Understanding Protein Function on a Genome-scale through the Analysis of Molecular Networks

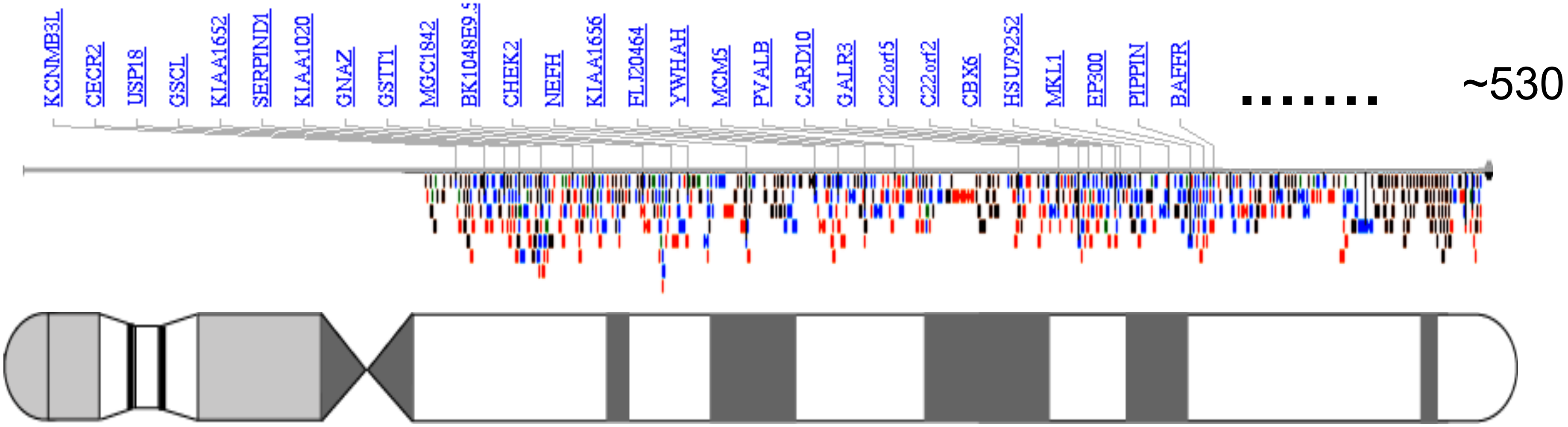
Mark B Gerstein
Yale

Slides at

Lectures.GersteinLab.org

(See Last Slide for References & More Info.)

The problem: Grappling with Function on a Genome Scale?



- 250 of ~530
  originally characterized on chr. 22
  [Dunham et al. Nature (1999)]

- >25K Proteins in Entire Human Genome
  (with alt. splicing)

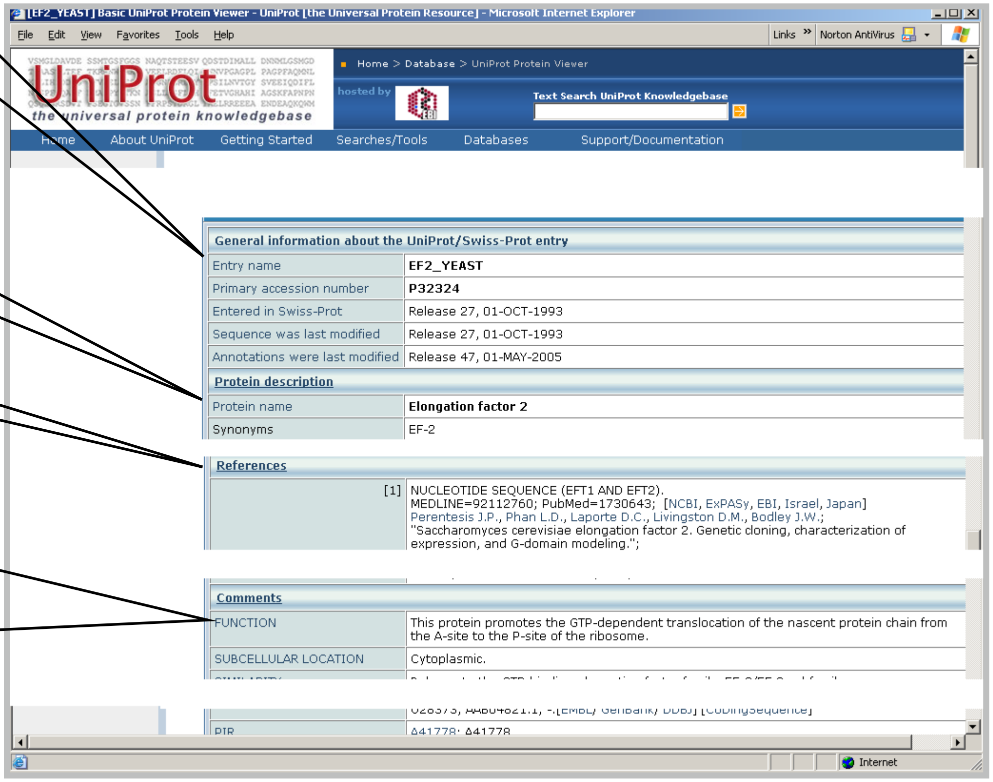
Traditional single molecule way to integrate evidence & describe function

EF2_YEAST

Descriptive Name:
Elongation Factor 2

Lots of references
to papers

Summary sentence describing function:
This protein promotes the GTP-dependent translocation of the nascent protein chain from the A-site to the P-site of the ribosome.


[EF2_YEAST] Basic UniProt Protein Viewer - UniProt [the Universal Protein Resource] - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Links » Norton AntiVirus
UniProt
the universal protein knowledgebase
Home > Database > UniProt Protein Viewer
hosted by
Text Search UniProt Knowledgebase
Home About UniProt Getting Started Searches/Tools Databases Support/Documentation
Internet

General information about the UniProt/Swiss-Prot entry

Entry name	**EF2_YEAST**
Primary accession number	**P32324**
Entered in Swiss-Prot	Release 27, 01-OCT-1993
Sequence was last modified	Release 27, 01-OCT-1993
Annotations were last modified	Release 47, 01-MAY-2005

Protein description

Protein name	**Elongation factor 2**
Synonyms	EF-2

References

[1] NUCLEOTIDE SEQUENCE (EFT1 AND EFT2).
MEDLINE=92112760; PubMed=1730643; [NCBI, ExPASy, EBI, Israel, Japan]
Perentesis J.P., Phan L.D., Laporte D.C., Livingston D.M., Bodley J.W.;
"Saccharomyces cerevisiae elongation factor 2. Genetic cloning, characterization of expression, and G-domain modeling.";

Comments

FUNCTION	This protein promotes the GTP-dependent translocation of the nascent protein chain from the A-site to the P-site of the ribosome.
SUBCELLULAR LOCATION	Cytoplasmic.

U28373; AAB04821.1; -.[EMBL/ GenBank/ DDBJ] [CodingSequence]

PIR	A41778; A41778

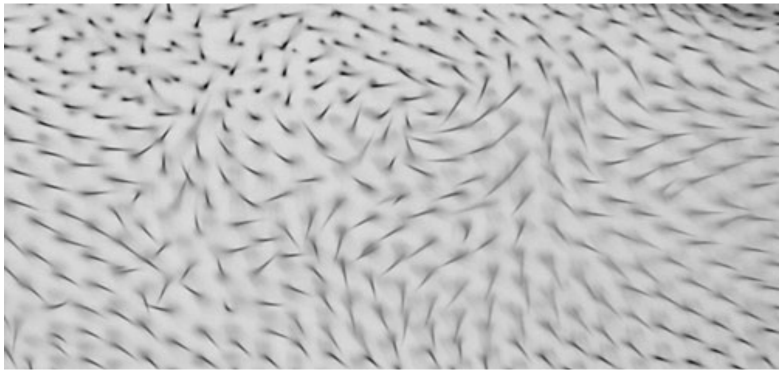
Some obvious issues in scaling single molecule definition to a genomic scale

- Fundamental complexities
  - ◊ Often >2 proteins/function
  - ◊ Multi-functionality:
    2 functions/protein
  - ◊ Role Conflation:
    molecular, cellular, phenotypic

Some obvious issues in scaling single molecule definition to a genomic scale

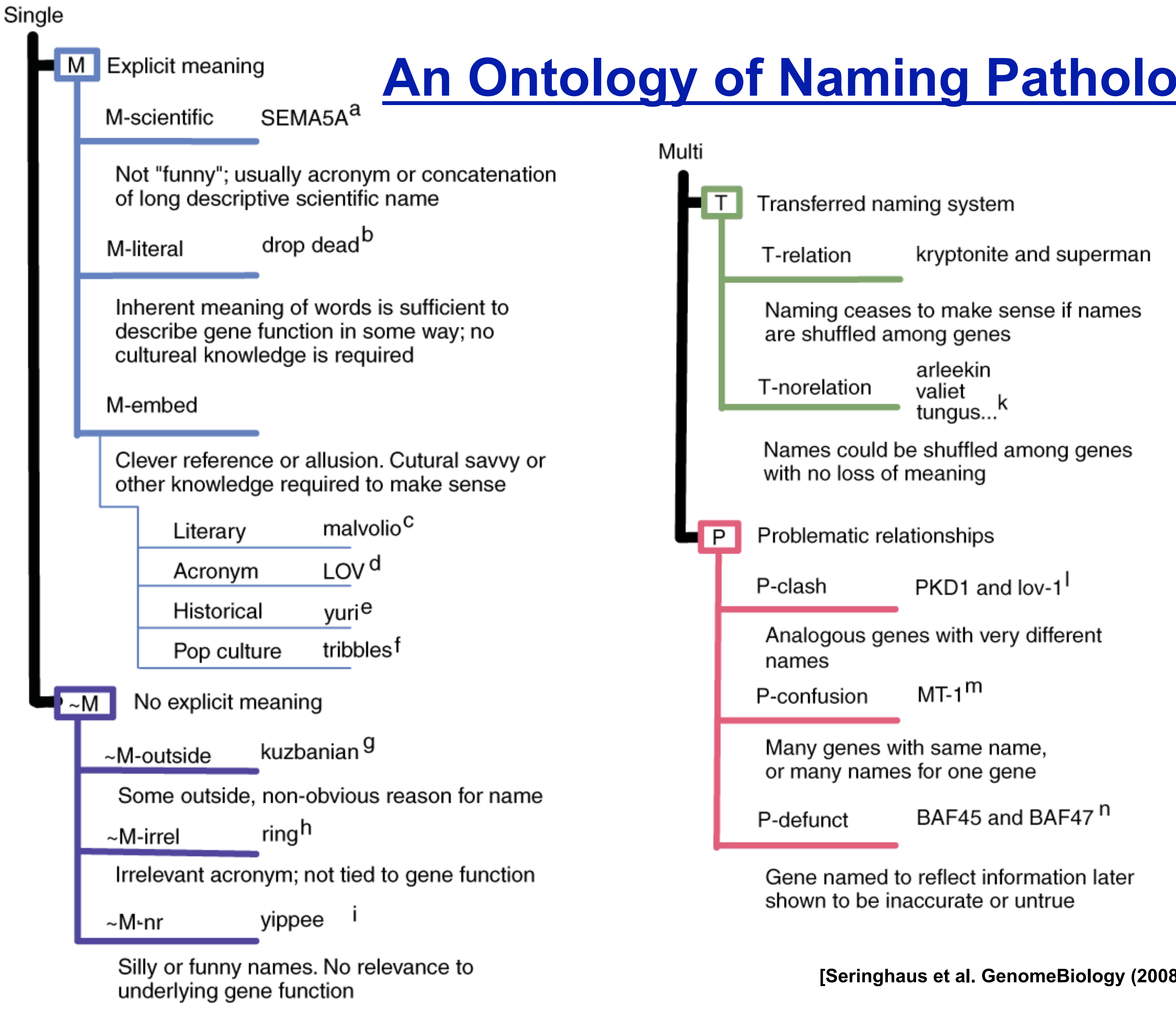
- Fundamental complexities
  - ◊ Often >2 proteins/function
  - ◊ Multi-functionality:
2 functions/protein
  - ◊ Role Conflation:
molecular, cellular, phenotypic

- Fun terms… but do they scale?....
  - ◊ **Starry night** (P Adler, ’94)



[Seringhaus et al. GenomeBiology (2008)]

An Ontology of Naming Pathologies



Single

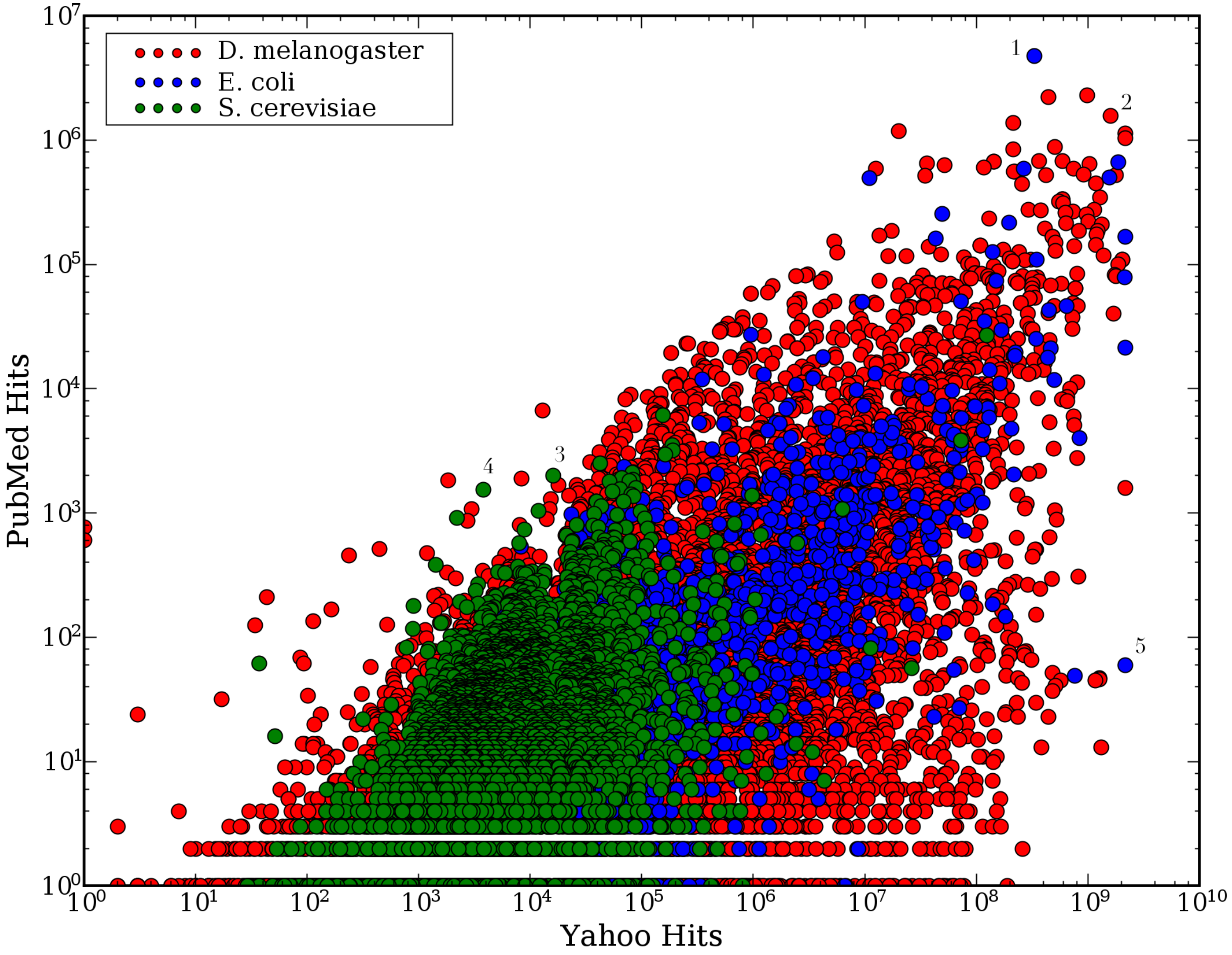
- M Explicit meaning
  - M-scientific SEMA5A[a]
    - Not "funny"; usually acronym or concatenation of long descriptive scientific name
  - M-literal drop dead[b]
    - Inherent meaning of words is sufficient to describe gene function in some way; no cultureal knowledge is required
  - M-embed
    - Clever reference or allusion. Cutural savvy or other knowledge required to make sense
      - Literary malvolio[c]
      - Acronym LOV[d]
      - Historical yuri[e]
      - Pop culture tribbles[f]
- ~M No explicit meaning
  - ~M-outside kuzbanian[g]
    - Some outside, non-obvious reason for name
  - ~M-irrel ring[h]
    - Irrelevant acronym; not tied to gene function
  - ~M-nr yippee [i]
    - Silly or funny names. No relevance to underlying gene function

Multi

- T Transferred naming system
  - T-relation kryptonite and superman
    - Naming ceases to make sense if names are shuffled among genes
  - T-norelation arleekin
valiet
tungus...[k]
    - Names could be shuffled among genes with no loss of meaning
- P Problematic relationships
  - P-clash PKD1 and lov-1[l]
    - Analogous genes with very different names
  - P-confusion MT-1[m]
    - Many genes with same name, or many names for one gene
  - P-defunct BAF45 and BAF47 [n]
    - Gene named to reflect information later shown to be inaccurate or untrue

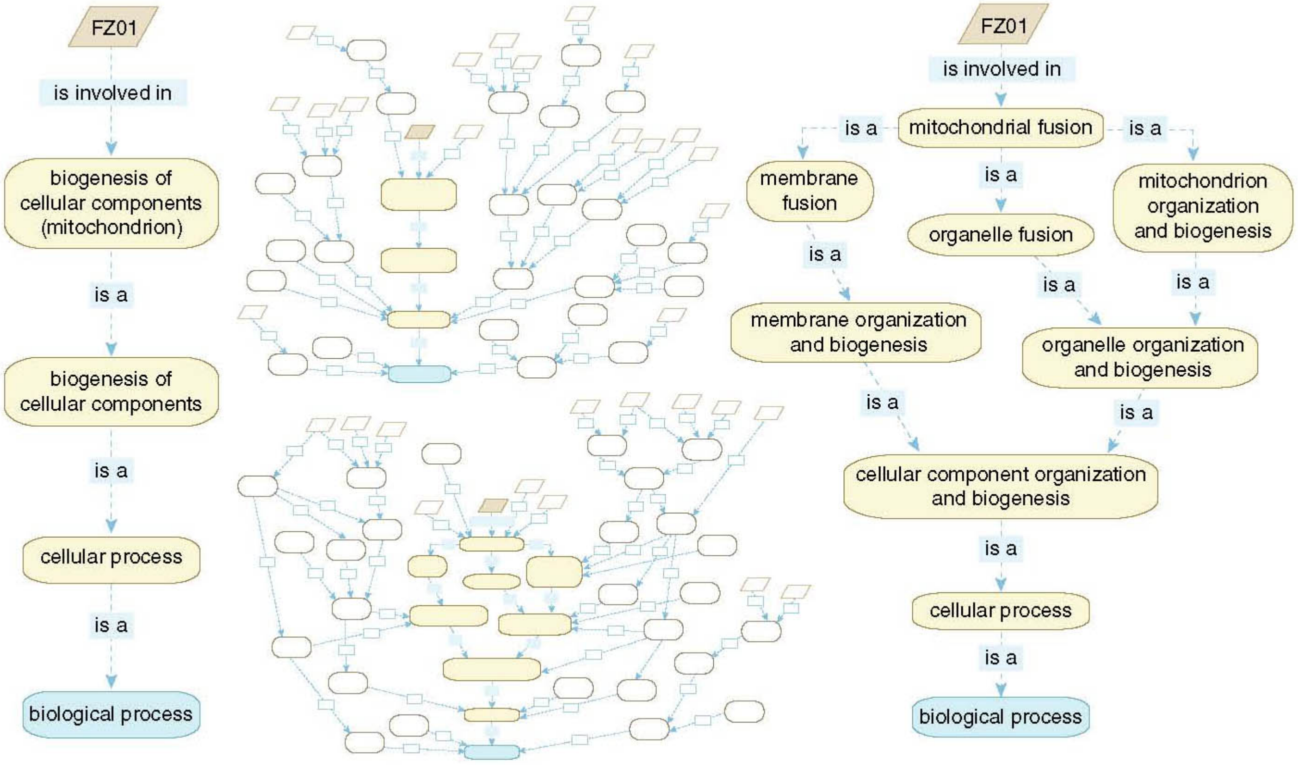
[Seringhaus et al. GenomeBiology (2008)]

Gene Name Skew



[Seringhaus et al. GenomeBiology (2008)]

Hierarchies & DAGs of controlled-vocab terms but still have issues...

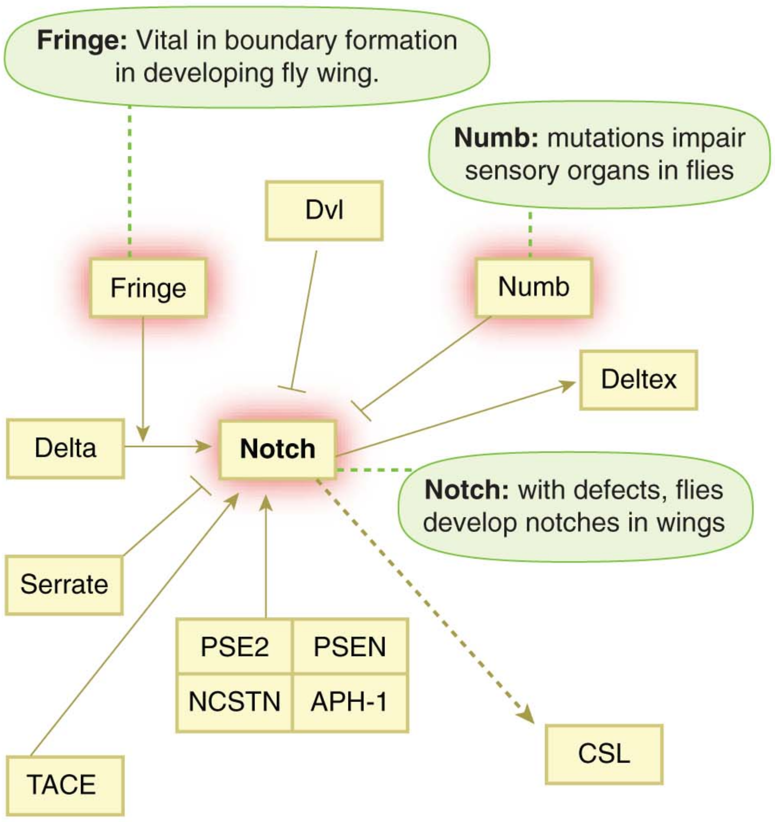

FZ01
is involved in
biogenesis of cellular components (mitochondrion)
is a
biogenesis of cellular components
is a
cellular process
is a
biological process
FZ01
is involved in
is a
mitochondrial fusion
is a
membrane fusion
is a
mitochondrion organization and biogenesis
organelle fusion
is a
membrane organization and biogenesis
is a
is a
organelle organization and biogenesis
is a
is a
cellular component organization and biogenesis
is a
cellular process
is a
biological process

MIPS (Mewes et al.)

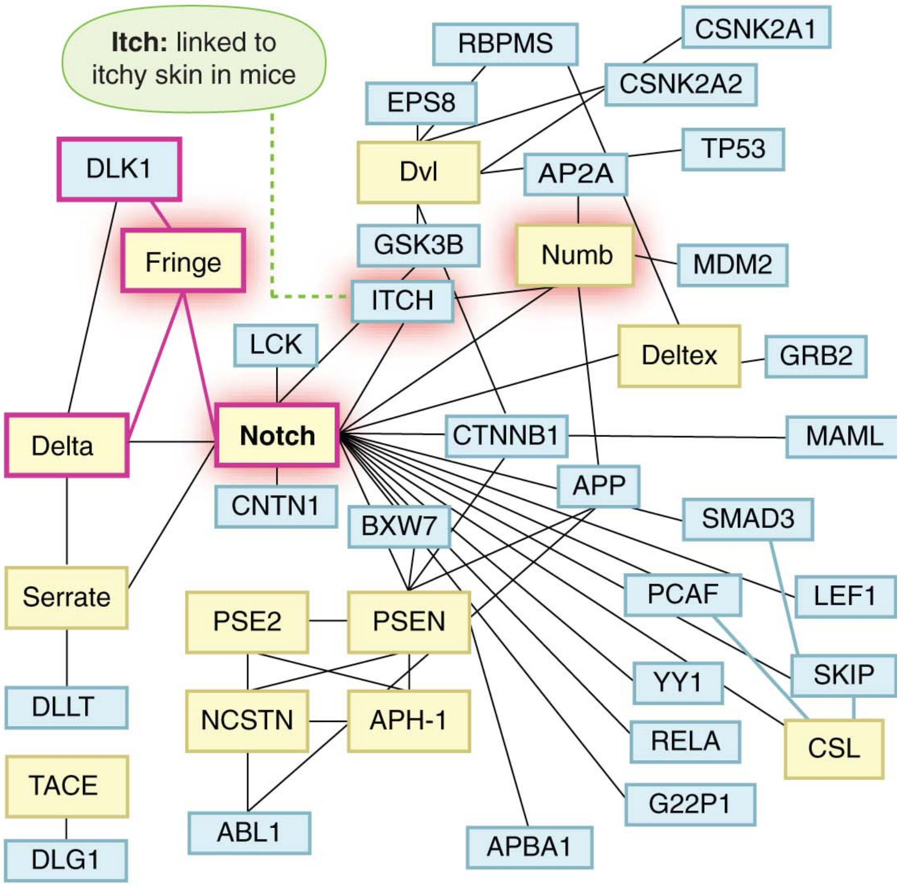
GO (Ashburner et al.)

