Patrik Edén
Complex Systems
Theoretical Physics;
Lund Stem Cell Center
patrik@thep.lu.se

Network Design I

- Random networks
- Degree distribution
- Random networks revisited
- Motifs

Are biological networks special?

Well, what is “special”?

- **Subjective**

  example: humans have 5-6 times as many genes
  as the bacteria *E.Coli*.
  Interestingly few! Complexity does not grow with network size
  as we expect, or humans are simply not as complex as we think.

- **Compare with other networks**

  Other fields (*e.g.*, social networks)
  Other biological networks
  - protein interaction versus protein regulation
  - different species

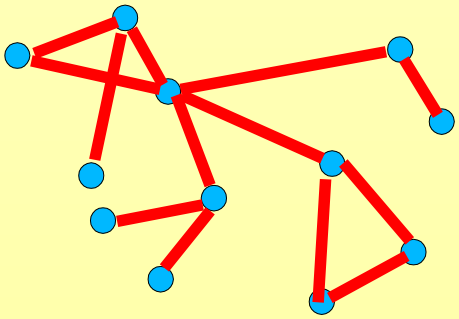
- **Compare with random artificial networks**

  Today's topic!

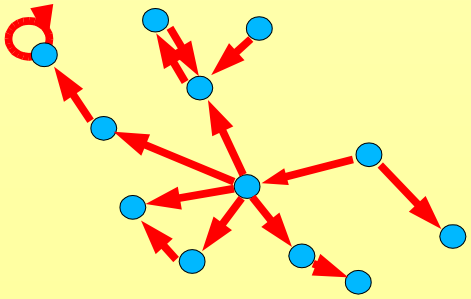
Random networks

Compare:
Biological network: N nodes, L links
Random network: N nodes, probability for a link p

What is p?



Undirected: $\frac{N(N-1)}{2}$ node pairs, $p = \frac{2L}{N(N-1)}$



Directed: N starting points, N destinations, $p = \frac{L}{N^2}$

Random network generation 1

Start with N nodes.
In every possible place, insert a link with probability p.
Compare with your real network.

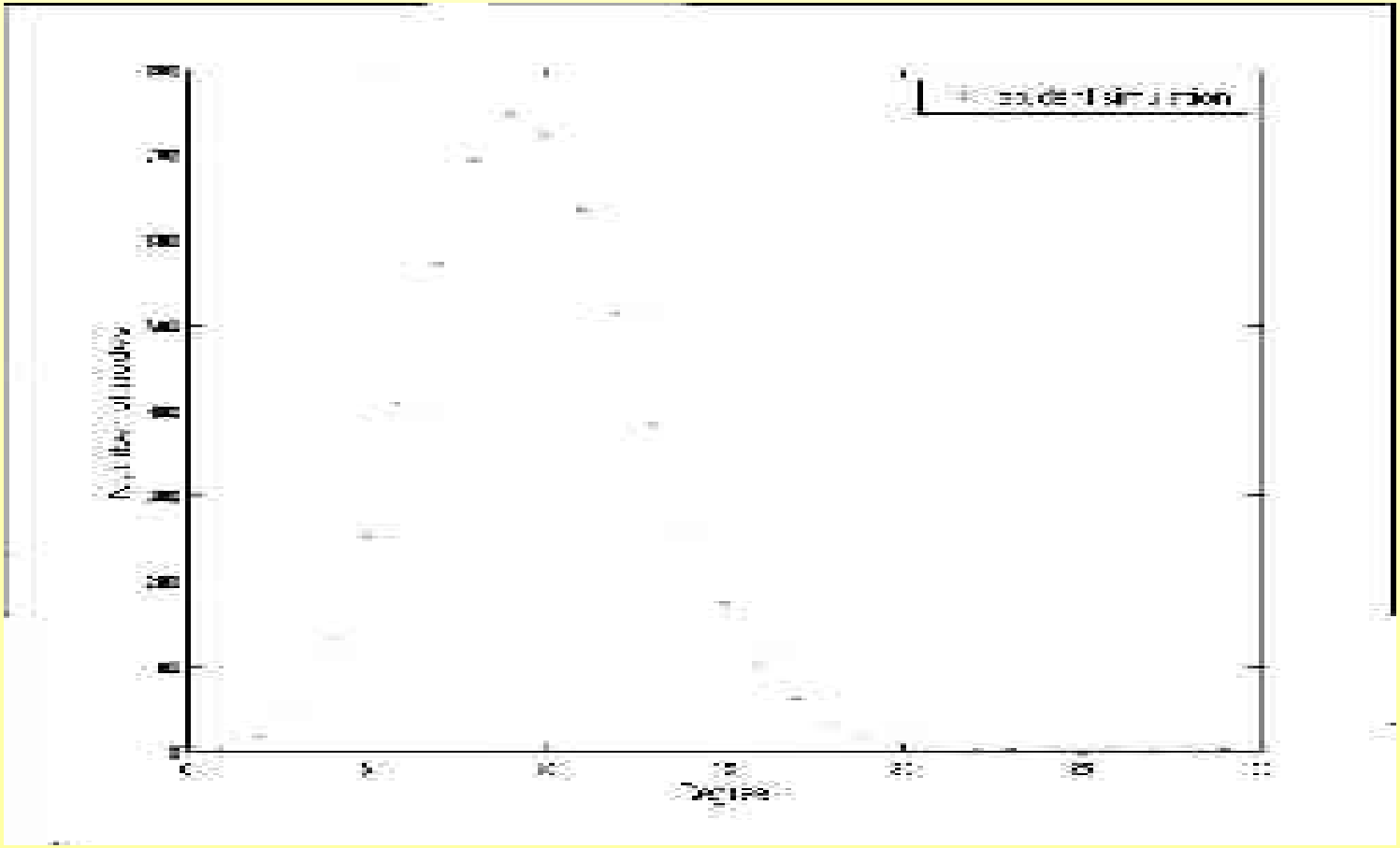
Degree distribution

Probability that a node has **k** links: binomial distribution.

