Network Motifs

Simple Building Blocks of Complex Networks

Moritz Bitterling

Universität Freiburg

13.12.2016

outline

introduction

genetics

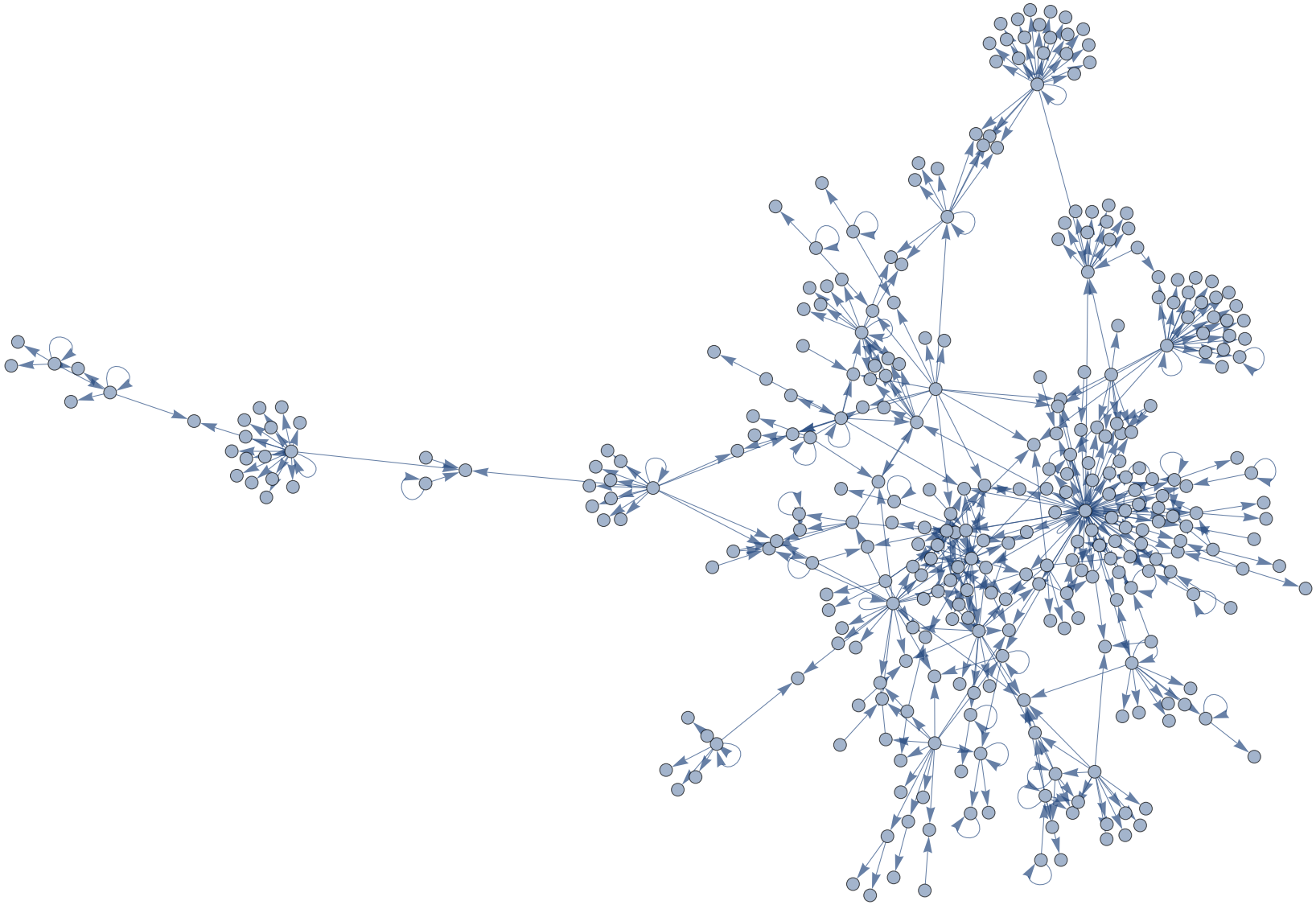
motifs

introduction
- motivation
- methods

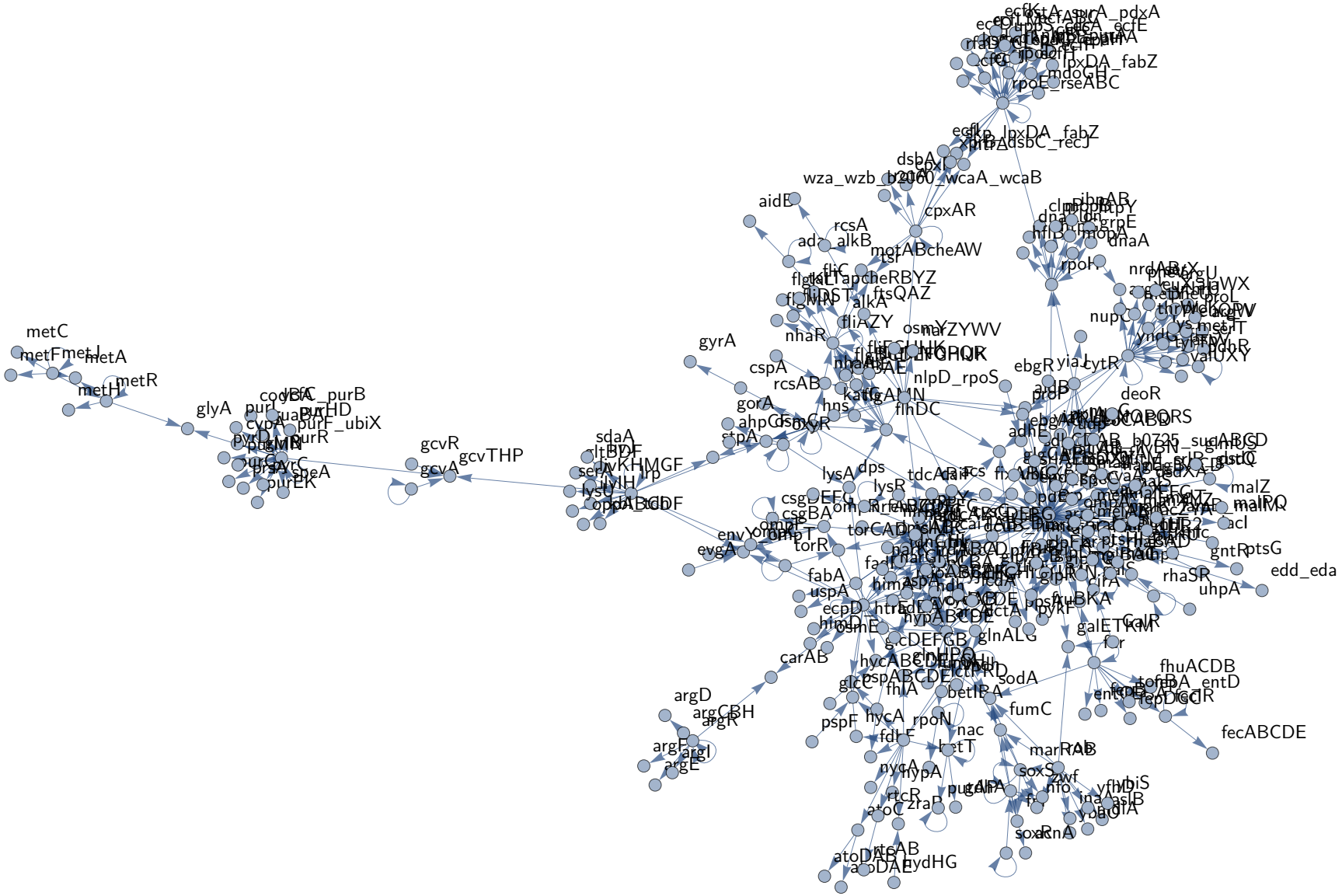
