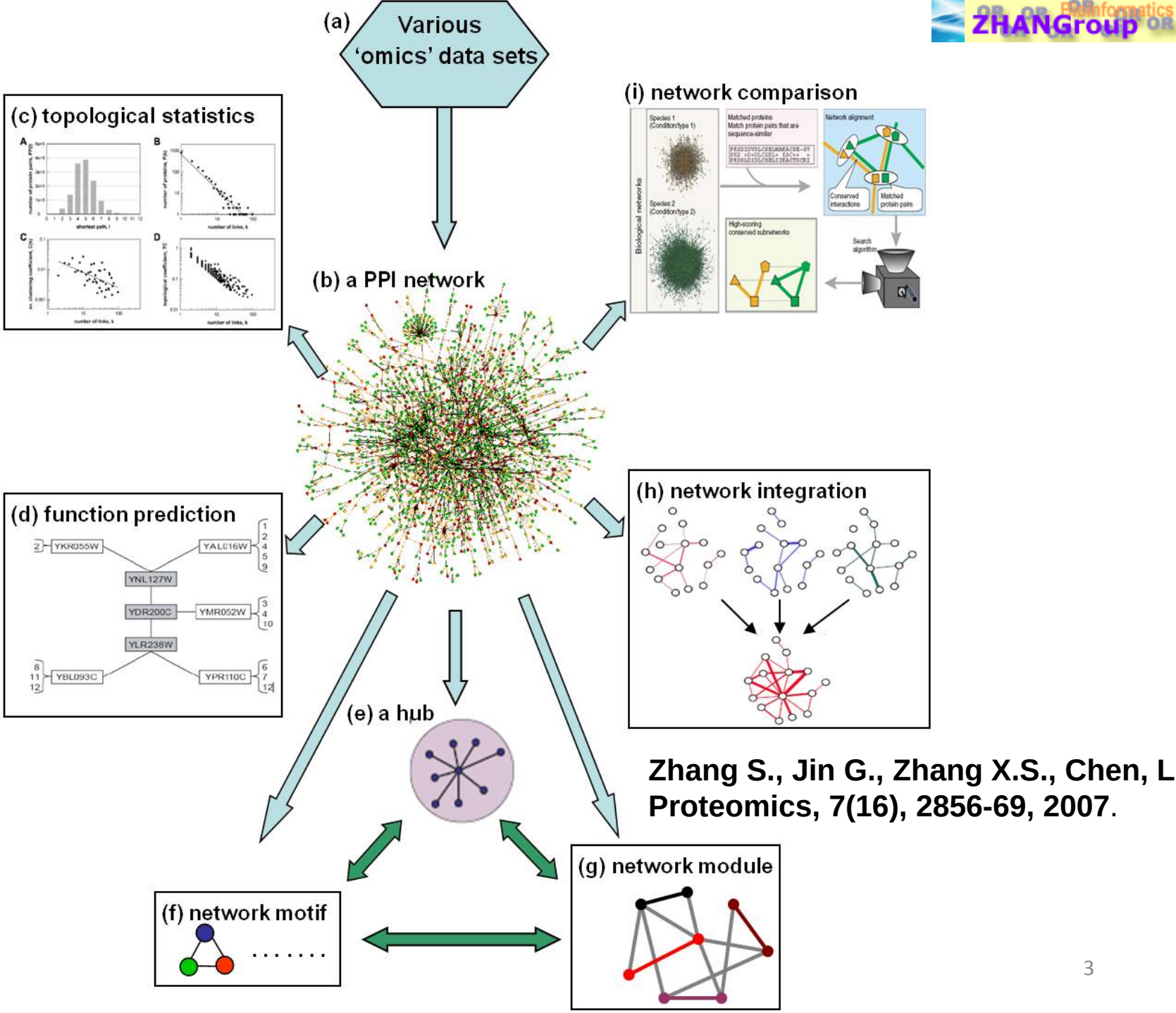
生物信息学与系统生物学

张世华

中国科学院数学与系统科学研究院

Bio-molecular network analysis

Shihua Zhang



Zhang S., Jin G., Zhang X.S., Chen, L
Proteomics, 7(16), 2856-69, 2007.

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

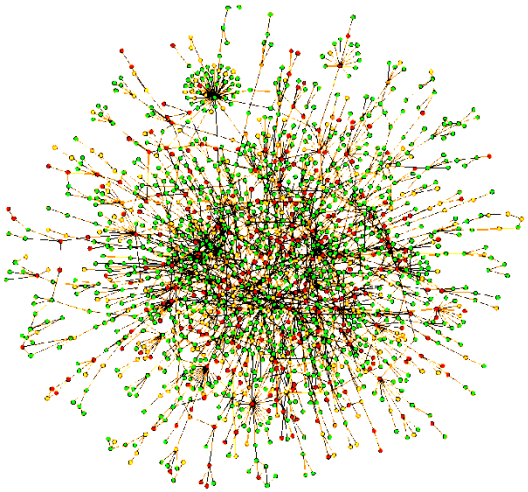
Subnetworks in different conditions

- 子网络分析 (Subnetworks)

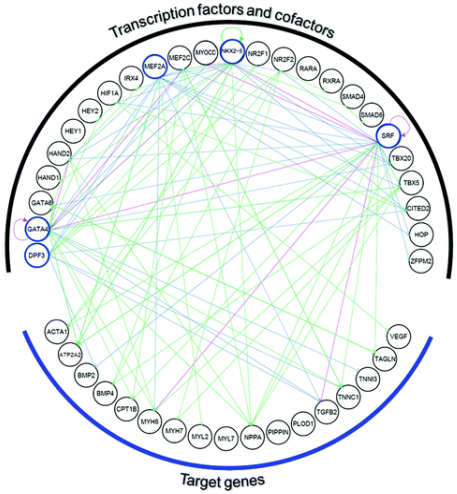
Aging and disease subnetwork

Evolution in TF subnetwork

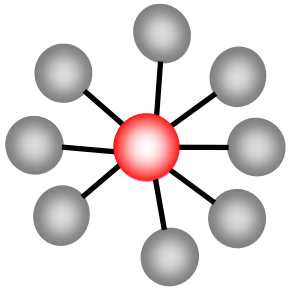
Different Types of Molecular Networks



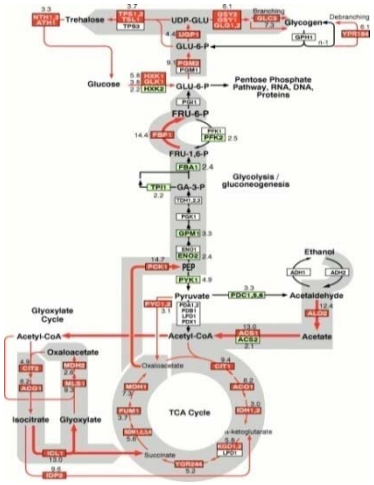
Protein-protein Interaction networks



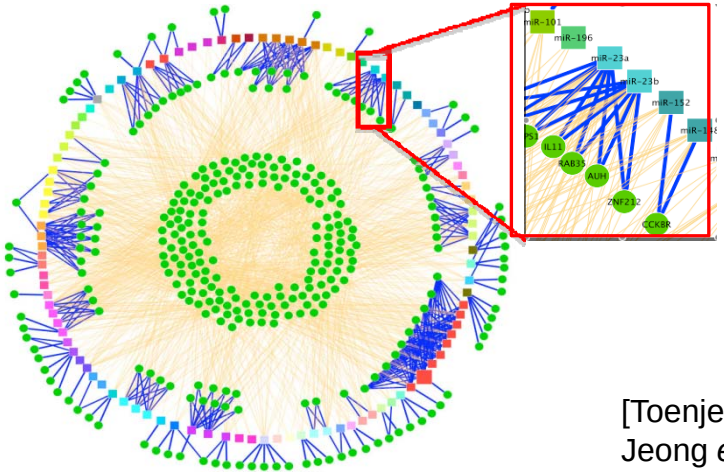
TF-target-gene Regulatory networks



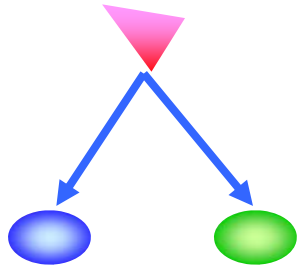
Undirected



Metabolic pathway networks



miRNA-target networks



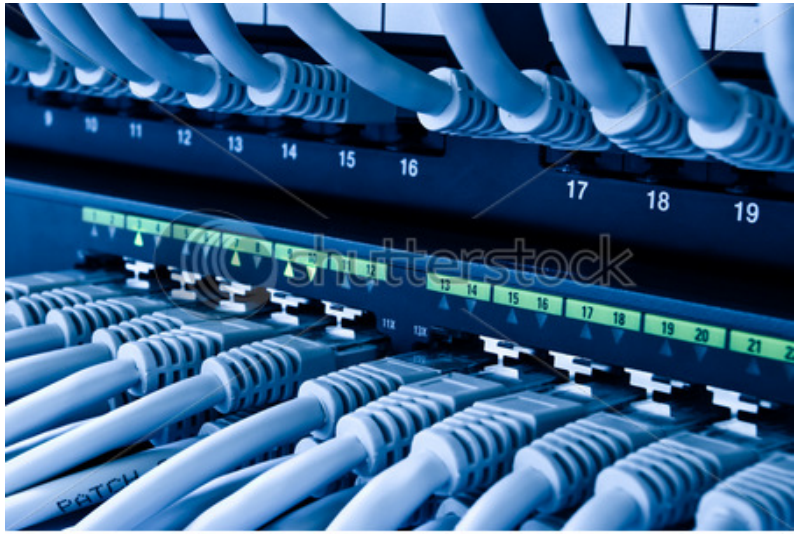
Directed

[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 Nodes in Networks: Hubs & Bottlenecks

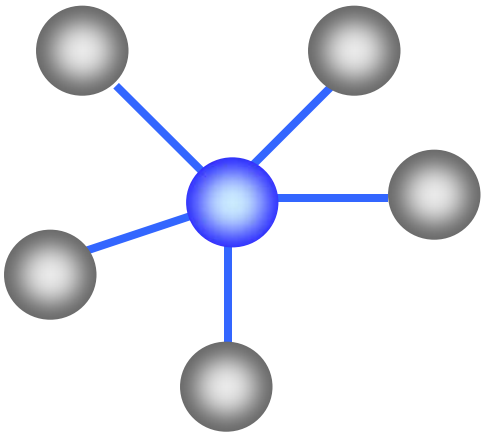
Which are key nodes in networks ? How do we locate them ?

Hub & bottleneck?



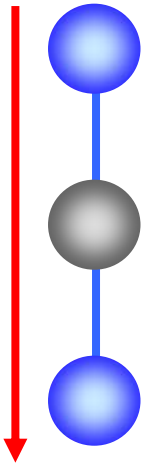
Global topological measures

Indicate the gross topological structure of the network



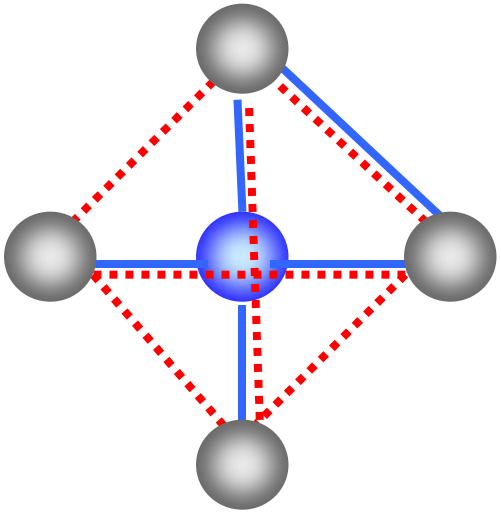
Degree (*K*)

5



Path length (*L*)

2

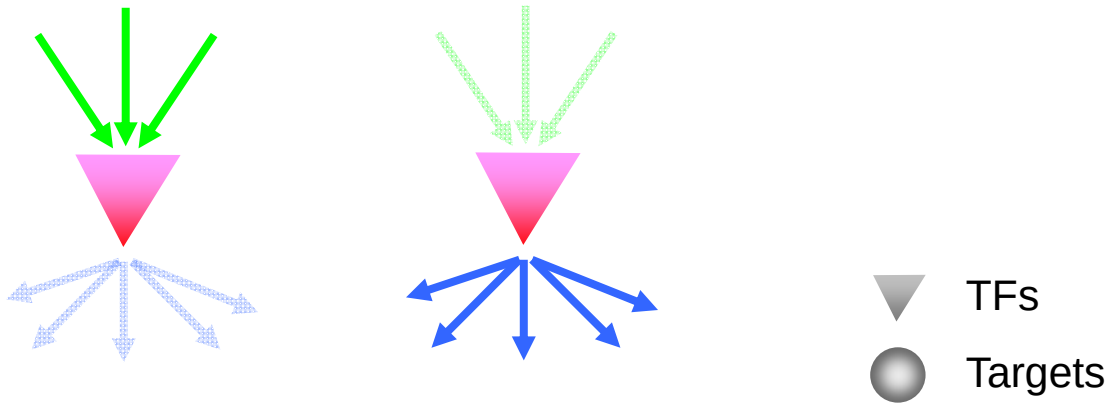


Clustering coefficient (*C*)

1/6

Interaction and co-expression networks are ***undirected***

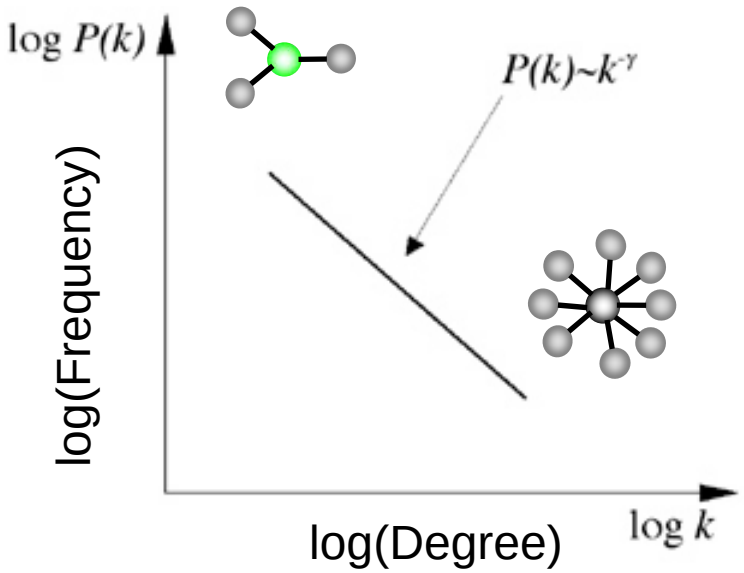
Global topological measures for directed networks



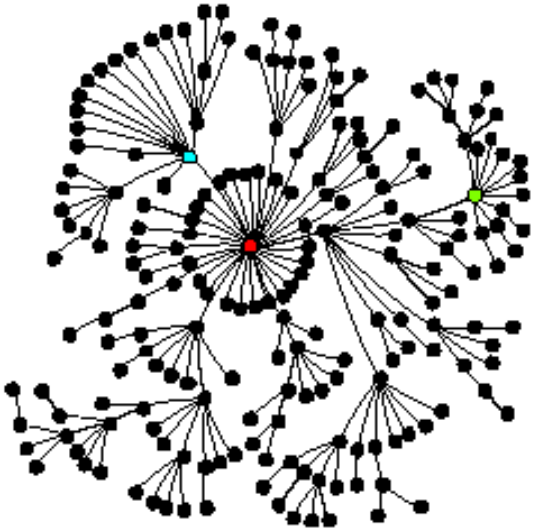
Regulatory and metabolic networks are ***directed***

Scale-free networks

Power-law distribution



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



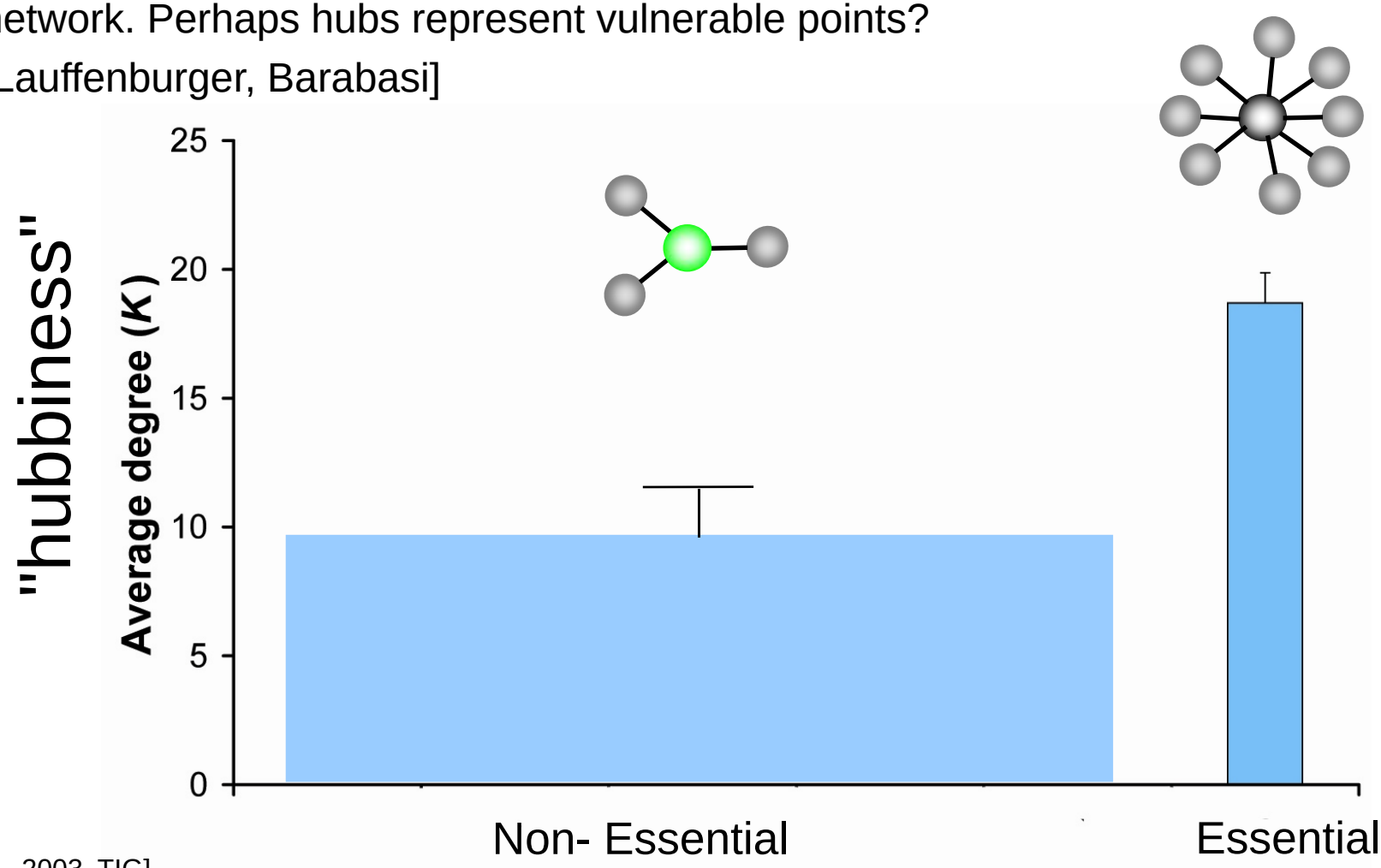
Hubs dictate the structure of the network

