The Abdus Salam
International Centre for Theoretical Physics

United Nations
Educational, Scientific
and Cultural Organization

International Atomic
Energy Agency

World Year of PHYSICS 2005

SMR.1656 - 1

School and Workshop on Structure and Function of Complex Networks

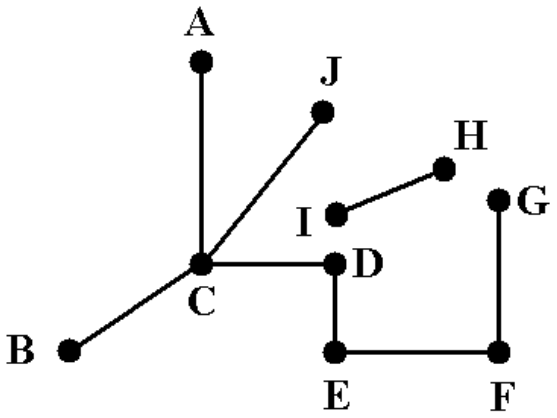
16 - 28 May 2005

The Structure of Biological Networks

Réka ALBERT
Department of Physics &
Huck Institute for the Life Sciences
Pennsylvania State University
122 Davey Laboratory
University Park, PA 16802
U.S.A.

The structure of biological networks

Réka Albert
Department of Physics and
Huck Institutes for the Life Sciences
Pennsylvania State University

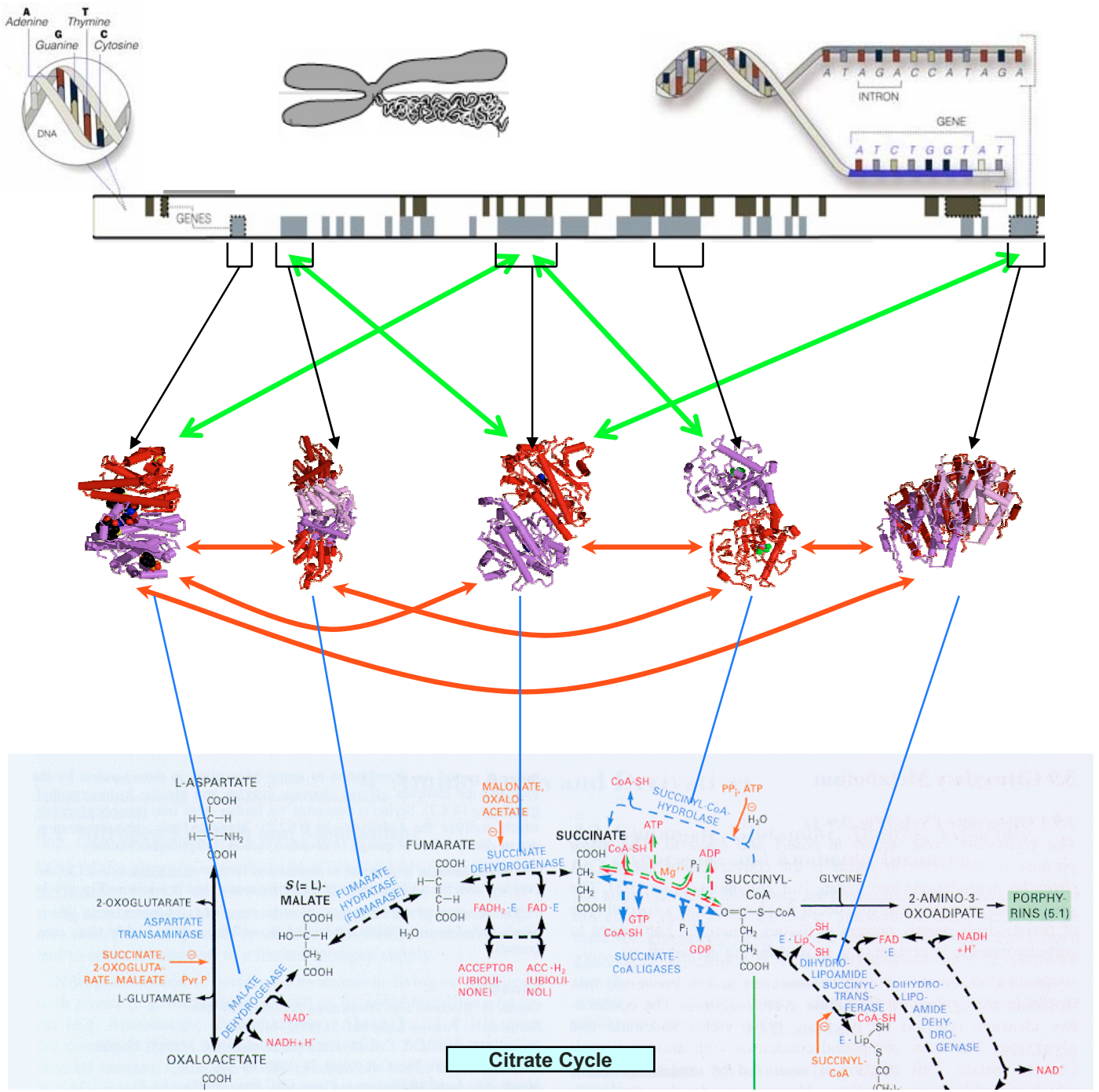


Network (graph): a set of nodes connected pairwise by edges

To be able to construct and analyze a cellular network, we need to clearly define what we identify as a node and what we represent with an edge.

The nodes and edges have to be at least similar to each other, e.g. represent the same type of cellular component (protein, chemical) or the same type of interaction (mass transfer, regulation).

Modifications possible: different types of nodes and edges, edge weights



GENOME

protein-gene interactions

PROTEOME

protein-protein interactions

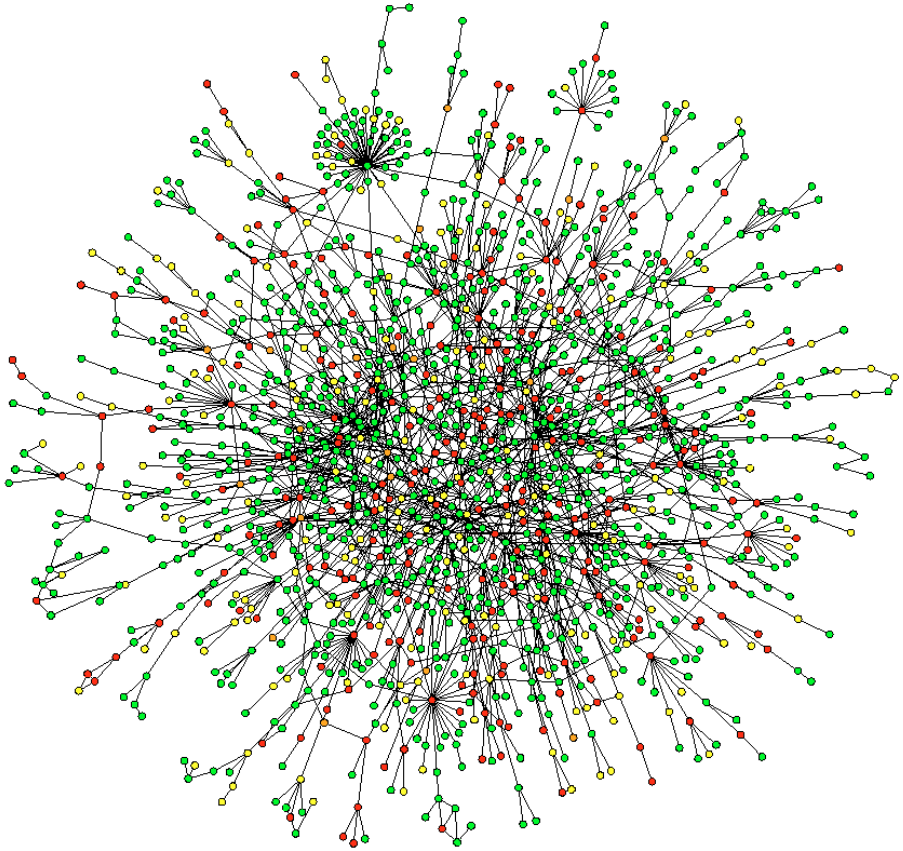
METABOLISM

Bio-chemical reactions

1. Protein interaction networks

nodes: proteins

edges: protein-protein interactions (binding)



Map of yeast protein-protein interactions, by Hawoong Jeong

Red: essential protein
Yellow: growth- affecting protein
Green: non-essential protein

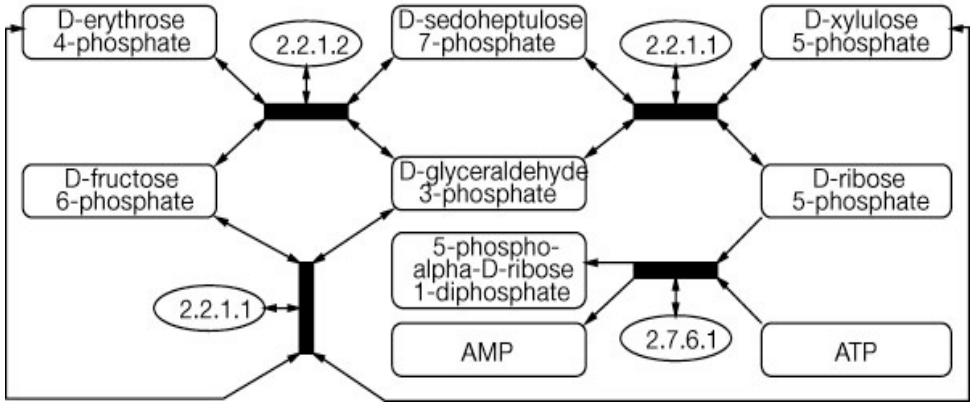
2. Biochemical reaction networks

Several types of nodes

- reactants (substrates) or products of the reactions
- enzymes – catalyze the reactions
- reactant-enzyme complex (“reaction node”)

Edges reflect reactions or catalysis (regulation)

