上海交通大学
SHANGHAI JIAO TONG UNIVERSITY

基于生物医学统计的信号处理

Institute of Media, Information, and Network

生命科学技术学院
School of Life Sciences and Biotechnology

生物网络分析重构

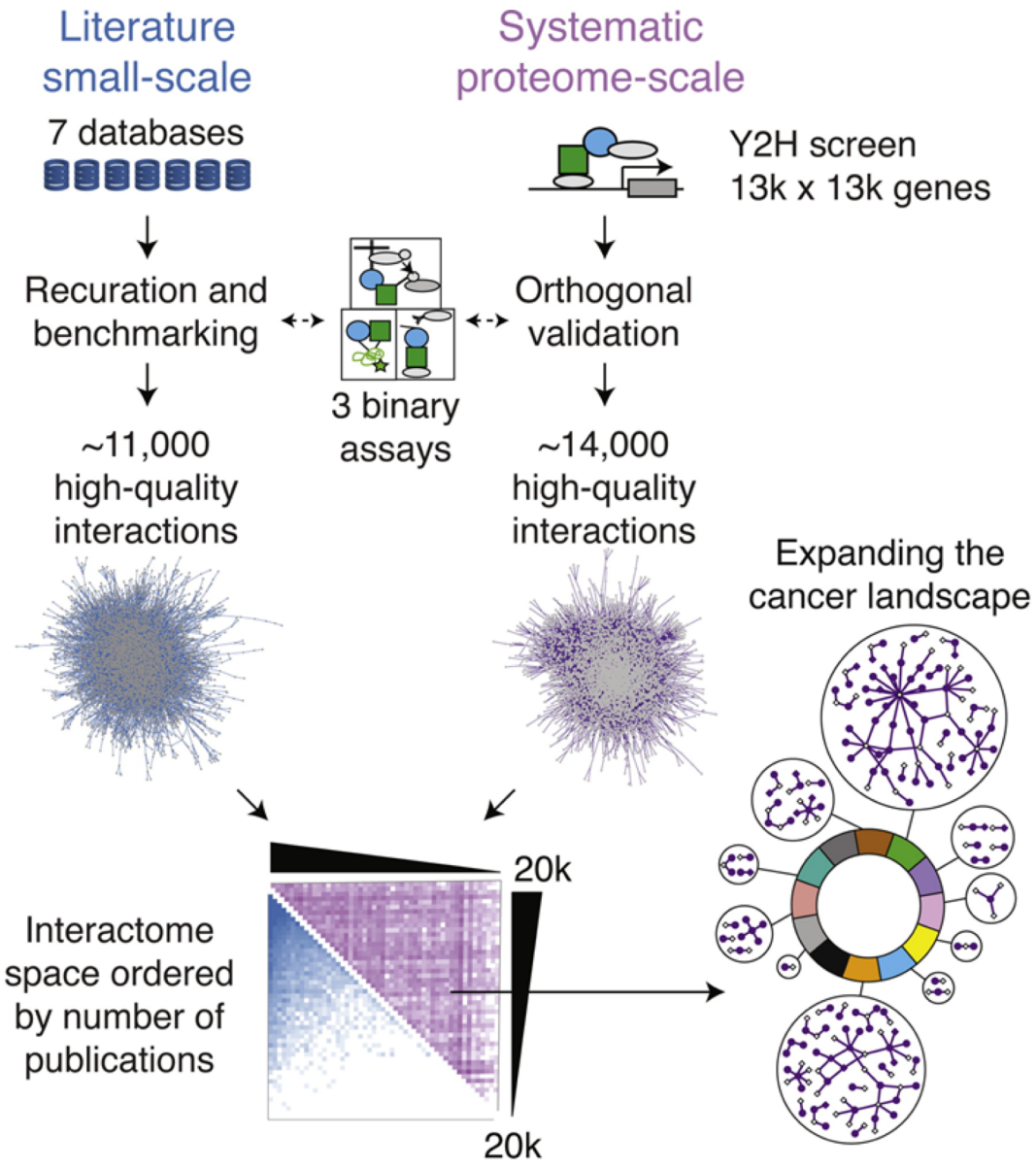
Jing Li
李　婧
jing.li@sjtu.edu.cn

June 22, 2020

Dept. Bioinformatics& Biostatistics
SJTU-Yale joint Center for Biostatistics
School of Life Sciences& Biotechnology, SJTU

A Proteome-Scale Map of the Human Interactome Network

A systematic map of 14,000 high-quality human binary protein-protein interactions.

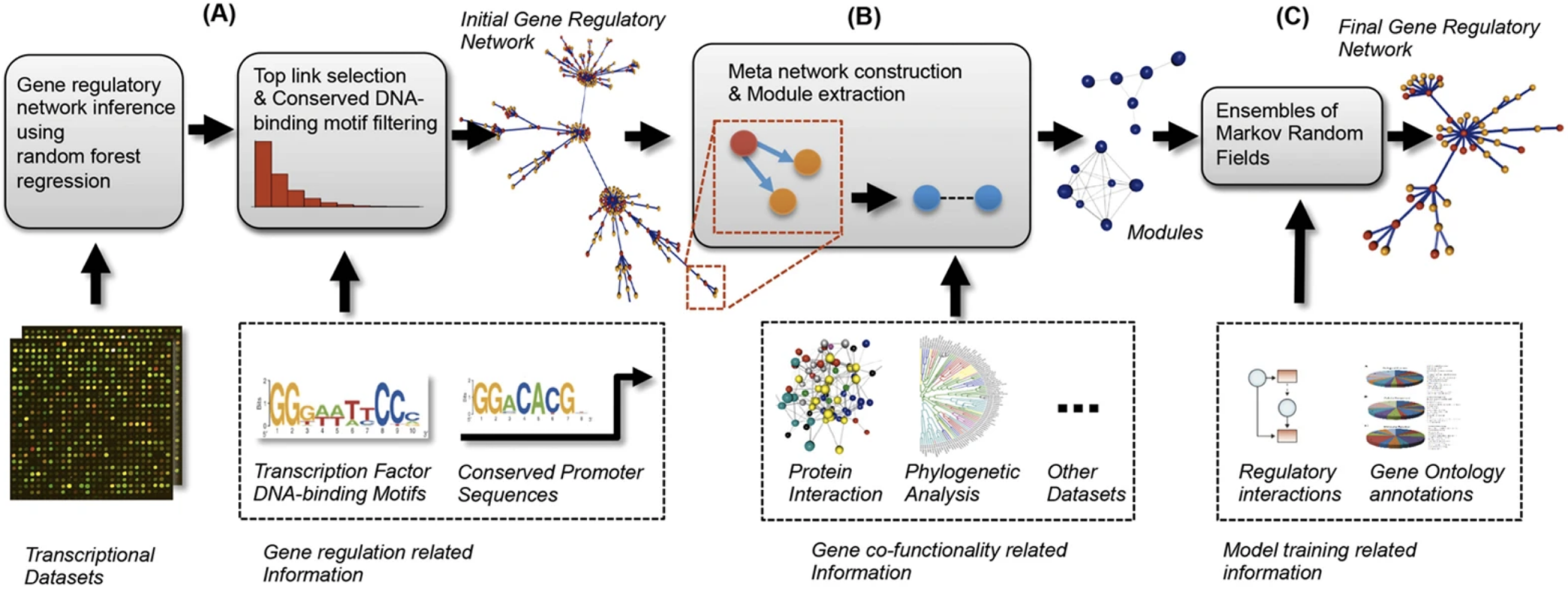


Cell 2014 159, 1212-1226

Other networks

Regulatory network

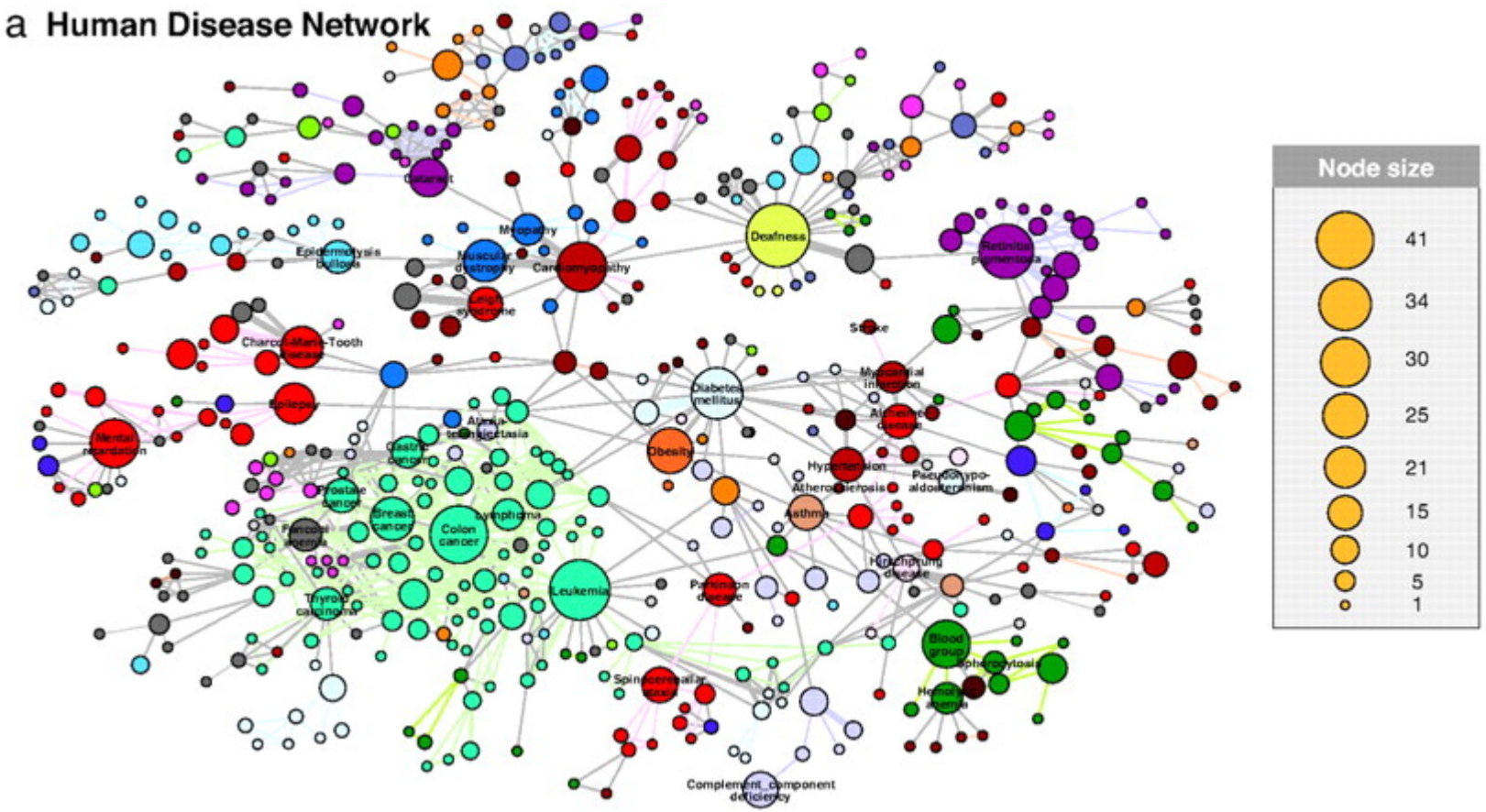
Transcription factor --- Targets



Scientific reports, 7, Article number: 41174 (2017)

Other networks

Disease Gene Networks



Goh et al. Proc Natl Acad Sci USA. (2007) 104:8685-90

Other networks

Drug-Target Networks

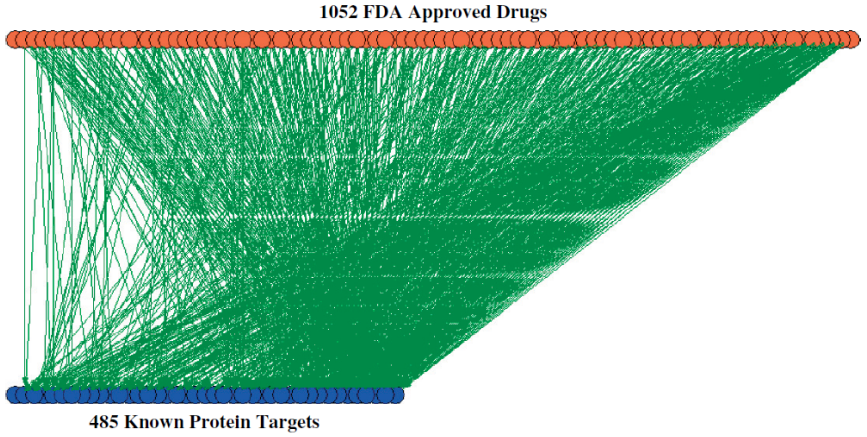
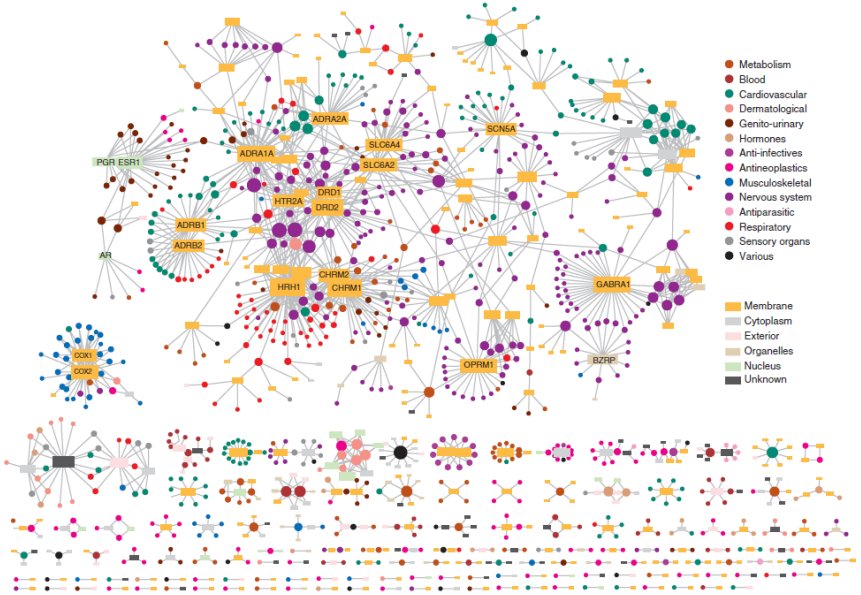


Fig 3. Visualization of the bipartite drug-target network extracted from DrugBank. Orange nodes represent drugs and blue nodes are known biomolecular targets. The network is made of 1537 nodes (1052 drugs and 485 targets) and 1815 interactions extracted from 2240 research articles.

Ma'ayan et al. Mt Sinai J Med (2007) 74:27



Yildirim et al. Nat Biotechnol. (2007) 25:1110

Metabolic networks

KEGG: Kyoto Encyclopedia of Genes a...

Go Clear

» Japanese

KEGG Home
Introduction
Overview
Release notes
Current statistics

KEGG Identifiers
Pathway maps
Brite hierarchies

KEGG XML

KEGG API

KEGG FTP

KegTools

GenomeNet

DBGET/LinkDB

Feedback

KEGG: Kyoto Encyclopedia of Genes and Genomes

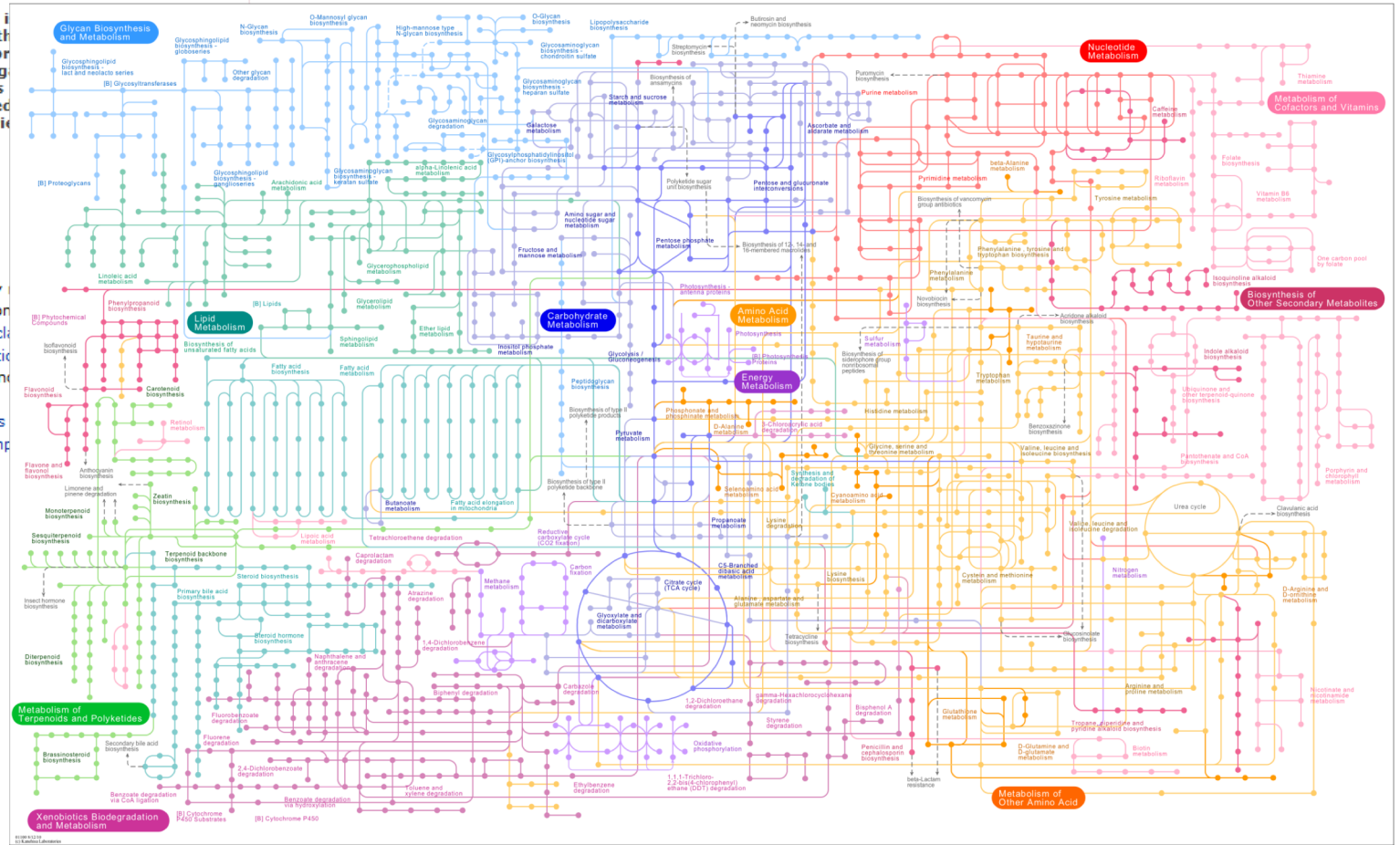
A grand challenge in the post-genomic era i
representation of the cell, the organism, th
biosphere, which will enable computational pr
complexity of cellular processes and org
genomic and molecular information. Towards
developing a bioinformatics resource named
research projects of the Kanehisa Laboratori
Center of Kyoto University and the Human
University of Tokyo.

