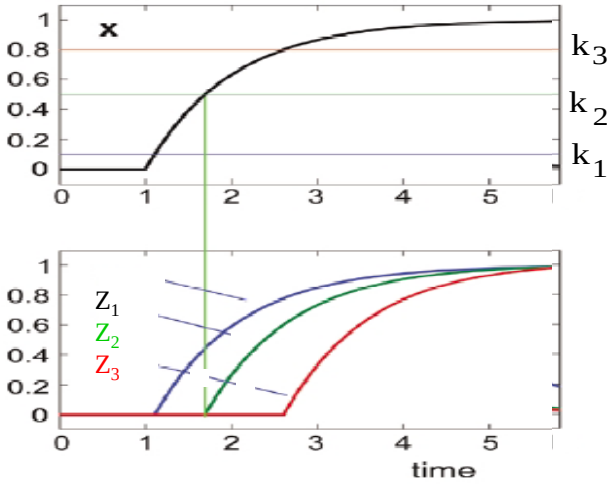




Mangan, Alon, PNAS, JMB, 2003

Single Input Module





$$X = X(t)$$

$$\frac{dZ_i}{dt} = \beta_i \theta(X - k_i) - Z_i$$

$$\begin{cases} k_1 < k_2 < \ldots < k_n \\ \beta_1 > \beta_2 > \ldots > \beta_n \end{cases}$$

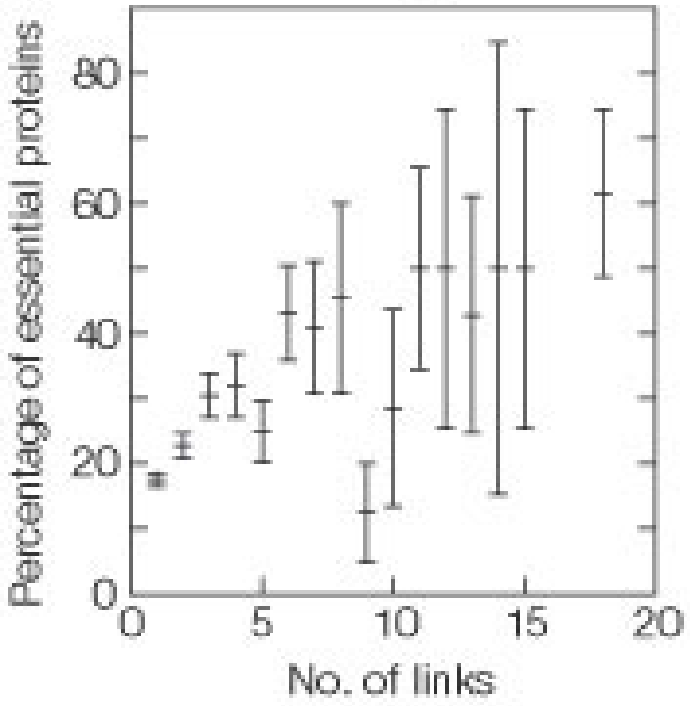
Temporal and expression level program generator

- The temporal order is encoded in a hierarchy of thresholds
- Expression levels hierarchy is encoded in hierarchy of promoter activities

Q4: Hubs in the interactome network are known to be very important to the network topology and function.
Considering the temporal aspect of the interactome, are all hubs equal?

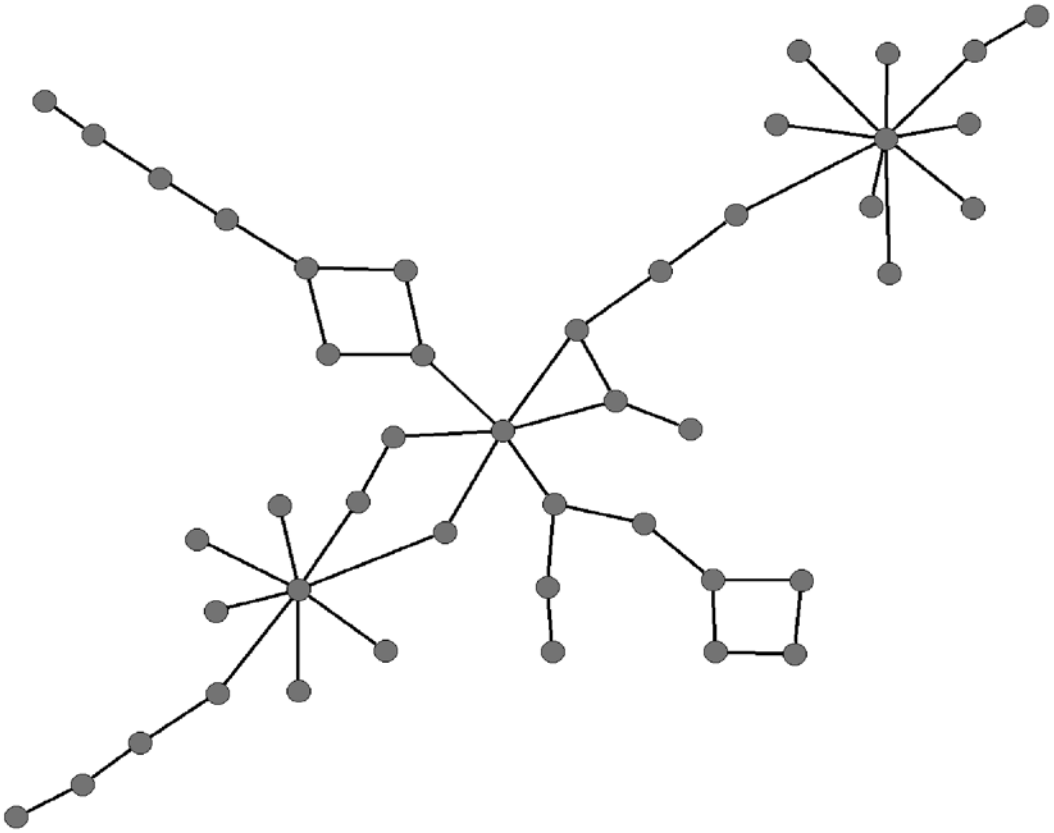
Yeast hubs are three-times more likely to be essential

Yeast Interactome mapped by Y2H is scale-free



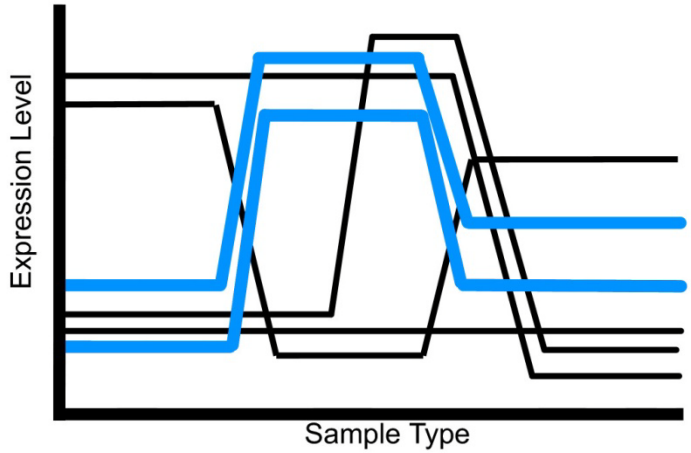
Jeong et al Nature 2001

Static view of the interactome network



Let’s introduce other dimension.

A Array Data



Correlation coefficients for all genes

B Similarity Matrix (correlation)

	G1	G2	G3	G4	G5	G6	G7	G8	G9	G10
G1	1	0.6	0.2	0.8	0.9	0.6	0.9	0.1	0.5	0.3
G2	0.6	1	0.9	0.1	0.2	0.6	1.0	0.1	0.3	0.4
G3	0.2	0.9	1	0.2	0.3	0.4	0.8	0.2	0.3	0.9
G4	0.8	0.1	0.2	1	0.9	0.9	0.8	0.3	0.6	0.0
G5	0.9	0.2	0.3	0.9	1	0.9	0.9	0.6	0.1	0.5
G6	0.6	0.6	0.4	0.9	0.9	1	0.6	0.2	0.7	0.1
G7	0.9	1.0	0.8	0.8	0.9	0.6	1	0.8	0.9	0.2
G8	0.1	0.1	0.2	0.3	0.6	0.2	0.8	1	0.9	0.2
G9	0.5	0.3	0.3	0.6	0.1	0.7	0.9	0.9	1	0.9
G10	0.3	0.4	0.9	0.0	0.5	0.1	0.2	0.2	0.9	1

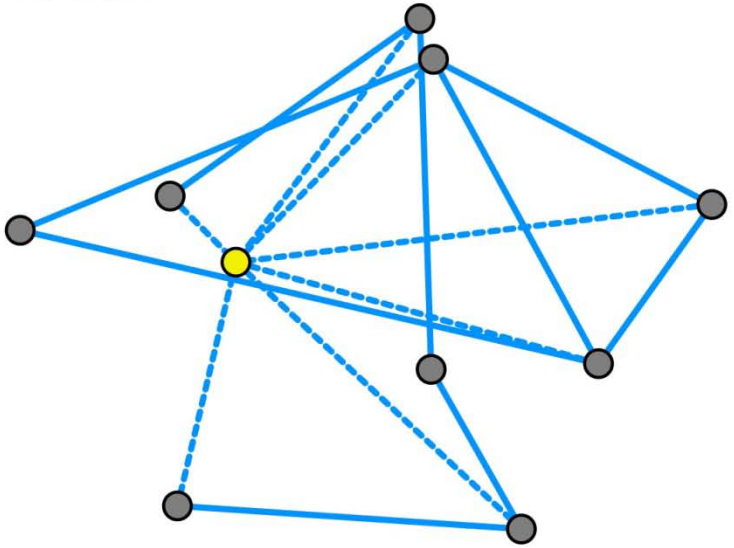
Threshold correlations into edges

C Adjacency Matrix

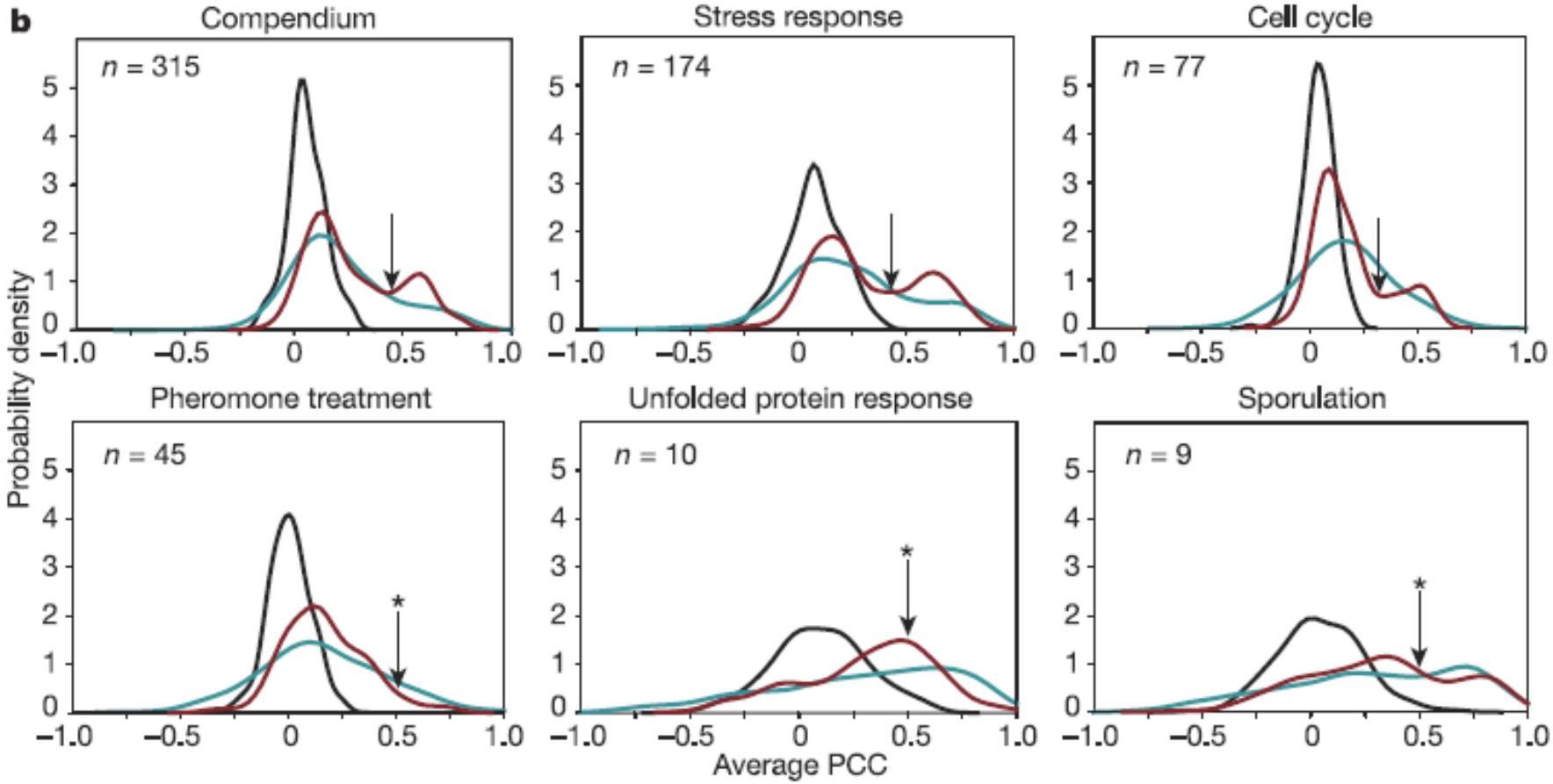
	G1	G2	G3	G4	G5	G6	G7	G8	G9	G10
G1	NA	O	O	E	E	O	E	O	O	O
G2	O	NA	E	O	O	O	E	O	O	O
G3	O	E	NA	O	O	O	E	O	O	E
G4	E	O	O	NA	E	E	E	O	O	O
G5	E	O	O	E	NA	E	E	O	O	O
G6	O	O	O	E	E	NA	O	O	O	O
G7	E	E	E	E	E	O	NA	E	E	O
G8	O	O	O	O	O	O	E	NA	E	O
G9	O	O	O	O	O	O	E	E	NA	E
G10	O	O	E	O	O	O	O	O	E	NA