$$\binom{N-1}{k} p^k (1-p)^{N-1-k}$$

Degree distribution

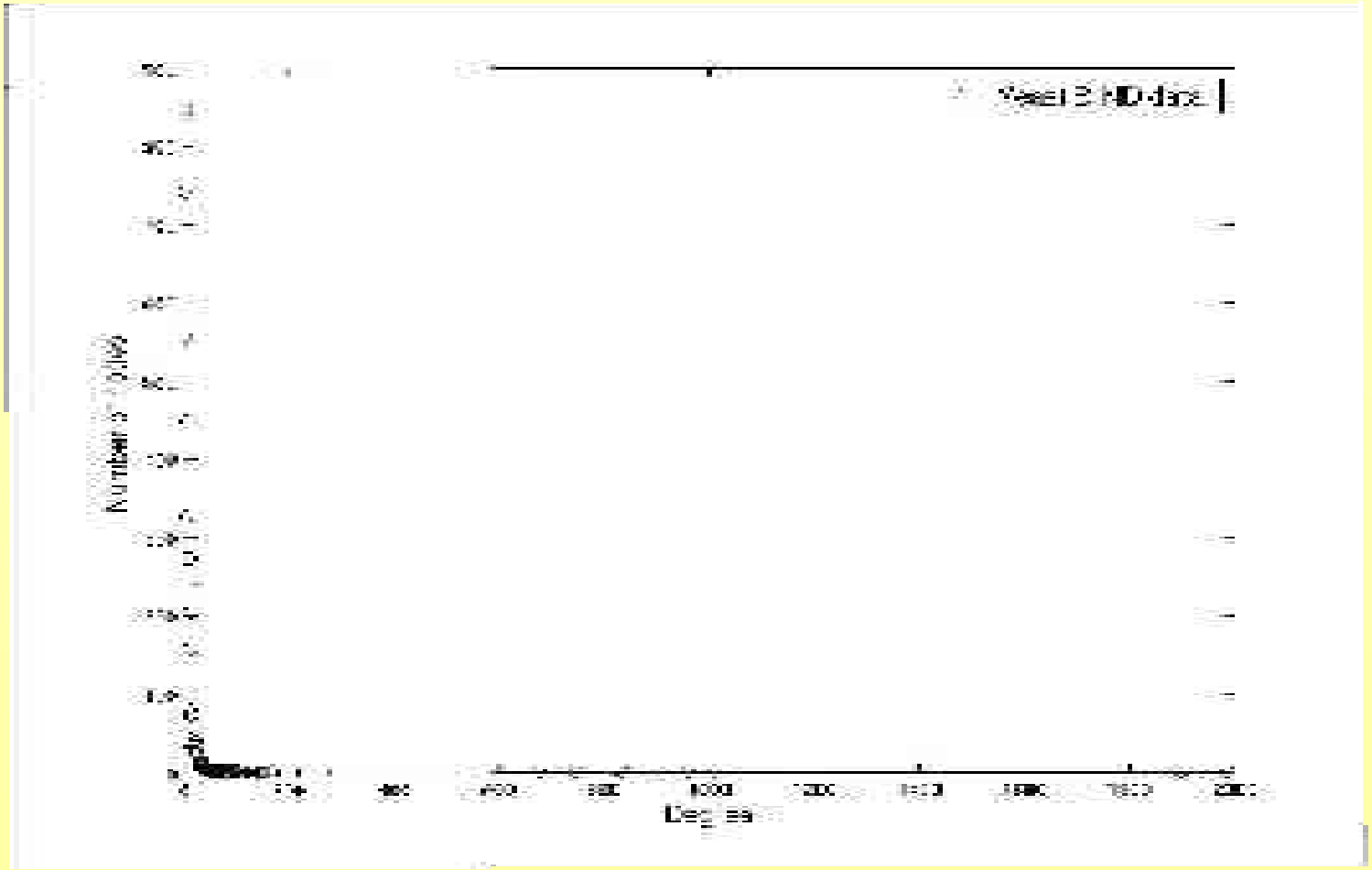
Random network, N=5800, L=28110



The network is not single connected, but almost.