[Barabasi]

Hubs tend to be Essential

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

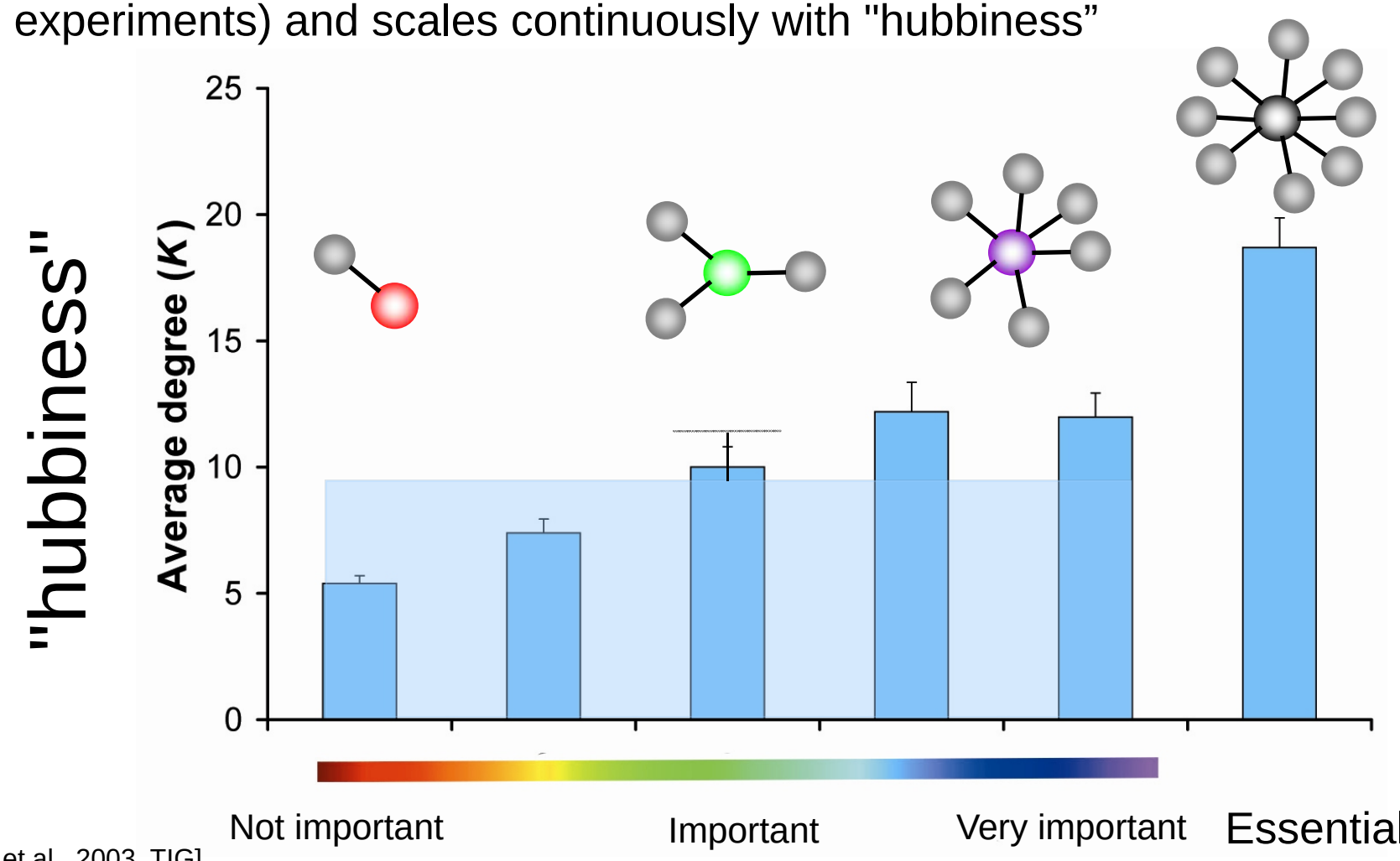
[Lauffenburger, Barabasi]



[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"



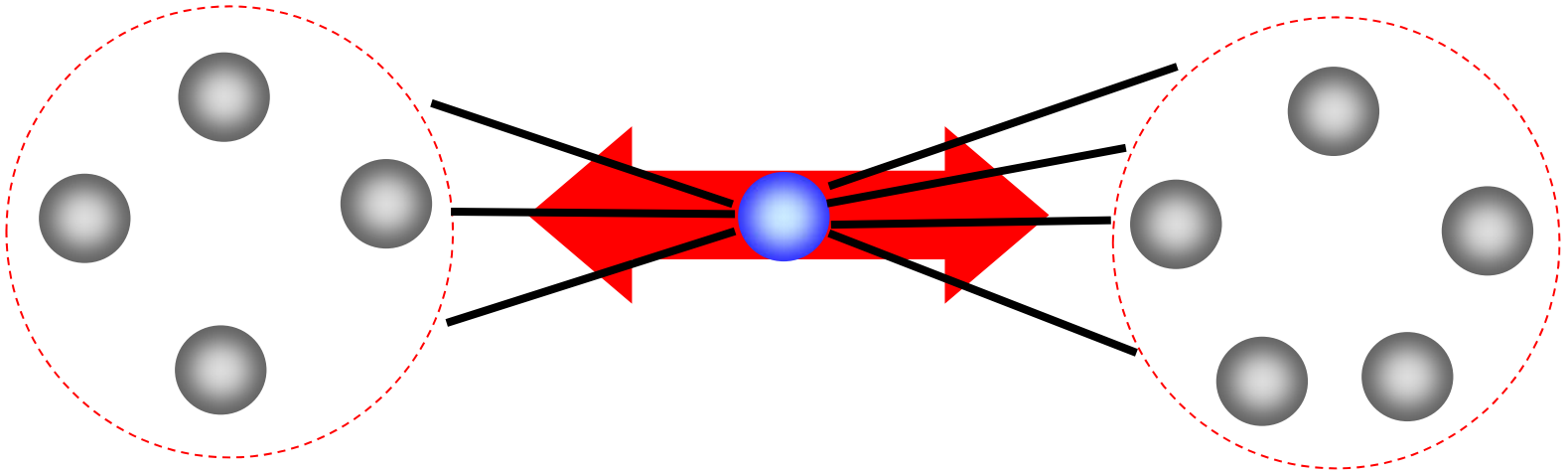
[Yu et al., 2003, TIG]

Another measure of Centrality: Betweenness centrality

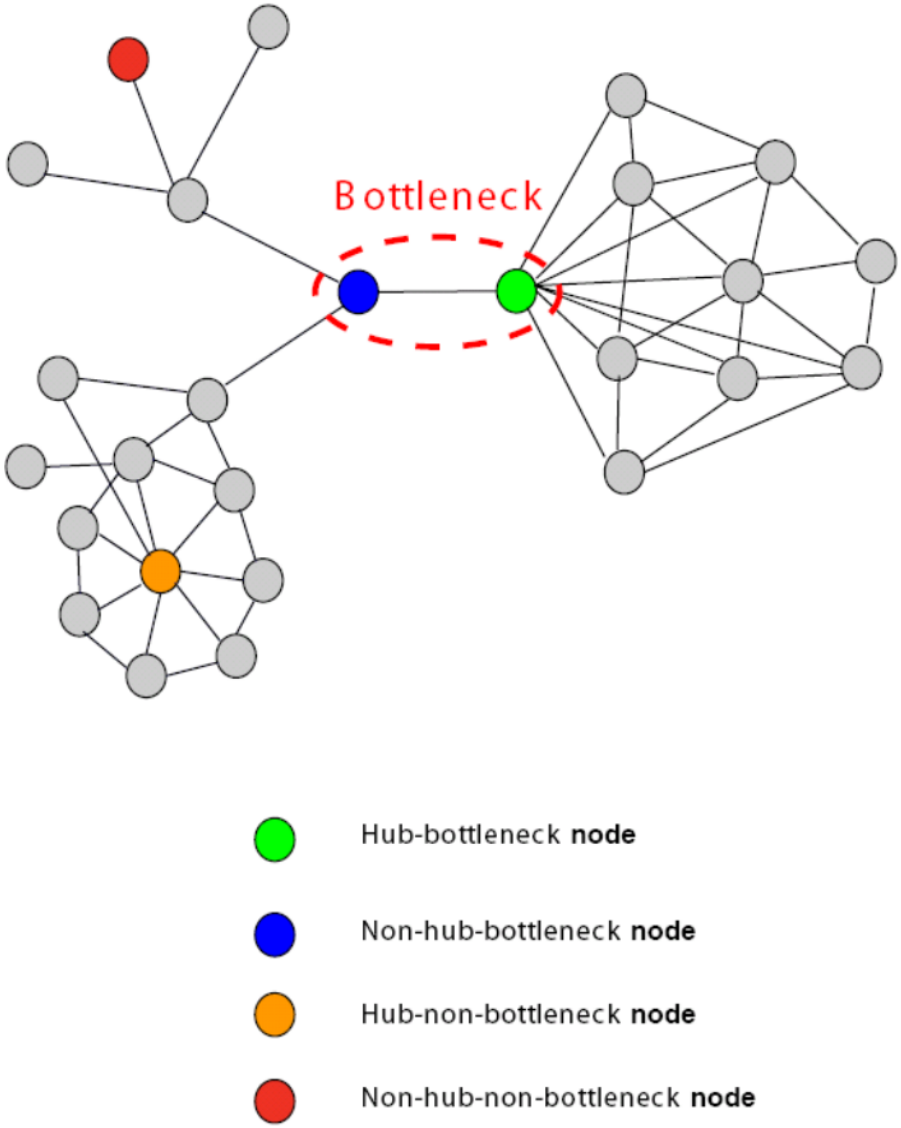
Betweenness of a node is the number of shortest paths of pairs of vertices that run through it -- a measure of information flow.

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

Girvan & Newman (2002) PNAS 99: 7821.

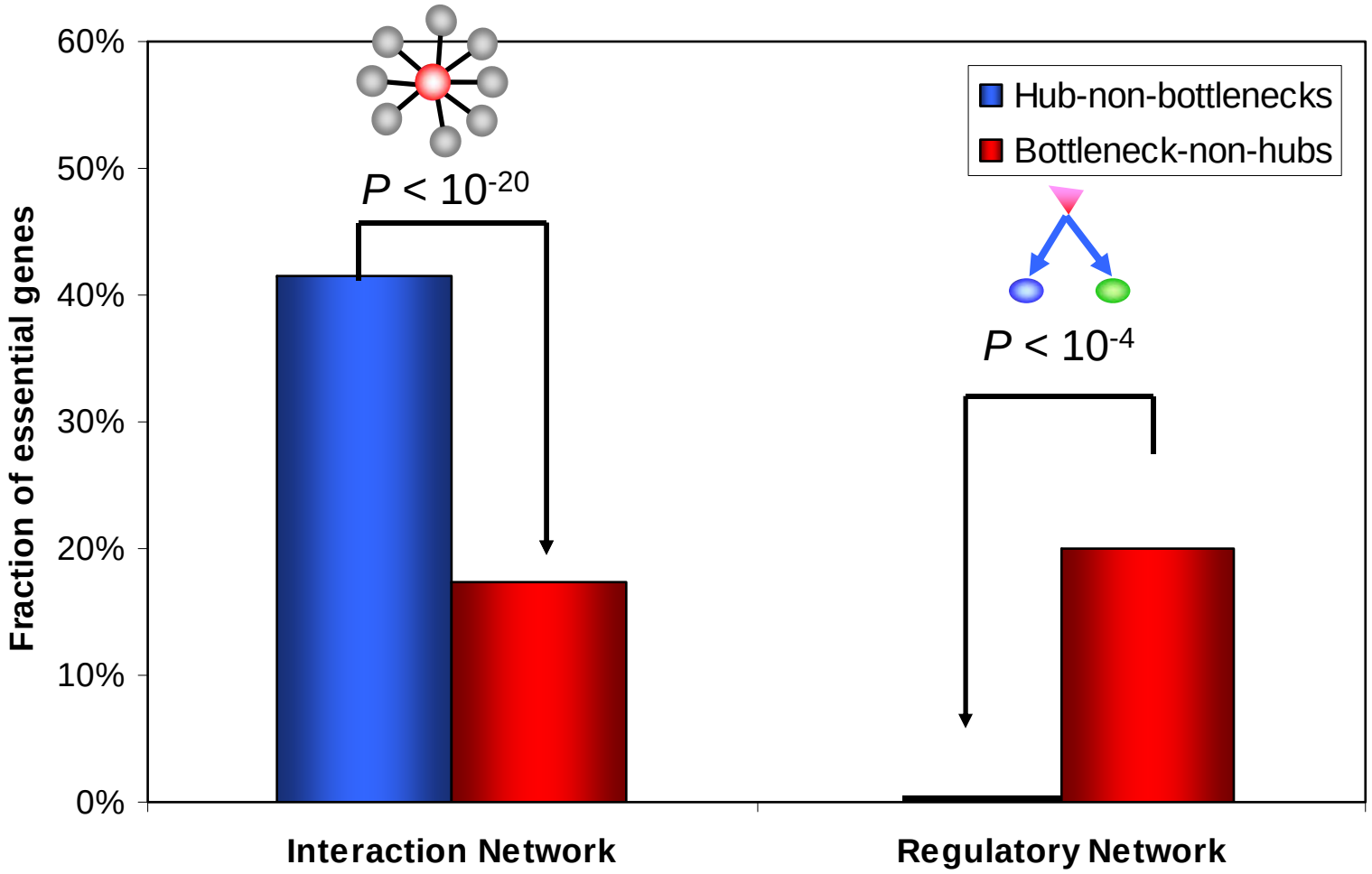


Bottlenecks & Hubs



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

Bottlenecks are what matters in regulatory networks

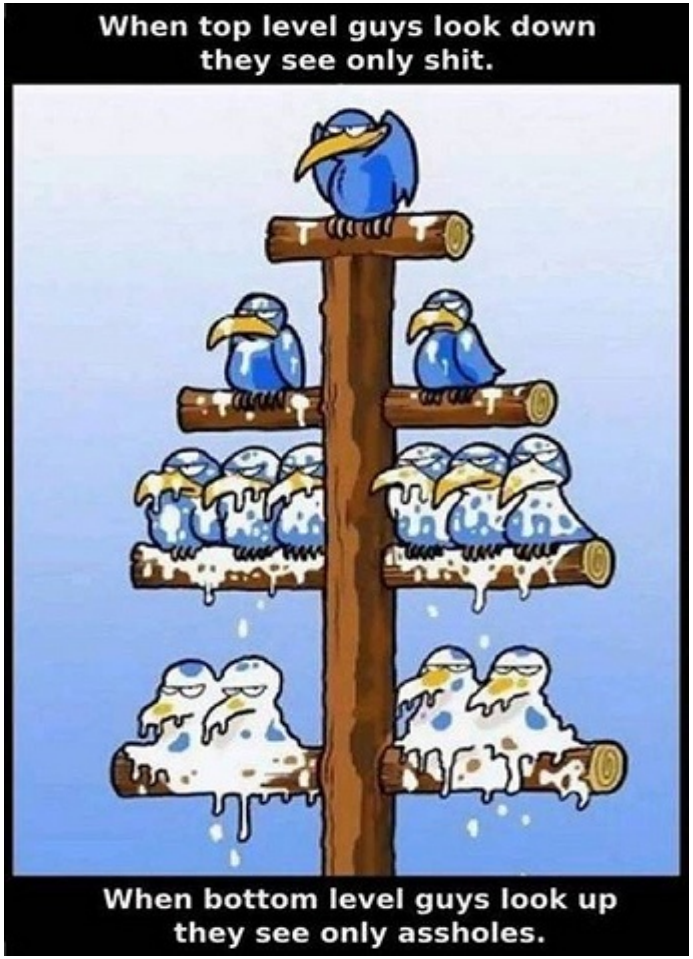


[Yu et al., Plos CB (2007)]

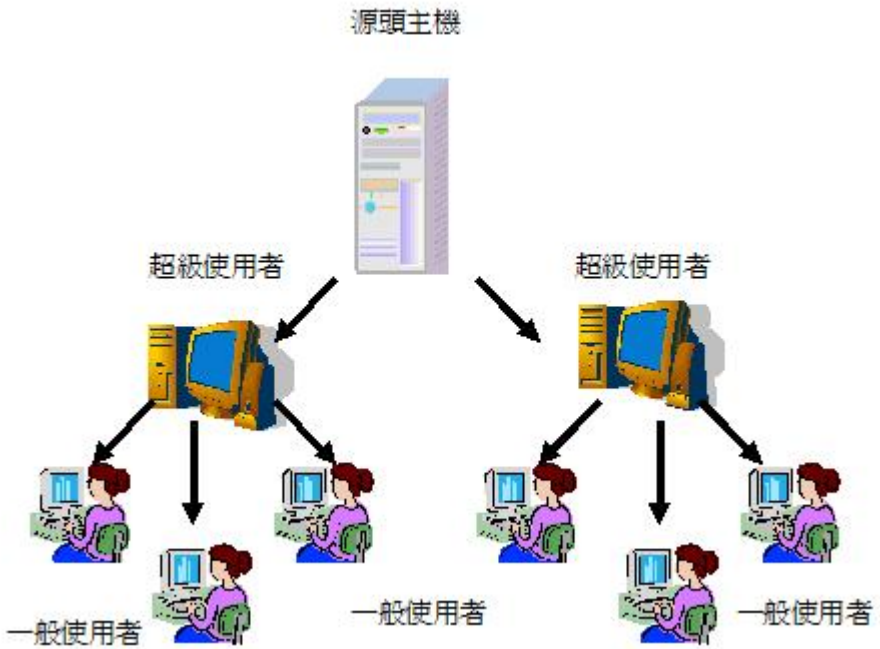
Q2: Does the Bio-molecular networks posses hierarchy structure