Draw network

D Network



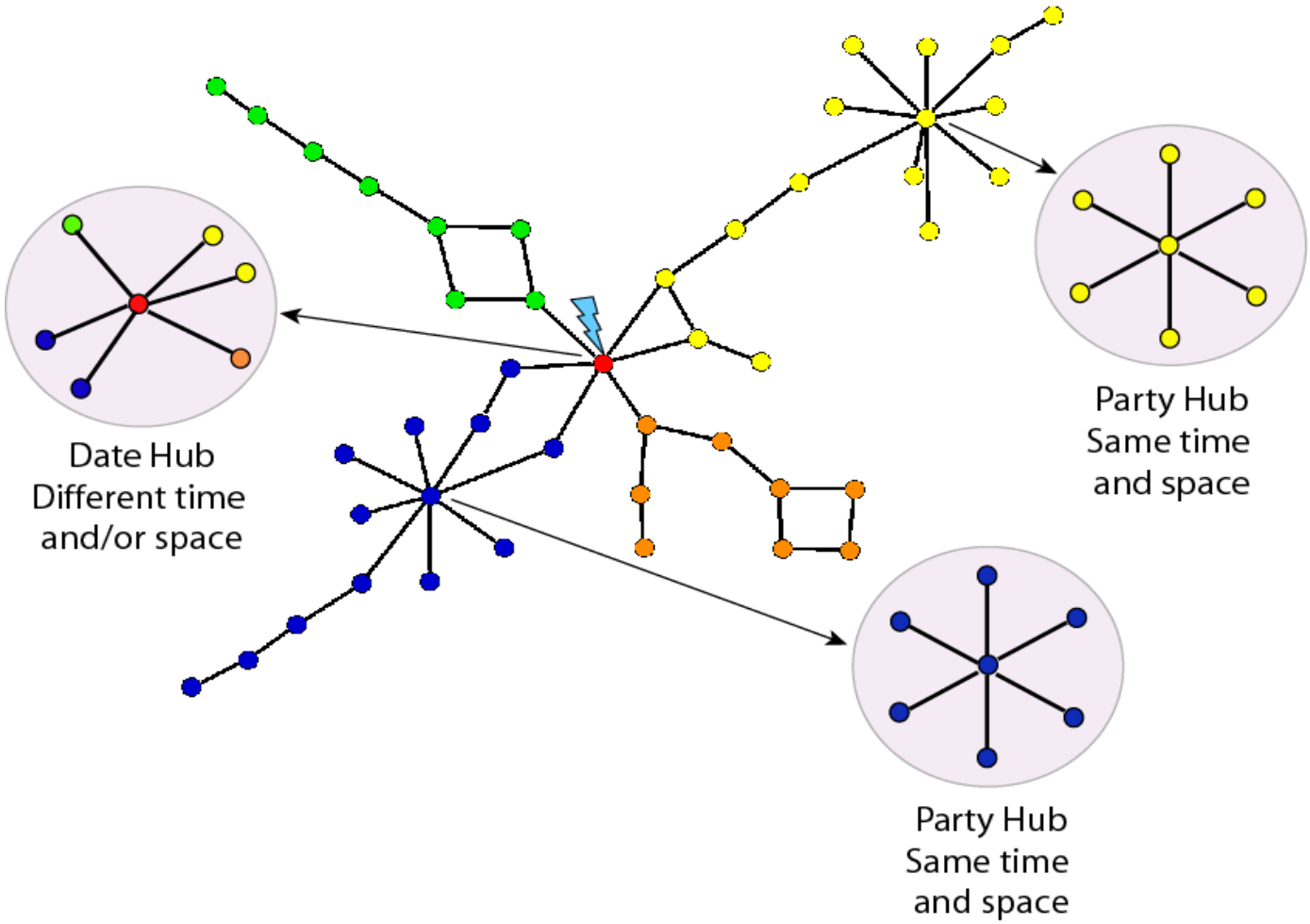
Co-expression in different conditions



-- hubs; -- non-hubs; -- randomized net

Are all hubs equal?

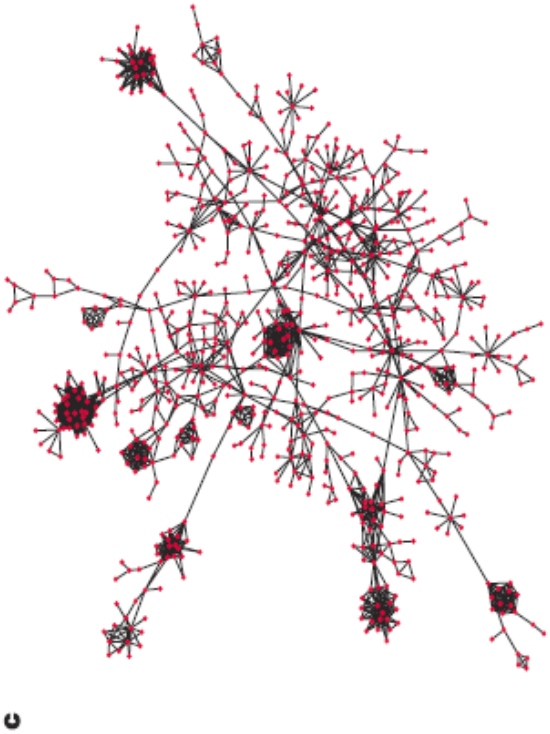
Dynamic or temporal aspects of interactome networks



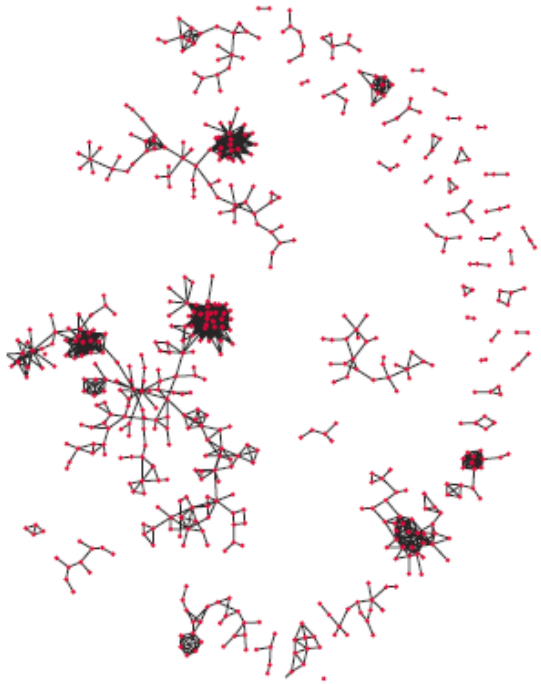
Han *et al.* Nature, 2004

Their Role in the Net

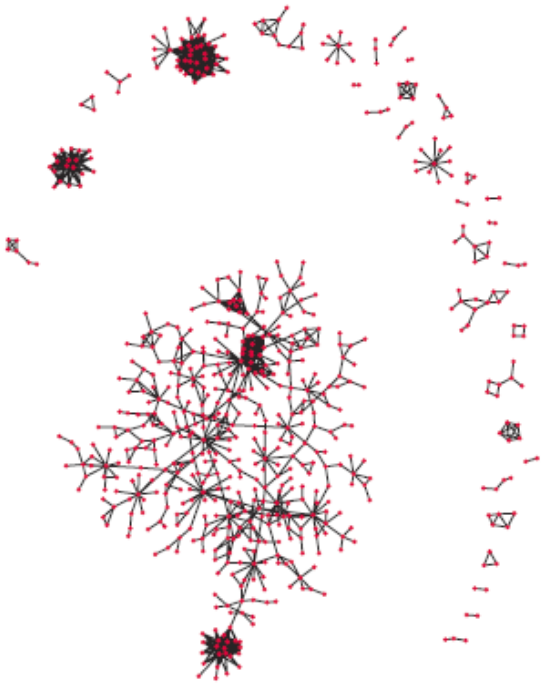
Full Net



No Date Hubs

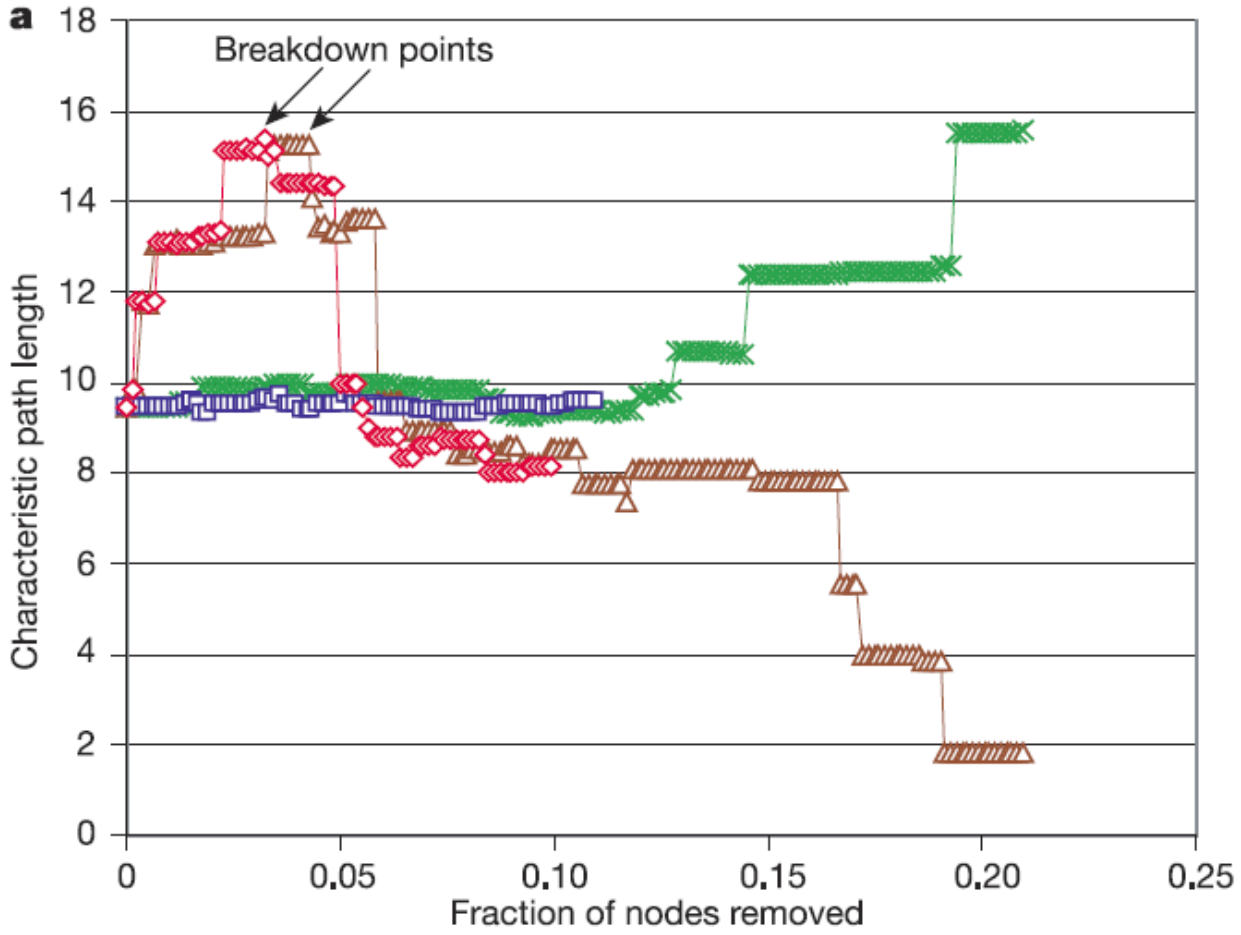


No Party Hubs



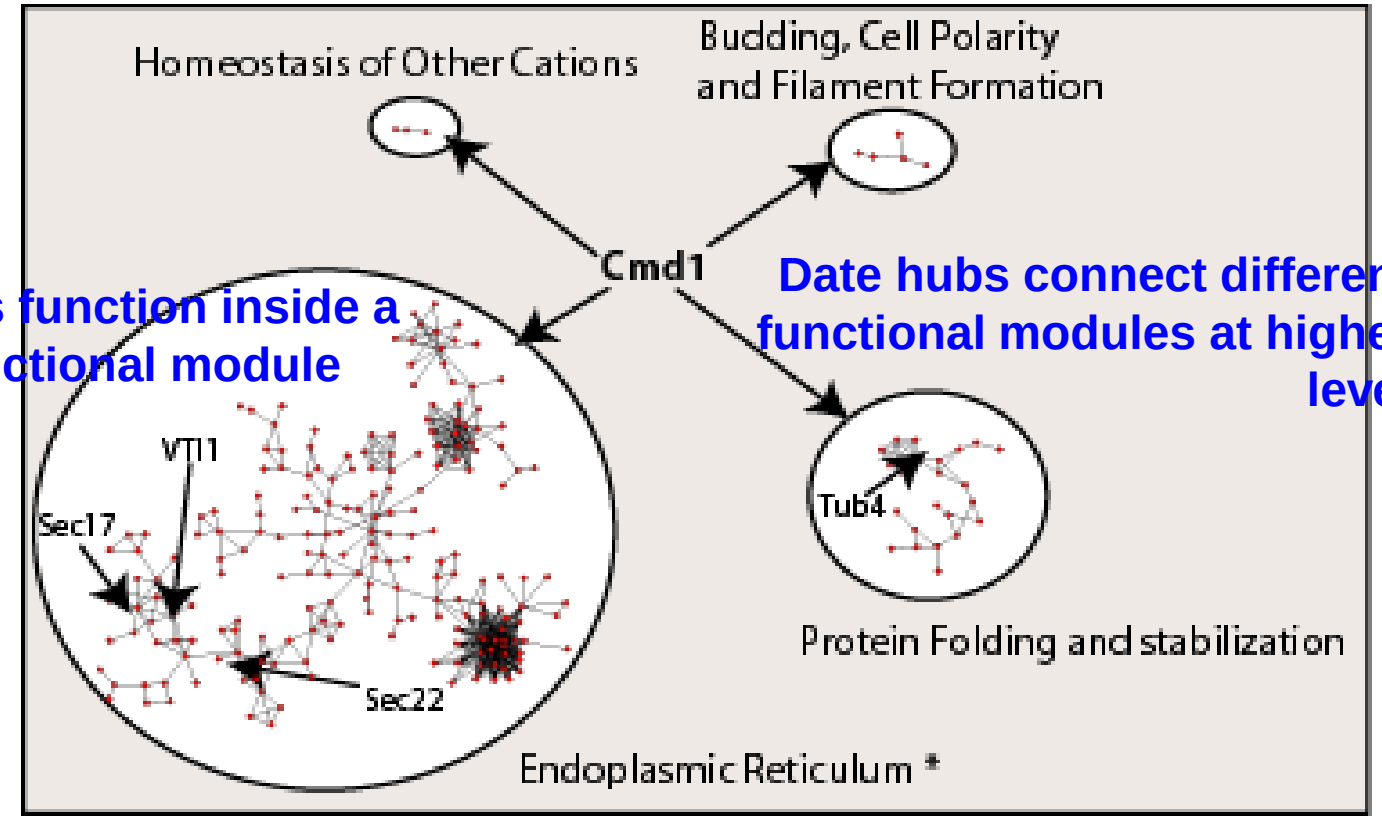
In silico simulation of node removal

-- Random
-- Hubs
-- Party
-- Date



Characteristic Path Length: For any connected graph G, the average distance between pairs of vertices is referred to as the graph's "characteristic path length"