[Seringhaus & Gerstein, Am. Sci. '08]

Networks (Old & New)


Fringe: Vital in boundary formation in developing fly wing.
Numb: mutations impair sensory organs in flies
Dvl
Fringe
Numb
Deltex
Delta
Notch
Notch: with defects, flies develop notches in wings
Serrate
PSE2
PSEN
NCSTN
APH-1
TACE
CSL

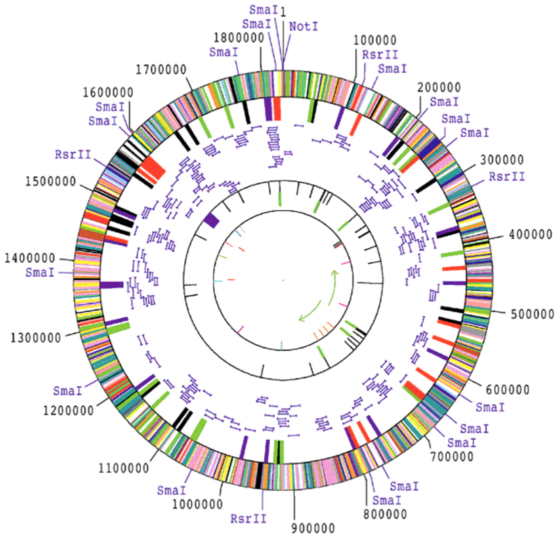
Classical KEGG pathway

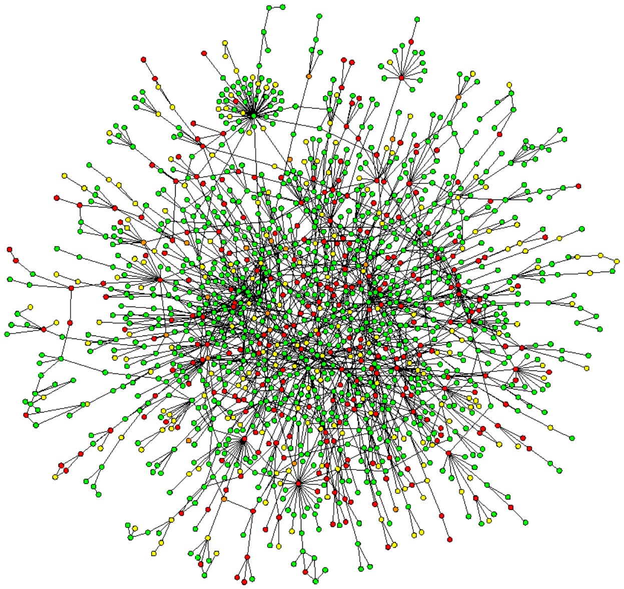


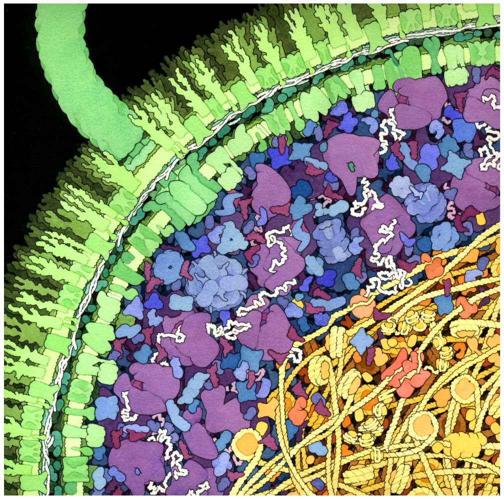
Same Genes in High-throughput Network

[Seringhaus & Gerstein, Am. Sci. '08]

Networks occupy a midway point in terms of level of understanding







1D: Complete Genetic Partslist

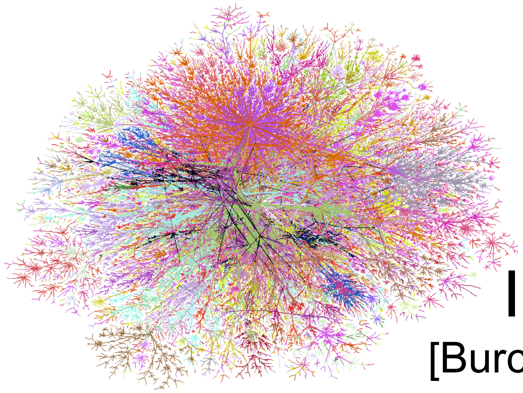
~2D: Bio-molecular Network Wiring Diagram

3D: Detailed structural understanding of cellular machinery

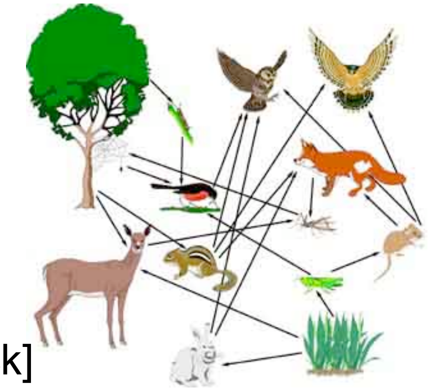
[Fleischmann et al., Science, 269 :496]

[Jeong et al. Nature, 41:411]

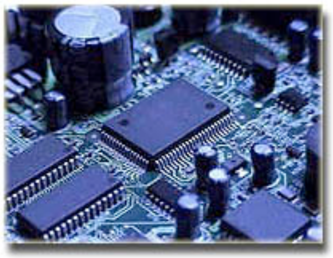
Networks as a universal language



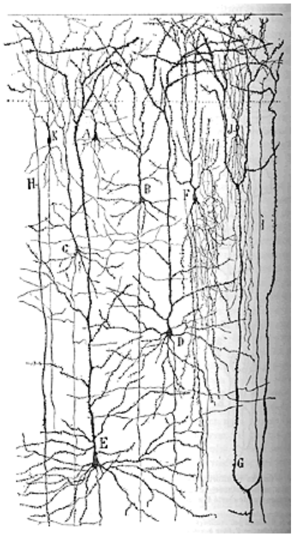
Internet
[Burch & Cheswick]



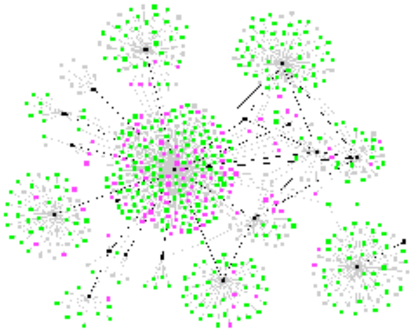
Food Web



Electronic Circuit

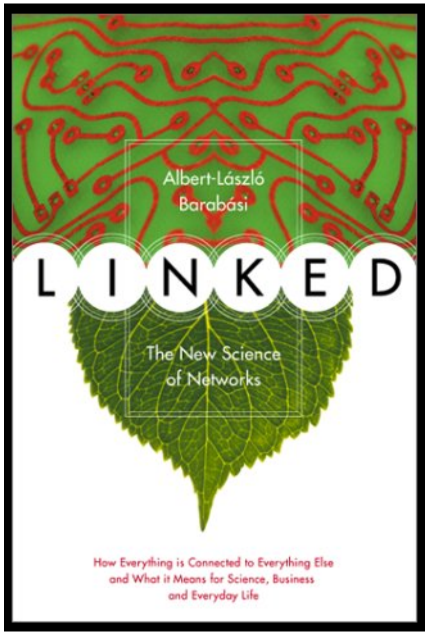


Neural Network
[Cajal]



Disease Spread
[Krebs]

Protein Interactions
[Barabasi]

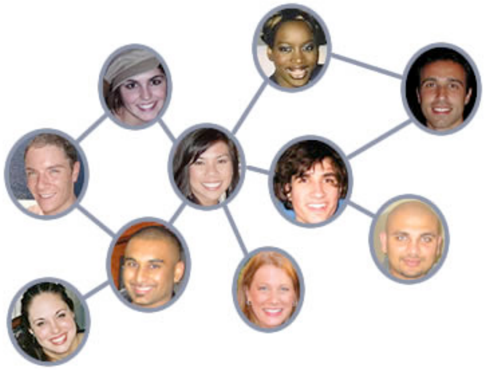


Albert-László Barabási

LINKED

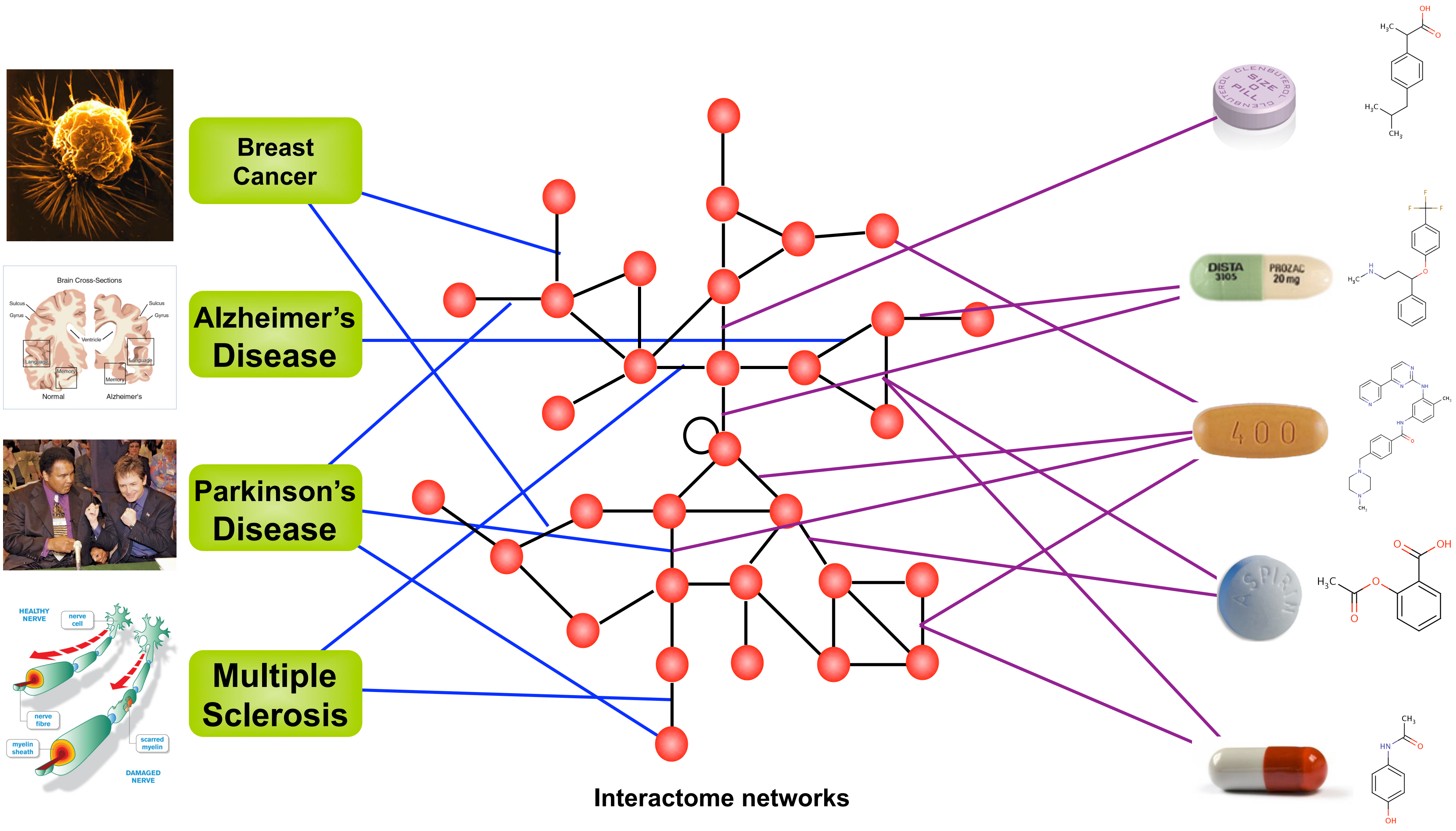
The New Science of Networks

How Everything is Connected to Everything Else and What it Means for Science, Business and Everyday Life



Social Network

Network pathology & pharmacology



Breast Cancer

Alzheimer's Disease

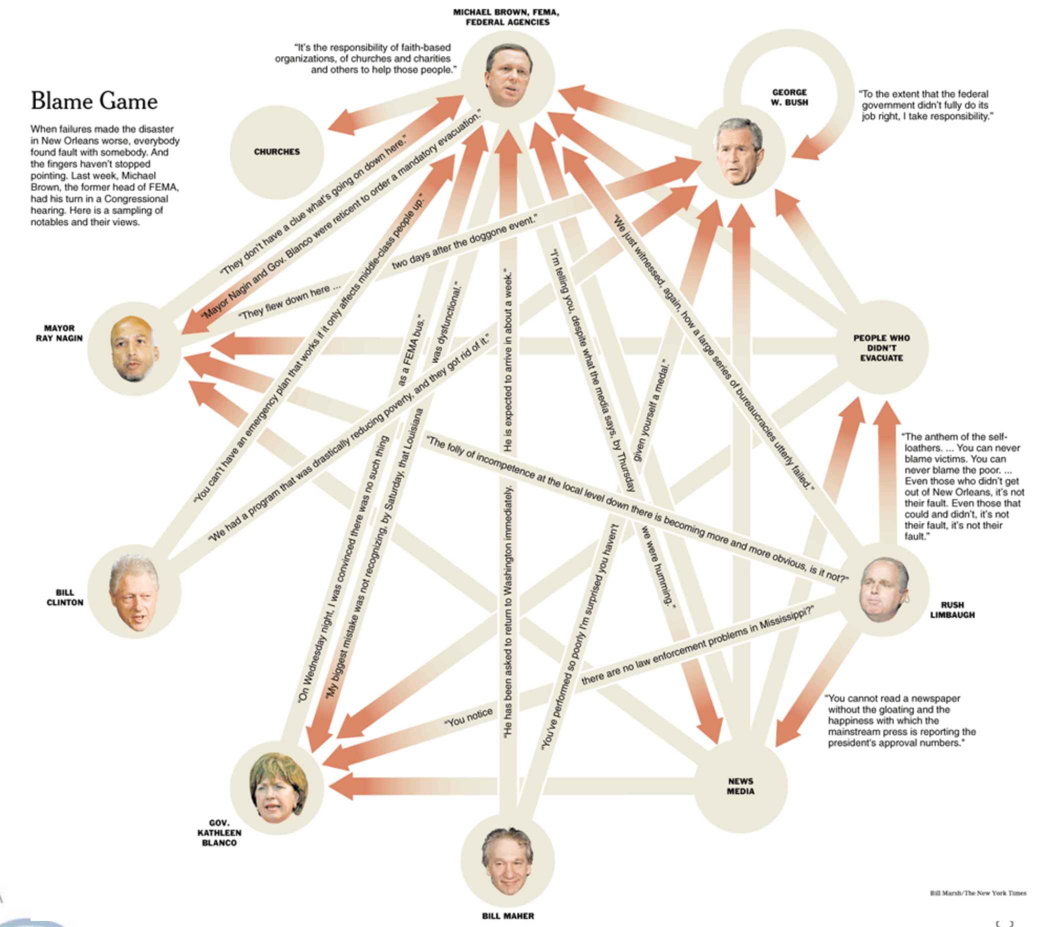
Parkinson's Disease

Multiple Sclerosis

Interactome networks

[Adapted from H Yu]

Using the position in networks to describe function



Blame Game

When failures made the disaster in New Orleans worse, everybody found fault with somebody. And the fingers haven't stopped pointing. Last week, Michael Brown, the former head of FEMA, had his turn in a Congressional hearing. Here is a sampling of notables and their views.

"It's the responsibility of faith-based organizations, of churches and charities and others to help those people."

"To the extent that the federal government didn't fully do its job right, I take responsibility."

"The anthem of the self-loathers. ... You can never blame victims. You can never blame the poor. ... Even those who didn't get out of New Orleans, it's not their fault. Even those that could and didn't, it's not their fault, it's not their fault."

"You cannot read a newspaper without the gloating and the happiness with which the mainstream press is reporting the president's approval numbers."

"They don't have a clue what's going on down here."

"Mayor Nagin and Gov. Blanco were reticent to order a mandatory evacuation."

"They flew down here ... two days after the doggone event."

"You can't have an emergency plan that works if it only affects middle-class people up."

"We had a program that was drastically reducing poverty, and they got rid of it."

"On Wednesday night, I was convinced there was no such thing as a FEMA bus."

"My biggest mistake was not recognizing, by Saturday, that Louisiana was dysfunctional."

"He has been asked to return to Washington immediately. He is expected to arrive in about a week."

"I'm telling you, despite what the media says, by Thursday we were humming."

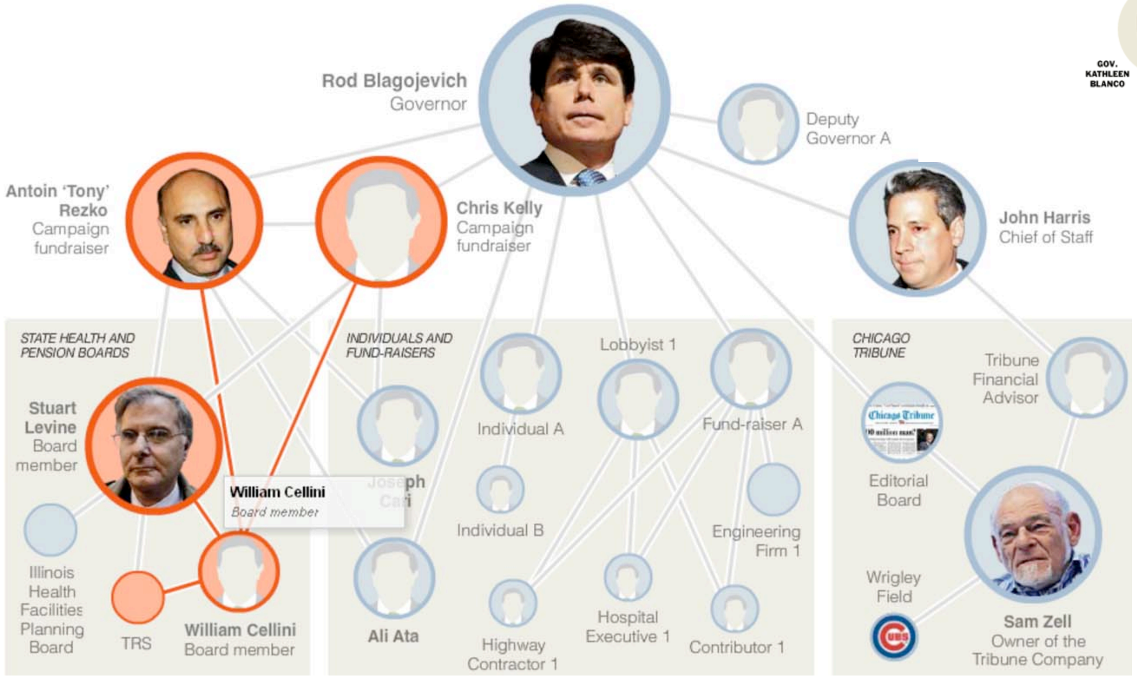
"You've performed so poorly I'm surprised you haven't given yourself a medal."

"We just witnessed, again, how a large series of bureaucracies utterly failed."

"The folly of incompetence at the local level down there is becoming more and more obvious, is it not?"

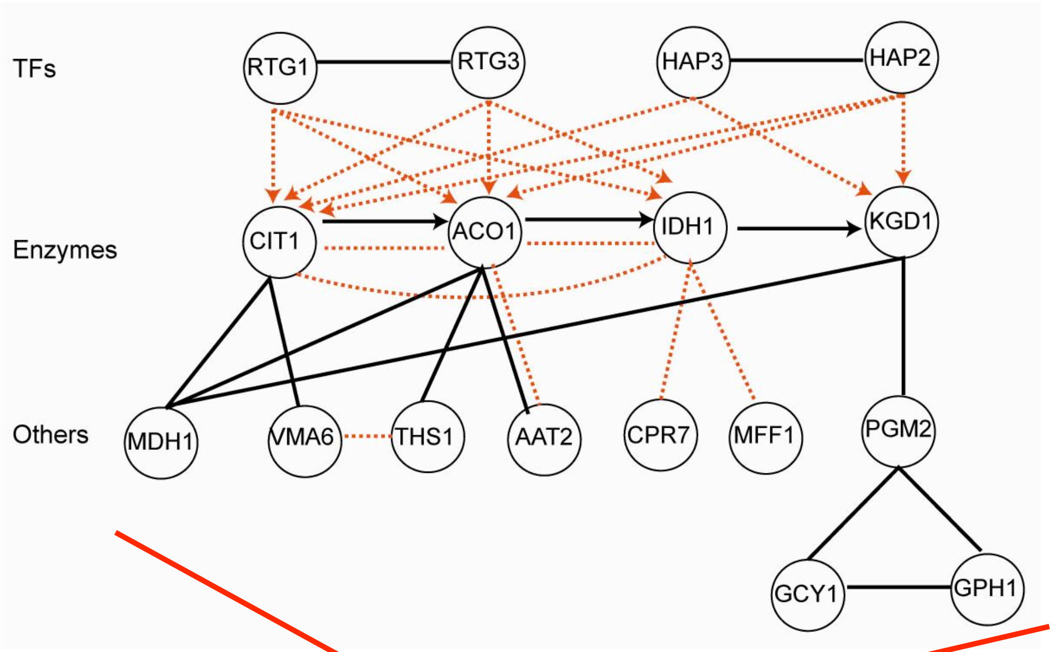
"You notice there are no law enforcement problems in Mississippi?"