Degree distribution

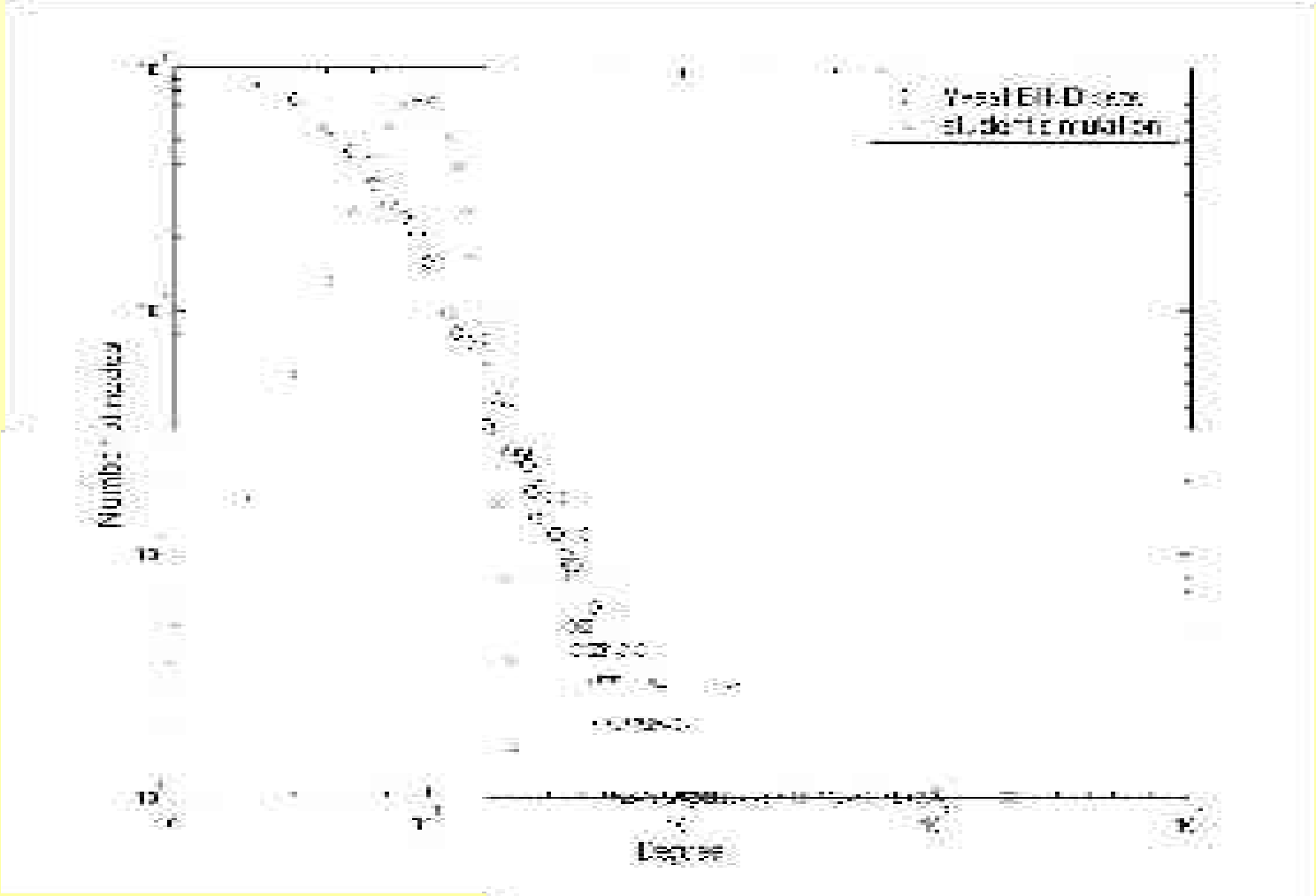
Yeast *(S. cerevisiae)* network, ***N***=5800, ***L***=28110



Note the huge range on the x-axis!

Degree distribution

comparison, N=5800, L=28110



The p-value of getting the yeast distribution by chance is 0.00000000000000000000...

Degree distribution

Roughly a straight line in log-log plot
"Scale-free network"

$p_k \sim k^{-\gamma}$

True for almost all social and biological networks.
(k-depedence differs. γ=1-3 common.)

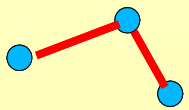
Not true for the random network discussed so far.

Biological networks contain more nodes with very many links than you expect by random.

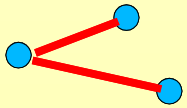
- Transcription factors controlling really many genes.
- Proteins interacting with really many other proteins.

Can we study more than degree distribution?

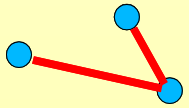
For example, how many connected 3-node groups have all 3 links?