Dynamic modular structure of yeast interactome



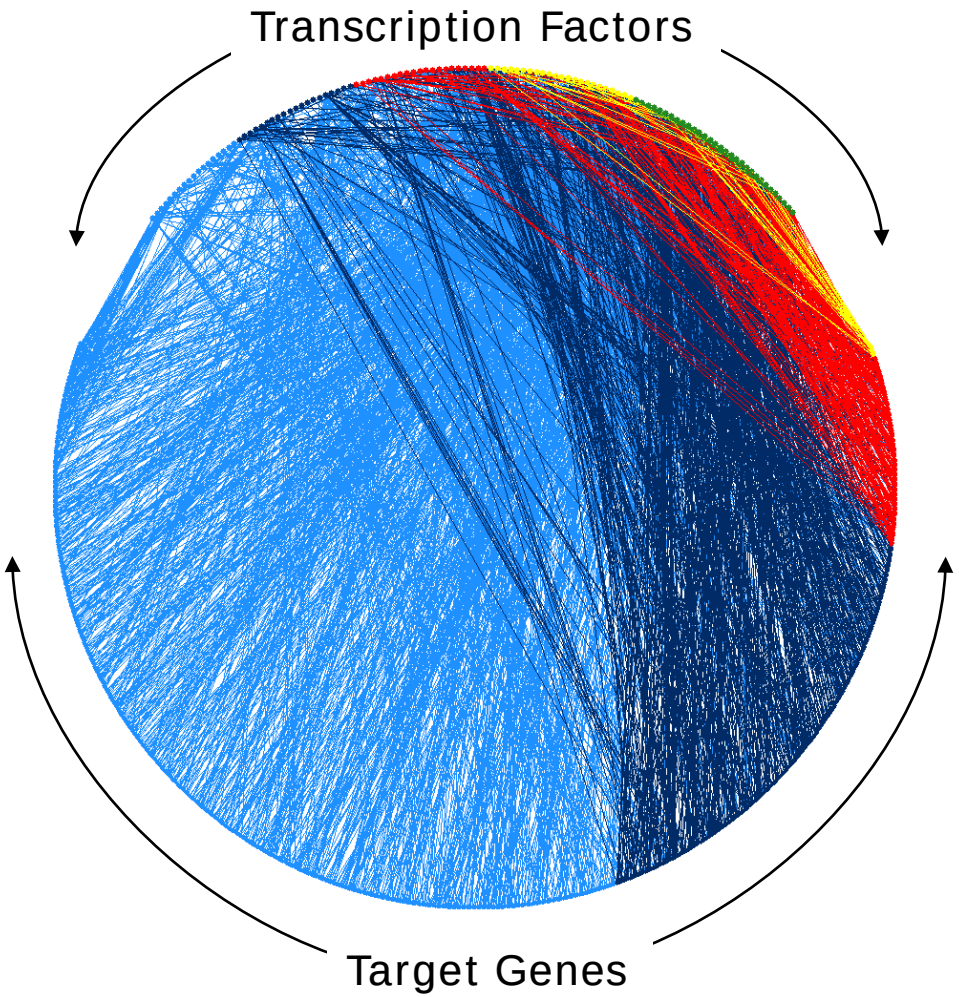
Party hubs function inside a certain functional module

Date hubs connect different functional modules at higher level

Han *et al*. Nature, 2004

Q5: Substructure in the interactome network are known to be very important to the network topology and function. Considering the condition aspect of the interactome, are all them equal?

Dynamic Yeast TF network



- Analysed network as a static entity
- But network is *dynamic*
  - Different sections of the network are active under different cellular conditions
- Integrate more gene expression data

[Luscombe et al, *Nature* (In press)]

Gene expression data

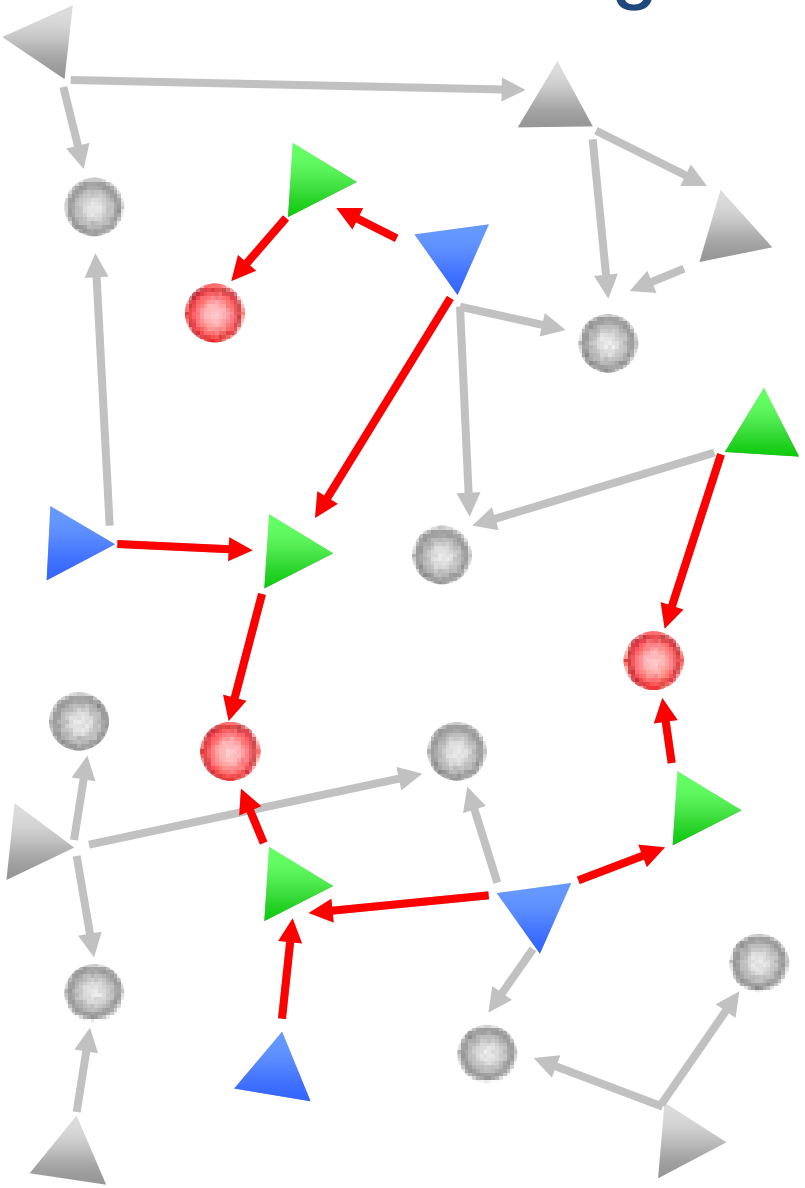
- Genes that are differentially expressed under five cellular conditions

Cellular condition	No. genes
Cell cycle	437
Sporulation	876
Diauxic shift	1,876
DNA damage	1,715
Stress response	1,385

- Assume these genes undergo transcription regulation

[Luscombe et al, *Nature*]

Backtracking to find active sub-network



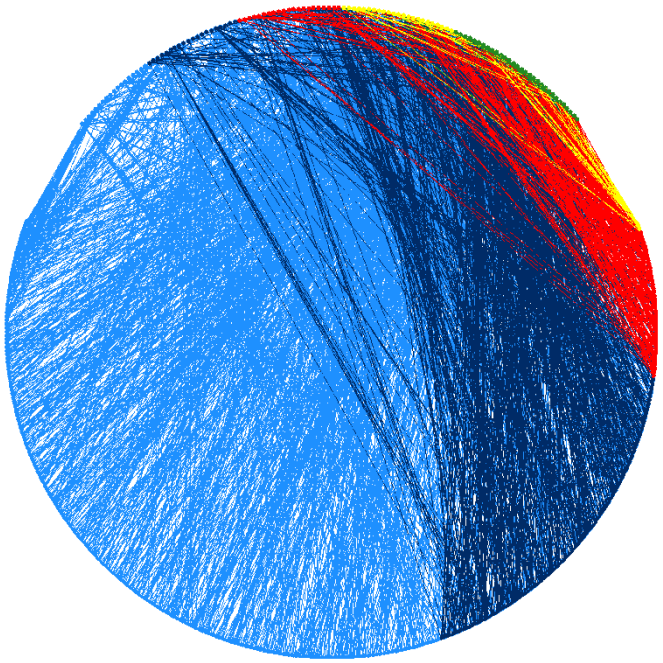
- Define differentially expressed genes
- Identify TFs that regulate these genes
- Identify further TFs that regulate these TFs

Active regulatory sub-network

[Luscombe et al, *Nature* (In press)]

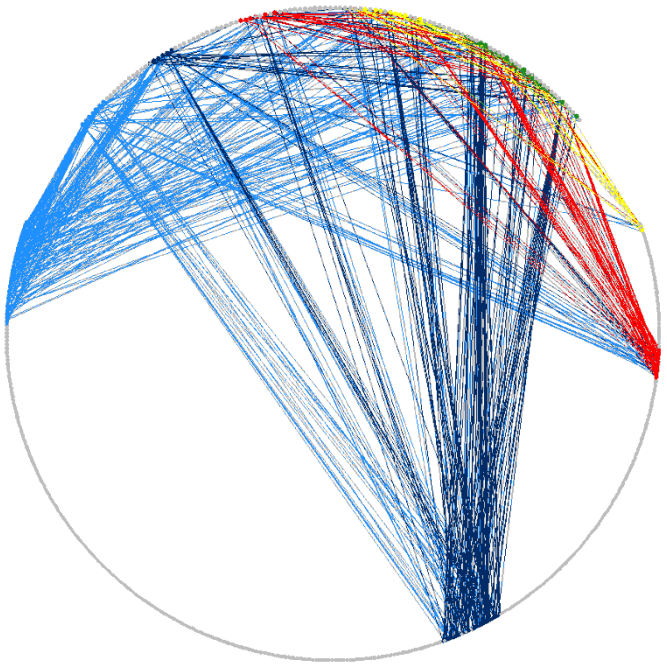
Network usage under cell cycle

complete network



- 142 TFs
- 3,420 genes
- 7,074 interactions

cell cycle sub-network

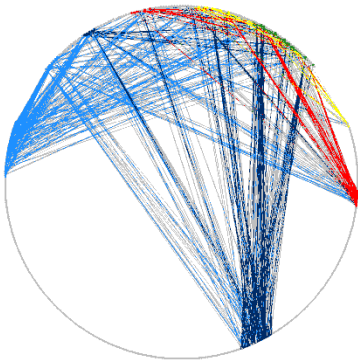


- 70 TFs
- 280 genes
- 550 interactions

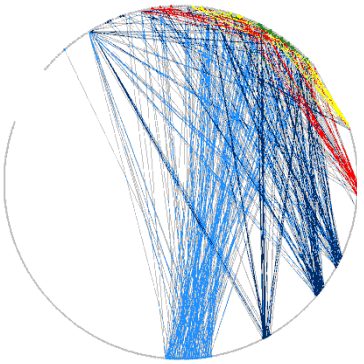
[Luscombe et al, *Nature* (In press)]

Network usage under different conditio

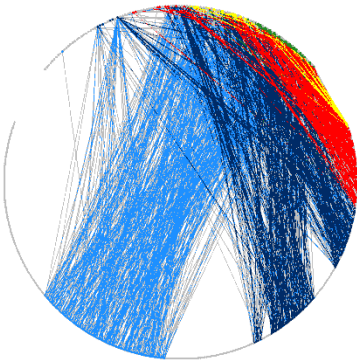
Cell cycle



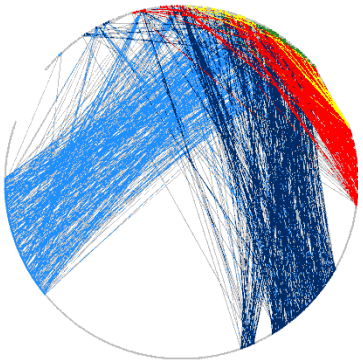
Sporulation



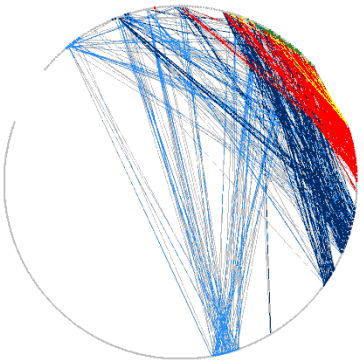
Diauxic shift



DNA damage



Stress



How do the networks change?

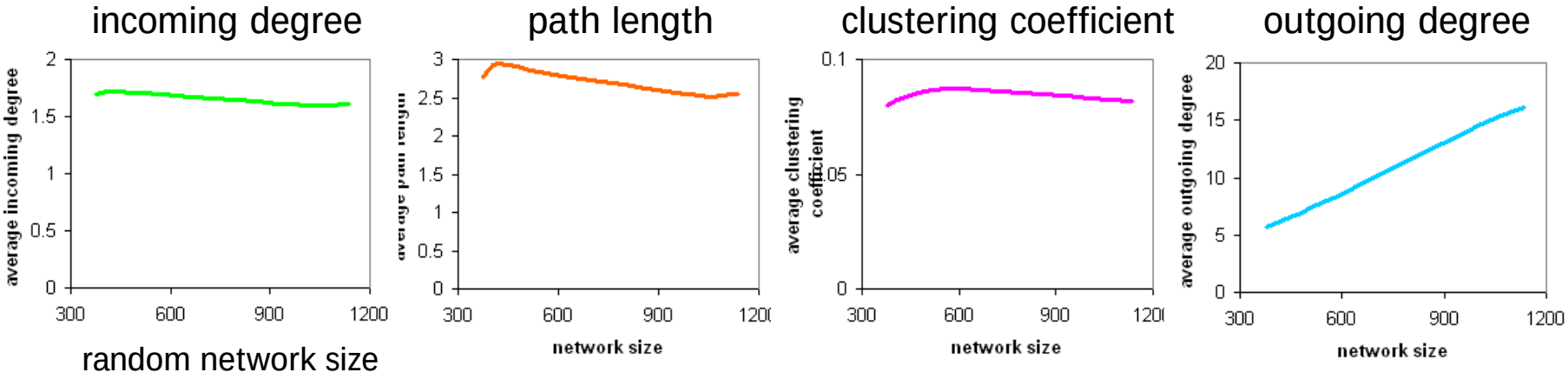
- topological measures
- network motifs

[Luscombe et al, *Nature* (In press)]

Our expectation

- Literature: Network topologies are perceived to be invariant
  - [Barabasi]
  - Scale-free, small-world, and clustered
  - Different molecular biological networks
  - Different genomes

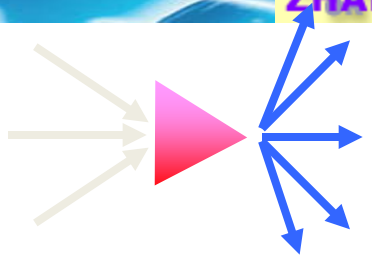
- Random expectation: Sample different size sub-networks from complete network and calculate topological measures

