1. Introduction
   - Gene regulation
   - Genomics and genome analyses
   - Hidden markov model (HMM)

2. Gene regulation tools and methods
   - Regulatory sequences and motif discovery
   - TF binding sites, microRNA target prediction

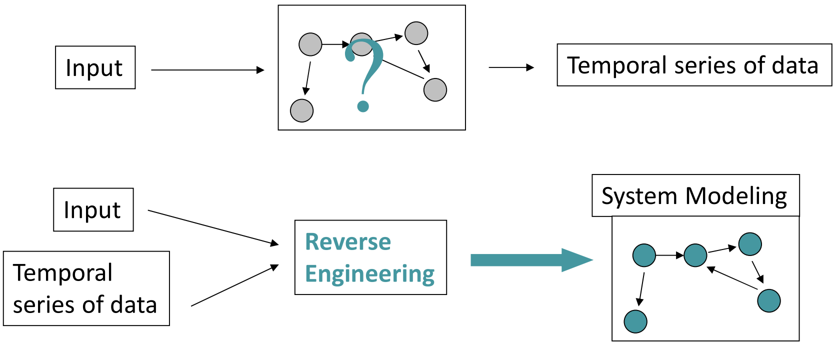
3. Technologies
   - Microarrays
   - Deep sequencing (RNAseq)
   - Single cell RNAseq spatial transcriptomics

4. Clustering
   - Unsupervized clustering (HCA,K-means, PCA, SOM)
   - Supervized clustering (classification)

5. Gene ontology, Pathways
   - Databases, tools
   - Over representation and enrichment analysis

6. Biomolecular networks
   - Network analysis and characteristics

Reverse Engineering



Input

Temporal series of data

Input

Temporal series of data

Reverse Engineering

System Modeling

Reverse Engineering

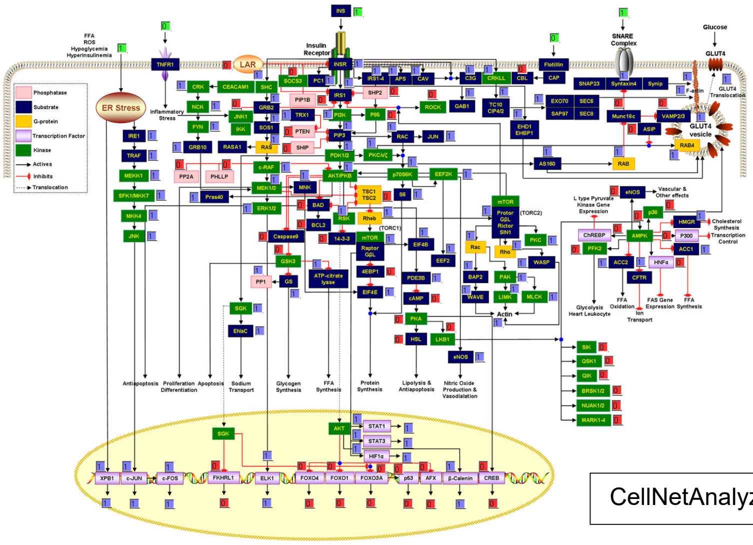
Considerations

- Inferential versus predictive power
- Parameters are static
- Genes and encoded proteins build a unit
- Instantenous versus synchronous model
- Model have to be stable

Solutions

- Boolean networks
- Differential equations
- Bayesian networks (conditional independence)

Boolean networks for sensitivity/robustness analysis



CellNetAnalyzer

Klemt A et al. BMC Systems Biology. 2007

Co-expression network analysis

Similarity measures

Correlation coefficient

$$S_{ij} = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sqrt{[\sum(x-\bar{x})^2][\sum(y-\bar{y})^2]}}$$

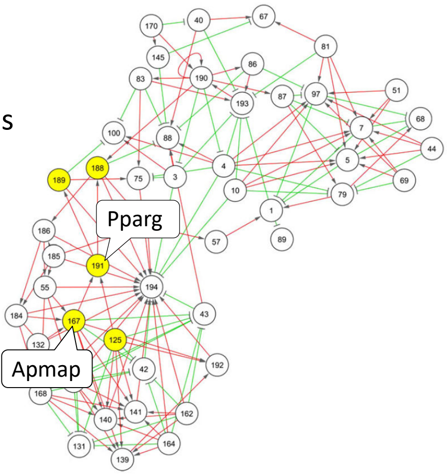
Mutual information

$$I(x, y) = S(x) + S(y) - S(x, y)$$

$$S(t) = -\sum_i p(t_i) \log p(t_i)$$

Gene association network

- MICO
- Discretizing expression profiles
  → groups of genes with identical profile
- REVEAL algorithm based on Mutual information
- Correlation