Main entry point to the KEGG web service

KEGG2	KEGG Table of Contents

Data-oriented entry points

KEGG PATHWAY	Pathway maps and pathway
KEGG BRITE	Functional hierarchies and on
KEGG DISEASE	Human diseases Disease cl
KEGG DRUG	Drugs ATC drug classificatio
KEGG ORTHOLOGY	KO system and ortholog ann
KEGG GENES	Genes and proteins
KEGG GENOME	Genomes KEGG organisms
KEGG COMPOUND	Chemical compounds Comp
KEGG GLYCAN	Glycans
KEGG REACTION	Reactions



Centrality

- Relative **importance of a node** in the graph
- Which nodes are in the "center" of a graph?
  - What do you mean by "center"?
  - Definition of "center" varies by context/purpose
- "There is certainly no unanimity on exactly what centrality is or on its conceptual foundations, and there is little agreement on the proper procedure for its measurement."
  ----- by Freeman, 1979

Centrality

- Real valued function on the nodes of a graph
- Structural index
- Applications:
  - ✓ How influential protein is in a PPI network?
  - ✓ How important a TF is?

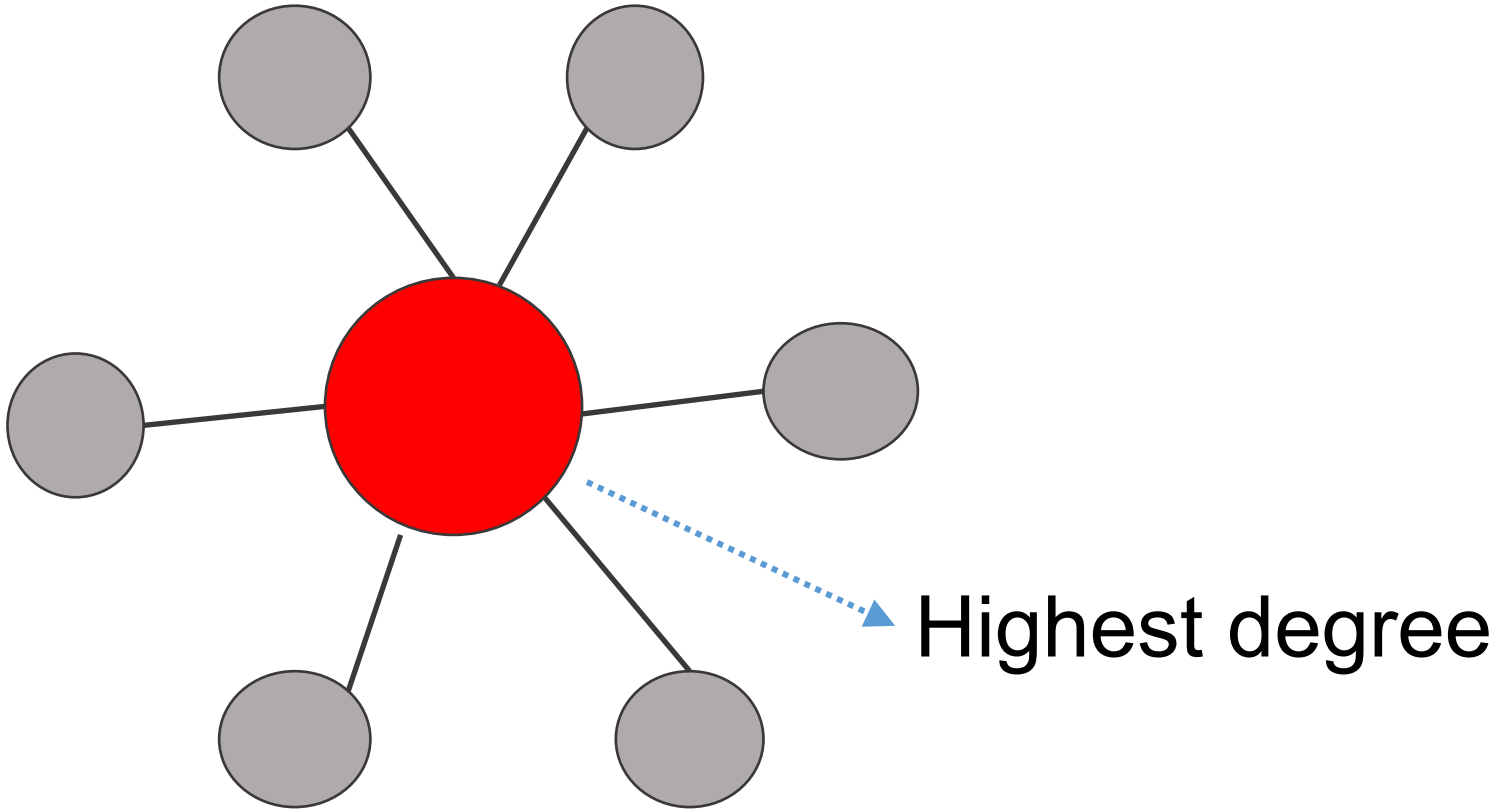
Centrality Measures

- Degree centrality
- Betweenness centrality
- Closeness centrality

Degree centrality

- Local measure of the importance of a node within a graph
- Sum of the weights of incident edges (in weighted graphs).
- Degree centrality assigns an importance score based simply on the number of links held by each node.
- Node with the highest degree is important
  - Index of exposure to what is flowing through the network
  - Hub genes
  - Gossip network: central actor more likely to hear a gossip

Degree centrality



Betweenness centrality

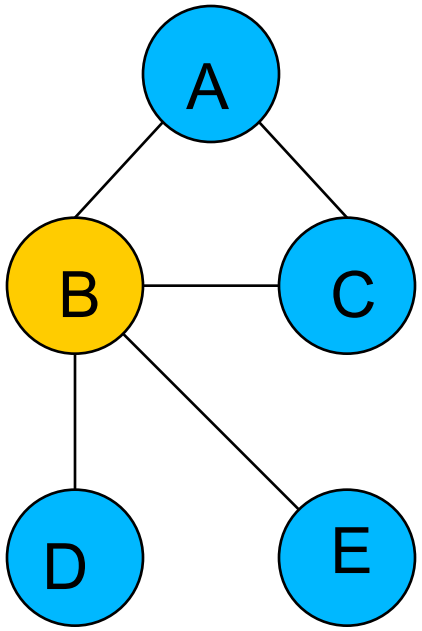
- Control on the optimal flow within a graph

- The number of shortest paths in the graph that pass through the node divided by the total number of shortest paths.

$$BC(k) = \sum_i \sum_j \frac{\rho(i,k,j)}{\rho(i,j)}, \quad i \neq j \neq k$$

where $\rho(i, j)$ is the total number of shortest paths from node i to node j and $\rho(i,k, j)$ is the number of those shortest paths that pass through k.

Betweenness centrality



- Shortest paths are:
  - AB, AC, <u>ABD</u>, <u>ABE</u>, BC, BD, BE, <u>CBD</u>, <u>CBE</u>, <u>DBE</u>

Betweenness centrality

- Nodes with a high betweenness centrality are interesting because they
  - control information flow in a network
  - may be required to carry more information
- And therefore, such nodes
  - may be the subject of targeted attack

Closeness centrality

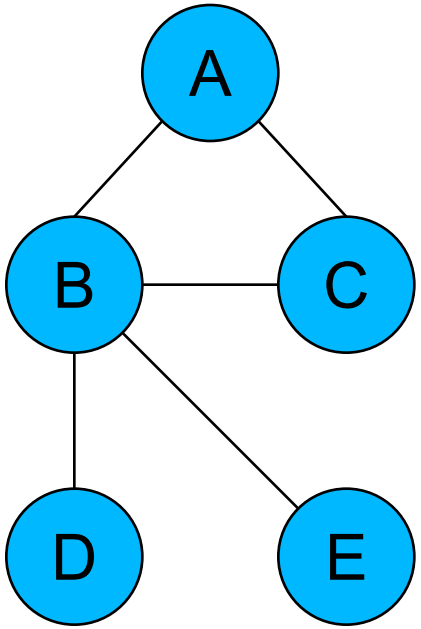
- How fast information can spread from one node to every other node
- A node is considered important if it is relatively close to all other nodes.
- The normalised inverse of the sum of topological distances in the graph.

Number of nodes

$$CC(i) = \frac{N-1}{\sum_j d(i,j)}$$

where d(i,j) is the distance (the number of edges in a shortest path) between vertices i and j.

Closeness centrality

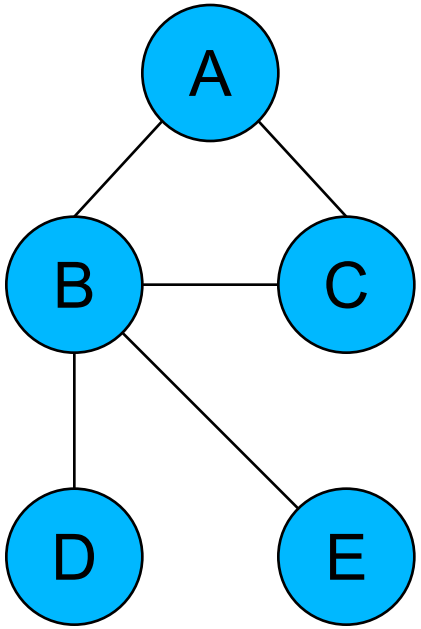


$$\begin{array}{c} \\ A \\ B \\ C \\ D \\ E \end{array} \begin{array}{c} \begin{array}{ccccc} A & B & C & D & E \end{array} \\ \begin{pmatrix} 0 & 1 & 1 & 2 & 2 \\ 1 & 0 & 1 & 1 & 1 \\ 1 & 1 & 0 & 2 & 2 \\ 2 & 1 & 2 & 0 & 2 \\ 2 & 1 & 2 & 2 & 0 \end{pmatrix} \end{array}$$

$$\sum_{j=1}^{n} d(i,j)$$

	Closeness
6	0.67
4	1.00
6	0.67
7	0.57
7	0.57

Closeness centrality

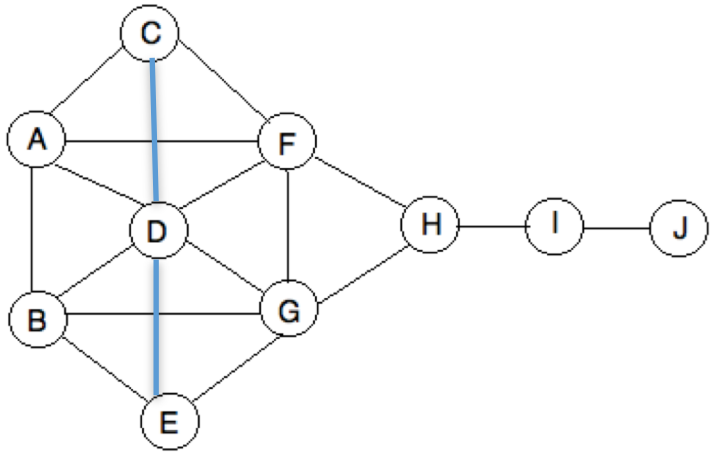


$$\begin{array}{c} \\ A \\ B \\ C \\ D \\ E \end{array} \begin{array}{c} \begin{array}{ccccc} A & B & C & D & E \end{array} \\ \begin{pmatrix} 0 & 1 & 1 & 2 & 2 \\ 1 & 0 & 1 & 1 & 1 \\ 1 & 1 & 0 & 2 & 2 \\ 2 & 1 & 2 & 0 & 2 \\ 2 & 1 & 2 & 2 & 0 \end{pmatrix} \end{array}$$

$\sum_{j=1}^{n} d(i,j)$	Closeness
6	0.67
4	1.00
6	0.67
7	0.57
7	0.57

- Node B is the most central one in spreading information from it to the other nodes in the network.

- <u>Example:</u>



Which node is most important?

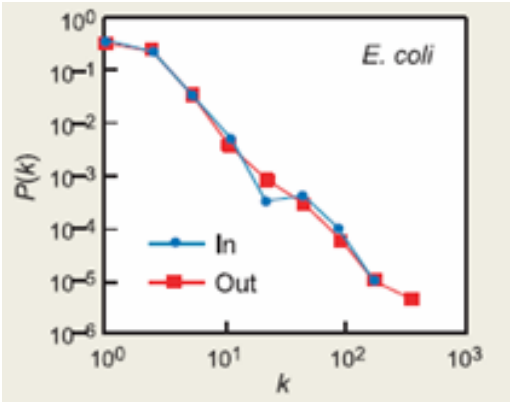
	Degree	Closeness	Betweeness
From highest	D	F, G	H
	F, G	D, H	F, G
to	A, B	A, B	I
	C, E, H	C, E	D
lowest	I	I	A, B
	J	J	C, D, J

“Real” Networks are “Scale Free”

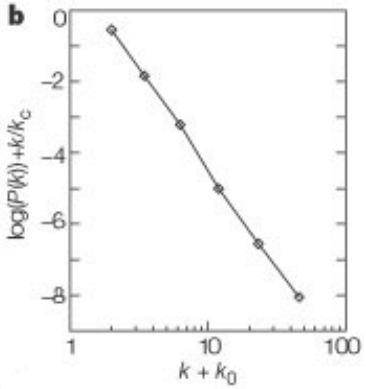

Barabasi and Albert. *Science* **286**, 509 (1999)

Barabasi et al. found that many real networks including the Internet and the WWW are **scale-free**. This means that the connectivity distribution of nodes fits a power-law.

They analyzed databases of metabolic networks in lower organisms and the protein-protein interactions map of the yeast proteome inferred from high-throughput yeast-2-hybrid screens. All shown to have scale-free connectivity distribution.

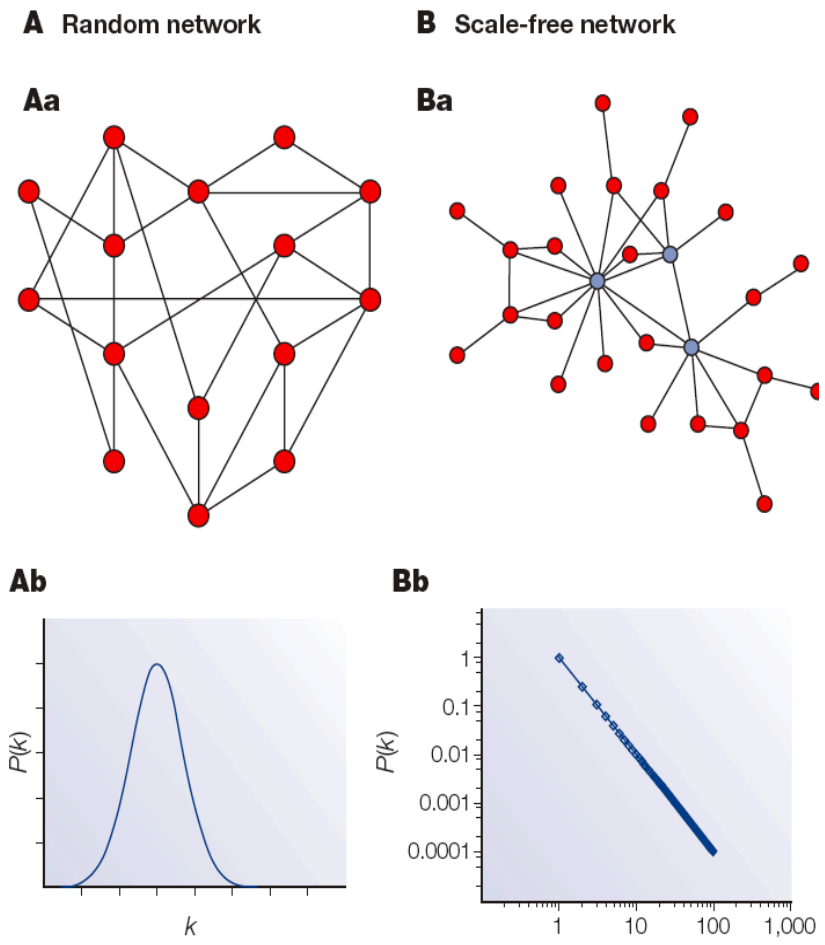


Jeong et al. *Nature* **407**, 651 (2000)



Jeong et al. *Nature* **411**, 41 (2001)

Degree Distribution



P(k) is probability of each degree k, i.e fraction of nodes having that degree.

For random networks, P(k) is normally distributed.