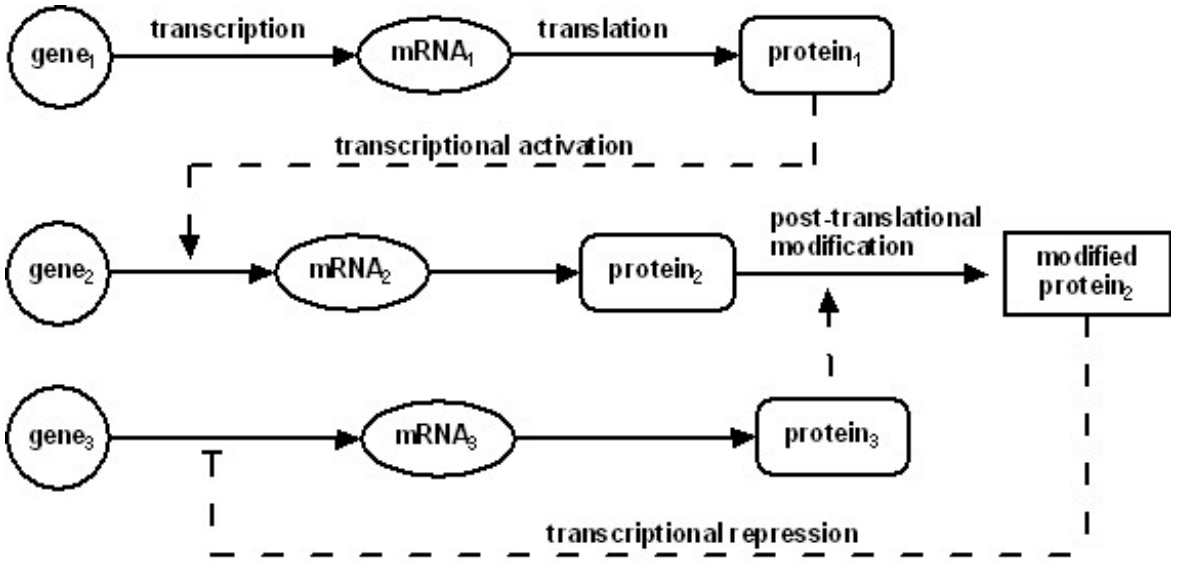
- one possibility: directed edges from reactants/enzymes to complex, from complex to products/enzyme



3. Gene regulatory networks

At least two types of nodes: mRNA , protein
Edges: mass flow (continuous) or regulation (dashed)
Regulatory edges acting on edges – similar to catalysis
Edges can be activating or inhibiting.

Transcription factor protein – DNA interaction represented as regulation, or protein- mRNA directed edge

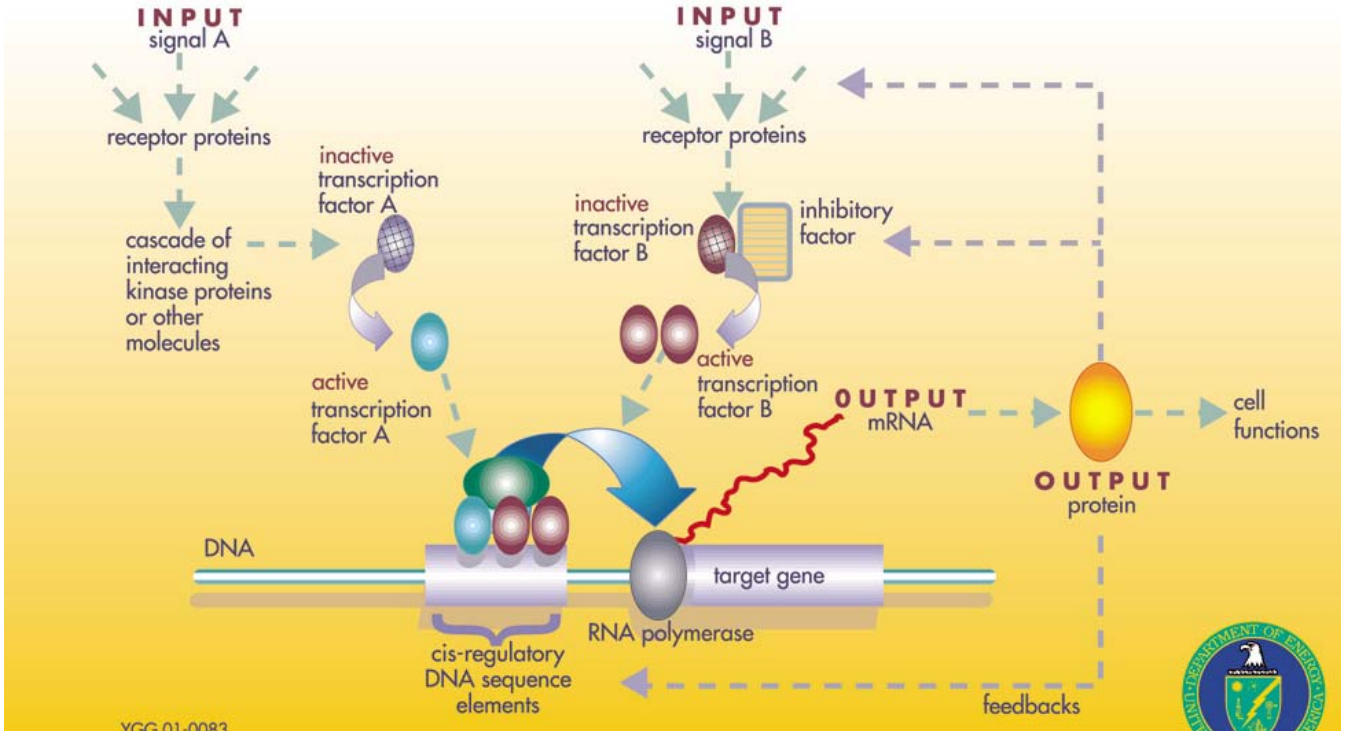


4. Signal transduction networks

Nodes: proteins, molecules

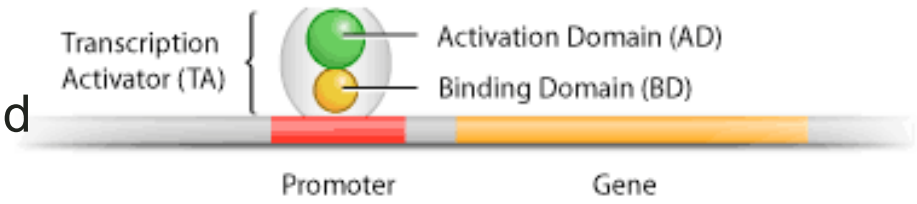
Edges: reactions and processes (e.g. ligand/receptor binding protein conformational changes); common to all is that they reflect information transfer

Signal transduction networks have defined inputs and outputs.

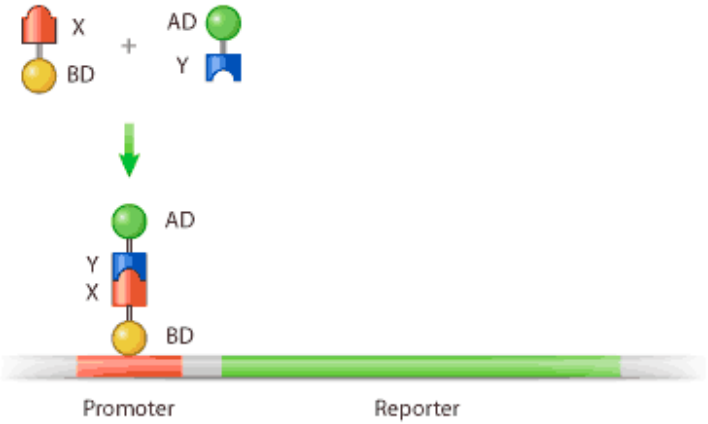


1. Protein-protein interactions are identified on the genomic level by using the yeast two-hybrid method

Transcription factors bind to the promoter regions of genes.
They have a DNA binding domain and an activation domain.



In the two-hybrid method the two domains are separated, and fused to two proteins.
If the two proteins interact by binding, the transcription factor activates the expression of a reporter gene.

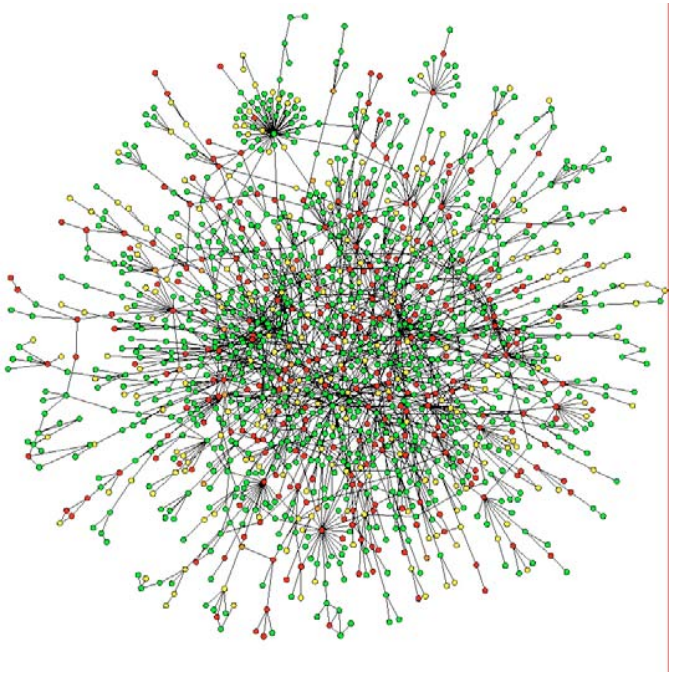


Systematic experiments with all proteins in a given organism lead to genome-wide protein interaction maps.

Protein interaction maps now contain thousands of nodes and edges

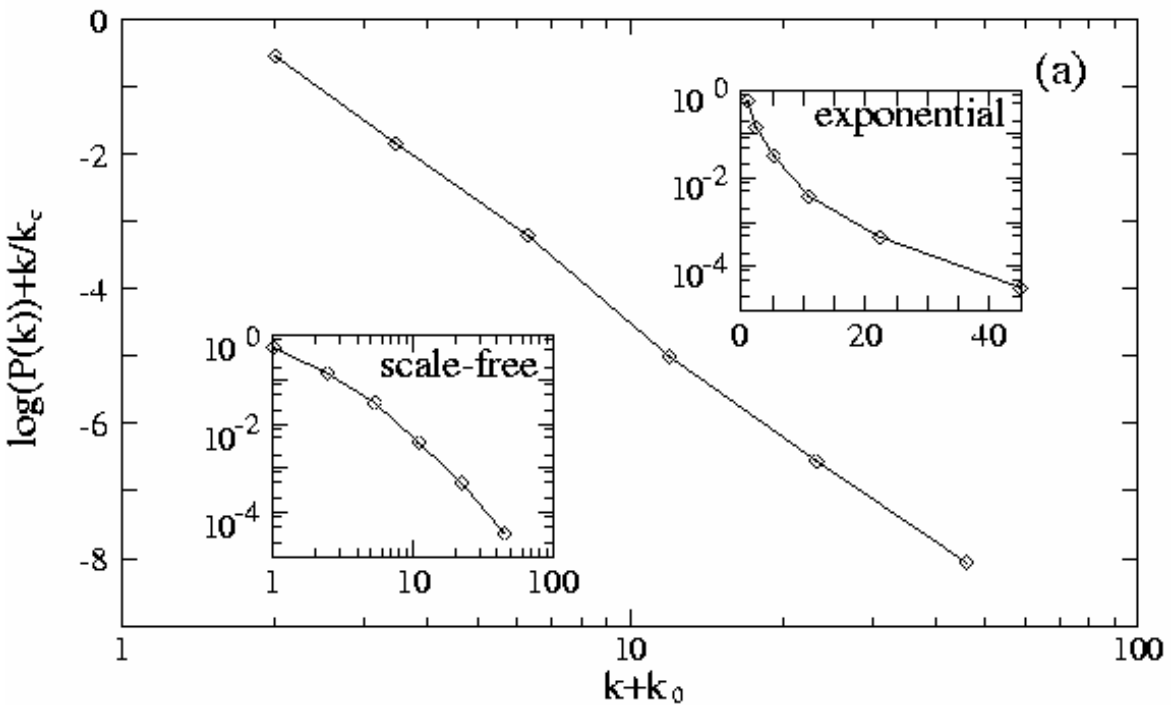
Ito (yeast): 8868 interactions between 3280 proteins
Uetz (yeast): 4480 interactions, 2115 proteins
Giot (Drosophila): 4780 interactions among 4679 proteins
Li (C. elegans): 5534 interactions, 3024 proteins

- Although usually tested in a given bait/prey setting, protein interactions are considered symmetrical
- All networks have giant connected components.
- The topological properties of diverse protein interaction networks are similar.