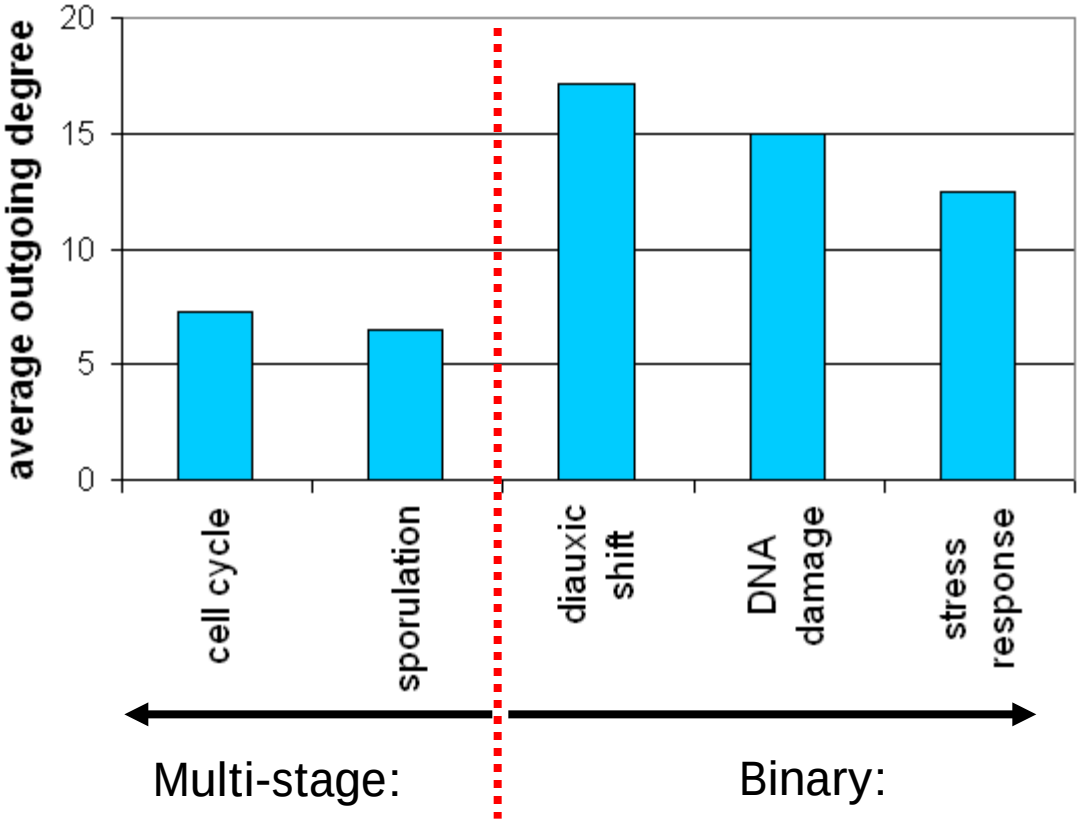


Measures should remain constant

[Luscombe et al, *Nature* (In press)]

Outgoing degree

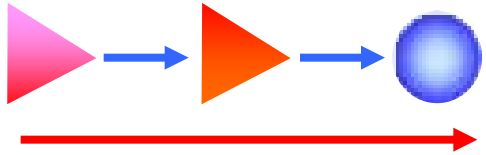


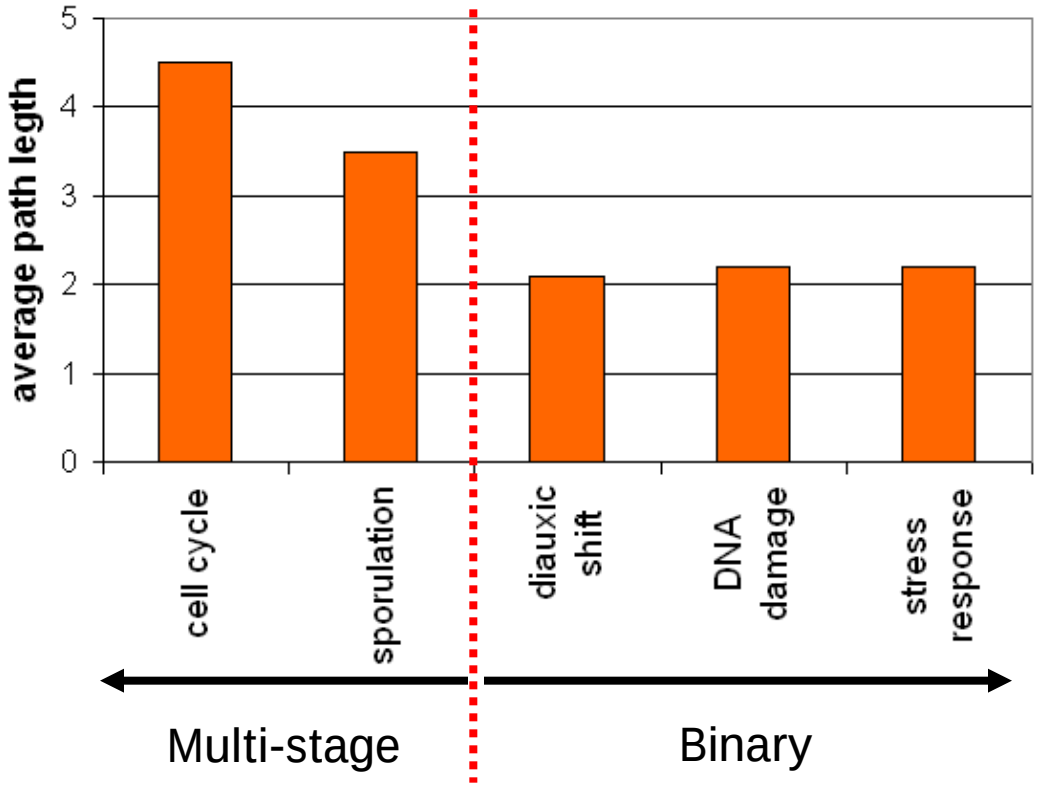


- "Binary conditions"
  →greater connectivity
- "Multi-stage conditions"
  →lower connectivity

[Luscombe et al, *Nature* (In press)]

Path length

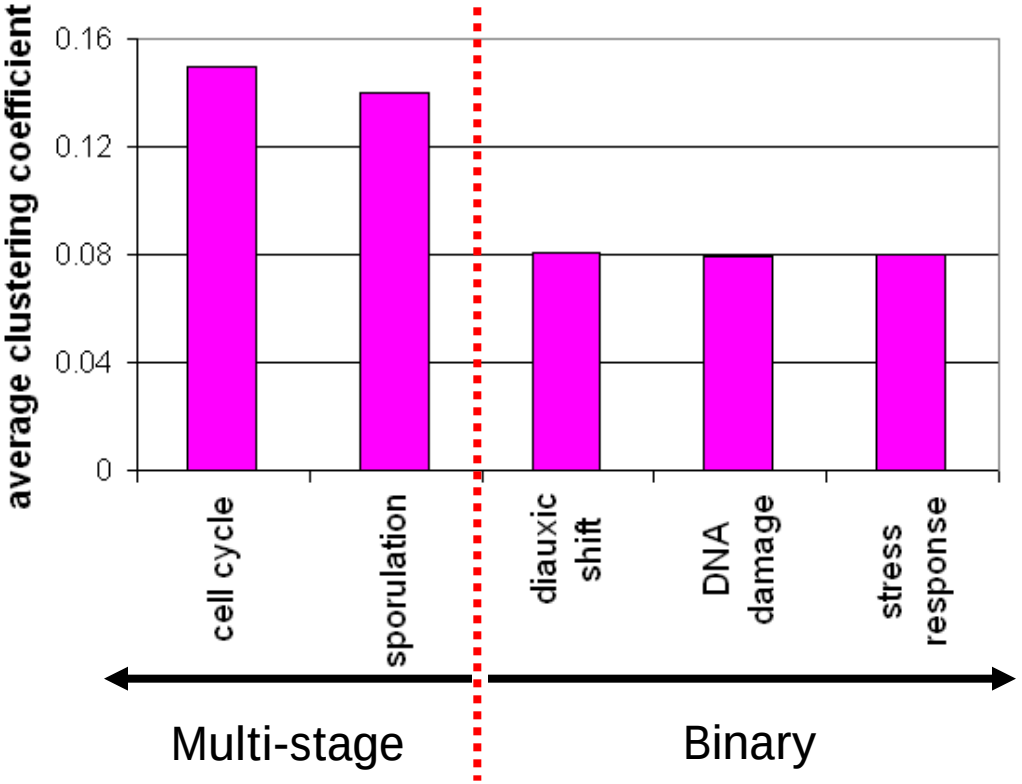




- "Binary conditions"
  - → shorter path-length
  - → "faster", direct action
- "Multi-stage" conditions
  - → longer path-length
  - → "slower", indirect action

[Luscombe et al, *Nature* (In press)]

Clustering coefficient



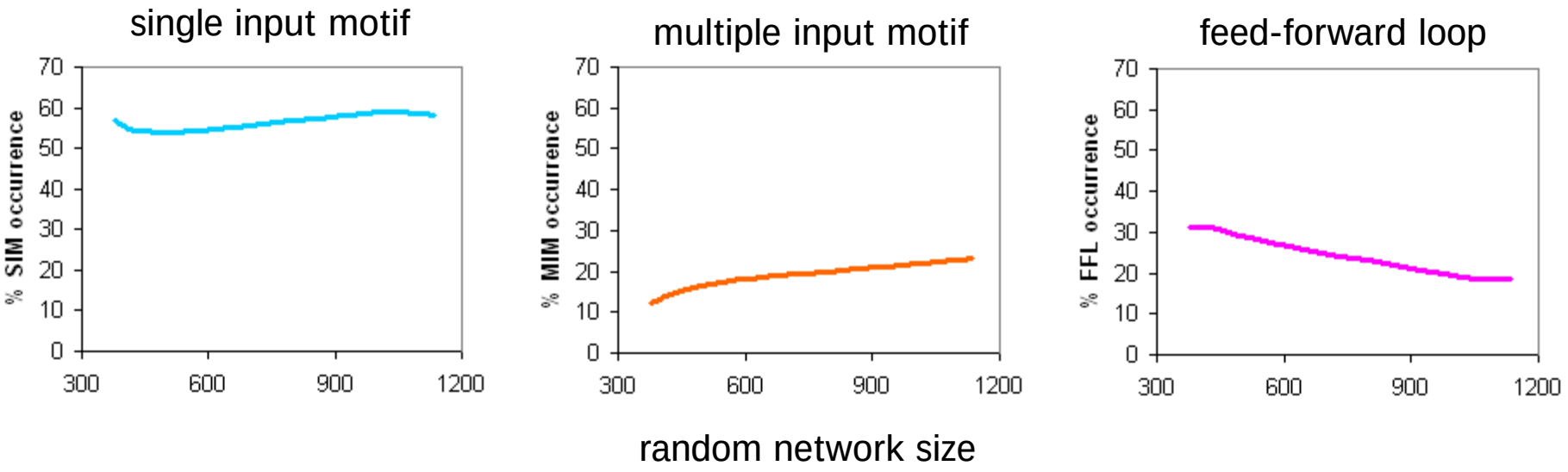
- "Binary conditions"
  → smaller coefficients
  → less TF-TF inter-regulation
- "Multi-stage conditions"
  → larger coefficients
  → more TF-TF inter-regulation

[Luscombe et al, *Nature* (In press)]

Our expectation

- Literature: motif usage is well conserved for regulatory networks across different organisms [Alon]

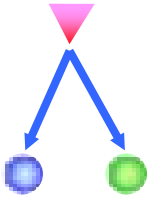
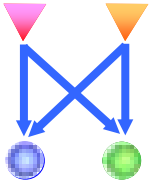
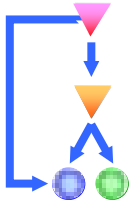
- Random expectation: sample sub-networks and calculate motif occurrence



Motif usage should remain constant

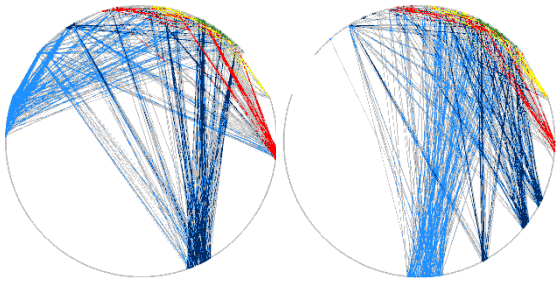
[Luscombe et al, *Nature* (In press)]

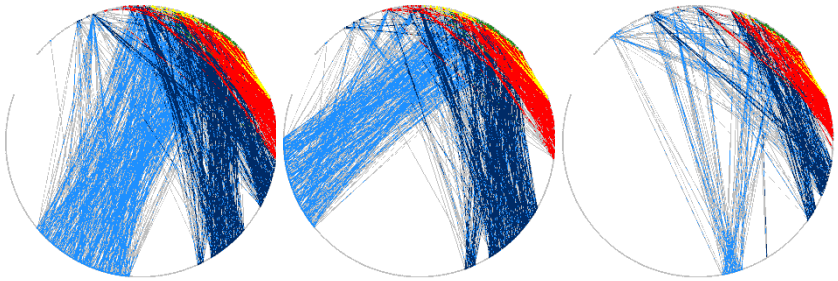
Network motifs

Motifs		Cell cycle	Sporulation	Diauxic shift	DNA damage	Stress response
SIM		32.0%	38.9%	57.4%	55.7%	59.1%
MIM		23.7%	16.6%	23.6%	27.3%	20.2%
FFL		44.3%	44.5%	19.0%	17.0%	20.7%

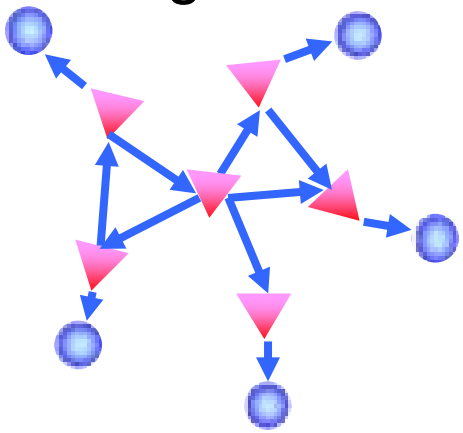
[Luscombe et al, *Nature* (In press)]

Summary of sub-network structures



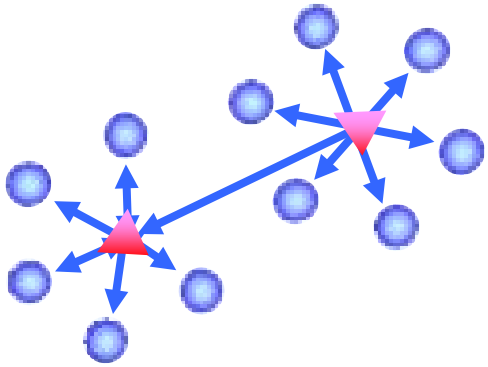


multi-stage conditions



- fewer target genes
- longer path lengths
- more inter-regulation between TFs

binary conditions



- more target genes
- shorter path lengths
- less inter-regulation between TFs

[Luscombe et al, *Nature* (In press)]

Q6: Aging and disease are known to be closely related.
Can we see this relationship in the interactome?

