269020 VO Computational Concepts in Biology II
Part: Biological Networks

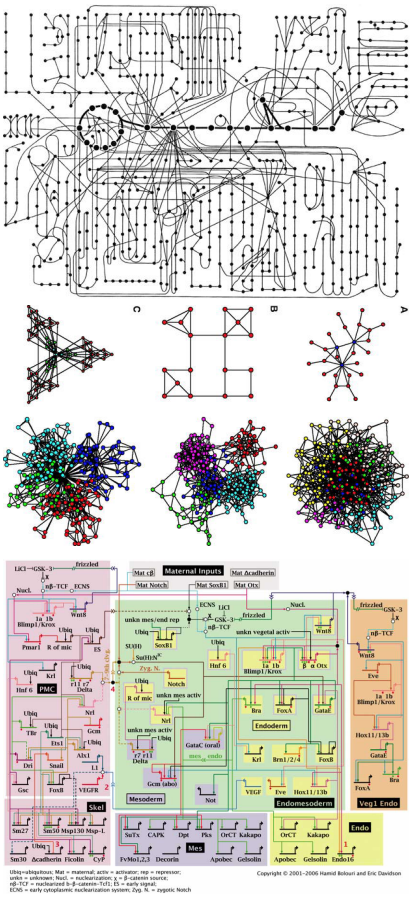
Stefanie Widder[‡] and Christoph Flamm[†]

[†]Institute for Theoretical Chemistry
[‡]CeMM - Center for Molecular Medicine of the
Austrian Academy of Sciences

swidder@cemm.oeaw.ac.at
xtof@tbi.univie.ac.at

SS2019

Biological Networks



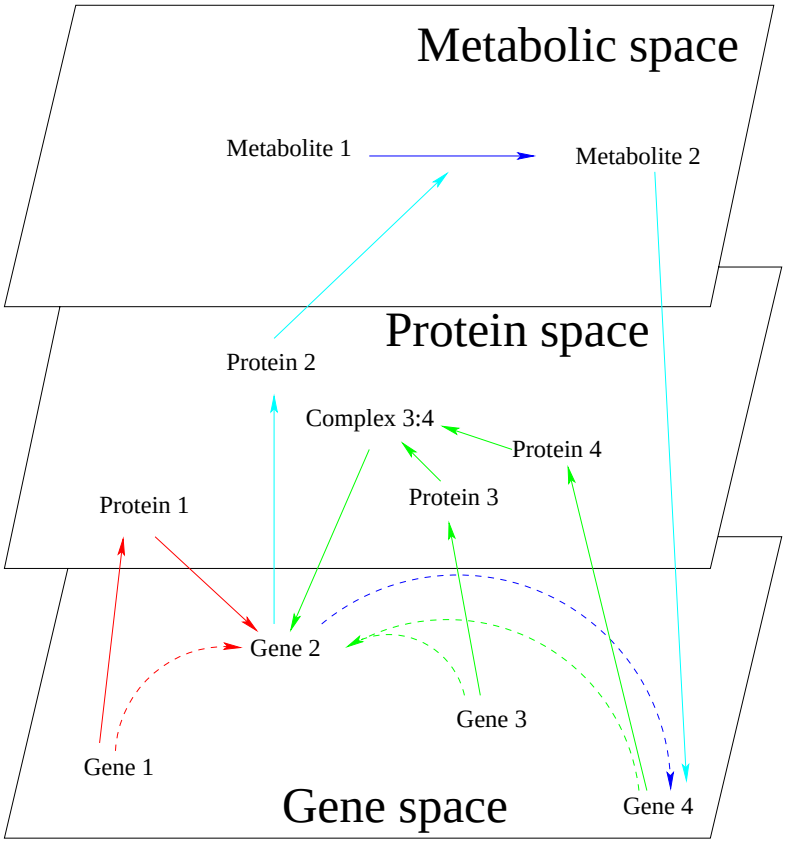
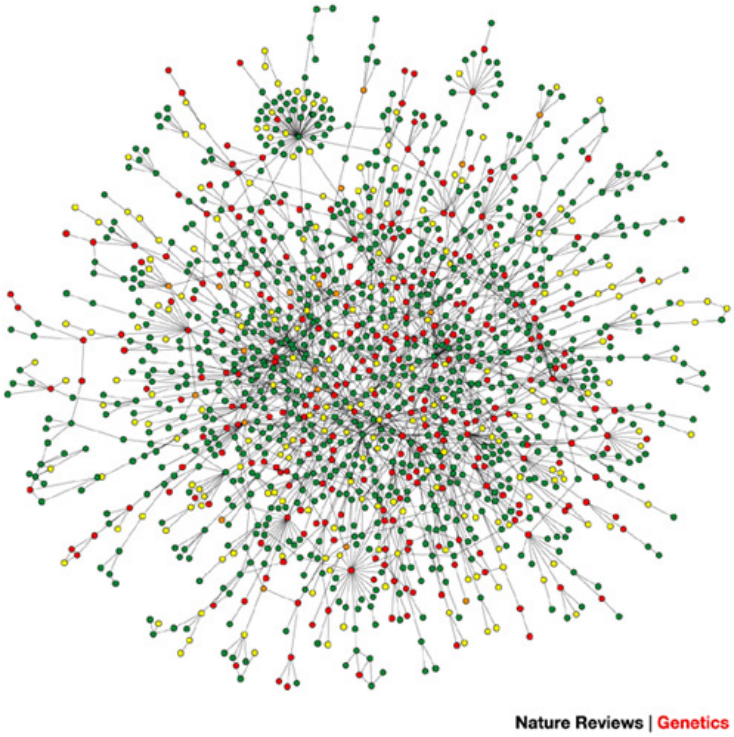


Figure adapted from Brazhnik, P et al. (2002), Gene networks: how to put the function in genomics, *Trends Biotech* **20**:467-472 | doi:10.1016/S0167-7799(02)02053-X