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

Hierarchy structure

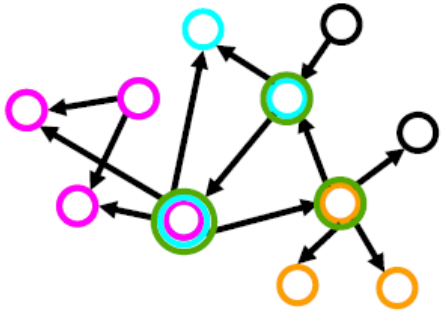


Management Hierachy

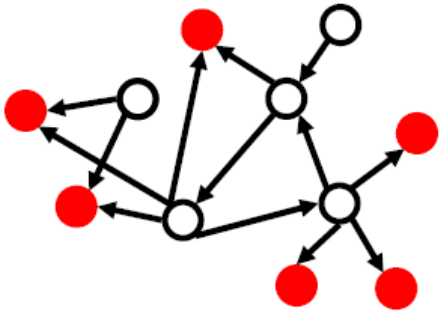


Determination of "Level" in Regulatory Network Hierarchy with Breadth-first Search

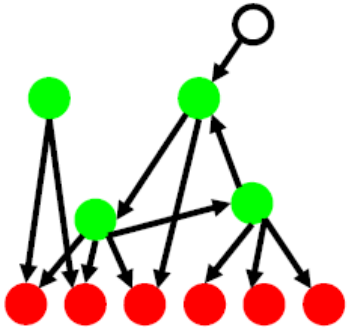
I. Example network with all 4 motifs



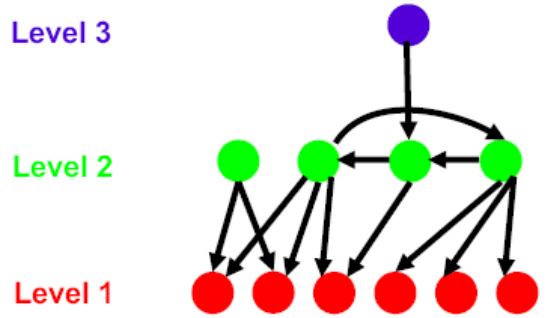
II. Finding terminal nodes (Red)



III. Finding mid-level nodes (Green)

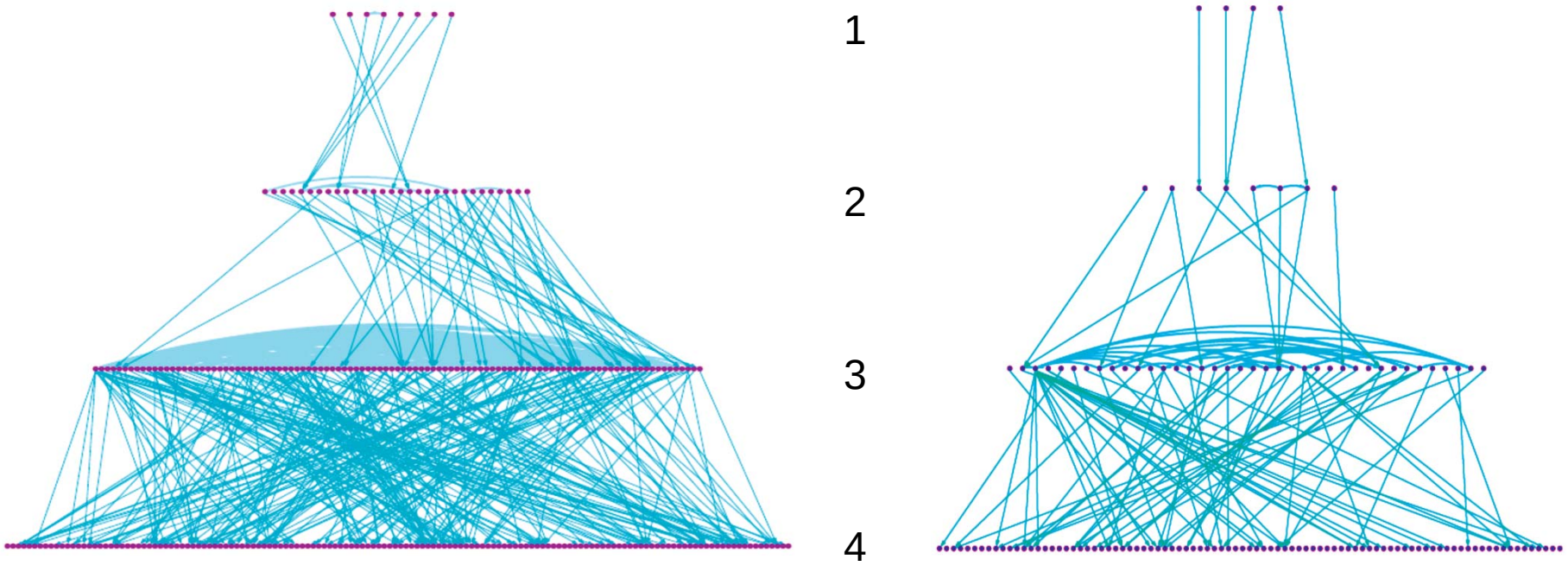


IV. Finding top-most nodes (Blue)



[Yu et al., PNAS (2006)]

Regulatory Networks have similar hierarchical structures

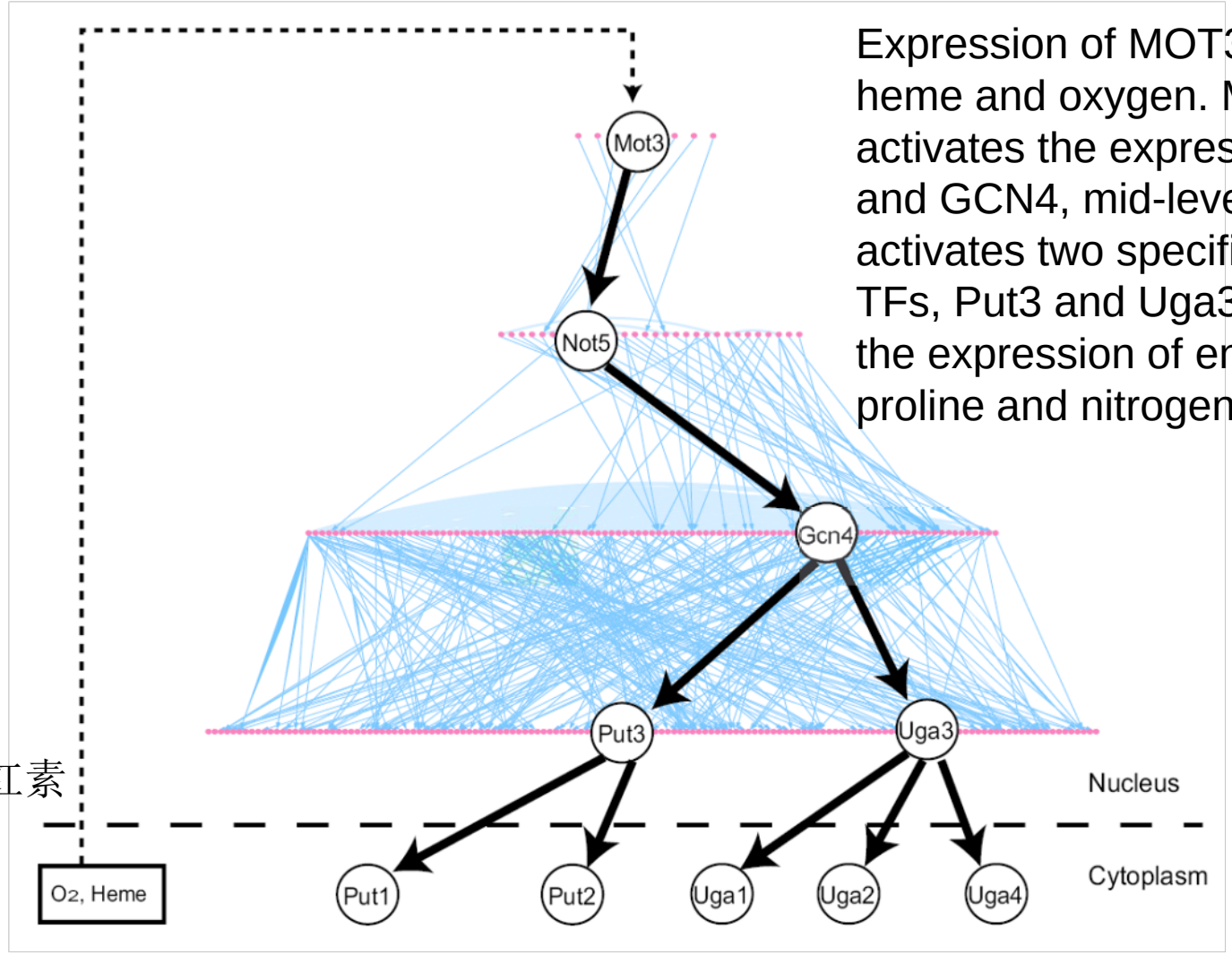


S. cerevisiae

E. coli

[Yu *et al.*, PNAS (2006)]

Example of Path Through Regulatory Network

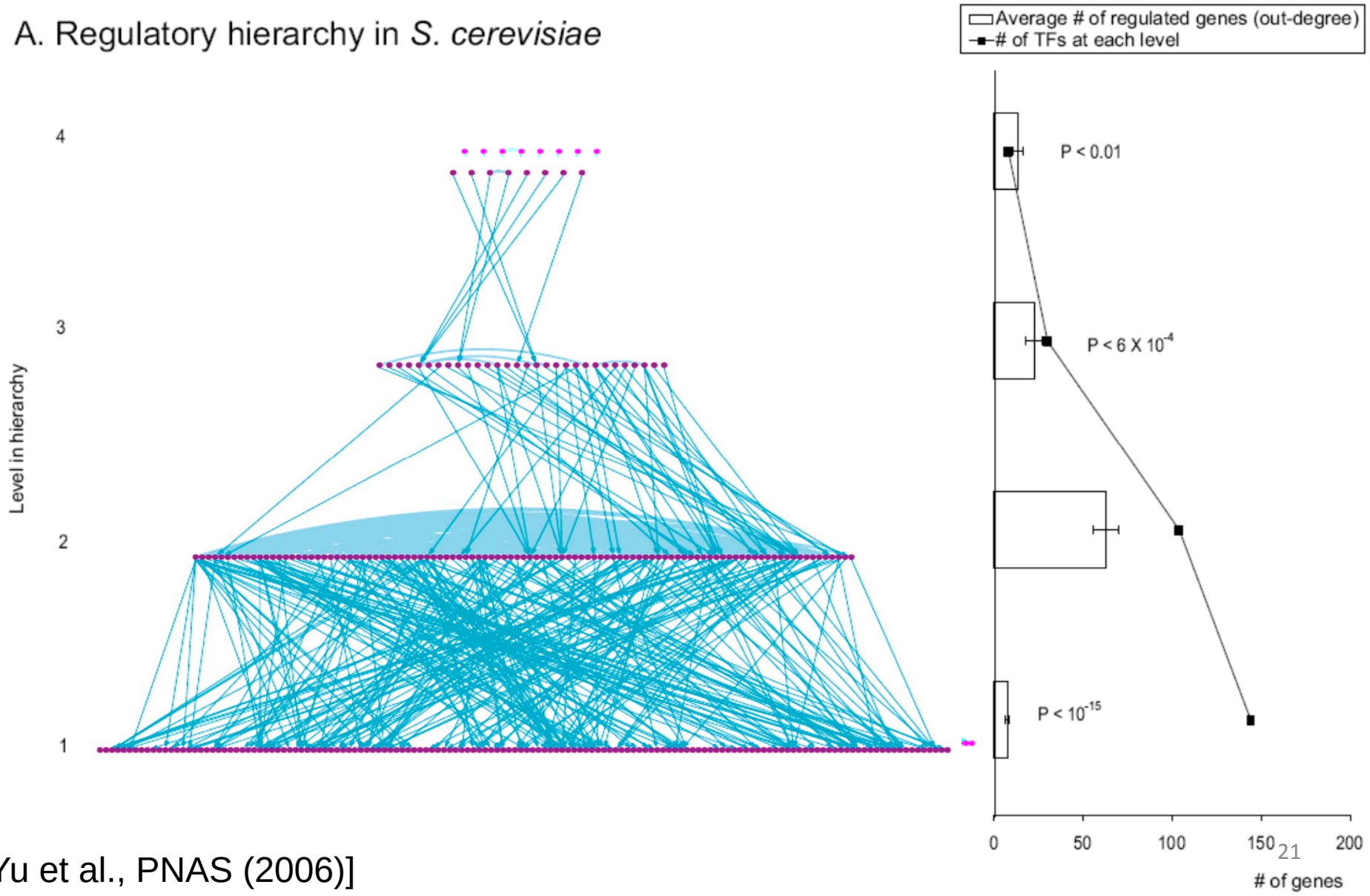


Expression of MOT3 is activated by heme and oxygen. Mot3 in turn activates the expression of NOT5 and GCN4, mid-level hubs. GCN4 activates two specific bottom-level TFs, Put3 and Uga3, which trigger the expression of enzymes in proline and nitrogen utilization.

亚铁血红素

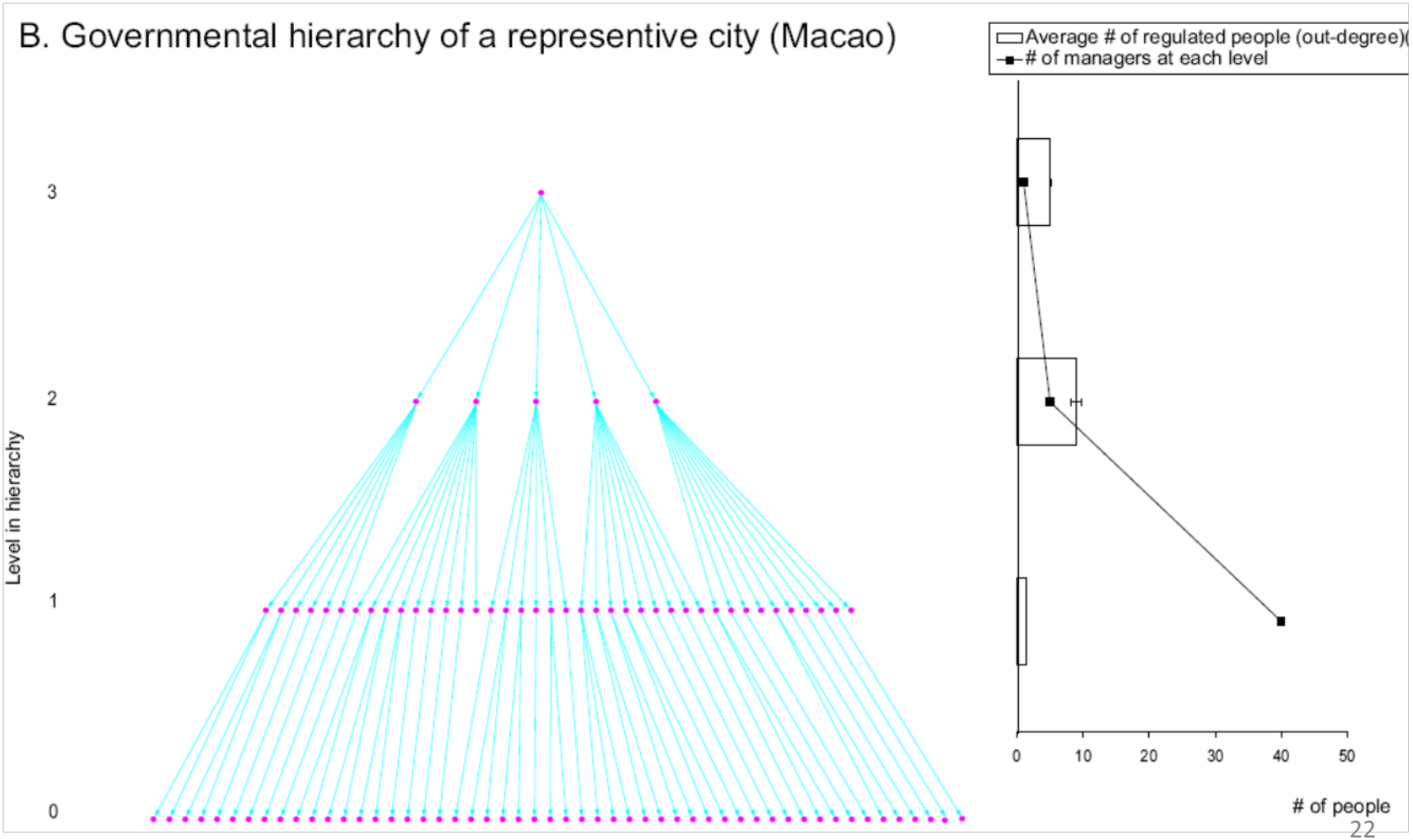
[Yu et al., PNAS (2006)]