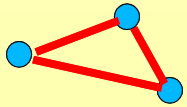
$p^2(1-p)$



$p^2(1-p)$



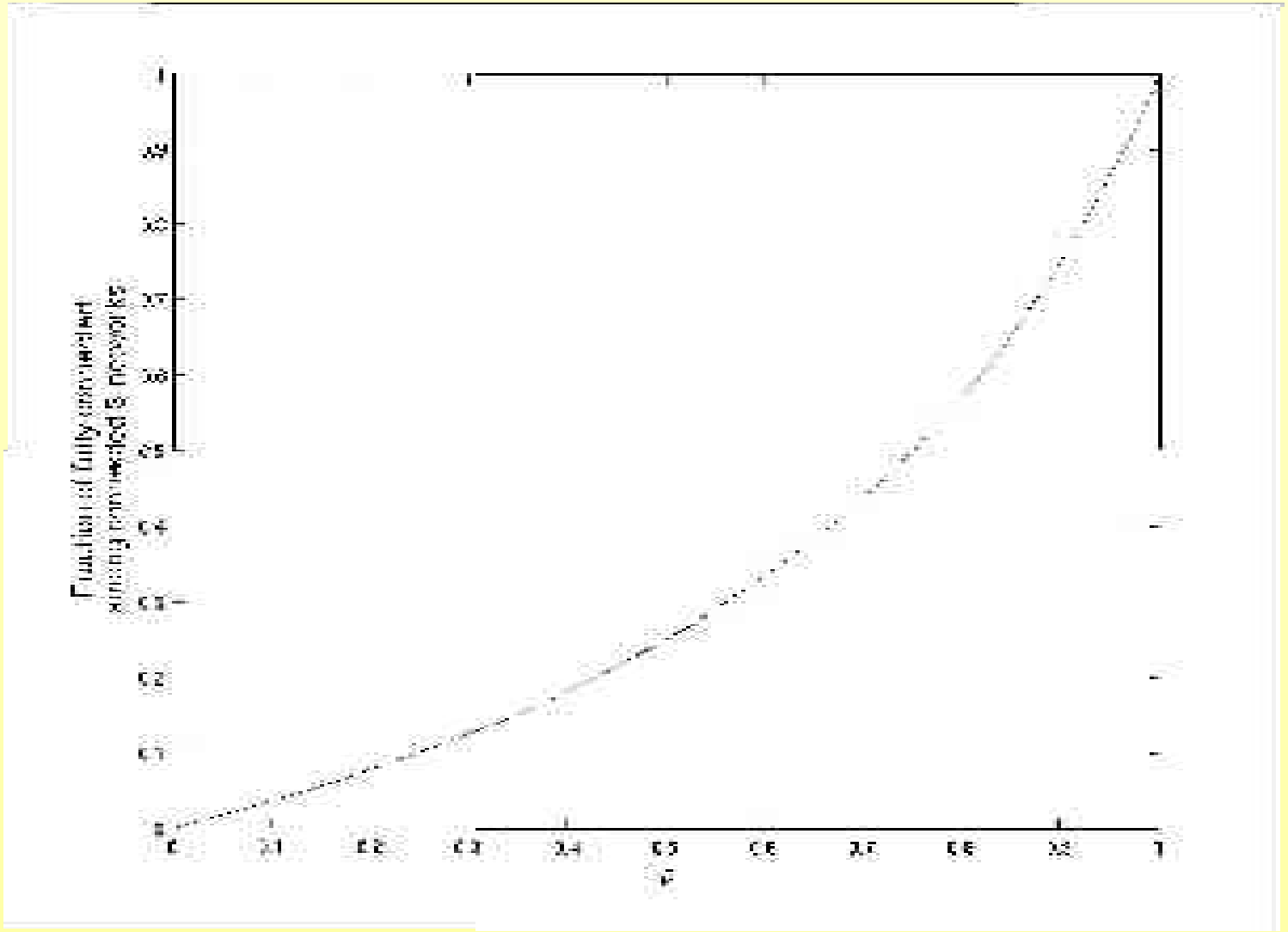
$p^2(1-p)$



p^3

Expected fraction complete triangles in random networks:

$$p^3/[3p^2(1-p)+p^3]=p/(3-2p)$$



Is this legitimate?

No!!

- Probability for a triangle depends on the probability for a link to be present
- The probability for a link to be present depends on the degree of the node in question
- **Have to ensure our random networks have the right degree distribution.**

Alternative: growing random networks

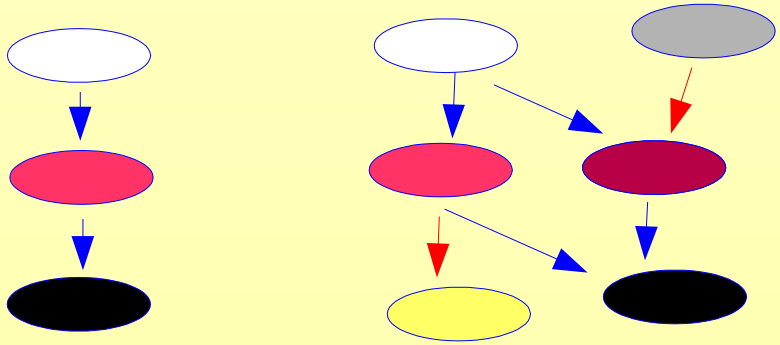
- Start with 2 nodes.
- Insert a link with probability p.
- Add a node.
- Insert a link to every other node, with a probability **that depends on the number of links the node already has** ("Many gets more").
- Stop when you have N nodes.

Fine-tune p and "many gets more" probabilities
to get roughly L links and correct degree distribution.

Only samples a subset of possible random networks
(e.g., the higher order connectivity discussed
in "network basics" will not really be random).

Alternative: mimic evolution

Copying + diversing



Can be tuned to give scale free networks,
but difficult to control what subset of random networks
that are generated

Heuristic method: randomize given network

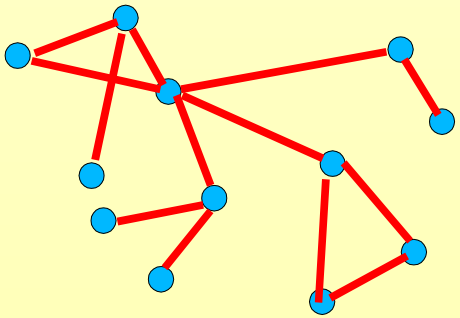
Flexible in preserving other known features of the network.

Believed to sample a large subset of networks
(not proven to my knowledge)

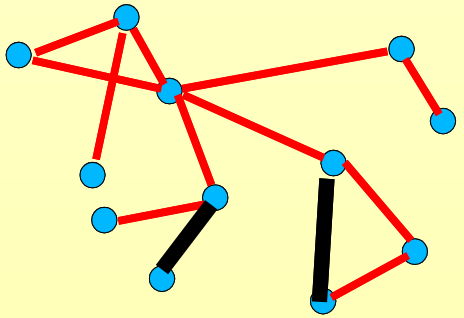
De facto standard

Conserves the degree of all nodes, *i.e.*,
keeps degree distribution of every kind of node
(transcription factors, other genes...)
(kinases, receptors, other protein subgroups...)