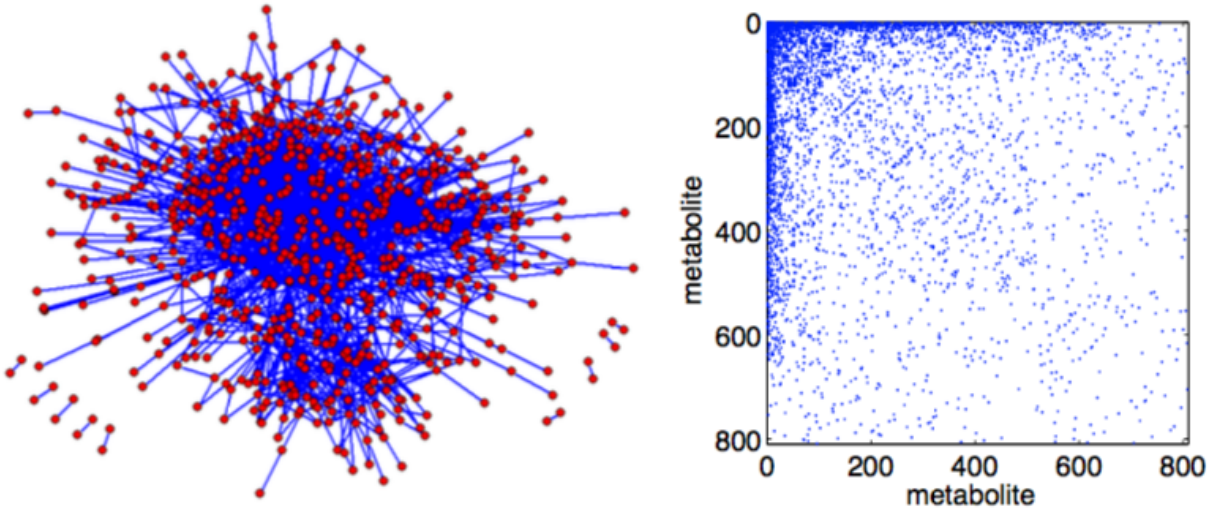
Yeast Protein Interaction Network



The vertex color reflects the phenotypic effect of a gene deletion mutant (green nonlethal, red lethal, orange slow growth, yellow unknown).

Figure adapted from Jeong, H et al. (2001), Lethality and centrality in protein networks, *Nature* **411**:41-42 | doi:10.1038/35075138

Yeast Metabolic Network



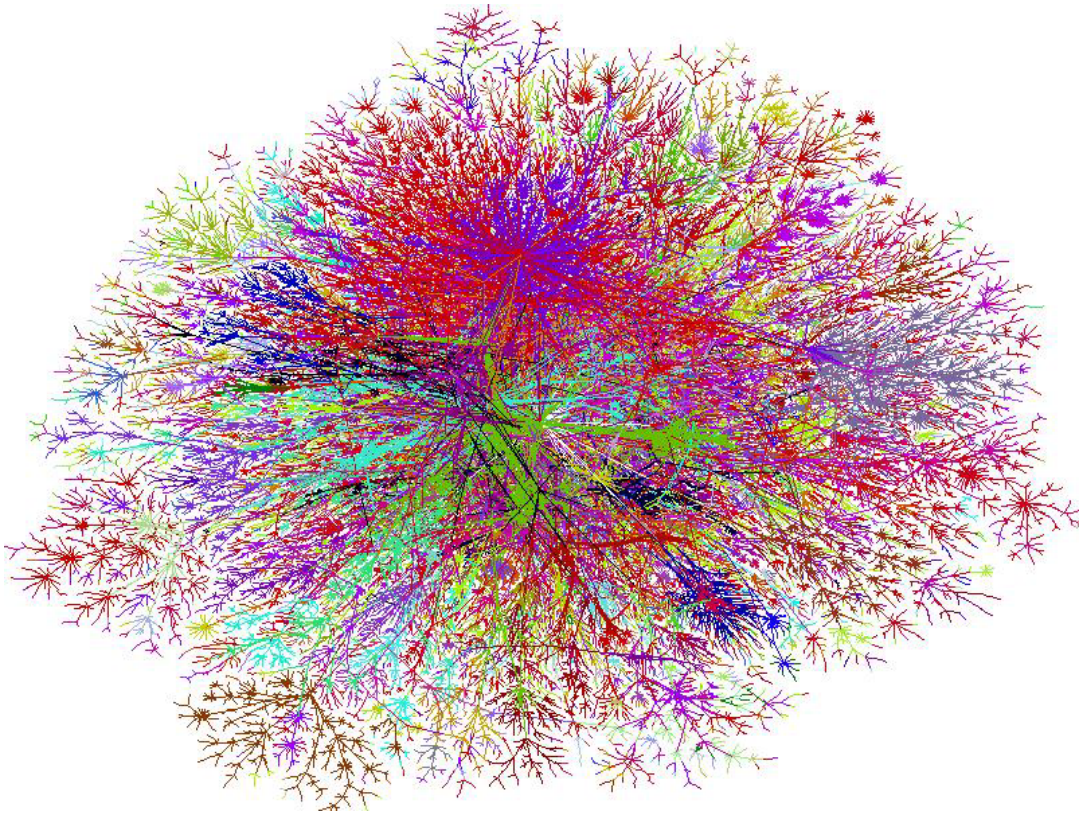
Reconstruction of the *S. cervisiae* metabolic network consisting of 810 metabolites (vertices) and 843 reactions (3419 edges).
Yeast Consensus Reconstruction (Yeast 5): 1418 metabolites, 2110 reactions with 918 verified *S. cerevisiae* genes.

Figure adapted from Förster J et al. (2003), Genome-Scale Reconstruction of the Saccharomyces cerevisiae Metabolic Network, *Genome Res* **13**:244-253 | doi:10.1101/gr.234503

Benjamin D Heavner BD et al (2012), Yeast 5 – an expanded reconstruction of the *Saccharomyces cerevisiae* metabolic network, *BMC Sys Biol* **6**:55 | doi:10.1186/1752-0509-6-55

Map of the Internet

Analytic approaches are needed to "look" at complex networks!



Map of Internet December 1998: colored by IP addresses By William R. Cheswick
http://www.cheswick.com/ches/map/

Language of Complex Networks

Complex dynamical systems are characterized by a large number of non-linearly interacting elements, giving rise to emergent properties (complication|complexity).

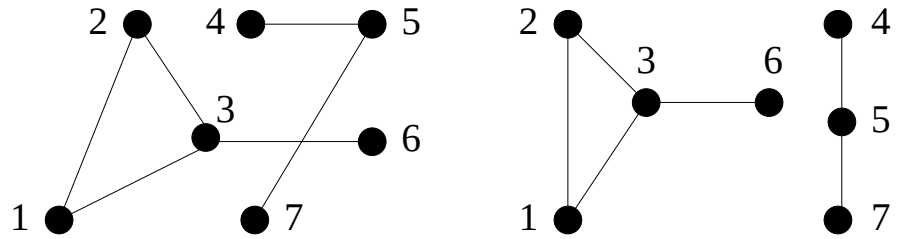
Translation of biological phenomena and processes into networks:

1. makes problems mathematically tractable
   (e.g. transcriptional regulation, metabolism, protein configurations)
2. uncovers
   - functional organization (network topology)
   - underlying design principles
   - unknown organizational principles
3. reveals crucial system properties
   - robustness, resilience
   - redundancy, modularity
   - functional dependencies among network elements
   - systems dynamics

Primer: Graphs

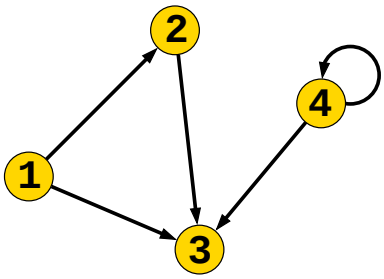
$$G = (V, E) \quad \text{where} \quad e_i = (v_i, v_j)$$

- A graph is a tupel of two sets, the vertex set and the edge set.
- The edge set members are tupels of vertex set members.
- Graphs preserve neighborhood relations.