Yeast Regulatory Hierarchy

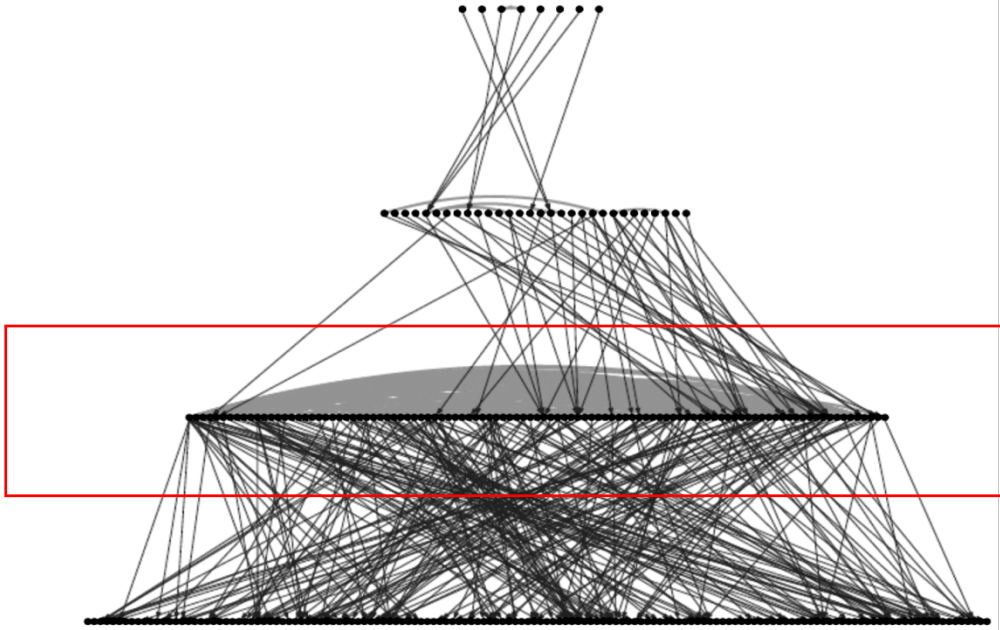


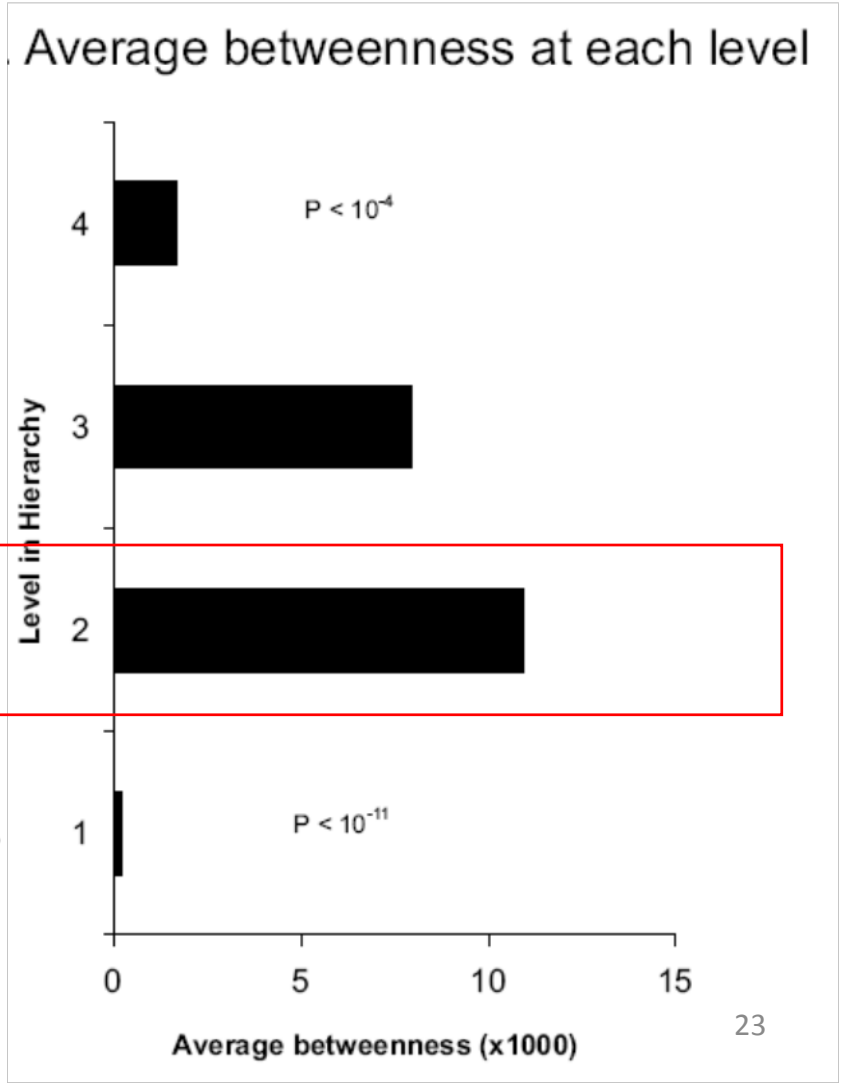
[Yu et al., PNAS (2006)]

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



Characteristics of Regulatory Hierarchy: Middle Managers are Information Flow Bottlenecks

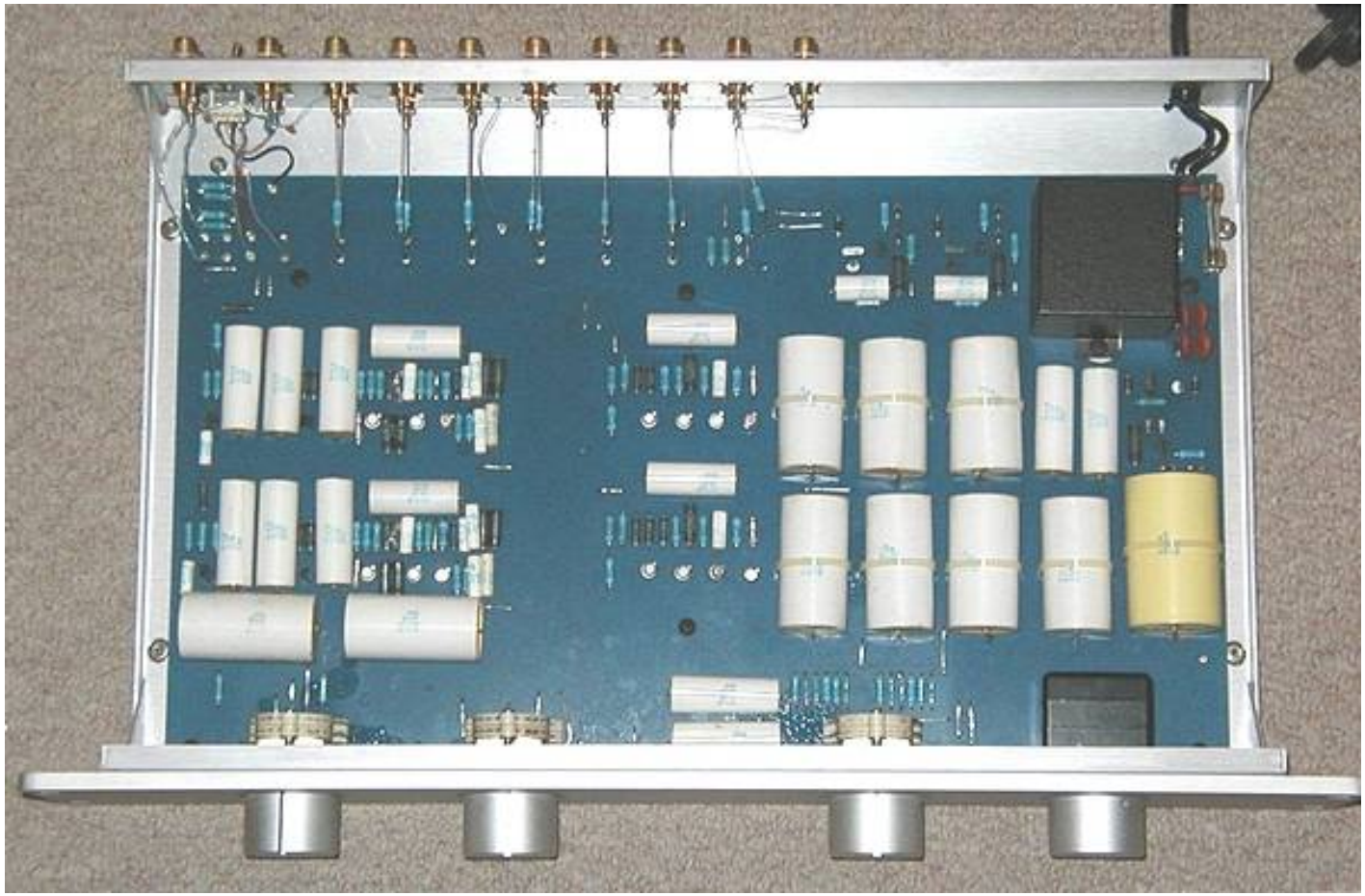




[Yu et al., PNAS (2006)]

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

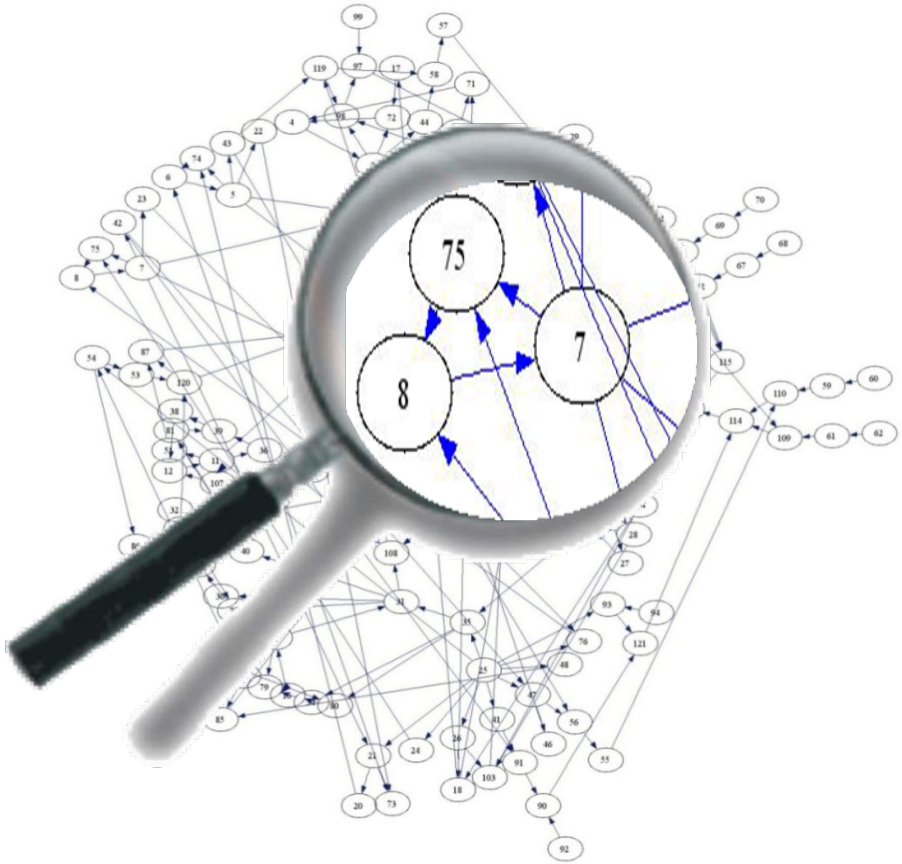
Where are they? How do we identify them? What does it mean?