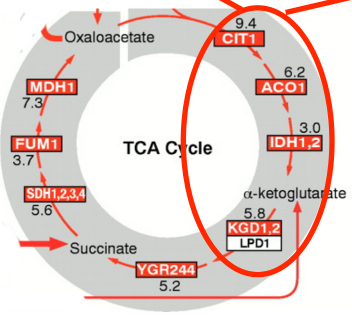
Bill Marsh/The New York Times



[NY Times, 2-Oct-05, 9-Dec-08]

Combining networks forms an ideal way of integrating diverse information





Part of the TCA cycle

→ **Metabolic pathway**

⋯→ **Transcriptional regulatory network**

— Physical protein-protein Interaction

⋯ Co-expression Relationship

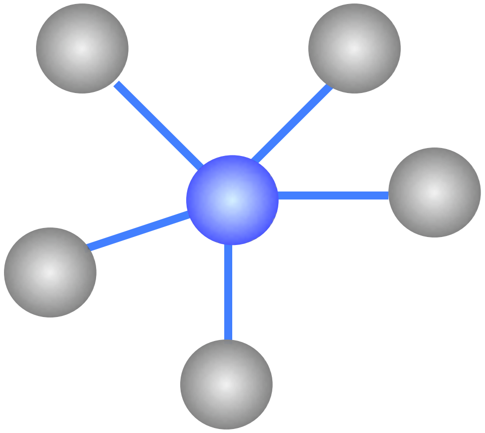
Genetic interaction (synthetic lethal)
Signaling pathways

Outline: Molecular Networks

- Why Networks?
- Network Structure: Key Positions
  - ◊ Hubs & Bottlenecks
  - ◊ Tops of a Hierachy
- Networks, Variation & the Environment
  - ◊ Which pathways change most with the environment

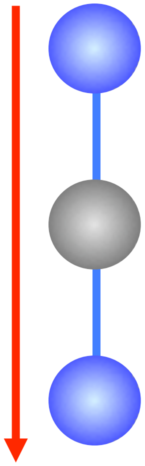
Global topological measures

Indicate the gross topological structure of the network



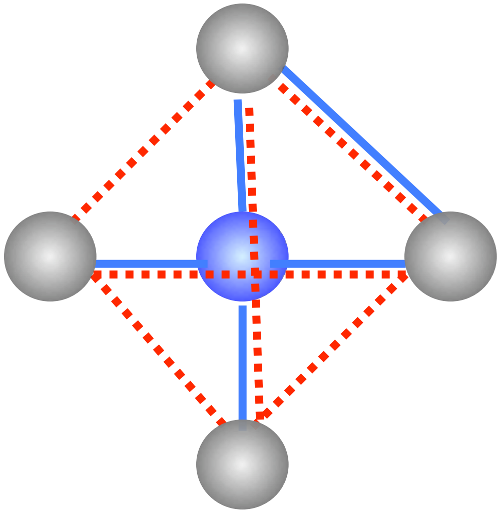
Degree (*K*)

5



Path length (*L*)

2



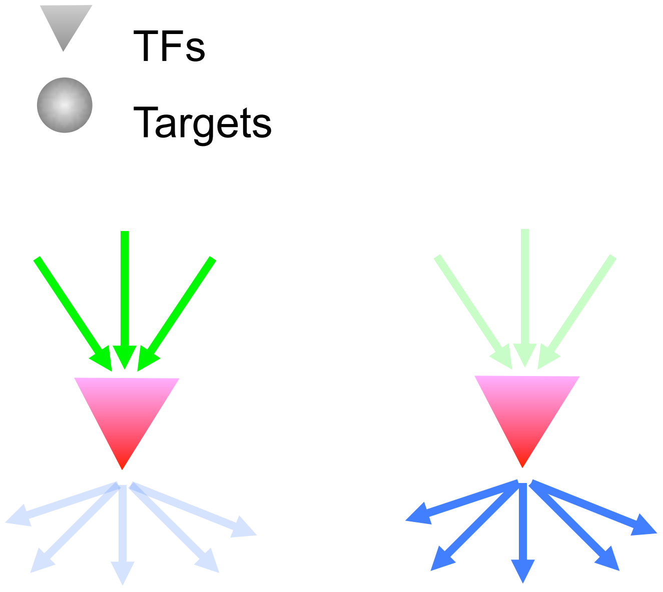
Clustering coefficient (*C*)

1/6

Interaction and expression networks are ***undirected***

[Barabasi]

Global topological measures for directed networks



▼ TFs

● Targets

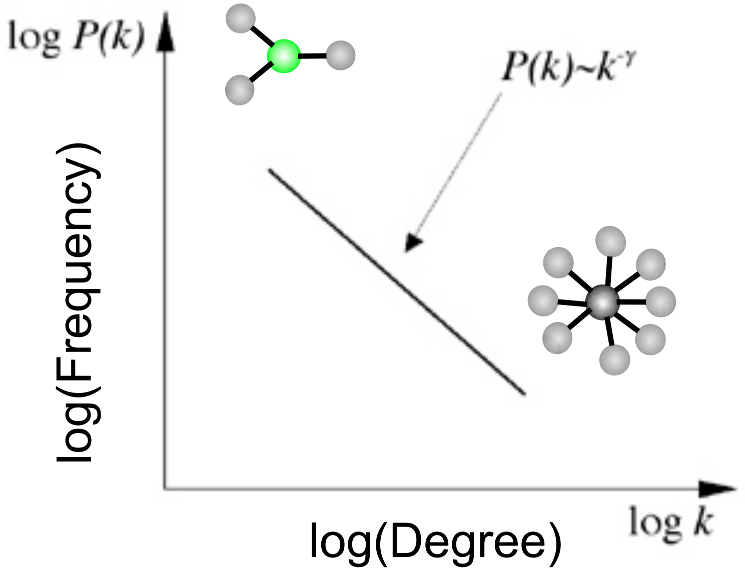
In-degree
3

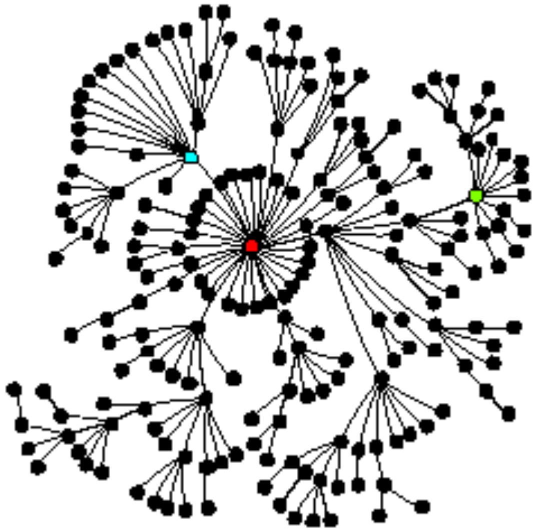
Out-degree
5

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

Scale-free networks

Power-law distribution





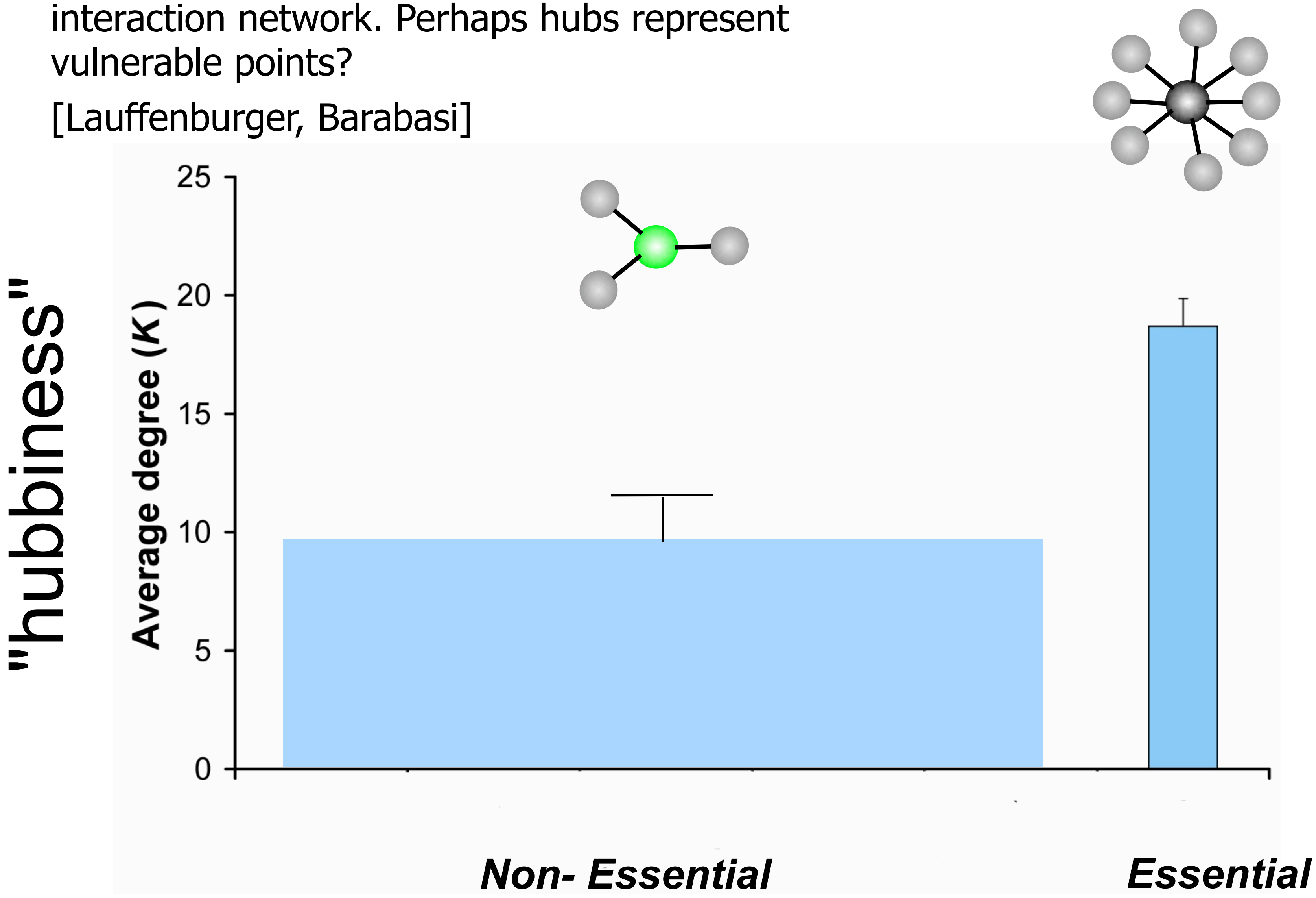
Hubs dictate the structure of the network

[Barabasi]

Hubs tend to be Essential

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

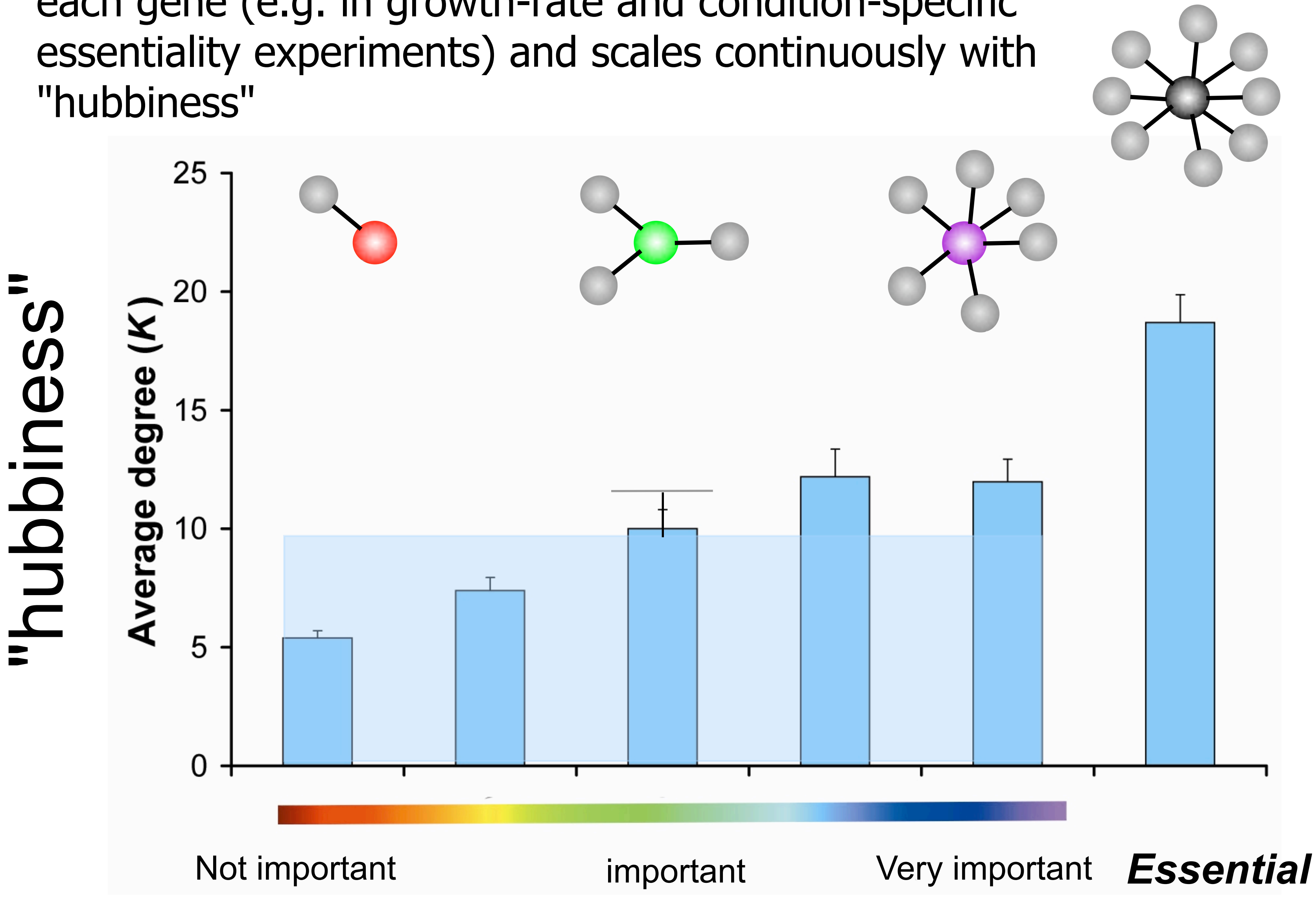
[Lauffenburger, Barabasi]



[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"



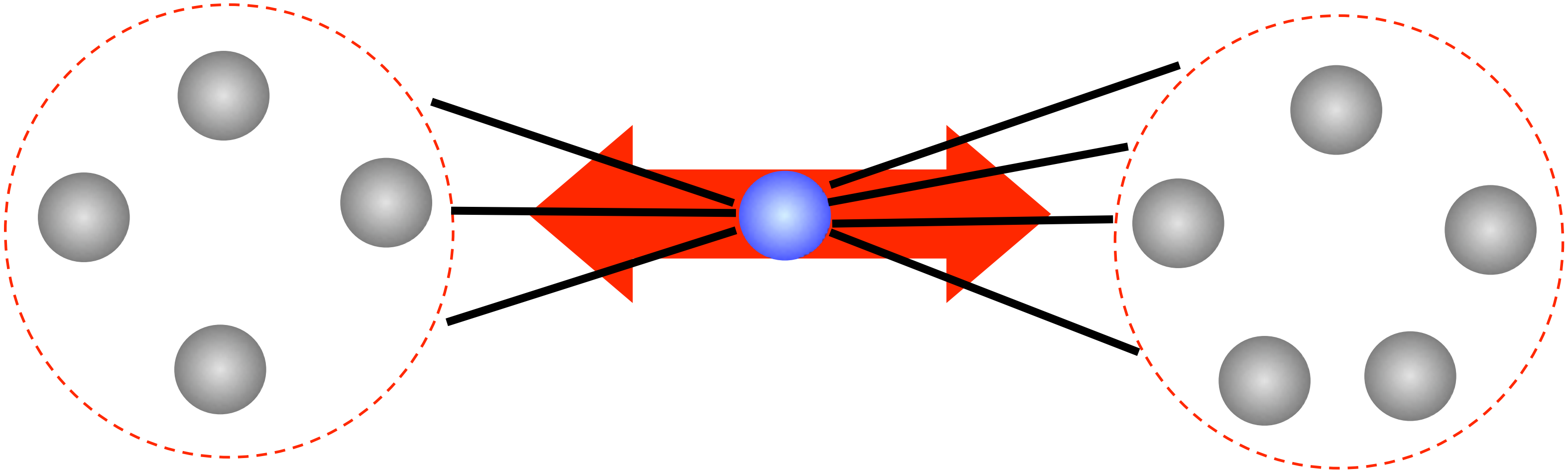
[Yu et al., 2003, TIG]

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.

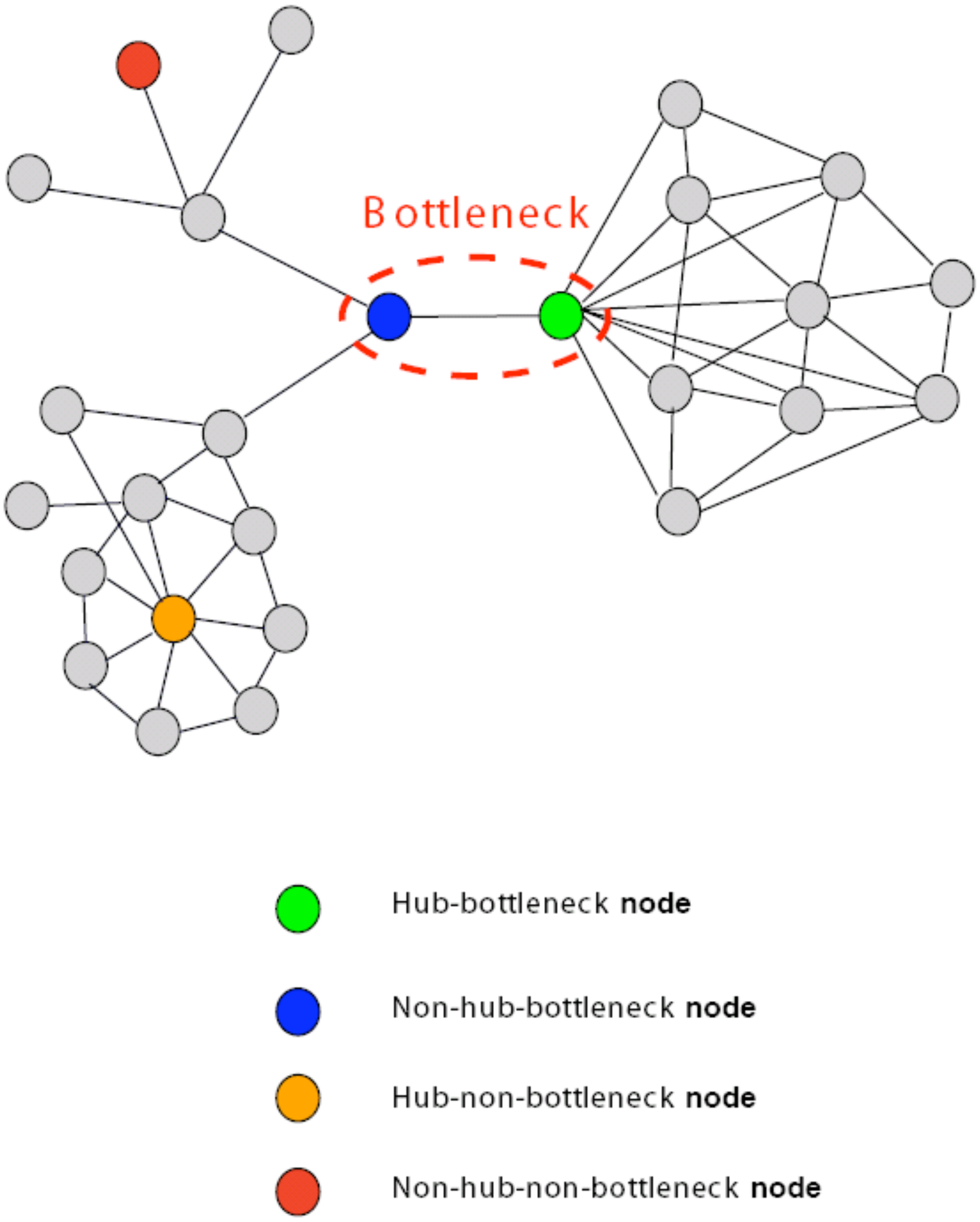
Girvan & Newman (2002) PNAS 99: 7821.



Betweenness centrality -- Bottlenecks

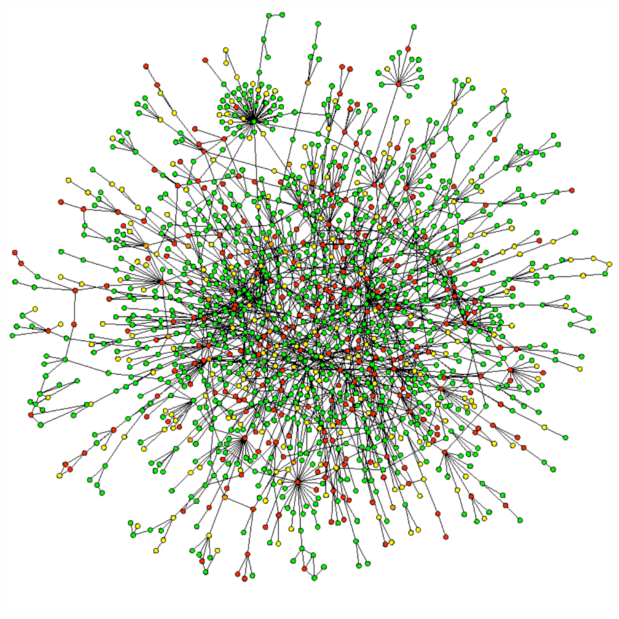
Proteins with high betweenness are defined as *Bottlenecks* (top 20%), in analogy to the traffic system

Bottlenecks & Hubs

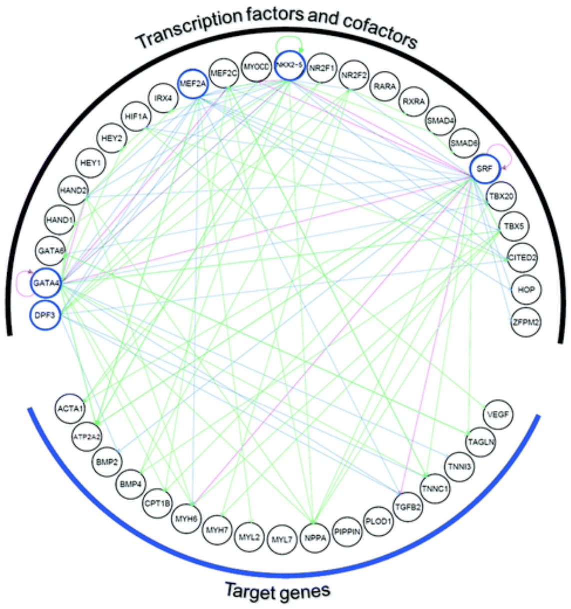


[Yu et al., PLOS CB (2007)]

Different Interactome networks

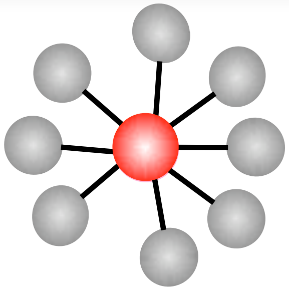


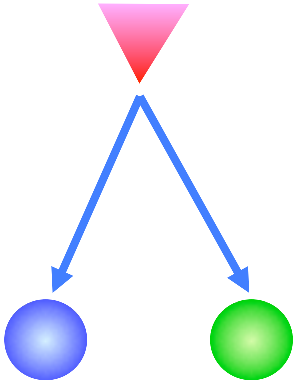
Interaction networks



Regulatory networks

Undirected



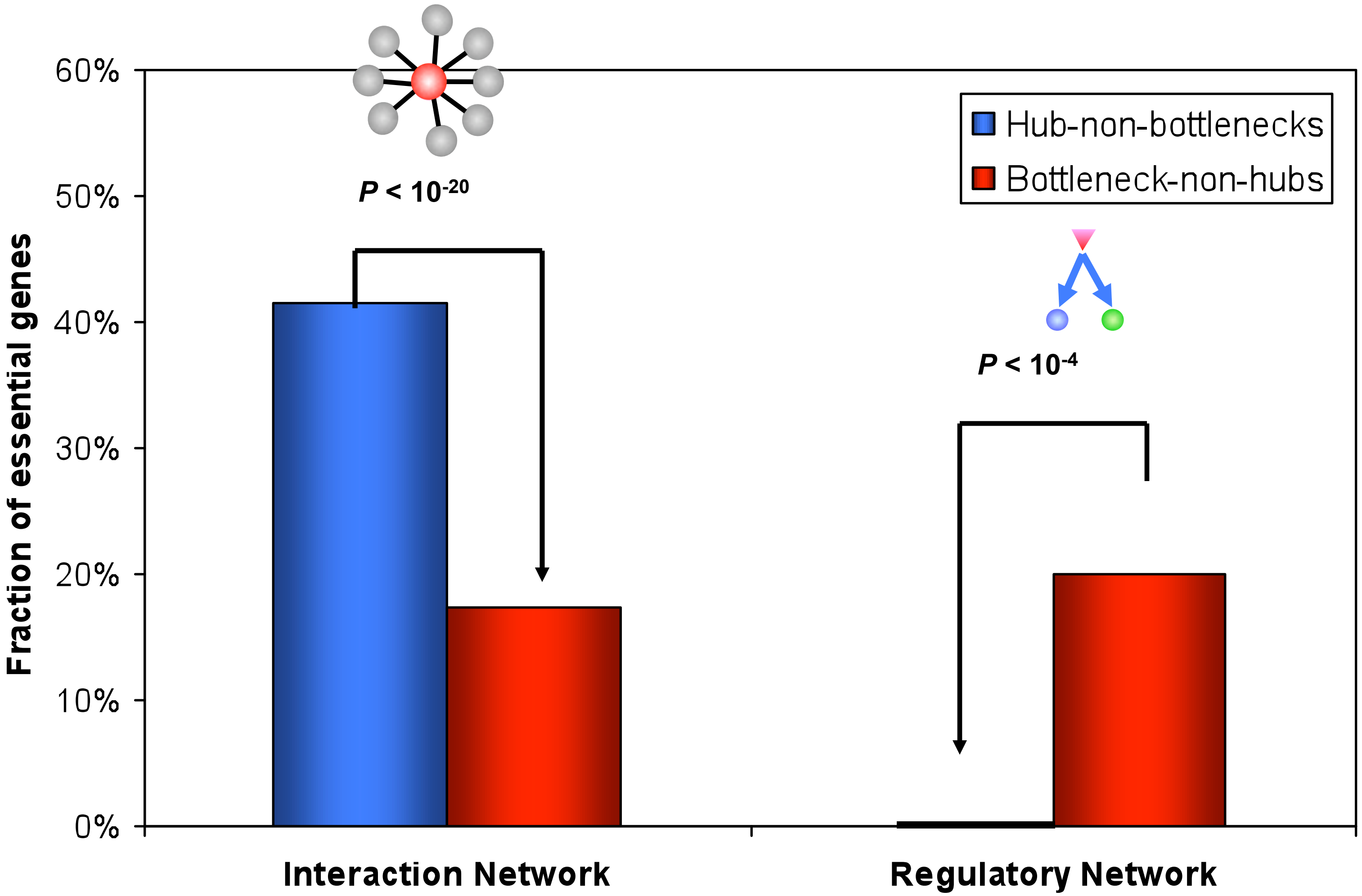


Directed

[Toenjes, *et al*, *Mol. BioSyst.* (2008)]