vertex set $V = \{1, 2, 3, 4, 5, 6, 7\}$
edge set $E = \{(1, 2), (2, 3), (1, 3), (3, 6), (4, 5), (5, 7)\}$

Primer: Graph Representations



graph drawing

$$\mathbf{A} = \begin{bmatrix} 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix}$$

adjacency matrix

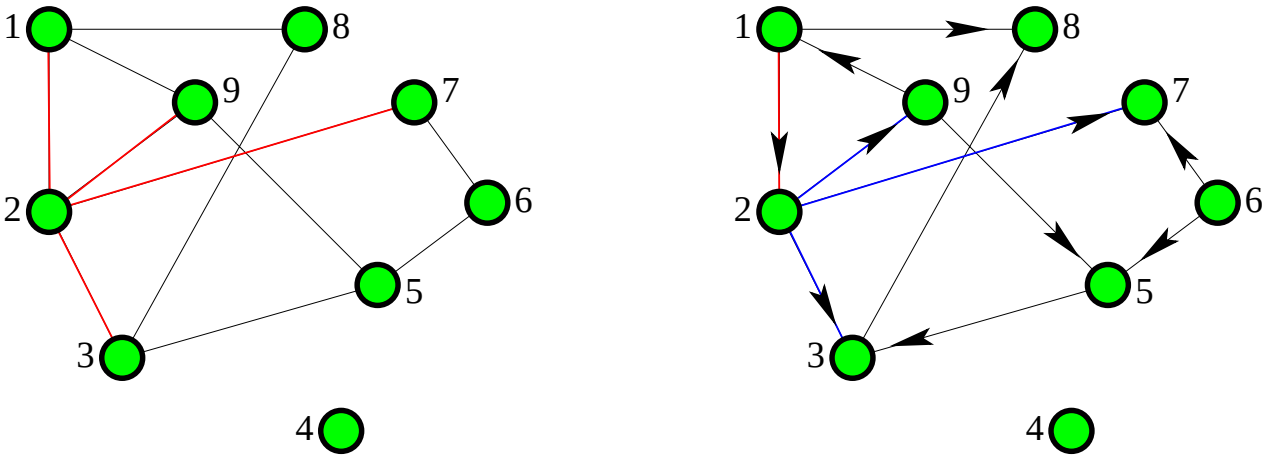
(1, 2, 3, 4)
(1, 2, 3, 4) → **(2, 3)**
(1, 2, 3, 4) → **(3)**
(1, 2, 3, 4) → **()**
(1, 2, 3, 4) → **(3, 4)**

adjacency list

While the adjacency list is very memory efficient, it is in general quicker to use the $|V| \times |V|$ adjacency matrix to determine if an edge (u, v) is present/absent in the graph.

Network Measure: Degree distribution

directed | undirected, in-degree | out-degree

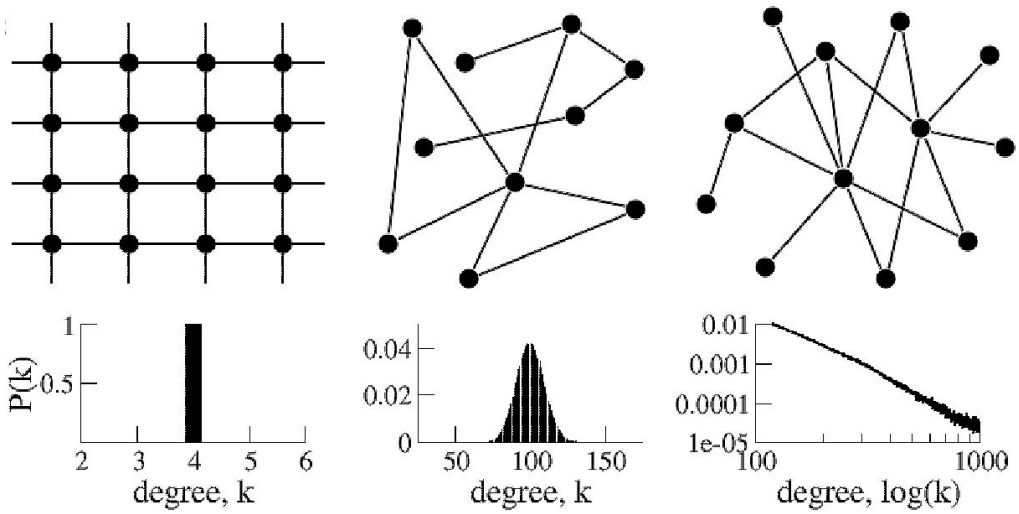


$$p(k) = \frac{|\{v|d(v) = k\}|}{N}$$

The degree distribution $p(k)$ measures the probability that a randomly chosen node in the network has degree k, $d(v)$ is the degree of node v, N is the total number of nodes in the graph.

Classifying Networks into Categories

In graphs without self-loops, the degree equals the number of neighbors of a vertex.



regular random scale-free

$$p(k) = k^{-\gamma}, \qquad \gamma \in \mathcal{R}^{+}$$

[Note: $\log\{p(k)\} = log\{k^{-\gamma}\} = -\gamma \cdot \log\{k\}$].

Power Laws are self-similar

$$f(x) = K \cdot x^{\alpha}$$

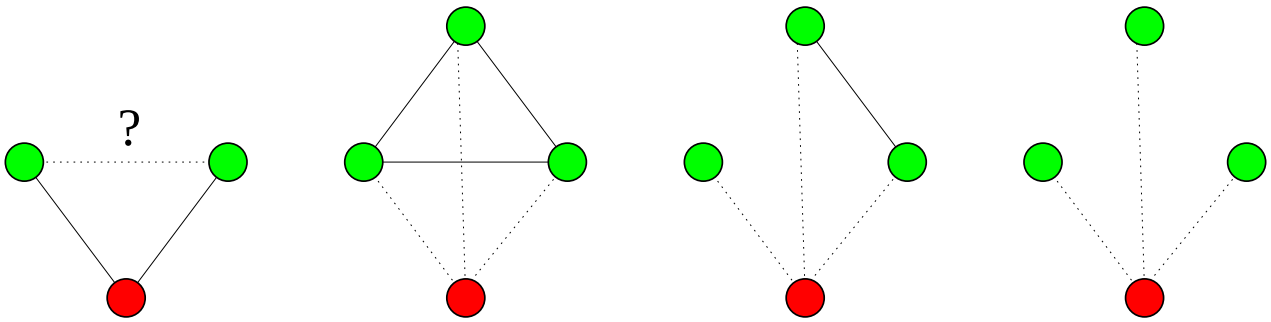
Scale-free denotes the lack of a meaningful average as power laws are self-similar. Overall system properties are similar at all levels of organization.

$$x \mapsto c \cdot x, \quad f(x) \mapsto f(c \cdot x) = K \cdot c^{\alpha} \cdot x^{\alpha} = c^{\alpha} f(x) \propto f(x)$$

A scale transformation leaves the form of the function invariant, it results in a proportional scaling of the function itself. Power laws are the only solution to this functional equation.