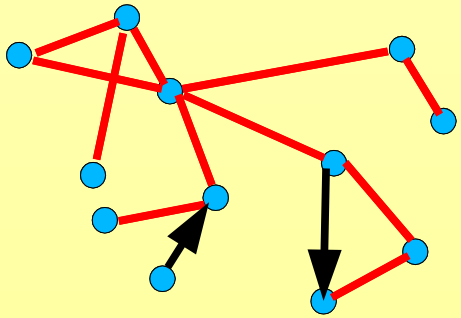
Randomization



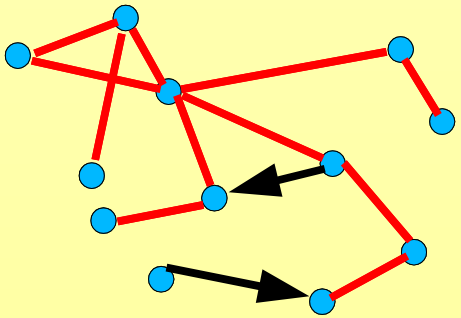
1a. Starting network



1b. Pick two links

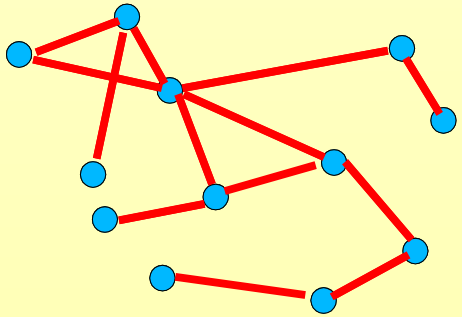


1c. If undirected, pick directions

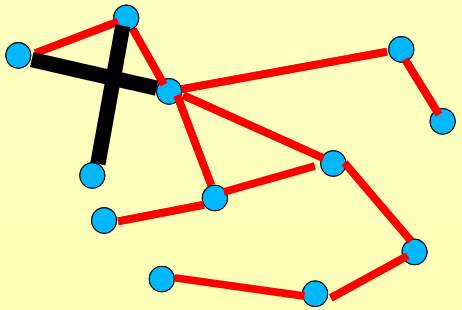


1d. Swap arrow heads

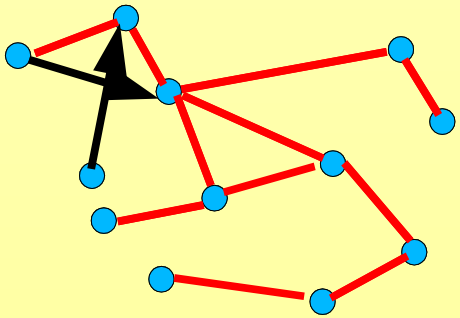
Randomization



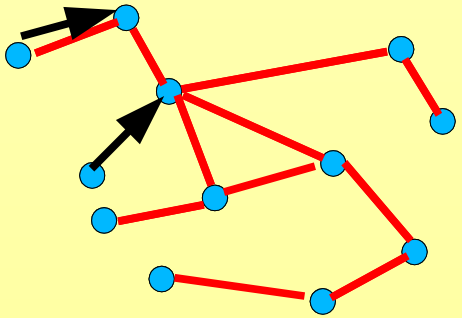
2a. Modified network



2b. Pick two new links

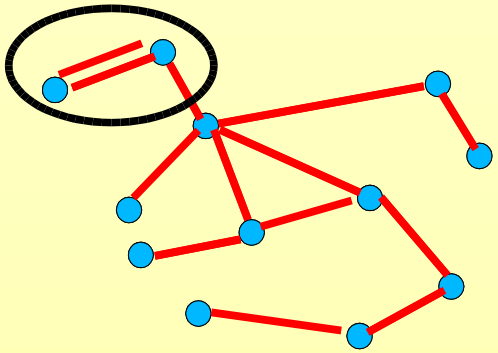


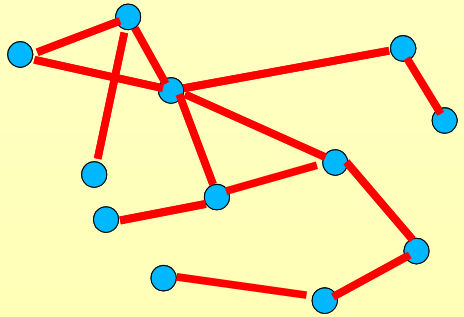
2c. Pick directions



2d. Swap arrow heads

Randomization





3a. Modified network
Is it acceptable?
No, double undirected link

3b. Go back to previously accepted network (2a).
Pick two new links...

Check for what ever you want to avoid:
Double undirected links
Double directed links in same direction
Self-self connection
Single connected network / many components.

“Other features”

- Functional classification of the gene
  (Are genes with high degree lethal? Part of known gene family?)
  This can be studied without random networks.

Using random networks with correct degree distribution,
we can ask:

Is our biological network different in…

- Motifs (small often occuring subgroups in the network)
- Degree distribution of neighbours to nodes of a fixed degree
  (the “higher order” connectivity from network basics)
- …

Motifs

- Big networks are never 100% identical
- Small subparts of networks might be, called “motifs”
- An example of a simple motif is the triangle

Triangles

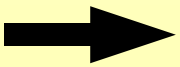
In directed networks

Triangles are feed forward loops or feed backward loops

Transcriptional regulation in *E. coli*
(nature genetics 2002, Shen-Orr)

Only feed forward loops are overrepresented

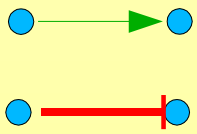
40 in real network
7 ± 4 in randomized networks



Z-score $(40-7)/4 = 8$
p-value ~ 10^{-17}

Regulation comes with a sign

- Activate (336 links). Turns on the gene.
- Repress (214 links). Turns off the gene.
- Dual (29 links). Sometimes activate, sometimes repress.

