Introduction to biological networks

Outline

- Measurements
- Analysis
- Modelling

Outline

- Measurements
  - + Expression
  - + Protein-protein interactions
  - + Protein-DNA interactions
- Analysis
- Modelling

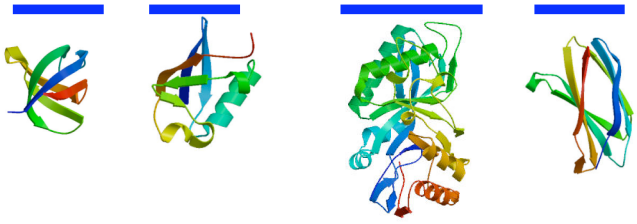
DNA

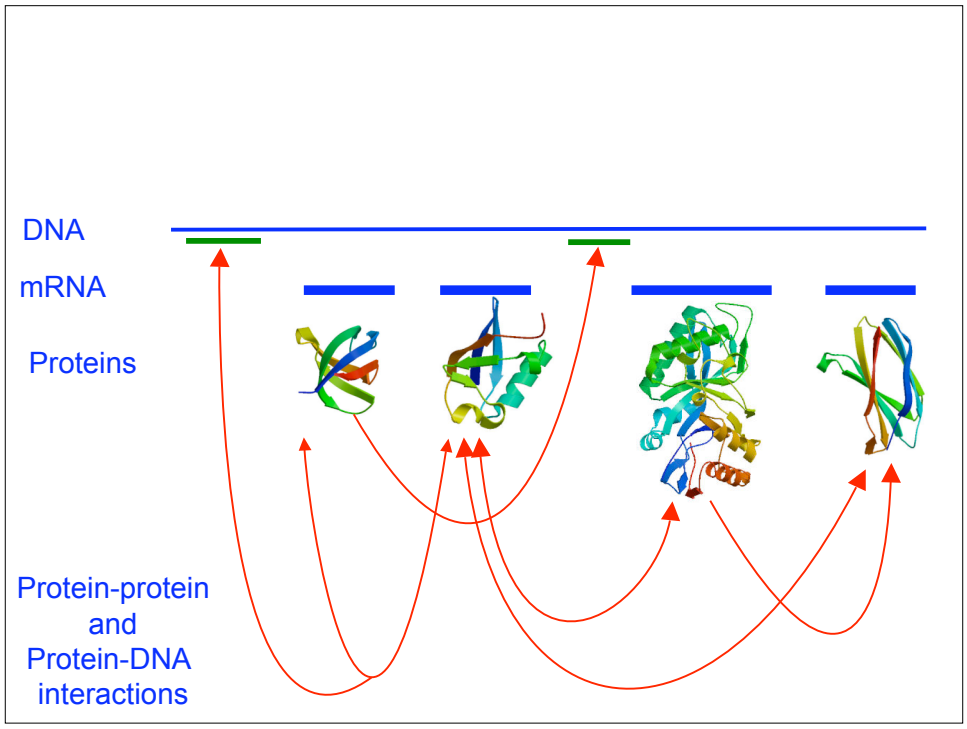
mRNA

DNA

mRNA

Proteins





Wanted measurements:

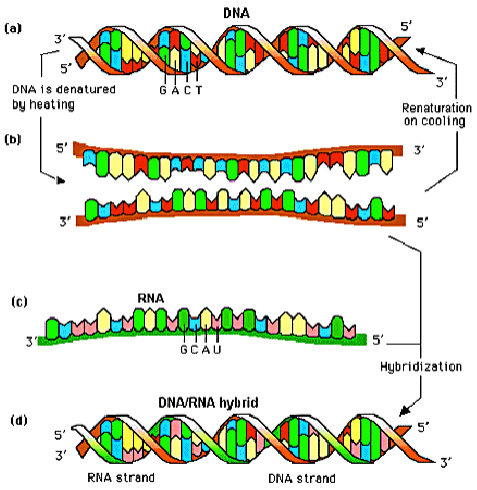
1. Concentration of mRNA RNA(t,**c**)
2. Concentration of protein Protein(t,**c**)
3. Protein interactions K(Protein1,Protein2|t,**c**)
4. Protein-DNA interactions K(Protein,Site|t,**c**)

Expression

Expression

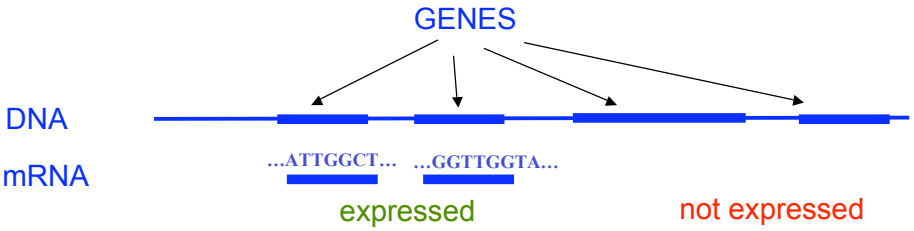
- mRNA level (average over many cells) at
  - different time points
  - different conditions
- Expression profiles
- Demo