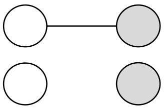


Bogner-Strauss et al. Cell Mol Life Sci. 2010

Adjacency function

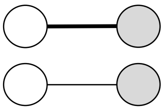
- unweighted

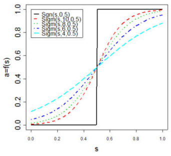
$$a_{ij} = \begin{cases} 1 & if\ s_{ij} \geq \tau \\ 0 & if\ s_{ij} < \tau \end{cases}$$



- weighted

$$a_{ij} = \frac{1}{1 + e^{-\alpha(s_{ij} - \tau_0)}}$$





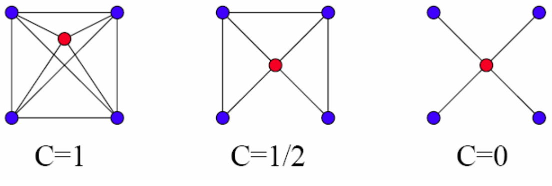
Network measures

Connectivity (degree)

$$k_i = \sum_{j=1}^{n} a_{ij}$$

Clustering coefficient

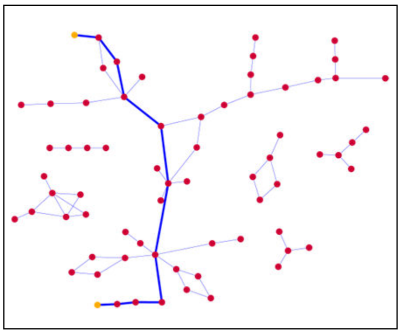
$$C_i = \frac{2n}{k_i\,(k_i - 1)}$$



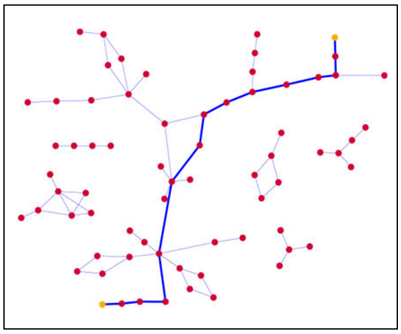
Path length and network diameter

A path is a sequence $\{x_1, x_2, \ldots, x_n\}$ such that $(x_1,x_2), (x_2,x_3), \ldots, (x_{n-1},x_n)$ are edges of the graph.

A closed path $x_n=x_1$ on a graph is called a graph cycle or circuit.

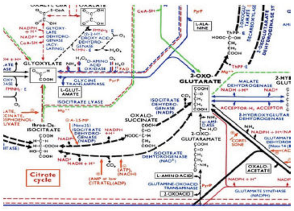


Shortest path between nodes



Longest shortest path

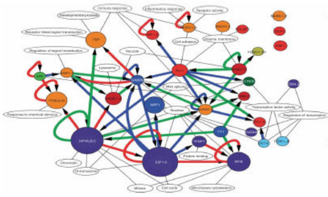
Different network representation



metabolic

Nodes metabolites and enzymes

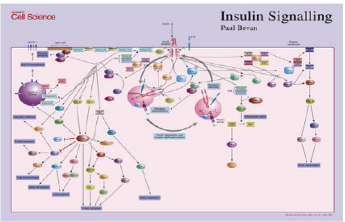
Edges chemical reactions



transcriptional

Nodes genes and transcription factors

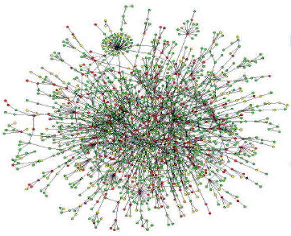
Edges Directed (Functional activation and inhibition)



signaling

Nodes proteins

Edges 'physical' interactions (binding and post-translational modifications)