H. Jeong et al.Nature 411, 41-42 (2001)
S.-H Yook, Z.N. Oltvai, A.-L. Barabasi, Proteomics 4, 928 (2004)

Degree distribution of the yeast protein network is a power law with exponential cutoff

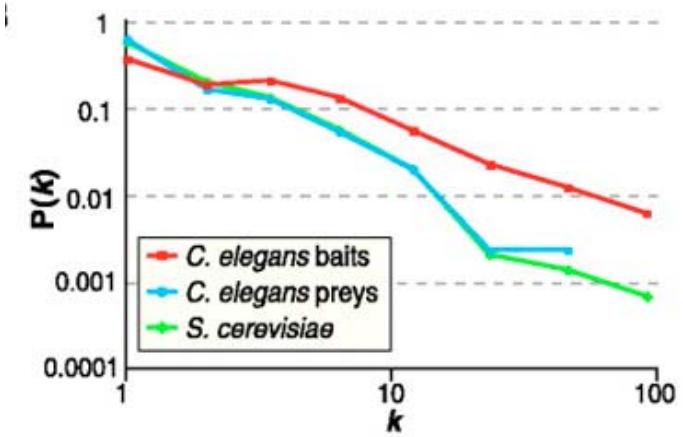


$$P(k) \sim (k + k_0)^{-\gamma} \exp(-\frac{k + k_0}{k_\tau})$$

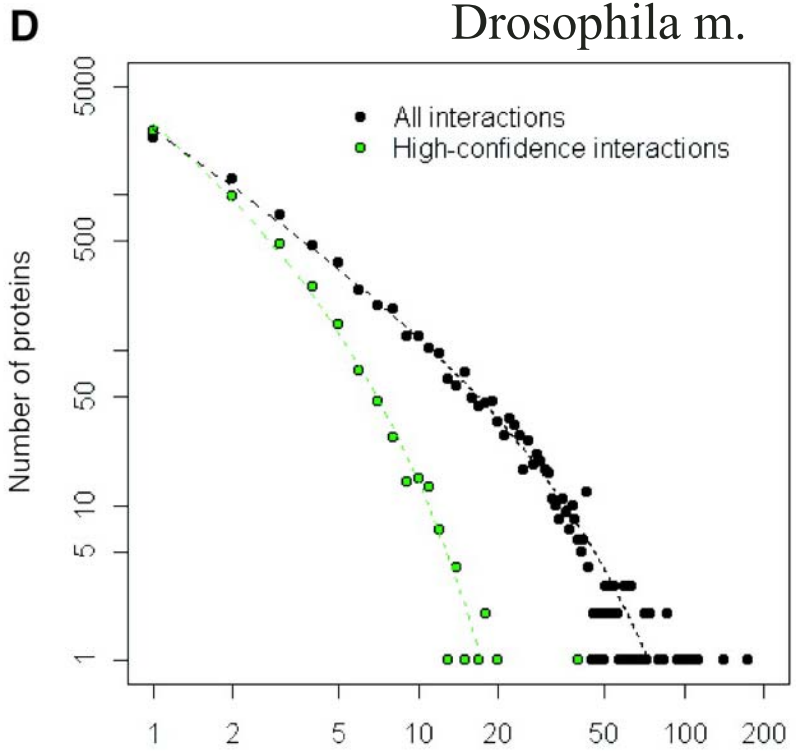
H. Jeong, S.P. Mason, A.-L. Barabasi, Z.N. Oltvai, Nature 411, 41-42 (2001)

Degree distribution of C. elegans and D. melanogaster protein networks

C. elegans



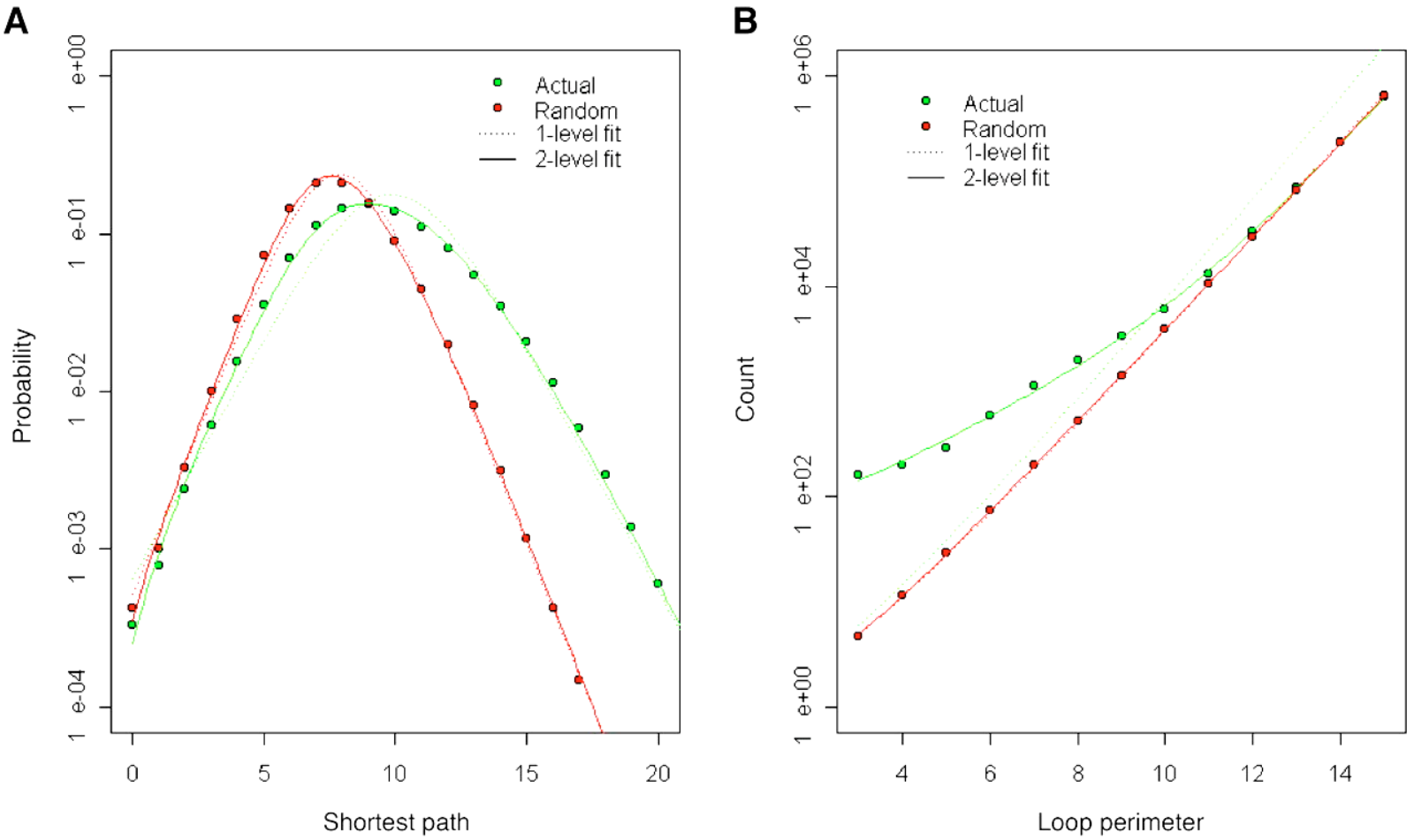
Drosophila m.



$$P(k) = Ak^{-\gamma} \exp(-\beta k)$$

The degree distribution gets closer to a power-law as more interactions are mapped.

Average path length larger, short cycles more abundant than in randomized networks

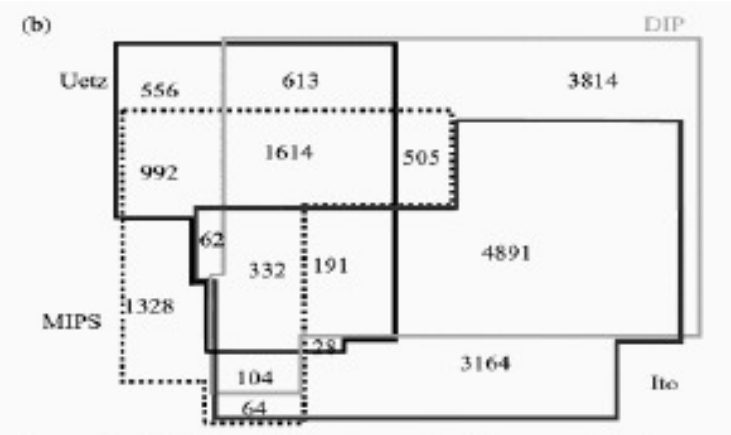


Randomization: swap the endpoints of two edges, node degrees stay the same.