For real networks the distribution is often a power-law:

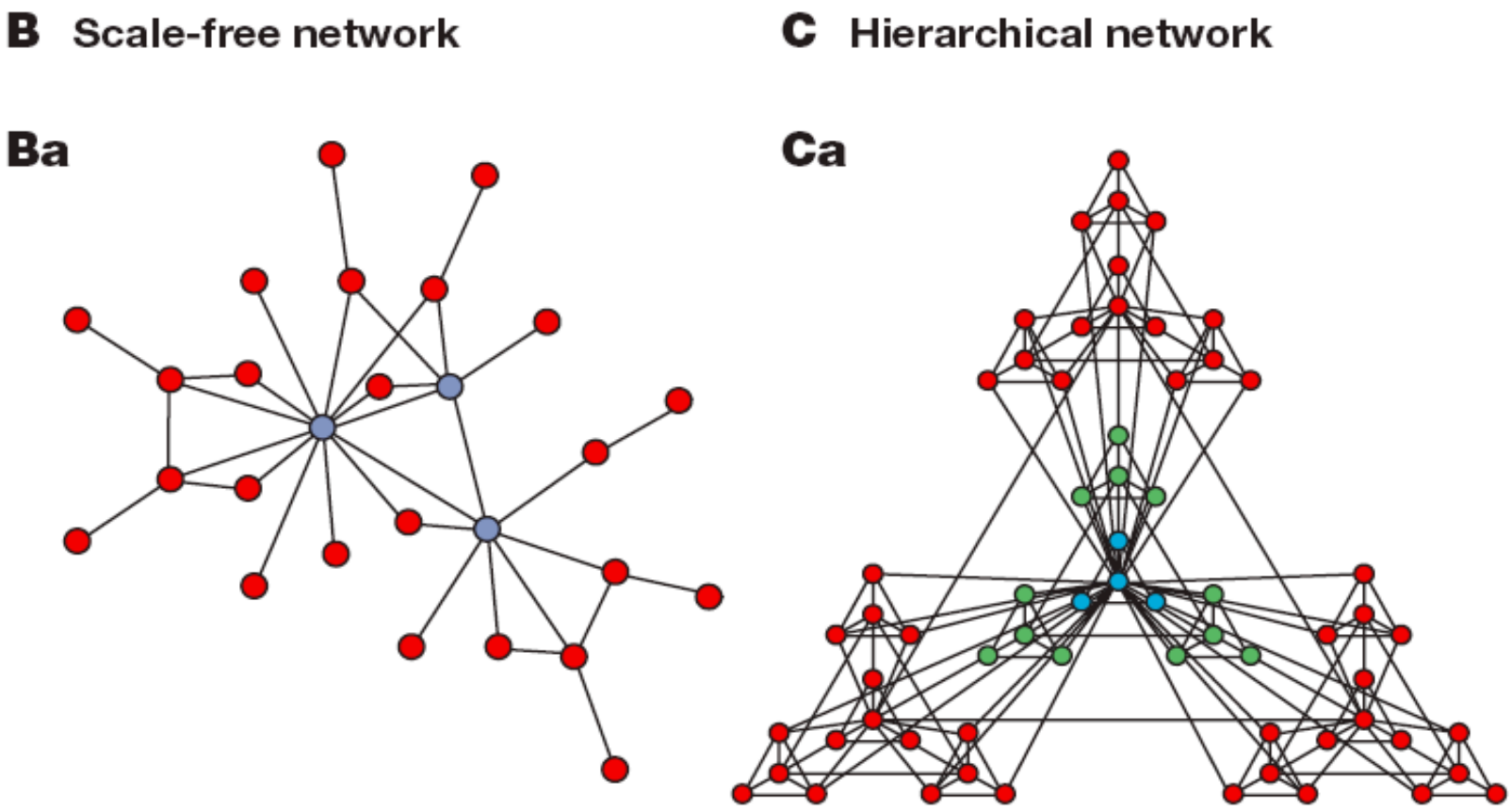
$P(k) \sim k^{-\gamma}$

Such networks are said to be scale-free

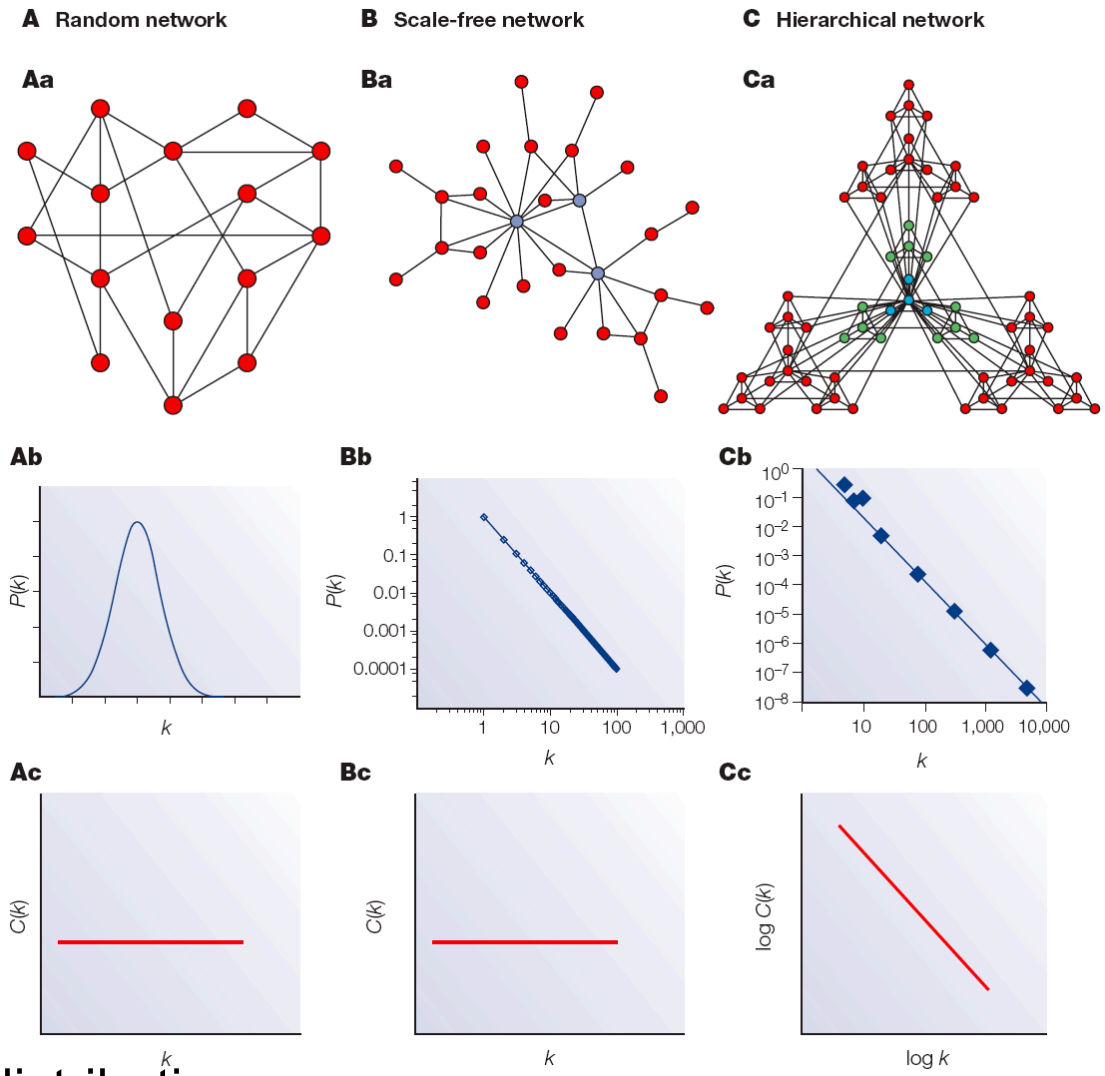
For most networks, $2 < \gamma < 3$

Nature Reviews Genetics **5**, 101-113 （2004）

Hierarchical Networks

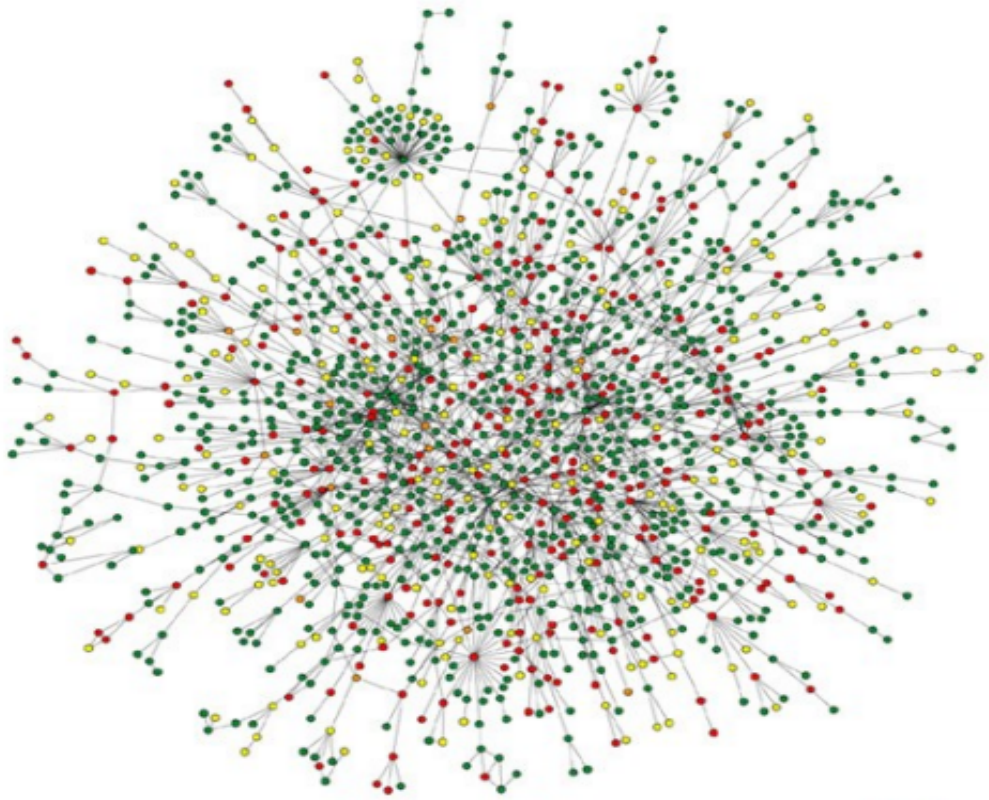


Detecting Hierarchical Organization



– P(k) degree distribution
– C(k) average clustering coefficient function

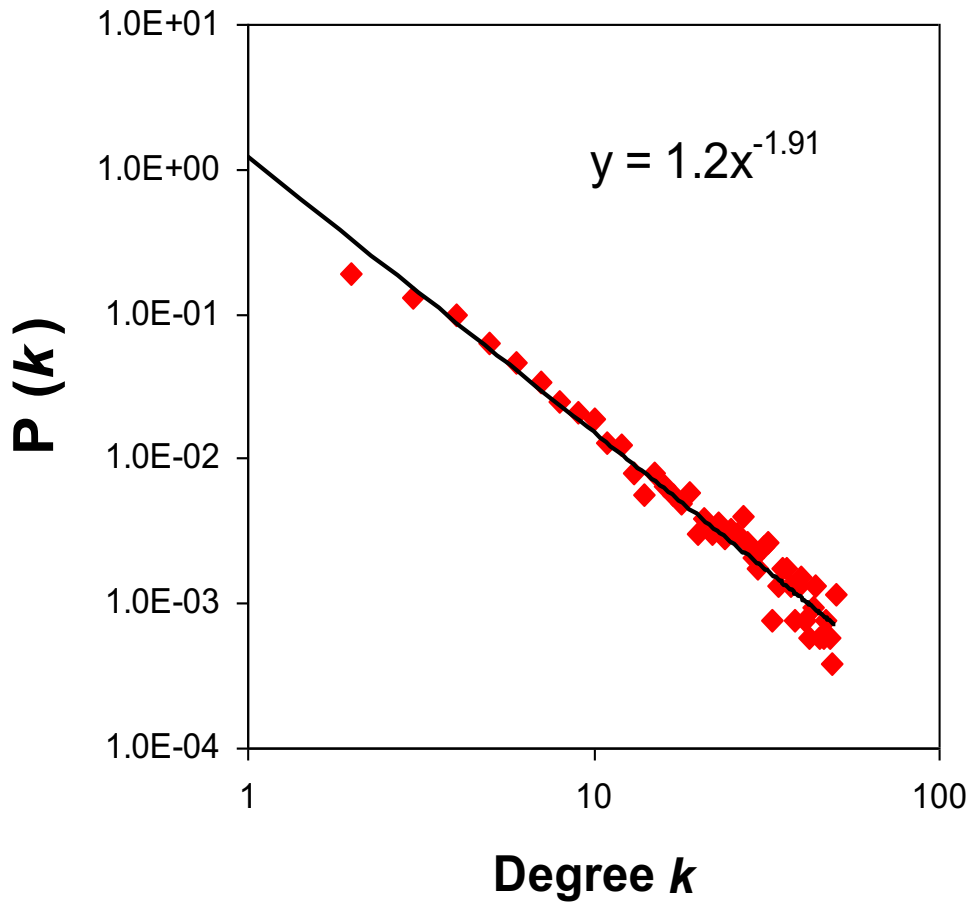
Cellular networks are scale-free

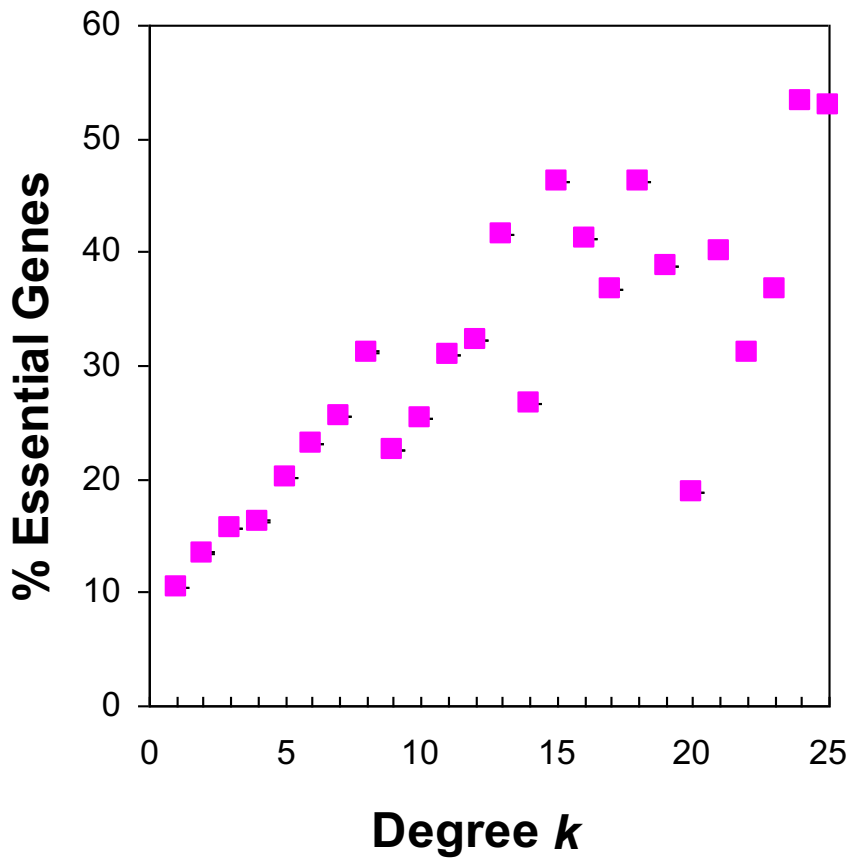


Scale-Free Networks are Robust

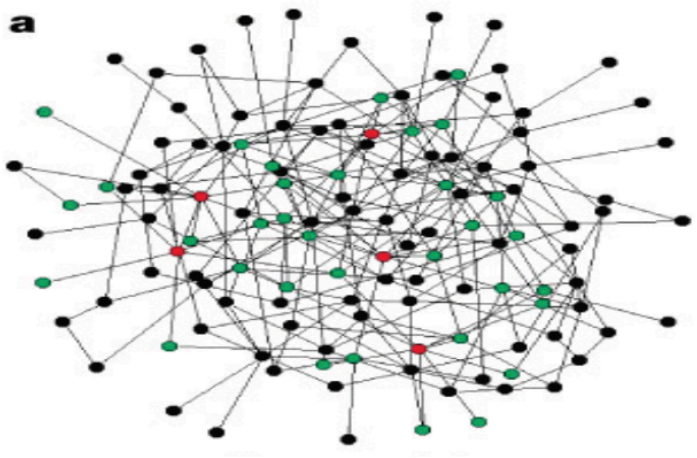
- Complex systems (cell, internet, social networks), are resilient to component failure
- Network topology plays an important role in this robustness
  - Even if ~80% of nodes fail, the remaining ~20% still maintain network connectivity
- *Attack vulnerability* if hubs are selectively targeted
- In yeast, only ~20% of proteins are lethal when deleted, and are 5 times more likely to have degree $k>15$ than $k<5$.

Knock-out Lethality and Connectivity



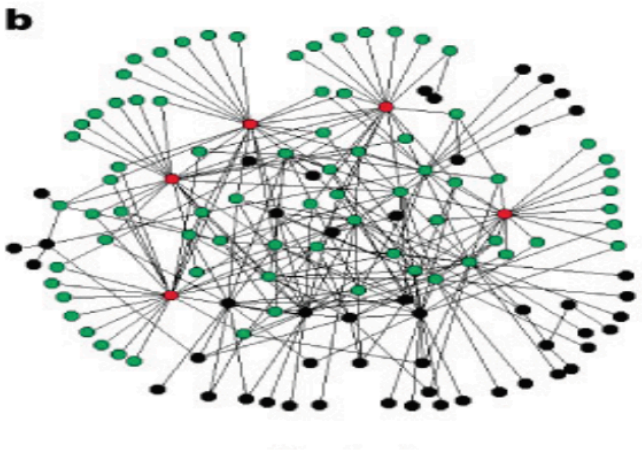


Random network *vs* scale-free network



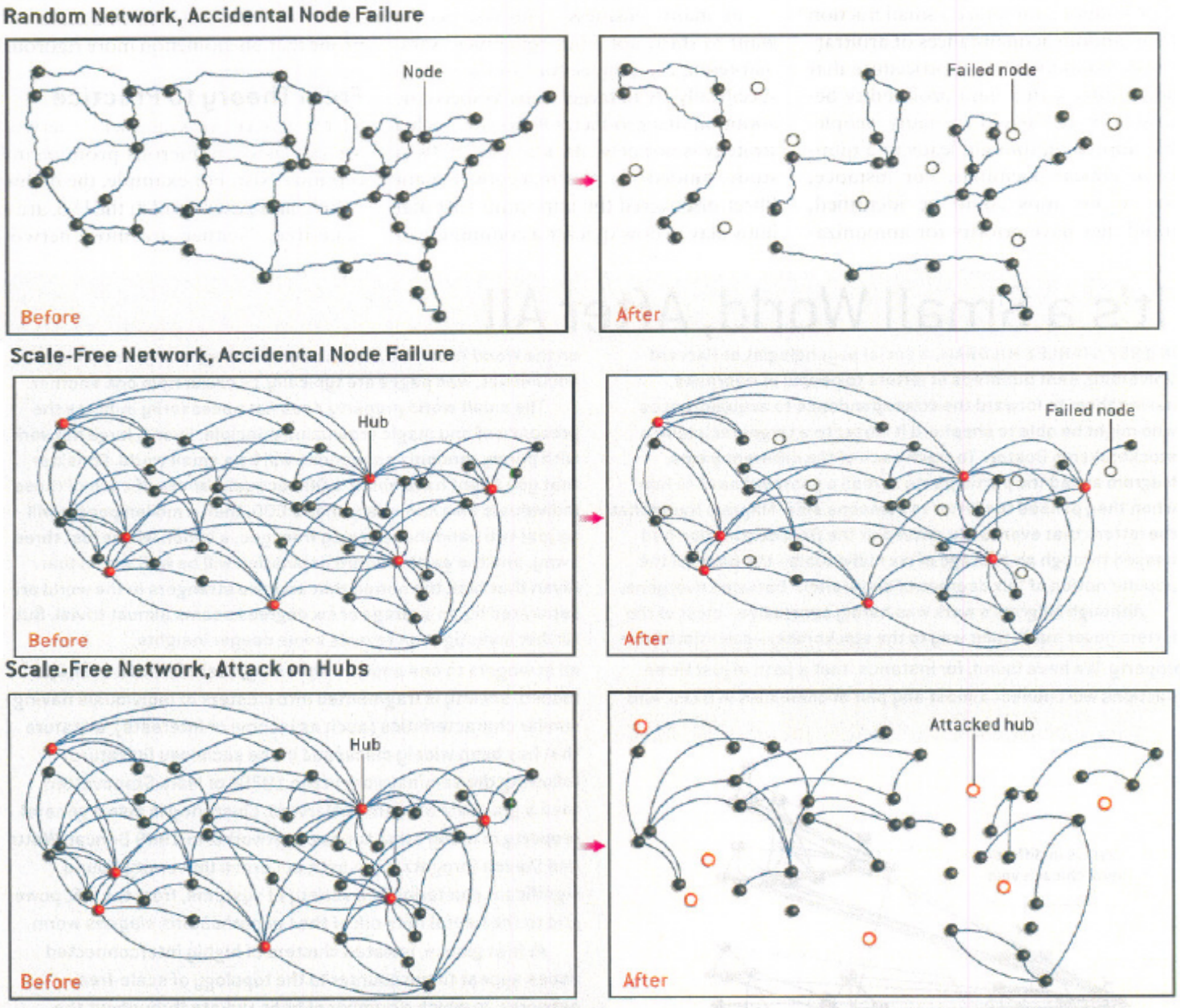
- Random network
- 130 nodes, 215 edges
- Homogeneous: most nodes have approximately the same number of links
- Five red nodes with the highest number of links reach 27% of the nodes

Albert et al., Nature, 406:378, 2000



- Scale-free network
- 130 nodes, 215 edges
- Heterogeneous: the majority of the nodes have one or two links but a few nodes have a large number of links
- Five red nodes with the highest degrees reach 60% of the nodes (**hubs**)

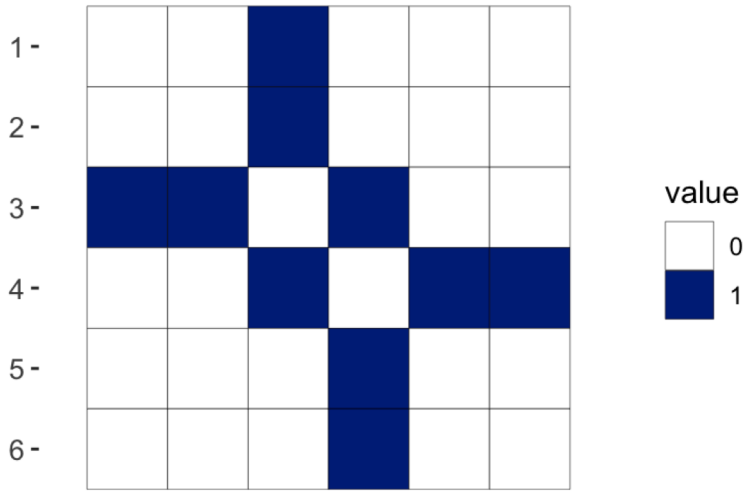
Degree (hubs)-attack



How to encode a graph

A **graph** is formed by a set of nodes or vertices (often called VV) and a set of edges between these vertices (EE). Edges EE are provided as unordered pairs of vertices in undirected graphs and ordered pairs for directed or oriented graphs.
An **adjacency matrix** AA is the matrix representation of EE. AA is a square matrix with as many rows as nodes in the graph. AA contains a non zero entry in the ith row and j^{th} column if there is an edge between the ith and j^{th} vertices.