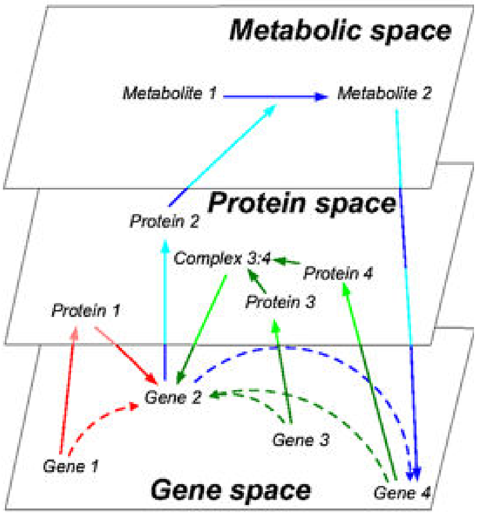


protein-protein Interaction PPI

Nodes proteins

Edges 'physical' interactions

Different levels of networks



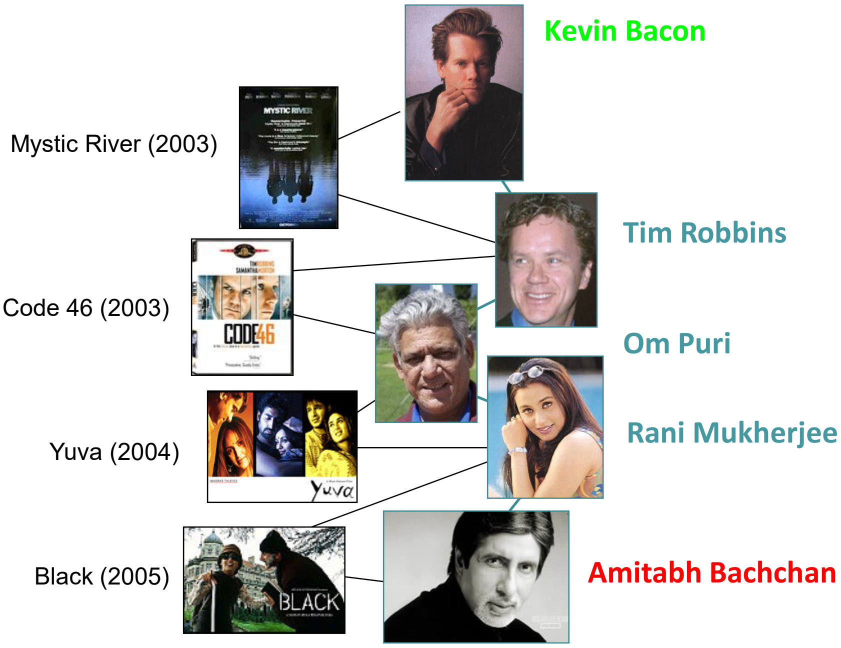
Small-world network

- Every node can be reached from every other by a small number of hops or steps
- High clustering coefficient and low mean-shortest path length (random graphs don’t necessarily have high clustering coefficients)
- Social networks, the Internet, and biological networks all exhibit small-world network characteristics
- Six degrees of separation

Six degrees of separation

- Pilgrim Experiment
  Random people from Nebraska were to send a letter (via intermediaries) to a stock broker in Boston. Could only send to someone with whom they were on a first-name basis
- Kevin Bacon Game
  Connect any actor to Kevin Bacon, by linking actors who have acted in the same movie
- Erdös Number
  Number of links required to connect scholars to Erdős, via co-authorship of papers

Kevin Bacon Game


Kevin Bacon

Mystic River (2003)

Tim Robbins

Code 46 (2003)

Om Puri

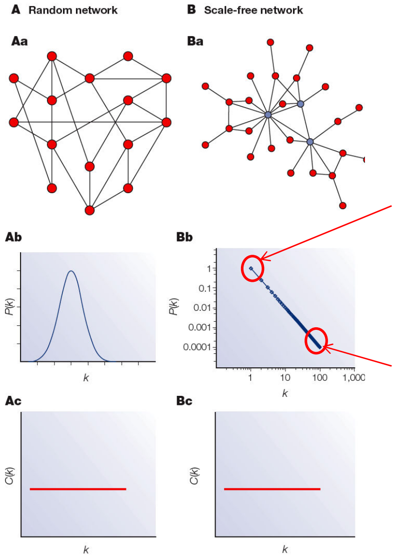
Rani Mukherjee

Yuva (2004)

Black (2005)

Amitabh Bachchan

Scale-free network follows a power law



Probability distribution

clustering coefficient

connectivity

Many nodes with few connections

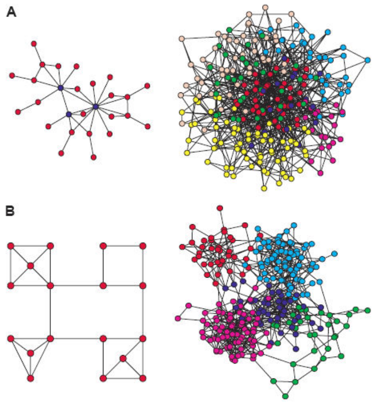
$$\boxed{p(k) \sim k^{-\gamma}}$$

few nodes with high connectivity

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
- Essential genes/proteins tend to be hubs
- Cellular networks are assortative, hubs tend not to interact directly with other hubs.

Complex network models



Scale-free network

Modular networks

Metabolic networks

- We have seen that the cellular functionality can be partitioned into a collection of modules.