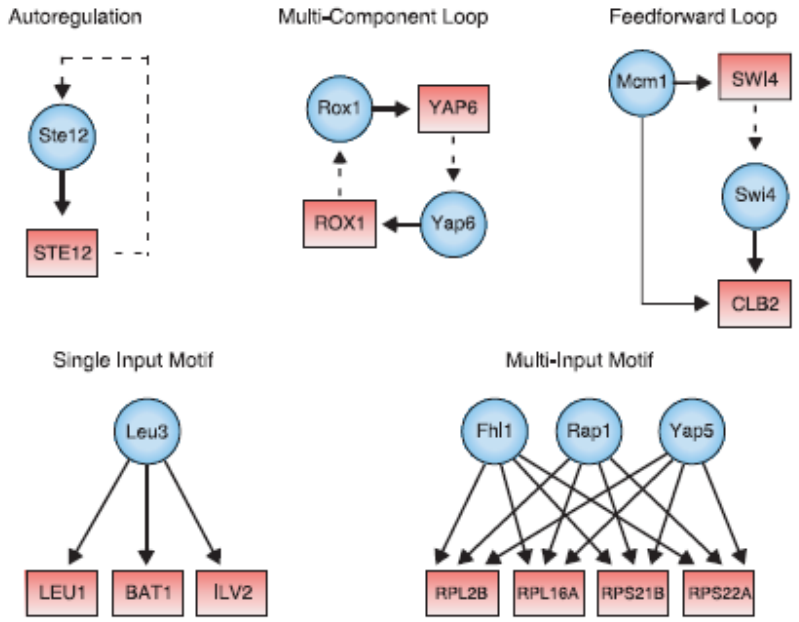


Circuit network

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

Network Motifs：Simple Building Blocks of Complex Networks



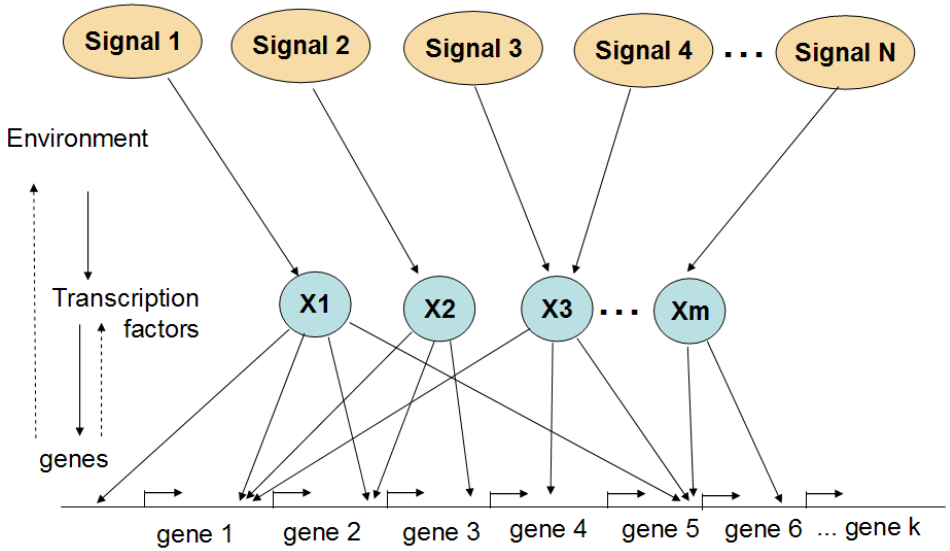


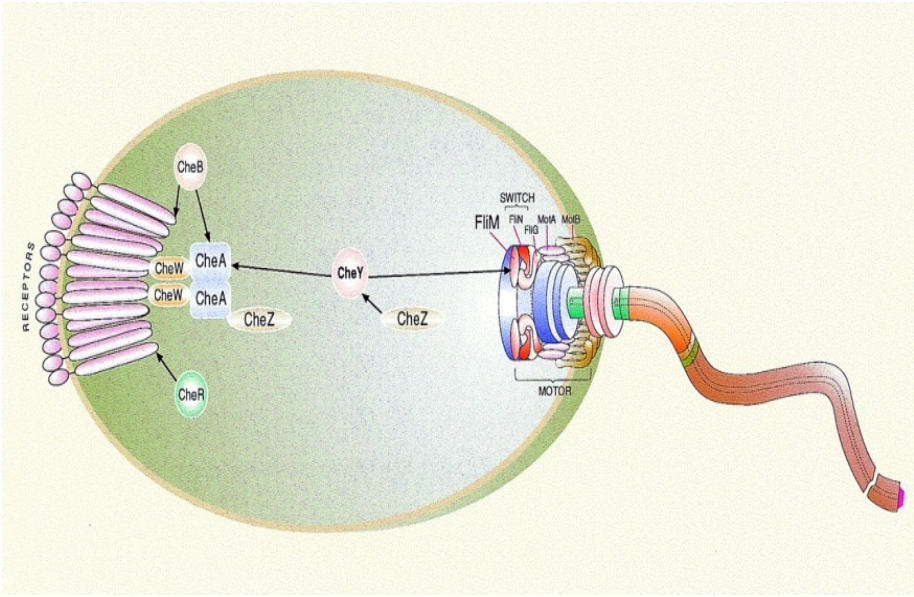
Science, 298:799-804, 2002

- 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

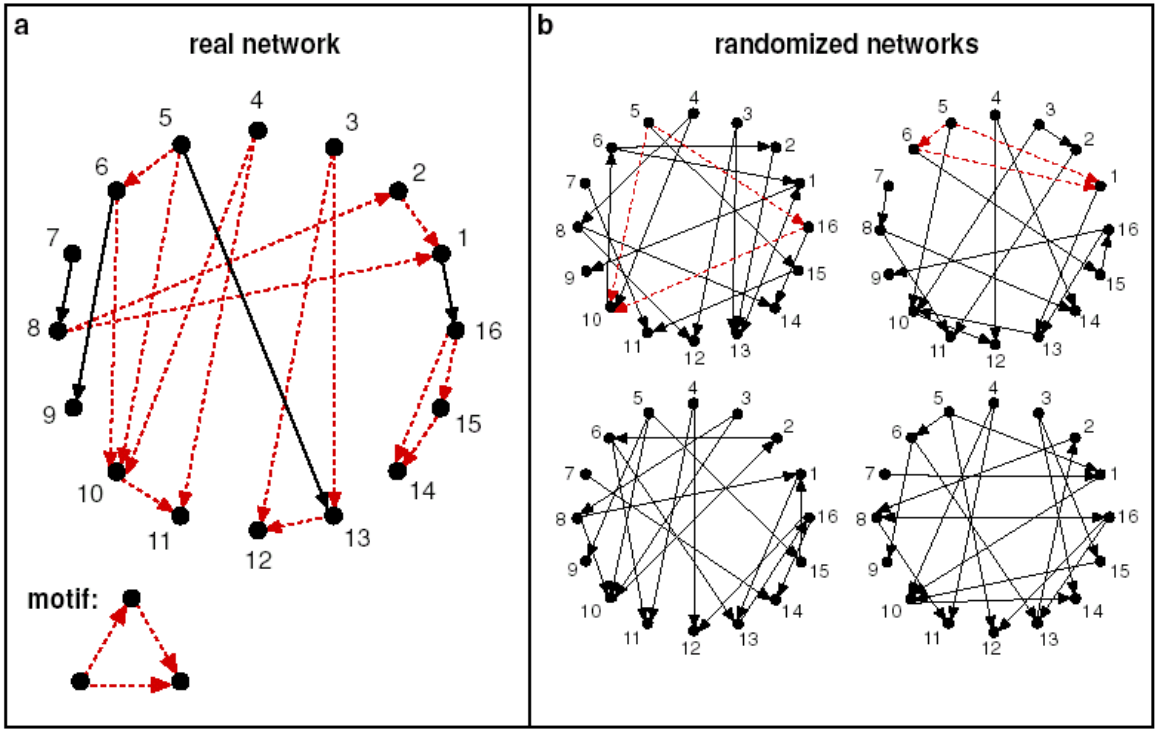




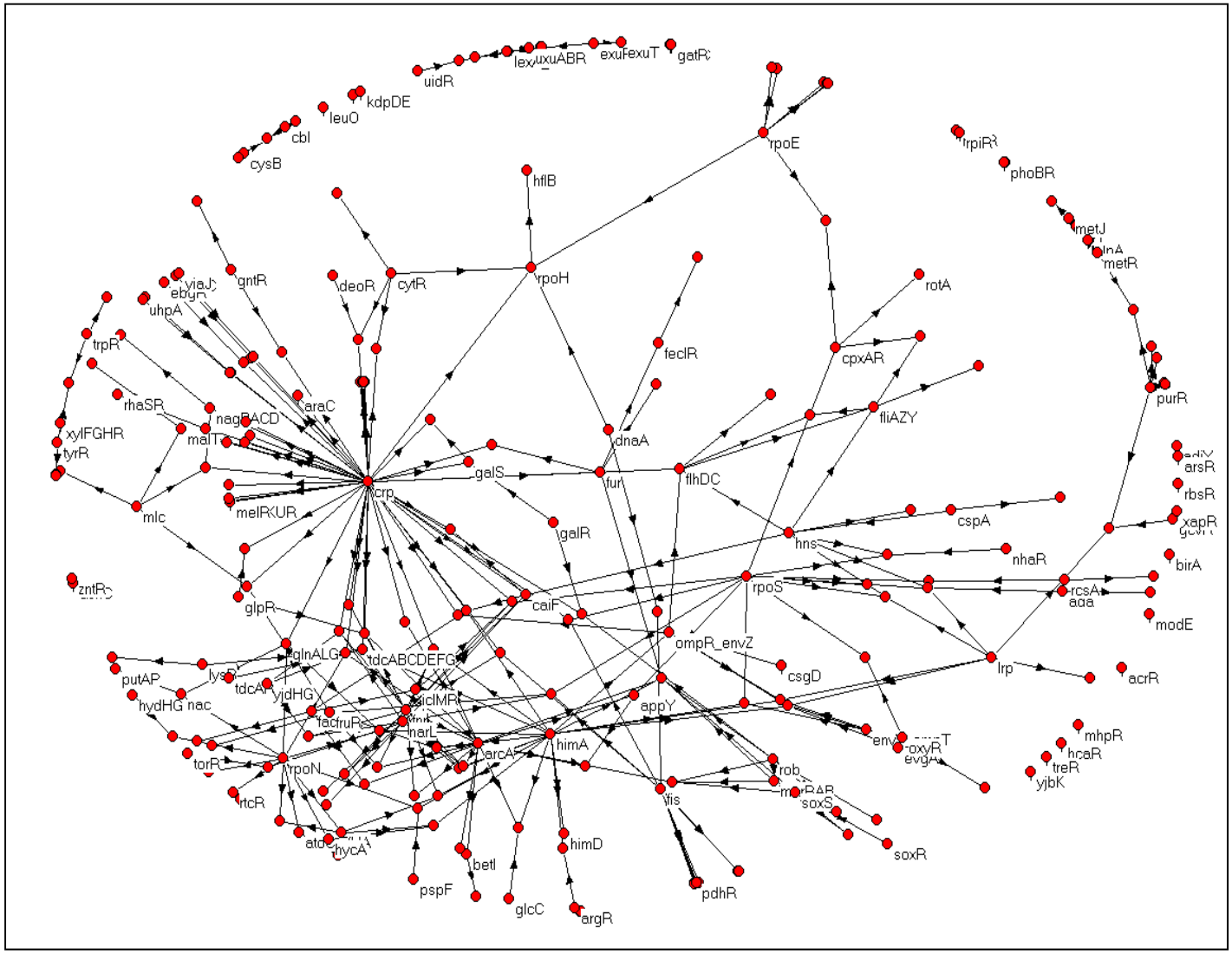
- E. Coli – 1000 protein types at any given moment >4000 genes (or possible protein types) – need regulatory mechanism to select the active set
- We are interested in the design principles of this network

Analyzing networks

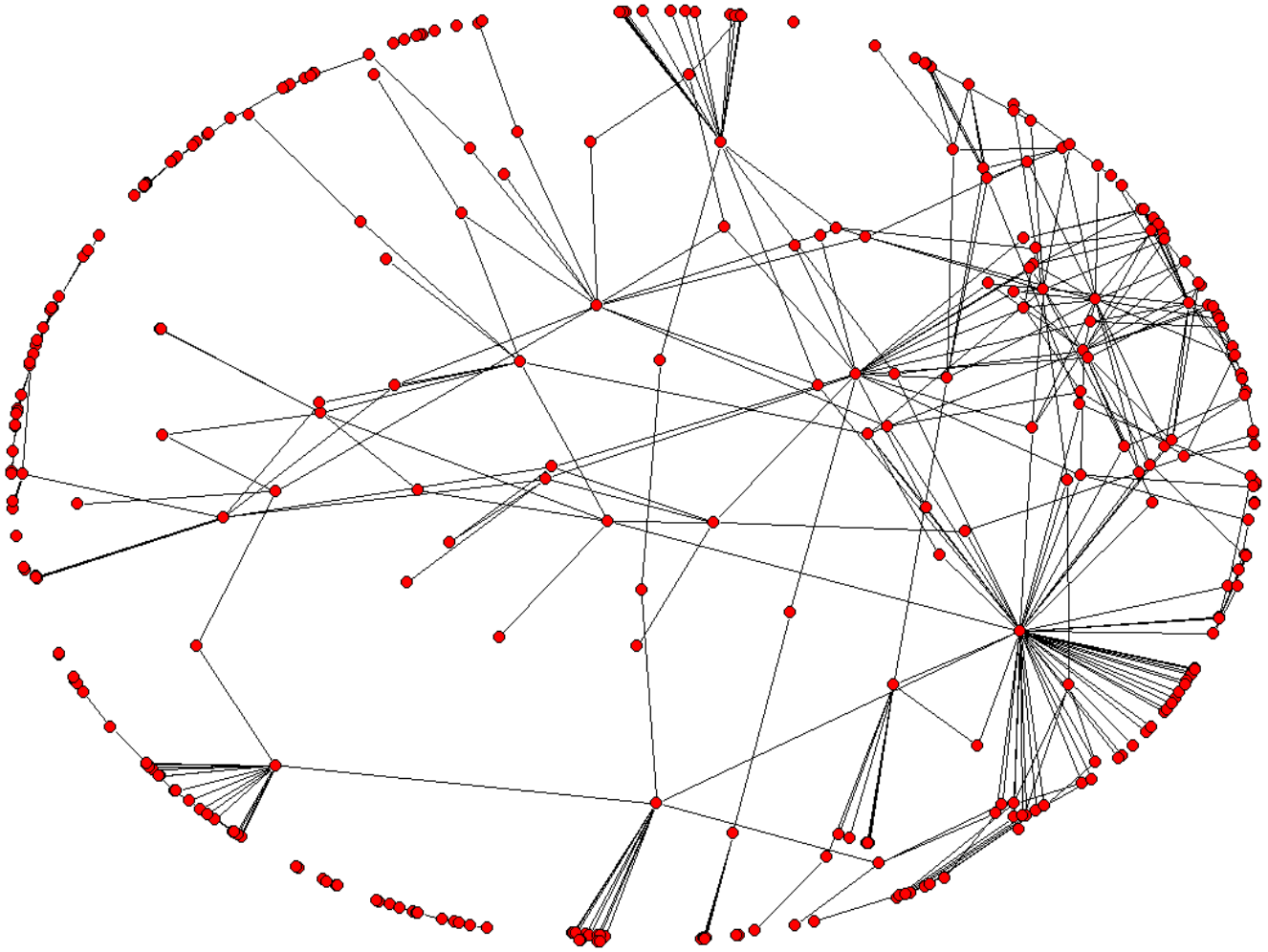
- The idea- **patterns that occur in the real network much more** than in a randomized network, must have functional significance.
- The **randomized networks** share the same number of edges and number of nodes, but edges are assigned at random



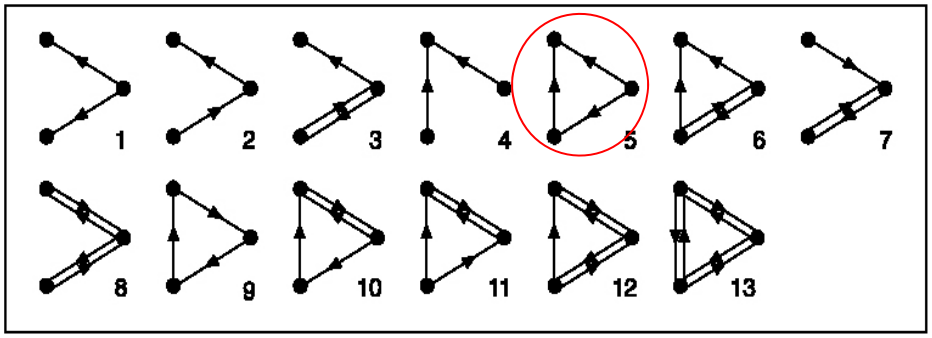
The known E. Coli transcription network

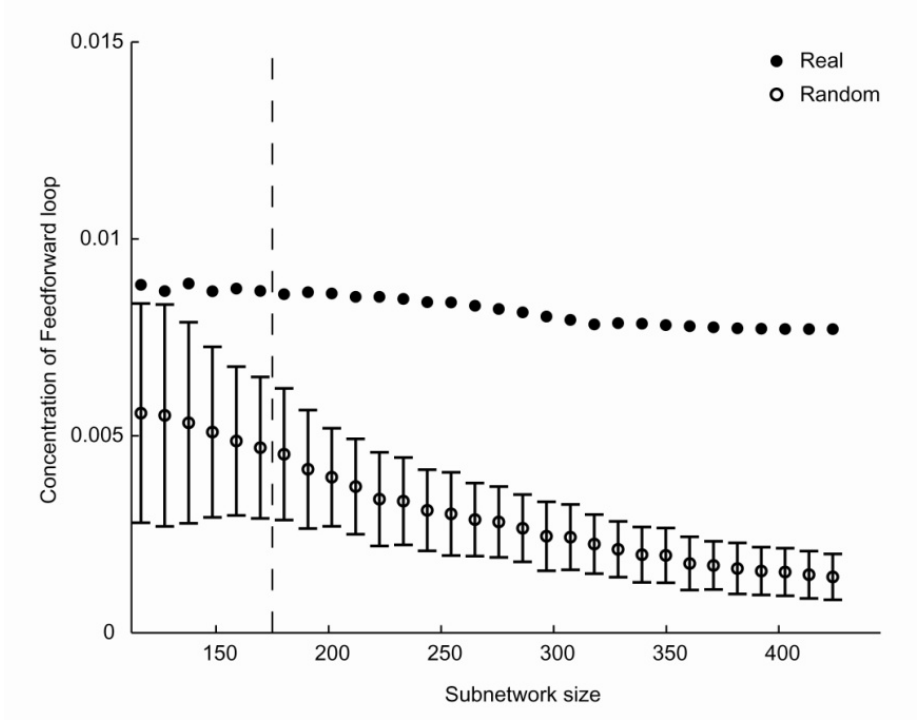


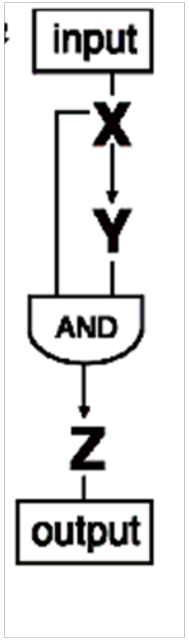
A random graph based on the same node statistics



3-node network motif – the feed-forward loop



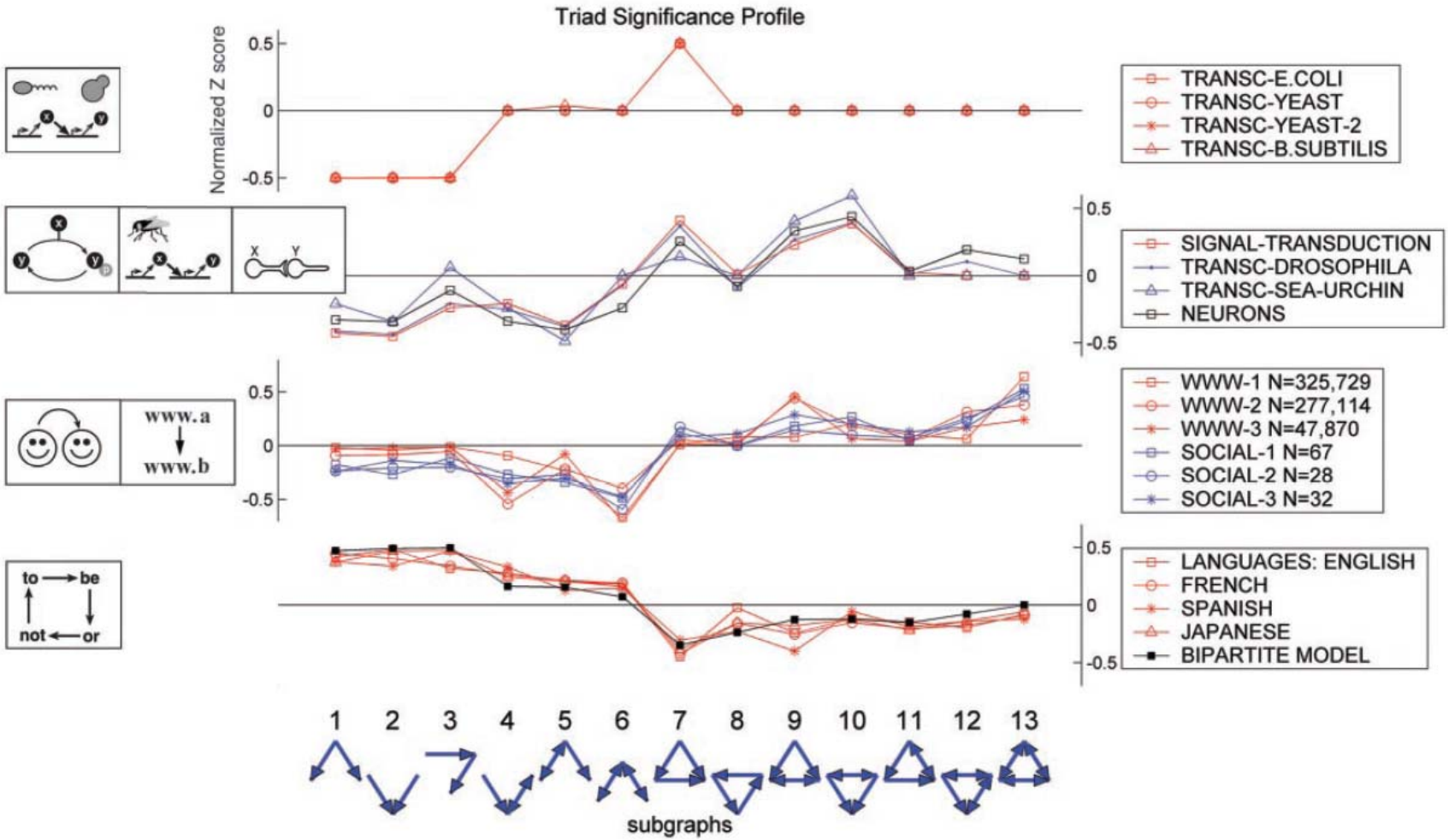




Nreal=40

Nrand=7±3

Superfamilies of Evolved and *Designed* Networks

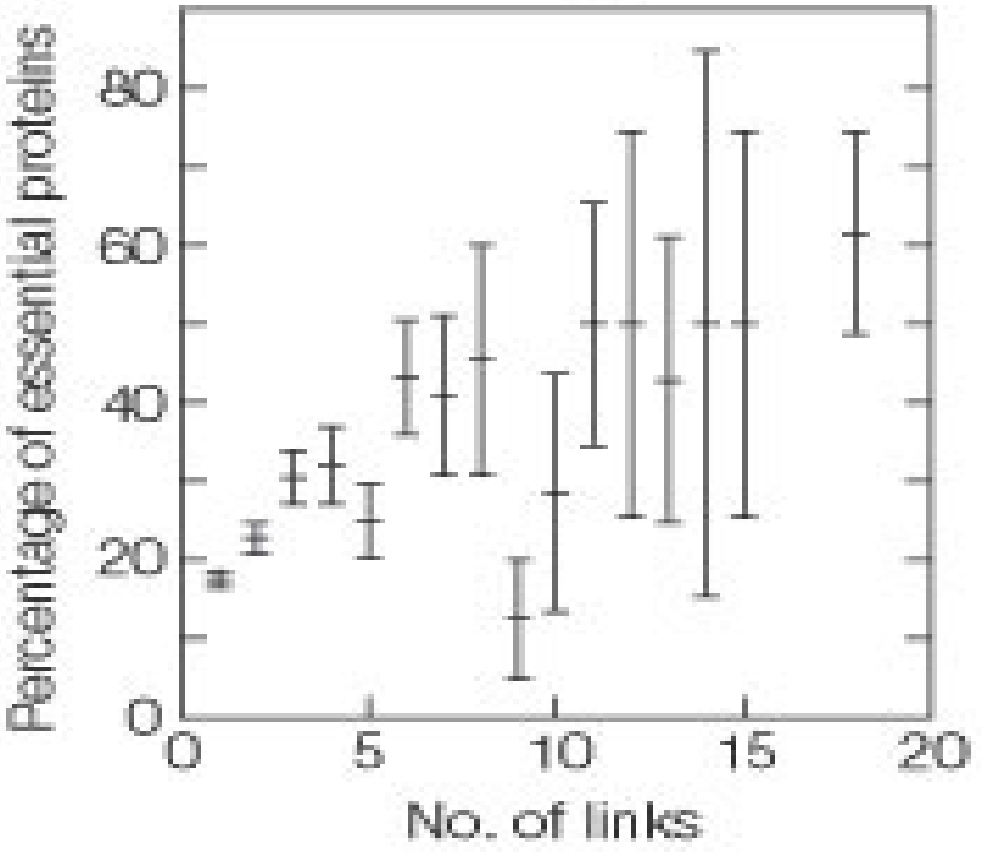


Milo R, et al. (2004), Science. 303:1538 –1542.

Q4: Are all hubs equal?
From the temporal aspect!!