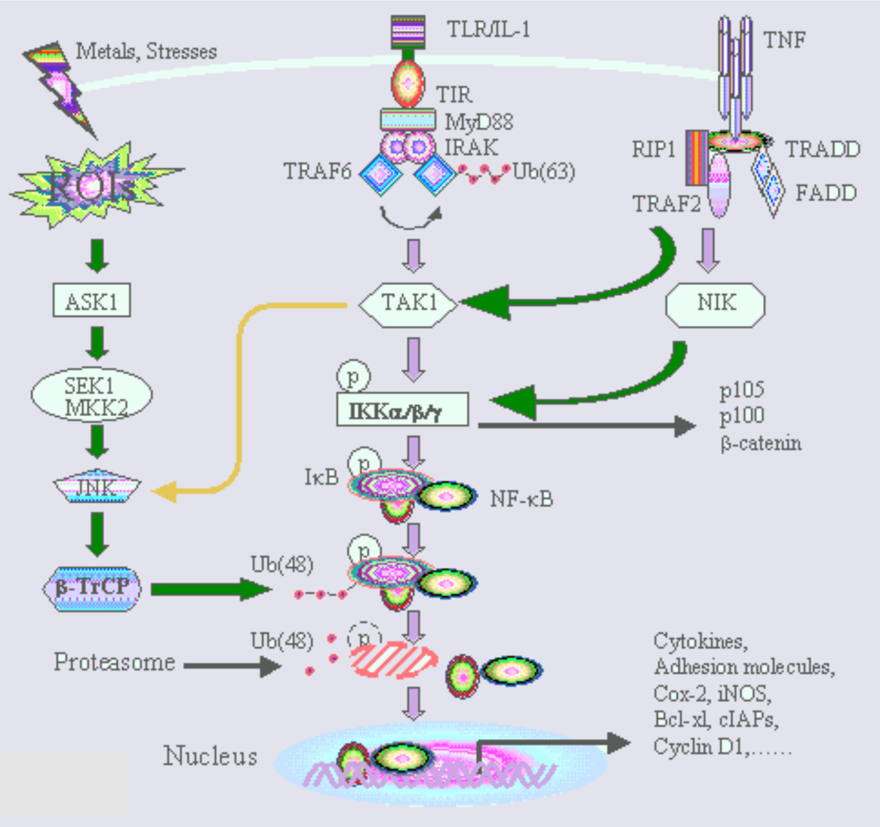
[Jeong *et al*, *Nature* (2001)]

Bottlenecks are what matters in regulatory networks



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

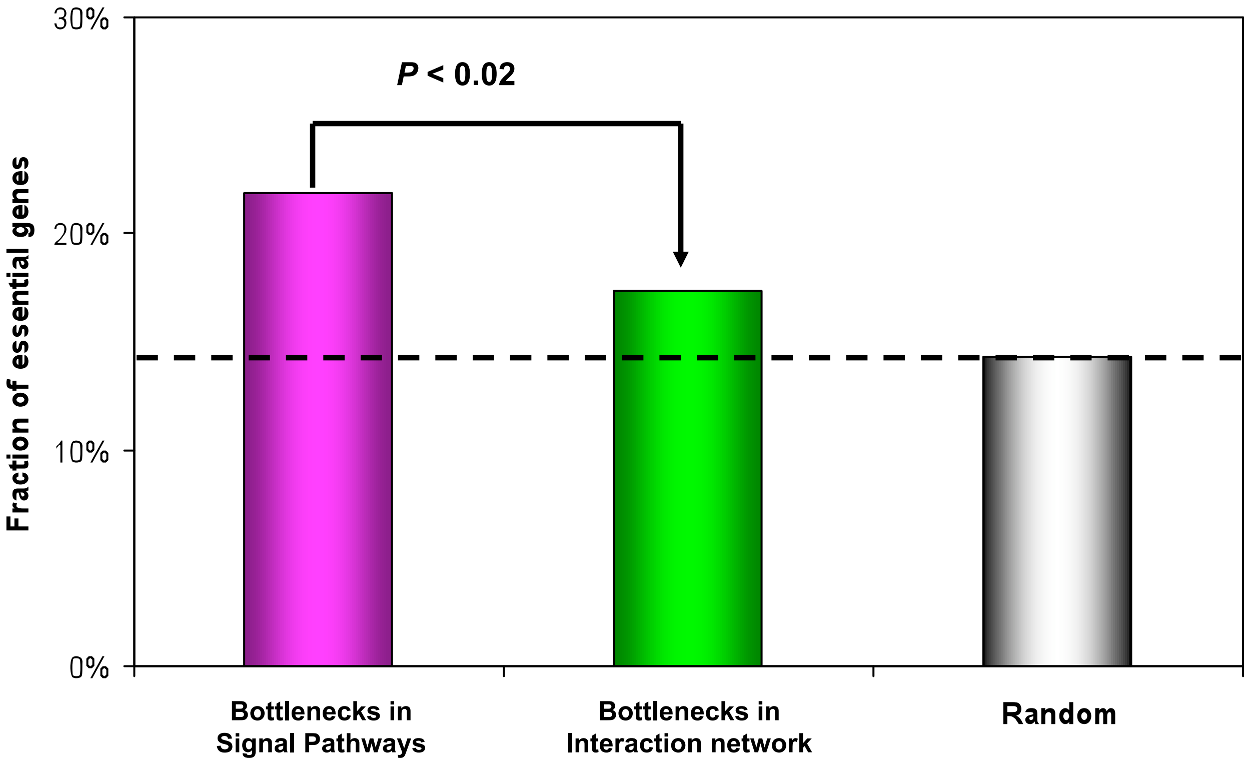
Signaling transduction pathways are directed

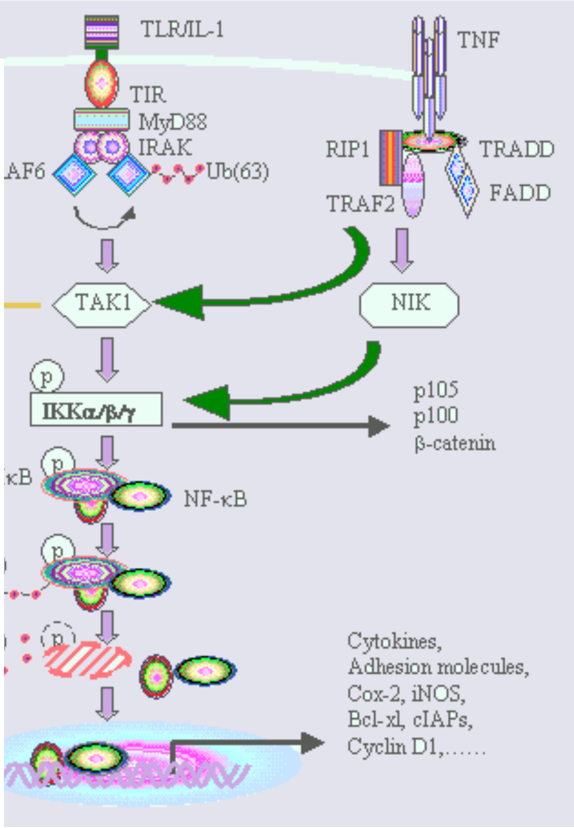


[Xianglin Shi]

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

Bottlenecks in signaling pathways are important





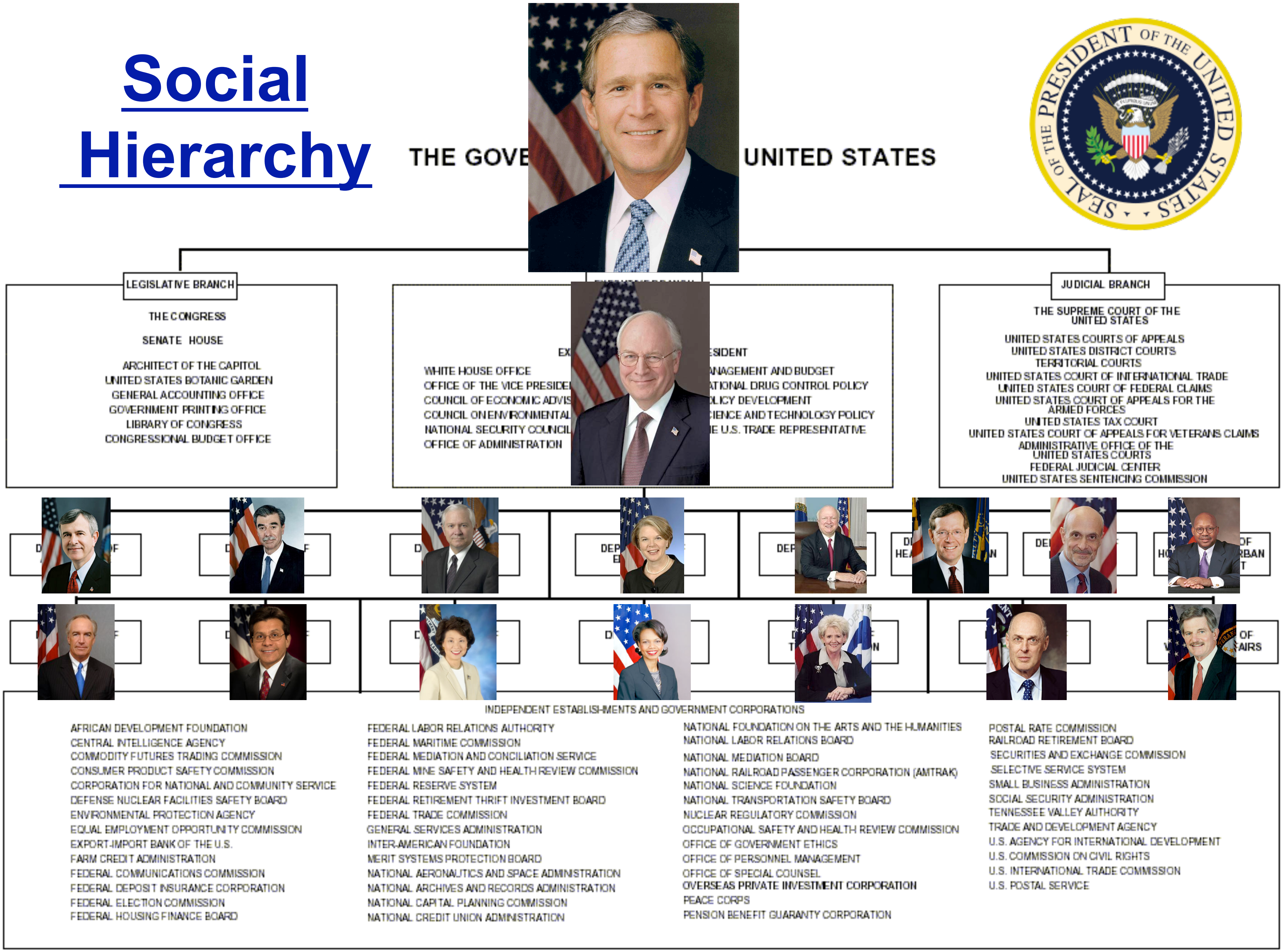
[Xianglin Shi]

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

Outline: Molecular Networks

- Why Networks?
- Network Structure: Key Positions
  - ◊ Hubs & Bottlenecks
  - ◊ Tops of a Hierachy
- Networks, Variation & the Environment
  - ◊ Which pathways change most with the environment

Social Hierarchy



THE GOVE

UNITED STATES

LEGISLATIVE BRANCH

THE CONGRESS

SENATE HOUSE

ARCHITECT OF THE CAPITOL
UNITED STATES BOTANIC GARDEN
GENERAL ACCOUNTING OFFICE
GOVERNMENT PRINTING OFFICE
LIBRARY OF CONGRESS
CONGRESSIONAL BUDGET OFFICE

EX
ESIDENT

WHITE HOUSE OFFICE
OFFICE OF THE VICE PRESIDE
COUNCIL OF ECONOMIC ADVIS
COUNCIL ON ENVIRONMENTAL
NATIONAL SECURITY COUNCIL
OFFICE OF ADMINISTRATION

NAGEMENT AND BUDGET
ATIONAL DRUG CONTROL POLICY
OLICY DEVELOPMENT
IENCE AND TECHNOLOGY POLICY
E U.S. TRADE REPRESENTATIVE

JUDICIAL BRANCH

THE SUPREME COURT OF THE UNITED STATES

UNITED STATES COURTS OF APPEALS
UNITED STATES DISTRICT COURTS
TERRITORIAL COURTS
UNITED STATES COURT OF INTERNATIONAL TRADE
UNITED STATES COURT OF FEDERAL CLAIMS
UNITED STATES COURT OF APPEALS FOR THE ARMED FORCES
UNITED STATES TAX COURT
UNITED STATES COURT OF APPEALS FOR VETERANS CLAIMS
ADMINISTRATIVE OFFICE OF THE UNITED STATES COURTS
FEDERAL JUDICIAL CENTER
UNITED STATES SENTENCING COMMISSION

D F
D F
D
DEP E
DEP
D HEA N
DE
HO OF RBAN T

F
D F
D F
D
T F N
V OF AIRS

INDEPENDENT ESTABLISHMENTS AND GOVERNMENT CORPORATIONS

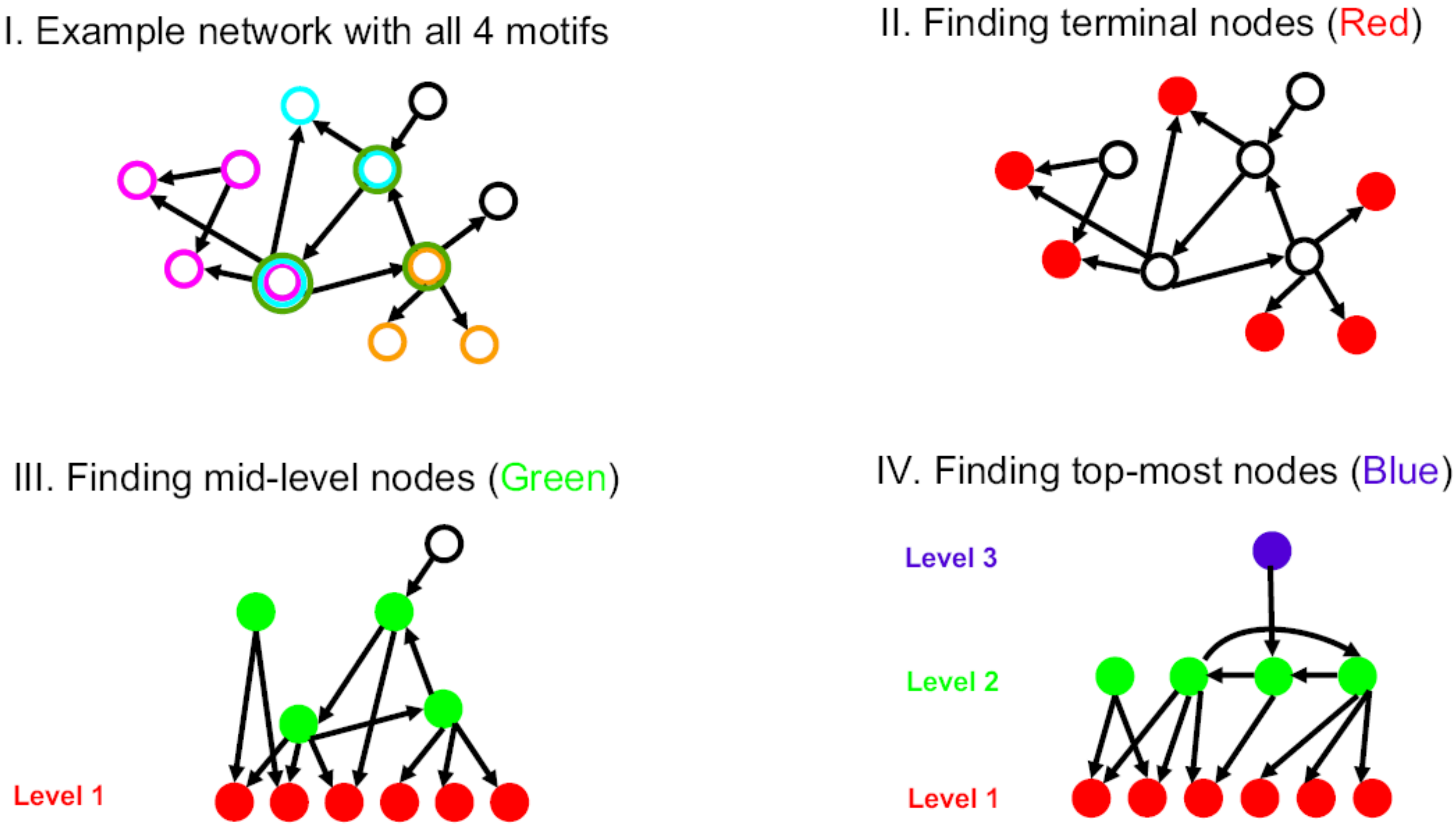
AFRICAN DEVELOPMENT FOUNDATION
CENTRAL INTELLIGENCE AGENCY
COMMODITY FUTURES TRADING COMMISSION
CONSUMER PRODUCT SAFETY COMMISSION
CORPORATION FOR NATIONAL AND COMMUNITY SERVICE
DEFENSE NUCLEAR FACILITIES SAFETY BOARD
ENVIRONMENTAL PROTECTION AGENCY
EQUAL EMPLOYMENT OPPORTUNITY COMMISSION
EXPORT-IMPORT BANK OF THE U.S.
FARM CREDIT ADMINISTRATION
FEDERAL COMMUNICATIONS COMMISSION
FEDERAL DEPOSIT INSURANCE CORPORATION
FEDERAL ELECTION COMMISSION
FEDERAL HOUSING FINANCE BOARD

FEDERAL LABOR RELATIONS AUTHORITY
FEDERAL MARITIME COMMISSION
FEDERAL MEDIATION AND CONCILIATION SERVICE
FEDERAL MINE SAFETY AND HEALTH REVIEW COMMISSION
FEDERAL RESERVE SYSTEM
FEDERAL RETIREMENT THRIFT INVESTMENT BOARD
FEDERAL TRADE COMMISSION
GENERAL SERVICES ADMINISTRATION
INTER-AMERICAN FOUNDATION
MERIT SYSTEMS PROTECTION BOARD
NATIONAL AERONAUTICS AND SPACE ADMINISTRATION
NATIONAL ARCHIVES AND RECORDS ADMINISTRATION
NATIONAL CAPITAL PLANNING COMMISSION
NATIONAL CREDIT UNION ADMINISTRATION

NATIONAL FOUNDATION ON THE ARTS AND THE HUMANITIES
NATIONAL LABOR RELATIONS BOARD
NATIONAL MEDIATION BOARD
NATIONAL RAILROAD PASSENGER CORPORATION (AMTRAK)
NATIONAL SCIENCE FOUNDATION
NATIONAL TRANSPORTATION SAFETY BOARD
NUCLEAR REGULATORY COMMISSION
OCCUPATIONAL SAFETY AND HEALTH REVIEW COMMISSION
OFFICE OF GOVERNMENT ETHICS
OFFICE OF PERSONNEL MANAGEMENT
OFFICE OF SPECIAL COUNSEL
OVERSEAS PRIVATE INVESTMENT CORPORATION
PEACE CORPS
PENSION BENEFIT GUARANTY CORPORATION

POSTAL RATE COMMISSION
RAILROAD RETIREMENT BOARD
SECURITIES AND EXCHANGE COMMISSION
SELECTIVE SERVICE SYSTEM
SMALL BUSINESS ADMINISTRATION
SOCIAL SECURITY ADMINISTRATION
TENNESSEE VALLEY AUTHORITY
TRADE AND DEVELOPMENT AGENCY
U.S. AGENCY FOR INTERNATIONAL DEVELOPMENT
U.S. COMMISSION ON CIVIL RIGHTS
U.S. INTERNATIONAL TRADE COMMISSION
U.S. POSTAL SERVICE

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



I. Example network with all 4 motifs

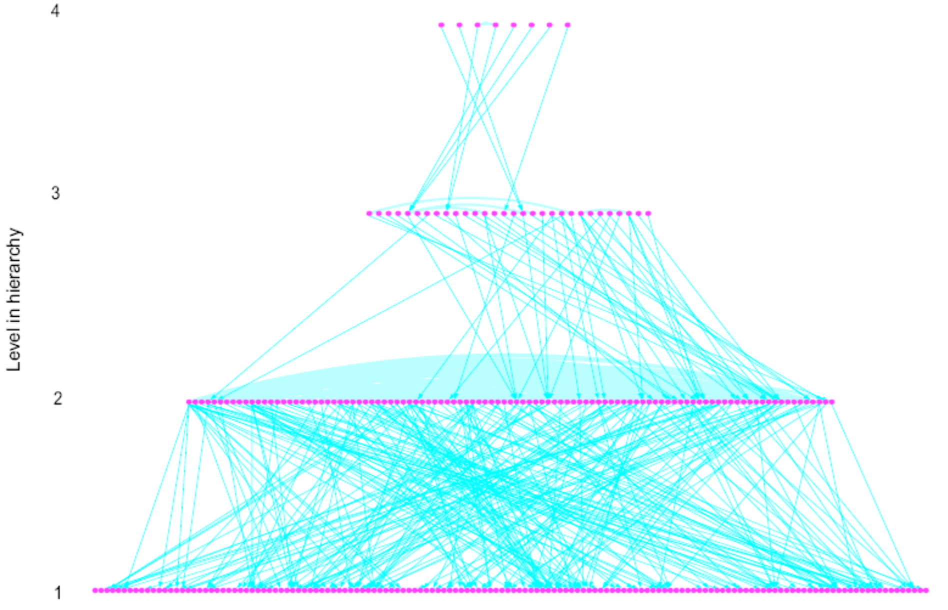
II. Finding terminal nodes (Red)

III. Finding mid-level nodes (Green)

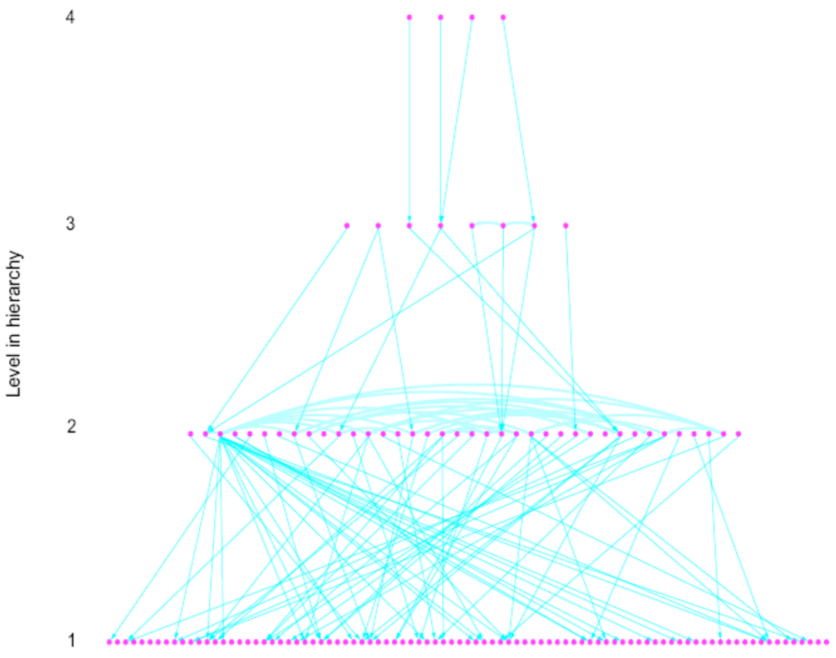
IV. Finding top-most nodes (Blue)