The bad news: protein interaction maps are far from perfect

- Protein interaction networks are incomplete – false negatives

- Little overlap (~7%) between maps constructed by different labs
- Est. coverage of Drosophila map is 21%, for C. elegans it is 10%

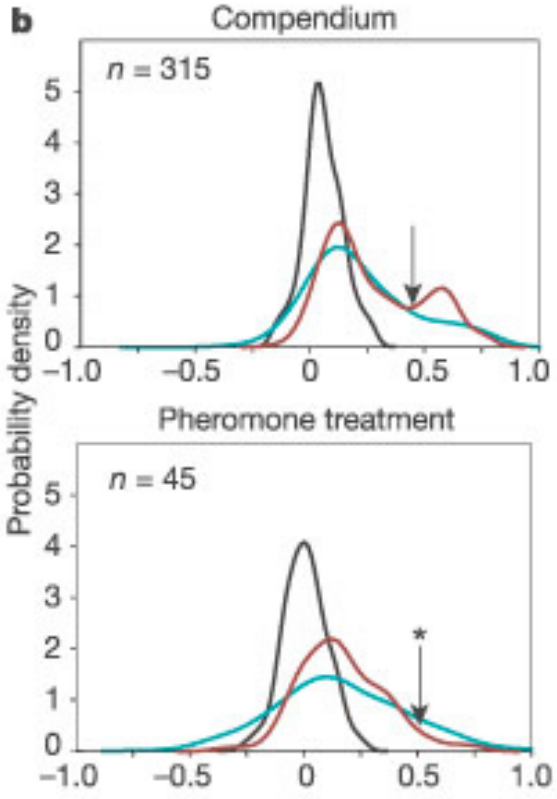


- A significant percentage (~20%) of interactions observed by the two-hybrid method are not biologically relevant - false positives

- Independent verification of interactions needs be done by alternative methods such as co-immunoprecipitation or co-affinity purification pull-down assays.

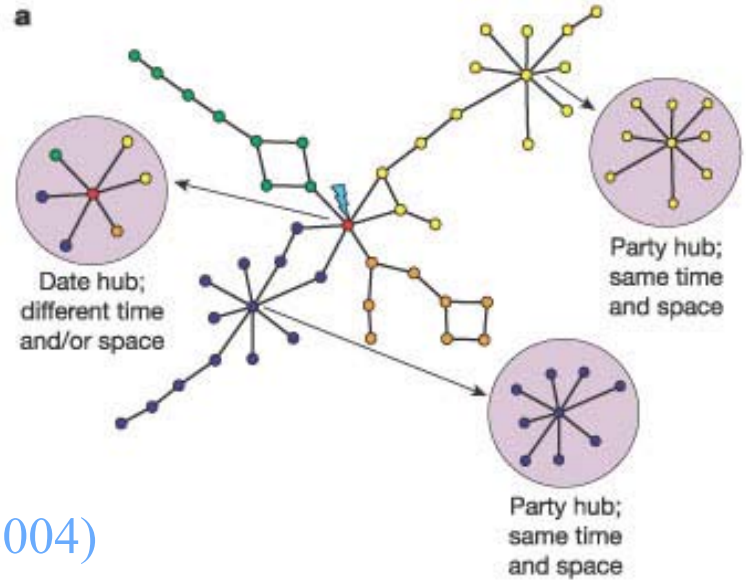
- These methods are small scale and slow, thus there is a need for prediction methods able to give a short list of candidates.

Not all observed interactions are simultaneously active



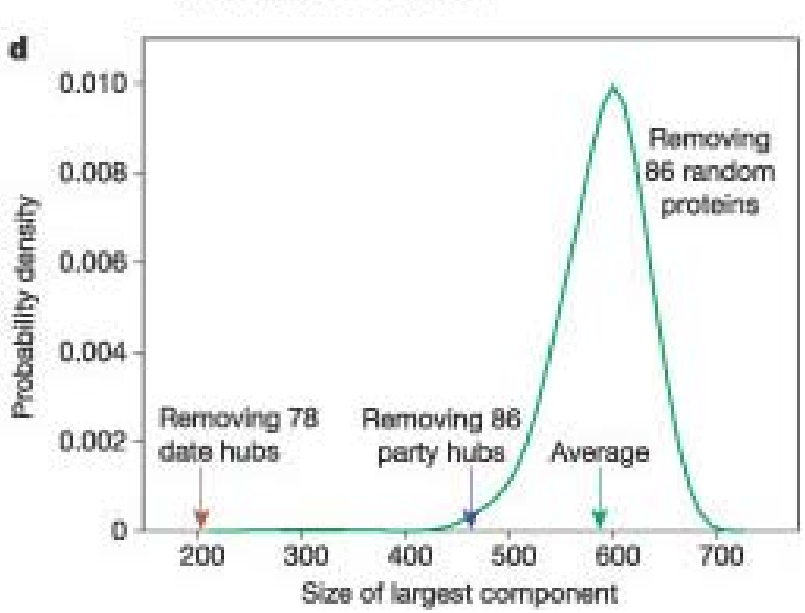
Calculate the correlation between the genes encoding the first neighbors of hub proteins.

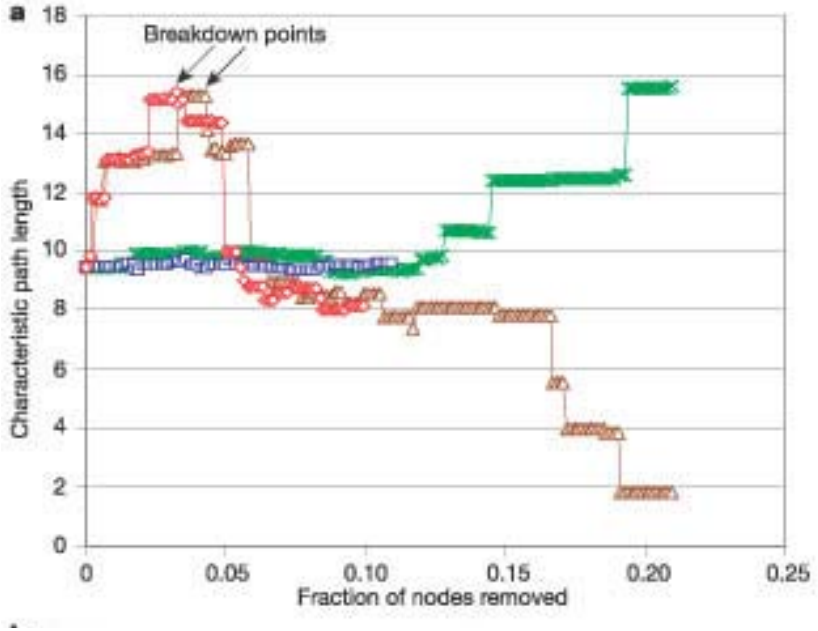
Two peaks – two different types of hubs.



Han et al, Nature 443, 88 (2004)

Loss of date hubs much more severe than loss of party hubs



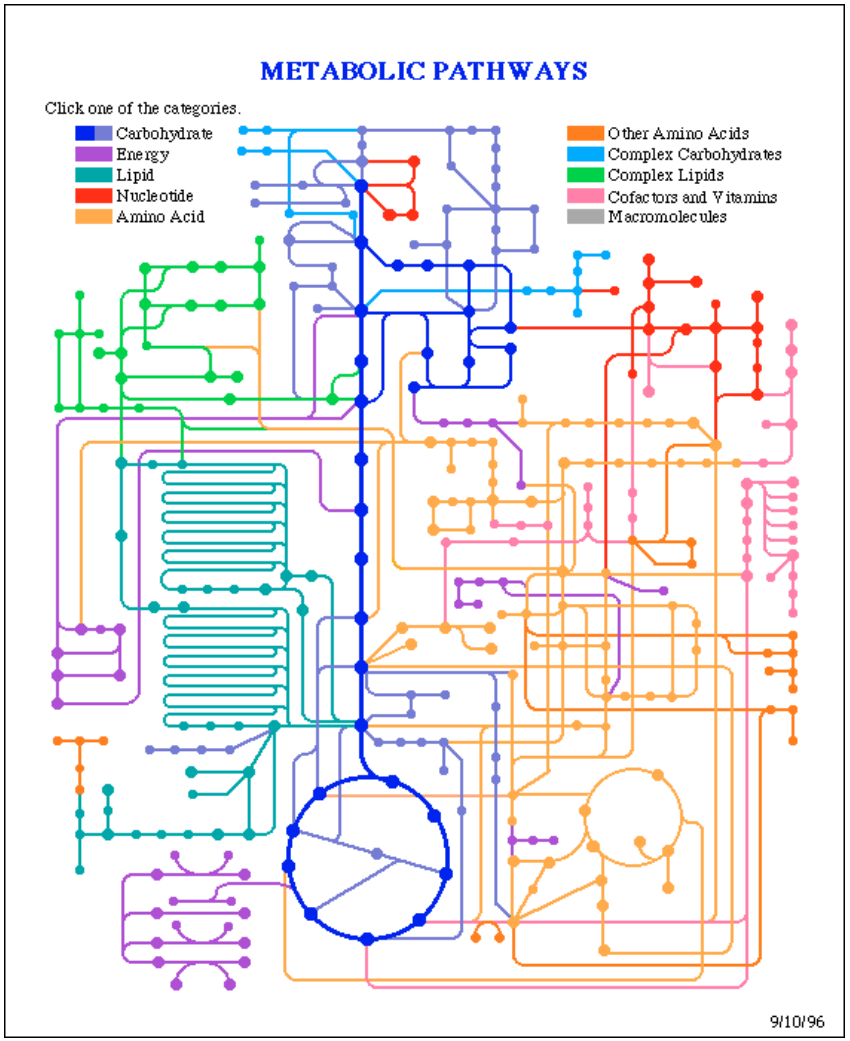


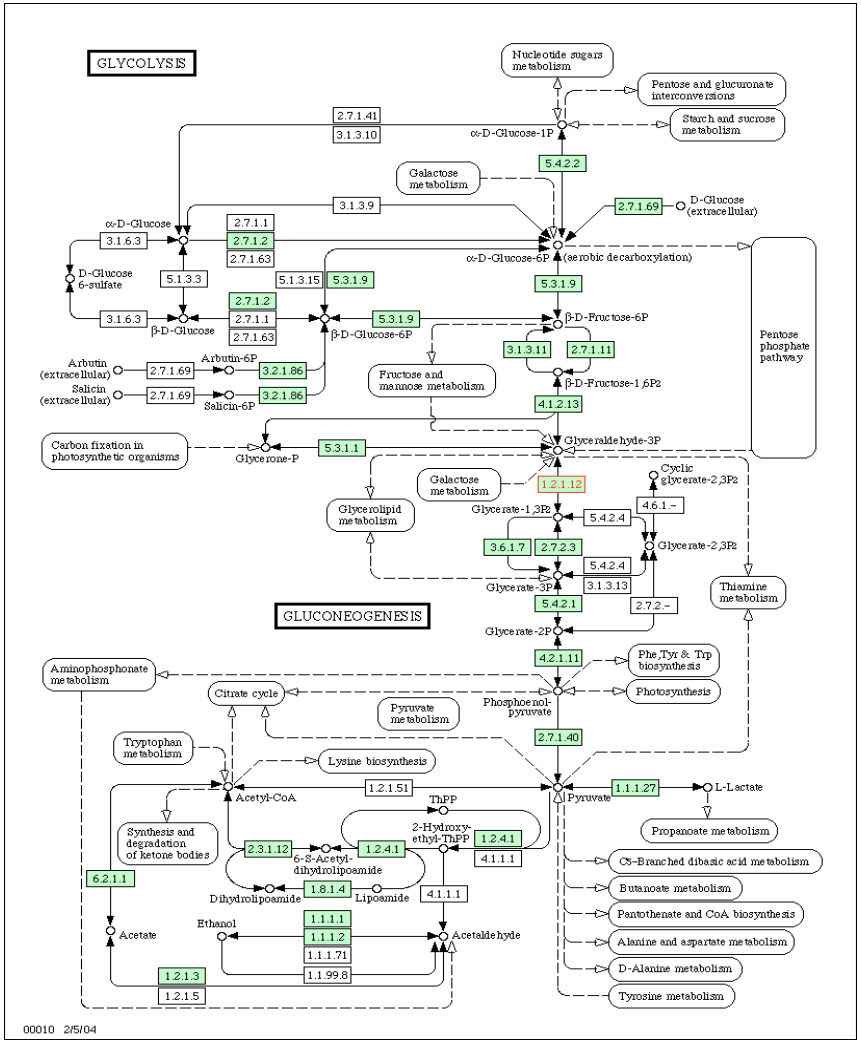
random node removal
preferential node removal