Demo http://www.bio.davidson.edu/courses/genomics/chip/chip.html



Nucleic Acid Hybridization

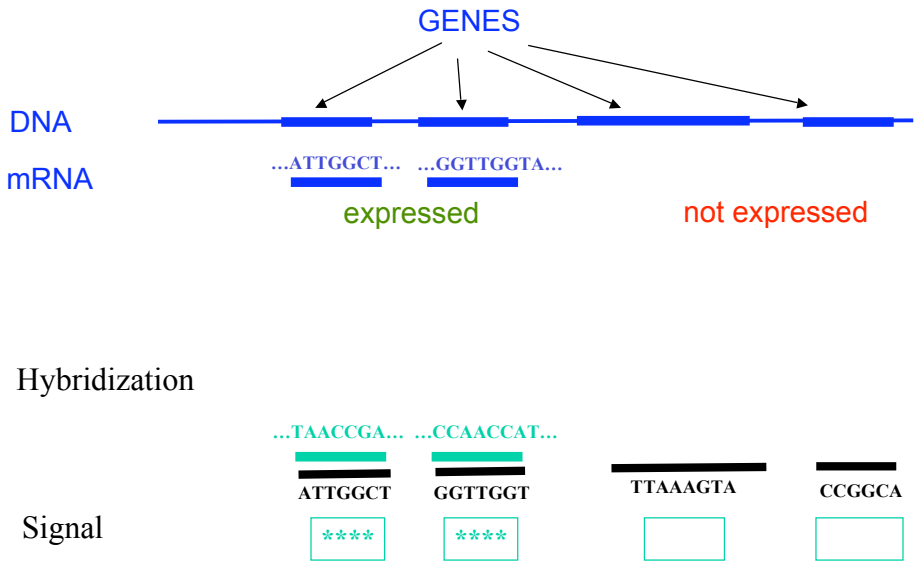
DNA chip measurements

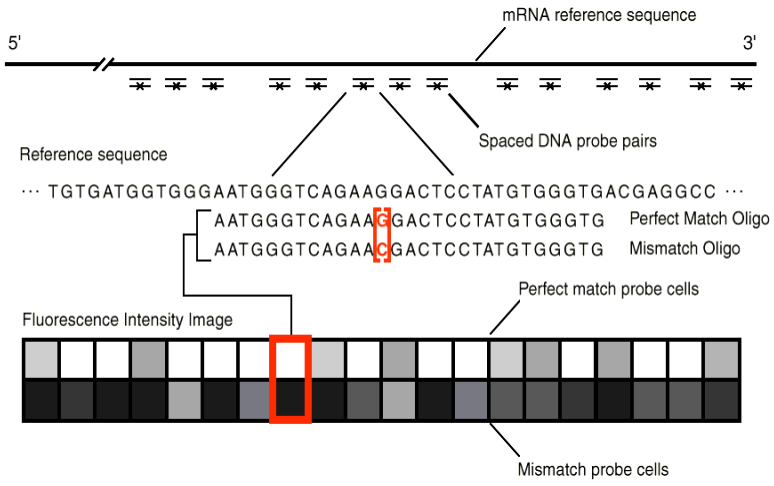


Make fluorescently labeled complements of all present mRNAs

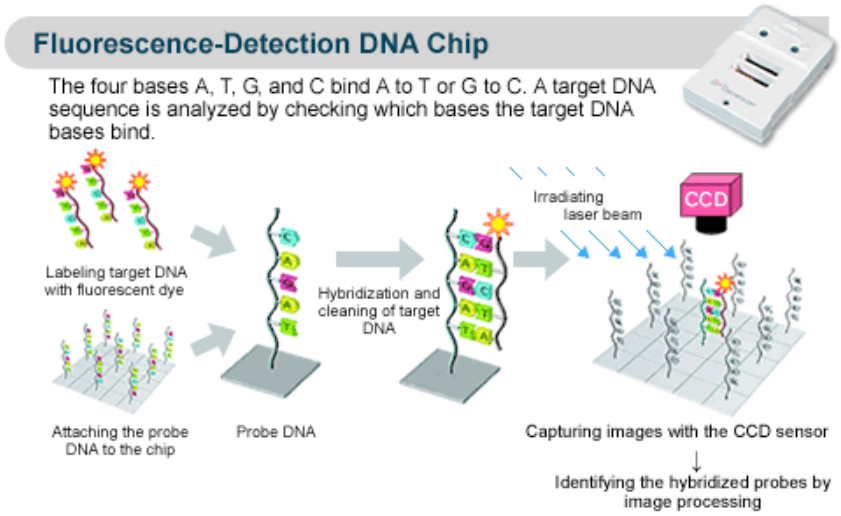
PROBES ON A CHIP

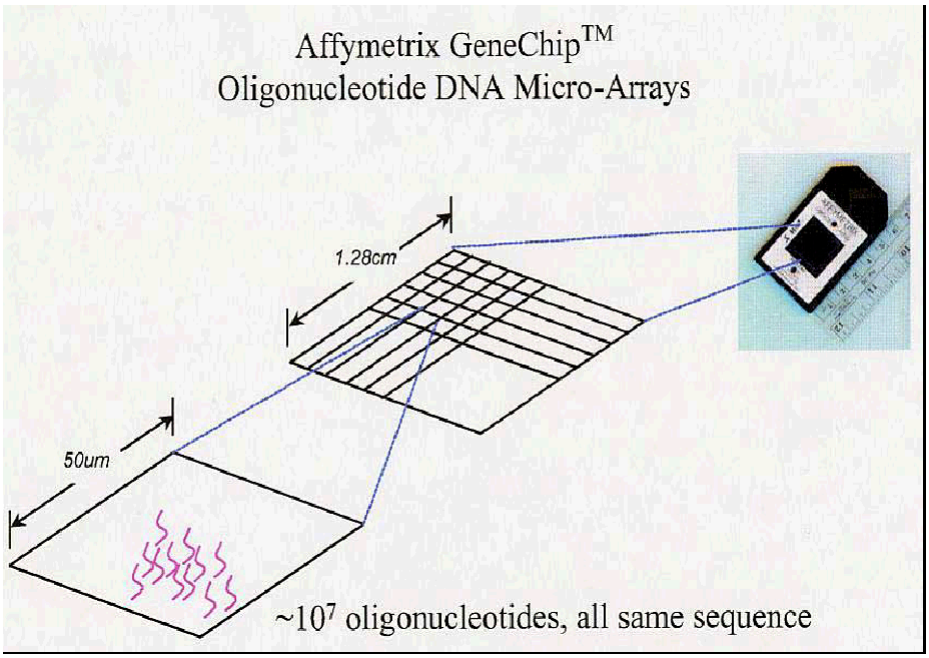
DNA chip measurements





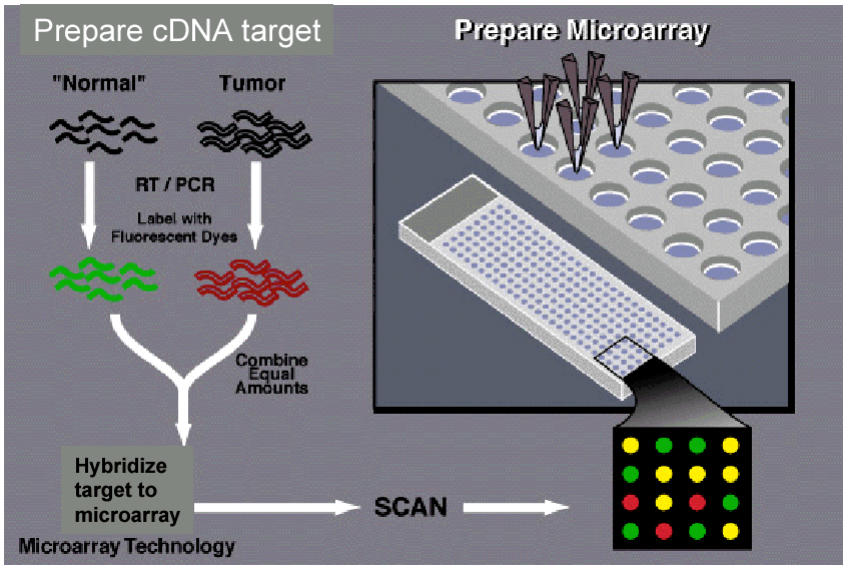
DNA chip


Fluorescence-Detection DNA Chip
The four bases A, T, G, and C bind A to T or G to C. A target DNA sequence is analyzed by checking which bases the target DNA bases bind.
Labeling target DNA with fluorescent dye
Attaching the probe DNA to the chip
Probe DNA
Hybridization and cleaning of target DNA
Irradiating laser beam
CCD
Capturing images with the CCD sensor
Identifying the hybridized probes by image processing

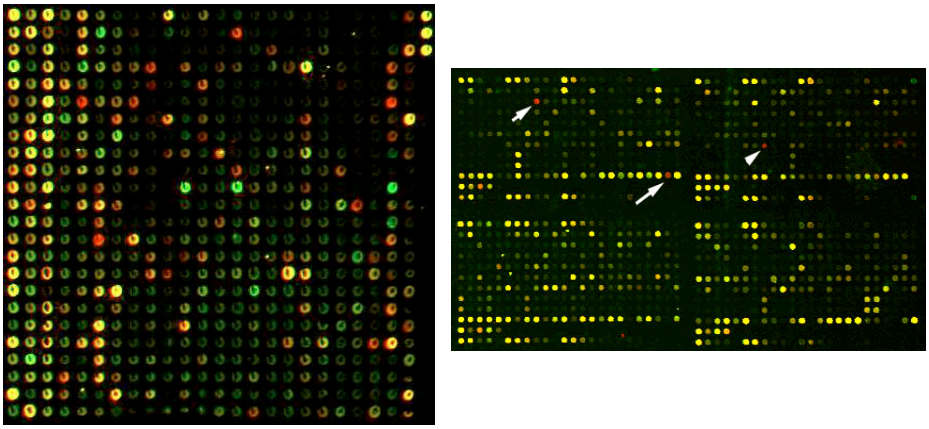

Affymetrix GeneChip™
Oligonucleotide DNA Micro-Arrays
1.28cm
50um
~10⁷ oligonucleotides, all same sequence

GeneChip® Probe Array

cDNA microarray expt



cDNA Spotted Microarrays

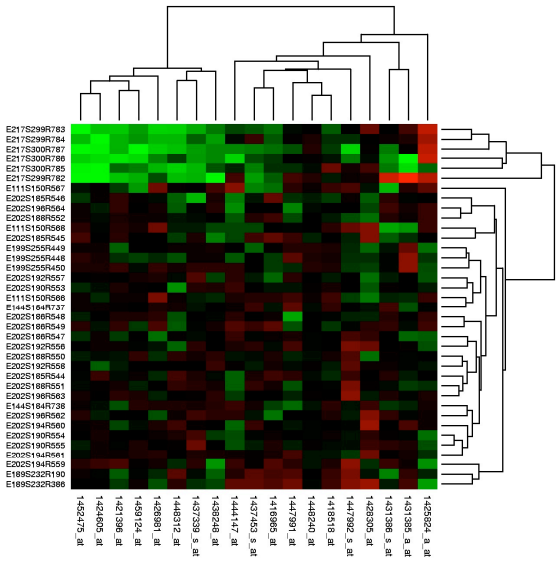


Outcome

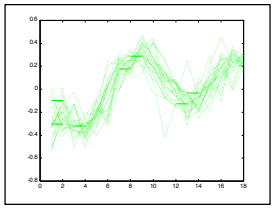
- mRNA level (average over many cells) at
  - different time points
  - different conditions
  - different tissues
  - normal and malignant samples

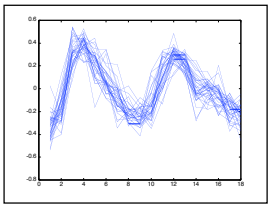
Expression profiles

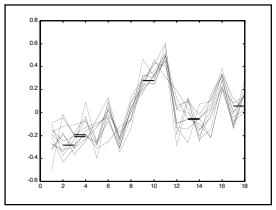
Expression profile

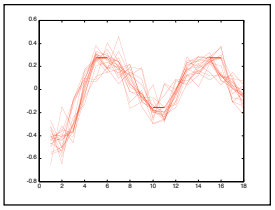


Expression profile: cell cycle









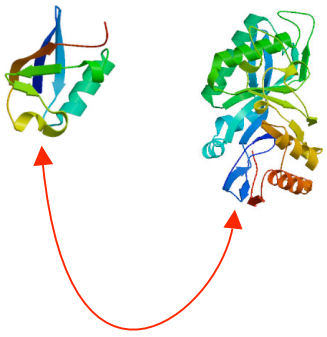
Research problems

1. Find genes that are differentially expressed, e.g. in response to perturbation.
2. Compare profiles and group genes or conditions that exhibit similar expression profiles: *clustering.*
3. Compare samples from two (or more) tissues and find features that can discriminate tissue, e.g. *expression signatures* of cancer types.
4. Given expression profiles from various perturbation experiments, *infer the regulatory network.*
5. Compare expression of different organisms *alignment of profiles/networks.*

Data

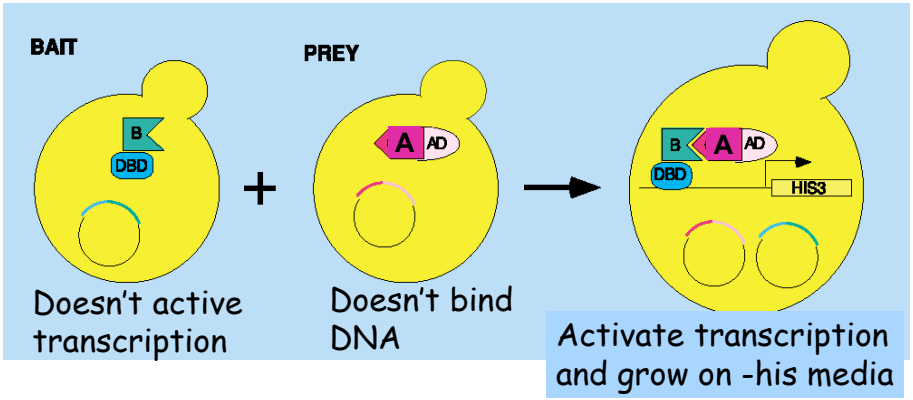
1. High coverage (all genes)
2. Average over large population of cells
3. Significant level of experimental noise
4. Hard cross-platform comparison
5. Few data-points

Protein-protein Interactions



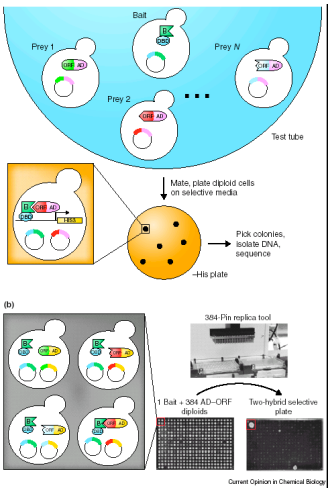
Protein-protein Interactions

Yeast two-hybrid assay:
Does a protein A interact with B ?



Protein-protein Interactions

Large scale yeast two-hybrid assay:
Find pairs of interacting proteins

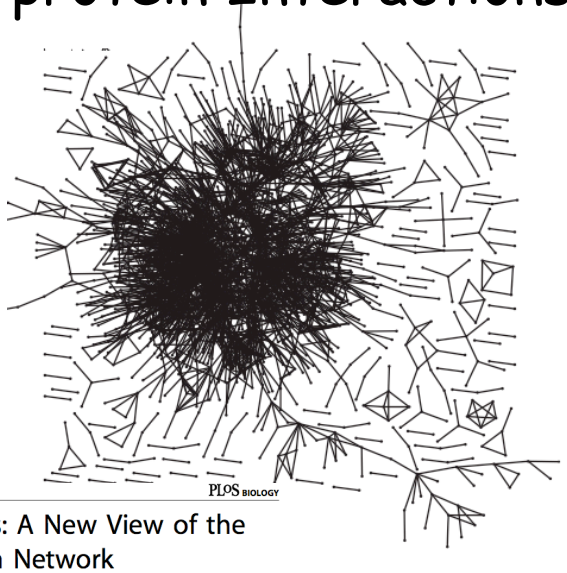


A comprehensive two-hybrid analysis to explore the yeast protein interactome.
Ito T et al, *PNAS* 2001

A comprehensive analysis of protein-protein interactions in *S. cerevisiae.*
P. Uetz et al, *Nature* 2000

Protein-protein Interactions

Baker's yeast:
9000 interactions
3000 proteins
(combined, 2006)



OPEN ACCESS Freely available online PLoS BIOLOGY

Stratus Not Altocumulus: A New View of the Yeast Protein Interaction Network

Nizar N. Batada[1,2*], Teresa Reguly[1], Ashton Breitkreutz[1], Lorrie Boucher[1,3], Bobby-Joe Breitkreutz[1], Laurence D. Hurst[2*], Mike Tyers[1,3*]

Research problems

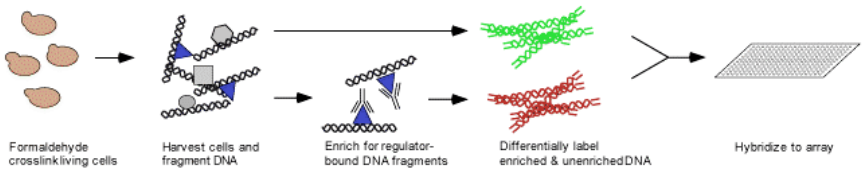
1. Characterize statistical properties of the network.
2. Connect statistical properties to biological function and evolution.
3. Reveal biologically important features of the network e.g. *clusters* or *motifs*.
4. Use networks to predict function of specific genes.
5. Compare/align networks.

Data

1. Many genes (up to 30%).
2. Measurements *in vivo*, but not in the endogenous cells.
3. Average over large population of cells.
4. High level of false-positives.
5. Non discrimination between direct and indirect interactions.
6. No quantitative measure of the interaction strength.