OPEN ACCESS Freely available online

PLoS COMPUTATIONAL BIOLOGY

Disease-Aging Network Reveals Significant Roles of Aging Genes in Connecting Genetic Diseases

Jiguang Wang[1,2], Shihua Zhang[1], Yong Wang[1], Luonan Chen[3,4]*, Xiang-Sun Zhang[1]*

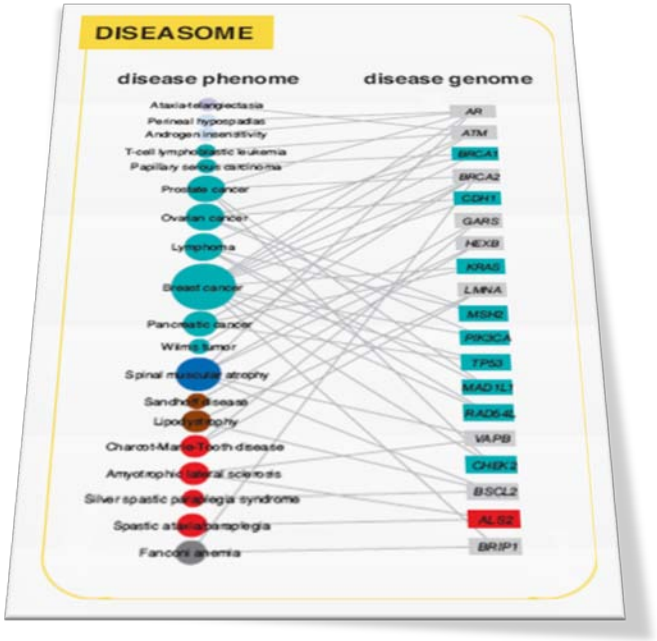
1 Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, China, **2** Graduate School of the Chinese Academy of Sciences, Beijing, China, **3** Institute of Systems Biology, Shanghai University, Shanghai, China, **4** Department of Electrical Engineering and Electronics, Osaka Sangyo University, Osaka, Japan

Abstract

One of the challenging problems in biology and medicine is exploring the underlying mechanisms of genetic diseases. Recent studies suggest that the relationship between genetic diseases and the aging process is important in understanding the molecular mechanisms of complex diseases. Although some intricate associations have been investigated for a long time, the studies are still in their early stages. In this paper, we construct a human disease-aging network to study the relationship among aging genes and genetic disease genes. Specifically, we integrate human protein-protein interactions (PPIs), disease-gene associations, aging-gene associations, and physiological system–based genetic disease classification information in a single graph-theoretic framework and find that (1) human disease genes are much closer to aging genes than expected by chance; and (2) diseases can be categorized into two types according to their relationships with aging. Type I diseases have their genes significantly close to aging genes, while type II diseases do not. Furthermore, we examine the topological characters of the disease-aging network from a systems perspective. Theoretical results reveal that the genes of type I diseases are in a central position of a PPI network while type II are not; (3) more importantly, we define an asymmetric closeness based on the PPI network to describe relationships between diseases, and find that aging genes make a significant contribution to associations among diseases, especially among type I diseases. In conclusion, the network-based study provides not only evidence for the intricate relationship between the aging process and genetic diseases, but also biological implications for prying into the nature of human diseases.

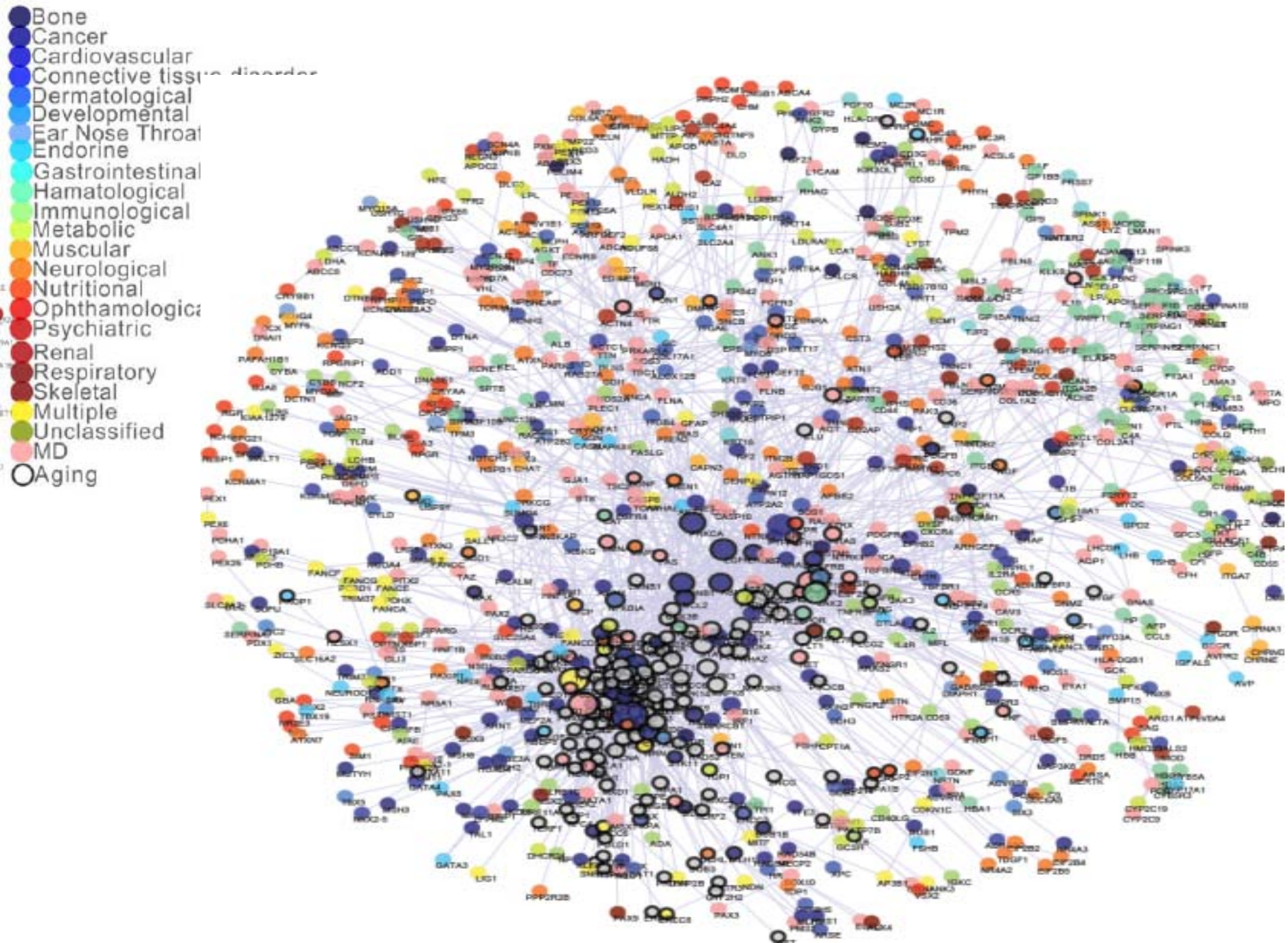
Aging

?



Disease

Association

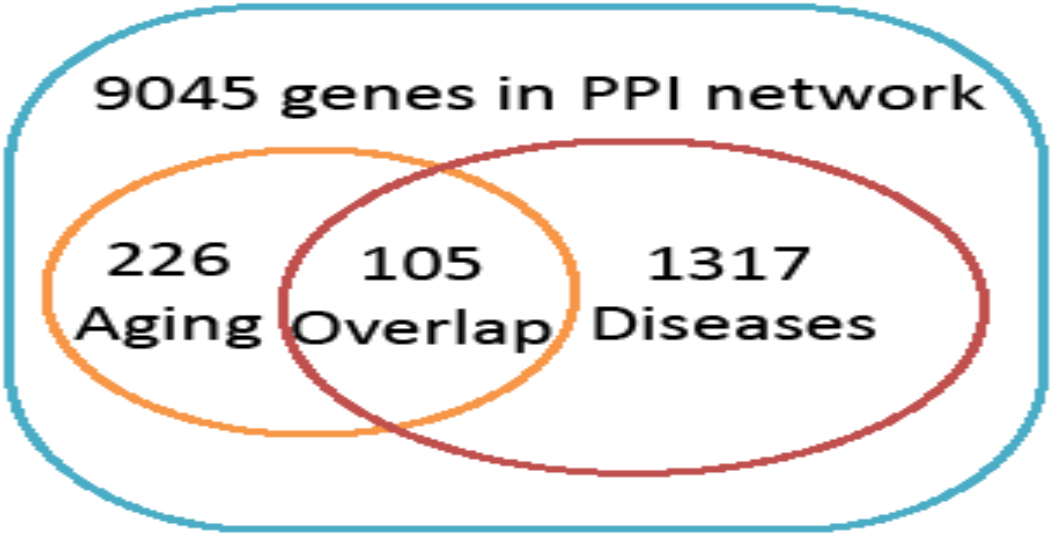

Bone
Cancer
Cardiovascular
Connective tissue disorder
Dermatological
Developmental
Ear Nose Throat
Endorine
Gastrointestinal
Hamatological
Immunological
Metabolic
Muscular
Neurological
Nutritional
Ophthamological
Psychiatric
Renal
Respiratory
Skeletal
Multiple
Unclassified
MD
Aging

Results

(1) Human disease genes are much closer to aging genes than expected by chance.

(2) Diseases can be categorized into two types according to their relationships with aging. Type I diseases have their genes significantly close to aging genes, while type II diseases do not.

(3) Aging genes make a significant contribution to associations among diseases.



Degree of aging genes	Average degree	Disease genes		
		Observed	Random	P-value
<20	9.38	2.51	1.99	7.3e-8
20-50	33.33	8.53	7.05	7.8e-7
50-100	69.27	17.49	14.52	1.9e-8
>100	139.81	33.86	28.82	1.4e-7

Results

(1) Human disease genes are much closer to aging genes than expected by chance.

(2) Diseases can be categorized into two types according to their relationships with aging. Type I diseases have their genes significantly close to aging genes, while type II diseases do not.

(3) aging genes make a significant contribution to associations among diseases.

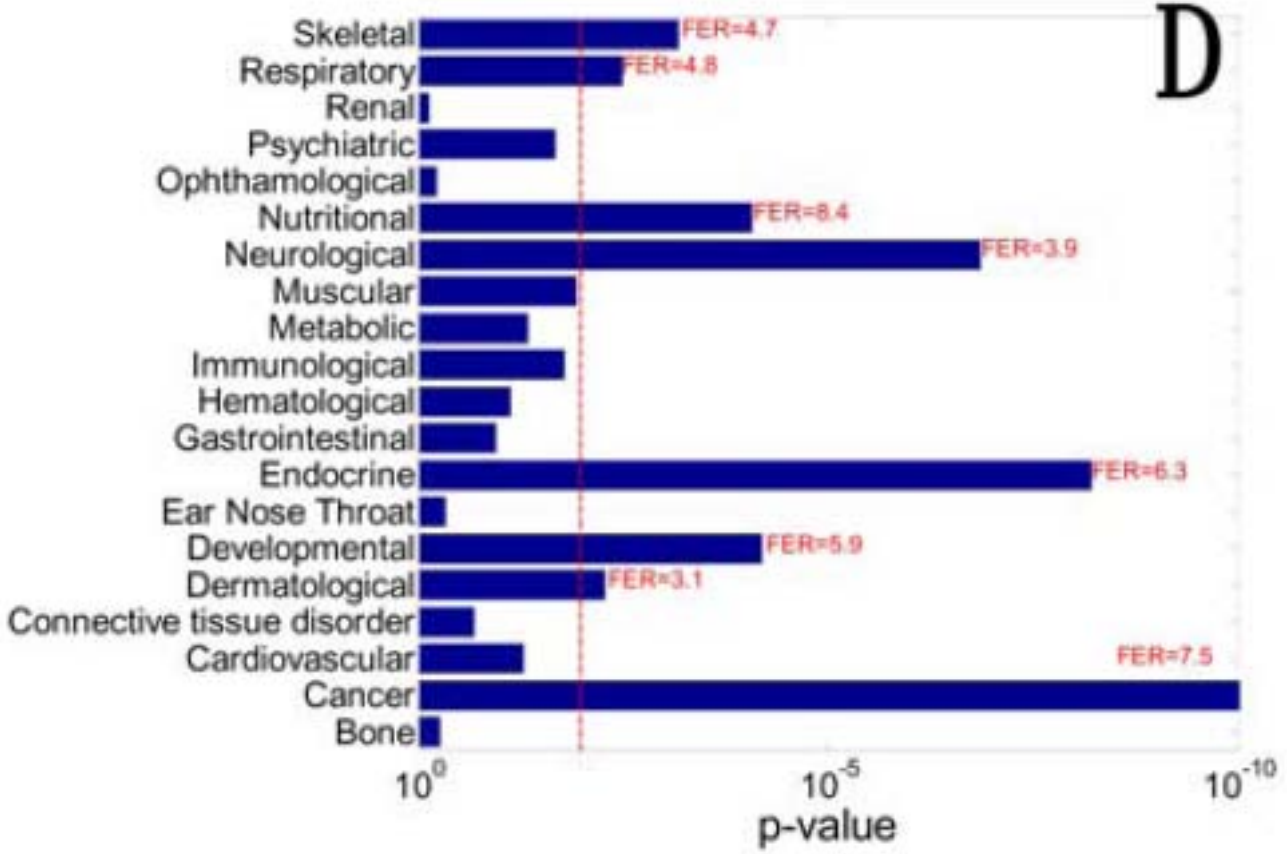

D
Skeletal
Respiratory
Renal
Psychiatric
Ophthamological
Nutritional
Neurological
Muscular
Metabolic
Immunological
Hematological
Gastrointestinal
Endocrine
Ear Nose Throat
Developmental
Dermatological
Connective tissue disorder
Cardiovascular
Cancer
Bone
FER=4.7
FER=4.8
FER=8.4
FER=3.9
FER=6.3
FER=5.9
FER=3.1
FER=7.5
10^0
10^-5
10^-10
p-value

Table 2. Different GOA enrichments of ARD and NARD.