date hub removal
party hub removal

Party hubs are inside connected modules that interact simultaneously.
Date hubs connect different modules.

Networks of chemical reactions





Metabolism: Sum of chemical processes by which energy is stored or released.

Reaction Stoichiometry

Reaction Pathway

A + B ⟶ C + D (1)

A + D ⟶ E (2)

B + C ⟶ F (3)

Stoichiometric Matrix (S)

Reactants (Substrates/Metabolites)	Reactions 1	2	3
A	-1	-1	0
B	-1	0	-1
C	1	0	-1
D	1	-1	0
E	0	1	0
F	0	0	1

S_{ij} = Number of molecules of substrate *i* participating in reaction *j*

$S_{ij} < 0$ if substrate *i* is a reactant in reaction *j*

$S_{ij} > 0$ if substrate *i* is a product in reaction *j*

i = 1,2,…,N = # of substrates = # rows

j = 1,2,…,M = # of reactions = # columns

Network Representation – Bi-partite Graph

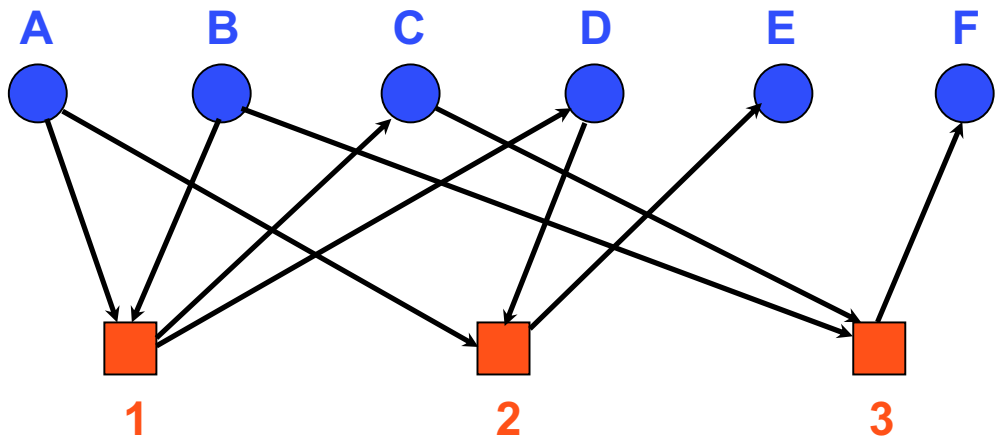
Reaction Pathway

A + B ⟶ C + D (1)

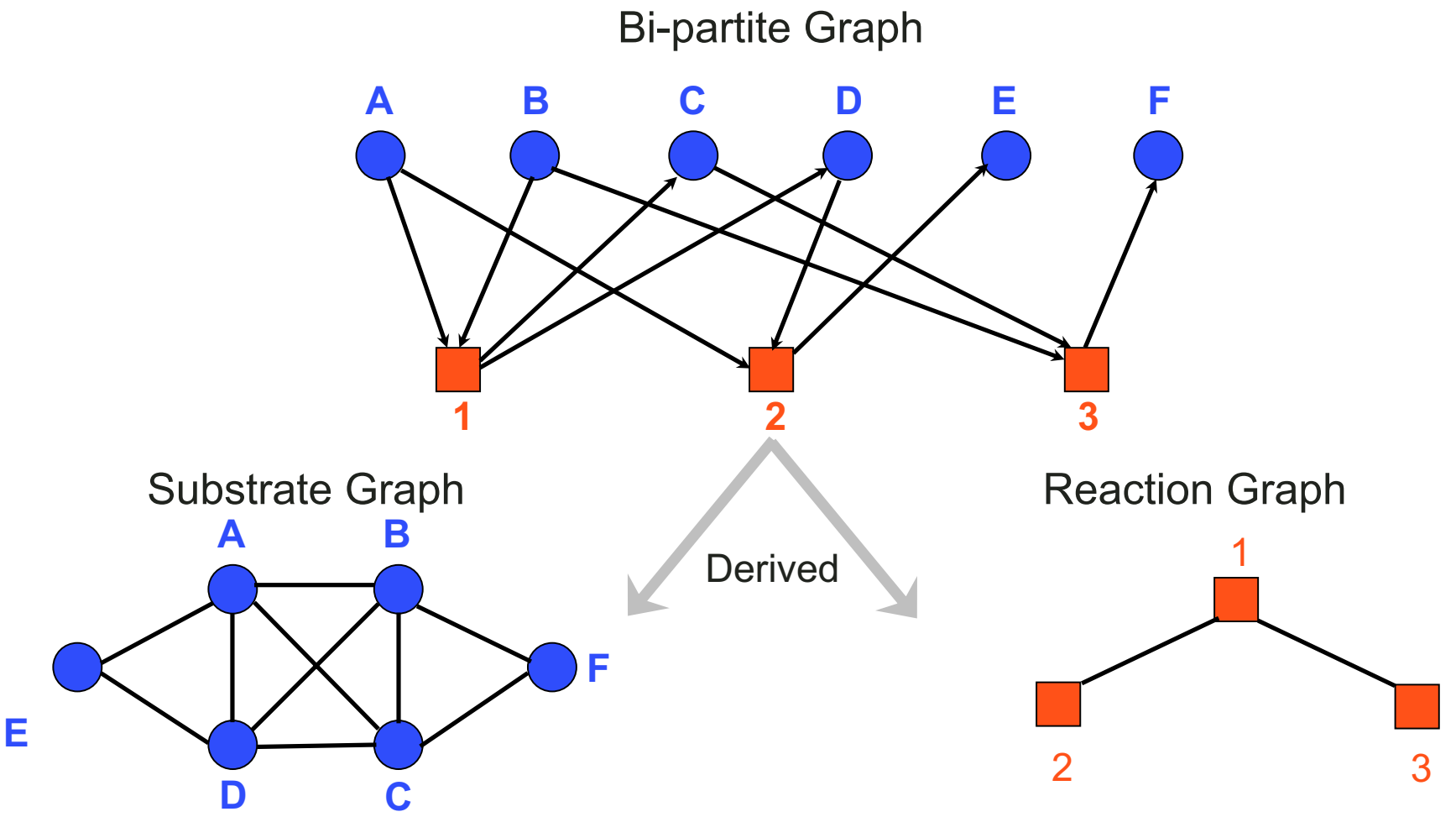
A + D ⟶ E (2)

B + C ⟶ F (3)

Bi-partite Graph (“S-Graph”)



- Two types of nodes:
  - Substrate Node
  - Reaction Node
- Directed arcs
- No direct arcs between nodes of the same type


Bi-partite Graph
A
B
C
D
E
F
1
2
3
Substrate Graph
A
B
E
F
D
C
Derived
Reaction Graph
1
2
3
Connect two substrates if they participate in the same reaction.
Connect two reactions if they share a substrate.

Key Properties of Metabolic Networks

- Metabolic networks are scale-free

P(k) = Probability that a given substrate participates in *k* reactions ≈ $k^{-\gamma}$

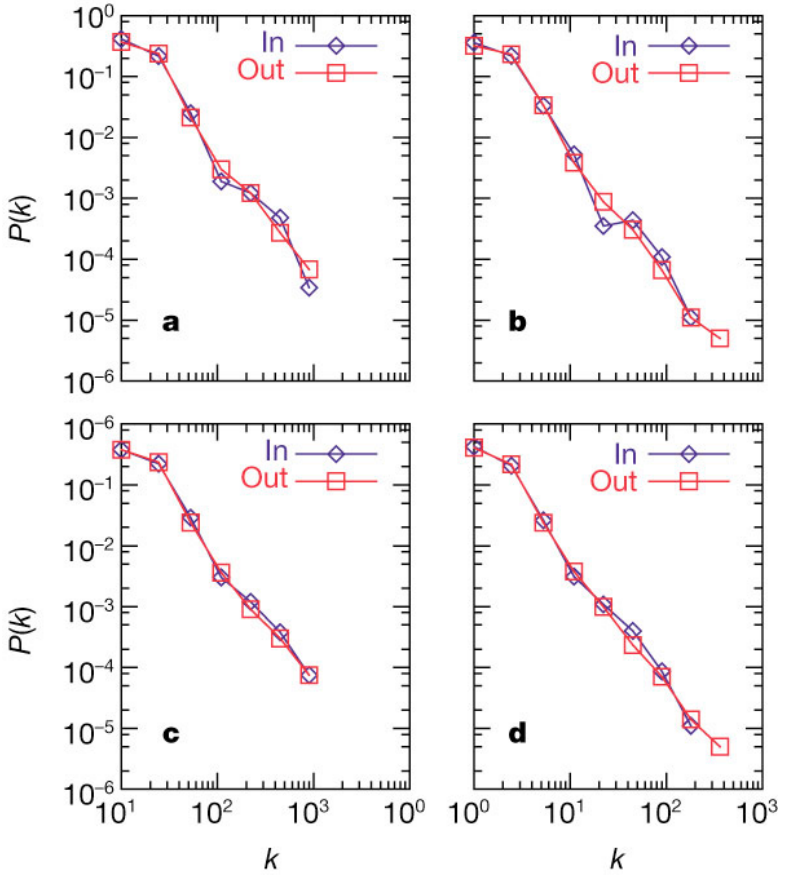
- In- and out- degree of substrate nodes in the bi-partite representation

$$P_{in}(k) \approx k^{-2.2}$$

$$P_{out}(k) \approx k^{-2.2}$$

- Existence of "hub" substrates such as ATP, ADP, NADP, NADPH
  (Carrier Metabolites)