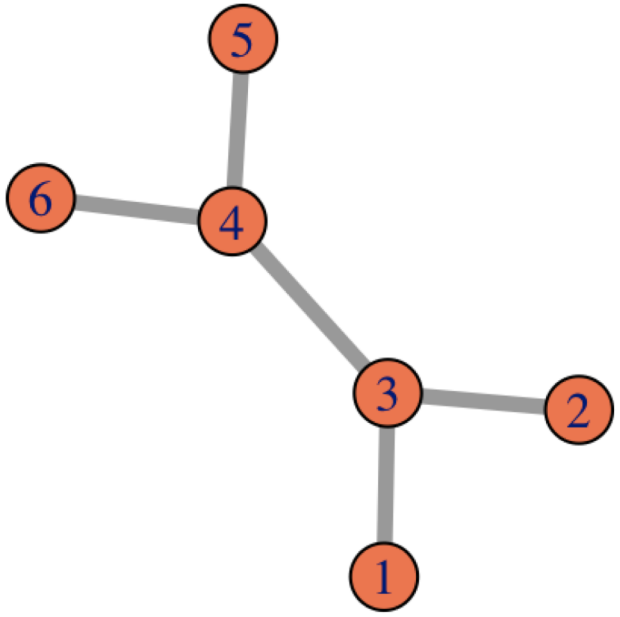
For graphs with undirected edges what is special about the adjacency matrix AA?



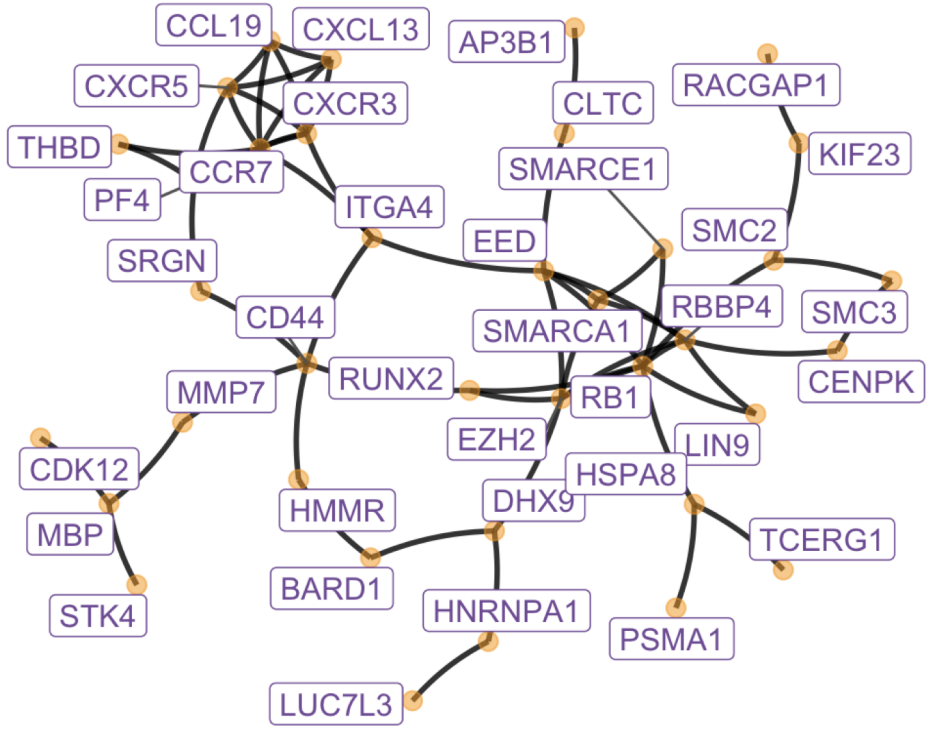
How to encode a graph

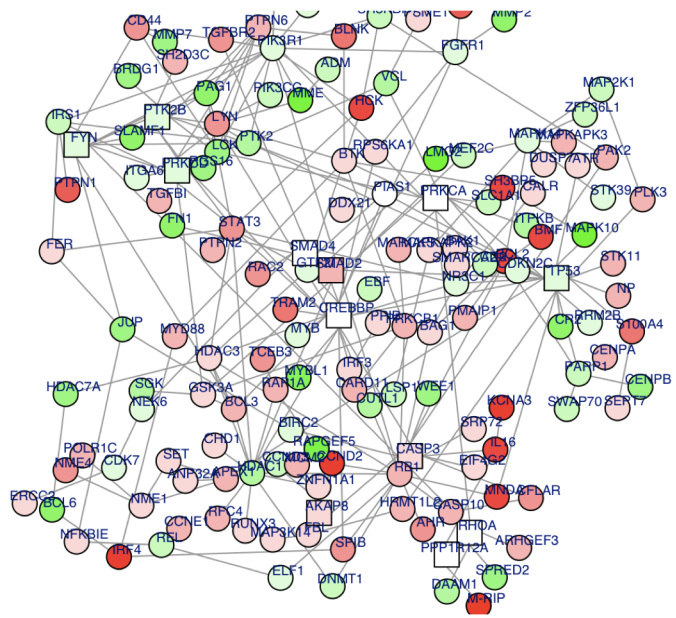
The adjacency matrix is symmetric

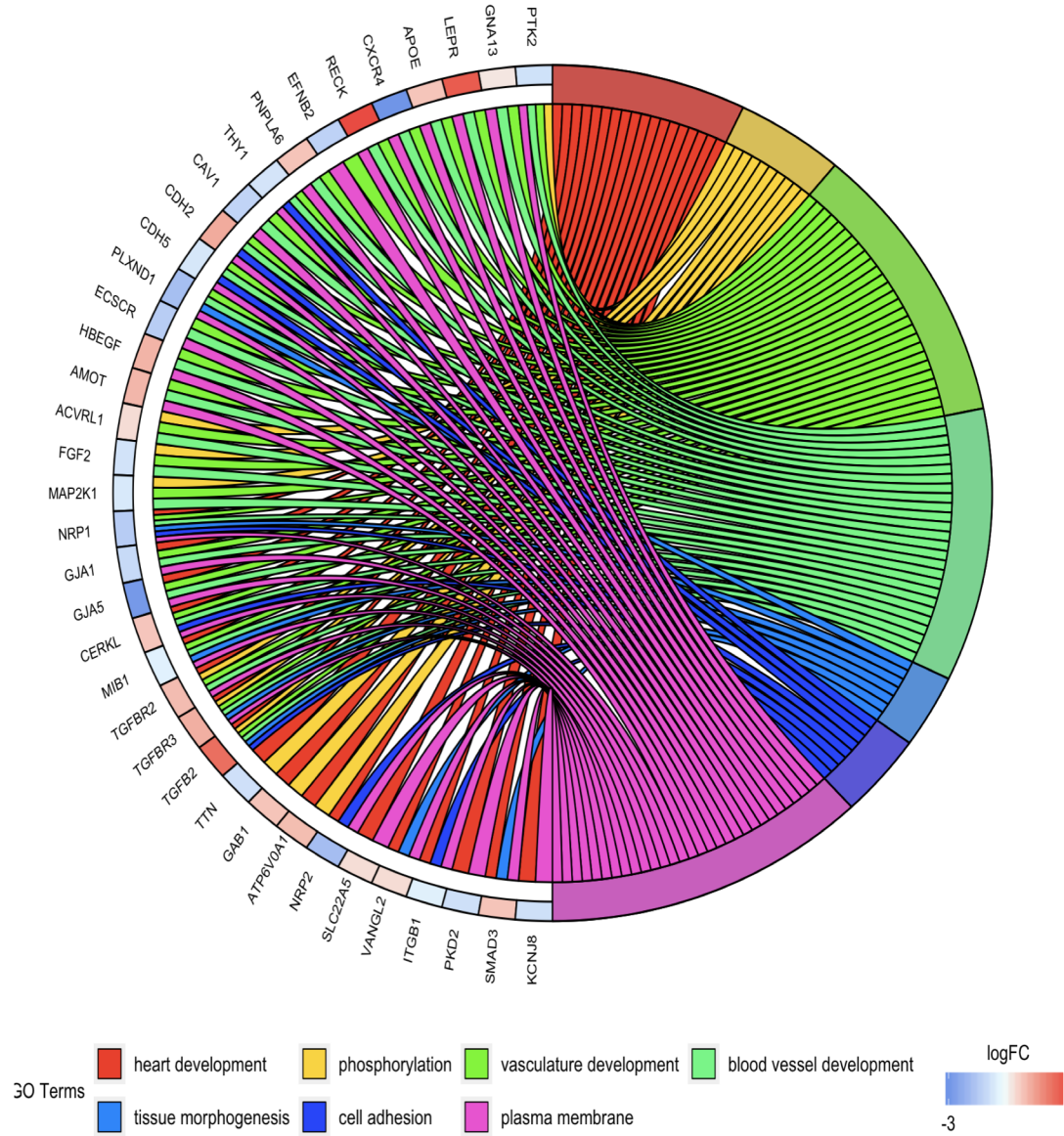
```
library("igraph")
edges1 = matrix(c(1,3,2,3,3,4,4,5,4,6),byrow = TRUE, ncol = 2)
g1 = graph_from_edgelist(edges1, directed = FALSE)
plot(g1, vertex.size = 25, edge.width = 5, vertex.color = "coral")
```



We may do graph statistics though the various packages, such as network, igraph.


CCL19
CXCL13
AP3B1
CXCR5
CXCR3
CLTC
RACGAP1
THBD
CCR7
KIF23
SMARCE1
PF4
ITGA4
EED
SMC2
SRGN
CD44
RBBP4
SMC3
SMARCA1
RUNX2
CENPK
MMP7
RB1
EZH2
LIN9
CDK12
HSPA8
DHX9
HMMR
MBP
TCERG1
BARD1
STK4
HNRNPA1
PSMA1
LUC7L3

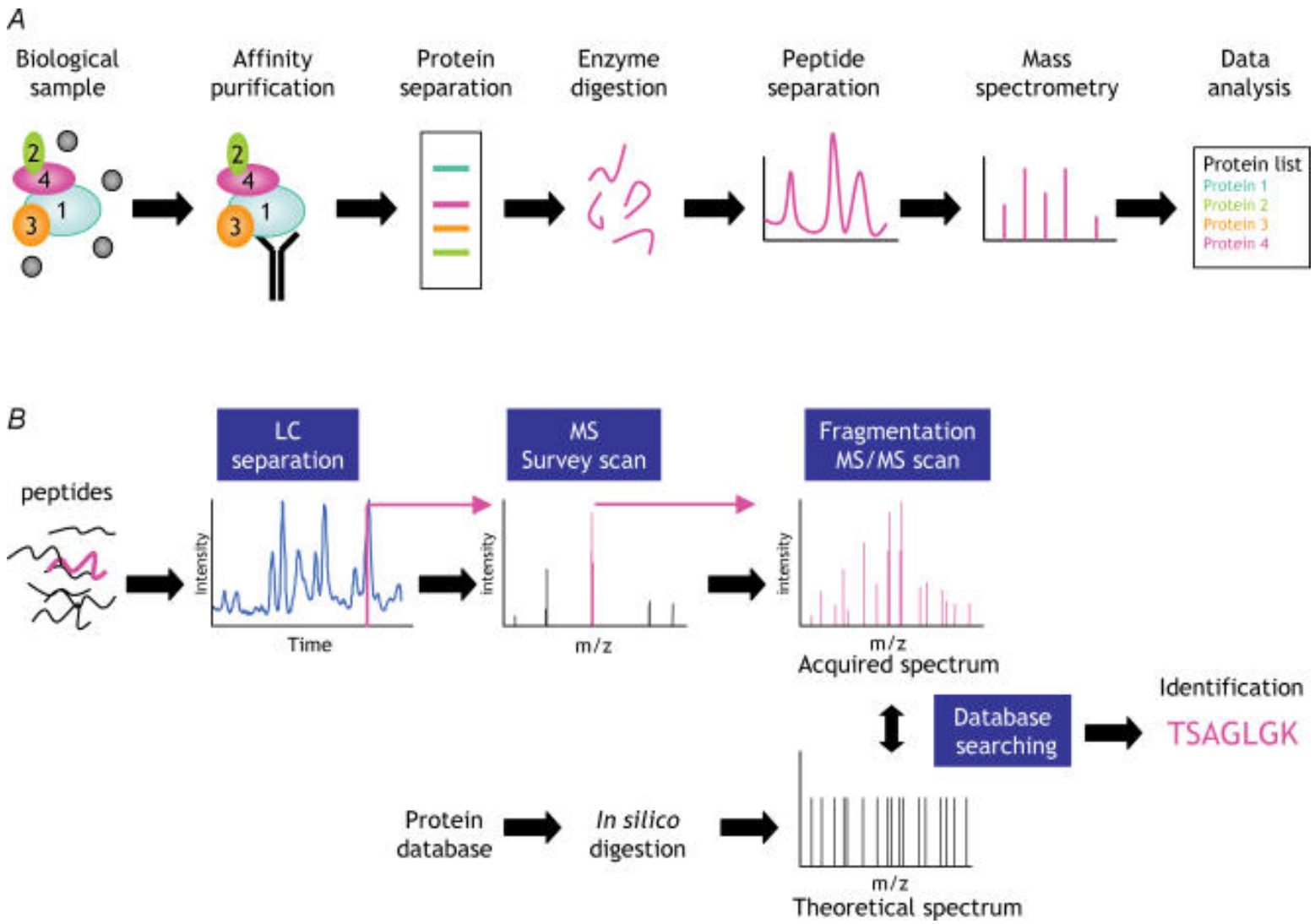



PTK2
GNA13
LEPR
APOE
CXCR4
RECK
EFNB2
PNPLA6
THY1
CAV1
CDH2
CDH5
PLXND1
ECSCR
HBEGF
AMOT
ACVRL1
FGF2
MAP2K1
NRP1
GJA1
GJA5
CERKL
MIB1
TGFBR2
TGFBR3
TGFB2
TTN
GAB1
ATP6V0A1
NRP2
SLC22A5
VANGL2
ITGB1
PKD2
SMAD3
KCNJ8
GO Terms
heart development
phosphorylation
vasculature development
blood vessel development
tissue morphogenesis
cell adhesion
plasma membrane
logFC
-3

Topics

- Network and network topology
- **Network reconstruction**
- Network application

(1) Mass spectrometry screening （Pull-down）



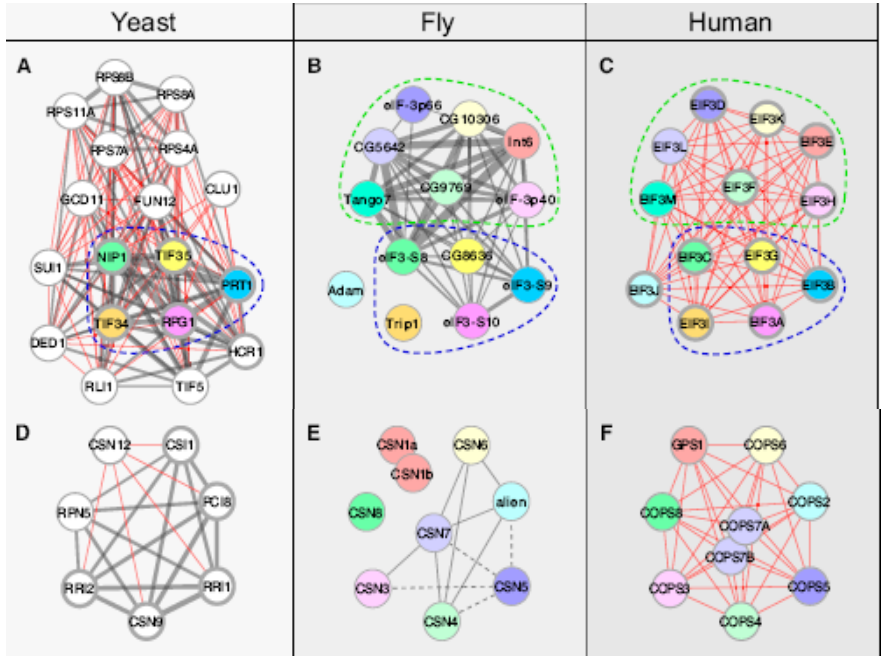
J Physiol. 2005 February 15; 563(Pt 1): 11–21.