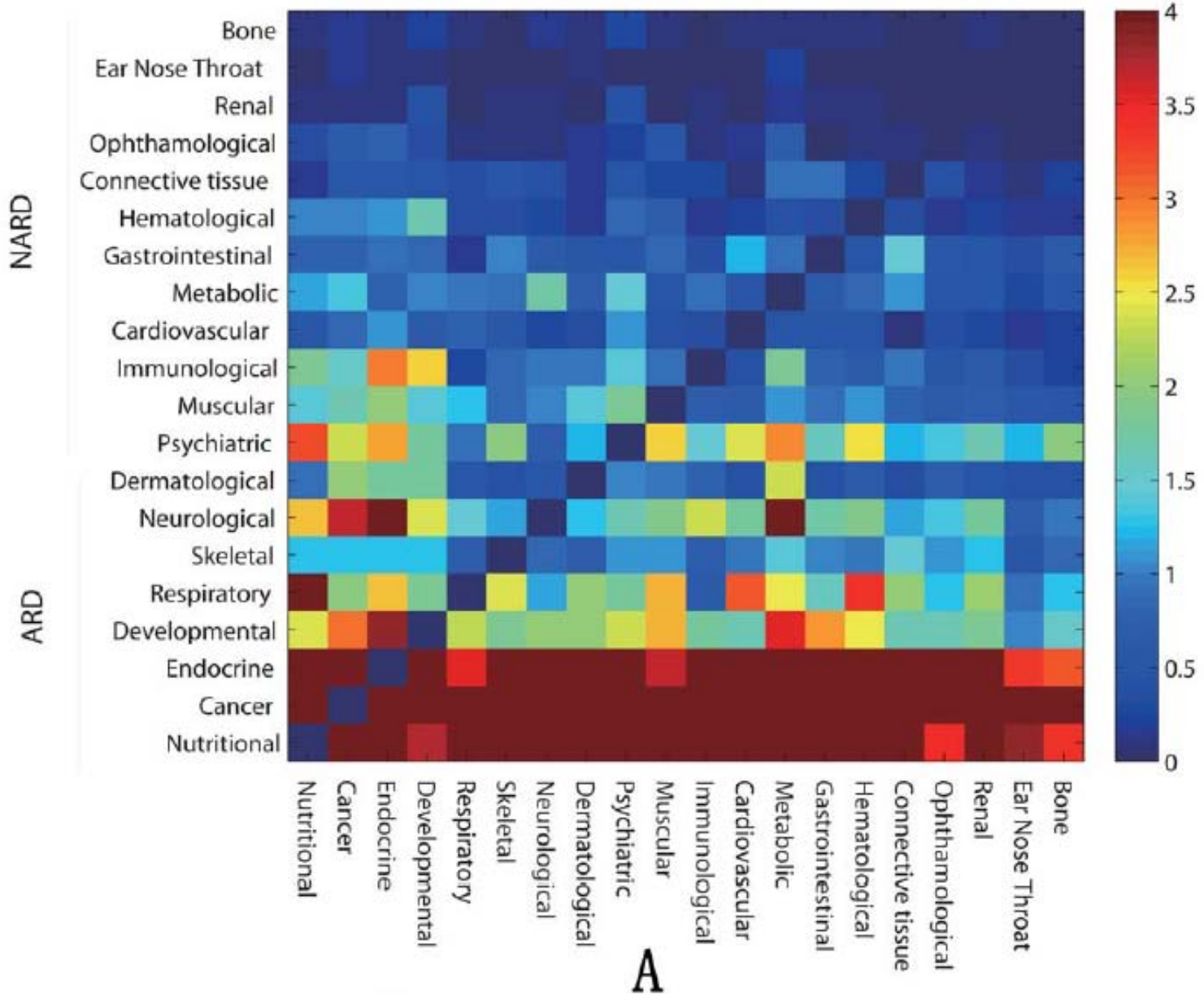
GO-ID	ARD		NARD		Description
	p-value	#Genes	p-value	#Genes	
3676	1.4e-4	156	1.1e-10(under)	68	nucleic acid binding
5634	3.2e-13	193	2.2e-7(under)	79	nucleus
6139	5.0e-19	194	3.7e-03(under)	113	nucleobase, nucleoside, nucleotide and nucleic acid metabolic proc
5622	1.1e-9	411	>0.01	391	intracellular
16301	2.4e-8	63	>0.01	44	oxidoreductase activity
30528	5.3e-15	112	>0.01	49	transcription regulator activity
43170	3.4e-11	313	>0.01	295	macromolecule metabolic process
3824	>0.01	206	1.6e-8	282	catalytic activity
5478	>0.01	58	3.9e-10	101	transporter activity
9055	>0.01	12	8.3e-7	56	catabolic process
9056	>0.01	29	2.5e-5	85	biosynthetic process
9405	>0.01	2	7.6e-7	20	cell surface
9929	>0.01	11	2.9e-7	60	ion transmembrane transporter activity
15075	>0.01	36	8.5e-6	37	channel activity
5941	>0.01	1	4.6e-4	6	unlocalized protein complex
16740	>0.01	76	1.2e-5	129	hydrolase activity
16787	>0.01	88	1.9e-5	20	lyase activity
16874	>0.01	13	1.4e-7	113	cell differentiation

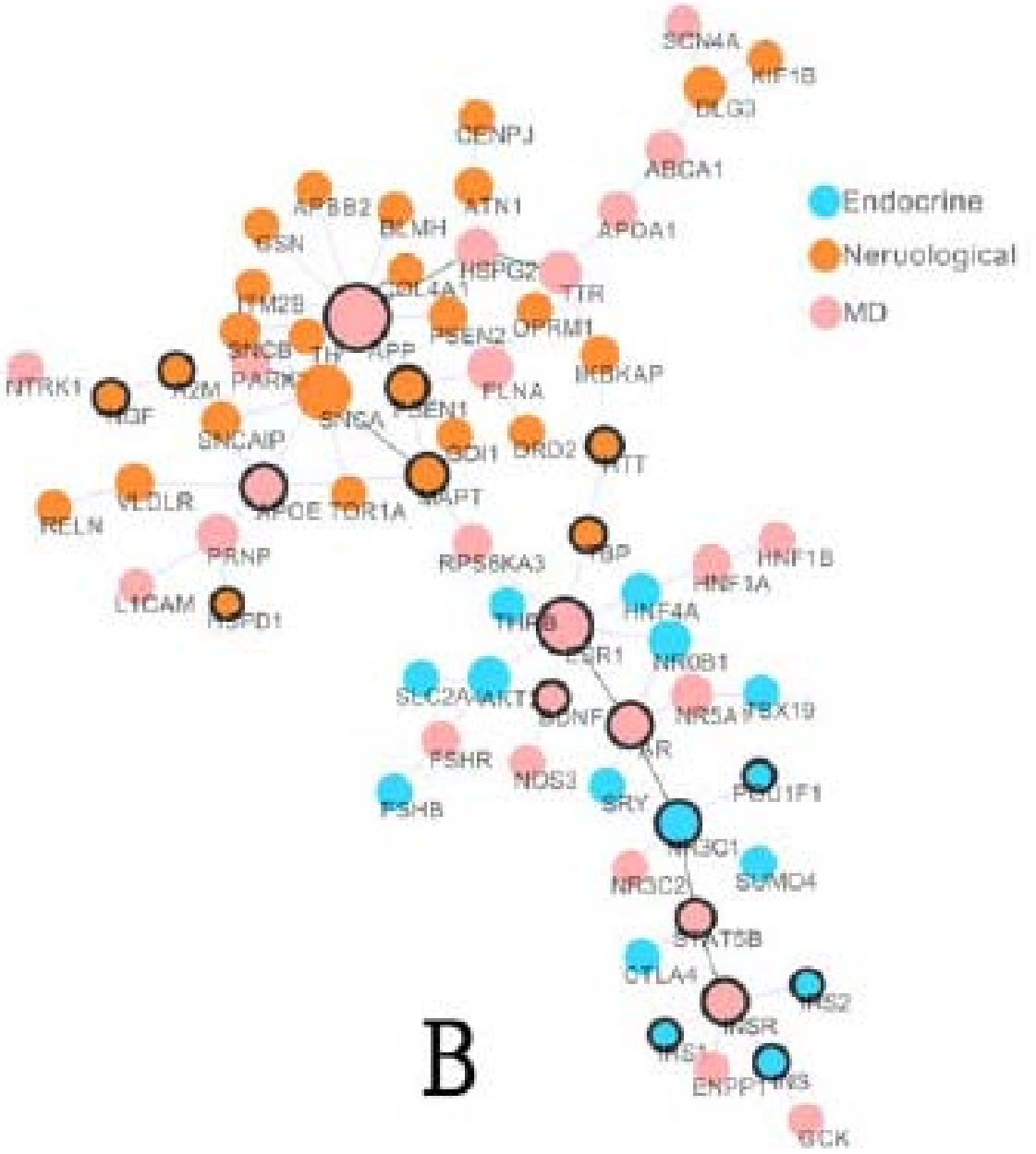
Results

(1) Human disease genes are much closer to aging genes than expected by chance.

(2) Diseases can be categorized into two types according to their relationships with aging. Type I diseases have their genes significantly close to aging genes, while type II diseases do not.

(3) aging genes make a significant contribution to associations among diseases.

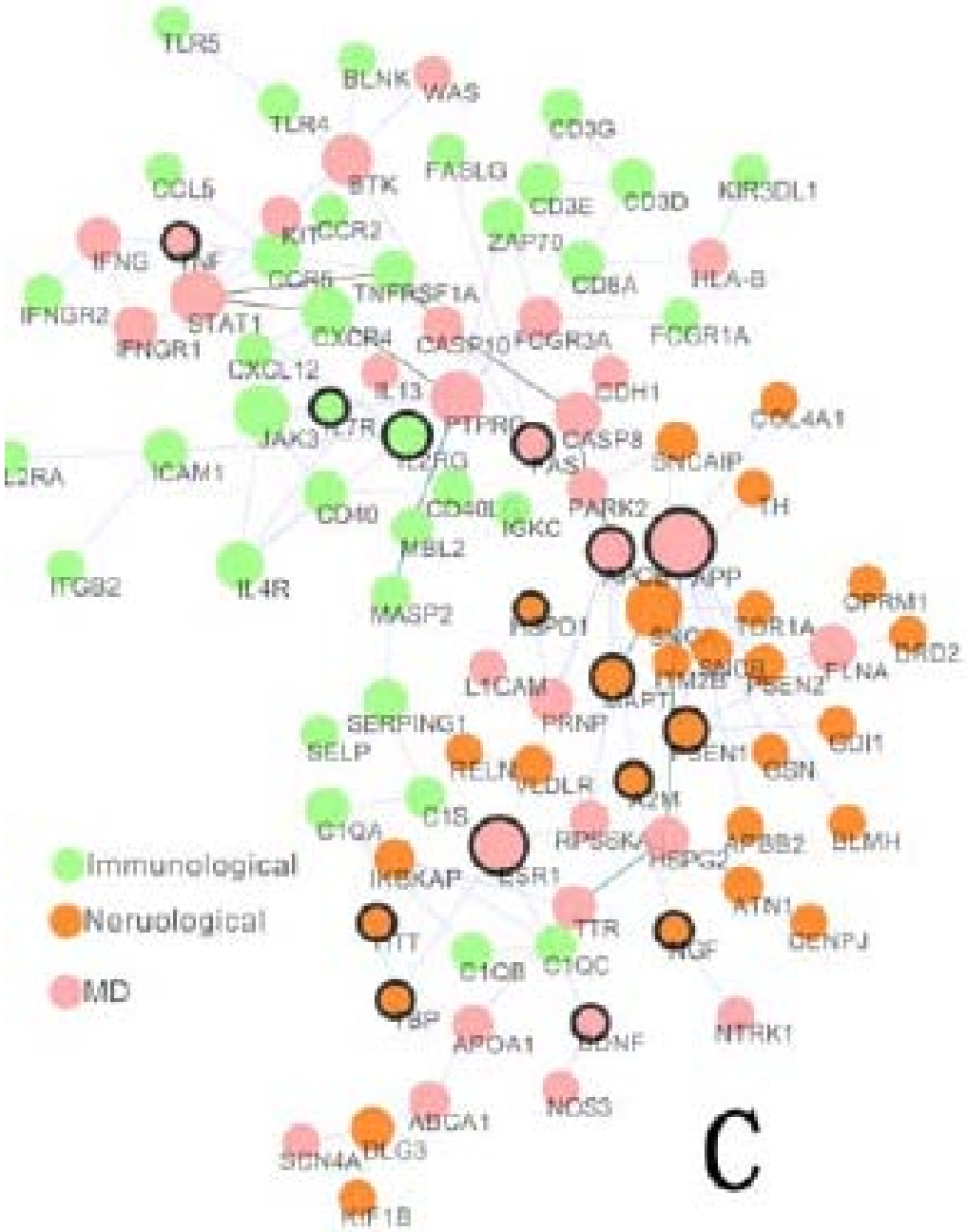

NARD
ARD
Bone
Ear Nose Throat
Renal
Ophthamological
Connective tissue
Hematological
Gastrointestinal
Metabolic
Cardiovascular
Immunological
Muscular
Psychiatric
Dermatological
Neurological
Skeletal
Respiratory
Developmental
Endocrine
Cancer
Nutritional
4
3.5
3
2.5
2
1.5
1
0.5
0
A


SCN4A
KIF1B
DLG3
CENPJ
ABCA1
APBB2
ATN1
APOA1
BCMH
GSN
HSPG2
TTR
PSEN2
OPRM1
IKBKAP
PLNA
SNCB
APP
PARK2
SNCA
PSEN1
SNCAIP
DRD2
HTT
RELN
VLDLR
APOE
TOR1A
APT
PRNP
L1CAM
RSPD1
RPS6KA3
HNF1B
HNF1A
HNF4A
ESR1
NR0B1
SLC2A
AKT1
BDNF
NR5A
PBX1
AR
FSHR
NDS3
SRY
FSHB
SUMO4
NR3C2
STAT5B
CTLA4
IRS2
INSR
ENPP1
GCK
Endocrine
Neruological
MD

内分泌

神经

B


TLR5
BLNK
WAS
TLR4
CD3G
FASLG
CCL5
BTK
CD3E
CD3D
KIR3DL1
IFNG
KIT
CCR2
ZAP70
TNF
CCR5
TNFRSF1A
CD8A
HLA-B
IFNGR2
STAT1
CXCR4
CASP10
FCGR3A
FCGR1A
IFNGR1
CXCL12
IL13
CDH1
JAK3
IL7R
PTPRC
CASP8
CXCL4A1
L2RA
ICAM1
IL2RG
FAS
SNCAIP
CD40
CD40L
PARK2
TH
MSKC
MBL2
ITGB2
IL4R
APP
MASP2
OPRM1
DRD1
DRD2
L1CAM
FLNA
PSEN2
SERPING1
PRNP
SELP
CDI1
RELN
LDLR
A2M
GSN
C1QA
C1S
RPS6KA
APBB2
BLMH
IRAK4
ESR1
HSPG2
ATN1
HTT
TTR
NGF
CENPJ
C1QB
C1QC
TSP
APOA1
BDNF
NTRK1
NOS3
ABCA1
SCN4A
DLG3
RIT1
Immunological
Neurological
MD
C

免疫

神经

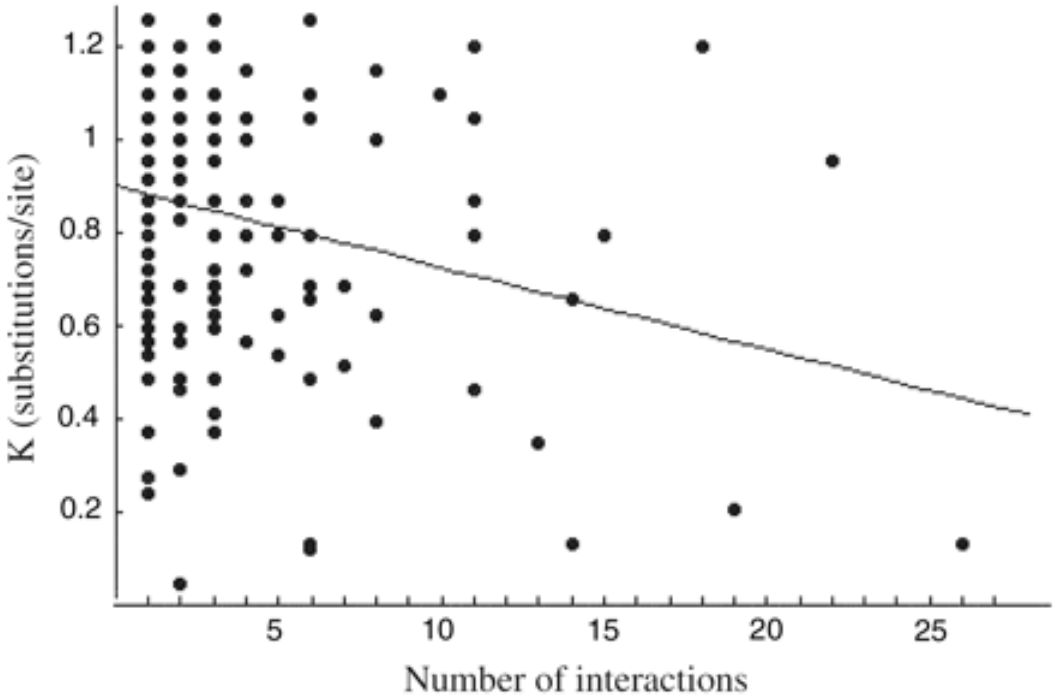
Q7: Regarding to evolution principles, is the subnetwork and the whole interactome the same?