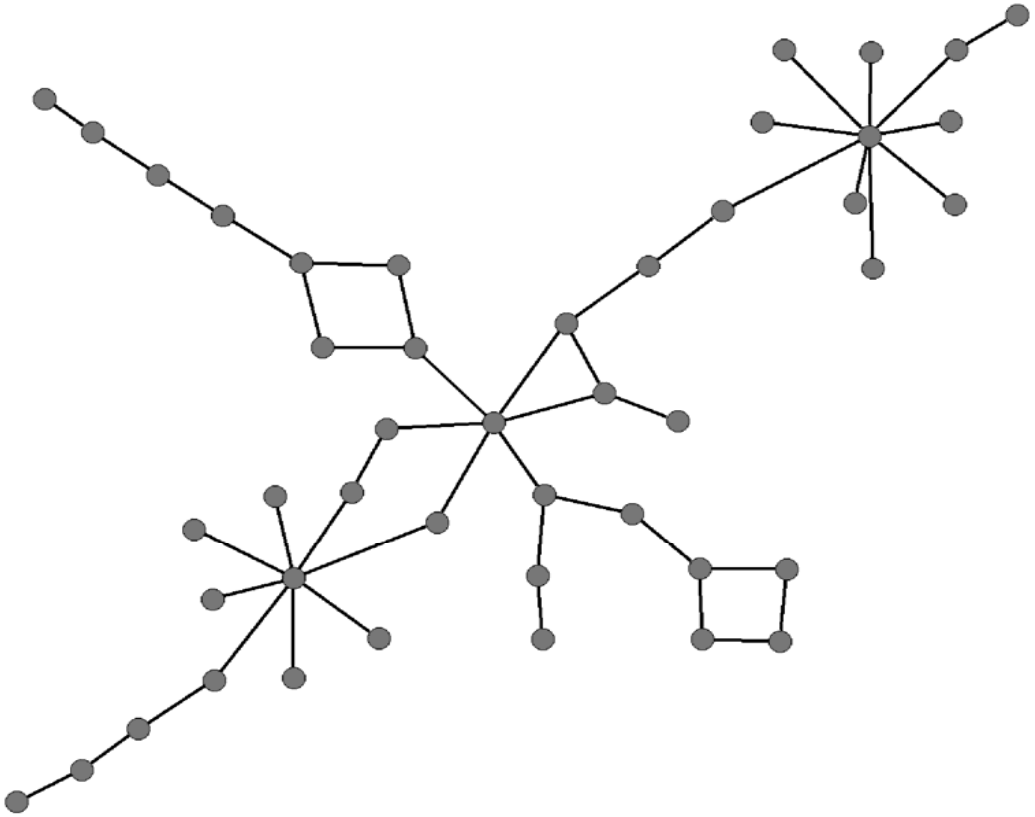
Yeast hubs are three-times more likely to be essential

Yeast Interactome mapped by Y2H is scale-free.



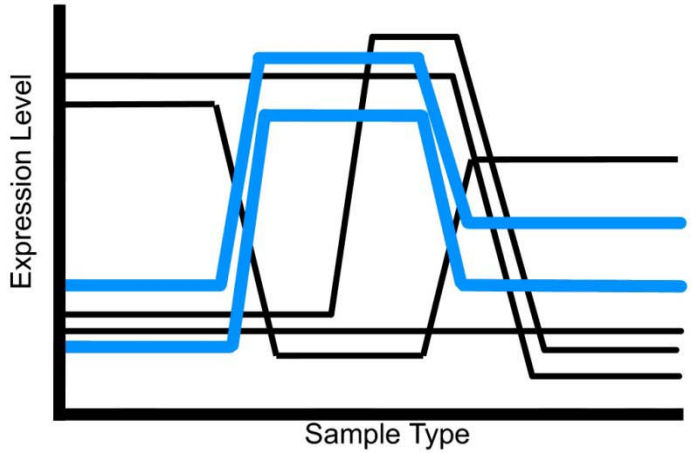
Jeong et al., Nature 2001

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

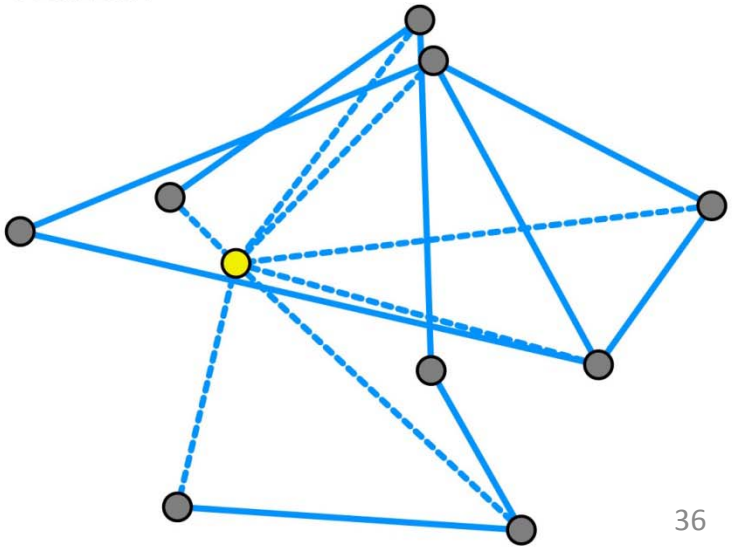
Threshold correlations into edges

C Adjacency Matrix

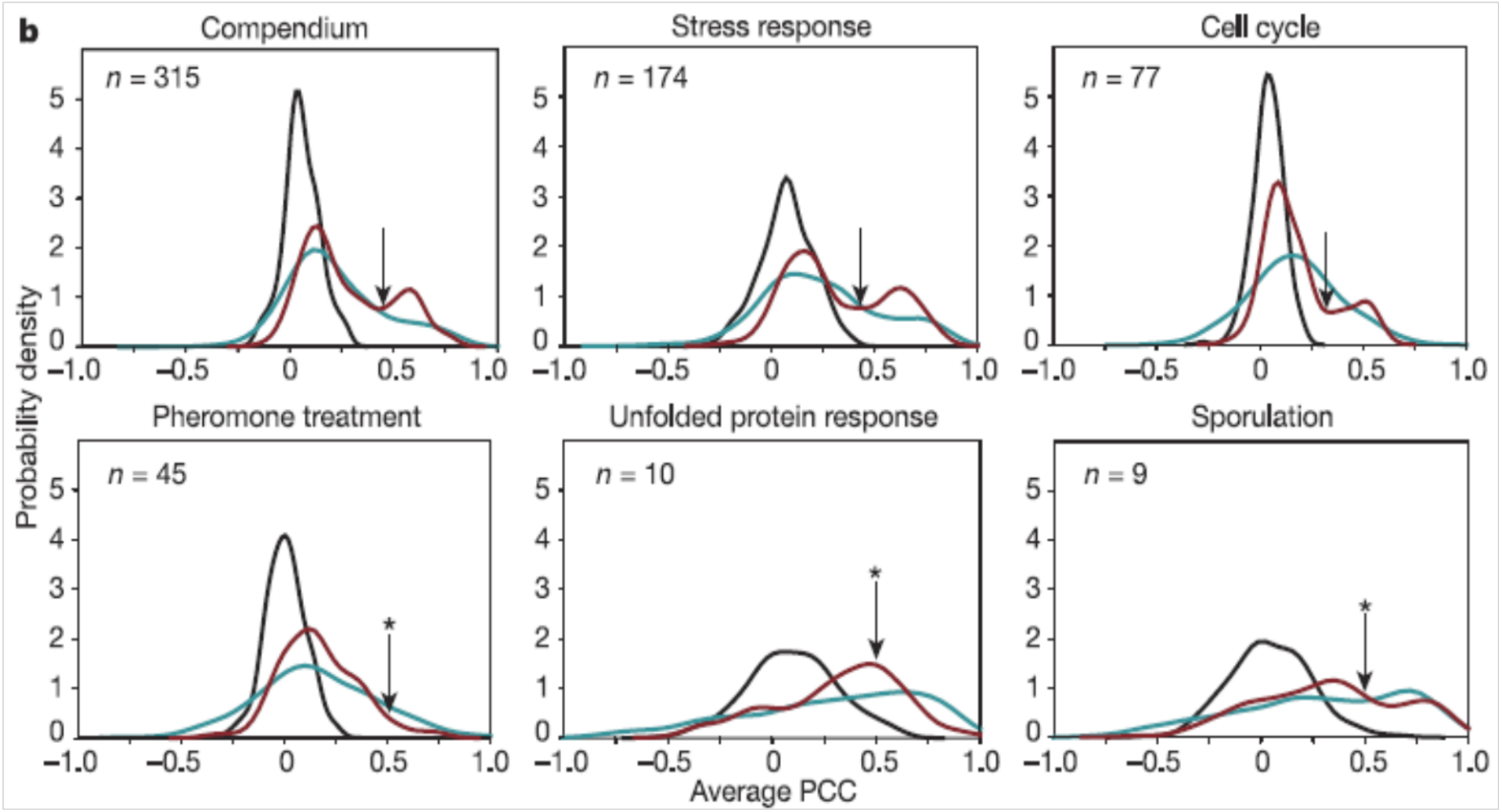
	G1	G2	G3	G4	G5	G6	G7	G8	G9	G10
G1	NA	O	O	E	E	O	E	O	O	O
G2	O	NA	E	O	O	O	E	O	O	O
G3	O	E	NA	O	O	O	E	O	O	E
G4	E	O	O	NA	E	E	E	O	O	O
G5	E	O	O	E	NA	E	E	O	O	O
G6	O	O	O	E	E	NA	O	O	O	O
G7	E	E	E	E	E	O	NA	E	E	O
G8	O	O	O	O	O	O	E	NA	E	O
G9	O	O	O	O	O	O	E	E	NA	E
G10	O	O	E	O	O	O	O	O	E	NA

Draw network

D Network



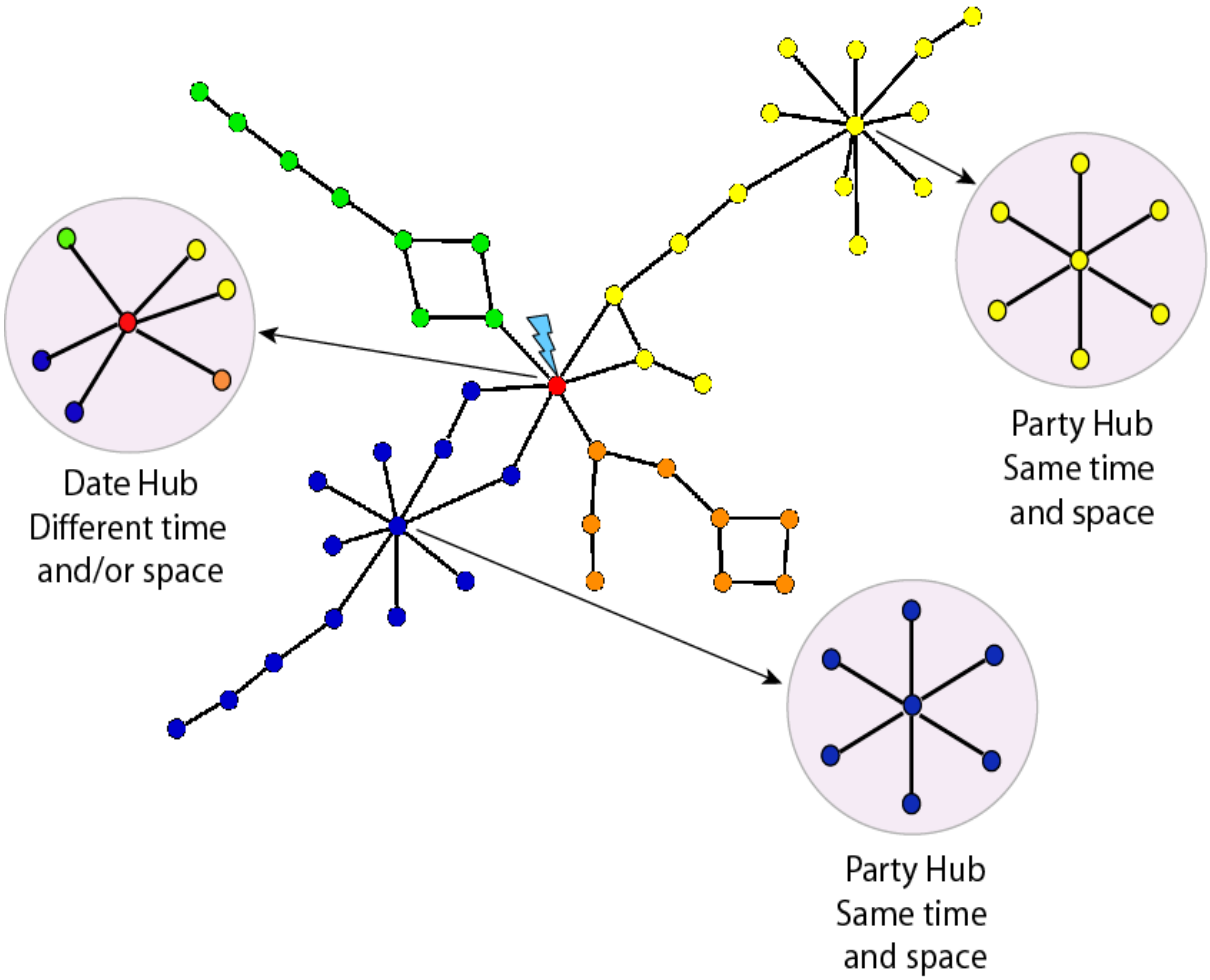
Co-expression in different conditions



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

Are all hubs equal?

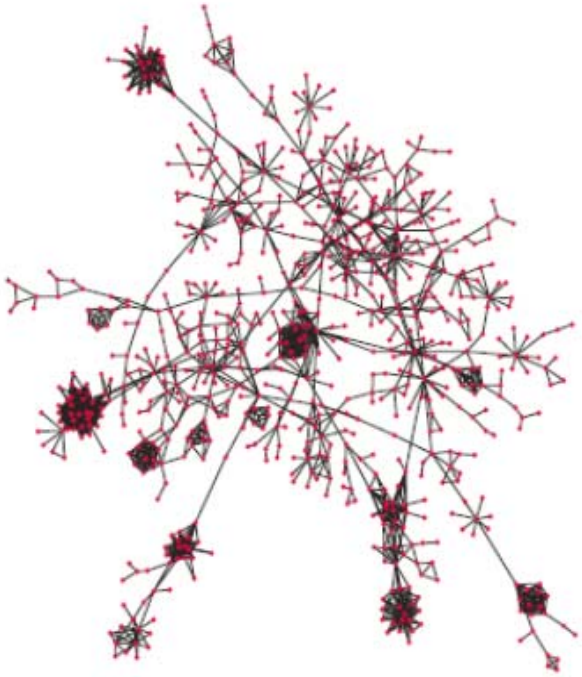
Dynamic or temporal aspects of interactome networks