genetics

motifs

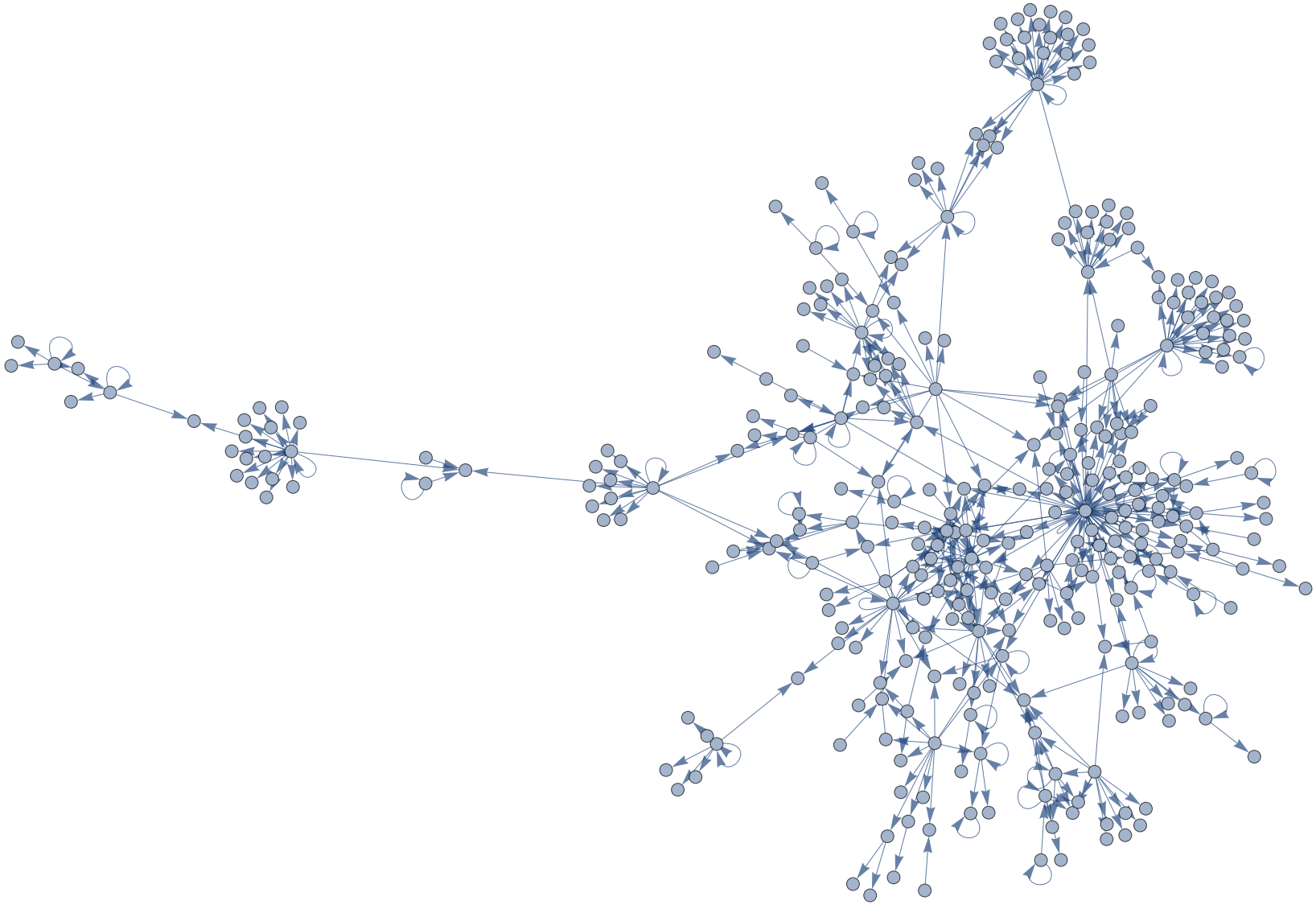
fraction of transcription network of E. coli