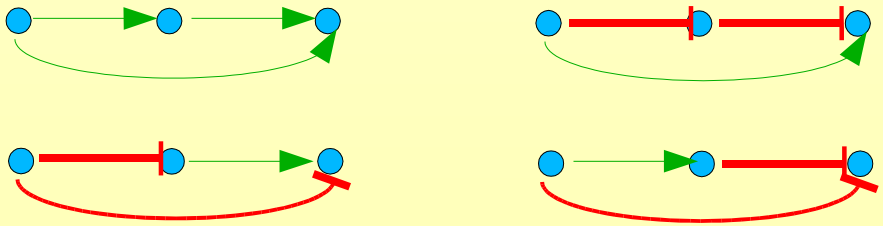


(This result for *E. coli* may change with experimental development, and might be different for eukariots.)

Transcriptional regulation

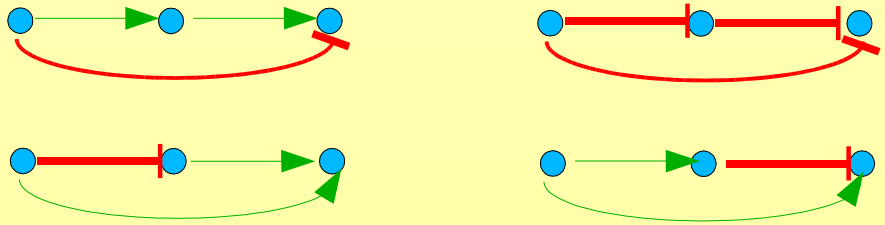
Two kinds fo feed forward loops

- Coherent



E. coli: 34 in real, 4.4 ± 3 in random, Z=(34-4.4)/3=9, p=10^{-22}

- Incoherent

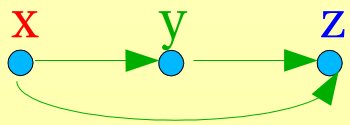


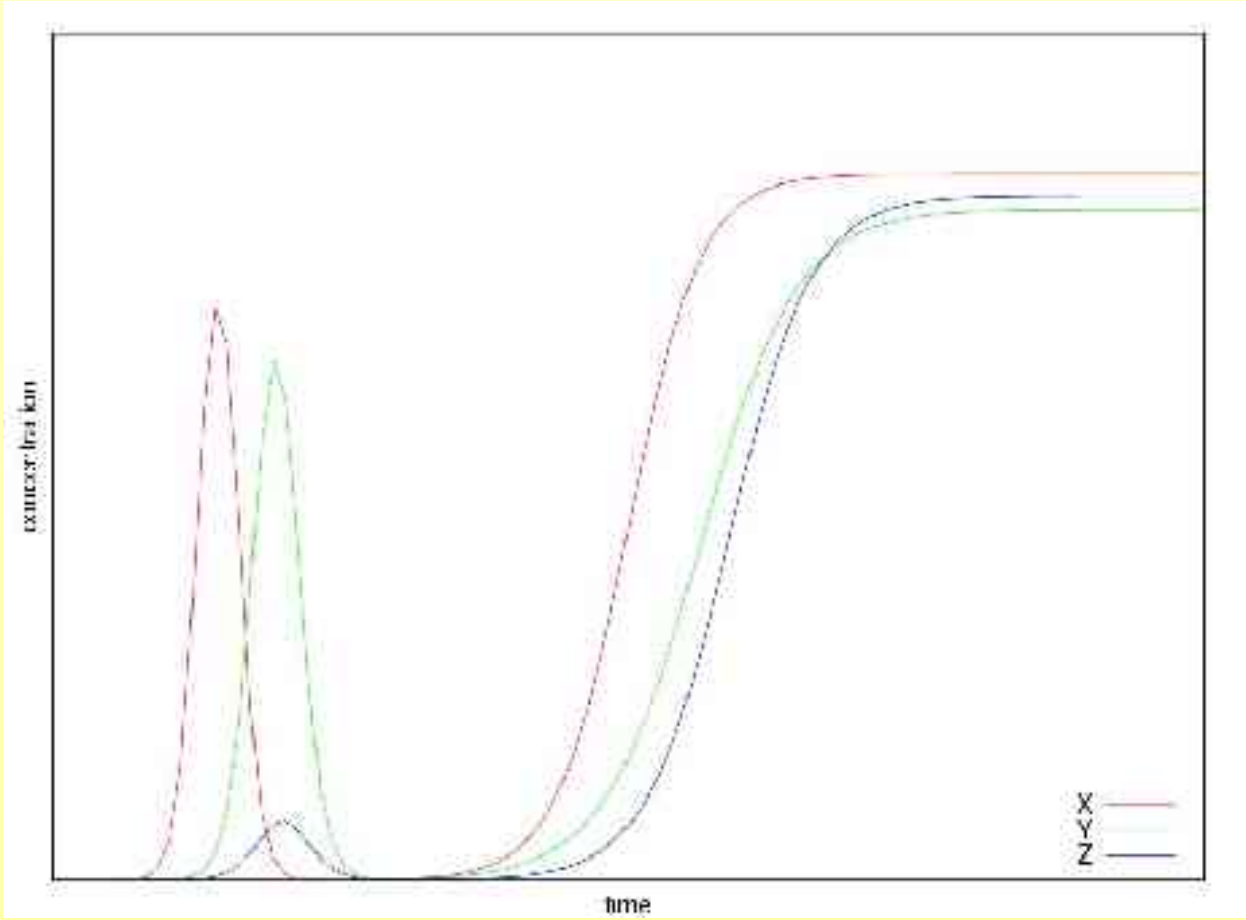
E. coli: *6* in real, 2.5 ± 2 in random, Z=(6-2.5)/2=1.7, p=0.04

Transcriptional regulation

Function of triangles (dynamics preview).