Network Measure: Clustering coefficient

Measures the probability that two vertices with a common neighbor are connected.

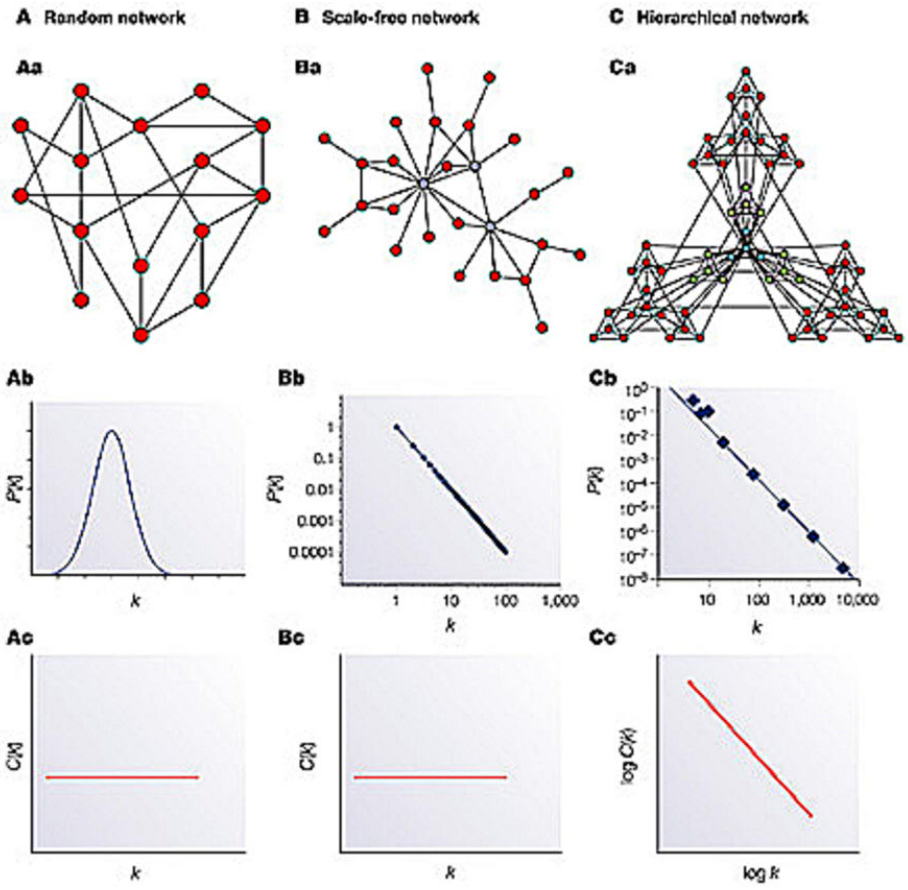


$$C_i = \frac{2 \cdot E_i}{k_i(k_i - 1)}$$

E_i is the number of all observed edges between k_i neighbors of vertex n_i. C_i quantifies the local order (local structure) in the network.

[Note: The maximum number of possible edges between k_i neighbors of vertex n_i is $E_{\max} = k_i(k_i - 1)/2$].

Discriminating between different Networks



Barabási AL and Oltvai ZN (2004), Network biology: understanding the cell's functional organization, *Nature Rev Gen* **5**:101-113 | doi:10.1038/nrg1272

Network Prototype Models

Purpose of prototype models to provide

- insight how observed features arise from construction rules.
  network measures → prototype models → global features
- null models (statistic significance of observed features).

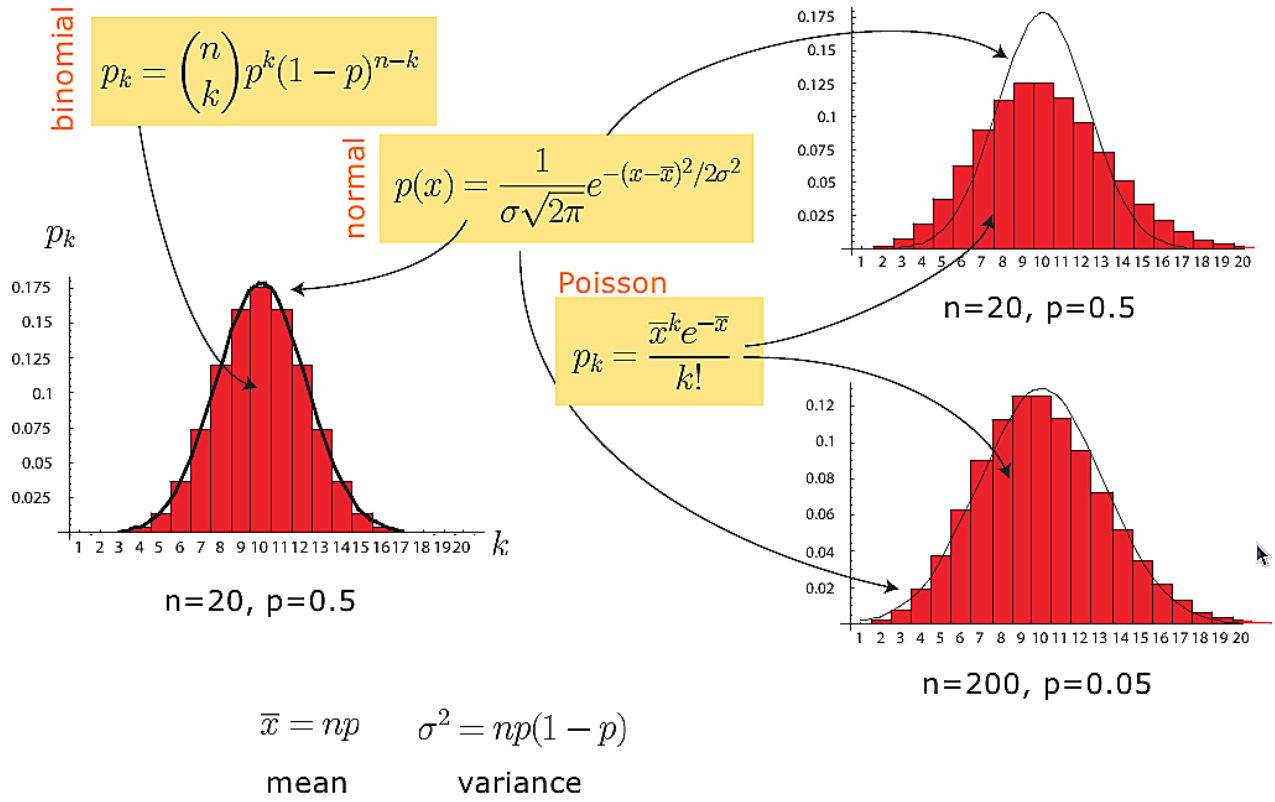
The most common prototype models are:

1. Erdös-Rényi (ER) Model
2. Watts-Strogatz (WS) Model
3. Barabási-Alberts (BA) Model

These are all random networks with particular features...