Protein-DNA interactions

Chromatin Immunoprecipitation (ChIP)



ChIP chip measurements

DNA

PROBES ON A CHIP

For promoters (~1-2Kb around start of a gene)

Or tiling array of the whole genome

Make fluorescently labeled complements of DNA fragments bound to the protein of interest

ChIP chip measurements

DNA

PROBES ON A CHIP

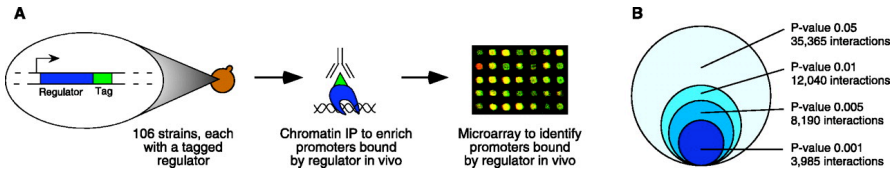
For promoters (~1-2Kb around start of a gene)

Or tiling array of the whole genome

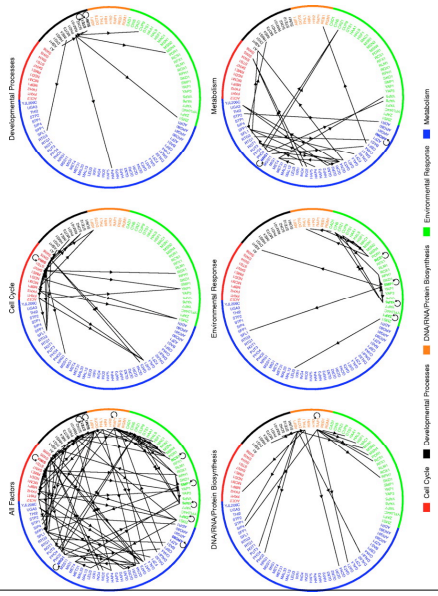
HYBREDIZATION

protein-DNA interactions

Chromatin Immunoprecipitation (ChIP)



Protein-DNA interactions

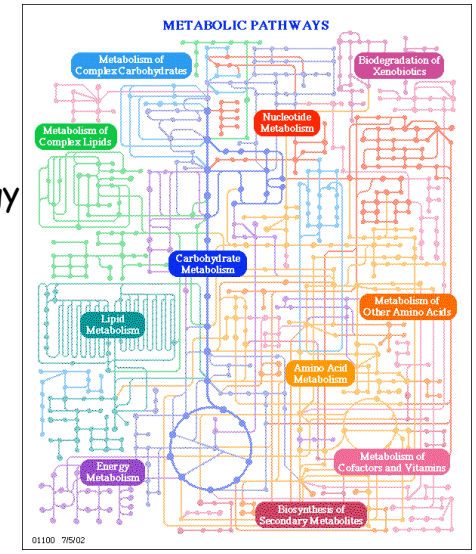


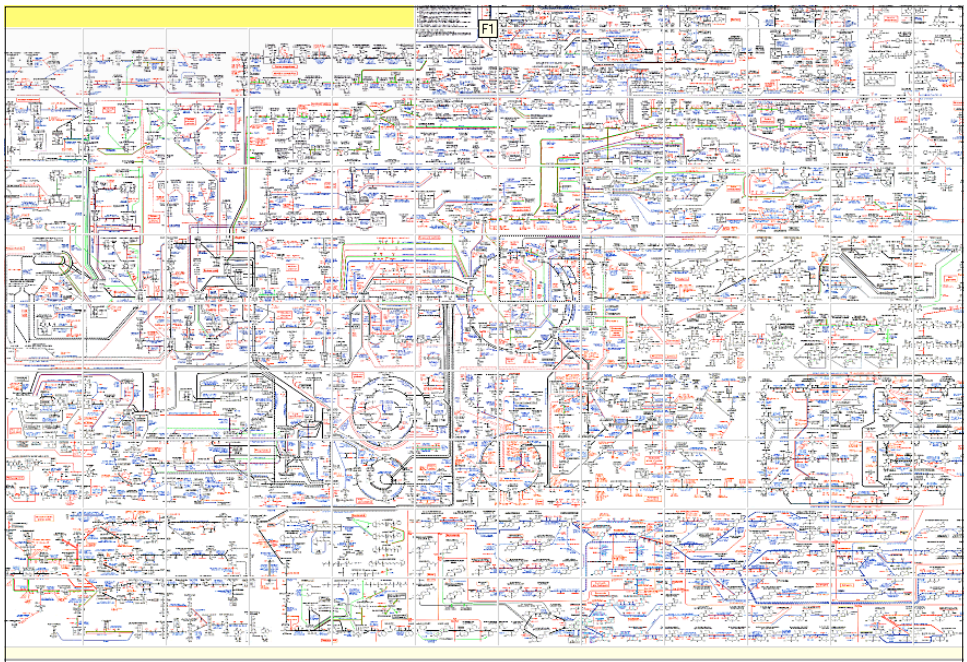
Research problems

1. Find motifs recognized by each DNA-binding protein.
2. Find genes regulated by these proteins.
3. Use networks to predict function of specific genes.
4. Characterize statistical properties of the network.
5. Compare/align networks.

Biochemical reactions

- Biochemical, metabolic reactions
  - "Chemical engine"
  - Determines cell physiology
  - Similar in all organisms.
- KEGG
- EcoCyc
- Metabolic Fluxes

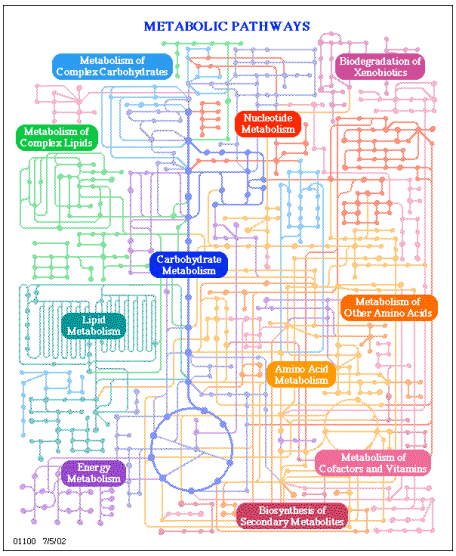




the other chart : Cellular and Molecular Processes

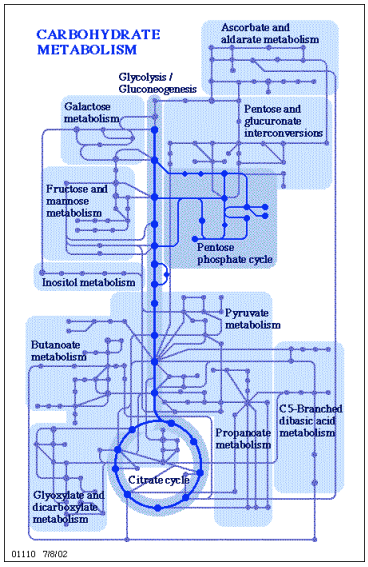
Metabolic Pathways

KEGG database

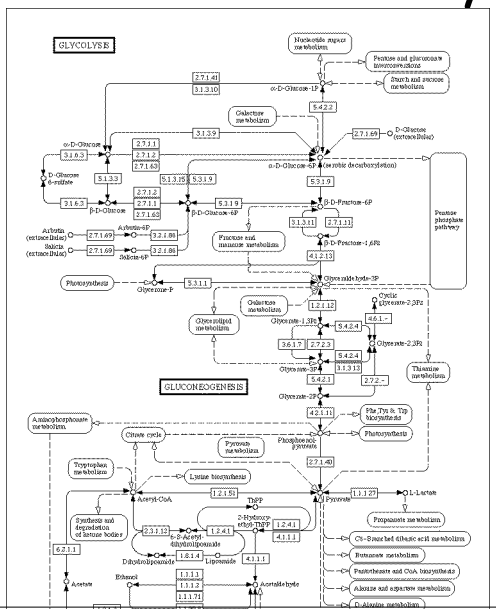


Metabolic Pathways

KEGG database

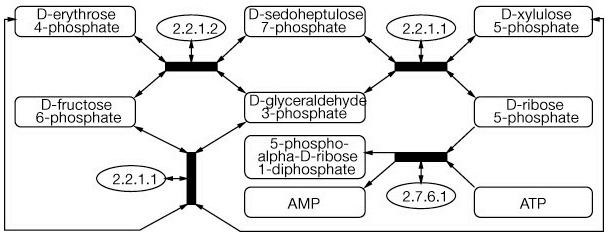


Metabolic Pathways



KEGG database

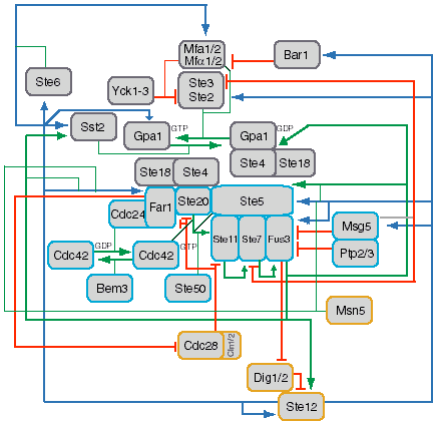
Metabolic Pathways



KEGG database

Examples of biological networks

- Regulation

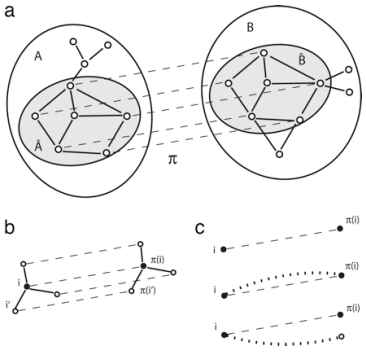


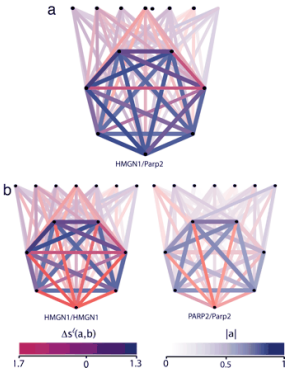
Drew Endy (MIT)

Outline

- Measurements
- Analysis
  - + Expression
  - + Interactions/Reactions

Alignment of networks





Cross-species analysis of biological networks by Bayesian alignment

Johannes Berg† and Michael Lässig

Institut für Theoretische Physik, Universität zu Köln, Zülpicherstrasse 77, 50937 Cologne, Germany

Clusters in PPI networks

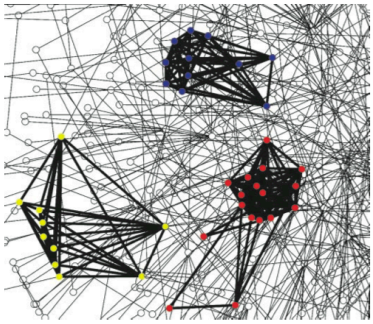


Fig. 2. Fragment of the protein network. Nodes and interactions in discovered clusters are shown in bold. Nodes are colored by functional categories in

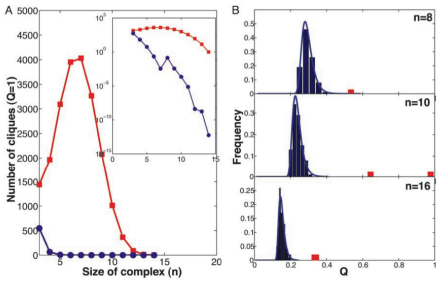


Fig. 1. Statistical significance of complexes and modules. (*A*) Number of complete cliques ($Q = 1$) as a function of clique size enumerated in the network of protein interactions (red) and in randomly rewired graphs (blue, averaged >1,000 graphs). *Inset* shows the same plot in log-normal scale. Note the dramatic enrichment in the number of cliques in the protein-interaction graph. Most of these cliques are parts of bigger complexes and modules. (*B*) Distribution of Q of clusters found by the MC search procedure in the randomly rewired graphs (blue bars). The blue line shows approximation of this distribution by the Fisher–Tippett extreme value distribution (EVD) with two fitted parameters. Red bars show complexes found in the original network of protein interactions. Sizes of the subgraphs are $n = 8$, 10, and 16. Note that real complexes have many more interactions than the tightest complexes found in randomly rewired graphs.