TF subnetwork Vs whole network

- We study evolutionary principles in the network of an important subset of proteins, the transcription factors (TFs).
- TFs are important regulators of cellular processes at the transcriptional level.
- The interactions and coordinated actions of multiple TFs in the TF network provide a primary mechanism for achieving fine-tuned transcriptional control in eukaryotes.

Well-known result

Hubs in the *S. cerevisiae* protein-protein interaction network tend to evolve more slowly than non-hubs



A protein's number of interaction partners exerts some influence on its evolutionary rate, most likely due to increased structural co-evolutionary constraints imposed by protein-protein interaction (negative selection) .

Fraser HB, Hirsh AE, Steinmetz LM, Scharfe C, Feldman MW (2002) Evolutionary rate in the protein interaction network. Science 296: 750-752.

Surprising findings

- hubs in the yeast TF network tend to evolve more quickly than non-hubs
- This result holds for all four major types of TF hubs:

1. Interaction hubs that interact with many other TFs
2. Regulatory in-degree hubs that are regulated by many TFs
3. Regulatory out-degree hubs that regulate many TFs
4. co-regulatory hubs that jointly regulate target genes (TGs) with many other TFs.

TF networks

- We collected 174 yeast TFs and assembled the whole-genome TF network based on three types of associations:
- protein-protein interactions among TFs (forming the TF interactome)
- transcriptional regulatory relationships among TFs (forming the TF transcriptional regulatory network)
- joint regulation of target genes among TFs (forming the TF co-regulatory network)

Evolutionary rate

- Evolutionary rate was measured as the K_A/K_S ratio calculated over alignments between the coding sequences of *S. cerevisiae* and their orthologs in *S. paradoxus* (the closest related yeast with a sequenced genome).
- K_A/K_S is the ratio of the rate of non-synonymous substitutions (K_A) to the rate of synonymous substitutions (K_S), and serves as an approximate measure of the strength of sequence selection acting on a protein (factoring out mutational background and translational selection).
- Smaller K_A/K_S values are associated with heightened purifying selection (reduced evolutionary rate), while larger values are associated with neutral or adaptive evolution (increased evolutionary rate).

同义与非同义的核苷酸替代

- 1. 同义替代：编码区的DNA序列，核苷酸的改变不改变编码的氨基酸的内容
- 2.非同义替代：核苷酸改变，从而改变编码氨基酸的内容

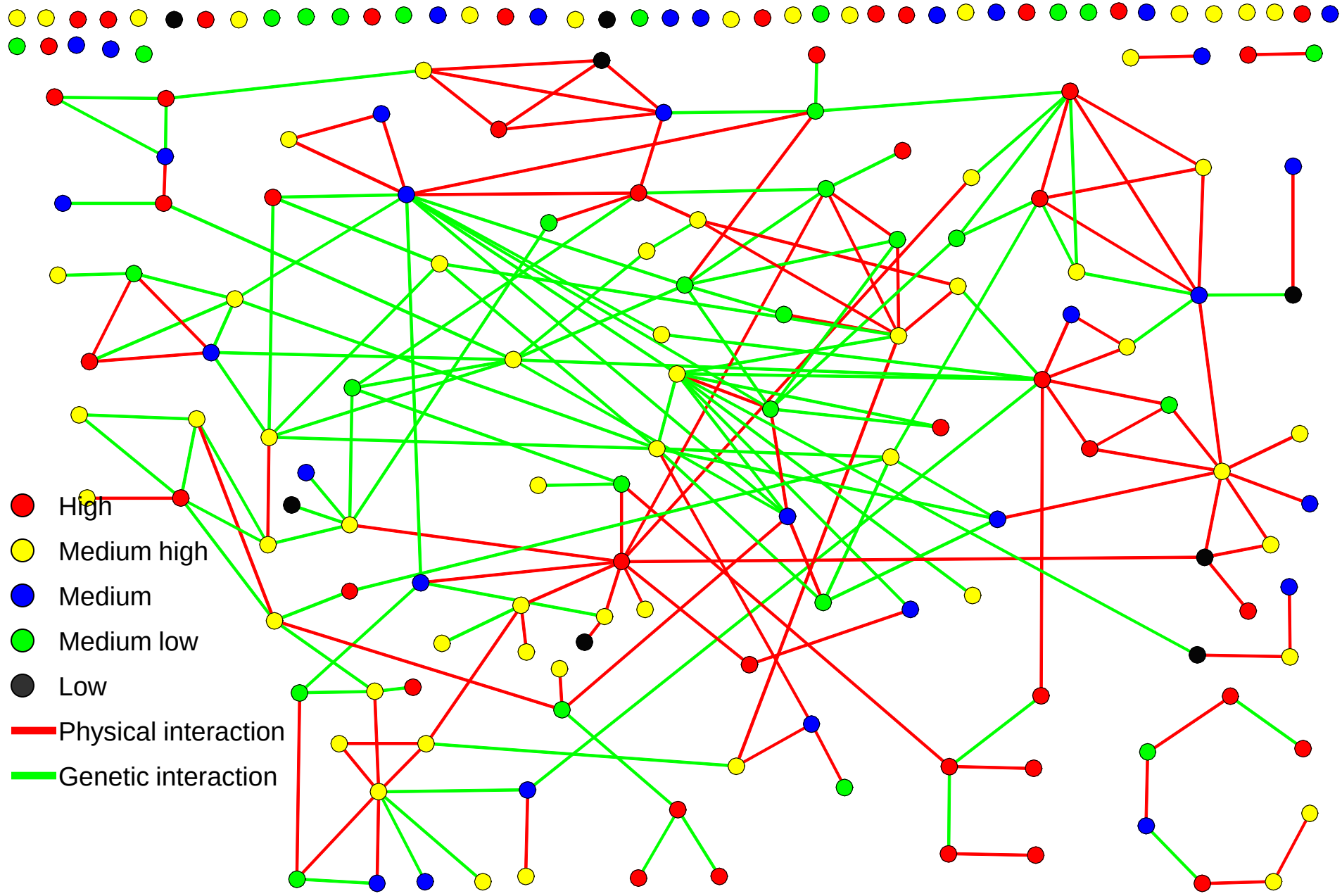
Ka/Ks：计算及含义

- 1. Ka：每个非同义位点的非同义替代数目
- 2. Ks：每个同义位点的同义替代数目
- 3. 一般计算公式：考虑序列上所有可能的同义位点(S)和非同义位点(N)，通过双序列比对发现存在的同义位点(S_d)和非同义位点(N_d)，存在：

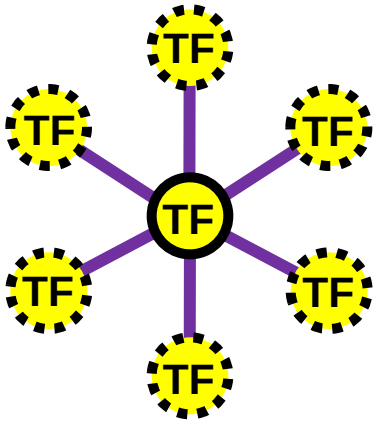
$$Ka/Ks = \frac{\frac{N_d}{N}}{\frac{S_d}{S}}$$

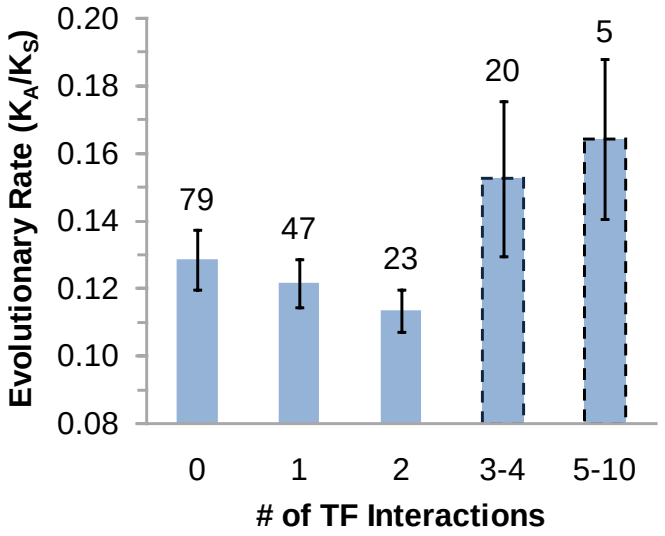
Ka/Ks：计算及含义 (2)

- 1. Ka/Ks ~ 1: 中性进化；
- 2. ka/Ks << 1: 阴性选择，净化选择；
- 3. ka/Ks >> 1: 阳性选择，适应性进化
- 4. 多数基因为中性进化，约1%的基因受到阳性选择->决定物种形成、新功能的产生。
- 5. PAML, MEGA等工具：计算Ka/Ks及统计显著性

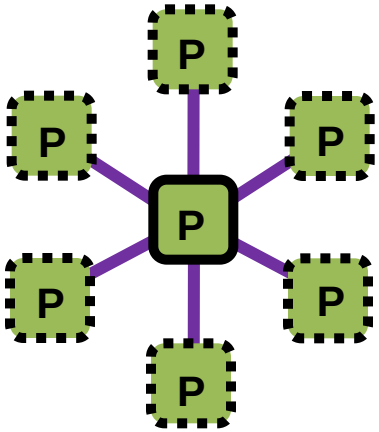

High
Medium high
Medium
Medium low
Low
Physical interaction
Genetic interaction

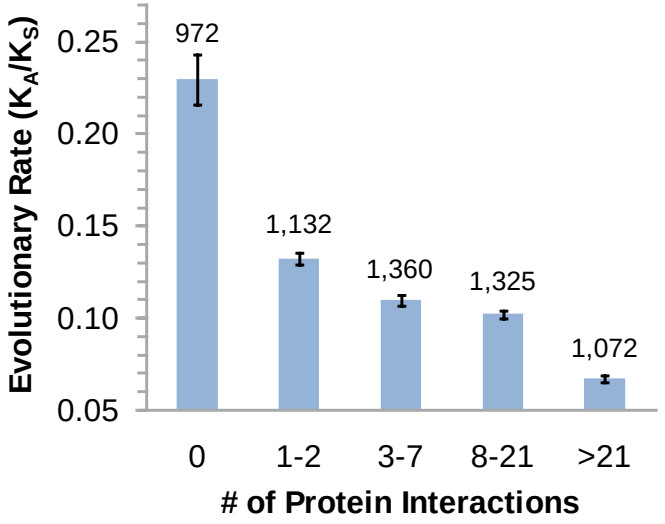
(a)


TF
TF
TF
TF
TF
TF
TF


Evolutionary Rate (K_A/K_S)
0.20
0.18
0.16
0.14
0.12
0.10
0.08
79
47
23
20
5
0
1
2
3-4
5-10
# of TF Interactions

(b)

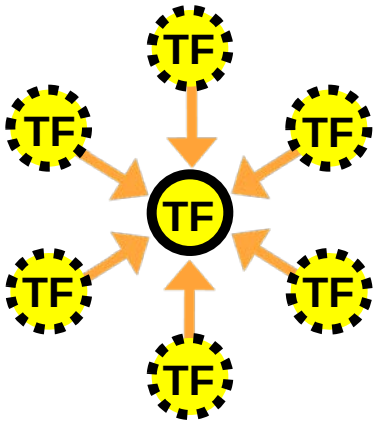

P
P
P
P
P
P
P

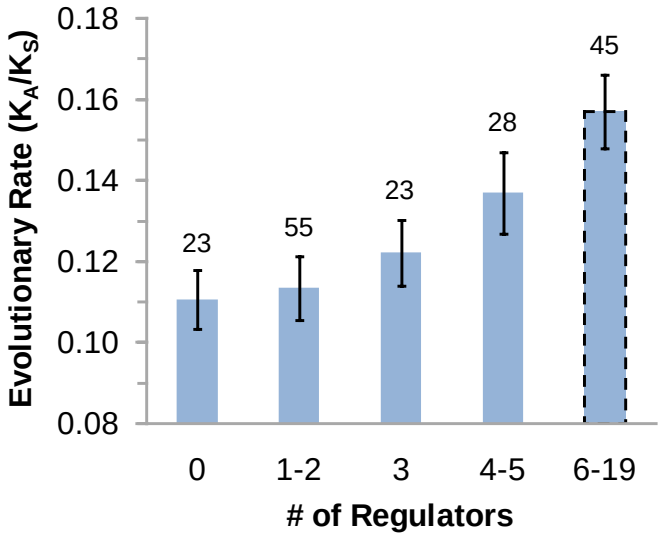

Evolutionary Rate (K_A/K_S)
0.25
0.20
0.15
0.10
0.05
972
1,132
1,360
1,325
1,072
0
1-2
3-7
8-21
>21
# of Protein Interactions

TF interaction hubs evolve fast

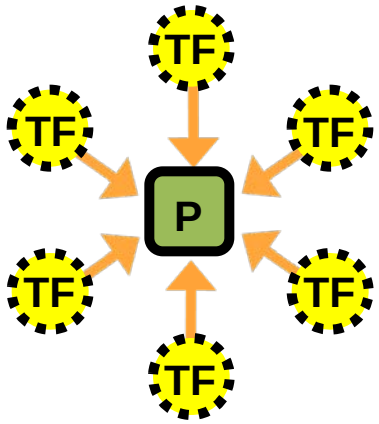
- The evolutionary rate of TF hubs is significantly greater on average than the evolutionary rate of TF non-hubs ($p = 0.04$).
- The mean of these sampled correlations between protein evolutionary rate and generic protein-protein interactions is significantly different from the observed correlation between TF evolutionary rate and TF-TF interactions ($p < 1.0 \times 10^{-6}$).
- We conclude that TF-TF interactions and generic protein-protein interactions evolve in very different ways: hubs in the protein interactome tend to evolve more slowly than non-hubs, whereas hubs in the TF interactome tend to evolve more quickly than non-hubs.