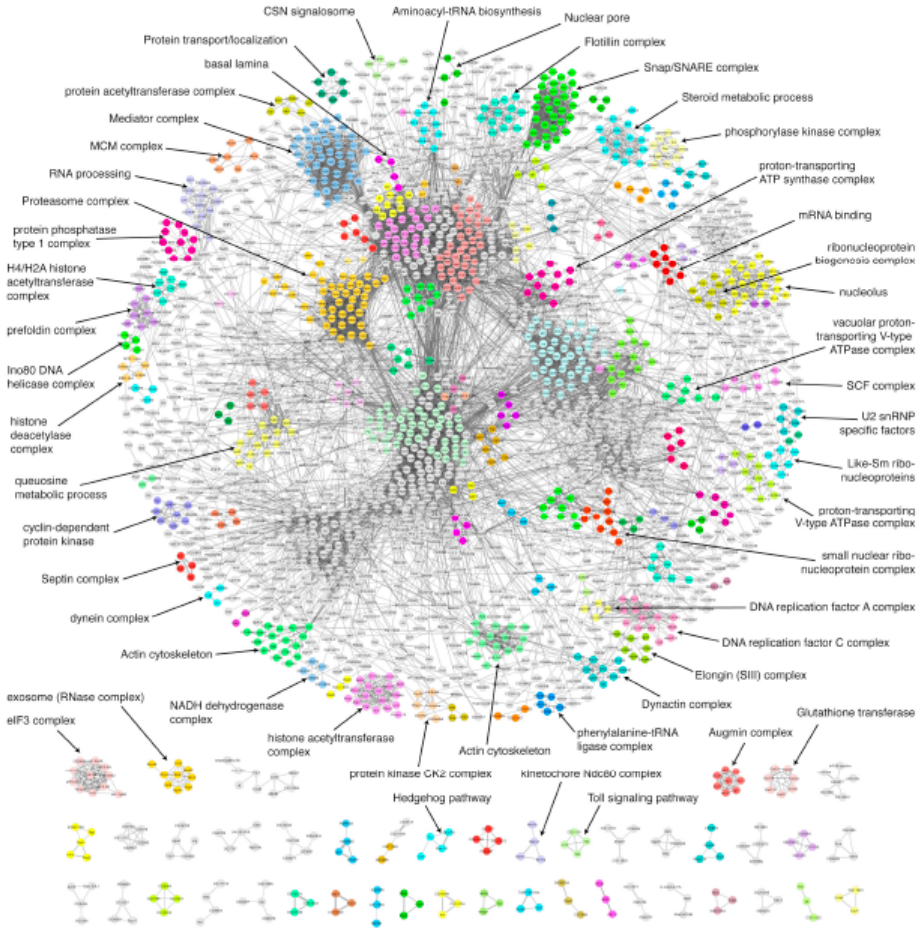
Cell

Resource

A Protein Complex Network of *Drosophila melanogaster*

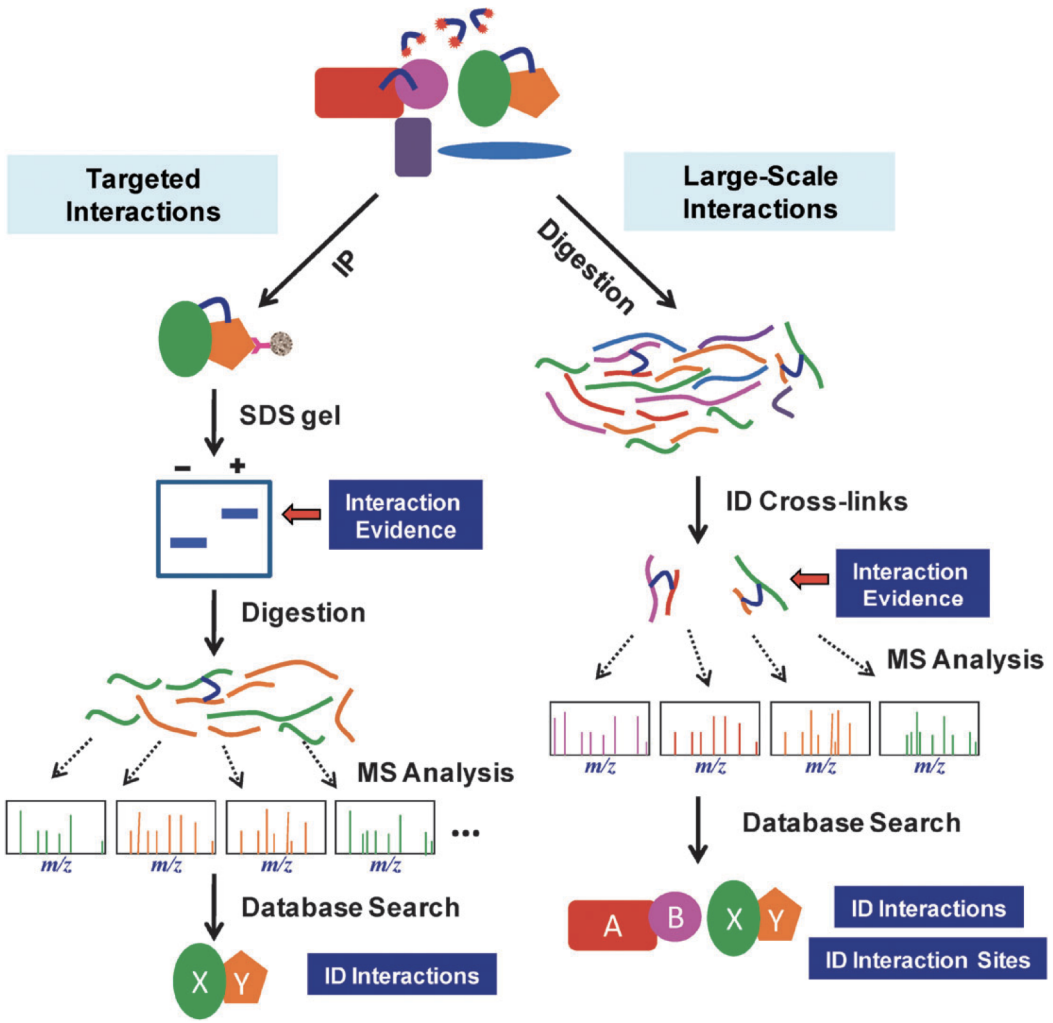
K.G. Guruharsha,[1,4] Jean-François Rual,[1,4] Bo Zhai,[1,4] Julian Mintseris,[1,4] Pujita Vaidya,[1] Namita Vaidya,[1] Chapman Beekman,[1] Christina Wong,[1] David Y. Rhee,[1] Odise Cenaj,[1] Emily McKillip,[1] Saumini Shah,[1] Mark Stapleton,[2] Kenneth H. Wan,[2] Charles Yu,[2] Bayan Parsa,[2] Joseph W. Carlson,[2] Xiao Chen,[2] Bhaveen Kapadia,[2] K. VijayRaghavan,[3] Steven P. Gygi,[1] Susan E. Celniker,[2] Robert A. Obar,[1,*] and Spyros Artavanis-Tsakonas[1,*]
[1]Department of Cell Biology, Harvard Medical School, Boston, MA 02115, USA
[2]Berkeley *Drosophila* Genome Project, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA
[3]National Centre for Biological Sciences, Tata Institute of Fundamental Research, Bangalore 560065, India
[4]These authors contributed equally to this work
*Correspondence: robert_obar@hms.harvard.edu (R.A.O.), artavanis@hms.harvard.edu (S.A.-T.)
DOI 10.1016/j.cell.2011.08.047





Cell. 2011 Oct 28;147(3):690-703.

(2) MS-based Cross-linking strategy for PPI detection



Nature Protocols 13, 2864–2889(2018)

PPI network identification

Direct: from an experiment

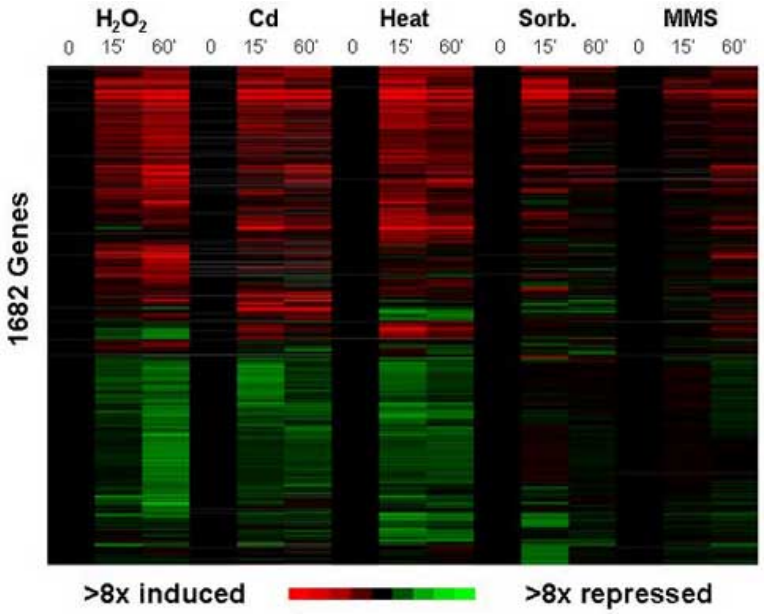
Indirect: **network reconstruction**

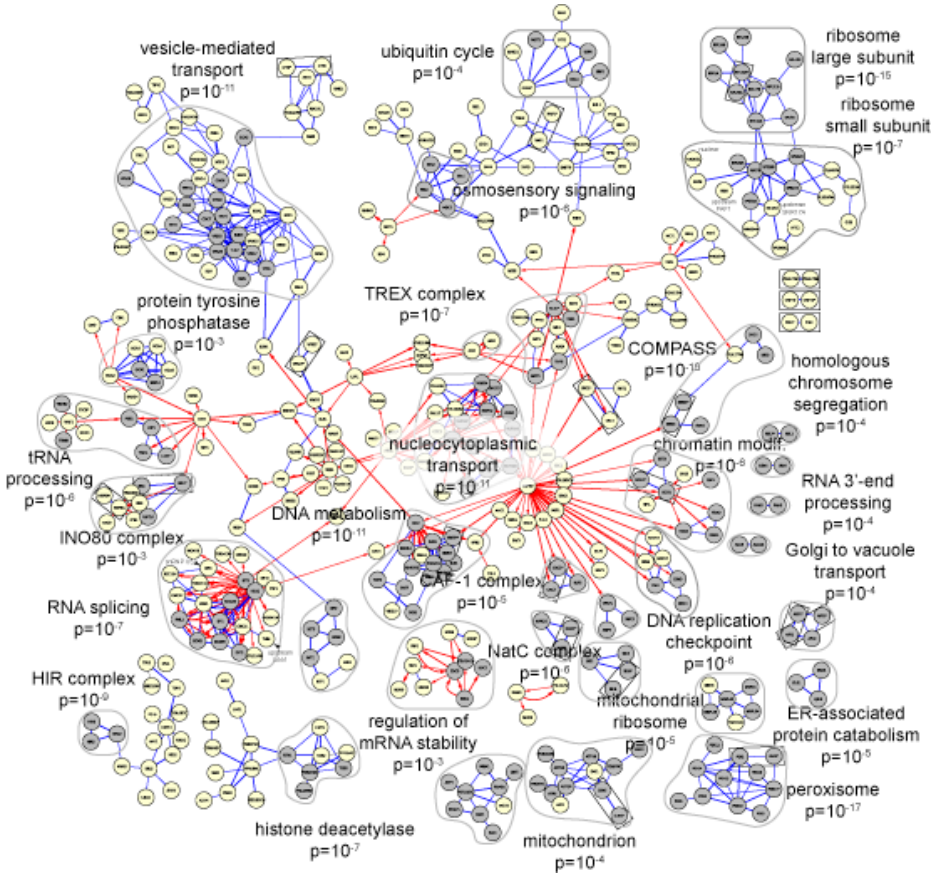
often models are needed (=

- gene co-expression
- …., …..

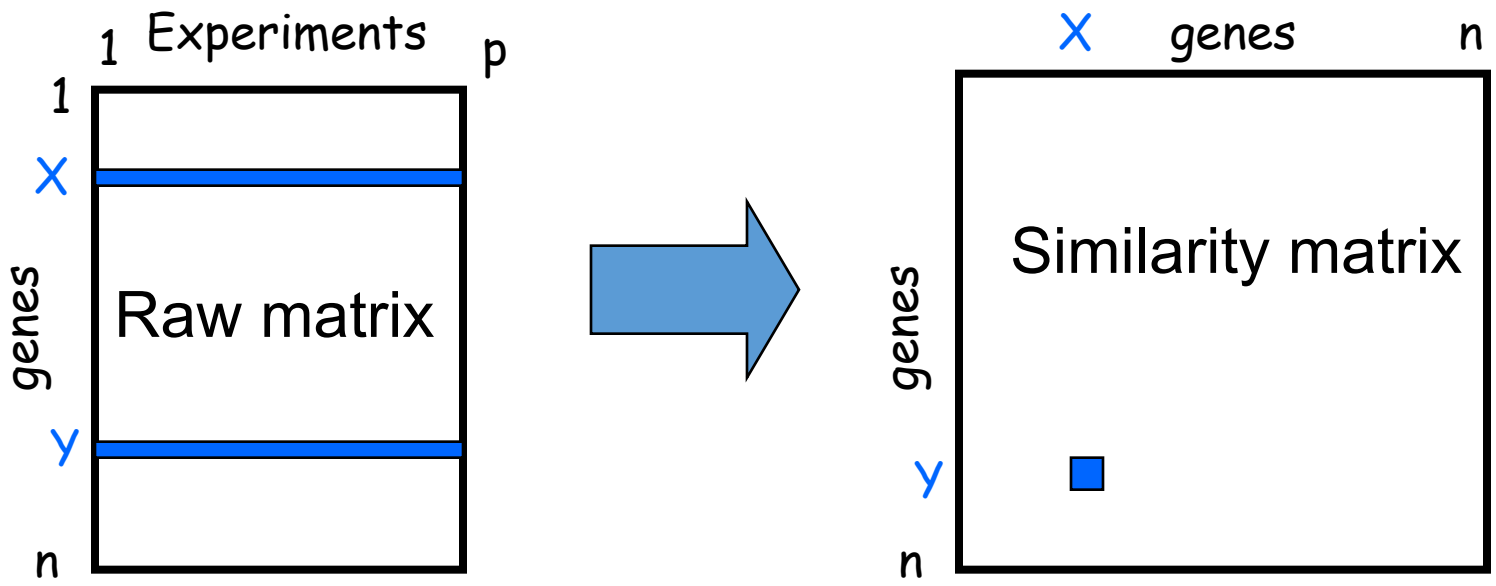
Co-expression Networks

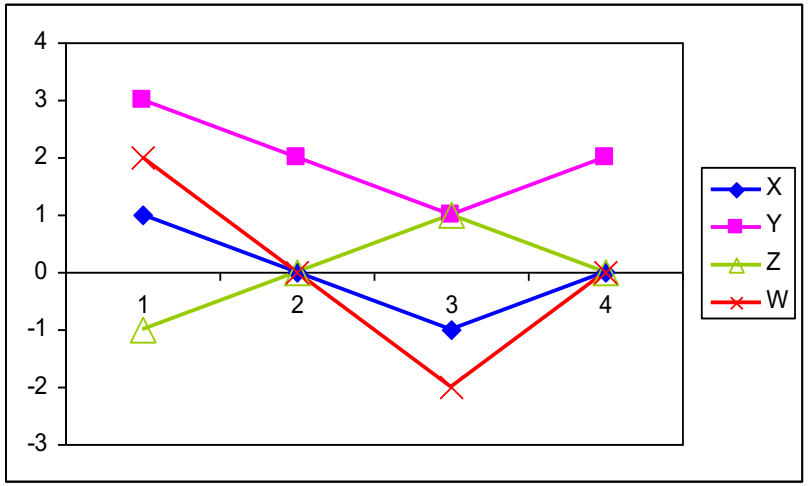
Nodes are connected if they have a significant pairwise expression profile association across environmental perturbations





Correlation: pairwise similarity





Correlation (X,Y) = 1

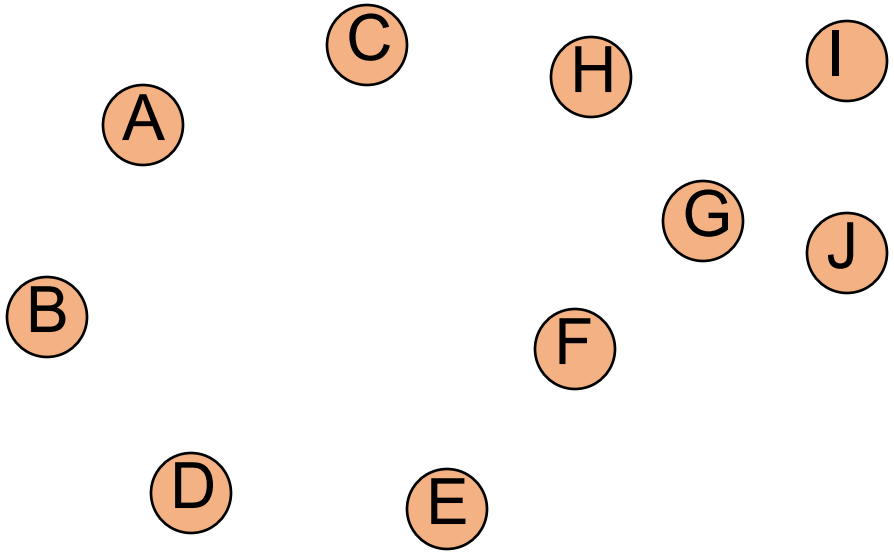
Correlation (X,Z) = -1

Correlation (X,W) = 1

Example: co-expression network

Correlation matrix

	A	B	c	D	E	F	G	H	I	J
A		0.91	0.72	0.84		0.78	0.88			
B	0.91									
C	0.72									
D	0.84									
E						0.94				
F	0.78				0.94		0.75			
G	0.88					0.75		0.92		
H							0.92			
I										0.98
J									0.98	



Correlation threshold, τ=1

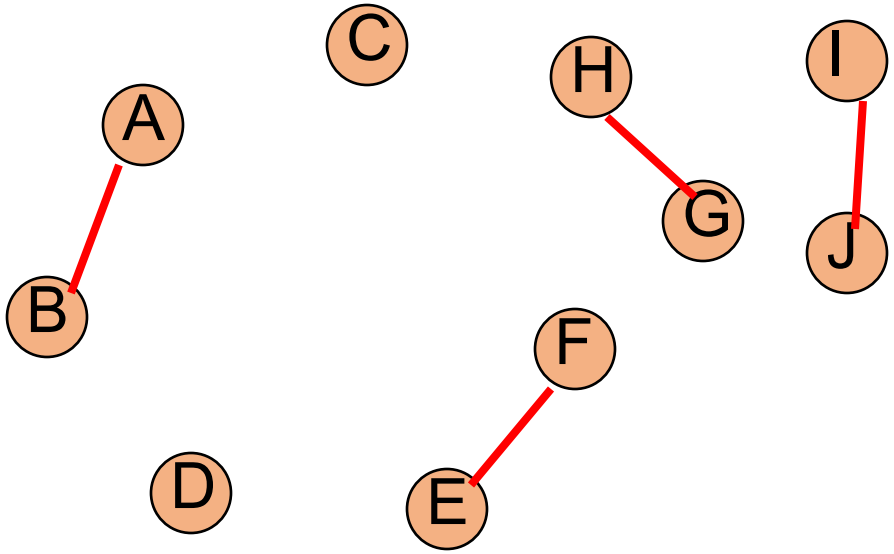
k	P(k)
0	10/10

At τ=1, there are no edges, so all nodes have degree (k) = 0

Example: co-expression network

Correlation matrix

	A	B	c	D	E	F	G	H	I	J
A		0.91	0.72	0.84		0.78	0.88			
B	0.91									
C	0.72									
D	0.84									
E						0.94				
F	0.78				0.94		0.75			
G	0.88					0.75		0.92		
H							0.92			
I										0.98
J									0.98	