- Each module is a discrete entity of several elementary components which perform an identifiable task
  → Modular network (Ci>)

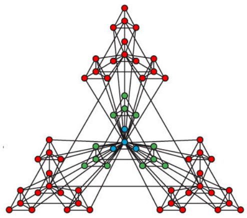
- But, it was demonstrated that the degree distribution follows a power law
  → Scale-free network (power law)

- Modular network ≠ Scale-free network
  → Hierarchical network

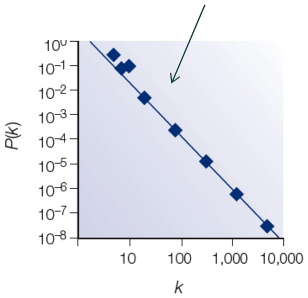
Ravasz et al.Science. 2002

Hierarchical networks

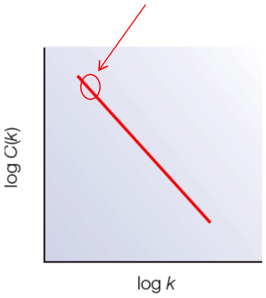
Network



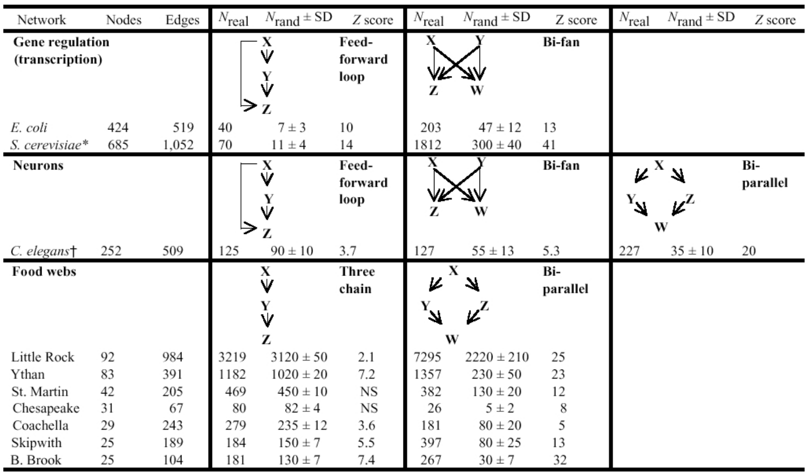
Power law



Modules (C_i↑)

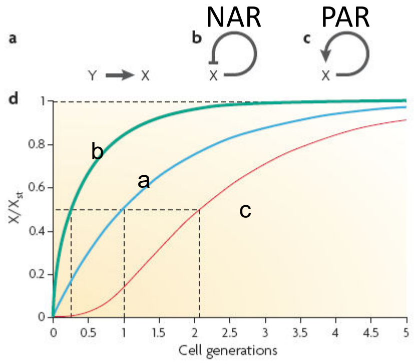


Network motifs



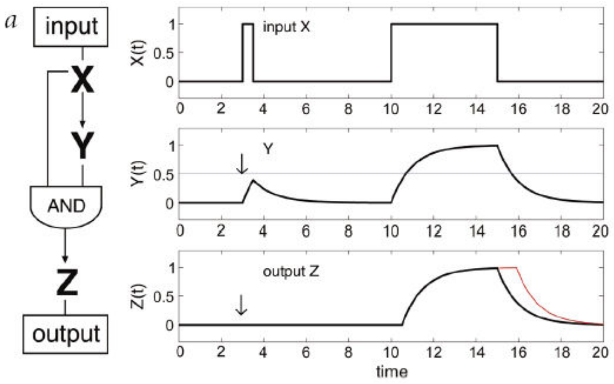
Network	Nodes	Edges	N_{real}	N_{rand} ± SD	Z score	N_{real}	N_{rand} ± SD	Z score	N_{real}	N_{rand} ± SD	Z score
Gene regulation (transcription)					**Feed-forward loop**			**Bi-fan**			
E. coli	424	519	40	7 ± 3	10	203	47 ± 12	13			
*S. cerevisiae**	685	1,052	70	11 ± 4	14	1812	300 ± 40	41			
Neurons					**Feed-forward loop**			**Bi-fan**			**Bi-parallel**
C. elegans†	252	509	125	90 ± 10	3.7	127	55 ± 13	5.3	227	35 ± 10	20
Food webs					**Three chain**			**Bi-parallel**			
Little Rock	92	984	3219	3120 ± 50	2.1	7295	2220 ± 210	25			
Ythan	83	391	1182	1020 ± 20	7.2	1357	230 ± 50	23			
St. Martin	42	205	469	450 ± 10	NS	382	130 ± 20	12			
Chesapeake	31	67	80	82 ± 4	NS	26	5 ± 2	8			
Coachella	29	243	279	235 ± 12	3.6	181	80 ± 20	5			
Skipwith	25	189	184	150 ± 7	5.5	397	80 ± 25	13			
B. Brook	25	104	181	130 ± 7	7.4	267	30 ± 7	32			

Negative and positive autoregulation



- NAR speeds up the response time of gene circuits
- NAR can reduce cell-cell variation in protein levels
- PAR works in the opposite way

Feedforward loop (FFL)



- Supress short signals
- Coherent and incoherent FFL