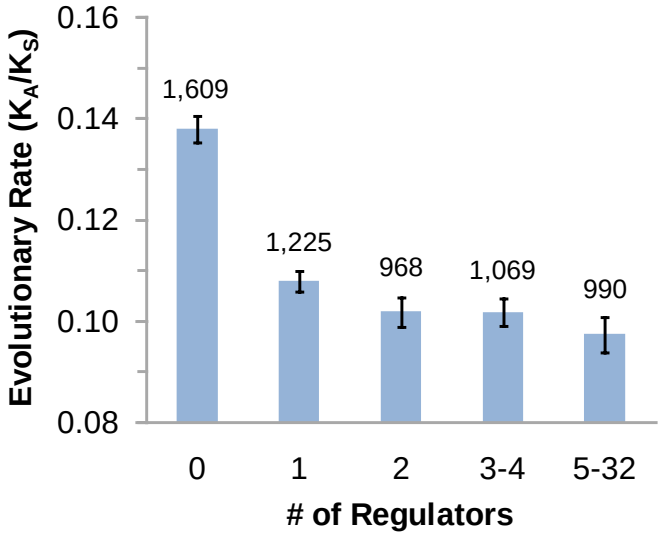
(a)


TF
TF
TF
TF
TF
TF
TF

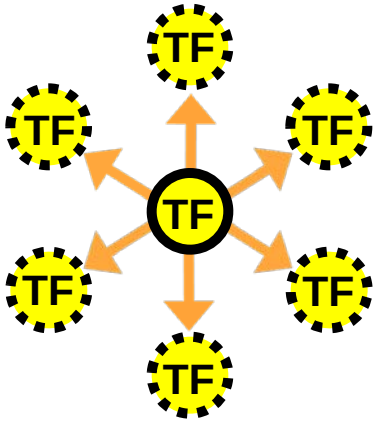

Evolutionary Rate (K_A/K_S)
0.18
0.16
0.14
0.12
0.10
0.08
23
55
23
28
45
0
1-2
3
4-5
6-19
# of Regulators

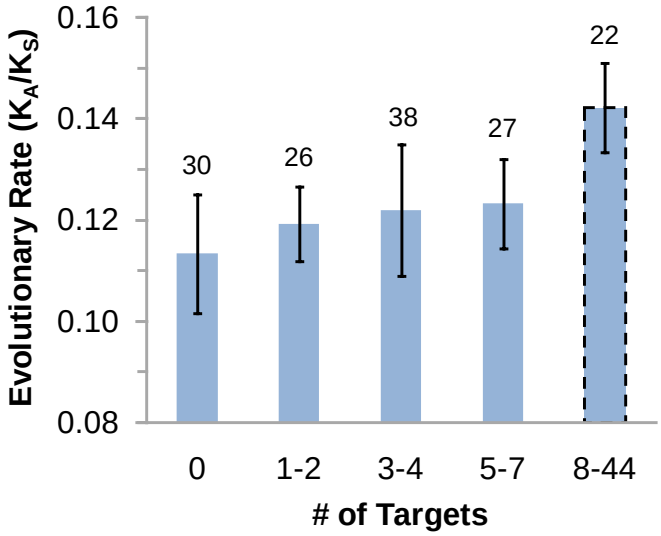
(b)


TF
TF
TF
P
TF
TF
TF

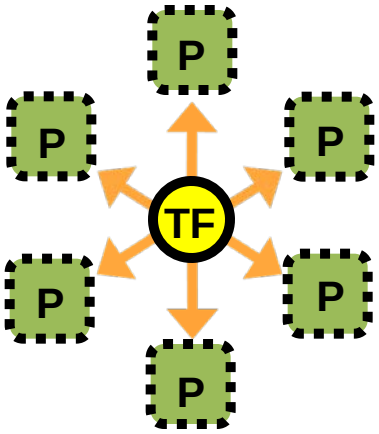

Evolutionary Rate (K_A/K_S)
0.16
0.14
0.12
0.10
0.08
1,609
1,225
968
1,069
990
0
1
2
3-4
5-32
# of Regulators

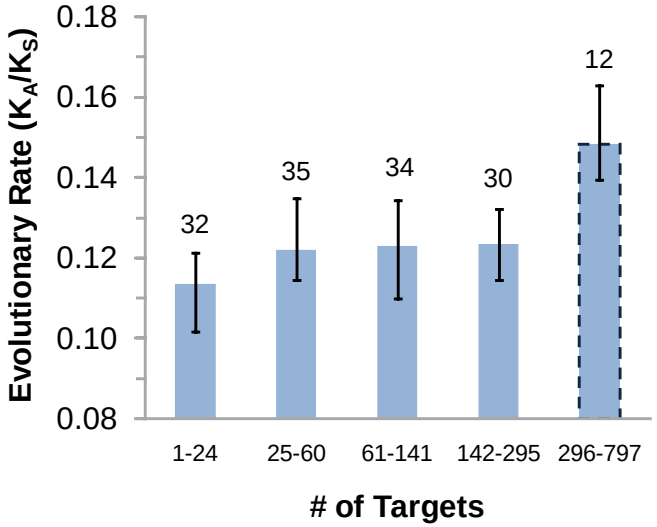
(a)

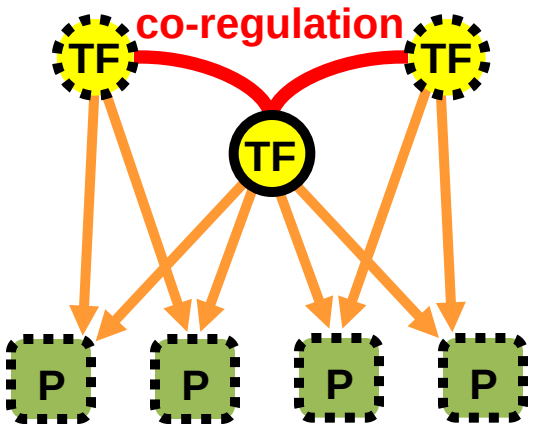

TF
TF
TF
TF
TF
TF
TF

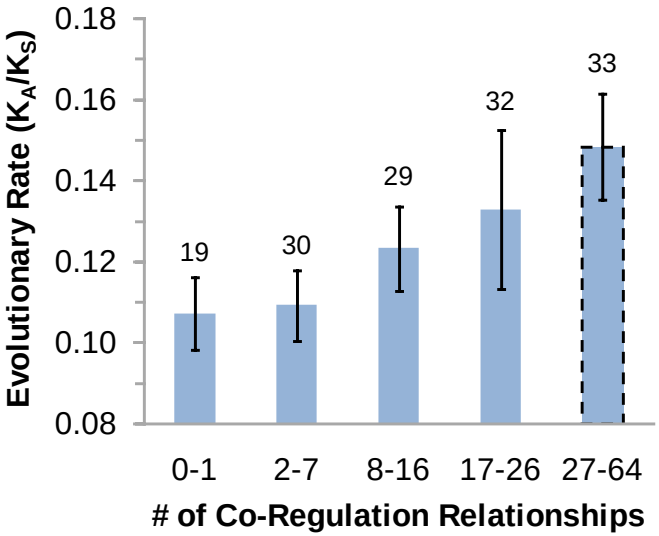

Evolutionary Rate (K_A/K_S)
0.16
0.14
0.12
0.10
0.08
30
26
38
27
22
0
1-2
3-4
5-7
8-44
# of Targets

(b)


P
P
P
P
P
P
TF

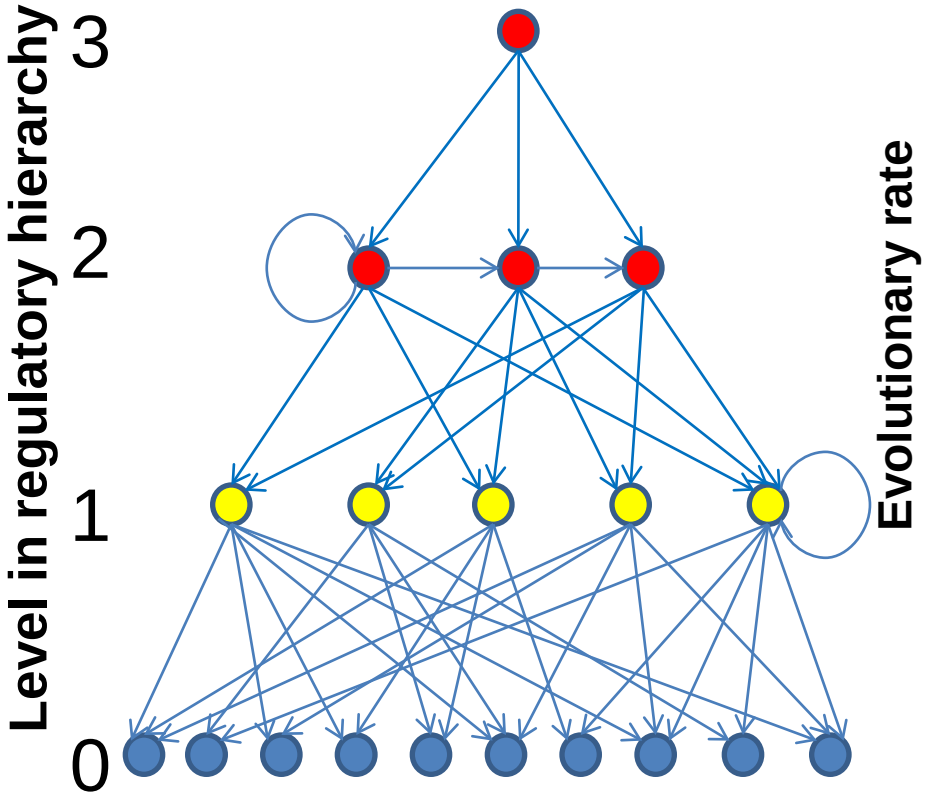

Evolutionary Rate (K_A/K_S)
0.18
0.16
0.14
0.12
0.10
0.08
32
35
34
30
12
1-24
25-60
61-141
142-295
296-797
# of Targets

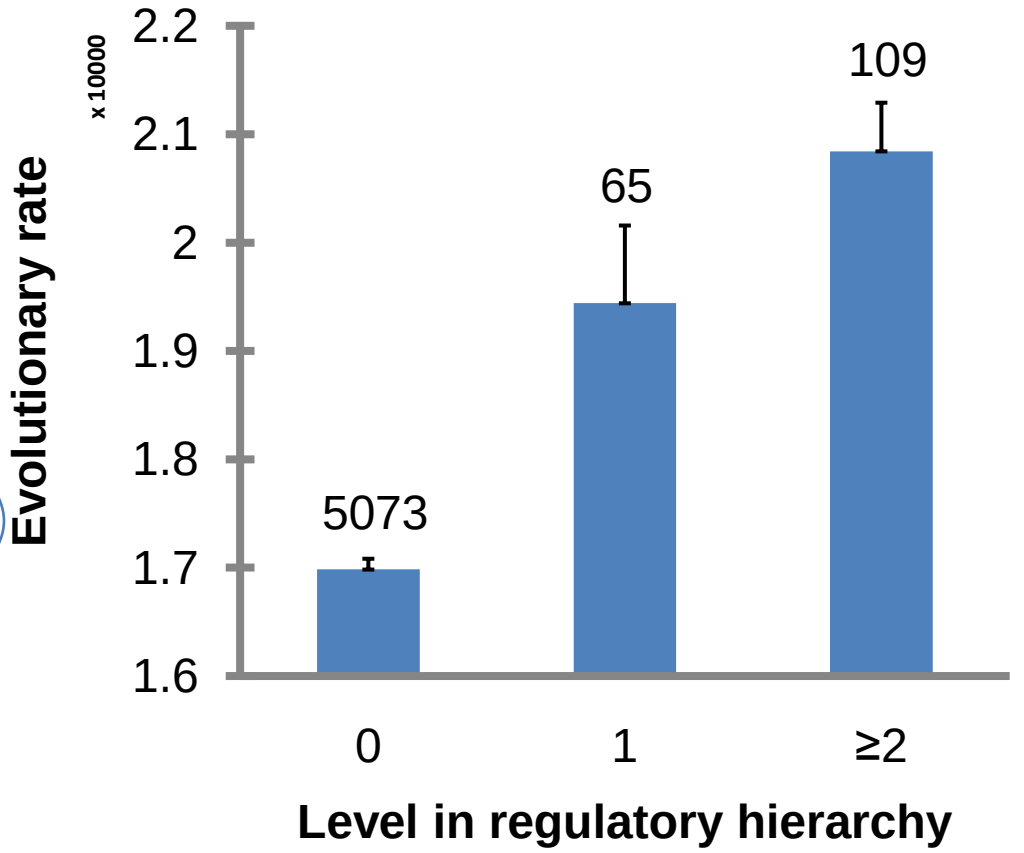

co-regulation
TF
TF
TF
P
P
P
P


Evolutionary Rate (K_A/K_S)
0.18
0.16
0.14
0.12
0.10
0.08
19
30
29
32
33
0-1
2-7
8-16
17-26
27-64
# of Co-Regulation Relationships

Network rewiring model

- We hypothesize that protein-protein interactions operate at a low level in the cellular network, and tend to be conserved during evolution.
- On the other hand, TF-TF associations operate at a high level in the cellular regulatory hierarchy, and tend to rewire during evolution.
- Protein-protein interactions are fundamental to the basic functions of a living cell; more interaction partners for a particular protein will lead to greater structural and functional constraint, resulting in negative selection.
- In contrast, TF-TF associations are more easily changed in evolution compared to protein-protein interactions. Positive selection acts to fix specific TF-TF associations that are beneficial to a particular organism in a particular environment. The rewiring of TF-TF associations also encourages adaptive TF evolution.


Level in regulatory hierarchy
3
2
1
0


Evolutionary rate
x 10000
2.2
2.1
2
1.9
1.8
1.7
1.6
5073
65
109
0
1
≥2
Level in regulatory hierarchy

Lesson learned

- We observe that while generic protein hubs tend to evolve more slowly than non-hubs, TF hubs tend to evolve more quickly than TF non-hubs.
- We made the surprising finding that two of the most important interactome subnetworks, the TF interactome and the protein interactome, are fundamentally different in terms of their function and evolution.
- Our work demonstrates a high degree of functional and evolutionary heterogeneity within biological networks, and highlights the rich insights that can be gained from modeling biomolecular subnetworks.

Take-home messages

- Network is powerful
- Network is a new platform
- Network can be dangerous
- More stories in network can be expected, but we need to ask a good question first!!!