Correlation threshold, τ=0.9

k	P(k)
0	2/10
1	8/10

At τ=0.9, there are 4 edges, so 8 nodes have degree (k) = 1

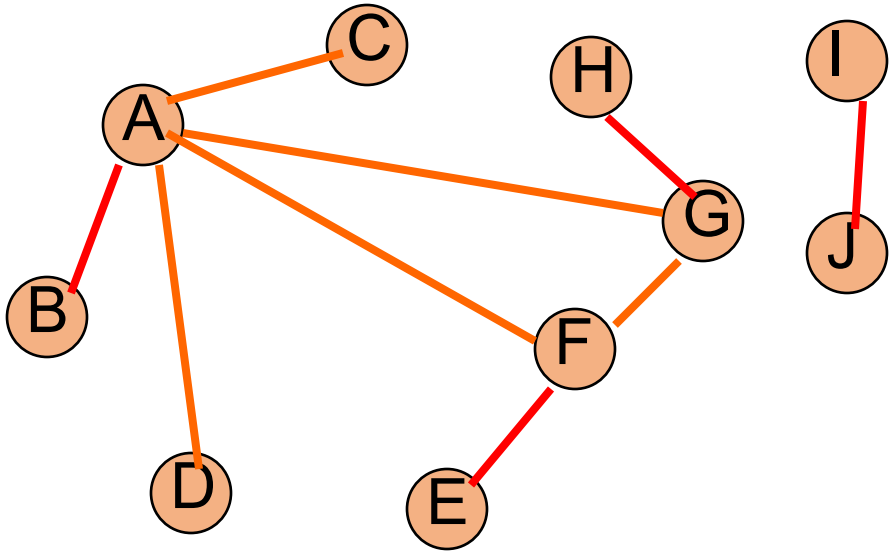
Example: co-expression network

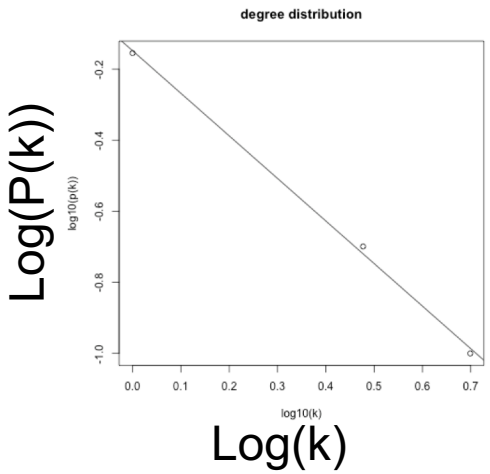
Correlation matrix

	A	B	c	D	E	F	G	H	I	J
A		0.91	0.72	0.84		0.78	0.88			
B	0.91									
C	0.72									
D	0.84									
E						0.94				
F	0.78				0.94		0.75			
G	0.88					0.75		0.92		
H							0.92			
I										0.98
J									0.98	

Correlation threshold, τ=0.7

k	P(k)
1	7/10
3	2/10
5	1/10





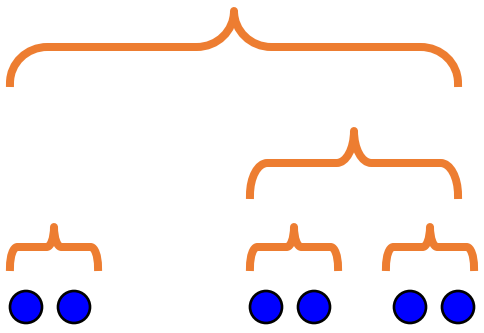
Clustering for network reconstruction

- Clustering: extract groups of genes that are tightly co-expressed over a range of different experiments.
- Pattern discovery
- No prior knowledge required

Clustering algorithms

- Inputs:
  - Similarity matrix
  - Number of clusters or some other parameters
- Many different classifications of clustering algorithms:
  - Hierarchical vs partitional
  - Heuristic-based vs model-based
  - Soft vs hard

Hierarchical Clustering



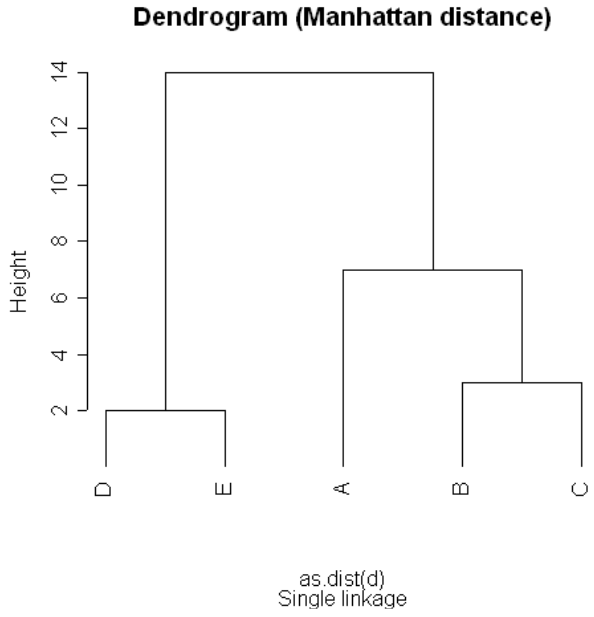
dendrogram

- Agglomerative (bottom-up)
- Algorithm:
  - Initialize: each item a cluster
  - Iterate:
    - select two most similar clusters
    - merge them
  - Halt: when required number of clusters is reached

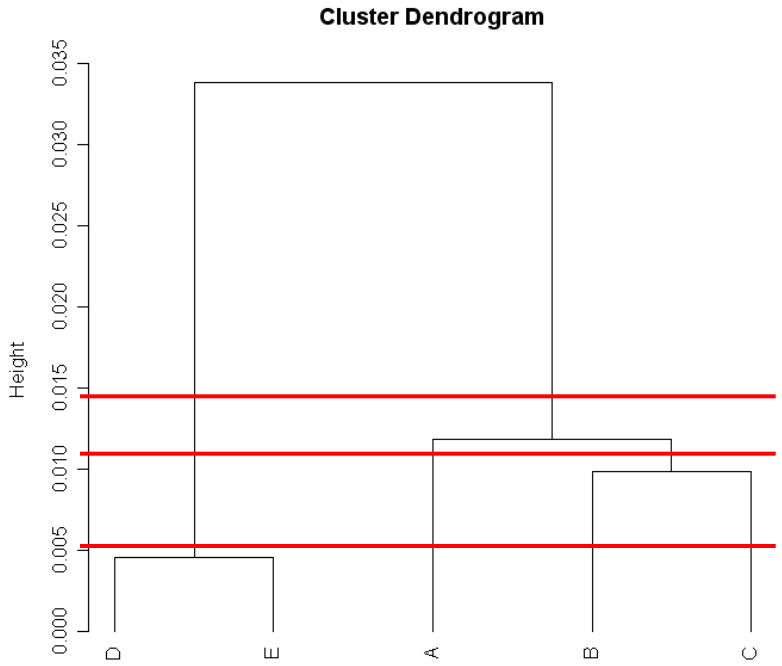
Hierarchical clustering for network reconstruction

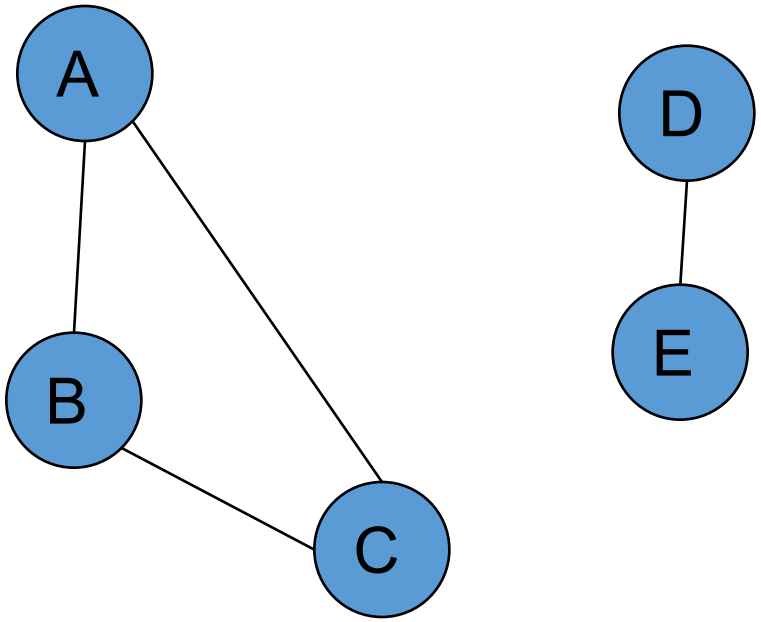
array \ gene	A	B	C	D	E
1	5	3	3	-1	0
2	3	1	2	-3	-4
3	2	0	-1	-3	-3
4	1	-1	0	-4	-4

log(Cy5/Cy3)

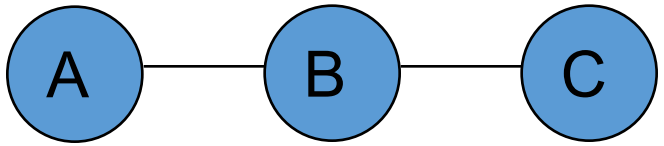


Hierarchical clustering for network reconstruction





True interactions?



Connect genes that are in the same cluster

Weighted Co-expression Networks

- SAGMB
- WGCNA

References:

- A general framework for weighted gene co-expression network analysis (Zhang, Horvath SAGMB 2005)
- WGCNA: an R package for weighted correlation network analysis. (Langfelder, Horvath BMC Bioinformatics 2008)

DISCUSSION

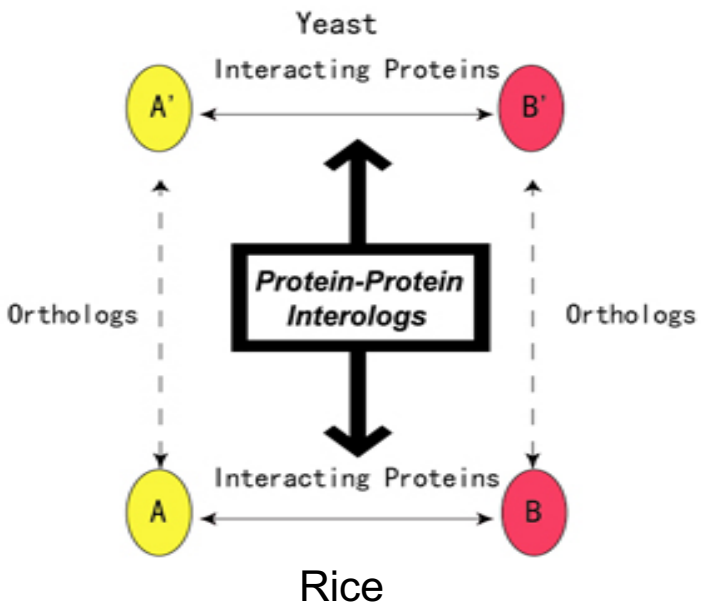
Any other type of data is helpful ?

Interolog Mapping: Orthologs

- Interest in Orthologs

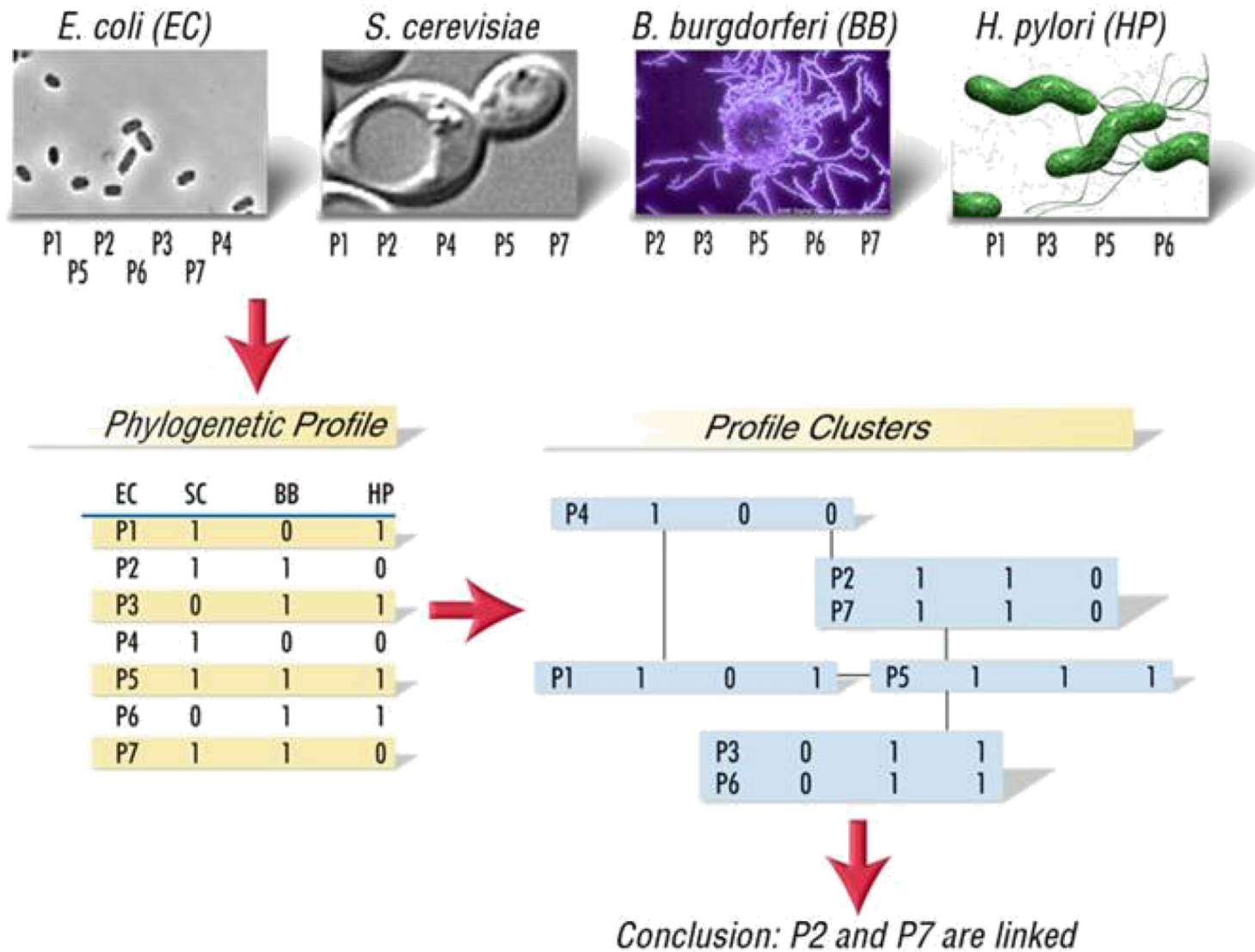
Maintain function → Maintain interactions

  - Key concept: If A and B interact in one spe orthologs A' and B' will interact
  - (A' & B') = "interologs" of (A & B)



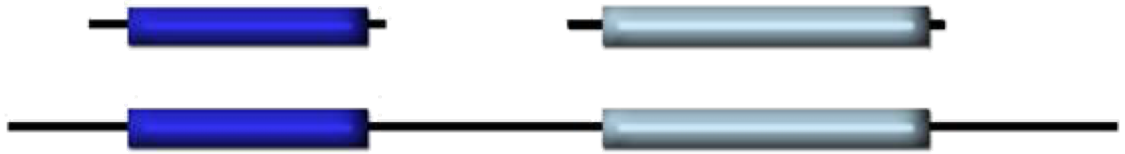
- Defining Orthologs
  - Loose definition: Top-blast hit
  - Stringent definition: Reciprocal top-blast hit
  - Not all orthologs can be found using above definitions

Gu et al (2011) PRIN: a predicted rice interactome network. BMC Bioinformatics.12:161, STRING Network



Pellegrini M, Marcotte EM, Thompson MJ, Eisenberg D, Yeates TO, Assigning protein functions by comparative genome analysis: protein phylogenetic profiles. *Proc Natl Acad Sci U S A*. 96(8):4285-8,. 1999

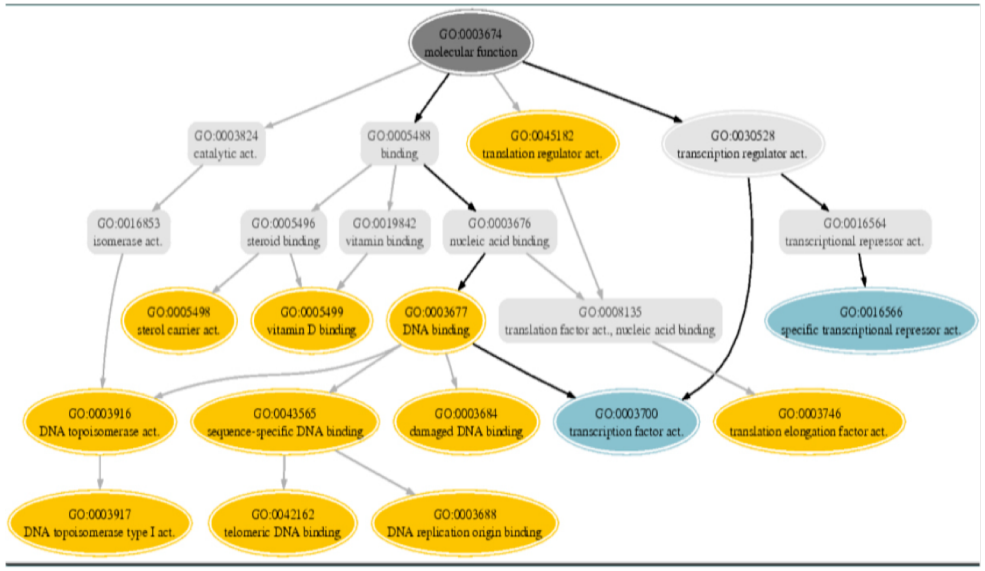
Protein Fusions



Monomeric proteins that are found fused in another organism are likely to be functionally related and physically interacting.