Albert R & Barabási A-L (2002), Statistical mechanics of complex networks, *Rev Mod Phys* **74**:47–97 | doi:10.1103/RevModPhys.74.47

Primer: Statistical distributions



$\overline{x} = np$ $\sigma^2 = np(1-p)$

mean variance

The Erdös-Rényi Model

Consists of N_V vertices connected by N_E undirected edges which are chosen randomly from the set of $N_V \cdot (N_V - 1)/2$ possible edges.

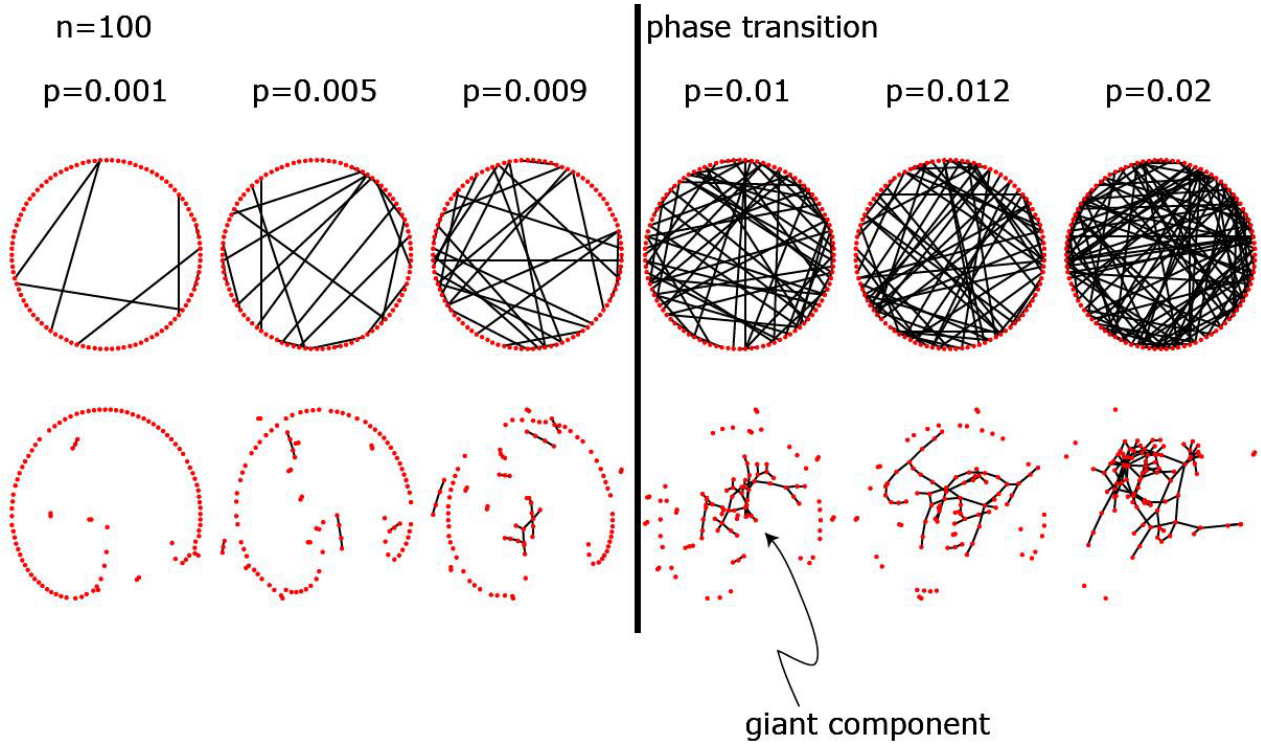
The probability p that two randomly chosen vertices are connected is thus

$$p = \frac{2 \cdot N_E}{N_V \cdot (N_V - 1)}$$

The degree distribution is given by a binomial distribution that becomes approximately **Poissonian** for large networks ($N_V \to \infty$).

The ER model exhibits the small-world property, i.e. over the percolation threshold the average path length l between nodes scales as $l \sim \log N_V$.

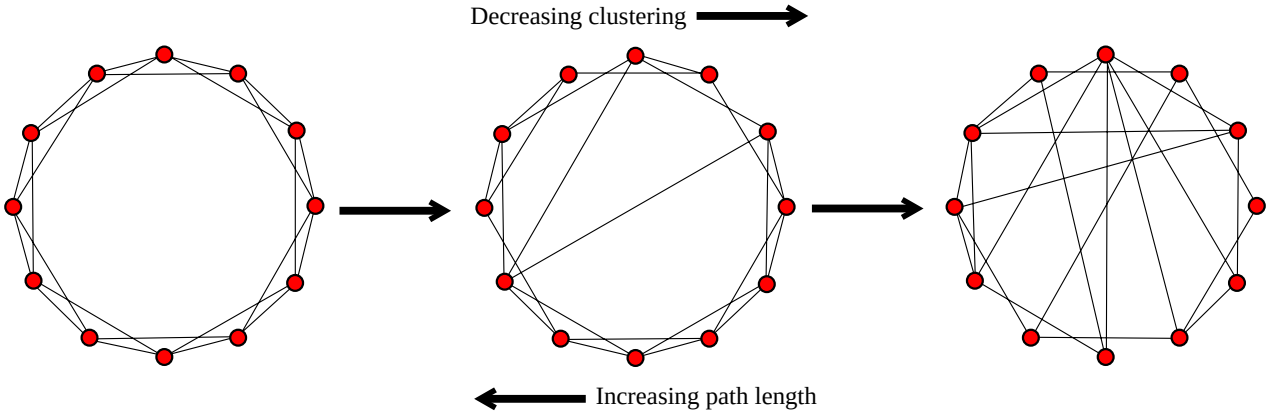
Phase Transition in the Erdös-Rényi Model



For small p the network is disconnected. At $p \sim 1/N_V$ (i.e. $\langle k \rangle \approx 1$) a phase transition to a giant component occurs. For $p \geq \log(N_V)/N_V$ all vertices are connected.

The Watts-Strogatz Model

While the ER model correctly reproduces the small world property, it fails to account for local clustering.



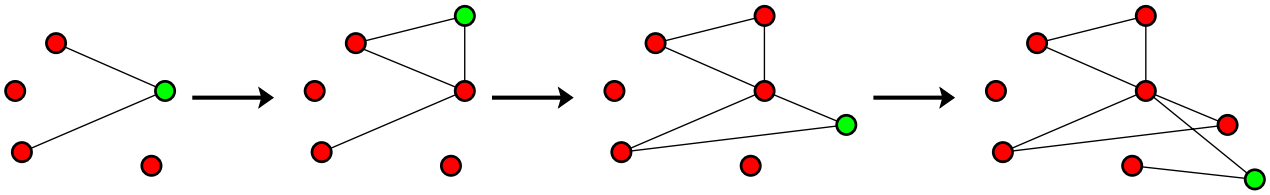
- Start with regular lattice-like network with local clusters.
- Introduce shortcuts with probability p_{new}.
- For increasing p_{new} the WS model approaches the ER model (for $p_{new} \to 1$ the ER model is recovered).