- Relatively small and constant (across organisms) network diameter

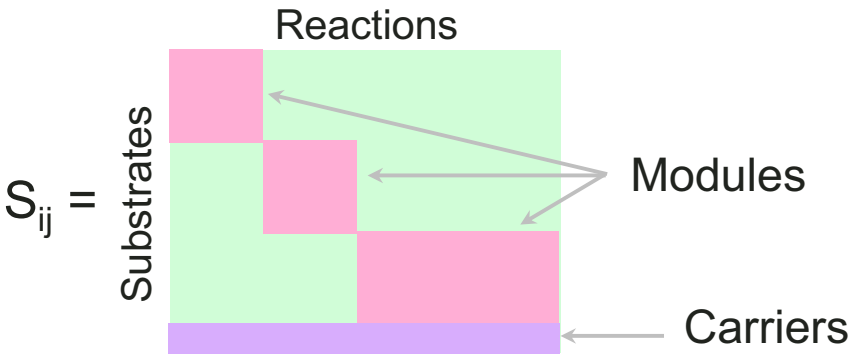


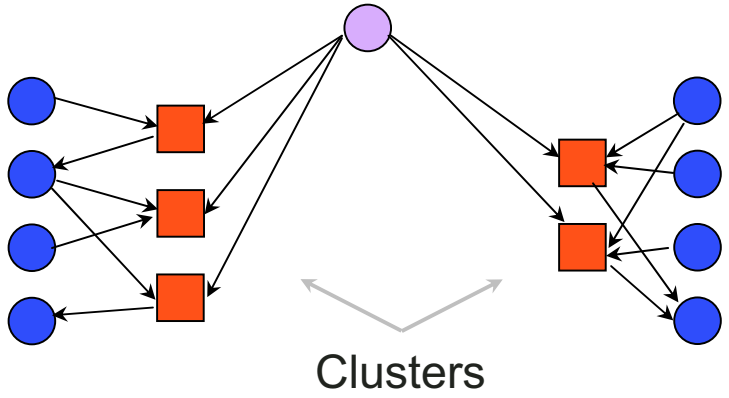
a: *A. fulgidus*
b: *E. coli*
d: *C. elegans*
e: Average (43 organisms)

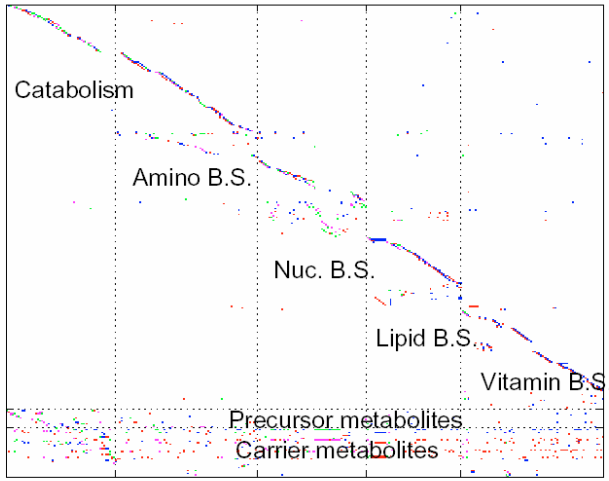
H. Jeong *et al.*, Nature 407, 651 (2000)

Key Properties of Metabolic Networks

- Networks are “modular” in nature

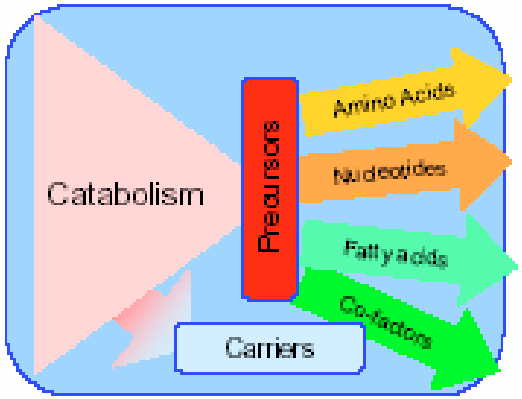






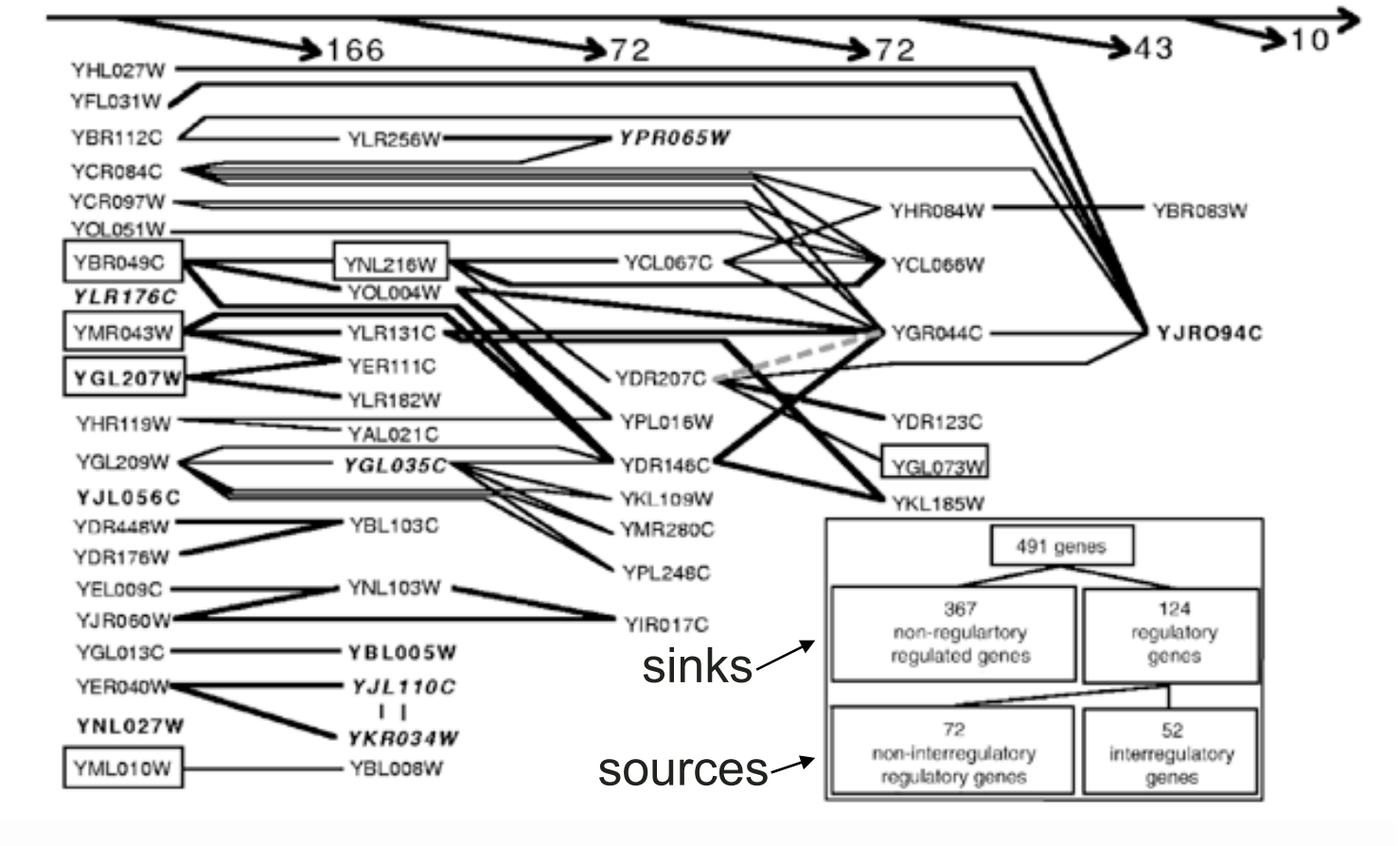
Stoichiometric matrix for *H. pylori*

“Bow-tie” structure of metabolism



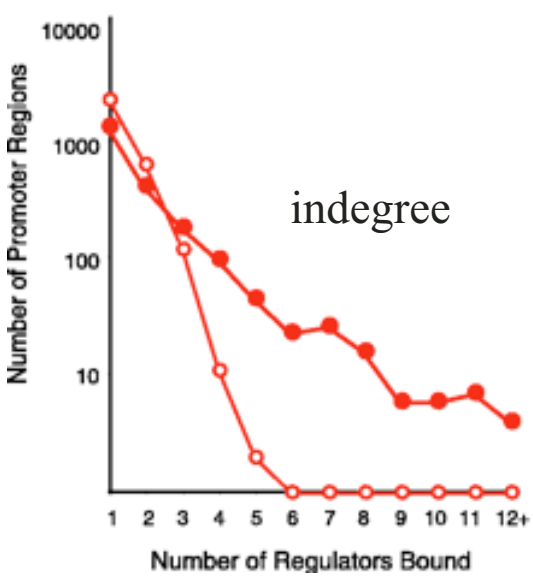
Csete and Doyle, Trends in Biotechnology, 22, 446 (2004)