Protein complexes and functional modules in molecular networks

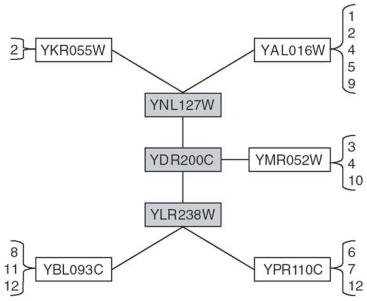
Victor Spirin and Leonid A. Mirny*

Harvard–MIT Division of Health Sciences and Technology, 16-343, Massachusetts Institute of Technology, 77 Massachusetts Avenue,

Nature Biotechnology **21**, 697 - 700 (2003)
Published online: 12 May 2003; | doi:10.1038/nbt825

Global protein function prediction from protein-protein interaction networks

Alexei Vazquez, Alessandro Flammini, Amos Maritan & Alessandro Vespignani



$$E = -\Sigma_{i,j} J_{ij}\, \delta(\sigma_i,\sigma_j) - \Sigma_i h_i(\sigma_i)$$

Clustering

STATEMENT OF THE PROBLEM

GIVEN DATA POINTS $\mathbf{X}_i$, i=1,2,...N, EMBEDDED IN D - DIMENSIONAL SPACE, IDENTIFY THE UNDERLYING STRUCTURE OF THE DATA.

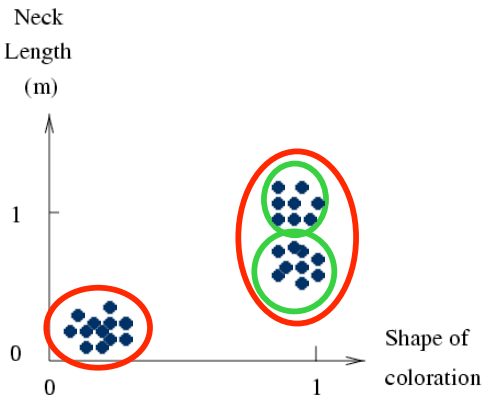
AIMS: PARTITION THE DATA INTO M CLUSTERS, POINTS OF SAME CLUSTER - "MORE SIMILAR"

- M ALSO TO BE DETERMINED!
- GENERATE DENDROGRAM,
- IDENTIFY SIGNIFICANT, "STABLE" CLUSTERS

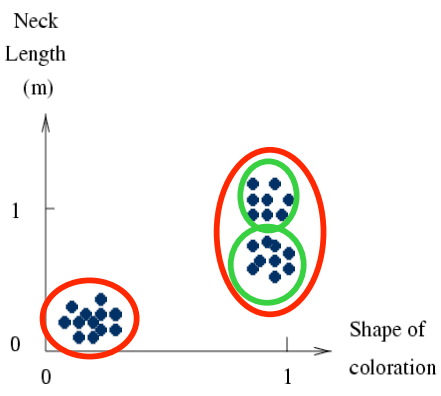
"ILL POSED":
- WHAT IS "MORE SIMILAR"?
- RESOLUTION

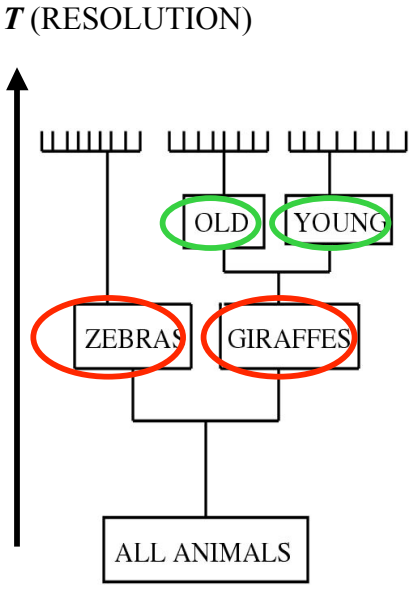
Eytan Domany@Weizmann

DEFINITION OF THE CLUSTERING PROBLEM

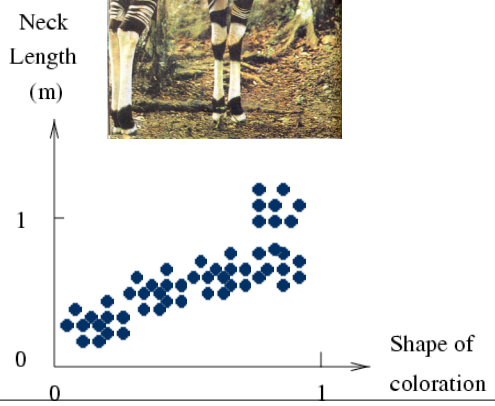


CLUSTER ANALYSIS YIELDS DENDROGRAM

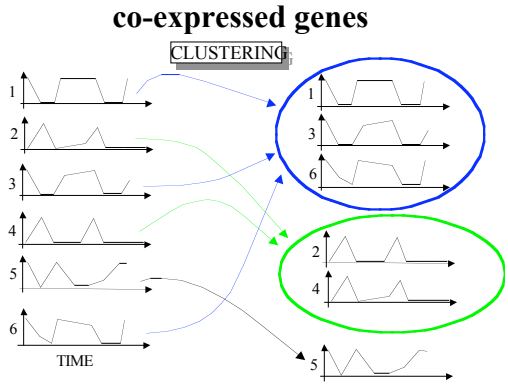




BUT WHAT ABOUT THE OKAPI?

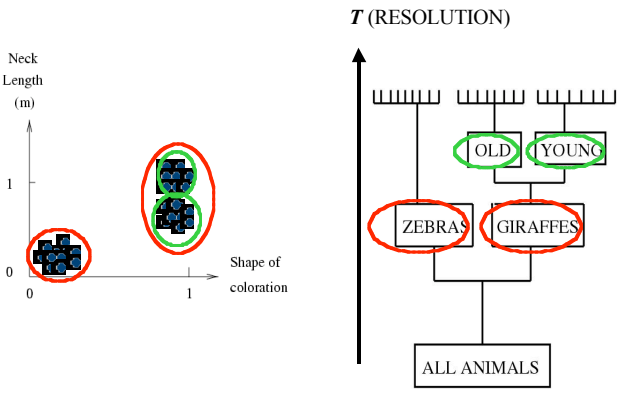


Gene expression



Clustering

CLUSTER ANALYSIS YIELDS DENDROGRAM



Eytan Domany@Weizmann

CLUSTERING METHODS

- AGGLOMERATIVE HIERARCHICAL
  - –AVERAGE LINKAGE (GENES: EISEN *ET. AL.,* PNAS 1998)
- CENTROID (REPRESENTATIVE)
  - –SELF ORGANIZED MAPS (KOHONEN 1997;
    (GENES: GOLUB *ET. AL.,* SCIENCE 1999)
  - --K-MEANS (GENES; TAMAYO *ET. AL.,* PNAS 1999)
- PHYSICALLY MOTIVATED
  - –DETERMINISTIC ANNEALING (ROSE *ET. AL.*,PRL 1990;
    GENES: ALON *ET. AL.*, PNAS 1999)
  - –SUPER-PARAMAGNETIC CLUSTERING (SPC)(BLATT *ET.AL.*
    GENES: GETZ *ET. AL.*, PHYSICA 2000,PNAS 2000)

Hierarchical (bottom-up) clustering

- Hierarchical agglomerative clustering:
  we sequentially merge the pair of "closest" points/clusters
- The procedure

1. Find two closest points (clusters) and merge them
2. Replace clusters with pseudo-points.
3. Proceed until we have a single cluster (all the points)

- Two prerequisites:

1. distance measure between two points
2. distance measure between clusters and way of replacing clusters with points(cluster linkage)

- No notion of optimality, greedy algorithm

Agglomerative Hierarchical Clustering

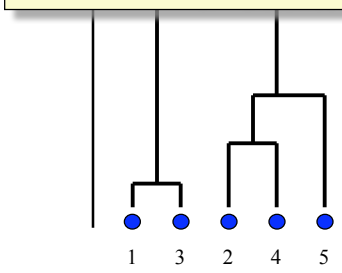
Need to define the distance between the new cluster and the other clusters.

Single Linkage: distance between closest pair.

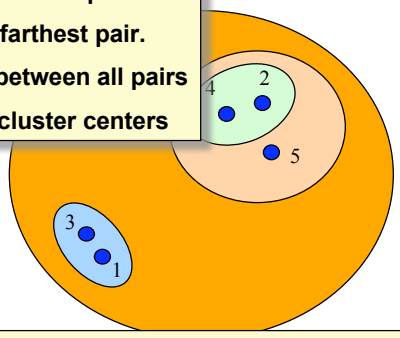
Complete Linkage: distance between farthest pair.

Average Linkage: average distance between all pairs

or distance between cluster centers



Dendrogram



The dendrogram induces a linear ordering of the data points

Clustering

a) Single linkage

$$d_{kl} = \min_{i \in C_k, j \in C_l} d(\mathbf{x}_i, \mathbf{x}_j)$$

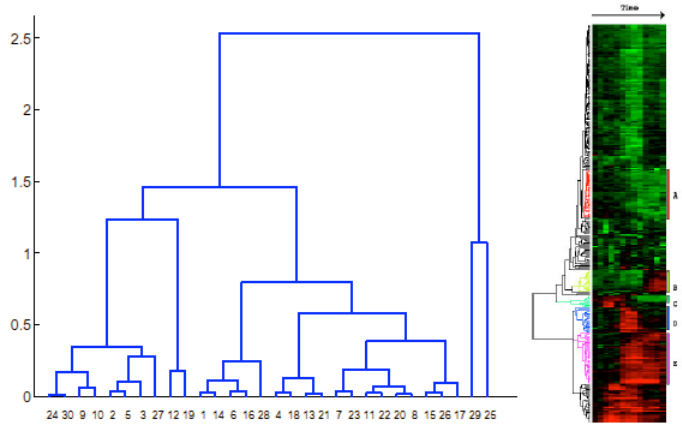
b) Average linkage

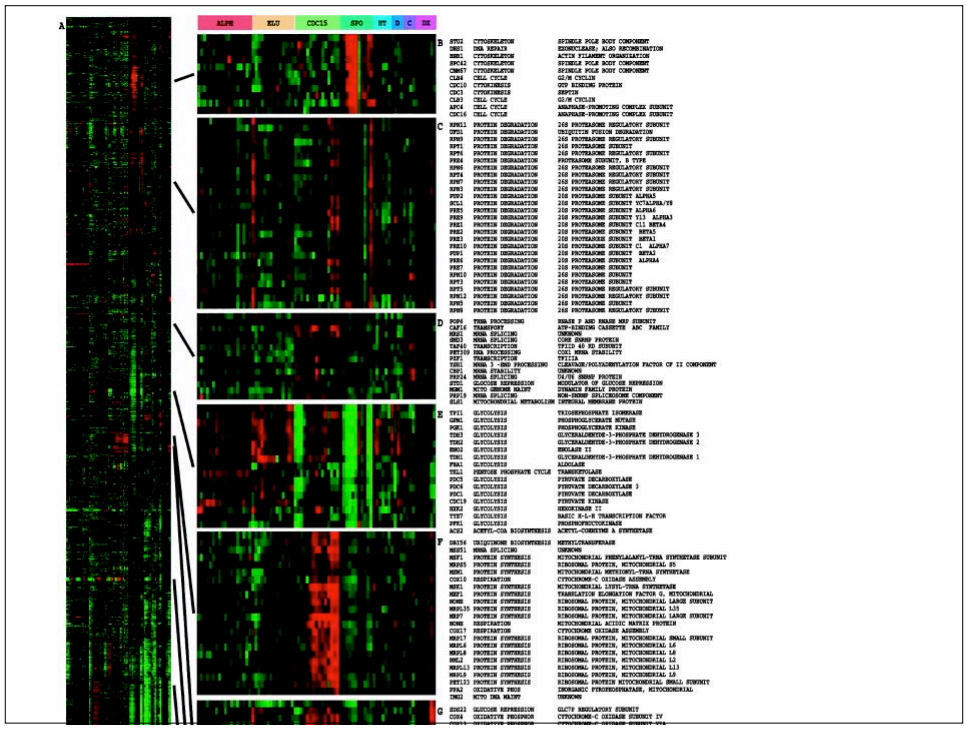
$$d_{kl} = \frac{1}{|C_l|\,|C_k|} \sum_{i \in C_k, j \in C_l} d(\mathbf{x}_i, \mathbf{x}_j)$$

c) Centroid linkage

$$d_{kl} = d(\bar{\mathbf{x}}_k, \bar{\mathbf{x}}_l), \quad \bar{\mathbf{x}}_l = \frac{1}{|C_l|} \sum_{i \in C_l} \mathbf{x}_i$$