A very small number of shortcuts $p_{new} \ll 1$ is sufficient to rapidly decrease the avarage path length and maintain local clustering.

Watts DJ & Strogatz SH (1998), Collective dynamics of 'small-world' networks, *Nature* **393**:440-442 | doi:10.1038/30918

The Barabási-Albert Model

ER and WS model do not capture the inhomogeneous degree distributions found in empirical networks.



The BA model is based on two essential ingredients:

- Growth process (new vertex is added each time step).
- Preferential attachment:

The probability that a vertex n_i receives a new edge is proportional to its degree k_i

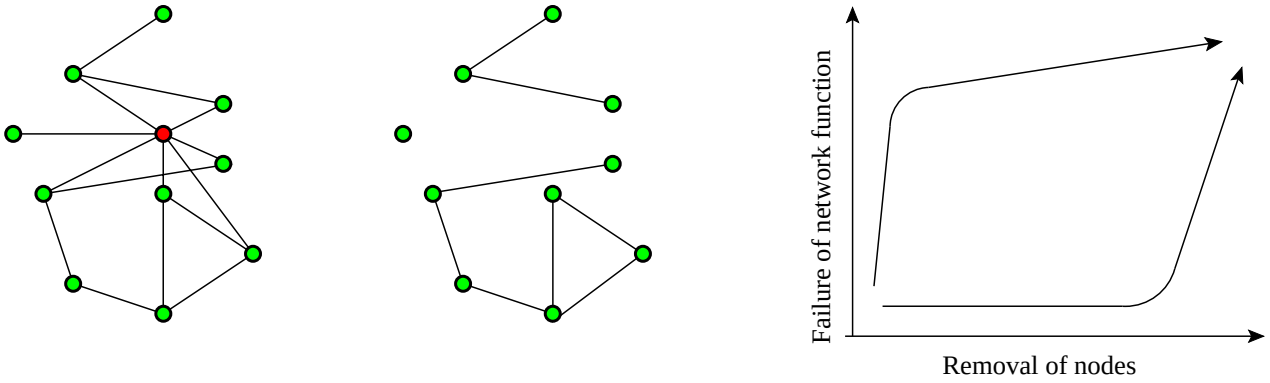
$$p(n_i) = \frac{k_i}{\sum_j k_j}$$

For $t \gg 1$ the degree distribution follows a power law $p(k) \sim k^{-\gamma}$.

Barabási A-L & Albert R (1999), Emergence of Scaling in Random Networks, *Science* **286**(5439):509-512 | doi:10.1126/science.286.5439.509

Tolerance against Node Removal

The topology of scale-free networks is dominated by a few highly connected hubs.



Therefore scale-free networks are robust against failures of arbitrary nodes but quite vulnerable if hubs are attacked.

Note that a comprehensive view of robustness must as well take dynamic aspects into account!

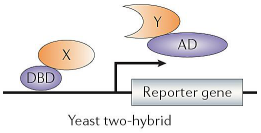
Barabási A-L (2012), Network science: Luck or reason, *Nature* **489**:507-508 | doi:10.1038/nature11486

The Interactom: Who interacts with whom?

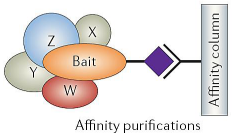
Provide a comprehensive list of protein-protein interactions.

Experimental approaches:

- chemical crosslinking / footprinting
- protein arrays
- flourescence resonance energy transfer (FRET)
- flourescence cross-correlation spectroscopy
- yeast two-hybrid system
- affinity purification



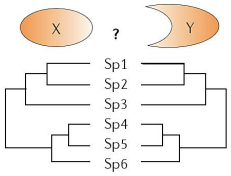
Yeast two-hybrid



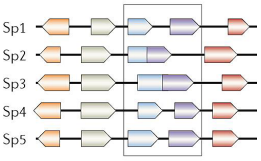
Affinity purifications

Computational approaches/predictions:

- Genomic context
- Co-evolutionary



Co-evolution

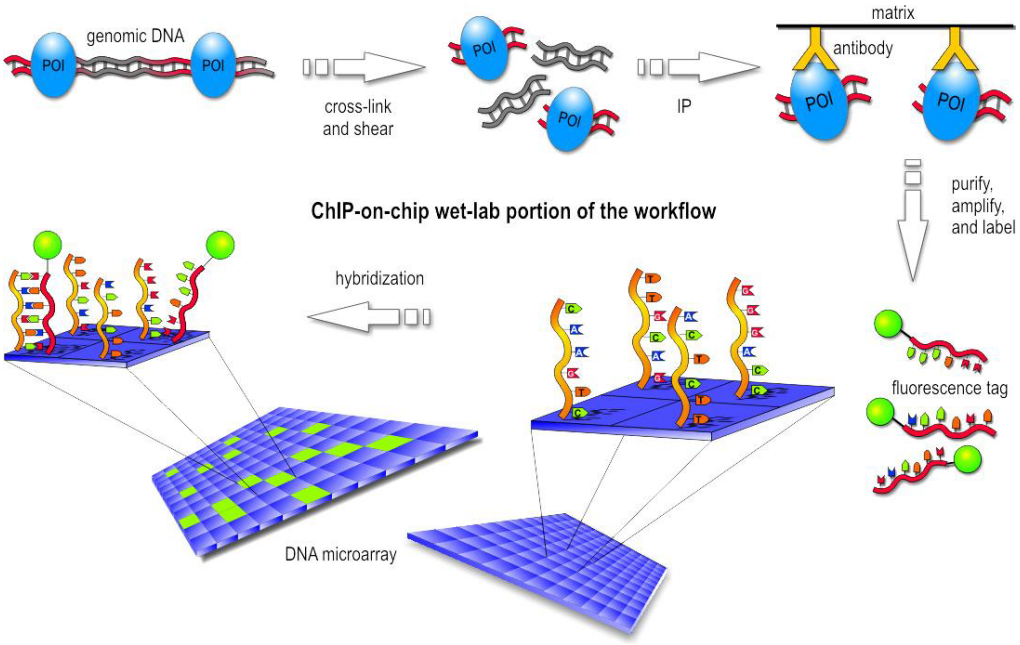


Genomic context

Sardiu ME and Washburn MP (2011), Building Protein-Protein Interaction Networks with Proteomics and Informatics Tools, *J Biol Chem* **286**:23645-23651 | doi:10.1074/jbc.R110.174052

Identification of protein DNA interactions (GRNs)

The **ChIP-on-chip** technique combines chromatin immunoprecipitation ("ChIP") with DNA microarray technology ("chip").