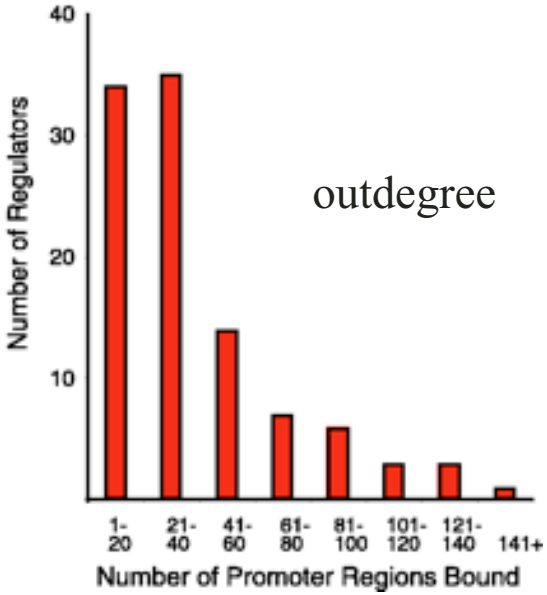
3. Genome-wide transcription networks

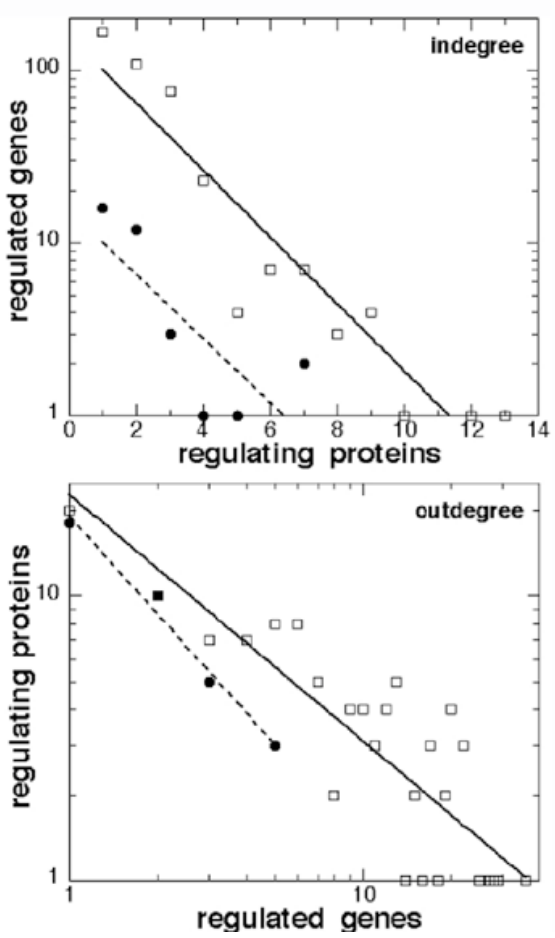


- Contract mRNA and protein into a single node, describe transcriptional regulation as a directed gene-gene edge, thick – activation, thin - inhibition
- Sources: TFs that are not regulated at the transcriptional level
- Sinks: non-TF genes, others are both regulators and regulated

Out-degree distribution long - tailed, in-degree distribution more limited



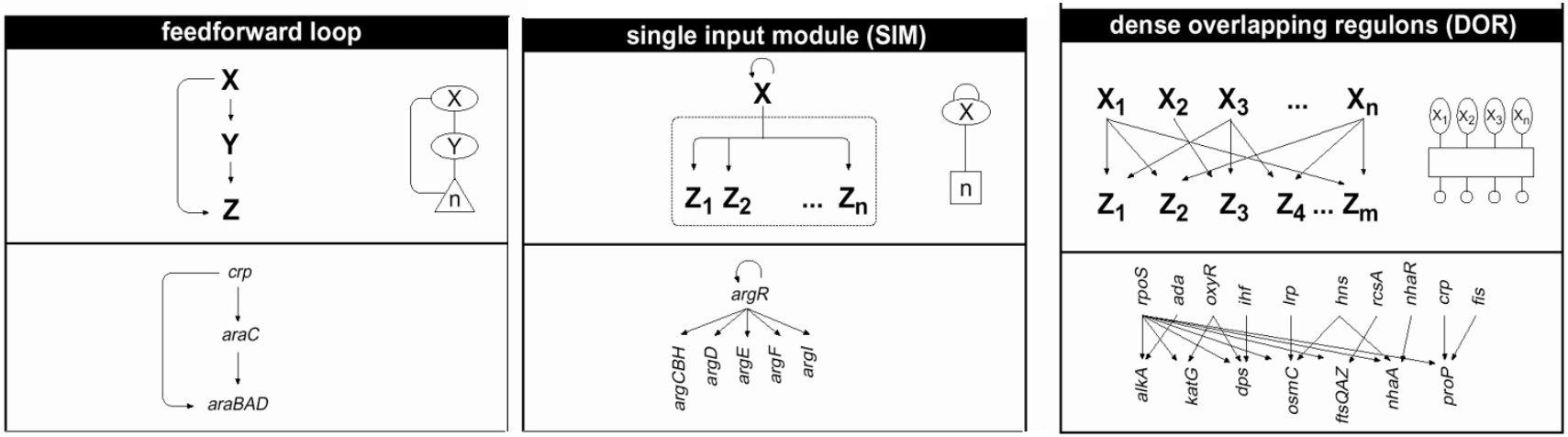




Guelzim et al, Nature Genetics 31, 60 (2002)
Lee et al, Science 298, 799 (2002)

S. cerevisiae

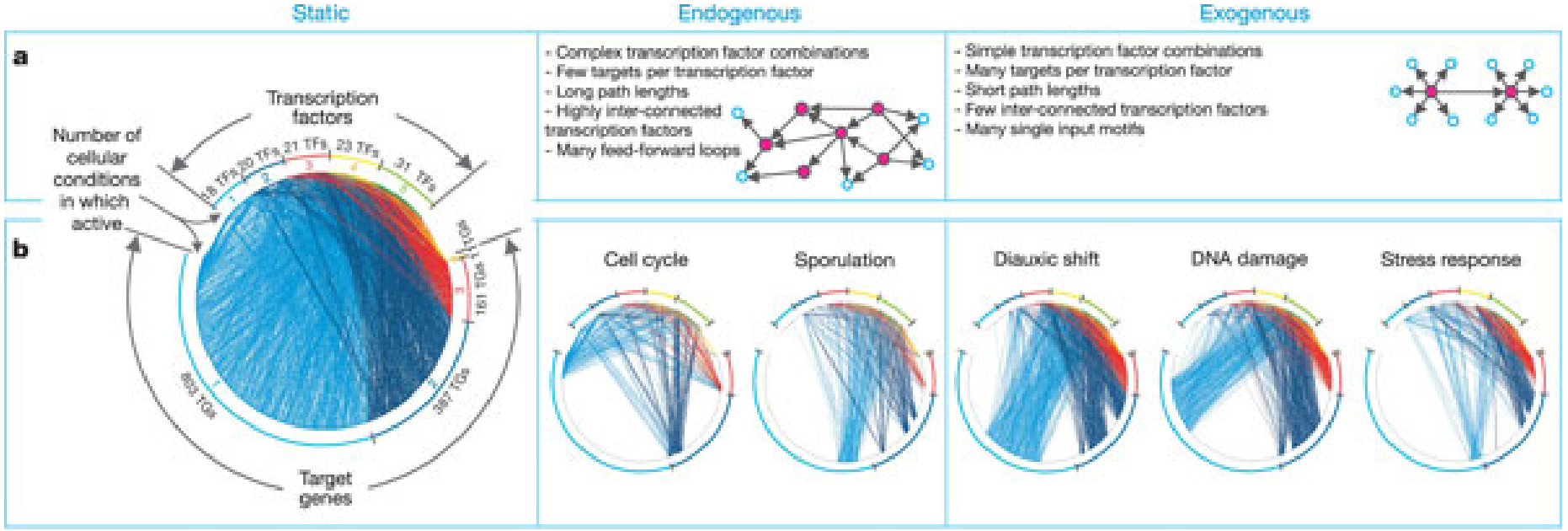
Abundant regulatory motifs



- Feedforward loop: convergent direct and indirect regulation
  possible role: noise filter
- Single input module: one TF regulates several genes
  possible role: temporal program
- Dense overlapping regulons: groups of genes regulated combinatorially

Shen – Orr et al., Nature Genetics (2002)

Condition-dependent transcription sub-networks

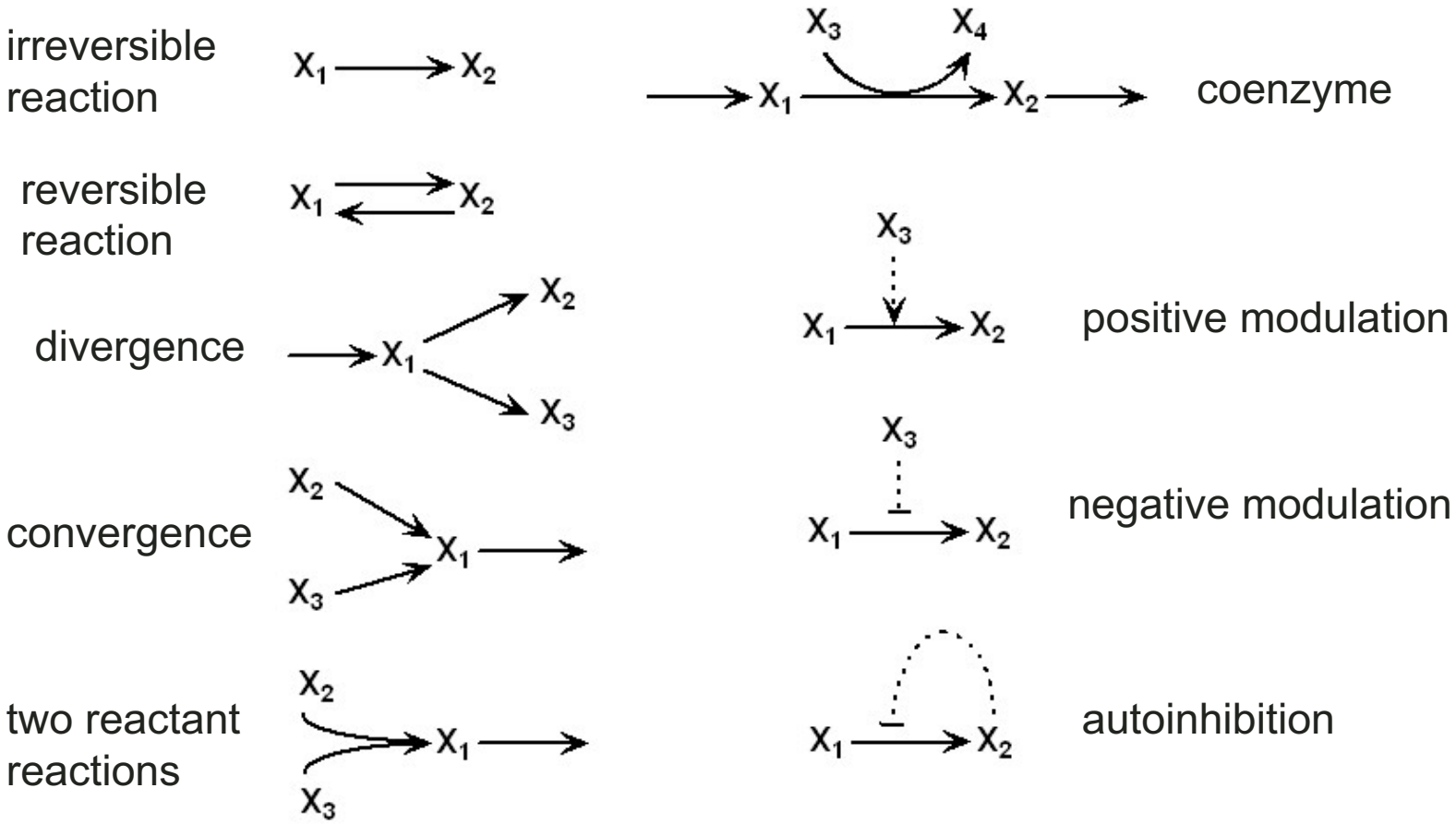


c

		Static		Endogenous		Exogenous		
Size	No. of transcription factors	142		70	74	71	72	63
	No. of target genes	3,420		280	257	748	678	362
	No. of regulatory interactions	7,074		550	481	1,217	1,082	566
Topological measures	In-degree ($<k_{in}>$)	2.1		2.0	1.9	1.6	1.6	1.6
	Out-degree ($<k_{out}>$)	49.8		7.9	6.5	17.1	15.0	9.0
	Path length (<l>)	4.7		4.5	3.4	2.1	2.0	2.2
	Clustering coefficient (<c>)	0.11		0.15	0.14	0.09	0.09	0.08
Motifs (%)	Single input (SIM)	1,748	(37.6%)	130 (32.0%)	117 (38.9%)	438 (57.4%)	462 (55.7%)	228 (59.1%)
	Multiple input (MIM)	325	(7.0%)	96 (23.7%)	50 (16.6%)	180 (23.6%)	226 (27.3%)	78 (20.2%)
	Feed-forward loop (FFL)	2,581	(55.5%)	180 (44.3%)	134 (44.5%)	145 (19.0%)	141 (17.0%)	80 (20.7%)
	Total	4,654		406	301	763	829	386