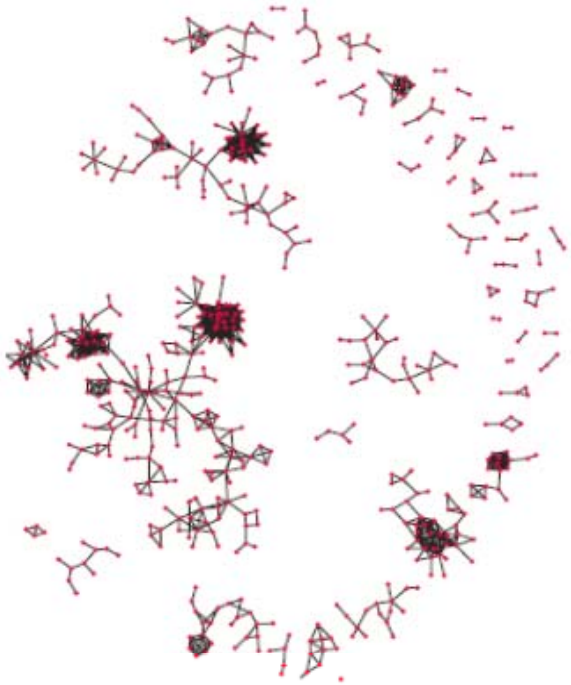
Han *et al.* Nature, 2004

Their Role in the Net

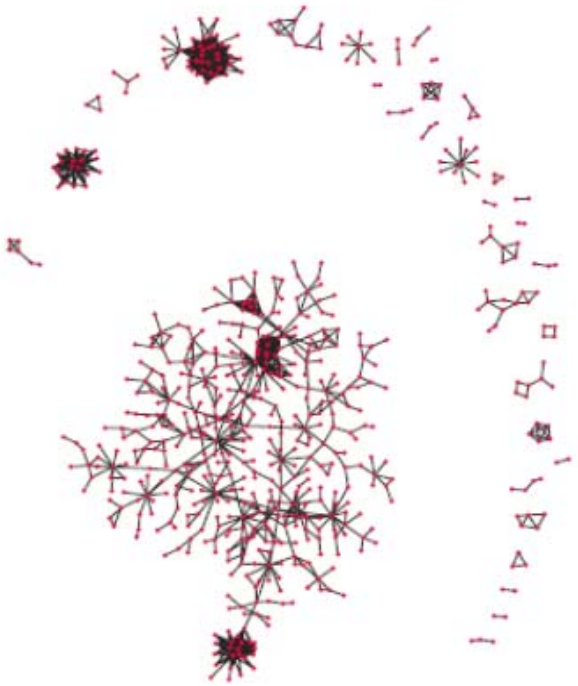
Full Net



No Date Hubs

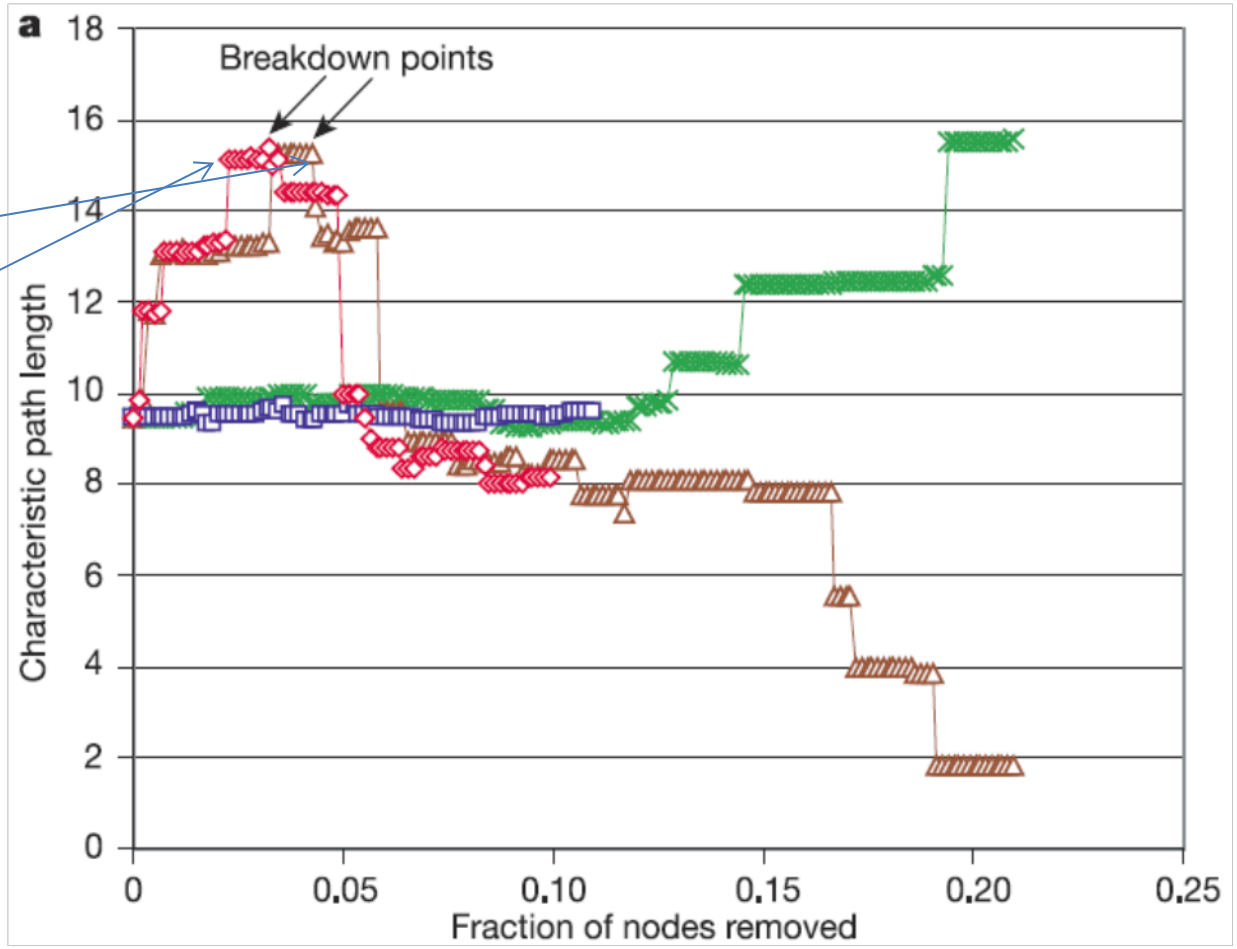


No Party Hubs



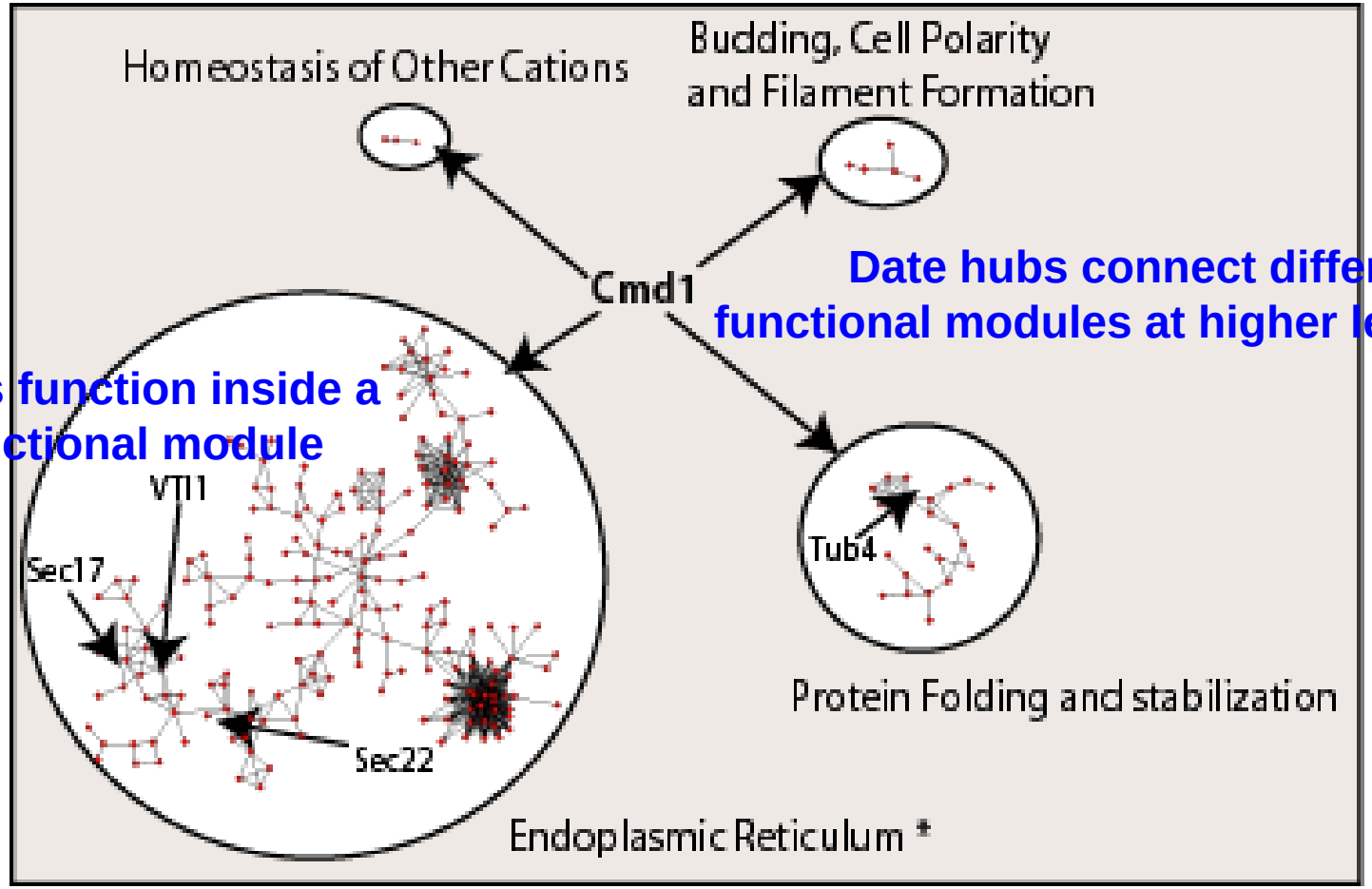
In silico simulation of node removal

-- 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

OPEN ACCESS Freely available online

PLoS one

Hubs with Network Motifs Organize Modularity Dynamically in the Protein-Protein Interaction Network of Yeast

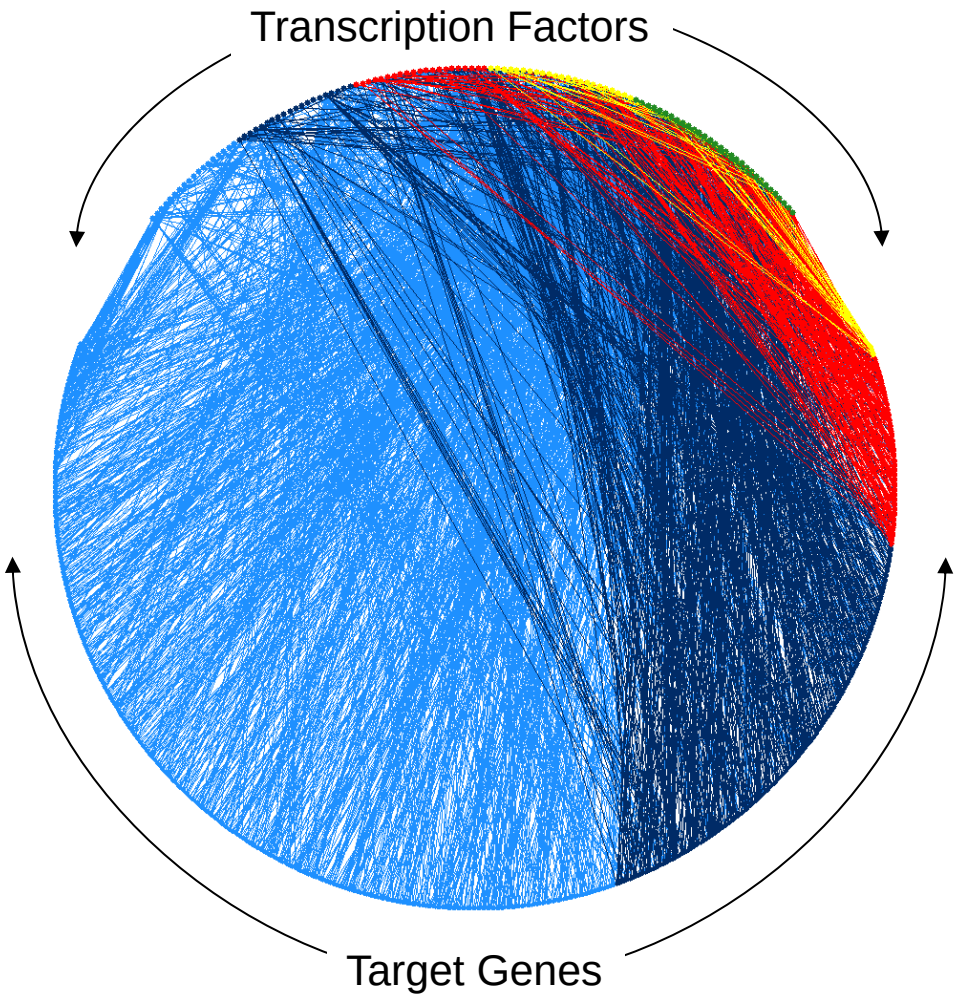
Guangxu Jin[1,2], Shihua Zhang[1,2], Xiang-Sun Zhang[1]*, Luonan Chen[3,4,5,6]*

1 Academy of Mathematics and Systems Science, Chinese Academy of Science, Beijing, China, 2 Graduate University of 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, 5 Exploratory Research for Advanced Technology (ERATO) Aihara Complexity Modelling Project, Japan Science and Technology Corporation (JST), Tokyo, Japan, 6 Institute of Industrial Science, The University of Tokyo, Tokyo, Japan

Background. It has been recognized that modular organization pervades biological complexity. Based on network analysis, 'party hubs' and 'date hubs' were proposed to understand the basic principle of module organization of biomolecular networks. However, recent study on hubs has suggested that there is no clear evidence for coexistence of 'party hubs' and 'date hubs'. Thus, an open question has been raised as to whether or not 'party hubs' and 'date hubs' truly exist in yeast interactome. ***Methodology.*** In contrast to previous studies focusing on the partners of a hub or the individual proteins around the hub, our work aims to study the network motifs of a hub or interactions among individual proteins including the hub and its neighbors. Depending on the relationship between a hub's network motifs and protein complexes, we define two new types of hubs, 'motif party hubs' and 'motif date hubs', which have the same characteristics as the original 'party hubs' and 'date hubs' respectively. The network motifs of these two types of hubs display significantly different features in spatial distribution (or cellular localizations), co-expression in microarray data, controlling topological structure of network, and organizing modularity. ***Conclusion.*** By virtue of network motifs, we basically solved the open question about 'party hubs' and 'date hubs' which was raised by previous studies. Specifically, at the level of network motifs instead of individual proteins, we found two types of hubs, motif party hubs (mPHs) and motif date hubs (mDHs), whose network motifs display distinct characteristics on biological functions. In addition, in this paper we studied network motifs from a different viewpoint. That is, we show that a network motif should not be merely considered as an interaction pattern but be considered as an essential function unit in organizing modules of networks.

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* (2004)]

Gene expression data

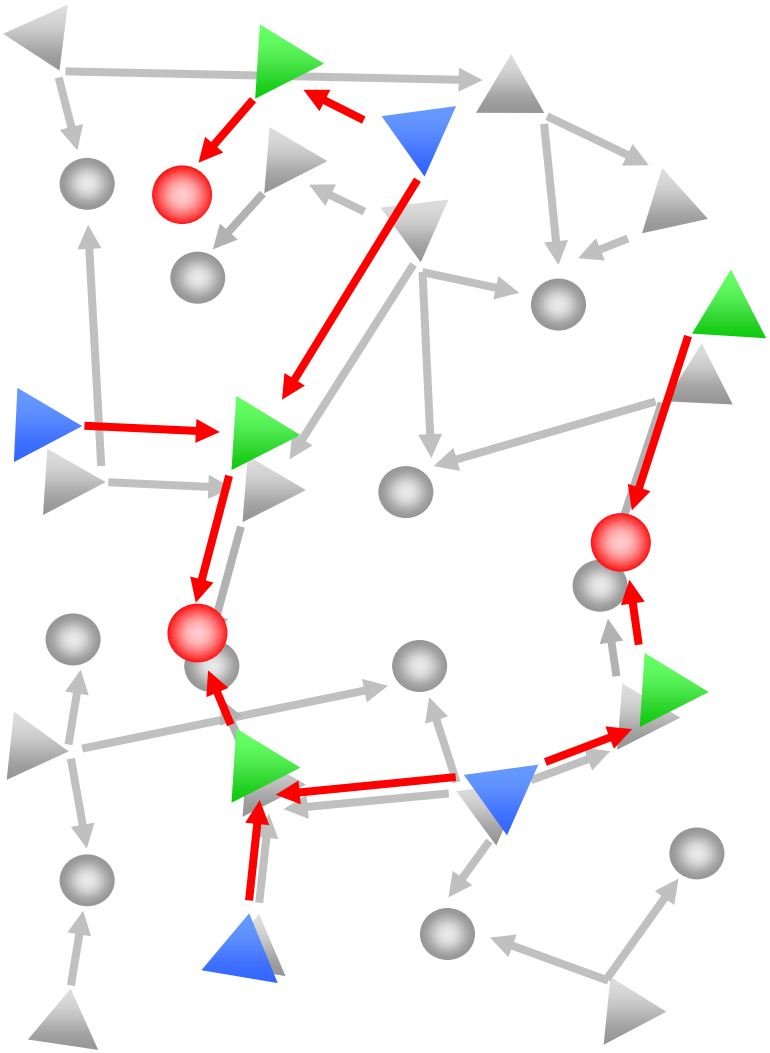
- Genes that are **differentially expressed** under five cellular conditions

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



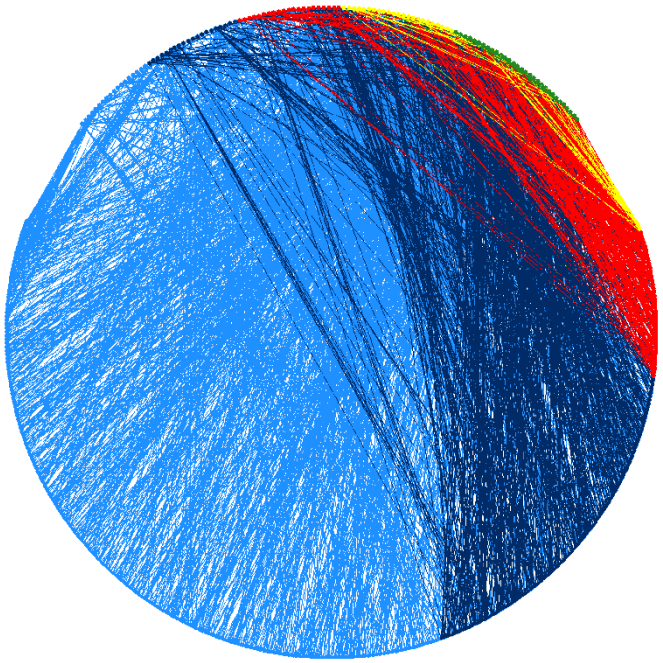
- Define differentially expressed genes
- Identify TFs that regulate these genes
- Identify further TFs that regulate these TFs

Active regulatory sub-network

[Luscombe et al, *Nature*, 2004]

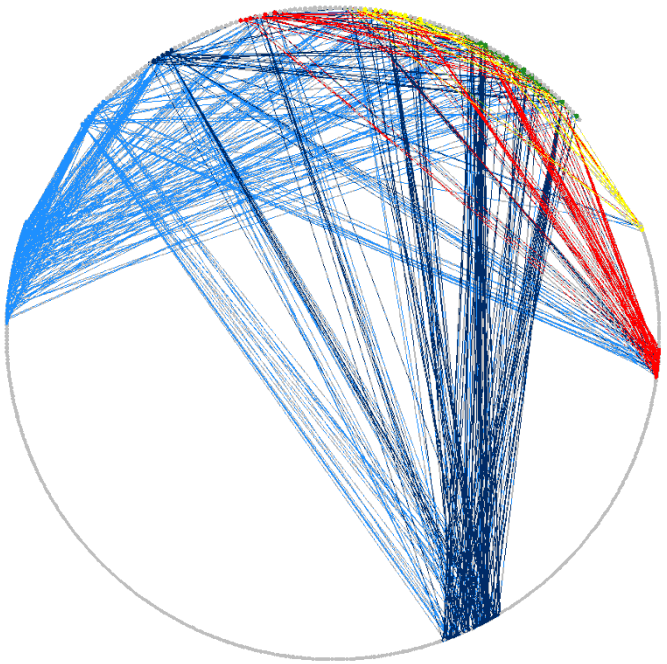
Network usage under cell cycle

complete network



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

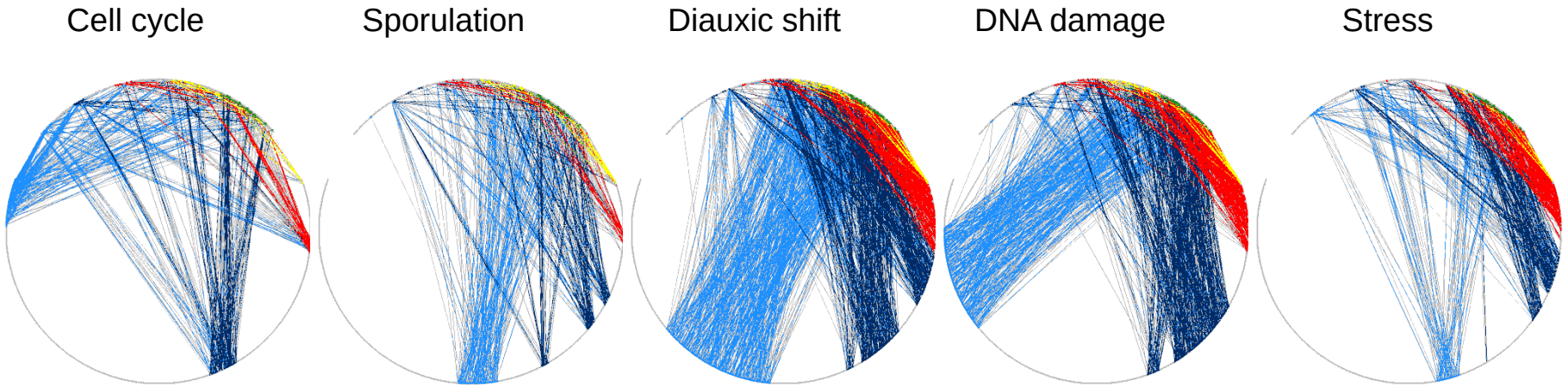
cell cycle sub-network



- 70 TFs
- 280 genes
- 550 interactions

[Luscombe et al, *Nature*, 2004]

Network usage under different conditions



How do the networks change?