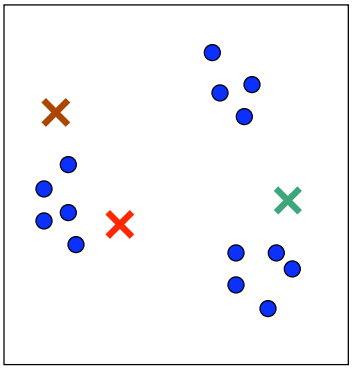
Clustering





K-means

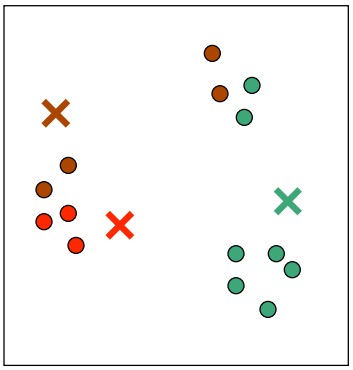
•Start with random positions of centroids.



Iteration = 0

K-means

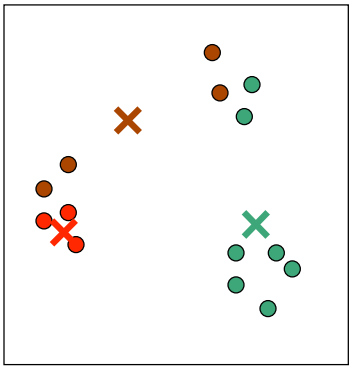
- Start with random positions of centroids.
- Assign data points to centroids: find closest centroid for each point



Iteration = 1

K-means

- Start with random positions of centroids.
- Assign data points to centroids: find closest centroid for each point
- Move centroids to center of assigned points



Iteration = 1

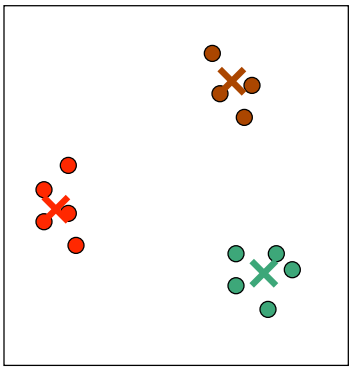
K-means

•Start with random positions of centroids.

•Assign data points to centroids: find closest centroid for each point

•Move centroids to center of assigned points

•Iterate till minimal cost: sum of distances to centroids



Iteration = 3

K-means - Summary

- Result depends on initial centroids' position
- Fast algorithm: compute distances from data points to centroids

- Must preset K
- Clusters are convex and comact
- Fails for non-spherical distributions

Super-paramagnetic clustering (SPC)

- Potts model
  - a spin in each node
  - connected spins inte

$$s_i = 1, \ldots q$$

$$J_{ij} > 0$$

$$\mathcal{H}(\mathcal{S}) = \sum_{\langle ij \rangle} J_{ij} \left(1 - \delta_{s_i, s_j}\right)$$

- Order Parameter

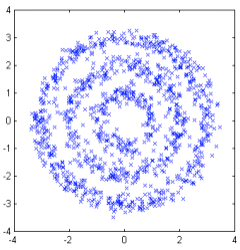
$$N_{\max}(\mathcal{S}) = \max\left\{N_1(\mathcal{S}),\ N_2(\mathcal{S}), \ldots\ N_q(\mathcal{S})\right\}$$

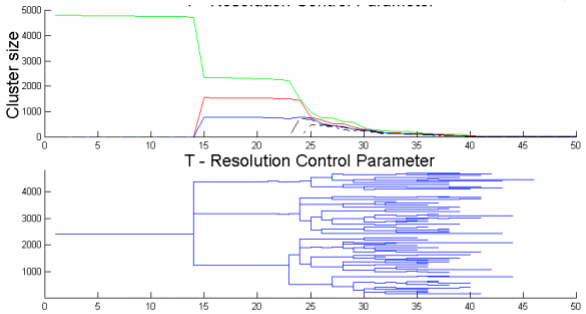
$$m(\mathcal{S}) = \frac{q\, N_{\max}(\mathcal{S}) - N}{(q-1)\, N}$$

Marcelo Blatt, Shai Wiseman, Eytan Domany 1997

3circles:

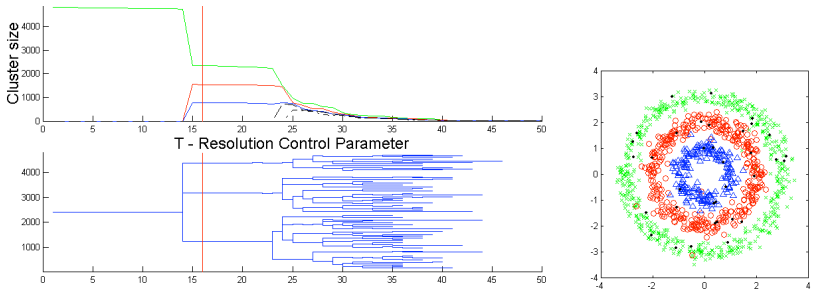
N=4800 POINTS IN D=2



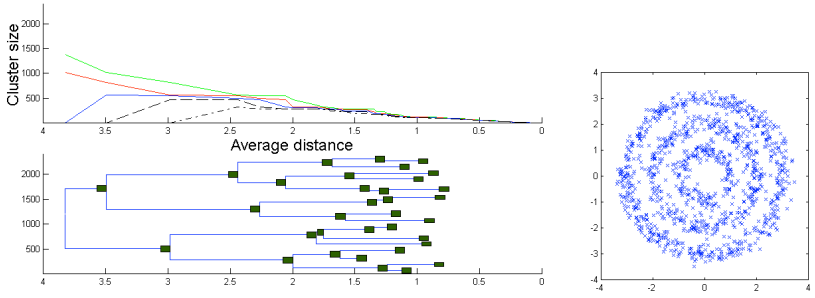


Stable clusters
"live" for large ΔT

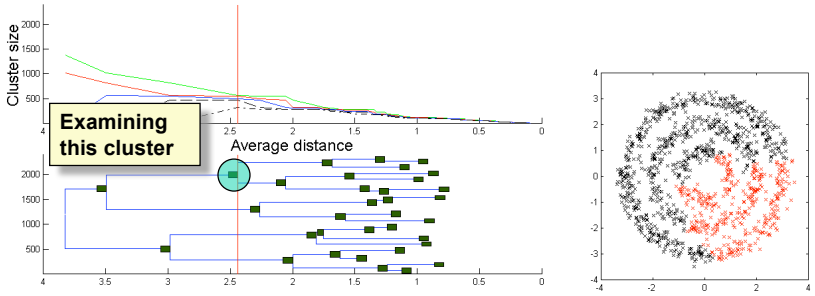
identifying stable clusters



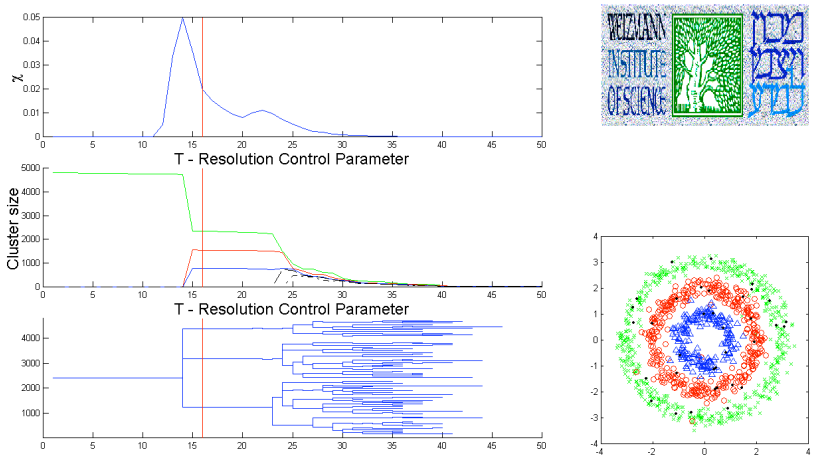
Same data - Average Linkage



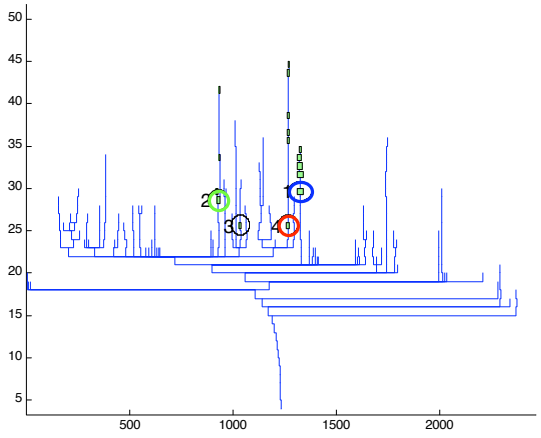
Same data - Average Linkage



Choosing a value for T

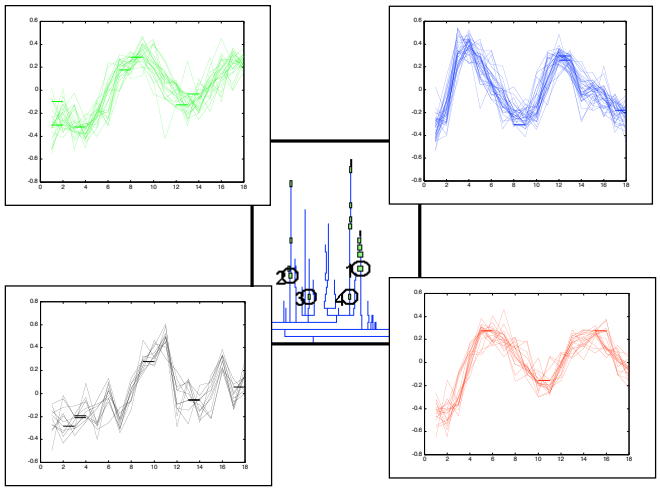


Chosen clusters



Cell-cycle clusters

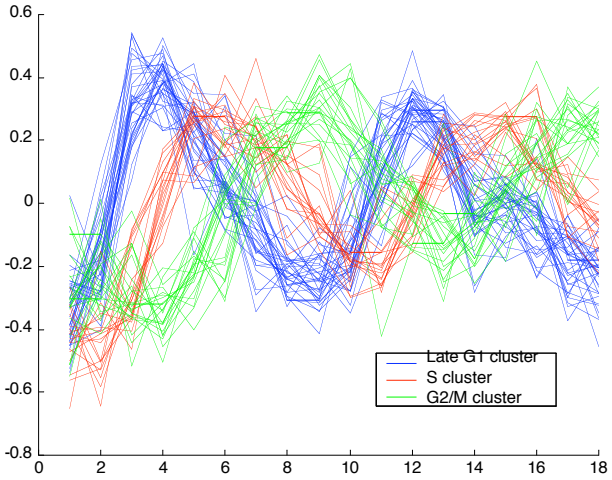
G2/M
Clb1,2
Swi5
Ace2



Late G1
Cln1,2
Clb5,6
Swi4

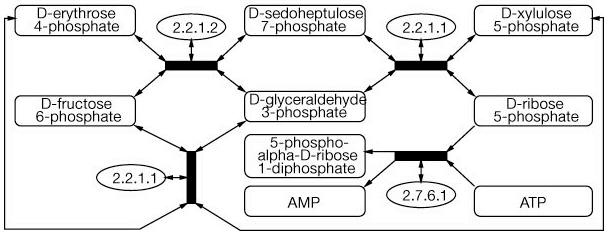
S
Histones

Progression of the cell-cycle



Statistical properties of biological networks

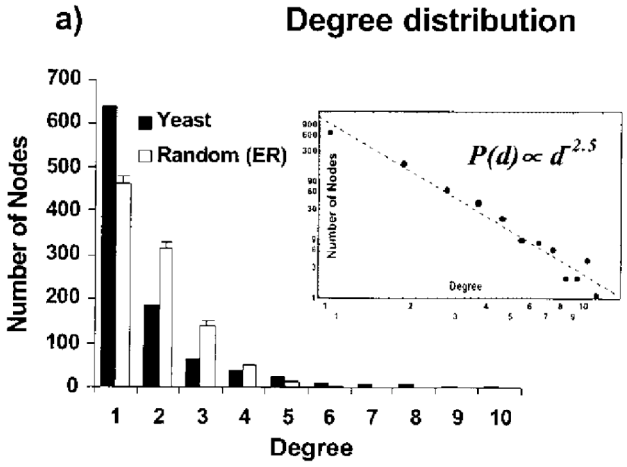
Metabolic Pathways



KEGG
database

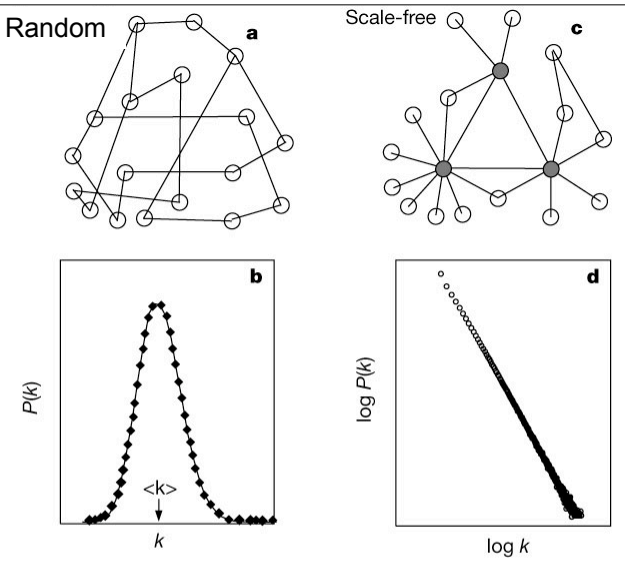
it's not a random graph!

Andreas Wagner
Mol. Biol. Evol. 18(7):1283–1292. 2001



Random vs power-law

Barabasi A et.al. Nature:411(2001)
Wagner A Mol Biol Evol:18(2001)



Other power-law networks

Barabasi A et.al. Nature:411(2001)

Other power-law networks:

- Metabolic network
- Network of social interactions:
  scientific collaborations, actors in films
- The Internet:
  links, physical connections

it's not a random graph!

Table 1
Comparison of Statistical Features Between Random Graphs and the Yeast Protein Interaction Network

		Random Graphs	
	Yeast	ER	PL ($\tau = 2.5$)
Whole graph			
Nodes	985	984.02 (10.39)	970.7 (81.57)
Degree	1.83	1.85 (0.98)	1.64 (1.76)
No. of components	163	108 (8)*	266.3 (30.6)*
Giant component			
Nodes	466	624.0 (38.7)*	336.9 (86)
Degree	2.3	2.07 (1.05)	2.50 (2.6)
Clustering coefficient ($\times 10^{-3}$)	22	0.59 (0.9)*	4.02 (2.3)*
Characteristic path length	7.14	15.88 (1.76)*	6.01 (1.14)

Wagner MBE 2000

Random vs power-law

Barabasi A et.al. Nature:411(2001)
Wagner A Mol Biol Evol:18(2001)

The network of protein-protein interactions (and other molecular biological networks) are power-law networks!

WHY?

- Power law networks are "better"…

  OR/AND

- Biological networks became power-law due to evolution.

Random

- Removal of a randomly picked node significantly increases the average path.
- All nodes are of equal "importance".

Power-law

- Removal of a random node slightly increases the average path.
- Removal of a highly-connected node leads to drastic increase of the average path!

POWER-LAW N[illegible]

- [illegible]ar[illegible] to random "attacks",
- [illegible] more sensitive to targeted attacks!

Let's check it

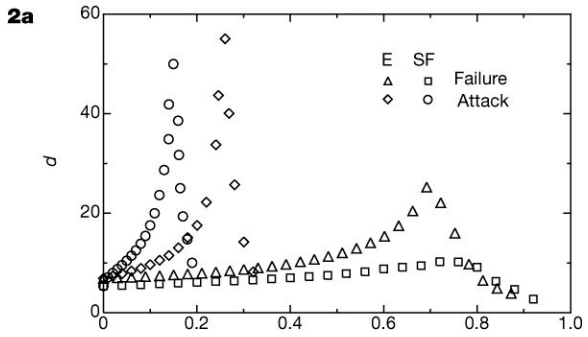


Figure 2 Changes in the diameter d of the network as a function of the fraction f of the removed nodes. **a**, Comparison between the exponential (E) and scale-free (SF) network models, each containing $N = 10{,}000$ nodes and 20,000 links (that is, $k = 4$).

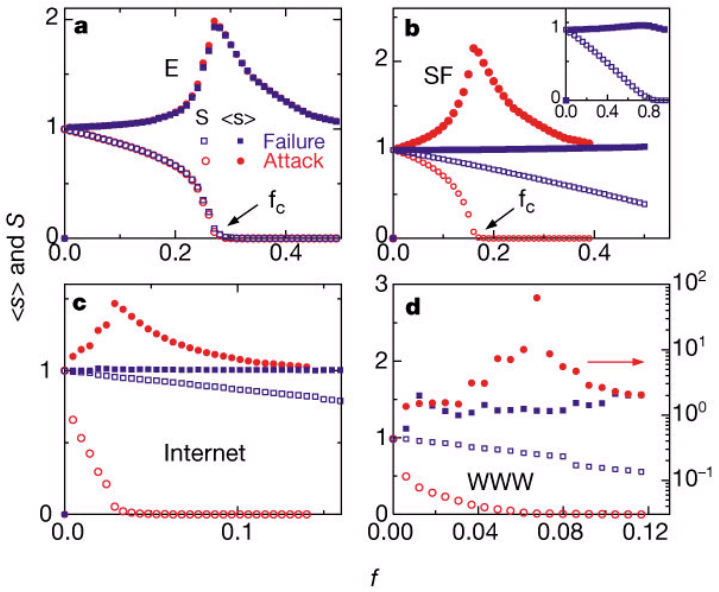


Figure 3 Network fragmentation under random failures and attacks. The relative size of the largest cluster S (open symbols) and the average size of the isolated clusters s (filled symbols) as a function of the fraction of removed nodes f for the same systems as in Fig. 2. The size S is defined as the fraction of nodes contained in the largest cluster (that is, $S = 1$

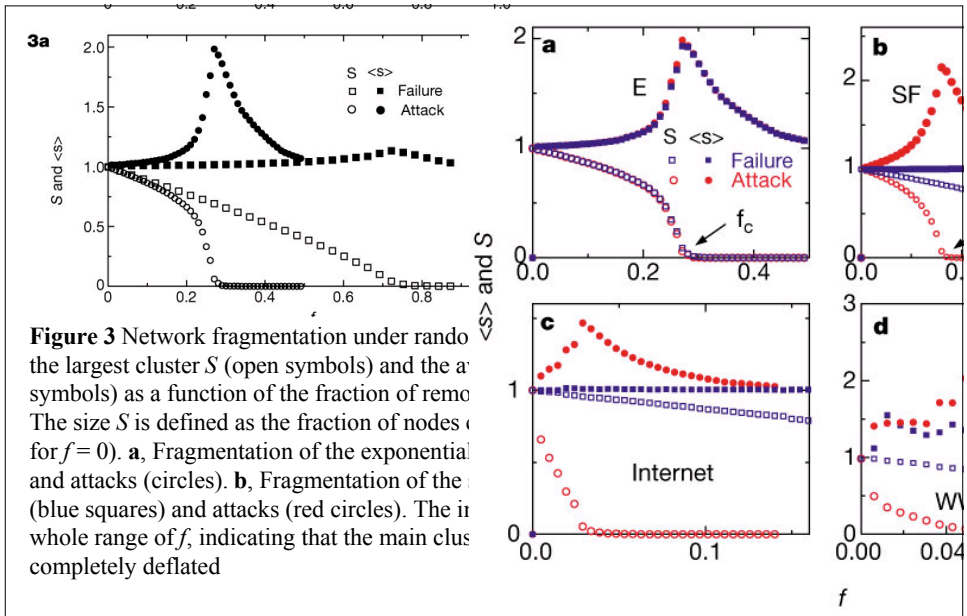


Figure 3 Network fragmentation under rando
the largest cluster S (open symbols) and the a
symbols) as a function of the fraction of remo
The size S is defined as the fraction of nodes
for $f = 0$). **a**, Fragmentation of the exponential
and attacks (circles). **b**, Fragmentation of the
(blue squares) and attacks (red circles). The i
whole range of f, indicating that the main clus
completely deflated

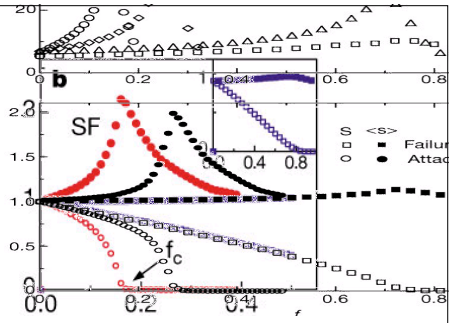


Figure 3 Network fragmentation under random failures and attacks. The relative size of the largest cluster S (open symbols) and the average size of the isolated clusters s (filled symbols) as a function of the fraction of removed nodes f for the same systems as in Fig. 2. The size S is defined as the fraction of nodes contained in the largest cluster (that is, $S = 1$ for $f = 0$). **a**, Fragmentation of the exponential network under random failures (squares) and attacks (circles). **b**, Fragmentation of the scale-free network under random failures (blue squares) and attacks (red circles). The inset shows the error tolerance curves for the whole range of f, indicating that the main cluster falls apart only after it has been completely deflated

Random | Power-law

Equally stable to random failures

More sensitive to attacks

POWER-LAW NETWORKS

- Tolerant to random "attacks",
- But more sensitive to targeted attacks!

Evolution of power-law graphs

1. Growth
2. Preferential attachment

~~Albert and Barabasi 2000~~

~~Herbert A. Simon 1955~~

Yule 1925

Evolution of power-law graphs

1. Growth
2. Preferential attachment

Alberts and Barabasi 2000

Herbert A. Simon 1967

Evolution of graphs

- Growth
  1. start with m_0 nodes
  2. add a node with m edges
  3. connect these edges to existing nodes

  at timestep t : $t+m_0$ nodes, tm edges

Evolution of graphs

- Preferential attachment

  Probability Π of connection to node i depends on the degree k_i of this node.