[Yu et al., PNAS (2006)]

Regulatory Networks have similar hierarchical structures



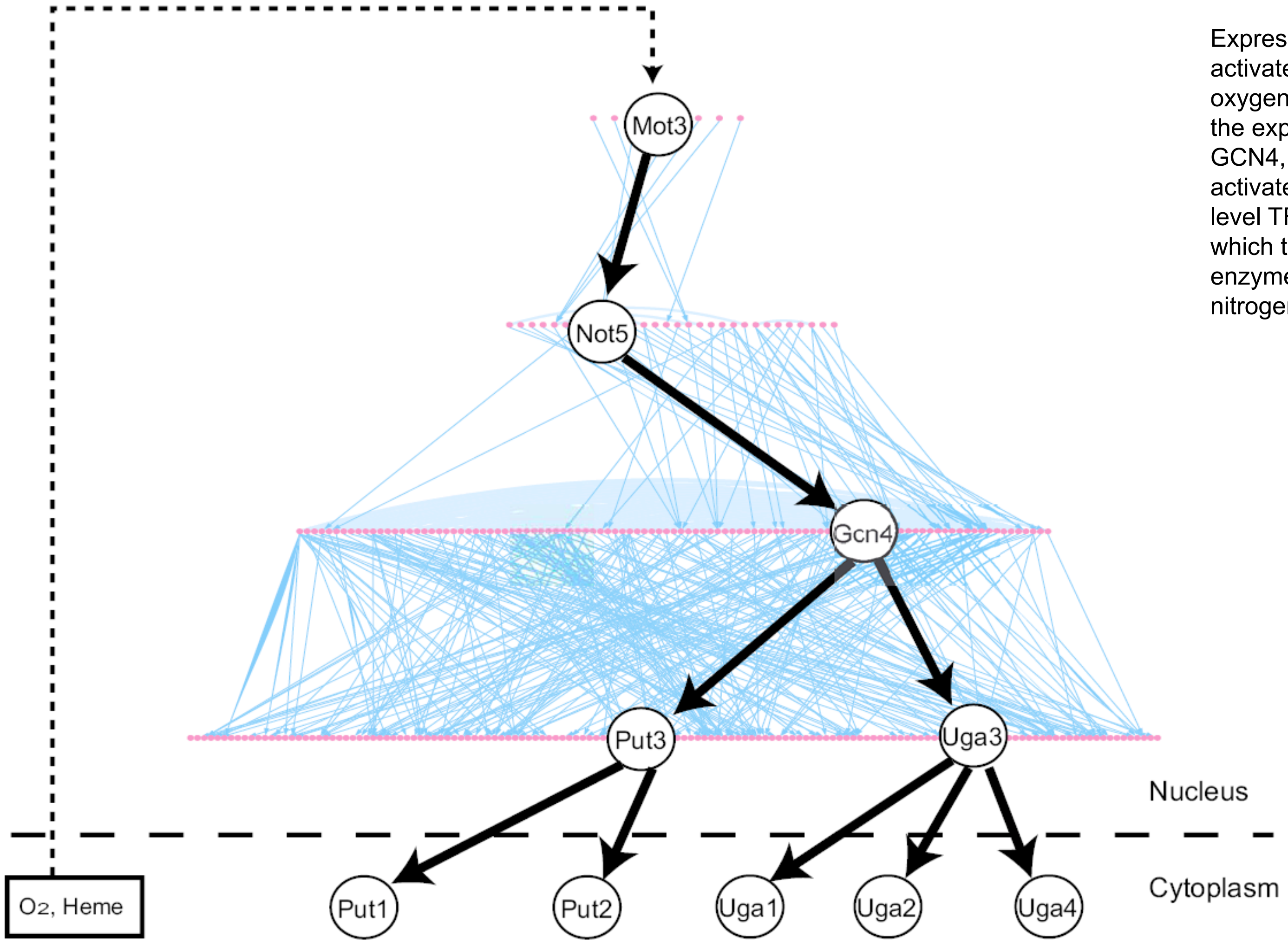
S. cerevisiae



E. coli

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

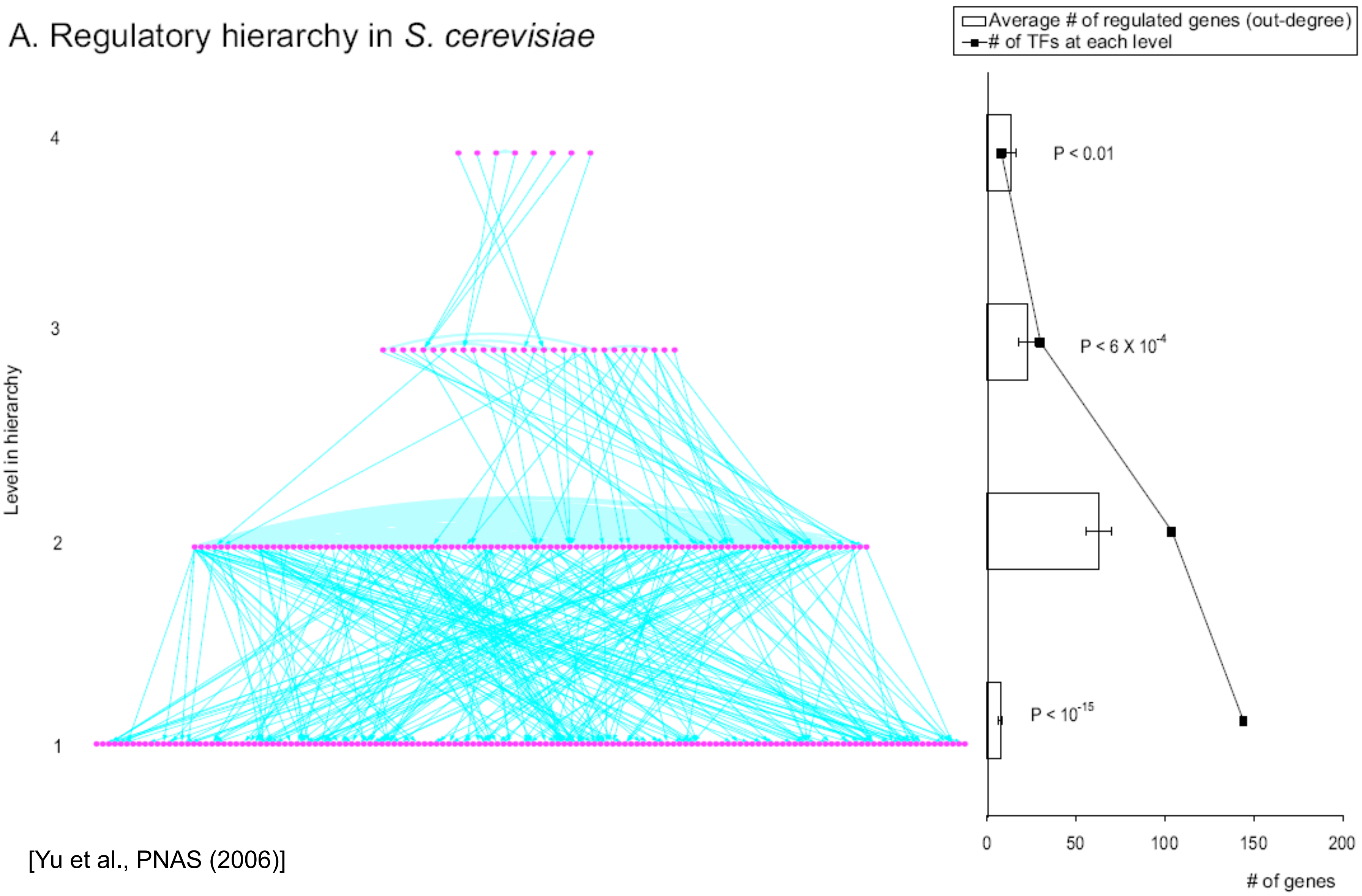
Example of Path Through Regulatory Network



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.

[Yu et al., PNAS (2006)]

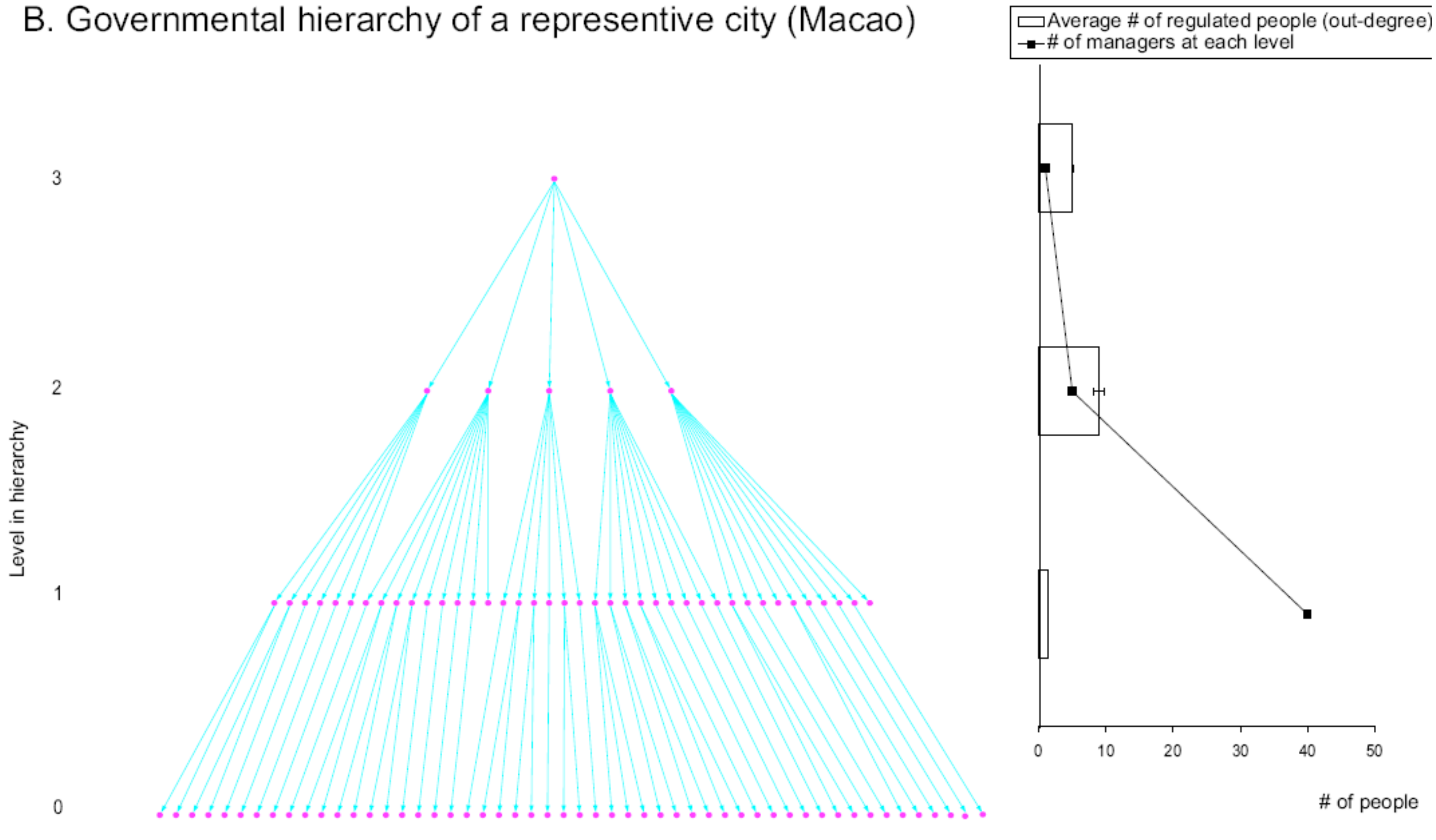
Yeast Regulatory Hierarchy: the Middle-managers Rule



A. Regulatory hierarchy in *S. cerevisiae*

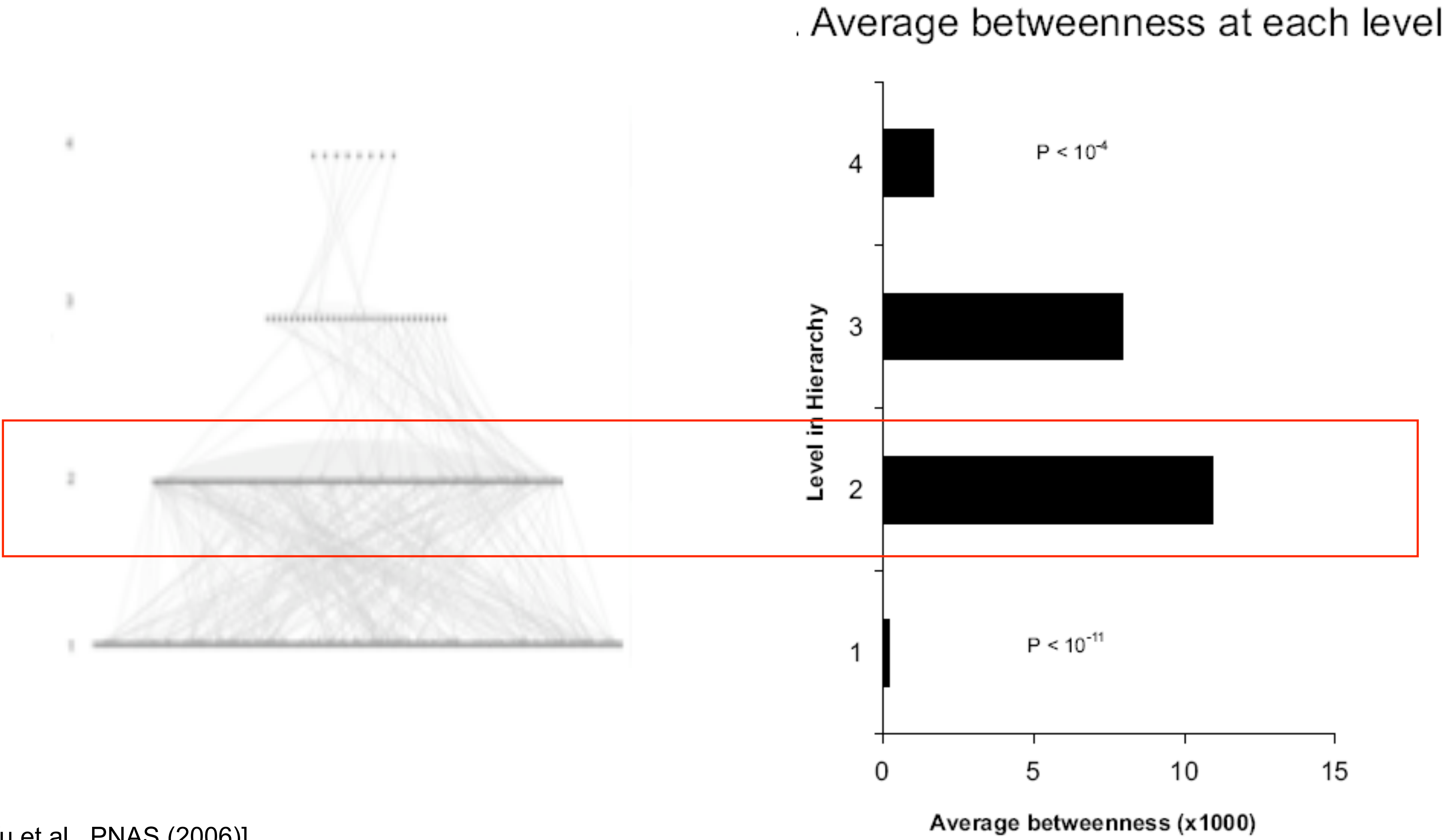
[Yu et al., PNAS (2006)]

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



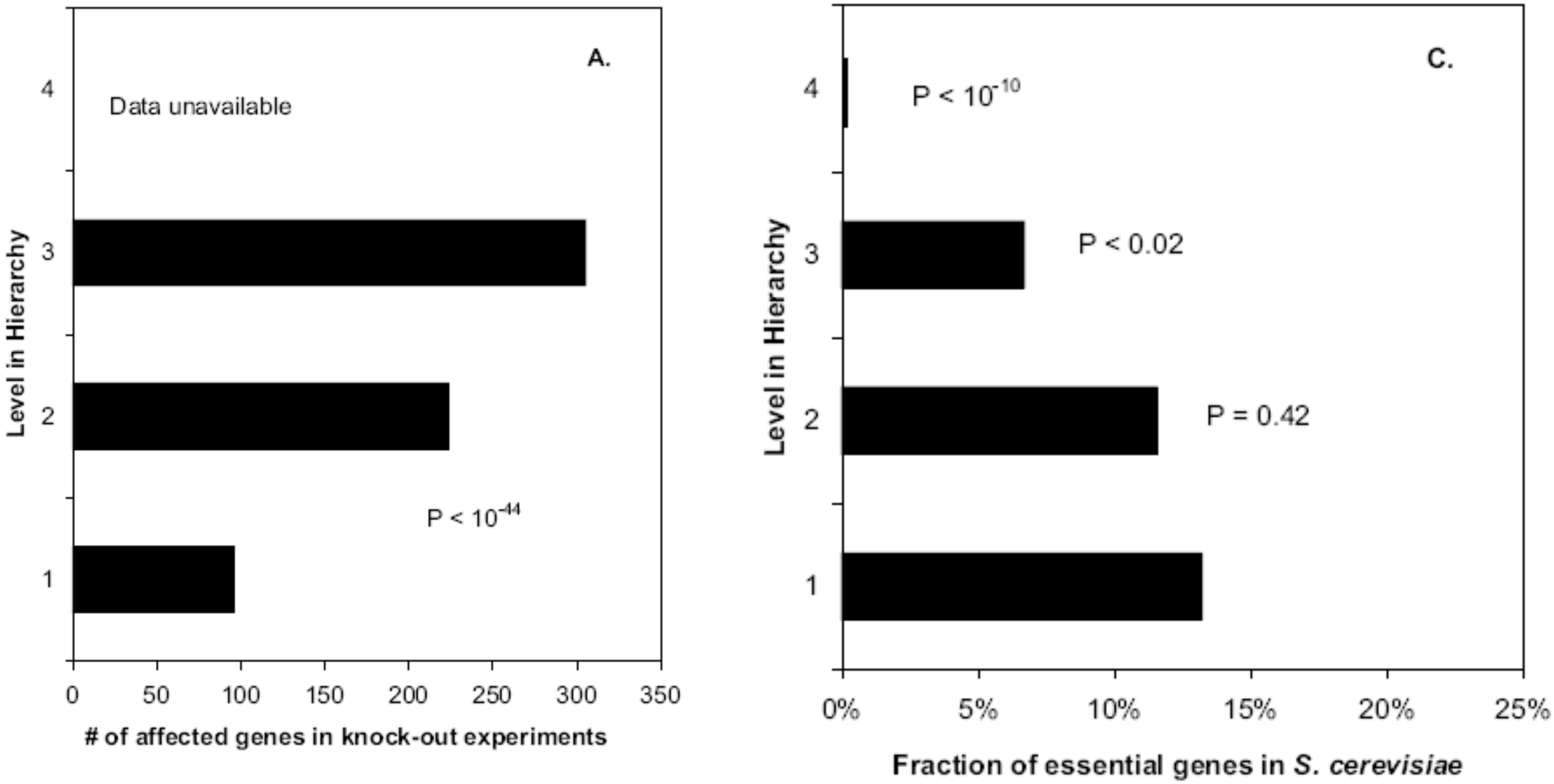
B. Governmental hierarchy of a representive city (Macao)

Characteristics of Regulatory Hierarchy: Middle Managers are Information Flow Bottlenecks



[Yu et al., PNAS (2006)]

Characteristics of Regulatory Hierarchy: The Paradox of Influence and Essentiality



[Yu et al., PNAS (2006)]

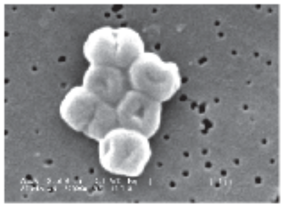
Outline: Molecular Networks

- Why Networks?
- Network Structure: Key Positions
  - ◊ Hubs & Bottlenecks
  - ◊ Tops of a Hierachy
- Networks, Variation & the Environment
  - ◊ Which pathways change most with the environment

What is metagenomics?

Genomics Approach

Culture Microbes

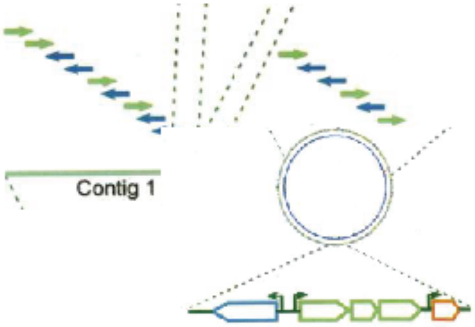


Extract DNA

Sequence

ATCGTATA
CGCGAAG
ACGTCTGA
AGTGCTGCT

Assemble and Annotate



PROBLEM: Estimated that less than 1% can be cultured in the lab

Metagenomics Approach

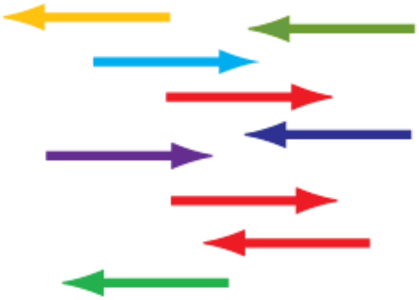
Collect Sample

Extract DNA

Sequence

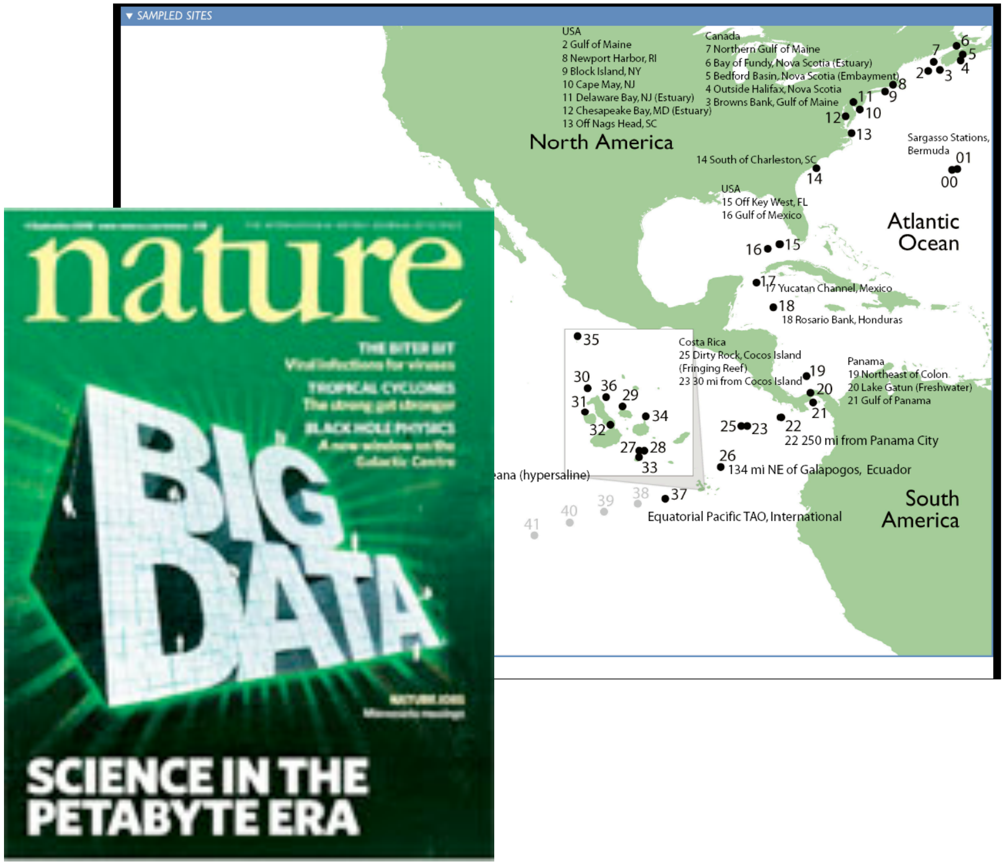
ATCGTGATAGATGATAGTAGA
ATGCTGCATGCATCTAGCACT
ACAGTAGCTAGCTACGTACTA
CAGCTGACTAGCTAGCTAGCT
ACGTAGCATGCTAGCTAGCAG
ACGTACGTAGCTAGCTAGCTAG
ACGTACGTACGTAGCTAGCATC
AGTCGACTGAGCCAGTGATGAT
ACGATGCATGAGCAGATGCTAC
AGATCGTAGCATGCTAGCATGCT
ACGTACGTAGCTAGCTAGCTAAG
AGCTAGCATGCTAGTAGCATGAG
ACGATGCTAGCTAGCTAGCTGATA
TCGATCAGCATGCTACGATGCAAG
ACGATCGATGCTAGCTAGCTAGCAT
AGCTAGCTAGTCAGCTAGCTAGATG

Partially Assemble and Annotate



PROBLEM: Lose information about which gene belongs to which microbe.