Marcotte EM, Pellegrini M, Ng HL, Rice DW, Yeates TO, Eisenberg D, Detecting protein function and protein-protein interactions from genome sequences. *Science* 285(5428):751-3, 1999

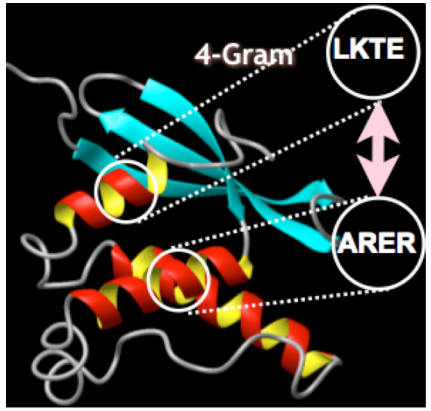
GO similarity (gene function)

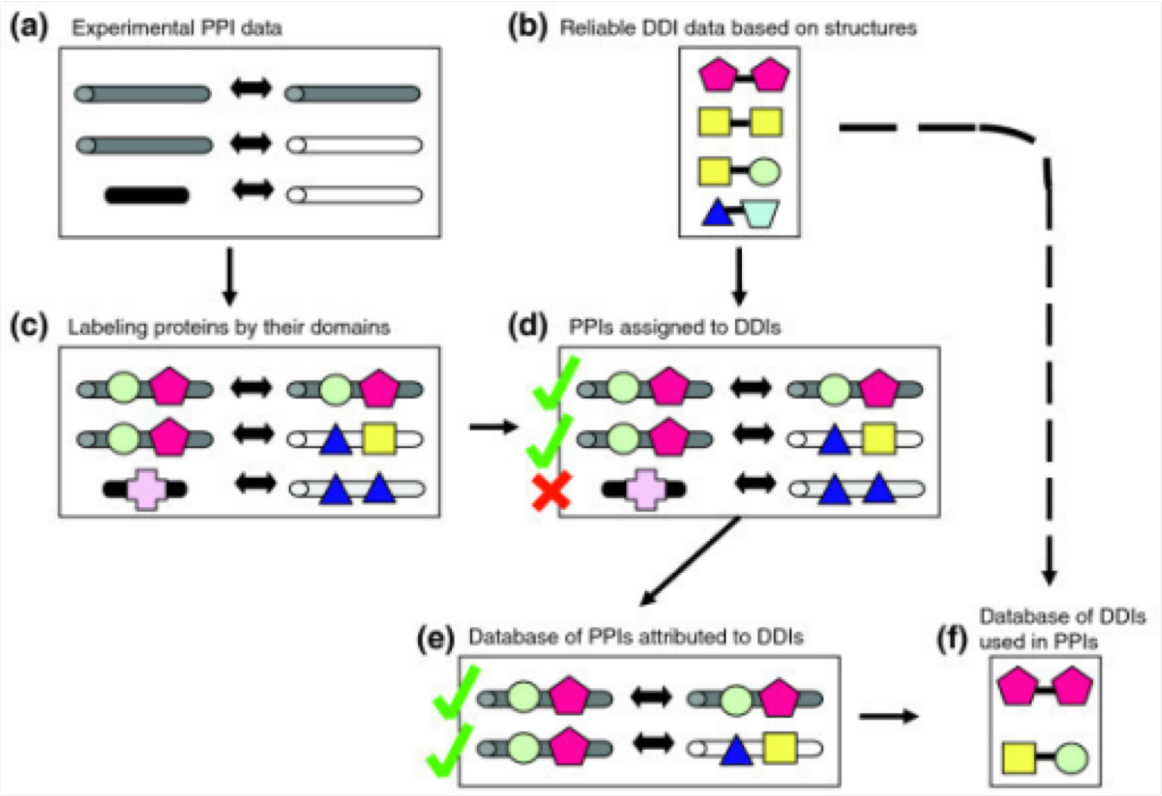


Proteins with the same biological function are more likely to physically interact than those without. In addition, proteins sharing a more specific annotation are more likely to interact than those sharing a commoner less specific annotation.

S. Jain, et al. BMC Bioinformatics, 11:562, 2010.

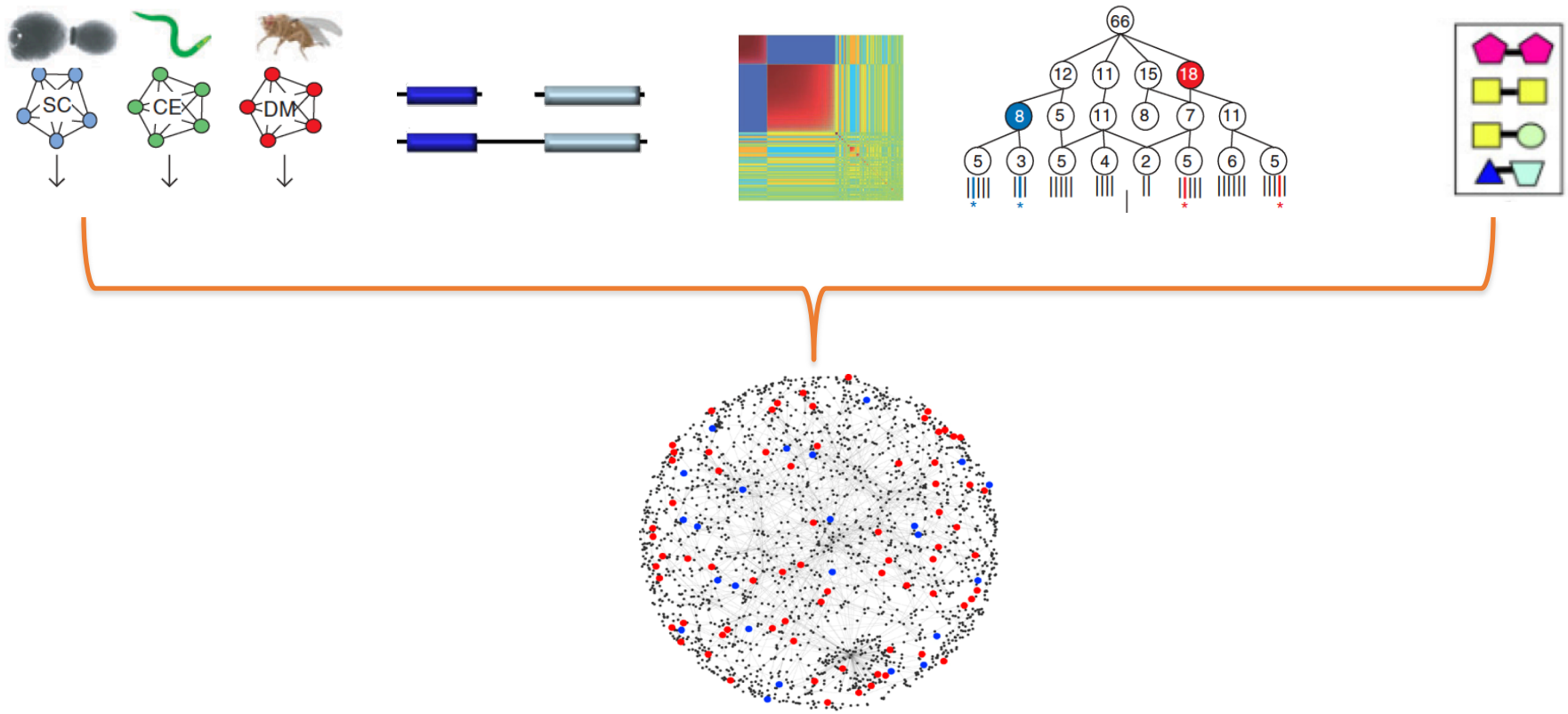
Interaction Domain(DDI)





Zhang et al, GAIA: a gram-based interaction analysis tool – an approach for identifying interacting domains in yeast. BMC Bioinformatics 2009, 10(Suppl 1):S60

Integrating data



Integrating experimental data

- Instead of using only one type of data use several types.

- This will:
  - give more supporting evidence that a protein performs a certain function.
  - reduce the number of false positive and false negative interactions.
  - give a more complete picture of the interactions between different elements involved in a certain biological process.

Probabilistic network approach

Bayesian Approach

Each "interaction" link between two proteins has a posterior probability of existence, based on the quality of supporting evidence.

.

Rhodes et al (2005). Nature Biotechnology 23(8):951-9.

Bayesian Approach

- A scalar score for a pair of genes is computed separately for each information source.
- Using gold positives (known interacting pairs) and gold negatives (known non-interacting pairs) interaction likelihoods for each information source is computed.
- The product of likelihoods can be used to combine multiple information sources
  - Assumption: A score from a source is independent from a score from another source.

Bayesian Approach for PPI Prediction

The posterior odds of interaction

The prior odds of interaction

$$O_{post} = O_{prior} \times L$$

Likelihood ratio

$$L = Pr(f_1 \ldots f_n \mid GSP) \div Pr(f_1 \ldots f_n \mid GSN)$$

$$\boxed{L = \prod_{i=1}^{i=n} Pr(f_i \mid GSP) \div Pr(f_i \mid GSN)}$$

$$O_{prior} = P(pos) \,/\, P(neg),$$

Constant value

Computing the likelihoods

- Partition the pair scores of an information source into bins and provide likelihoods for score-ranges

- E.g. Using the microarray information source and using Pearson correlation for scoring protein pairs you may get scores between -1 and 1. You want to know what is the likelihood of interaction for a protein pair that gets a Pearson correlation of 0.9.

Partitioning the scores

pearson corr.	Likelihood （*L*）
(0.8,1.0]	
(0.6,0.8]	
(0.4,0.6]	
(0.2,0.4]	
(0.0,0.2]	
(-0.2,0.0]	
(-0.4,-0.2]	
(-0.6,-0.4]	
(-0.8,-0.6]	
[-1.0,-0.8]	

Computing the likelihood

- P(Interaction | Score) / P (Interaction)

L = ----------

P(~Interaction | Score) / P (~Interaction)

$$\Pr(f_i \mid GSP) \div \Pr(f_i \mid GSN)$$

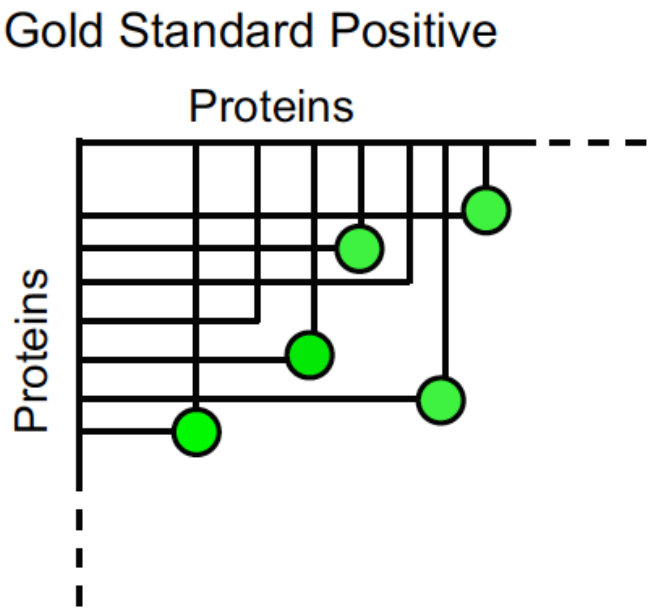
Example:

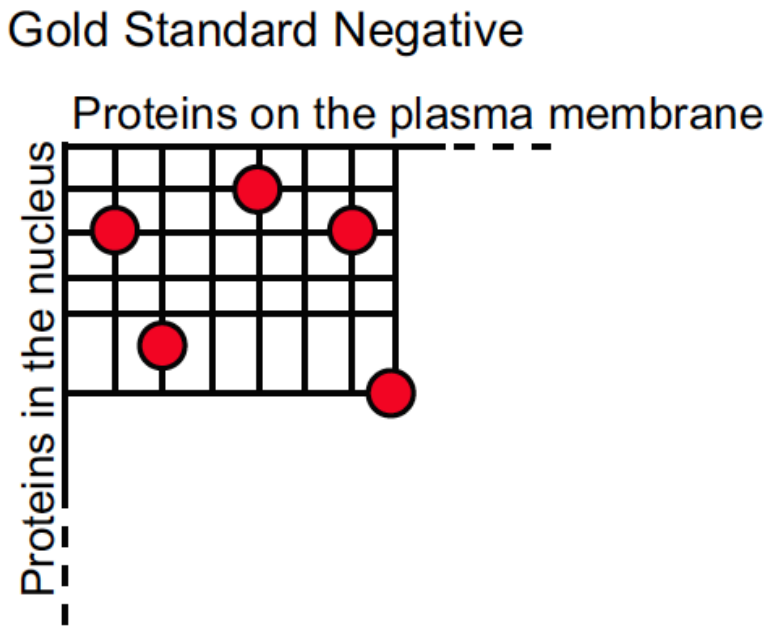
Partitioning the scores

pearson corr.	P(exp|pos)	P(exp|neg)	Likelihood （*L*）
(0.8,1.0]	0.4	0.05	0.4/0.05=8
(0.6,0.8]			
(0.4,0.6]			
(0.2,0.4]	0.15	0.12	0.15/0.12=1.25
(0.0,0.2]			
(-0.2,0.0]			
(-0.4,-0.2]			
(-0.6,-0.4]			
(-0.8,-0.6]			
[-1.0,-0.8]			

Training data sets

- A Bayesian networks approach for predicting PPI





von Mering, et al Nucleic Acids Res. 2005

Example

- A Bayesian networks approach for predicting protein-protein interactions from genomic data

Data type	Dataset		# protein pairs	Used for ...
Experimental interaction data	In-vivo pull-down	Gavin et al.	31,304	Integration of experimental interaction data (PIE)
		Ho et al.	25,333	
	Yeast two-hybrid	Uetz et al.	981	
		Ito et al.	4,393	
Other genomic features	Expression	Rosetta compendium	19,334,806	De novo prediction (PIP)
		Cell cycle	17,467,005	
	Biological function	GO biological process	3,146,286	
		MIPS function	6,161,805	
	Essentiality		8,130,528	
Gold standards	Positives	Proteins in the same MIPS complex	8,250	Training & testing
	Negatives	Proteins separated by localization	2,708,746	

Jansen *et al.* (2003) *Science*.

Some data … …

Expression correlation		# protein pairs	Gold standard overlap		*P(exp|pos)*	*P(exp|neg)*	*L*
			pos	*neg*			
Values	0.9	617	16	45	2.10E-03	1.68E-05	124.93
	0.8	4,127	137	563	1.80E-02	2.10E-04	85.50
	0.7	14,979	530	2,117	6.96E-02	7.91E-04	87.97
	0.6	36,145	1,073	5,597	1.41E-01	2.09E-03	67.36
	0.5	81,102	1,089	14,459	1.43E-01	5.40E-03	26.46
	0.4	189,369	993	35,350	1.30E-01	1.32E-02	9.87
	0.3	444,757	1,028	83,483	1.35E-01	3.12E-02	4.33
	0.2	1,016,105	870	183,356	1.14E-01	6.85E-02	1.67
	0.1	2,205,895	739	368,469	9.71E-02	1.38E-01	0.70
	0	8,118,256	894	1,244,477	1.17E-01	4.65E-01	0.25
	-0.1	2,345,009	164	408,562	2.15E-02	1.53E-01	0.14
	-0.2	1,038,181	63	203,663	8.27E-03	7.61E-02	0.11
	-0.3	399,554	13	84,957	1.71E-03	3.18E-02	0.05
	-0.4	131,361	3	28,870	3.94E-04	1.08E-02	0.04
	-0.5	40,759	2	8,091	2.63E-04	3.02E-03	0.09
	-0.6	15,289	-	2,134	0.00E+00	7.98E-04	0.00
	-0.7	6,795	-	807	0.00E+00	3.02E-04	0.00
	-0.8	1,886	-	261	0.00E+00	9.76E-05	0.00
	-0.9	55	-	12	0.00E+00	4.49E-06	0.00
Sum		16,090,241	7,614	2,675,273	1.00E+00	1.00E+00	1.00

Some data … …

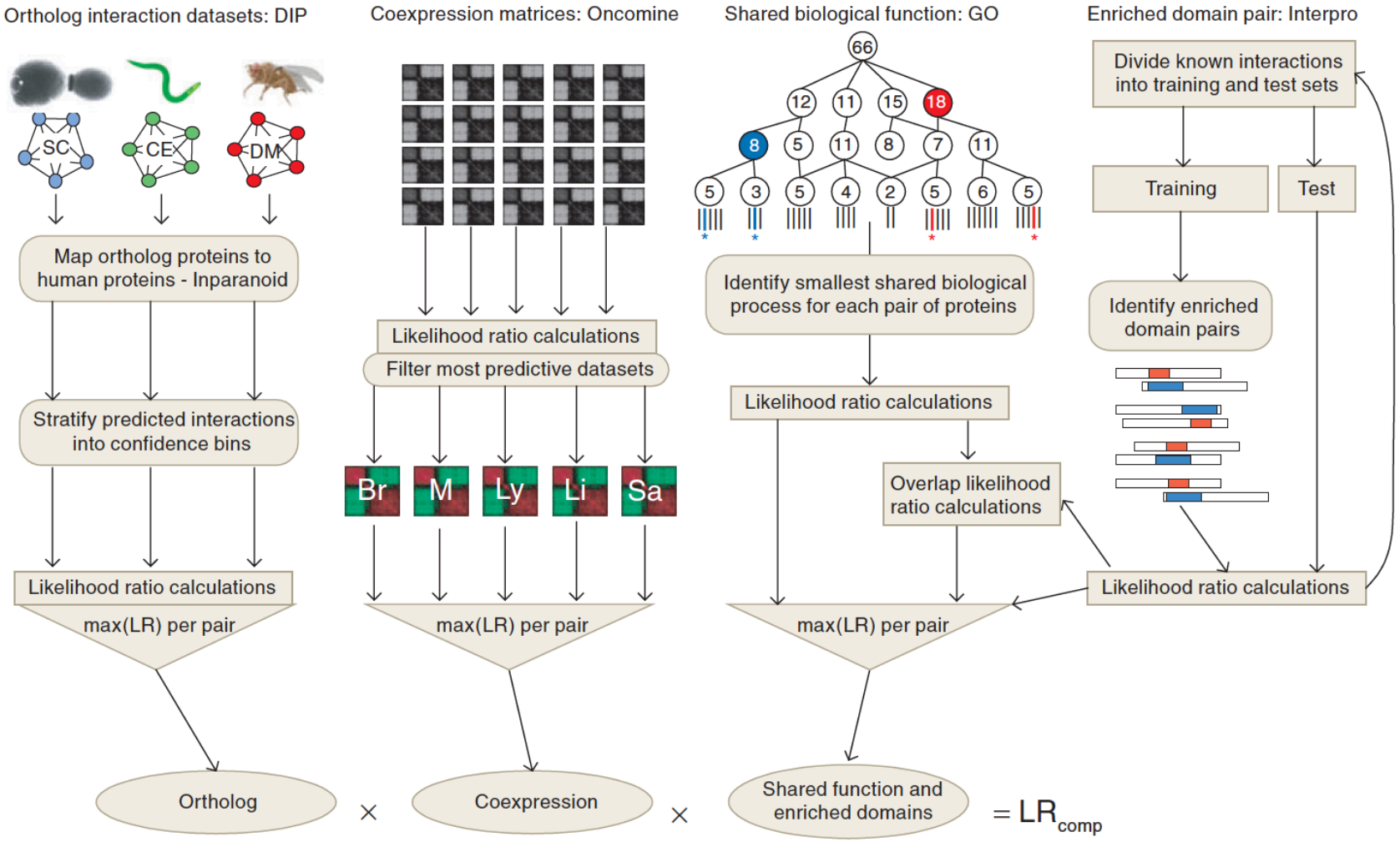
Essentiality		# protein pairs	Gold-standard overlap		*P(Ess|pos)*	*P(Ess|neg)*	*L*
			pos	*neg*			
Values	EE	301,088	1,114	81,924	5.18E-01	1.43E-01	3.63
	NE	2,481,701	624	285,487	2.90E-01	4.98E-01	0.58
	NN	4,771,865	412	206,313	1.92E-01	3.60E-01	0.53
Sum		7,554,654	2,150	573,724	1.00E+00	1.00E+00	1.00

GO biological process similarity		# protein pairs	Gold standard overlap		*P(GO|pos)*	*P(GO|neg)*	*L*
			pos	*neg*			
Values	1 – 9	4,789	88	819	1.17E-02	1.27E-03	9.22
	10 -- 99	20,467	555	3,315	7.38E-02	5.14E-03	14.36
	100 – 1000	58,738	523	10,232	6.95E-02	1.59E-02	4.38
	1000 -- 10000	152,850	1,003	28,225	1.33E-01	4.38E-02	3.05
	10000 – Inf	2,909,442	5,351	602,434	7.12E-01	9.34E-01	0.76
Sum		3,146,286	7,520	645,025	1.00E+00	1.00E+00	1.00

If we set L_{cut}=600, given protein A and B, their expression correlation=0.85, GO similarity=156, Essentiality value=EE, is there interaction between them?