  E.g. $$\Pi(k_i) = \frac{k_i}{\sum_j k_j}$$

  "Rich gets richer"

Yule model

Growth of biological genera (families)

1. New species evolve at a constant rate
2. Out of new m species, one diverges to form a new family

equivalent to:

Measure time in the number of families

At each time step:

1. a new family is created.
2. m species are placed in existing families with prob. ~ to the number of species in each family.

Yule model

Measure time in the number of families

At each time step:

1. a new family is created.
2. m species are placed in existing families with prob. ~ to the number of species in each family.

$$p_k = \frac{k-1}{k+1+1/m} p_{k-1}$$

$$p_k \to k^{-\alpha} \quad \alpha = 2 + 1/m$$

[Alberts and Barabasi, 2000] = [Yule, 1925 for m=1]

Better evolution of graphs

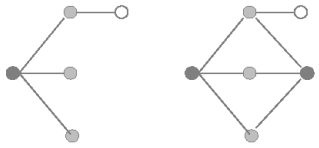
A. Wagner, M.Lassig, A.Maritan etc

- Gene duplication
- Mutations
- Preferential attachment

More biological neutral evolution of graphs

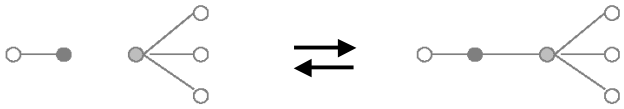
A.Wagner, M.Lassig, A.Maritan, S.Redner etc.

- Gene duplication



- Mutations

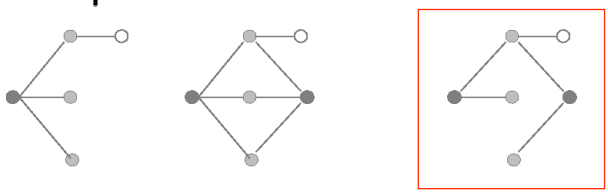
a)



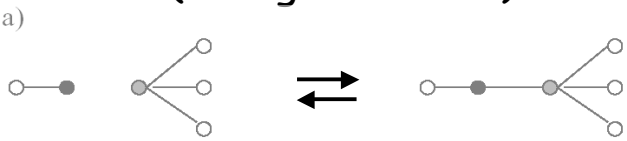
More biological neutral evolution of graphs

A.Wagner, M.Lassig, A.Maritan, S.Redner etc.

- Gene duplication



- Mutations (rich gets richer)



=> Broad (not power-law) distribution!

More biological neutral evolution of graphs

- Gene duplication and re-wiring

Infinite-Order Percolation and Giant Fluctuations in a Protein Interaction Network

J. Kim[1], P. L. Krapivsky[2], B. Kahng[1], and S. Redner[2]

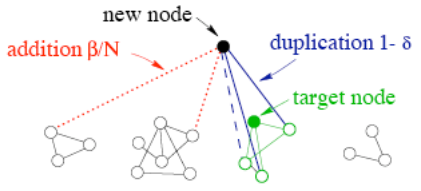
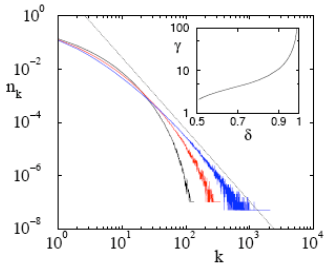


FIG. 1. Growth steps of the protein interaction network: The new node duplicates 2 out of the 3 links between the target node (shaded) and its neighbors. Each successful duplication occurs with probability $1-\delta$ (solid lines). The new node also attaches to any other network node with probability β/N (dotted lines). Thus 3 previously disconnected clusters are joined by the complete event.



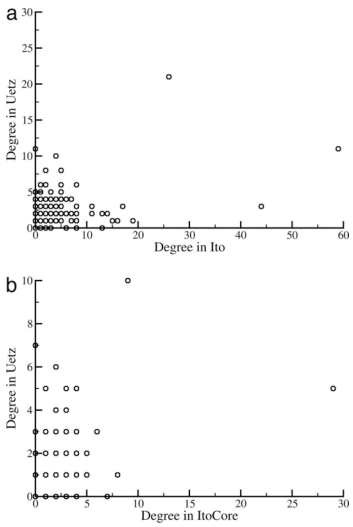
A biophysical model of apparent power-law

A simple physical model for scaling in protein–protein interaction networks

PNAS | January 10, 2006

Eric J. Deeds*, Orr Ashenberg†, and Eugene I. Shakhnovich‡§

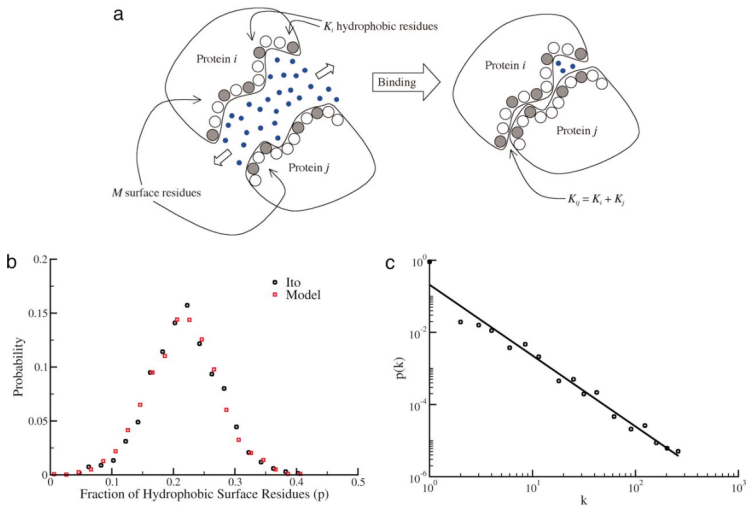
Correlation between PPI networks



Deeds E. J. et.al. PNAS 2006;103:311-316

PNAS

A physical model for PPI measurements



Deeds E. J. et.al. PNAS 2006;103:311-316

PNAS

A biophysical model of apparent power-law

A simple physical model for scaling in protein–protein interaction networks

PNAS | January 10, 2006

Eric J. Deeds*, Orr Ashenberg†, and Eugene I. Shakhnovich‡§

Binding constant $Kd=\exp(-\Delta G/kT)$

Bound in experiment: $Kd<Kc$

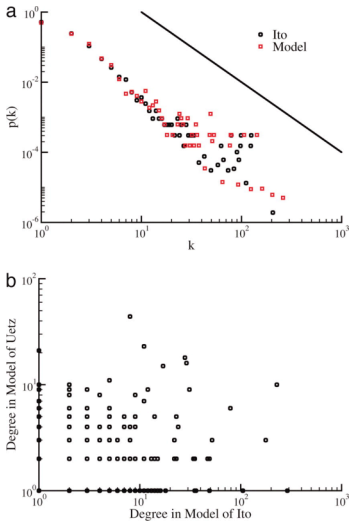
If ΔG is distributed normally

$$P(\Delta G) \sim \exp(-(\Delta G-\langle\Delta G\rangle)^2/(2s^2))$$

then

$$P(Kd) \sim \exp((\log(Kd)-\mu)^2/(2s^2))$$ <-- looks like power-law!

Degree distributions and correlations for model PPI networks



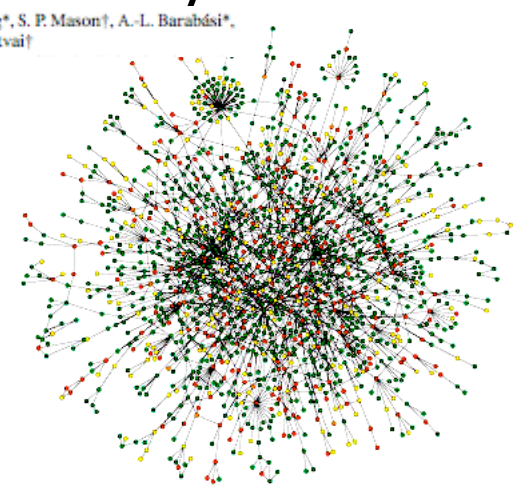
Deeds E. J. et.al. PNAS 2006;103:311-316

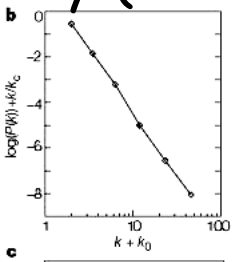
PNAS

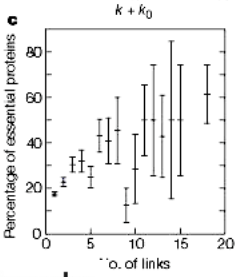
Connecting to biology

Lethality and centrality (2001)

H. Jeong*, S. P. Mason†, A.-L. Barabási*,
Z. N. Oltvai†



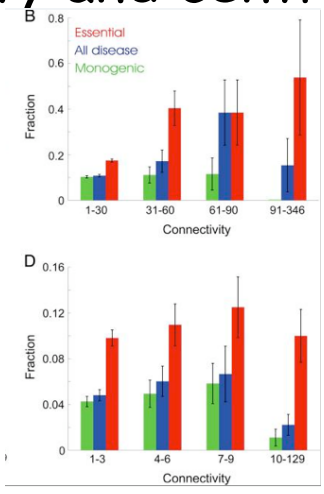




Lethality and centrality in protein networks

The most highly connected proteins in the cell are the most important for its survival.

Lethality and centrality (2008)



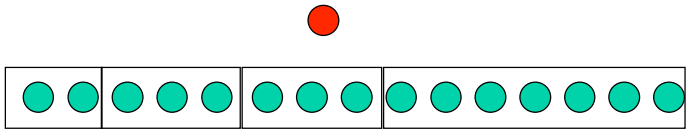
Network properties of genes harboring inherited disease mutations

Igor Feldman*, Andrey Rzhetsky*†‡, and Dennis Vitkup*‡

PNAS | March 18, 2008

Generalization of evolution by duplication and attachment

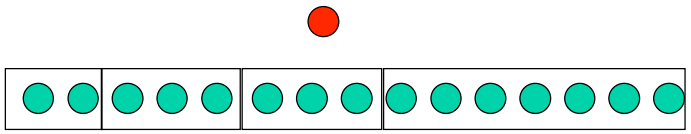
For fixed parameters, $\gamma \in \mathbf{R}$, $0 \leq p < 1$ and a positive integer $k > 1$, begin with k bins, each containing one ball and then introduce balls one at a time. For each new ball, with probability p, create a new bin and place the ball in that bin; with probability $1 - p$, place the ball in an existing bin, such that the probability the ball is placed in a bin is proportional to m^γ, where m is the number of balls in that bin.



Generalization of evolution by duplication and attachment

	Finite Polya process $p = 0$	Infinite Polya process $0 < p < 1$	
$\gamma > 1$	one bin dominates	one bin dominates	
$\gamma = 1$	Polya's urn problem	power law distribution	$f_i \propto i^{(-1+1/(1-p))}$
$0 < \gamma < 1$	all bins grow at the same rate asymptotically	exponentially decreasing assuming $(*)$	$f_i \propto i^{-\gamma} e^{-K i^{1-\gamma}/(1-\gamma)}$
$\gamma = 0$			$f_i \propto (K+1)^{-i}$
$\gamma < 0$			$f_i = O\left(((i-1)!)^\gamma / K^i\right)$

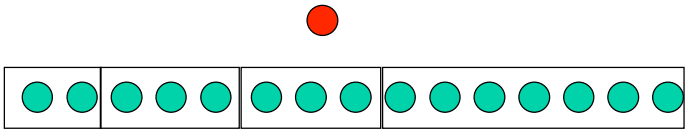
TABLE 1. **The distribution of bin sizes.**
f_i is the limit of the fraction of bins with i balls and $K = \frac{p}{1-p} \sum_{i=1}^{\infty} f_i \, i^\gamma$.