ChIP-on-chip allows to study binding sites of DNA-binding proteins on a genome-scale basis. Other methods: DNAse-seq, ATAK-seq distinguishes binding sites and open chromatin regions

Figure from wikipedia.org

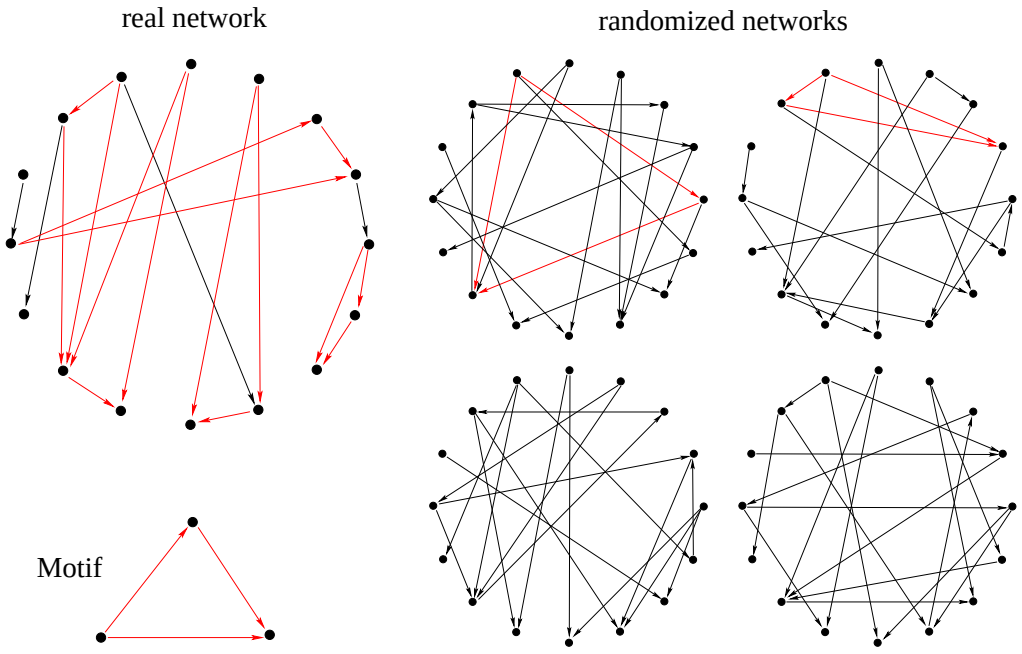
Weaknesses of Interactom Screens

1. A poor overlap was observed between screens run by different groups on the same organism.
2. High numbers of false positives (30-60%) and false negatives (40-80%) in the screens.
3. All biophysically possible interactions are discovered not necessarily only those occuring in biology (Q: What constitutes a false positive?).
4. Many already known interactions are missed. (This is due to under representation, mainly owed to the biological properties of proteins, e.g. aggregation upon overexpression).
5. Transient interactions or interactions depending on post-trancriptional modification events (e.g. phosphorylation, ubiquitination) are not captured by the current methods.

Russell RB and Aloy P (2008), Targeting and tinkering with interaction networks, *Nature Chem Biol* **4**(11):666-673 | doi:10.1038/nchembio.119

Motif Detection and Network Dynamics

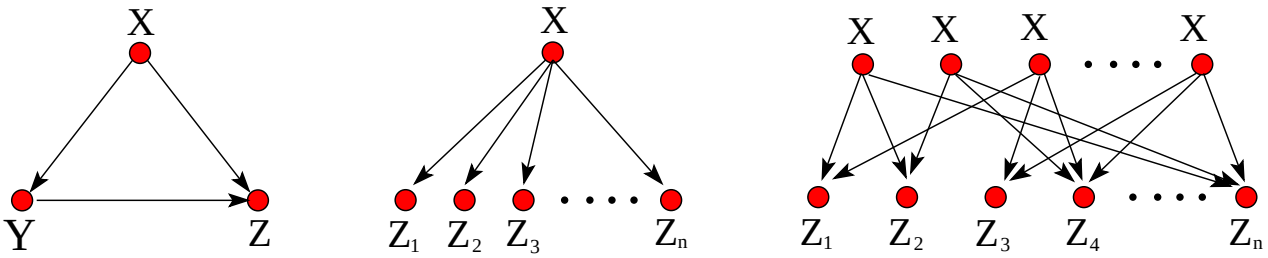
1. Represent the network of interest as a directed graph.
2. Determine the frequency of particular subgraphs.
3. Compare againset ensemble of randomized networks.



Shen-Orr SS et al (2002), Network motifs in the transcriptional regulation network of *Escherichia coli*, *Nature Genetics* **31**:64-68 | doi:10.1038/ng881

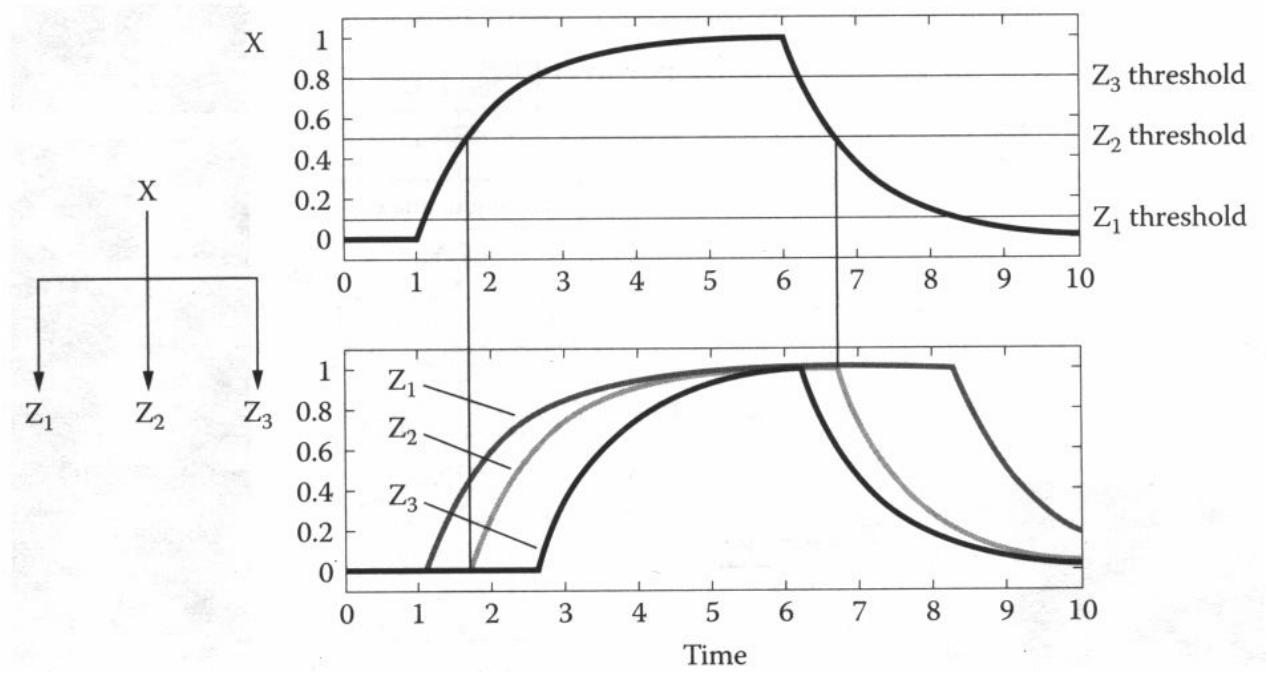
Transcription regulation motifs in *E. coli*