Coherent feed forward:





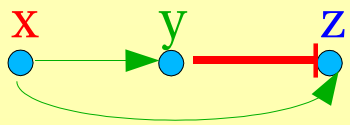
Noise reduction (z indpendent of sudden burst in x)

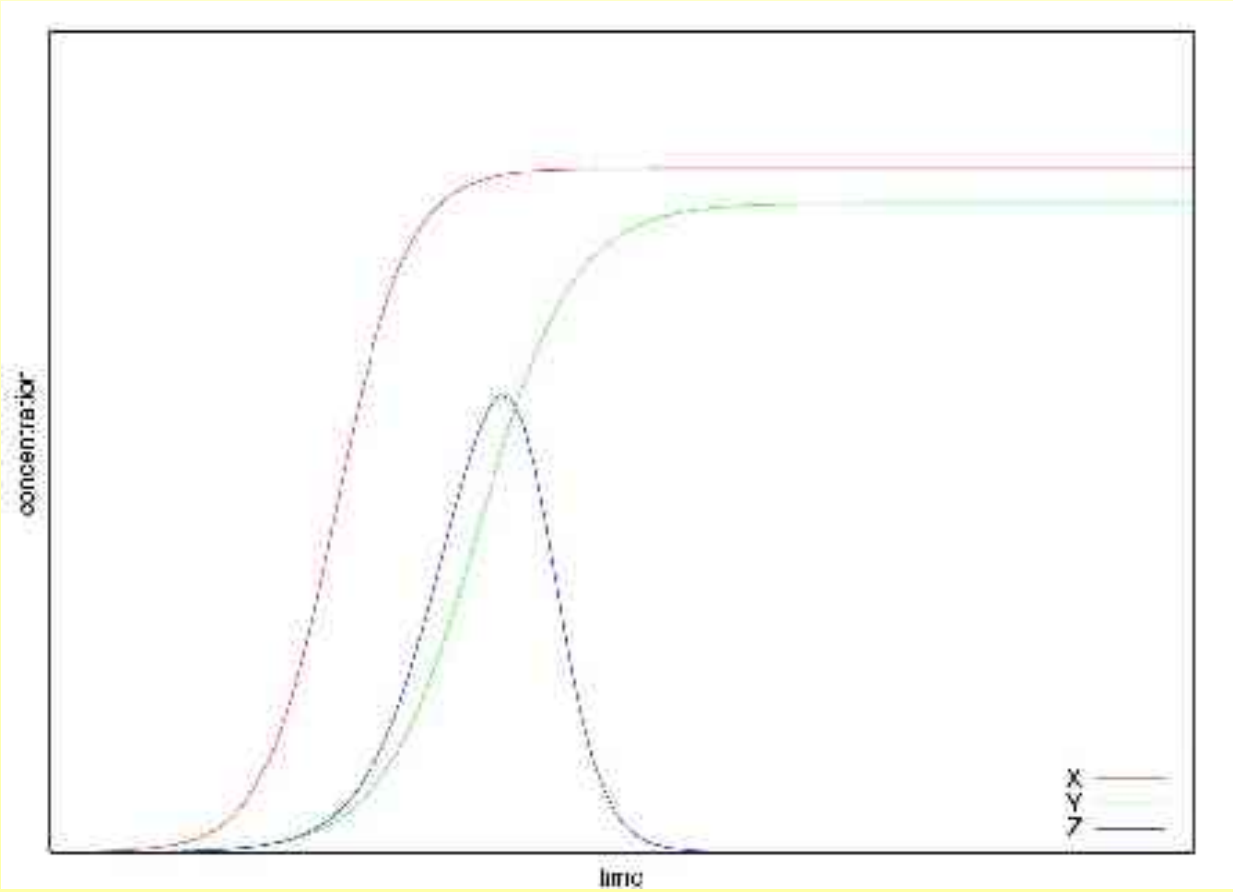
Cannot be achieved with less than 3 nodes (to my knowledge)

Transcriptional regulation

Function of triangles (dynamics preview).

Incoherent feed forward:



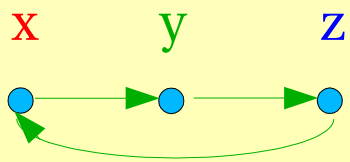


Transient response (z only responds while x increases)
Cannot be achieved with less than 3 nodes (to my knowledge)

Transcriptional regulation

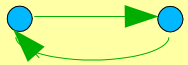
Function of triangles (dynamics preview).

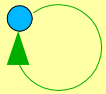
Positive feed backward: Toggle switch.



Temporary external signal activates x (or y or z)
=> all three remain active
Temporary external signal inactivates x (or y or z)
=> all three remain inactive

Simpler alternatives:

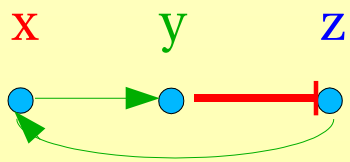




Transcriptional regulation

Function of triangles (dynamics preview).