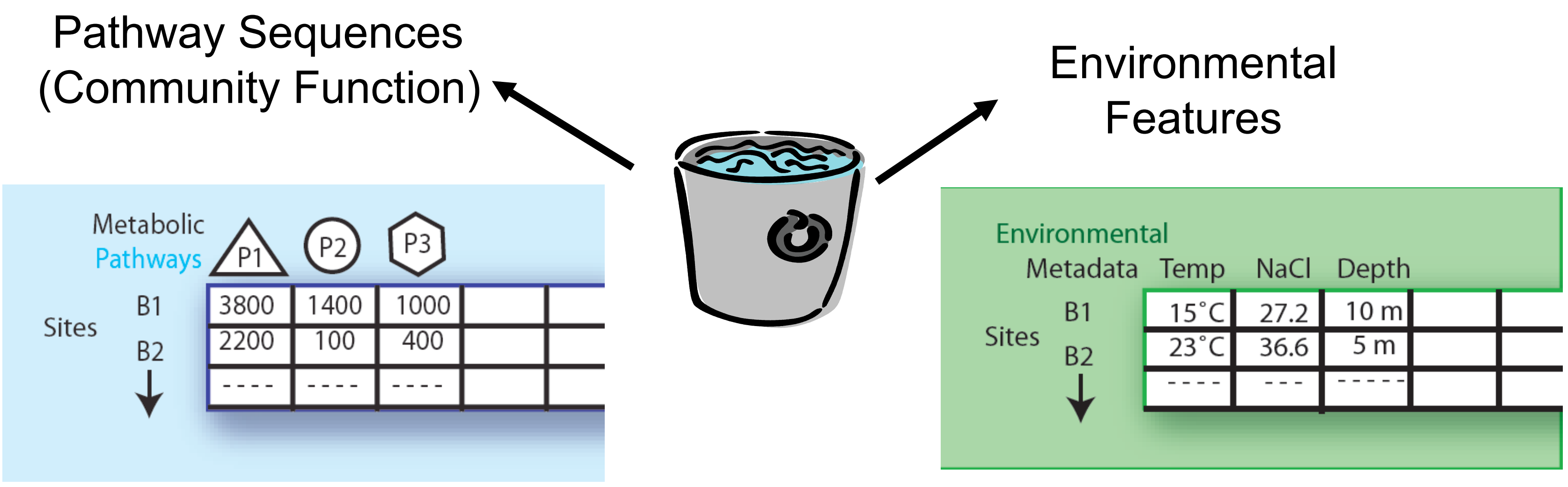
Global Ocean Survey Statistics (GOS)


SAMPLED SITES
USA
2 Gulf of Maine
8 Newport Harbor, RI
9 Block Island, NY
10 Cape May, NJ
11 Delaware Bay, NJ (Estuary)
12 Chesapeake Bay, MD (Estuary)
13 Off Nags Head, SC
Canada
7 Northern Gulf of Maine
6 Bay of Fundy, Nova Scotia (Estuary)
5 Bedford Basin, Nova Scotia (Embayment)
4 Outside Halifax, Nova Scotia
3 Browns Bank, Gulf of Maine
North America
Sargasso Stations, Bermuda
01
00
14 South of Charleston, SC
USA
15 Off Key West, FL
16 Gulf of Mexico
Atlantic Ocean
17 Yucatan Channel, Mexico
18 Rosario Bank, Honduras
Costa Rica
25 Dirty Rock, Cocos Island (Fringing Reef)
23 30 mi from Cocos Island
Panama
19 Northeast of Colon
20 Lake Gatun (Freshwater)
21 Gulf of Panama
22 250 mi from Panama City
26 134 mi NE of Galapogos, Ecuador
ana (hypersaline)
Equatorial Pacific TAO, International
South America
nature
THE BITER BIT
Viral infections for viruses
TROPICAL CYCLONES
The strong get stronger
BLACK HOLE PHYSICS
A new window on the Galactic Centre
BIG DATA
SCIENCE IN THE PETABYTE ERA

6.25 GB of data
7.7M Reads
1 million CPU hours
to process

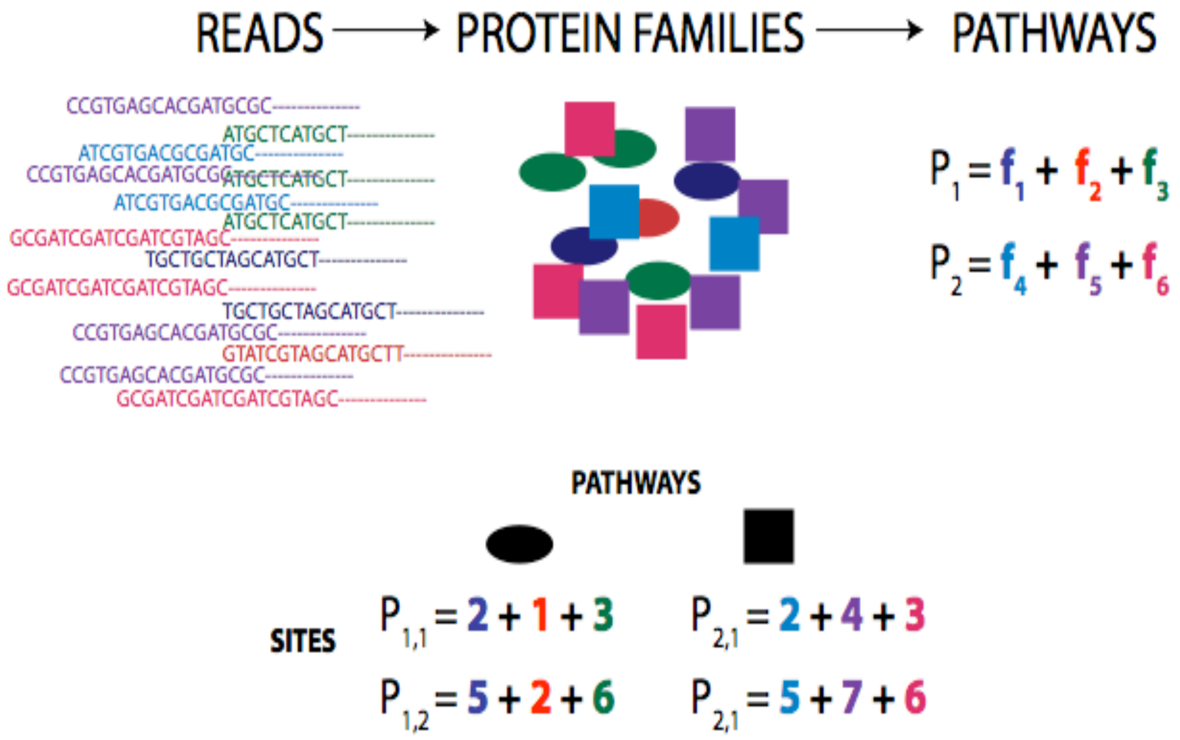
Rusch, et al., PLOS Biology 2007

Expressing data as matrices indexed by site, env. var., and pathway usage



Metabolic Pathways		P1	P2	P3		
Sites	B1	3800	1400	1000		
	B2	2200	100	400		
	↓	----	----	----		

Environmental Metadata		Temp	NaCl	Depth		
Sites	B1	15˚C	27.2	10 m		
	B2	23˚C	36.6	5 m		
	↓	----	---	-----		



$P_1 = f_1 + f_2 + f_3$

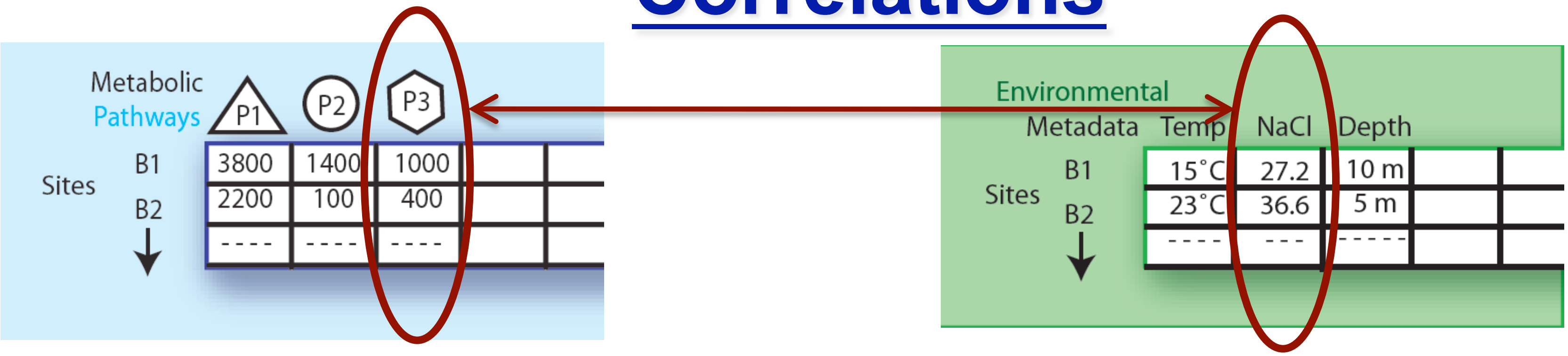
$P_2 = f_4 + f_5 + f_6$

$P_{1,1} = 2 + 1 + 3$ $P_{2,1} = 2 + 4 + 3$

$P_{1,2} = 5 + 2 + 6$ $P_{2,1} = 5 + 7 + 6$

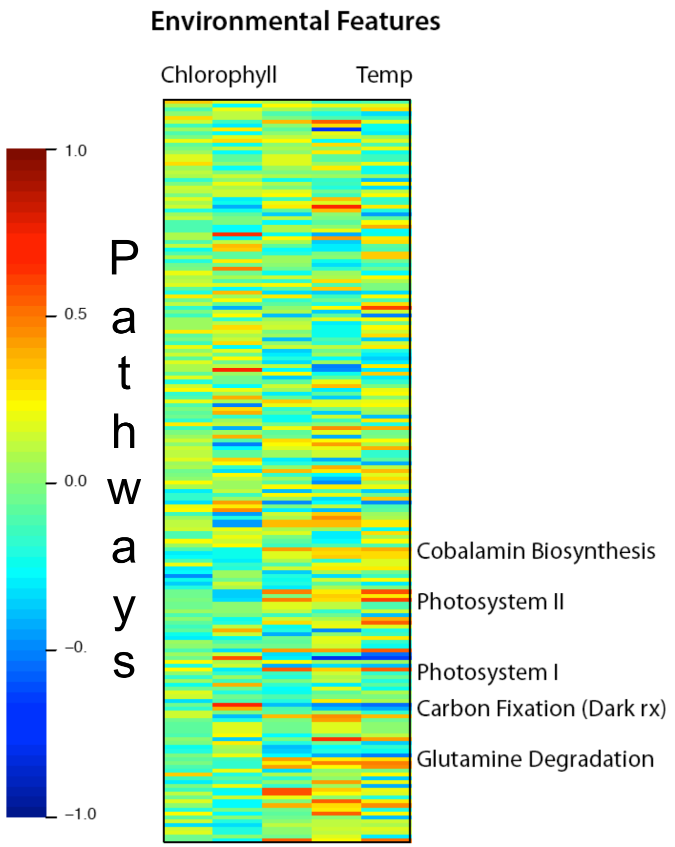
[Rusch et. al., (2007) PLOS Biology;
Gianoulis et al., PNAS (in press, 2009]

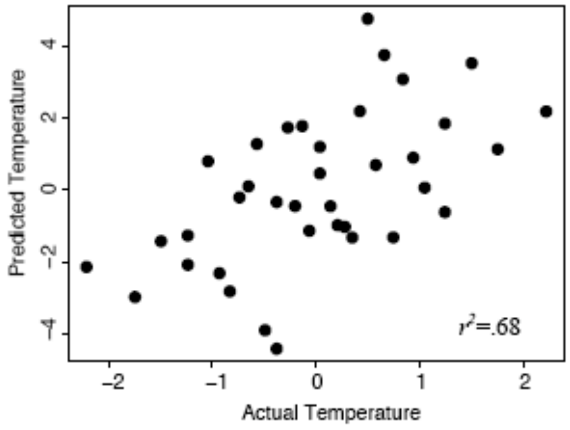
Simple Relationships: Pairwise Correlations



	Metabolic Pathways	P1	P2	P3
Sites	B1	3800	1400	1000
	B2	2200	100	400
		----	----	----

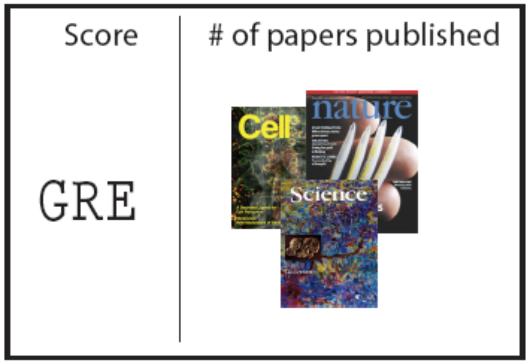
	Environmental Metadata	Temp	NaCl	Depth
Sites	B1	15˚C	27.2	10 m
	B2	23˚C	36.6	5 m
		----	---	-----

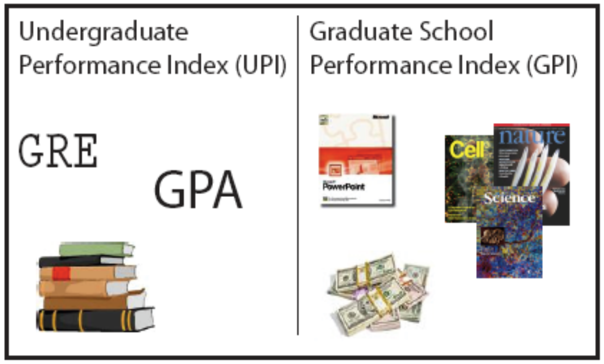




[Gianoulis et al., PNAS (in press, 2009)]

Canonical Correlation Analysis: Simultaneous weighting



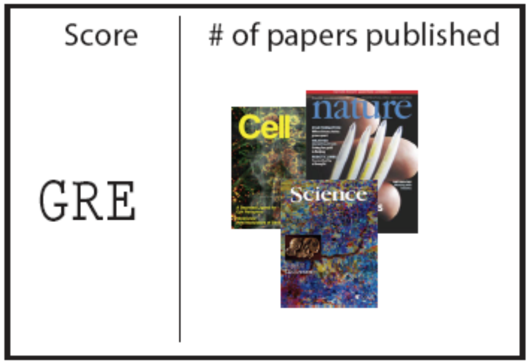

Undergraduate Performance Index (UPI)
Graduate School Performance Index (GPI)
GRE
GPA

UPI = a GRE + b + c GPA

GPI = a' + b' + c'

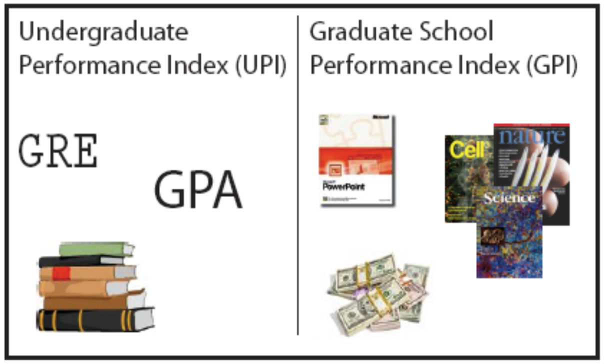
[Gianoulis et al., PNAS (in press, 2009)]

Canonical Correlation Analysis: Simultaneous weighting



Score | # of papers published

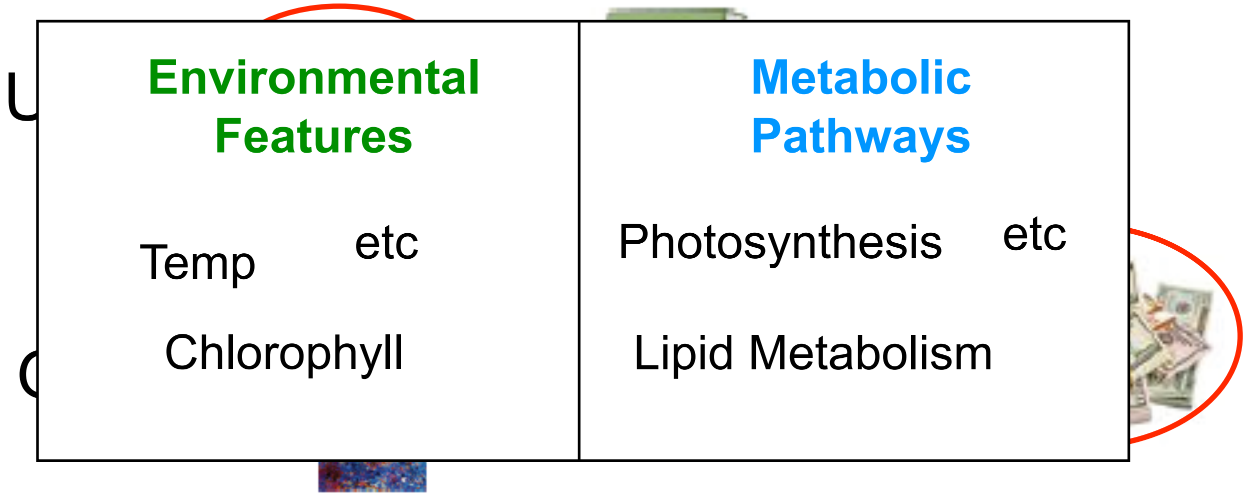
GRE



Undergraduate Performance Index (UPI) | Graduate School Performance Index (GPI)

GRE

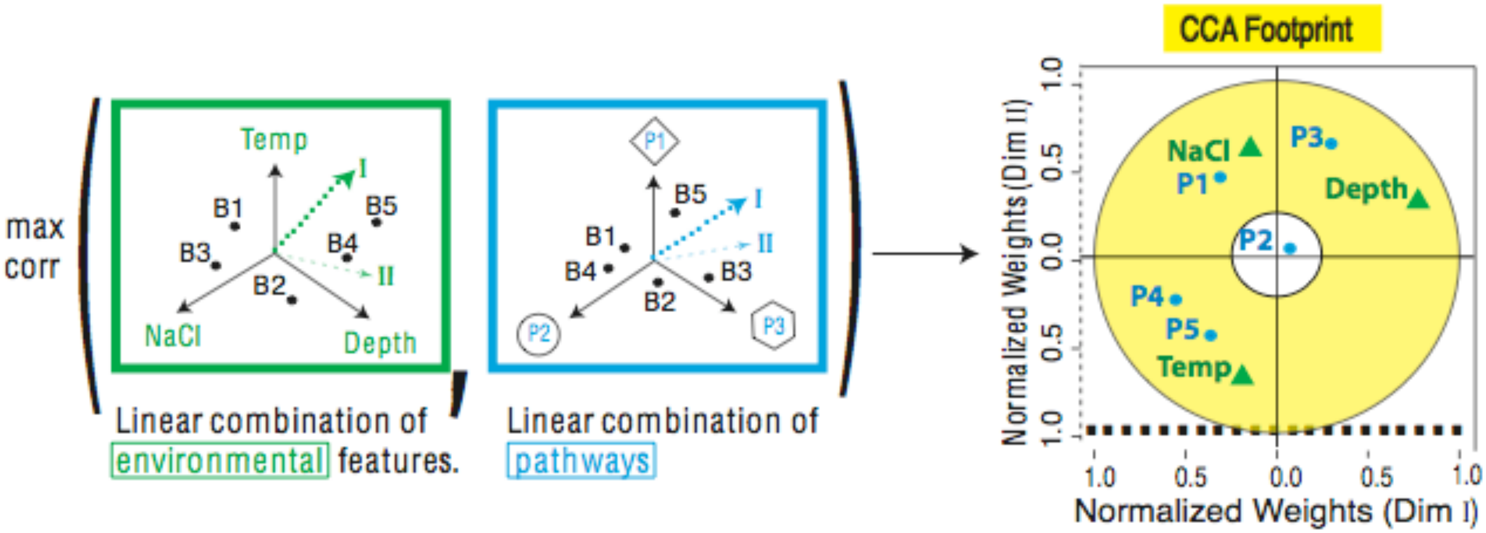
GPA



Environmental Features	Metabolic Pathways
Temp etc	Photosynthesis etc
Chlorophyll	Lipid Metabolism

[Gianoulis et al., PNAS (in press, 2009)]

Environmental-Metabolic Space

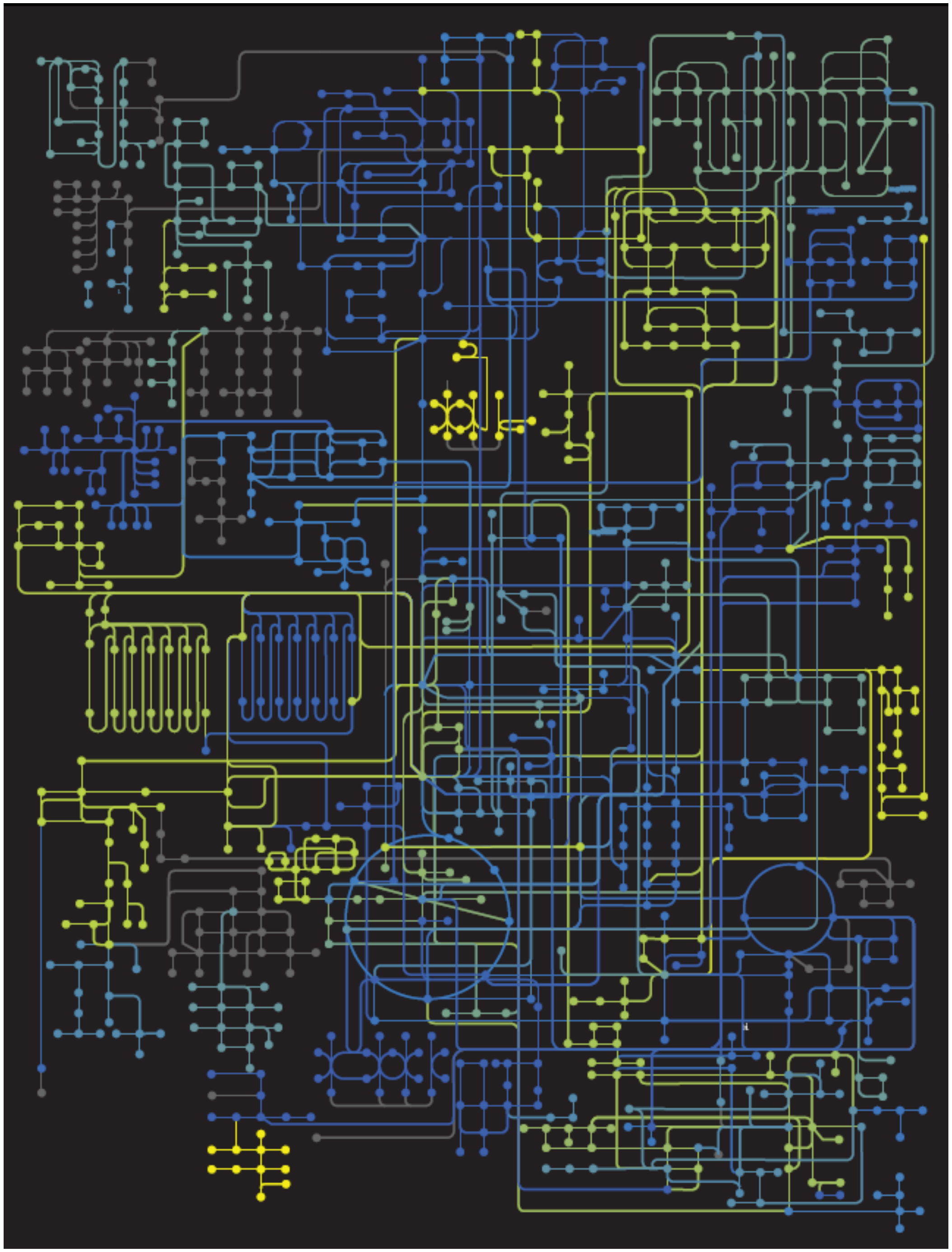


The goal of this technique is to interpret cross-variance matrices
We do this by defining a change of basis.