L=85.50*3.63*4.38=1359.40 > 600

More examples



Rhodes, *et al*. Probabilistic model for the human protein-protein interaction network, *Nature Biotechnology* (2005)

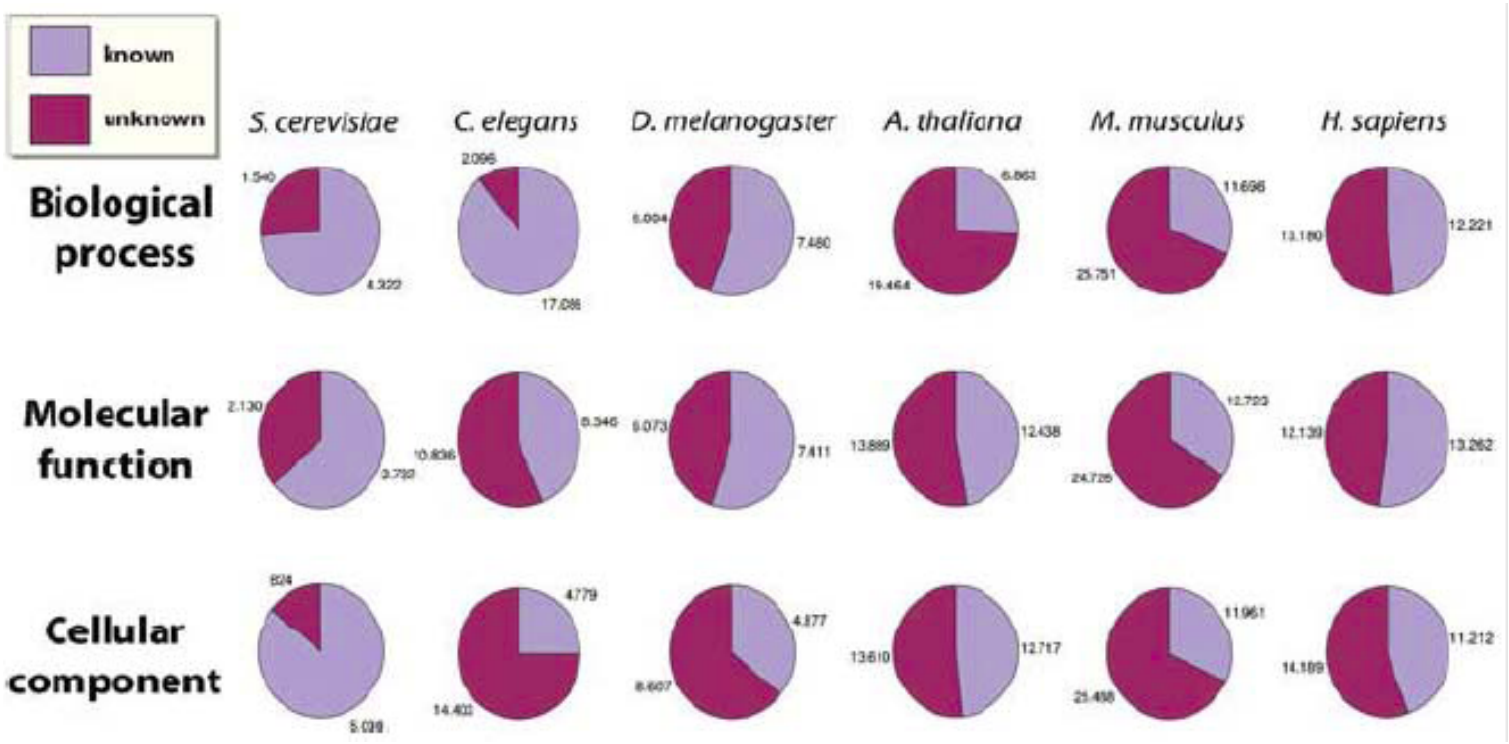
Topics

- Network and network topology
- Network reconstruction
- Network application

Network application in biomedicine

- Gene/Protein functional annotation
- Key/disease gene prioritization
- Network-based biomarker or molecular signature

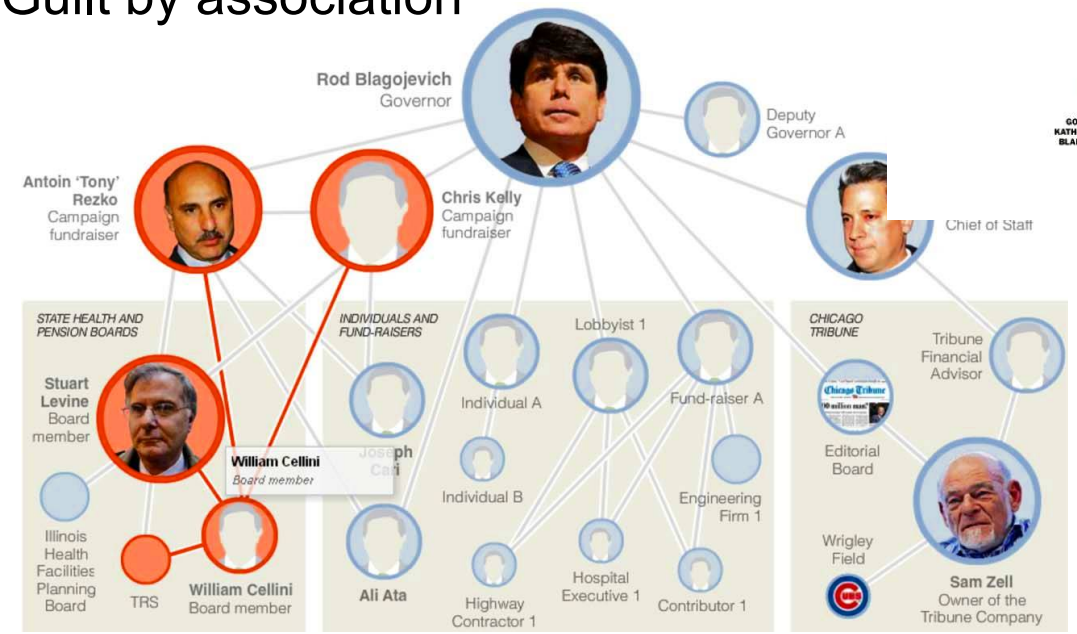
Protein function prediction

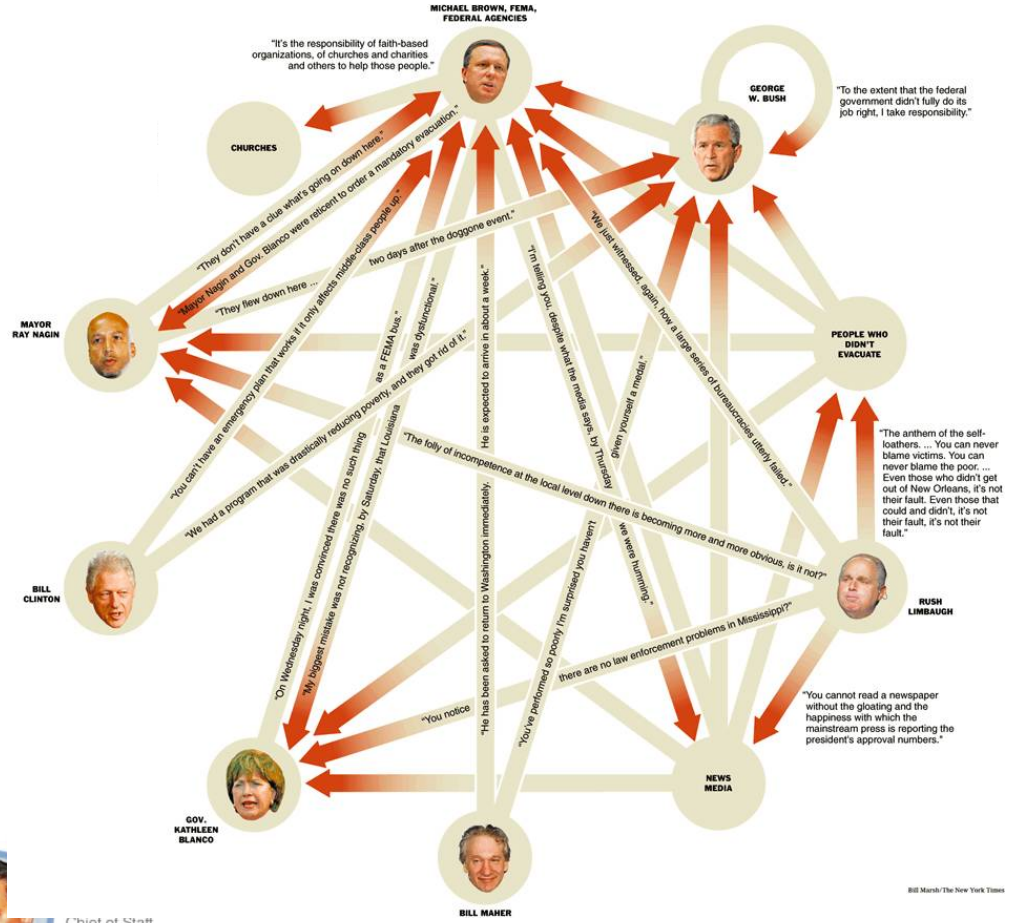


What can we do with these molecular networks?

Using the position in networks to describe function

Guilt by association

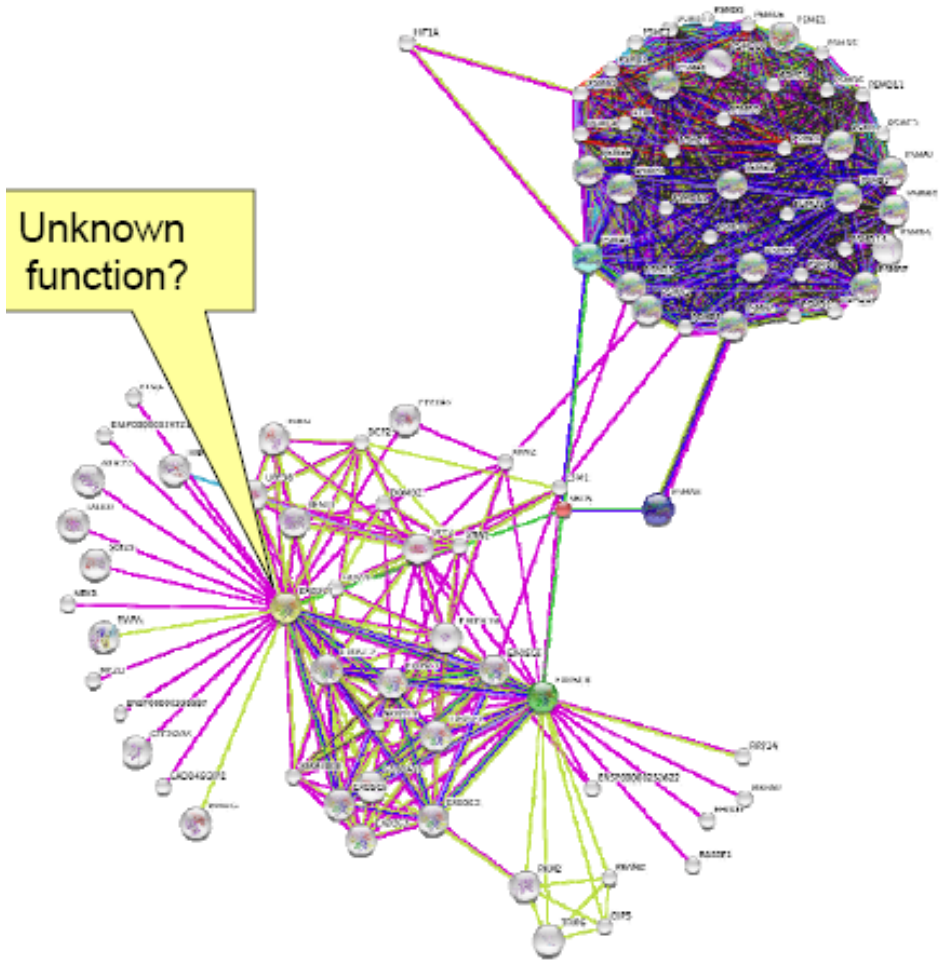




Finding the causal regulator
(the "Blame Game")

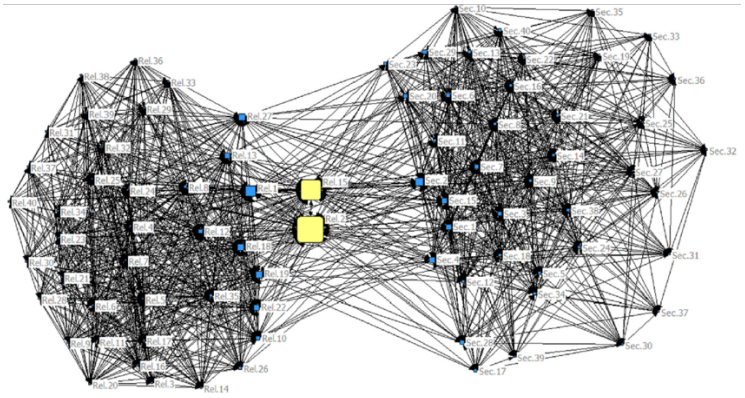
Courtesy of Mark Gerstein

Protein function prediction



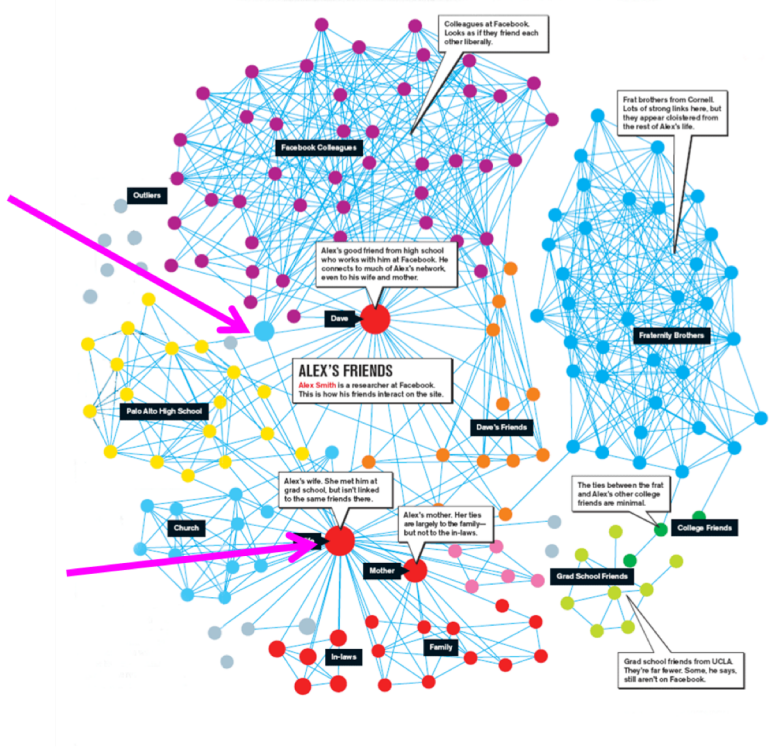
- *Majority voting (Local)*
- *K-neighborhood*
  - -Generalized majority voting
- Chi square
  - - Functional enrichment among the k-neighborhood as measured by the chi square score
- Module-assisted
  - - Module identification
  - - Functional enrichment evaluation

Key/disease gene prioritization

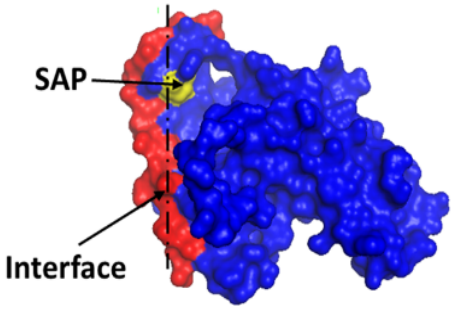


Degree ?

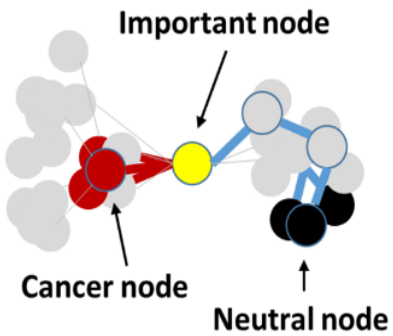
Betweenness ?



Network-based risk evaluation for SAPs in cancer



L	E	R	R	V	G	H	Q	A	Y	S
L	E	R	R	V	G	H	Q	A	Y	S
L	E	Q	R	V	G	H	Q	A	Y	S
L	E	R	R	V	G	H	Q	A	Y	S
L	E	R	K	V	G	Y	Q	F	Y	S
L	E	R	K	V	G	Y	Q	A	Y	S
L	E	R	K	V	G	Y	Q	A	Y	S



I-score

S-score

T-score

Condel

$$WAS = \frac{\sum_i S_i * W_i}{\sum_i W_i}$$

Meta-score

Rank	SAP	Meta-score	I-score	S-score	T-score
1	KRAS:G12D	0.97	1	1.00	0.90
2	CTSS:R113W	0.95	1	0.87	0.96
3	RTN4:A71T	0.74	0	1.00	0.90
… …	… …	… …	… …	… …	… …

Big Challenges

- Strategies for explaining unmatched spectrum (30%-50%)
- How to refine genome annotation using proteome and transcriptome data
- How to do protein identification when the reference genome is unknown
- Omics integration and clinical application
- Single cell , meta-omics, pan-omics, phen-omics ?

Thanks