- topological measures
- network motifs

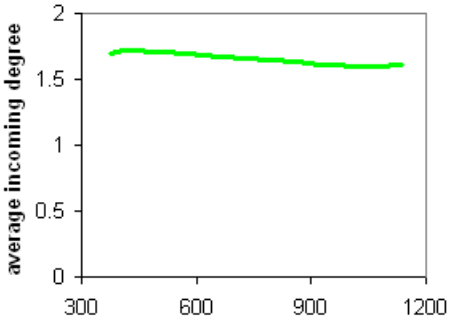
[Luscombe et al, *Nature*, 2004]

Our expectation

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

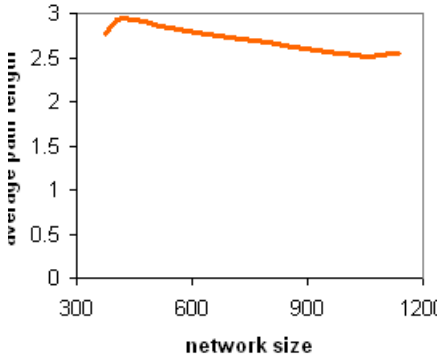
- Random expectation: Sample different size sub-networks from complete network and calculate topological measures

incoming degree

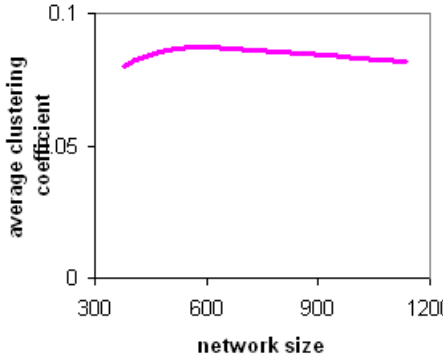


random network size

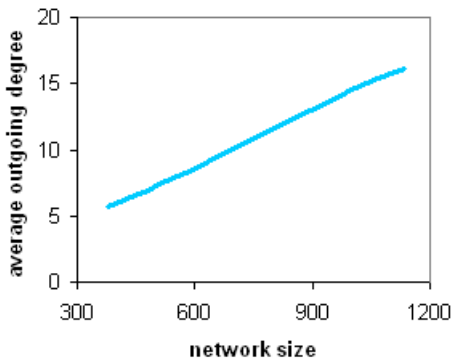
path length



clustering coefficient



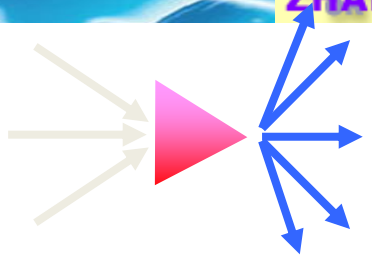
outgoing degree

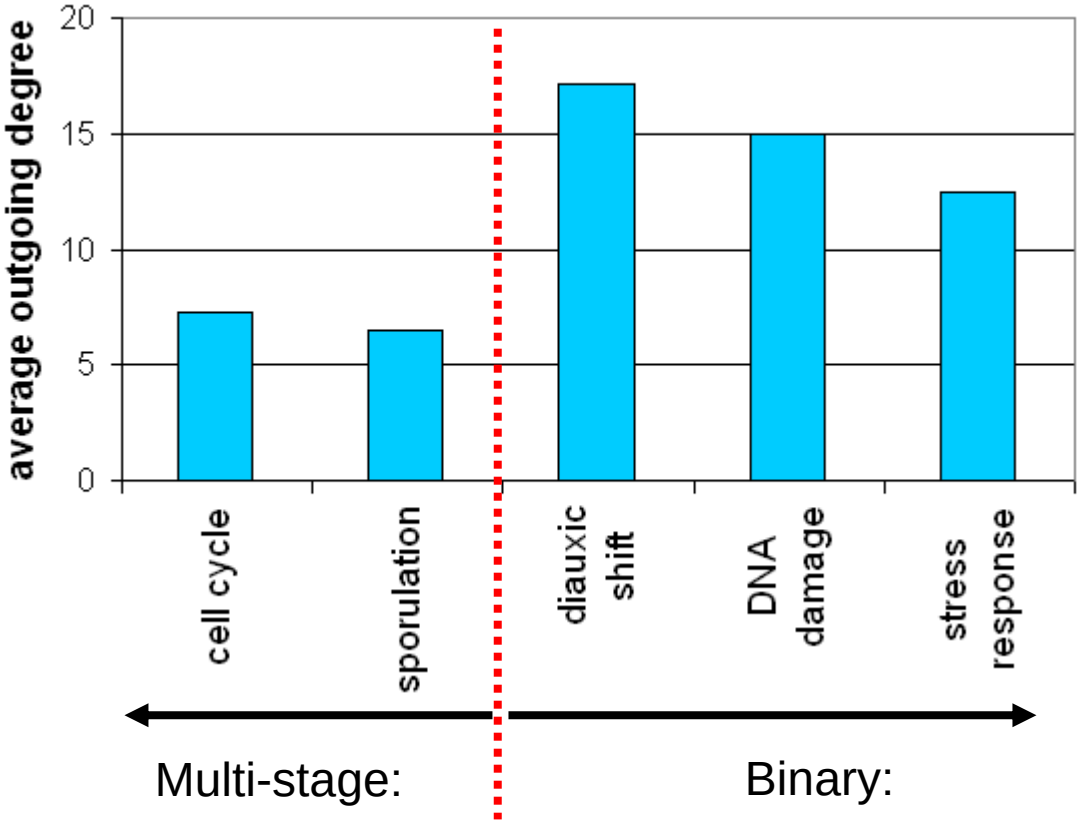


Measures should remain constant

[Luscombe et al, Nature, 2004]

Outgoing degree

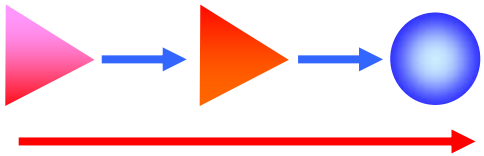


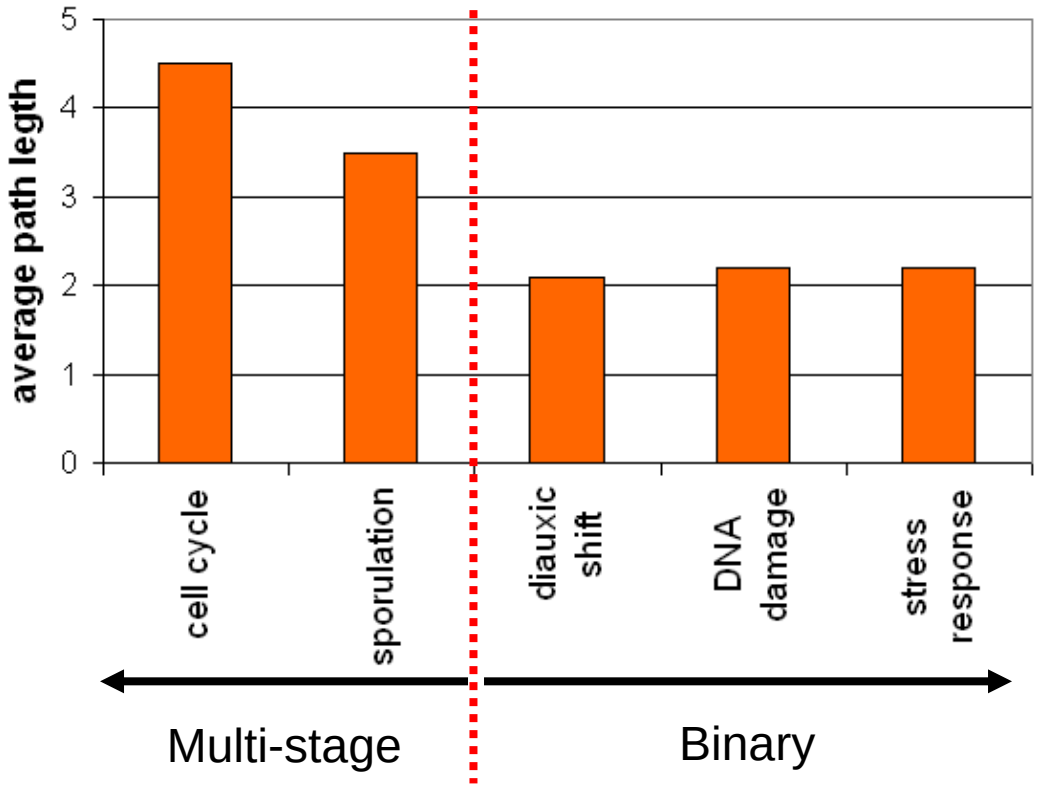


- "Binary conditions"
  → greater connectivity
- "Multi-stage conditions"
  → lower connectivity

[Luscombe et al, Nature, 2004]

Path length

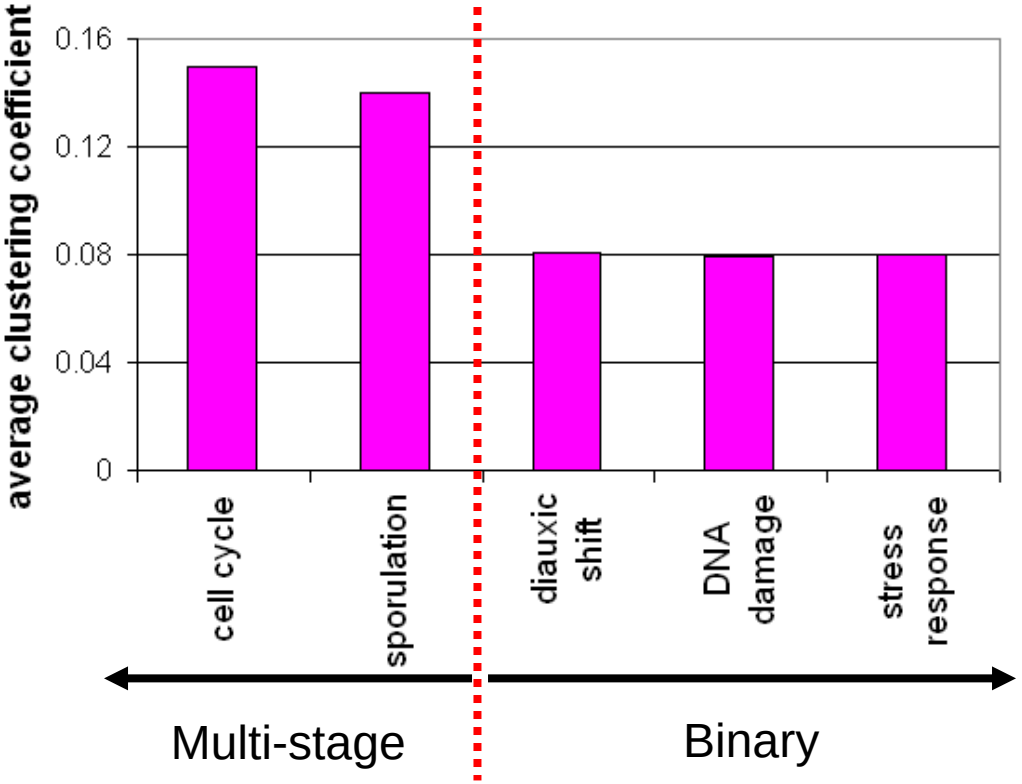




- "Binary conditions"
  - → shorter path-length
  - → "faster", direct action
- "Multi-stage" conditions
  - → longer path-length
  - → "slower", indirect action

[Luscombe et al, Nature, 2004]

Clustering coefficient



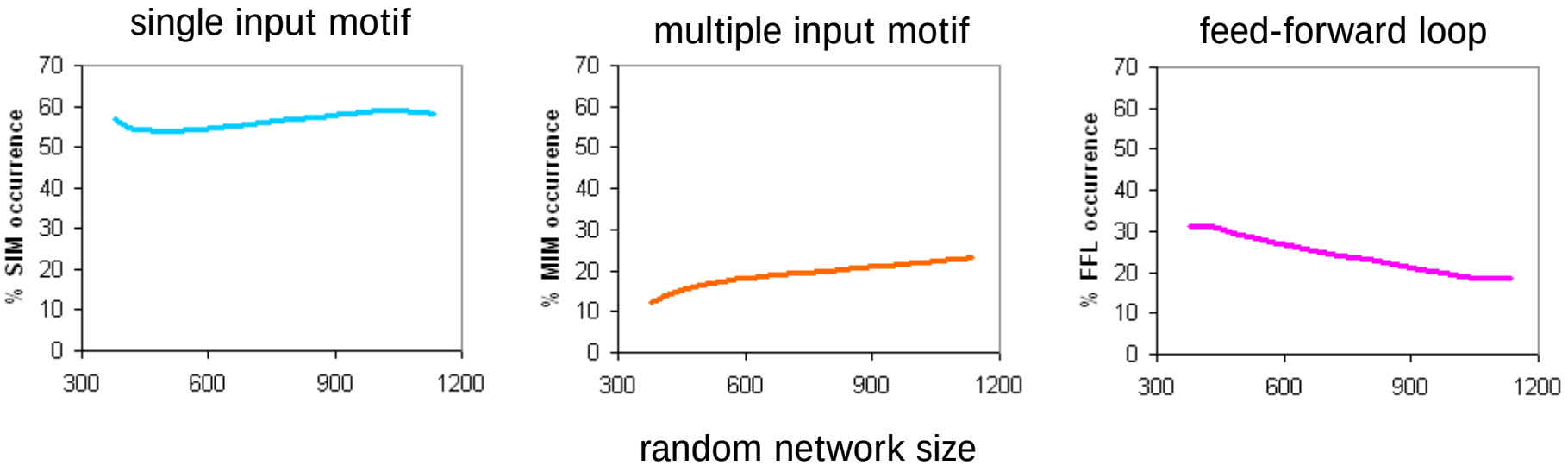
- "Binary conditions"
  → smaller coefficients
  → less TF-TF inter-regulation
- "Multi-stage conditions"
  → larger coefficients
  → more TF-TF inter-regulation

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

Our expectation

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

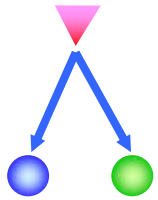
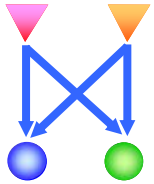
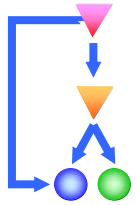
- Random expectation: sample sub-networks and calculate motif occurrence



Motif usage should remain constant

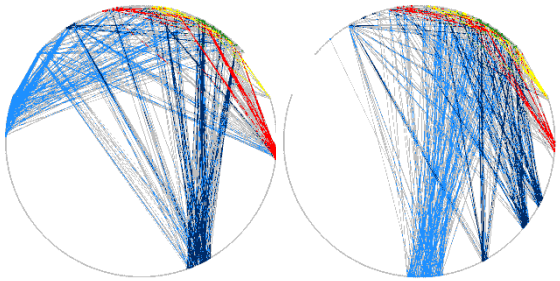
[Luscombe et al, Nature, 2004]

Network motifs

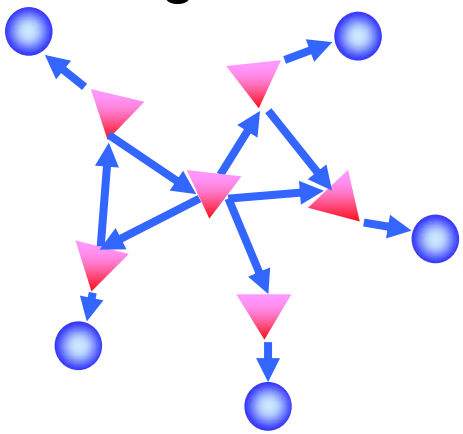
Motifs		Cell cycle	Sporulation	Diauxic shift	DNA damage	Stress response
SIM		32.0%	38.9%	57.4%	55.7%	59.1%
MIM		23.7%	16.6%	23.6%	27.3%	20.2%
FFL		44.3%	44.5%	19.0%	17.0%	20.7%

[Luscombe et al, Nature , 2004]

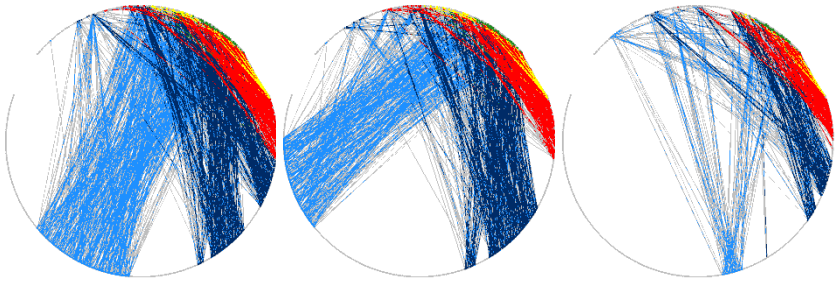
Summary of sub-network structures



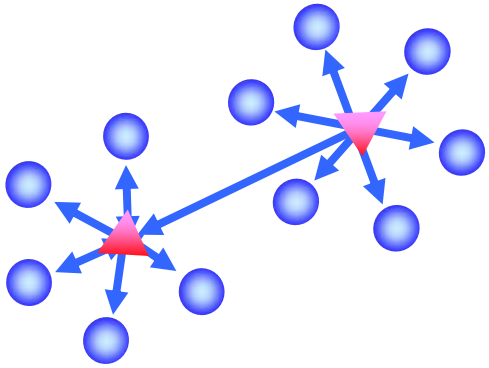
multi-stage conditions



- 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,* 2004]

- The End!