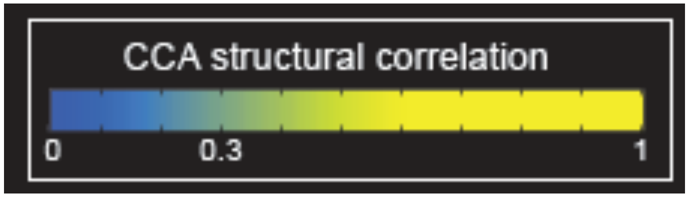
$$\text{Given } X=\{x_1,x_2,....,x_n\} \text{ and } Y=\{y_1,y_2,...,y_m\}$$

$$C=\begin{matrix}\Sigma_X & \Sigma_{X,Y} \\ \Sigma_Y & \Sigma_{Y,X}\end{matrix} \qquad \max_{a,b} Corr(U,V)=\frac{a'\Sigma_{12}\,b}{\sqrt{a'\Sigma_{11}a}\sqrt{b'\Sigma_{22}\,b}}$$

[Gianoulis et al., PNAS (in press, 2009)]

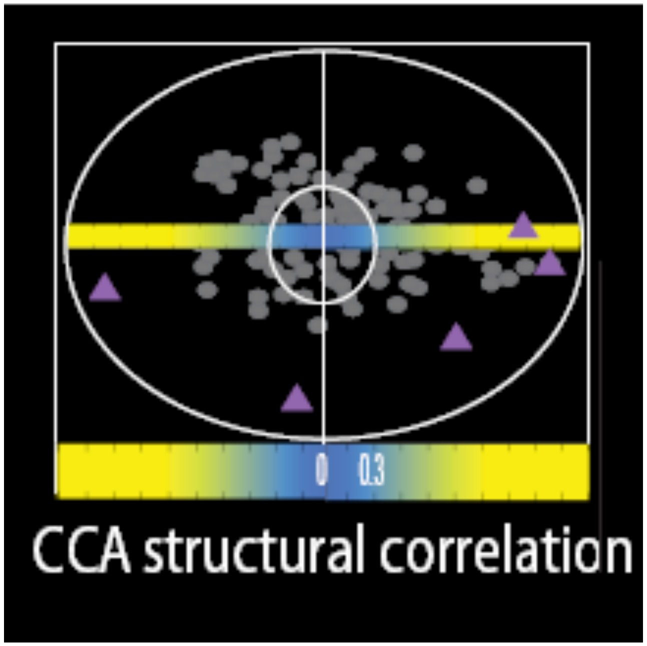


Strength of Pathway co-variation with environment



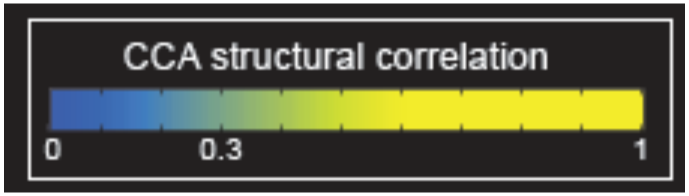
Environmentally invariant

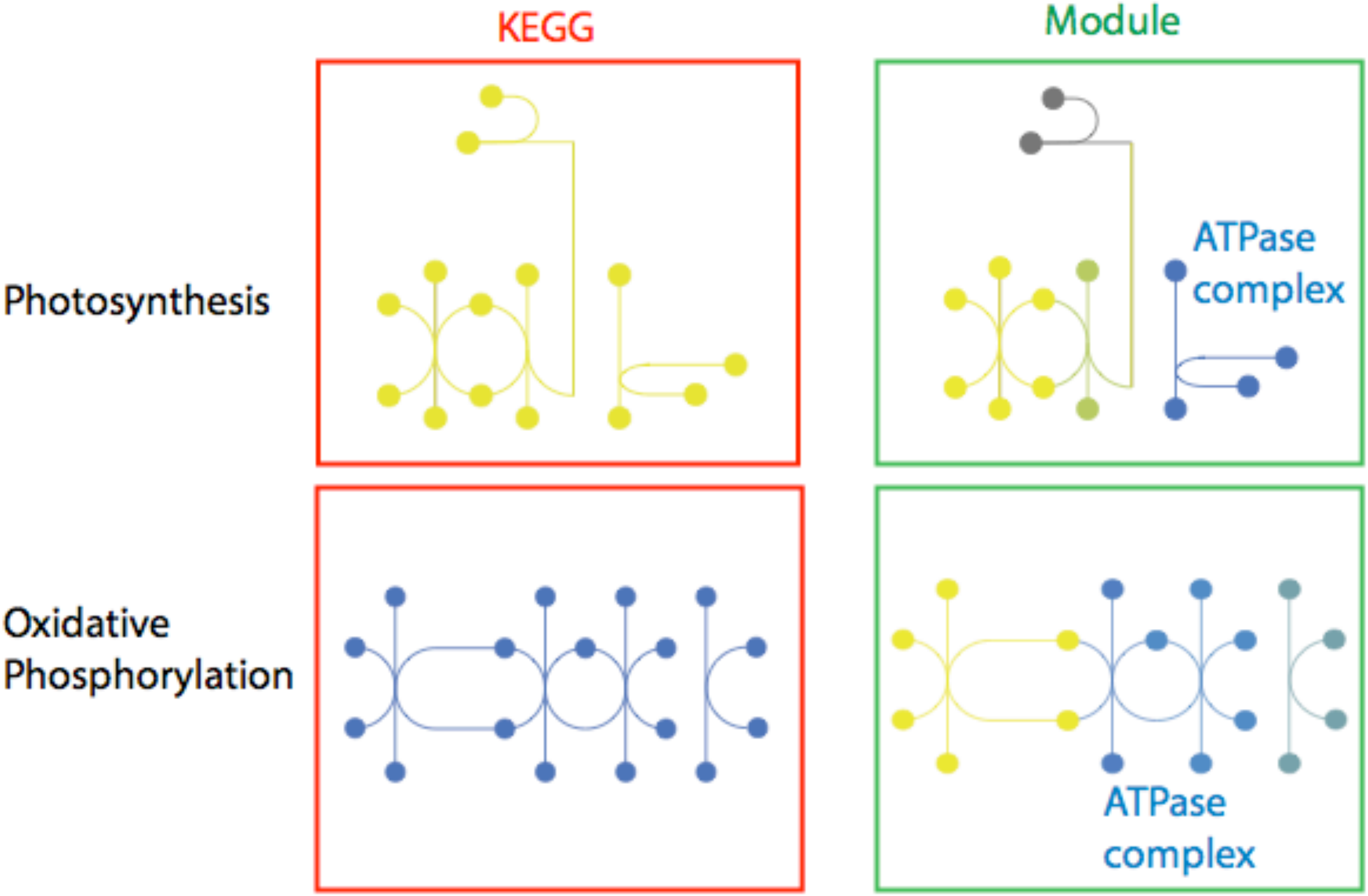
Environmentally variant



[Gianoulis et al., PNAS (in press, 2009)]

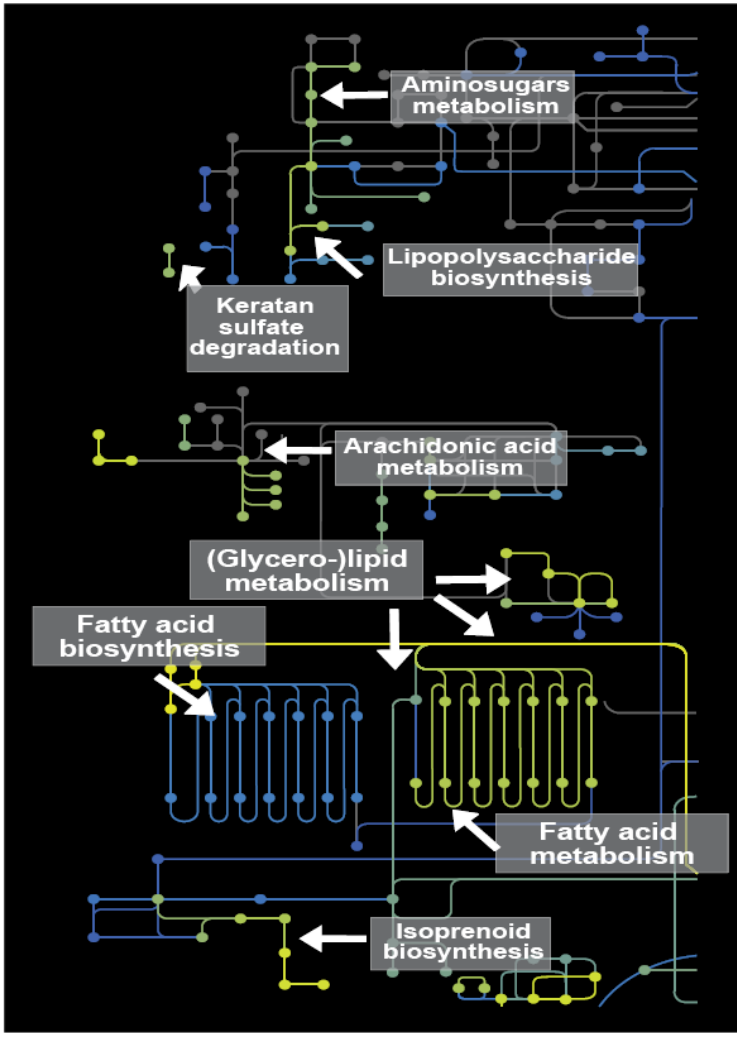
Conclusion #1: energy conversion strategy, temp and depth

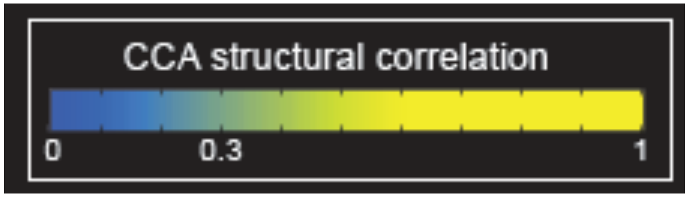




[Gianoulis et al., PNAS (in press, 2009)]

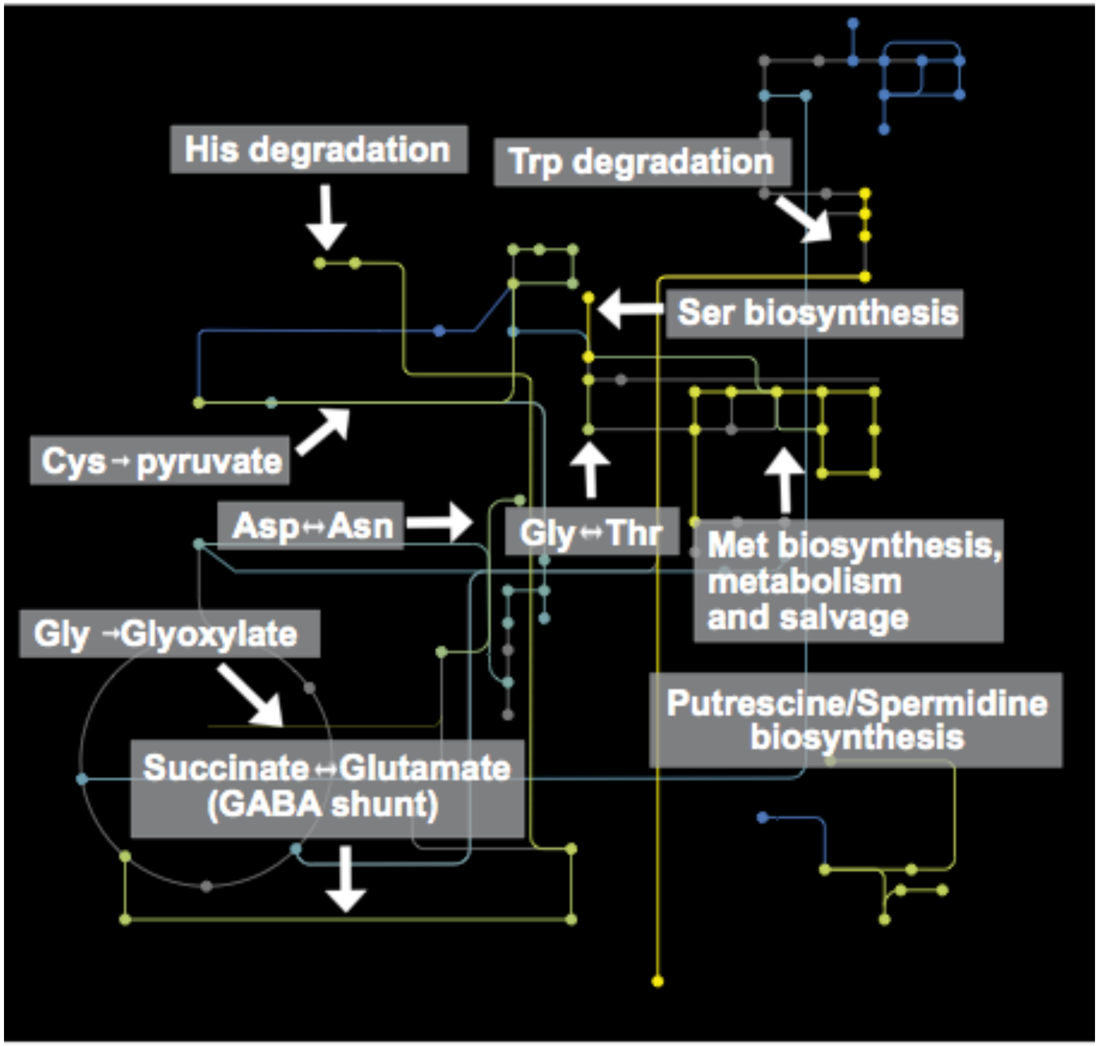
Conclusion #2: Outer Membrane components vary the environment


Aminosugars metabolism
Lipopolysaccharide biosynthesis
Keratan sulfate degradation
Arachidonic acid metabolism
(Glycero-)lipid metabolism
Fatty acid biosynthesis
Fatty acid metabolism
Isoprenoid biosynthesis



[Gianoulis et al., PNAS (in press, 2009)]

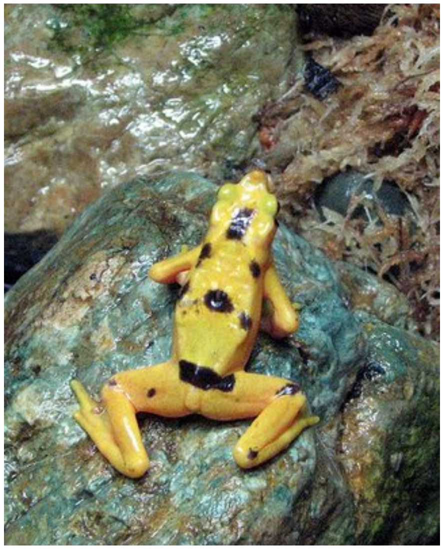
Conclusion #3: Covariation of AA biosynthesis and Import


His degradation
Trp degradation
Ser biosynthesis
Cys→pyruvate
Asp↔Asn
Gly↔Thr
Met biosynthesis, metabolism and salvage
Gly →Glyoxylate
Putrescine/Spermidine biosynthesis
Succinate↔Glutamate (GABA shunt)

Why is their fluctuation
in amino acid metabolism?
Is there a feature(s) that
underlies those that are
environmentally-variant
as opposed to those which are not?

[Gianoulis et al., PNAS (in press, 2009)]

Biosensors:
Beyond Canaries in a Coal Mine

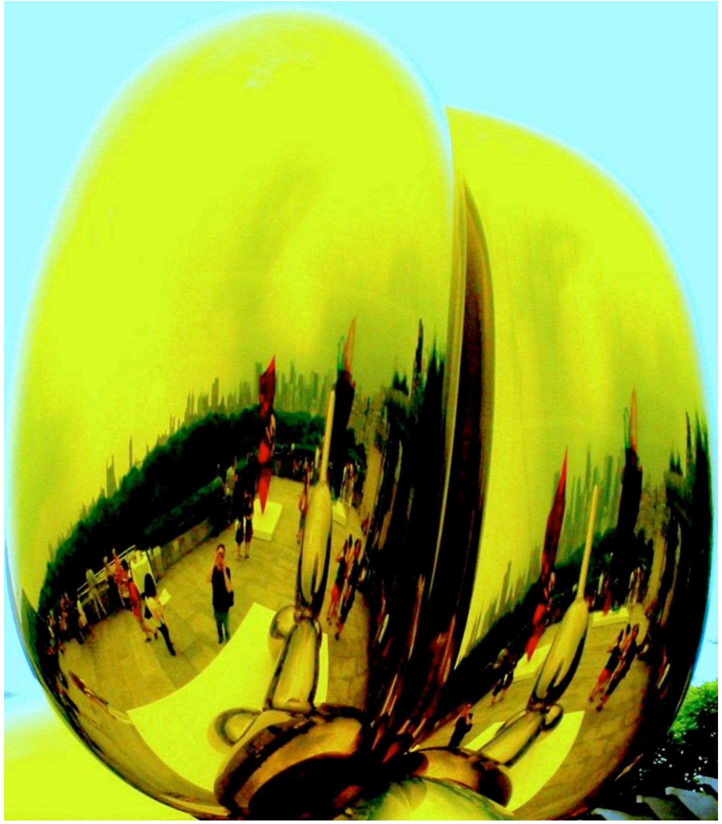


[Gianoulis et al., PNAS (in press, 2009)]

Conclusions

- Developing Standardized Descriptions of Protein Function
- Gene Naming
- Betweenness is an important global network statistic
  - ◊ Bottlenecks are more correlated with essentiality than hubs in regulatory networks
- Regulatory Network Hierarchies
  - ◊ Middle managers dominate, sitting at info. flow bottlenecks
  - ◊ Paradox of influence and essentiality
  - ◊ Topmost proteins sit at center of interaction network

Conclusions: Networks Dynamics across Environments



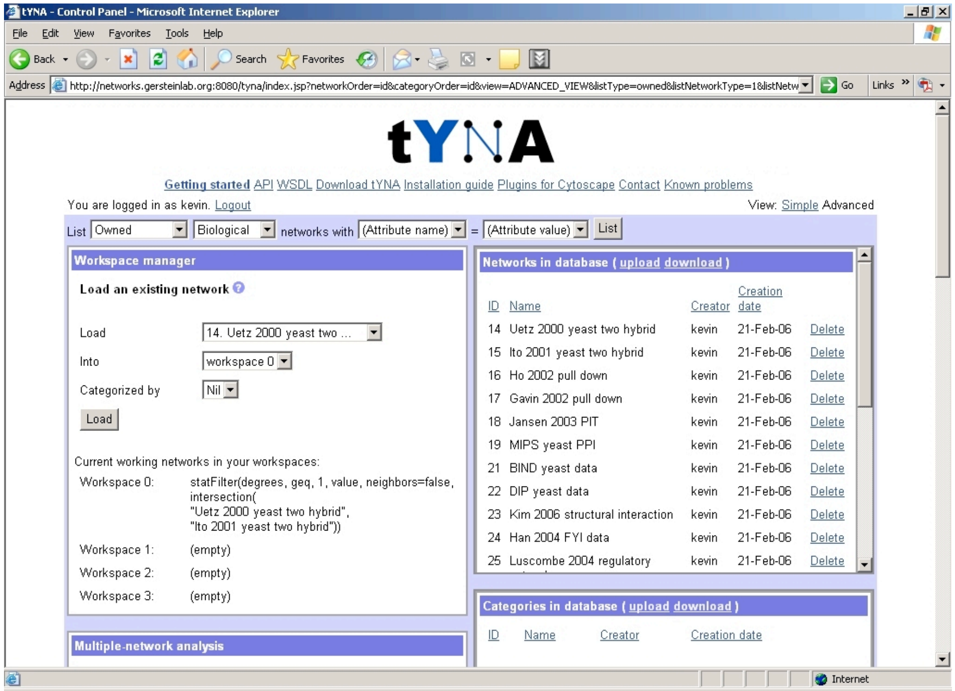
- Developed and adapted techniques to connect quantitative features of environment to metabolism.
- Applied to available aquatic datasets, we identified footprints that were predictive of their environment (potentially could be used as biosensor).
- Strong correlation exists between a community's energy conversion strategies and its environmental parameters (e.g. temperature and chlorophyll).
- Suggest that limiting amounts of cofactor can (partially) explain increased import of amino acids in nutrient-limited conditions.

TopNet
Topology of Networks

- an automated web tool

tYNA

(vers. 2 : "TopNet-like Yale Network Analyzer")


tYNA - Control Panel - Microsoft Internet Explorer

Address http://networks.gersteinlab.org:8080/tyna/index.jsp?networkOrder=id&categoryOrder=id&view=ADVANCED_VIEW&listType=owned&listNetworkType=1&listNetw

tYNA

Getting started API WSDL Download tYNA Installation guide Plugins for Cytoscape Contact Known problems

You are logged in as kevin. Logout

View: Simple Advanced

List Owned Biological networks with (Attribute name) = (Attribute value) List

Workspace manager

Load an existing network

Load: 14. Uetz 2000 yeast two ...
Into: workspace 0
Categorized by: Nil
Load

Current working networks in your workspaces:

Workspace 0: statFilter(degrees, geq, 1, value, neighbors=false, intersection("Uetz 2000 yeast two hybrid", "Ito 2001 yeast two hybrid"))
Workspace 1: (empty)
Workspace 2: (empty)
Workspace 3: (empty)

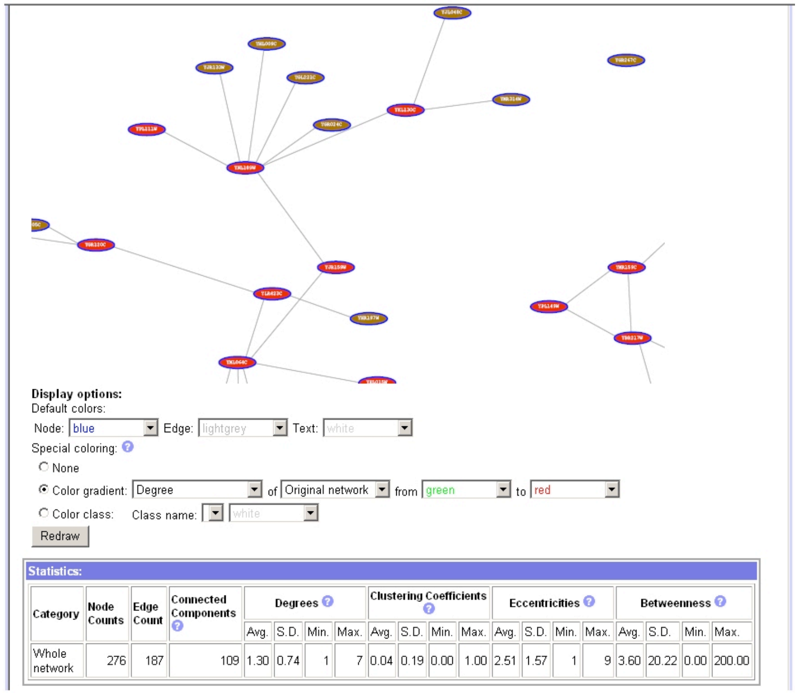
Multiple-network analysis

Networks in database (upload download)

ID	Name	Creator	Creation date	
14	Uetz 2000 yeast two hybrid	kevin	21-Feb-06	Delete
15	Ito 2001 yeast two hybrid	kevin	21-Feb-06	Delete
16	Ho 2002 pull down	kevin	21-Feb-06	Delete
17	Gavin 2002 pull down	kevin	21-Feb-06	Delete
18	Jansen 2003 PIT	kevin	21-Feb-06	Delete
19	MIPS yeast PPI	kevin	21-Feb-06	Delete
21	BIND yeast data	kevin	21-Feb-06	Delete
22	DIP yeast data	kevin	21-Feb-06	Delete
23	Kim 2006 structural interaction	kevin	21-Feb-06	Delete
24	Han 2004 FYI data	kevin	21-Feb-06	Delete
25	Luscombe 2004 regulatory	kevin	21-Feb-06	Delete

Categories in database (upload download)

ID	Name	Creator	Creation date


Display options:
Default colors:
Node: blue Edge: lightgrey Text: white
Special coloring:
○ None
● Color gradient: Degree of Original network from green to red
○ Color class: Class name: white
Redraw

Statistics:

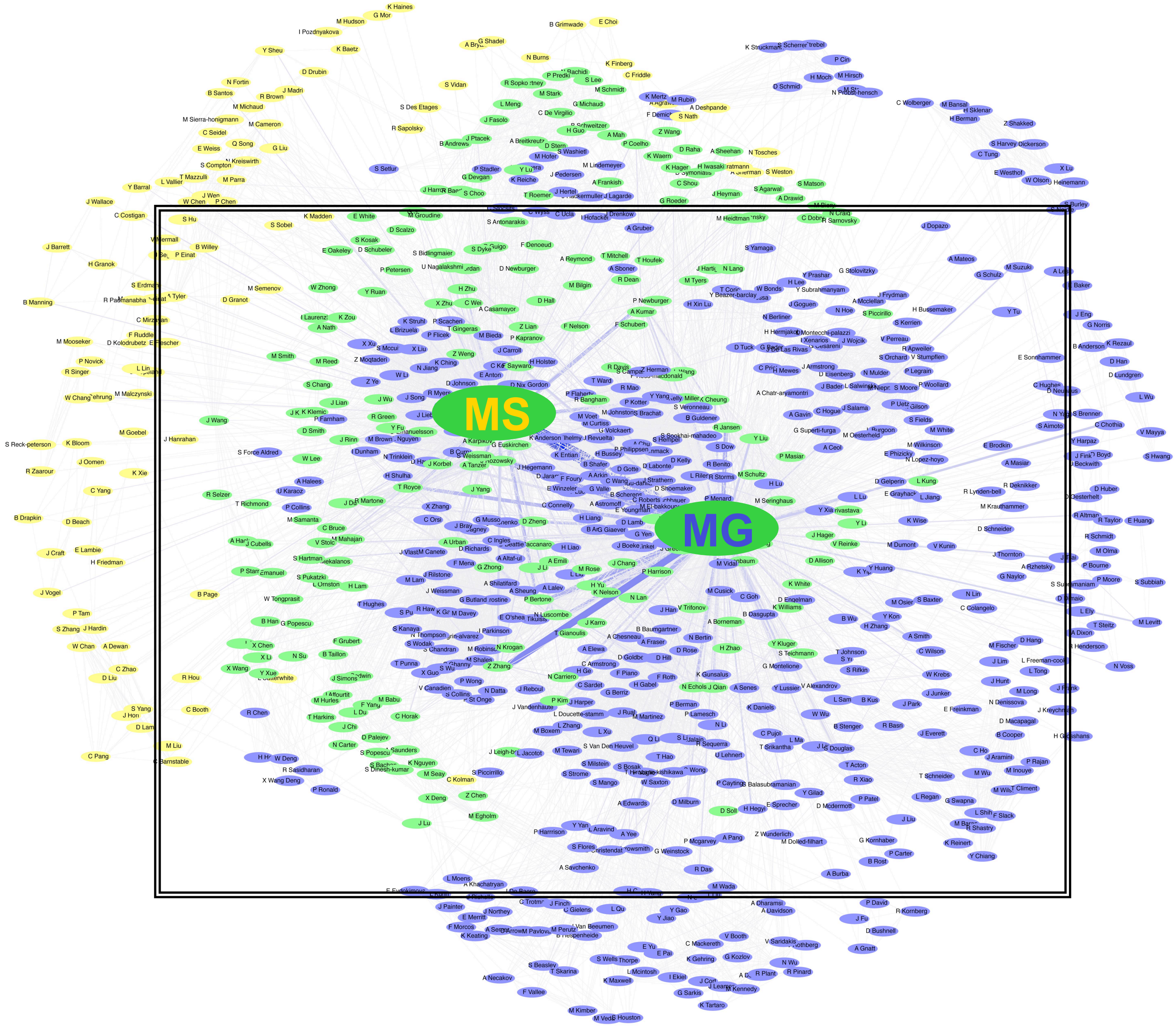
Category	Node Counts	Edge Count	Connected Components	Degrees				Clustering Coefficients				Eccentricities				Betweenness			
				Avg.	S.D.	Min.	Max.	Avg.	S.D.	Min.	Max.	Avg.	S.D.	Min.	Max.	Avg.	S.D.	Min.	Max.
Whole network	276	187	109	1.30	0.74	1	7	0.04	0.19	0.00	1.00	2.51	1.57	1	9	3.60	20.22	0.00	200.00

Normal website + Downloaded code (JAVA)
+ Web service (SOAP) with Cytoscape plugin

[Yu et al., NAR (2004); Yip et al. Bioinfo. (2006);
Similar tools include Cytoscape.org, Idekar, Sander et al]

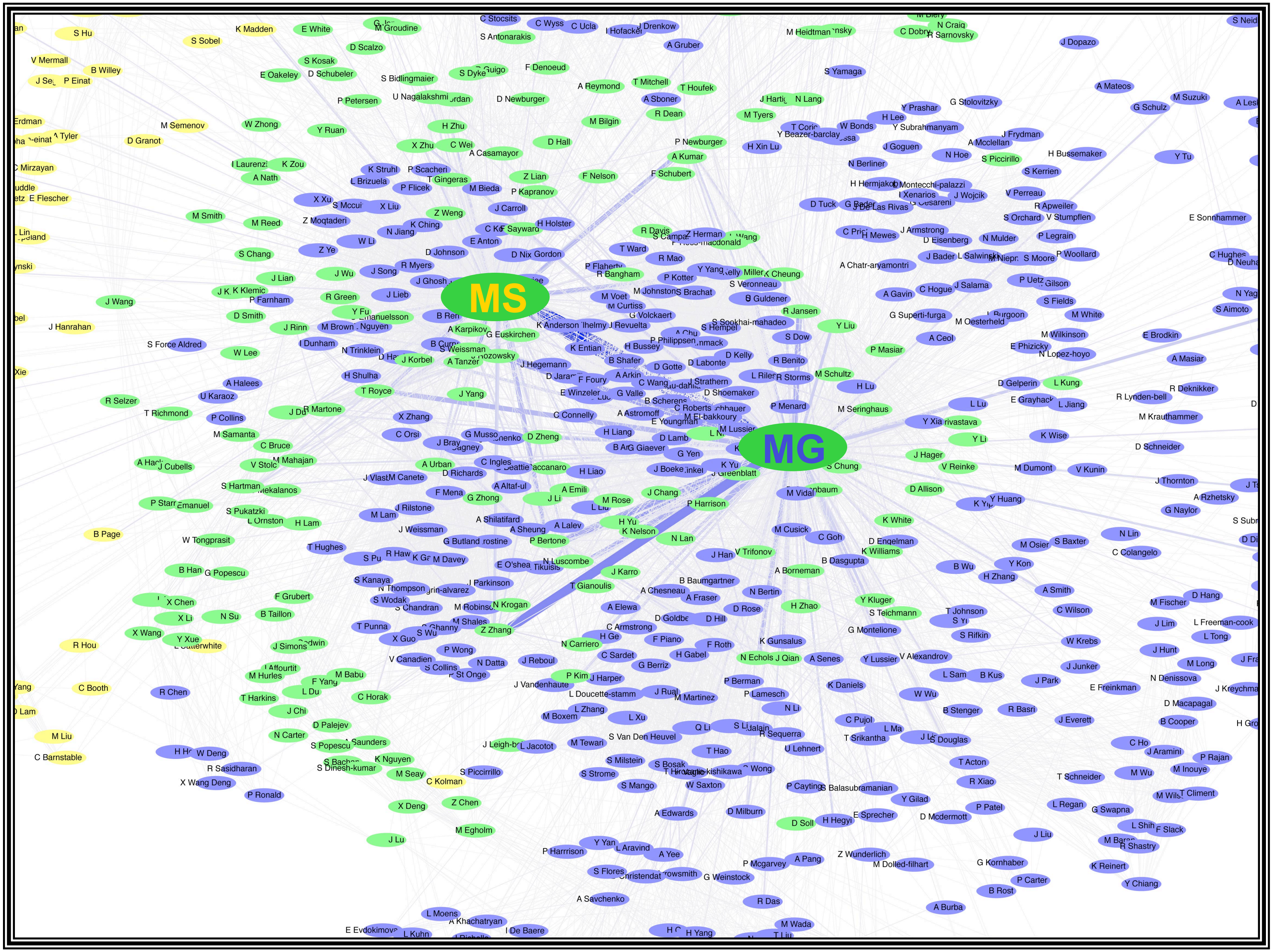
Acknowledgements

TopNet.GersteinLab.org



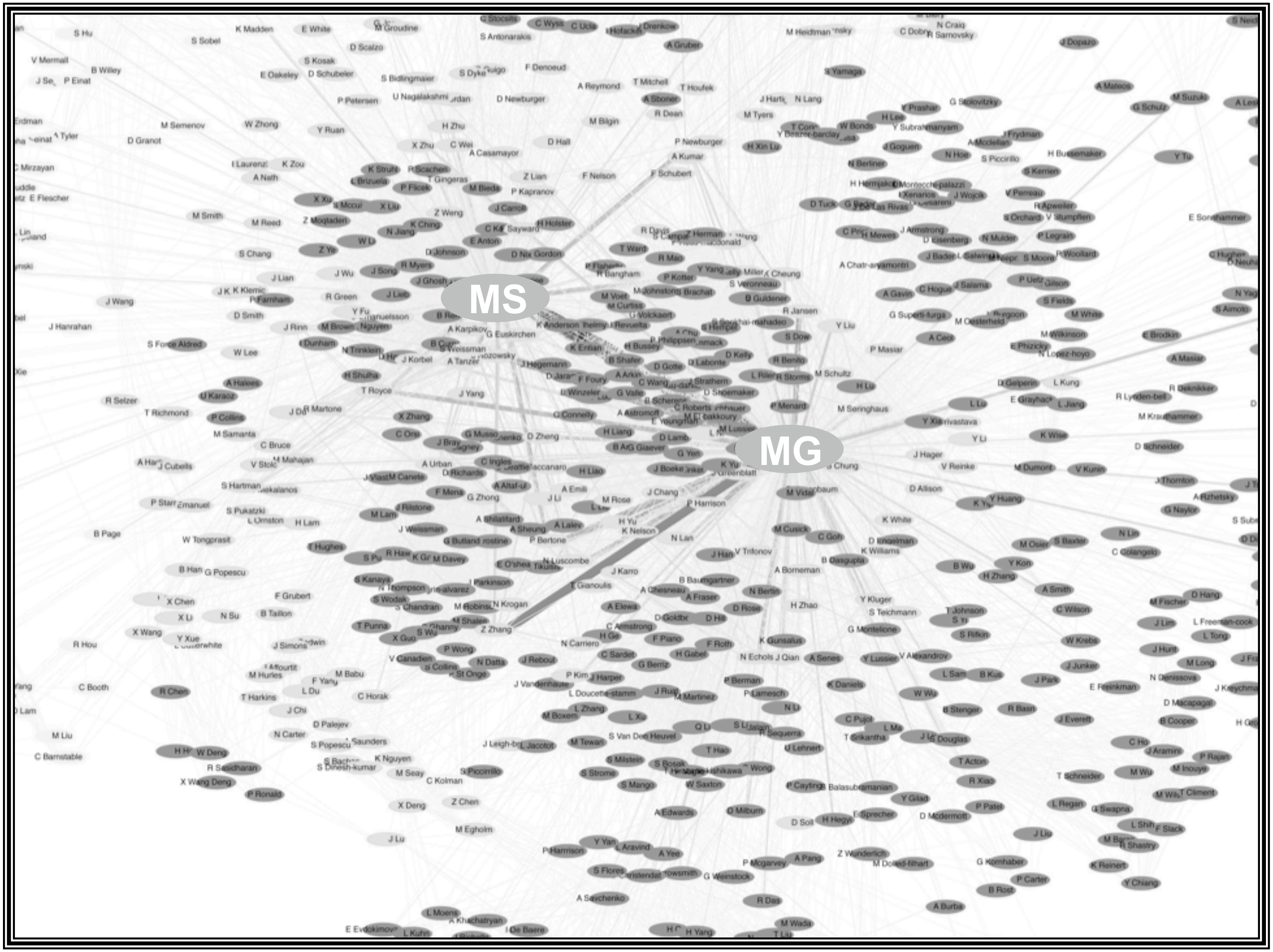
Acknowledgements

TopNet.GersteinLab.org



Acknowledgements

TopNet.GersteinLab.org

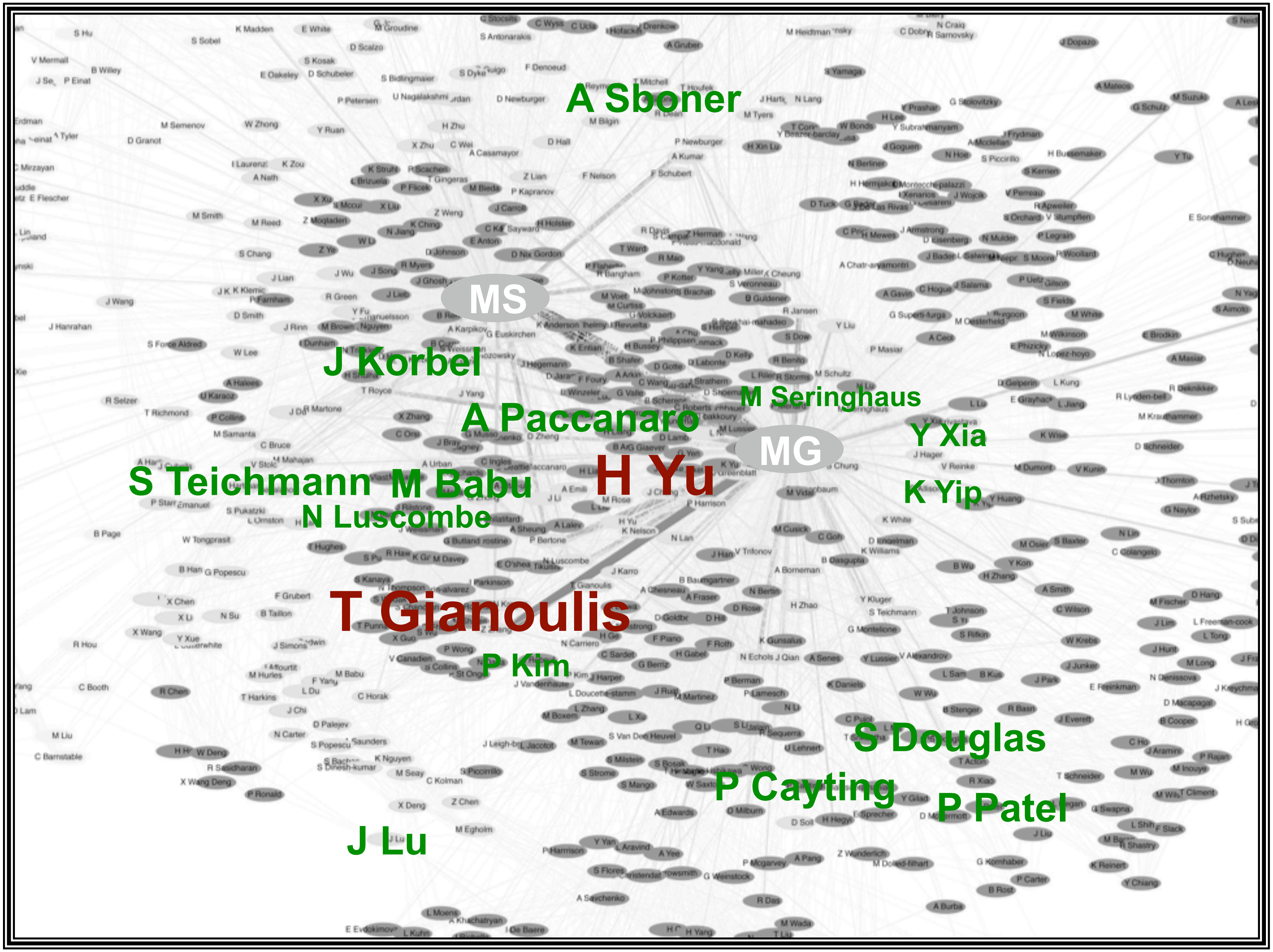


P Bork, J Raes

Acknowledgements

TopNet.GersteinLab.org

Job opportunities currently for postdocs & students



A Sboner

J Korbel

M Seringhaus

A Paccanaro

Y Xia

S Teichmann

M Babu

H Yu

K Yip

N Luscombe

T Gianoulis

P Kim

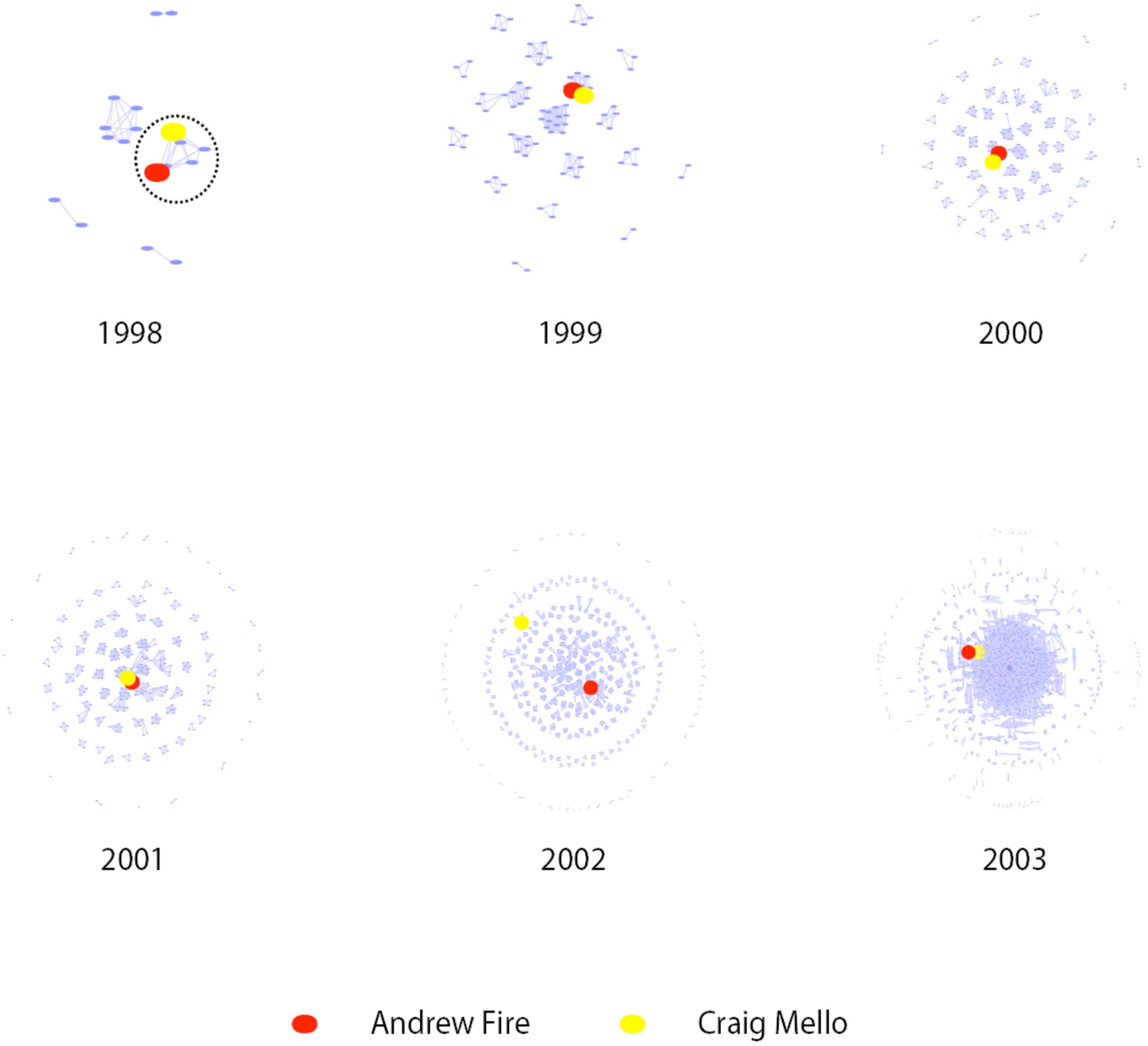
S Douglas

P Cayting

P Patel

J Lu

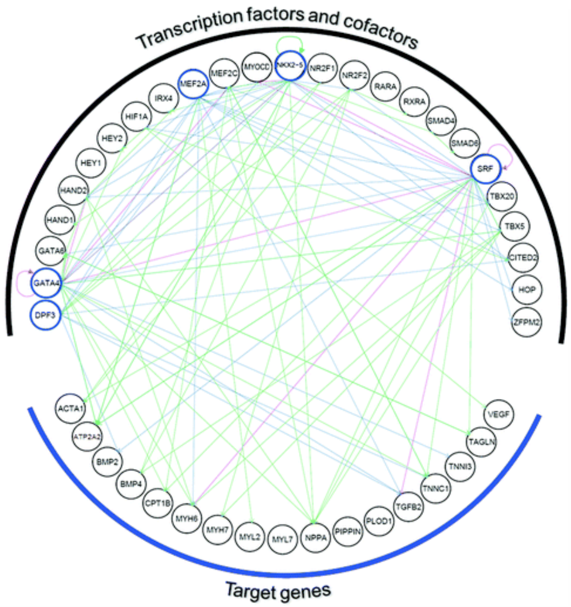
RNAi: Birth of a Field in the Literature Culminating in the 2006 Nobel



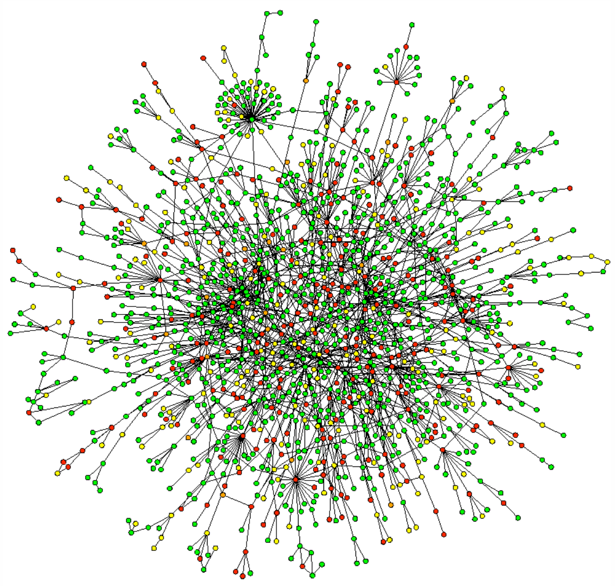
Source:
Gerstein & Douglas.
PLoS Comp. Bio. 3:e80
(2007)
PubNet.GersteinLab.org

Extra

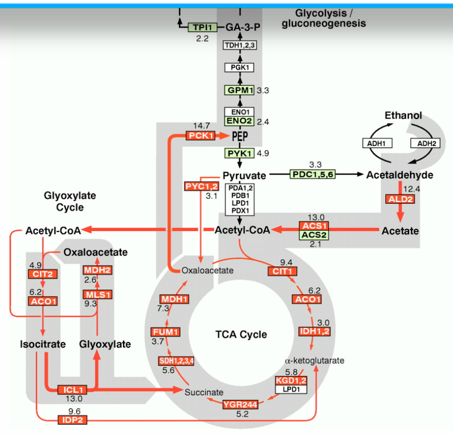
Types of Networks



Regulatory networks



Interaction networks



Metabolic networks

Nodes: proteins or genes
Edges: interactions

[Horak, et al, Genes & Development, 16:3017-3033]

[DeRisi, Iyer, and Brown, Science, 278:680-686]

[Jeong et al, Nature, 41:411]

More Information on this Talk

TITLE: Understanding Protein Function on a Genome-scale through the Analysis of Molecular Networks

SUBJECT: Networks

DESCRIPTION:
National Academy of Engineering, Meeting at Columbia U, 2009.04.14, 14:00-14:30; [I:**NAECU**] (Short networks talk, incl. the following topics:
why networks w. **amsci*, funnygene*, bottleneck*, nethierarchy*, metagenomics*, tyna* + topnet*,** & **pubnet*** . Fits into 30' w. 5' questions. PPT works on mac & PC and has many photos w. EXIF tag **kwtimewemet** .)

(Paper references in the talk were mostly from Papers.GersteinLab.org. The above topic list can be easily cross-referenced against this website. Each topic abbrev. which is starred is actually a papers "ID" on the site. For instance,
the topic **pubnet*** can be looked up at
http://papers.gersteinlab.org/papers/pubnet)

PERMISSIONS: This Presentation is copyright Mark Gerstein, Yale University, 2008. Please read permissions statement at http://www.gersteinlab.org/misc/permissions.html . Feel free to use images in the talk with PROPER acknowledgement (via citation to relevant papers or link to gersteinlab.org).
.

PHOTOS & IMAGES. For thoughts on the source and permissions of many of the photos and clipped images in this presentation see http://streams.gerstein.info . In particular, many of the images have particular EXIF tags, such as **kwpotppt** , that can be easily queried from flickr, viz: **http://www.flickr.com/photos/mbgmbg/tags/kwpotppt** .