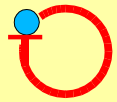
Negative feed backward: Stable state



External signal activates/represses x (or y or z)

=> Negative feedback restores all three to approximately the previous concentrations.

Simpler alternatives:



Transcriptional regulation

Function of triangles (dynamics preview). Summary

Coherent feed forward: Noise reduction
Incoherent feed forward: Transient response
Positive feed backward: Toggle switch
Negative feed backward: Stable state

Note: These functions are just examples.
They depend on time scales

(is x->y->z *much* slower than x->z?)

and logical combinations of signals

(is z activated by [x AND y] or [x OR y]?)

In particular, negative feedback with large time delay becomes an oscillator (dynamics lectures, computer exercise).

Transcriptional regulation

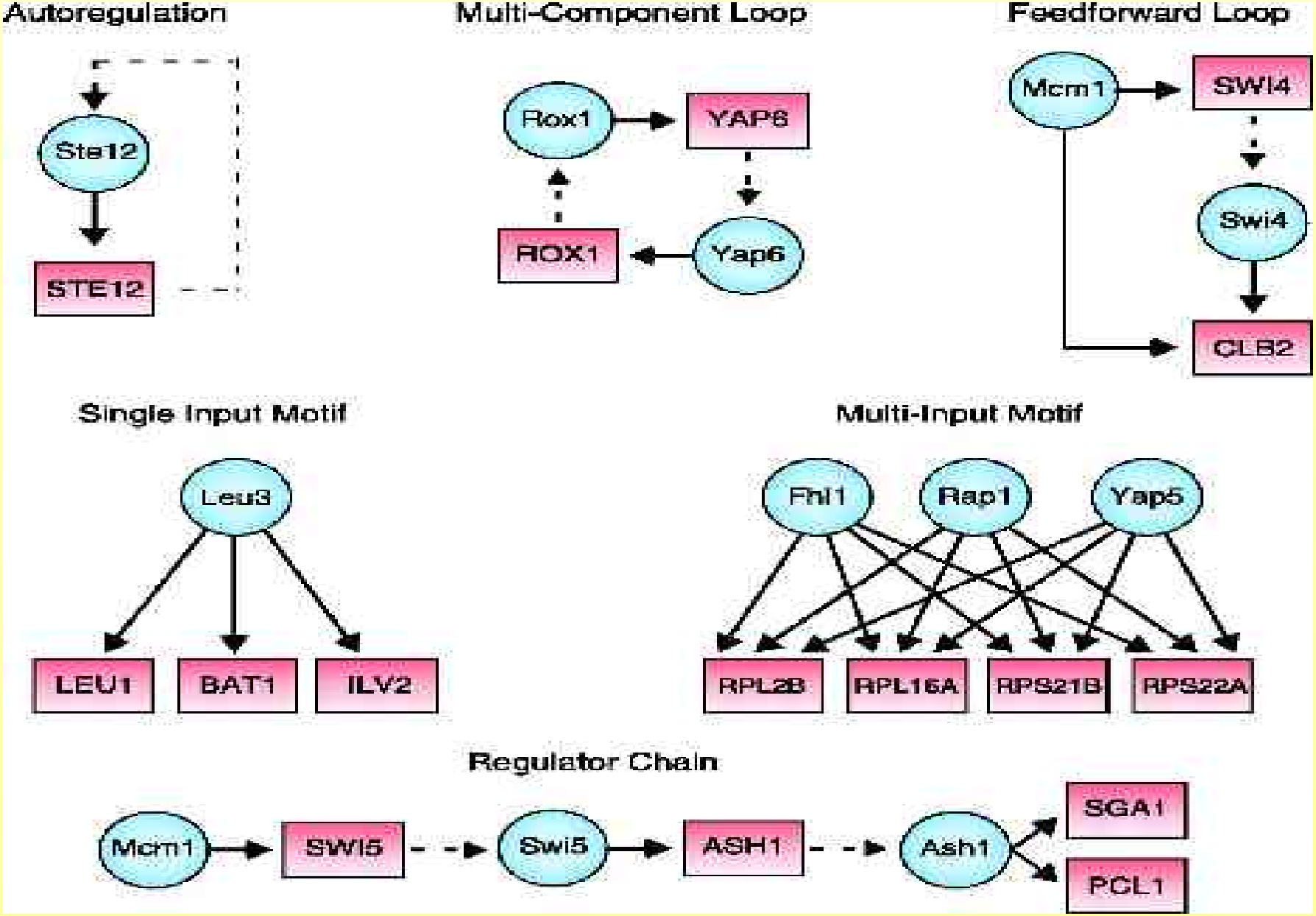
Function of triangles (dynamics preview). Comment

The feed forward triangles solve tasks that cannot be solved by simpler means.

The feed backward triangles can be replaced by simpler (smaller) motifs.

Our systems biology finding
(overrepresentation in *E. coli* transcriptional regulation of feed forward triangles, but not of feed backward)
makes biological sense!

Transcriptional regulation, other motifs (Science 2002, Lee)



Dynamics of autoregulations (Kasper's design II)

Digital or analog networks?
Digital safe but expensive.
“What is conc.?” vs. “Is it above threshold?”
Simple feedback loop in protein interaction
can be both analog and digital.

Modules or mess?
Define module from dinosaur skeleton.
Make analogy to electric circuits. Modules.
Show messy square-root of “evolutionary programming”
Rare, evolved programs often give modules.
Probability for module solution could depend on parameters,
cannot be directly adapted to real biological systems.
Modules safe, but more expensive?