Generalization of evolution by duplication and attachment

	Finite Polya process $p = 0$	Infinite Polya process $0 < p < 1$	
$\gamma > 1$	one bin dominates	one bin dominates	
$\gamma = 1$	Polya's urn problem	power law distribution	$f_i \propto i^{(-1+1/(1-p))}$
$0 < \gamma < 1$	all bins grow at the same rate asymptotically	exponentially decreasing assuming $(*)$	$f_i \propto i^{-\gamma} e^{-K i^{1-\gamma}/(1-\gamma)}$
$\gamma = 0$			$f_i \propto (K+1)^{-i}$
$\gamma < 0$			$f_i = O(((i-1)!)^{\gamma}/K^i)$

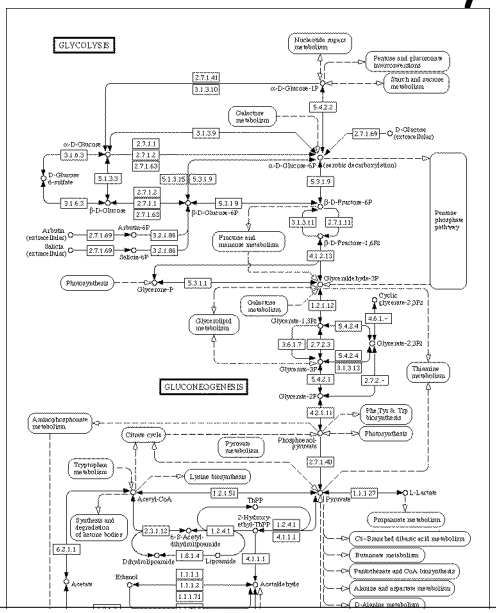
TABLE 1. **The distribution of bin sizes.**
f_i is the limit of the fraction of bins with i balls and $K = \frac{p}{1-p} \sum_{i=1}^{\infty} f_i \, i^{\gamma}$.



http://www.math.uah.edu/stat/applets/PolyaExperiment.xhtml

Metabolic Pathways

KEGG database



Flux Balance Analysis

No accumulation of intermediates

=

# of molecules in = # of molecules out

=

Vin+Vout = 0

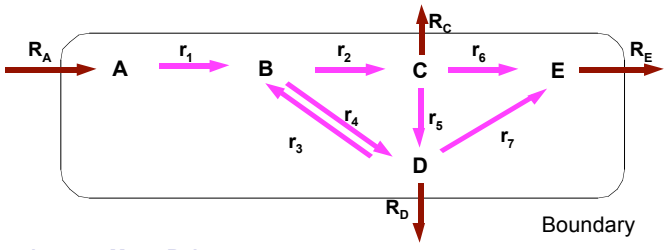
Example:

2A+B -> 3D

D+C->E

$(2V_A+V_B)/3 = V_E$

Flux Balance Analysis



Steady state Mass Balance

A:-r1=-R_A
B:-r1+r4-r2-r3=0
C:+r2-r5-r6=+R_C
D:+r3+r5-r4-r7=+R_D
E:+r6+r7=+R_E

S.v = b

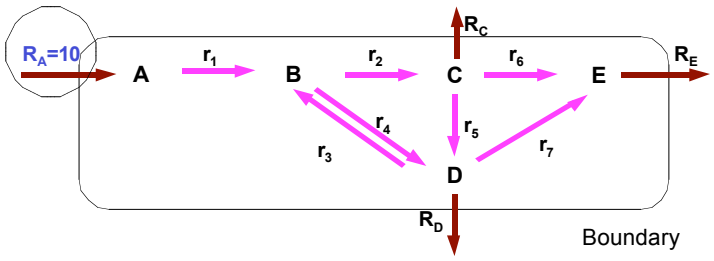
$$\begin{bmatrix} r_1 & r_2 & r_3 & r_4 & r_5 & r_6 & r_7 \end{bmatrix}$$

$$\begin{bmatrix} -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & -1 & -1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & -1 & -1 & 0 \\ 0 & 0 & 1 & -1 & 1 & 0 & -1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix} = \begin{bmatrix} -R_A \\ 0 \\ R_C \\ R_D \\ R_E \end{bmatrix}$$

Internal fluxes

Transportation fluxes

Flux Balance Analysis



Steady state Mass Balance

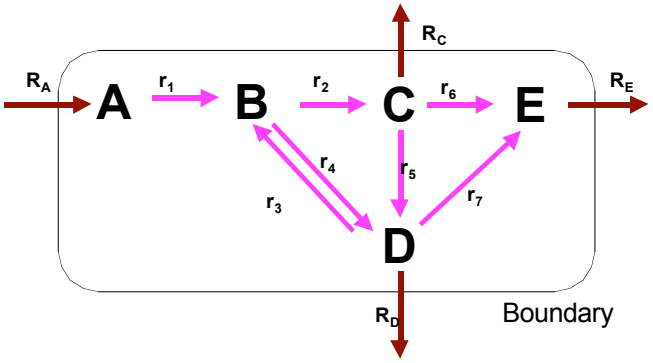
A:-r1=-10
B:-r1+r4-r2-r3=0
C:r2-r5-r6-R_C=0
D:r3+r5-r4-r7-R_D=0
E:r6+r7-R_E=0

$$\begin{bmatrix} r_1 & r_2 & r_3 & r_4 & r_5 & r_6 & r_7 & R_C & R_D & R_E \\ -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & -1 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & -1 & -1 & 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 & 1 & 0 & -1 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & -1 \end{bmatrix} = \begin{bmatrix} -10 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

10 Unknown fluxes

1 Known fluxes

Flux Balance Analysis



Total Number of fluxes = 11
Total numer of known flux = 1
Total number of Metabolites = 5
Total number of d.f = 11-1+5=5
(i.e 5 possible solutions for this reaction network)

Flux Balance Analysis

- If cells optimize their growth rate then we need to find a solution that maximizes growth.
- Growth = biomass/time

```
+1BIOM-0.582GLY-0.0485MethylTHF-0.25GLN-     45.135ATP+44.96ADP+ 44.96Pi    -
0.25GLU-0.176PHE-0.131TYR-0.205SER-0.054TRP-0.229ASP-0.229ASN-0.326LYS-
0.087CYS-0.146MET-0.241THR-0.276ILE-0.21PRO-0.281ARG-0.488ALA-0.402VAL-
0.428LEU-0.09HIS-0.203GTP-0.136UTP-0.126CTP-0.0247dATP-0.0254dGTP-
0.0254dCTP-0.0247dTTP-0.00258PS -0.09675PE-0.02322PG-0.00645CL-
0.00785LPS-0.0276Pept-0.0341PTRSC-0.007SPRMD-0.154Glycogen;
```

Millimoles of metabolites present in 1 gm (dry wt.) of biomass

Flux Balance Analysis

- Input: stoichiometric matrix
  optimization function (biomass)
- Calculations:

$$\frac{d\mathbf{X}}{dt} = \mathbf{S} \bullet \mathbf{v} - \mathbf{b} = 0$$

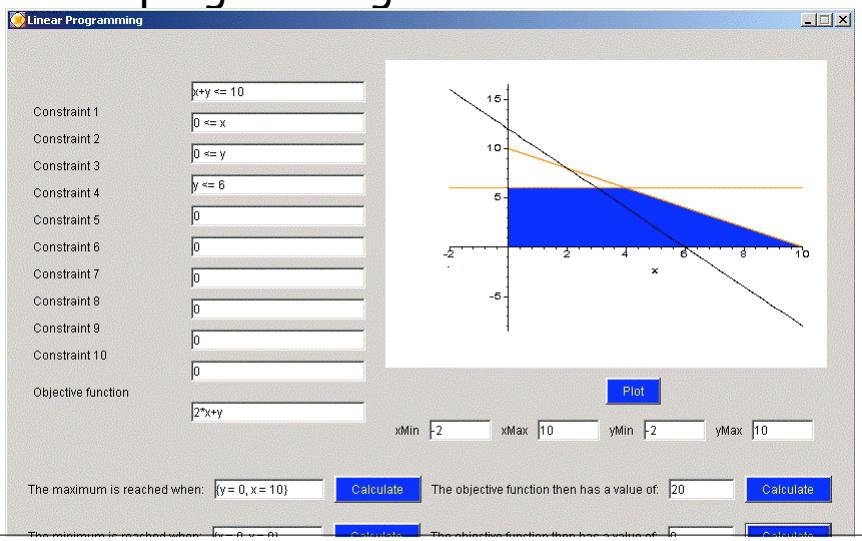
Maximize Z

$$Z = \sum c_i \cdot v_i = \mathbf{c} \bullet \mathbf{v}$$

- Output: fluxes, growth rate

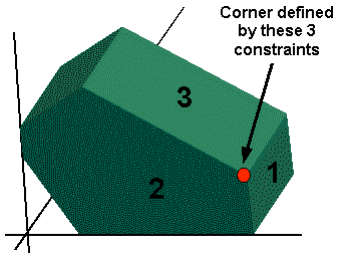
Flux Balance Analysis

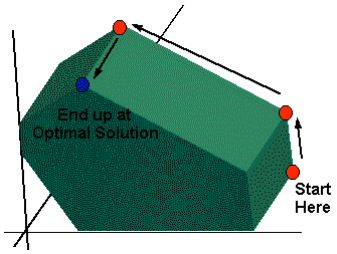
- Linear programming



Flux Balance Analysis

- Linear programming



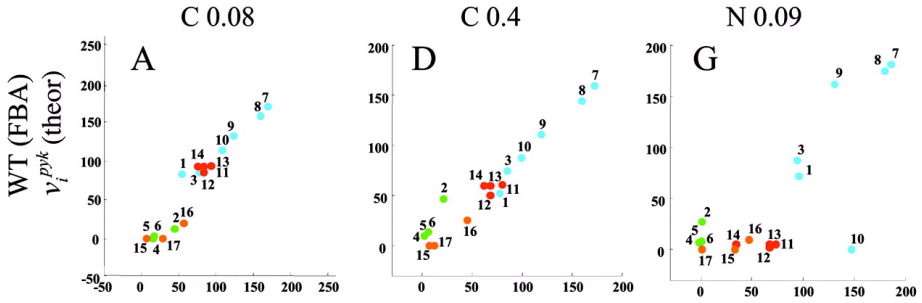


Flux Balance Analysis

- Effects of external conditions
- Effect of mutations
- Predictive cell physiology

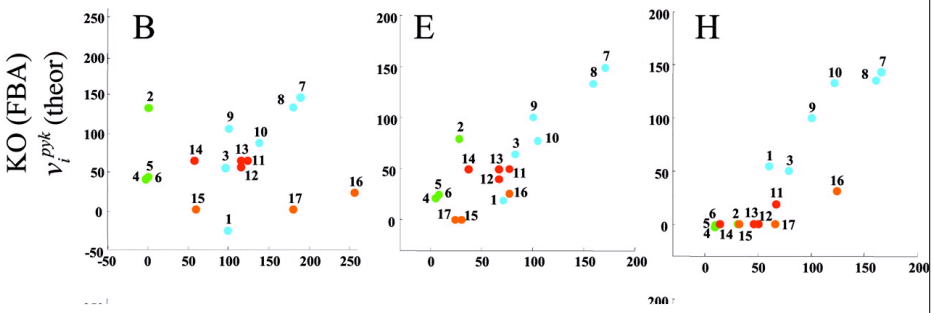
Flux Balance Analysis

- Effect of C and N starvation



Flux Balance Analysis

- Effect of mutations and starvation



Networks

- Structure and dynamics of some biological network can be studied experimentally
  (partially and with lots of mistakes!)

- Networks don't look like random graphs, more like power-law graphs.
  - results of neutral evolution
  - results of selection