Motif	Freq_{real}	Freq_{rand}	P-value
coherent FFL	34	4.40 ± 3	$P < 0.001$
incoherent FFL	6	2.50 ± 2	$P \sim 0.03$
SIM	68	28.00 ± 2	$P < 0.01$
DORs	203	57.00 ± 14	$P < 0.01$
cycles	0	0.18 ± 0.6	$P \sim 0.8$



Shen-Orr SS et al (2002), Network motifs in the transcriptional regulation network of *Escherichia coli*, *Nature Genetics* **31**:64-68 | doi:10.1038/ng881

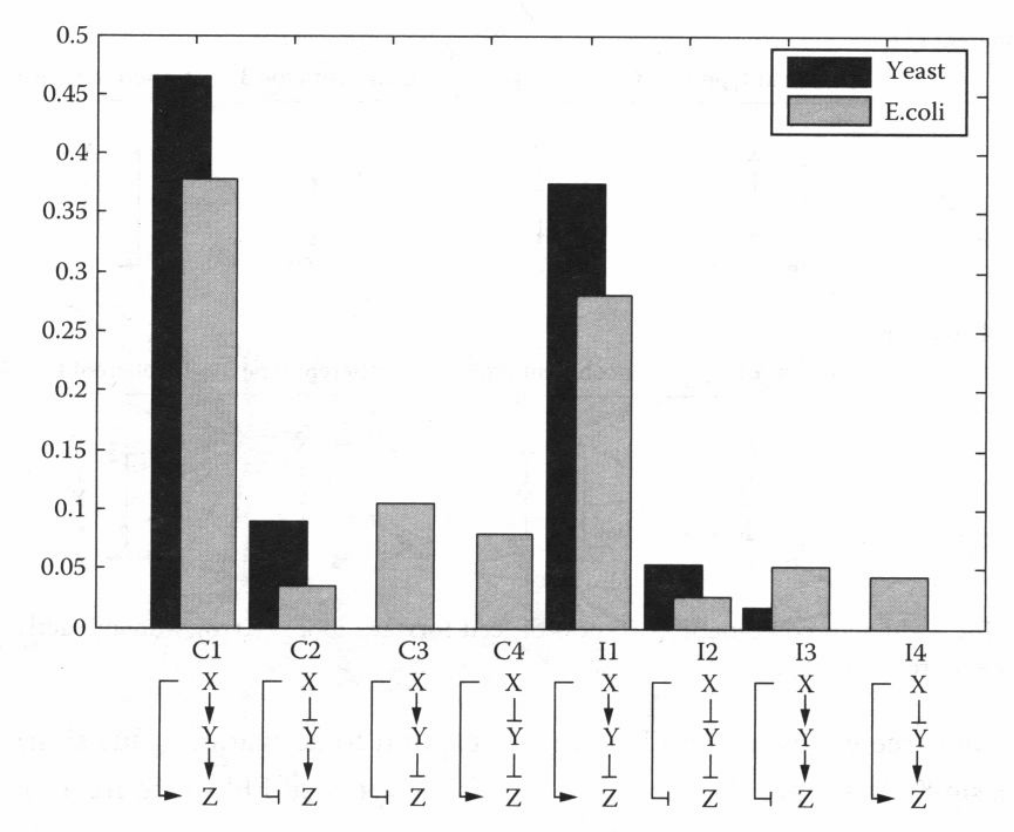
SIMs generate a temporal program of expression



1. Frequent in metabolic pathways regulation (e.g. arginine system).
2. Temporal order of genes matches their functional order.
3. Economic design! Proteins are not produced before they are needed.

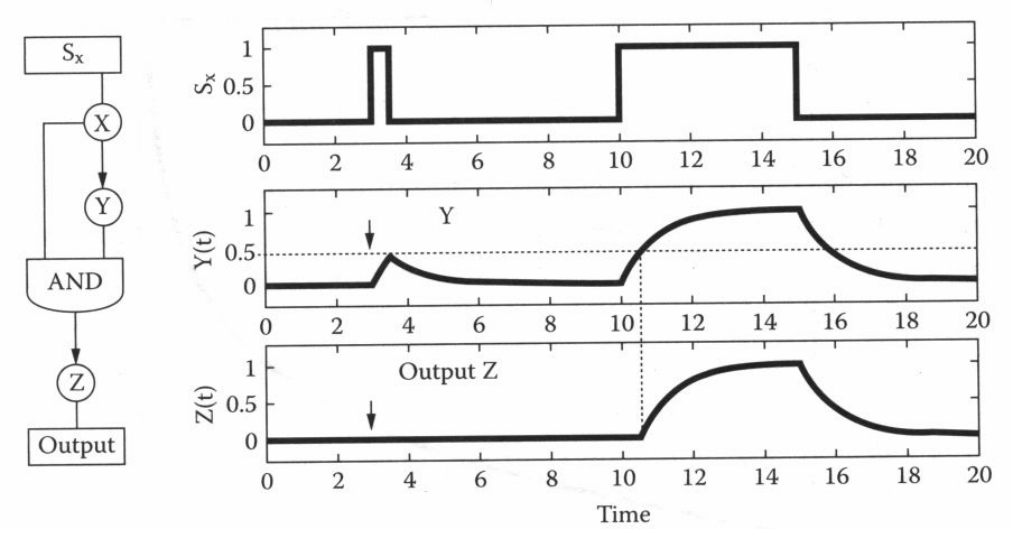
Zaslaver A et al (2004), Just-in-time transcription program in metabolic pathways, *Nature Genetics* **36**:486-491 | doi:10.1038/ng1348

Relative abundance of FFL types



Mangan S et al (2006), The incoherent feed-forward loop accelerates the response-time of the gal system of Escherichia coli, *J Mol Biol* **356**:1073-1081 | doi:10.1016/j.j mb.2005.12.003

Dynamics of the coherent feed-forward-loop motif



The motif function: filter out brief spurious pulses of signal

- Sign-sensitive delay function (signal on/off ⇒ delay/no delay).
- Responds only to persistent signals

Alon, U (2007), Network motifs: theory and experimental approaches, *Nature Rev Genetics* **8**:450-461 | doi:10.1038/nrg2102

Further Reading


Alon U

Network motifs: Theory and experimental approaches.

Nature Reviews Genetics, **8**:450–461, 2007.


Albert R, Barabási A-L

Statistical mechanics of complex networks.

Reviews of Modern Physics, **74**:47–97, 2002.


Barabási A-L, Zoltán NO

Network Biology: Understanding the cell's functional organization.

Nature Reviews Genetics, **5**:101–113, 2004.


Newman MEJ

The structure and function of complex networks.

SIAM Review, **45**(2):167–256, 2003.


Watts DJ, Fell DA

Collective dynamics of 'small-world' networks.

Nature, **393**:440–442, 1998.