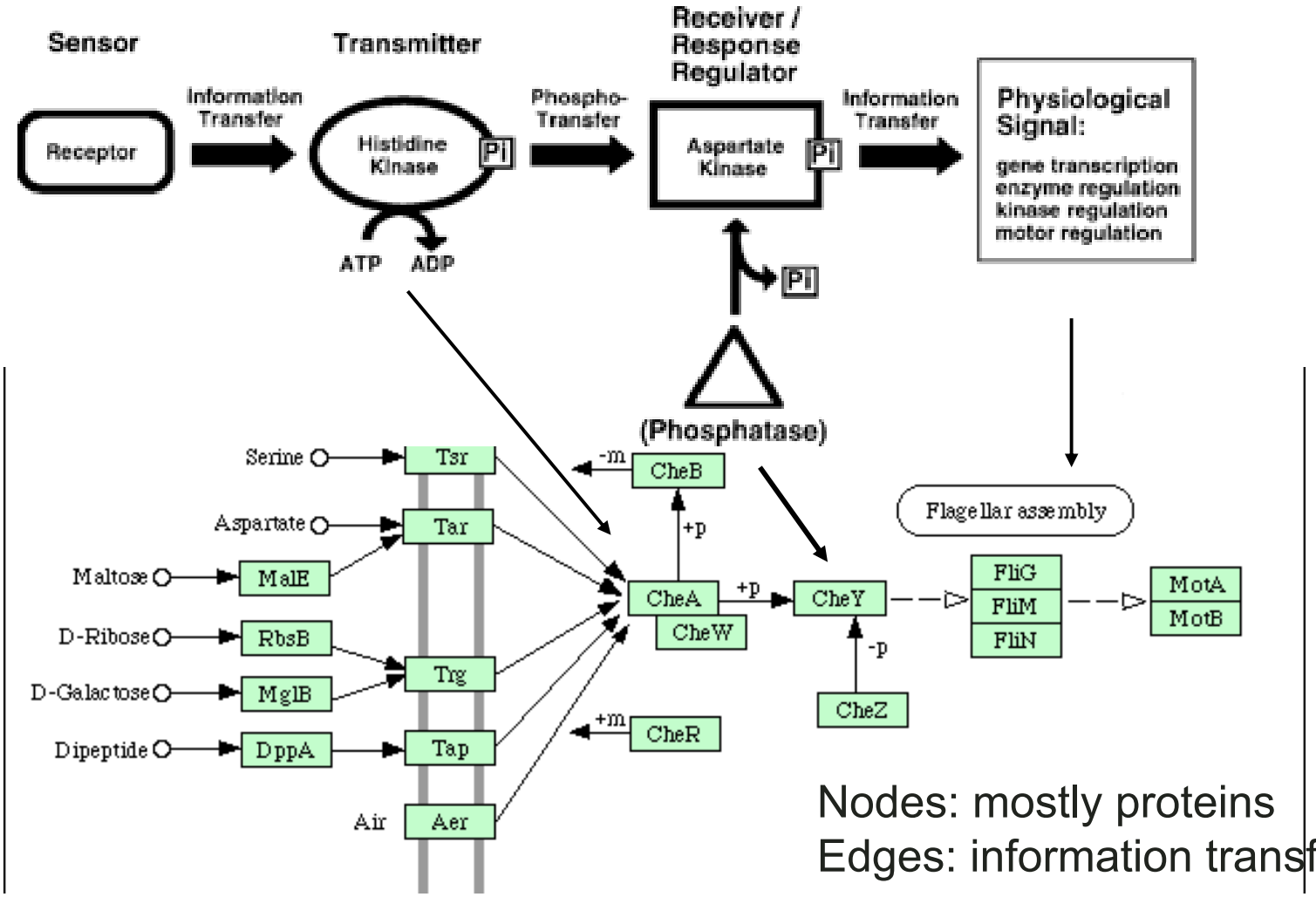
Luscombe et al, Nature 431, 308 (2004)

Representation of chemical reactions+ regulation



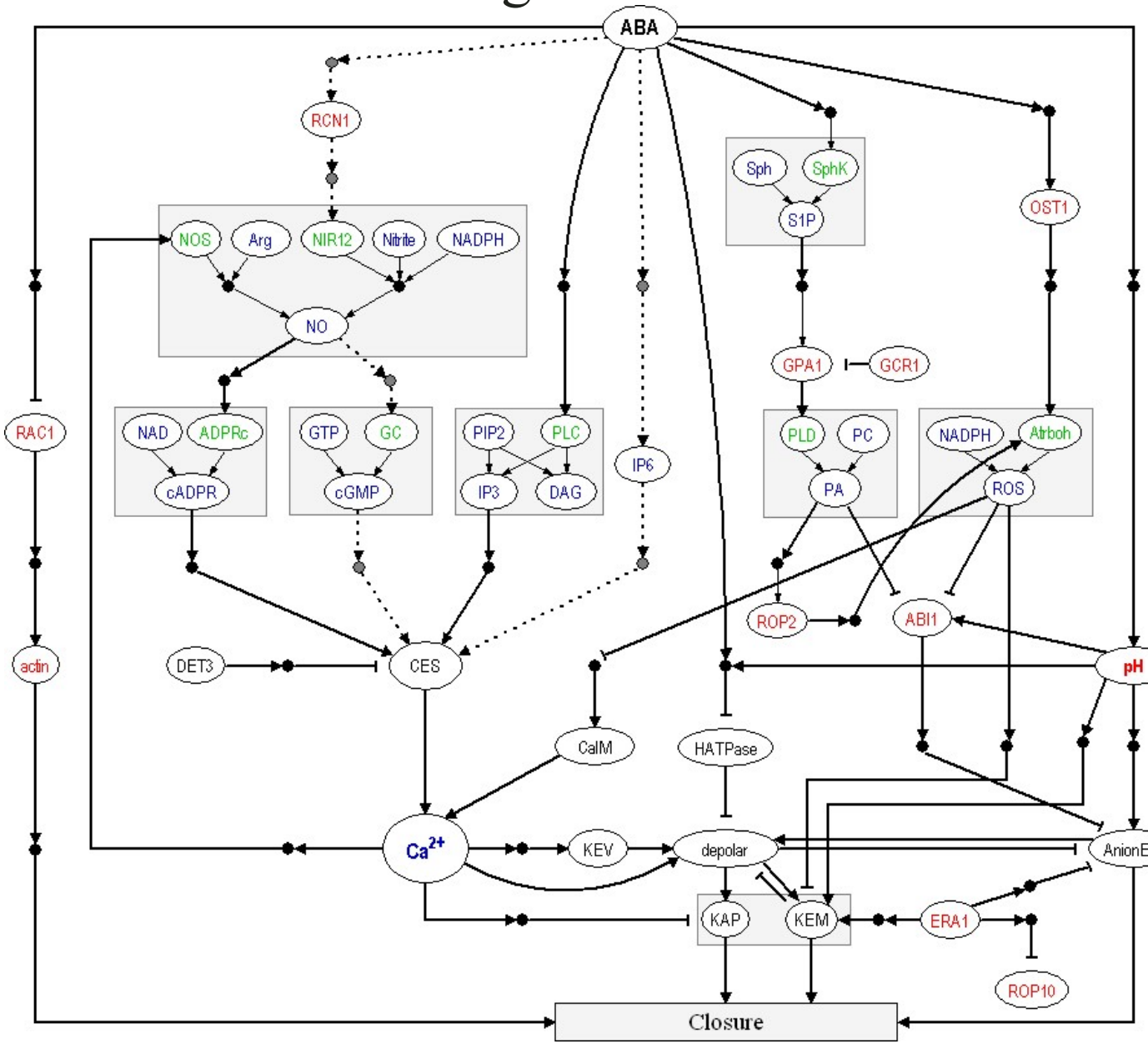
E. O. Voit, Computational Analysis of Biochemical Systems

4. Signal transduction pathways



Nodes: mostly proteins
Edges: information transfer

ABA signal transduction network

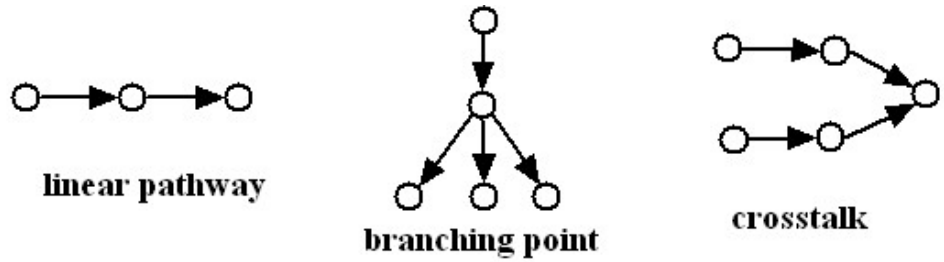


Green: enzymes
Black: transport
Blue: small molecules
Red: sign. transd. proteins
Black points: unknown intermediary nodes

Song Li, PSU

Network analysis needs to be complemented by dynamics

- Topology intertwined with function and dynamics
- Not all interactions are realized (active) at the same time!
- Topological analysis needs to be focused towards answering function – oriented questions
- Dynamical modeling is necessary to investigate the timecourse of the processes represented by networks



synergy

filter

homeostasis

amplification