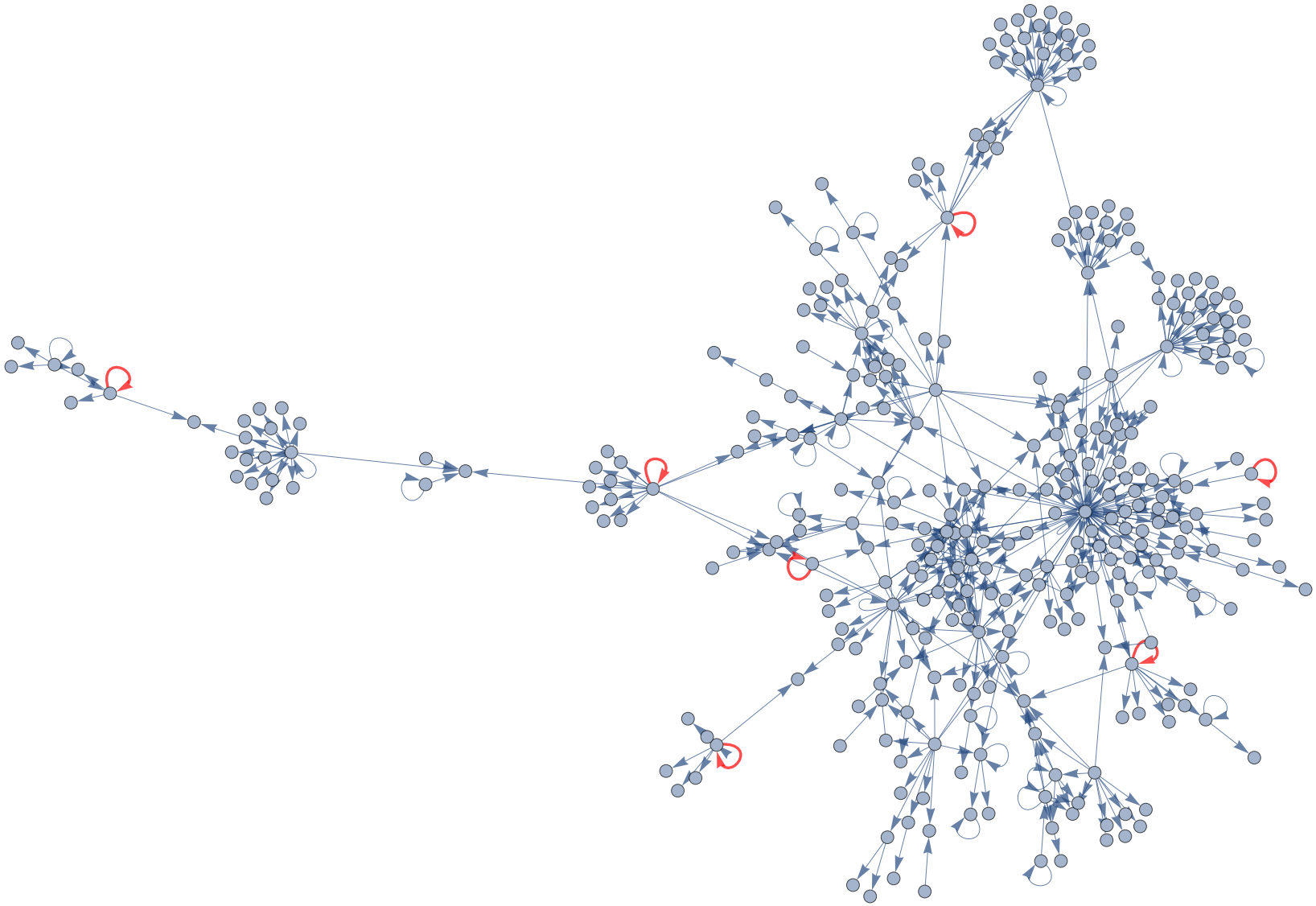
fraction of transcription network of E. coli


metC
metF
metL
metA
metR
metH
glyA
purL
cvpA
purF_ubiX
purR
purEK
gcvR
gcvA
gcvTHP
sdaA
lysU
lysT
oppABCDF
gyrA
cspA
rcsAB
gorA
ahpC
hns
flhDC
nlpD_rpoS
ebgR
yiaJ
lytR
deoR
lysA
dps
lysR
tdcA
csgDEFG
csgBA
torCAD
torR
evgA
fabA
uspA
ecpD
himD
osmE
carAB
argD
argCBH
argR
argF
argI
argE
cpxAR
wza_wzb_b2060_wcaA_wcaB
aidB
rcsA
alkB
motABcheAW
tsr
flhDCcheRBYZ
ftsQAZ
alkA
fliAZY
nhaR
osmYZYWV
narP
fliFGHIJK
nhaA
ydeP
ydeO
ydeH
ydeI
ydeS
ydeT
ydeU
ydeV
ydeW
ydeX
ydeY
ydeZ
ydiV
ydiU
ydiT
ydiS
ydiR
ydiQ
ydiP
ydiO
ydiN
ydiM
ydiL
ydiK
ydiJ
ydiI
ydiH
ydiG
ydiF
ydiE
ydiD
ydiC
ydiB
ydiA
ydhX
ydhW
ydhV
ydhU
ydhT
ydhS
ydhR
ydhQ
ydhP
ydhO
ydhN
ydhM
ydhL
ydhK
ydhJ
ydhI
ydhH
ydhG
ydhF
ydhE
ydhD
ydhC
ydhB
ydhA
ydgX
ydgW
ydgV
ydgU
ydgT
ydgS
ydgR
ydgQ
ydgP
ydgO
ydgN
ydgM
ydgL
ydgK
ydgJ
ydgI
ydgH
ydgG
ydgF
ydgE
ydgD
ydgC
ydgB
ydgA
ydfX
ydfW
ydfV
ydfU
ydfT
ydfS
ydfR
ydfQ
ydfP
ydfO
ydfN
ydfM
ydfL
ydfK
ydfJ
ydfI
ydfH
ydfG
ydfF
ydfE
ydfD
ydfC
ydfB
ydfA
ydeX
ydeW
ydeV
ydeU
ydeT
ydeS
ydeR
ydeQ
ydeP
ydeO
ydeN
ydeM
ydeL
ydeK
ydeJ
ydeI
ydeH
ydeG
ydeF
ydeE
ydeD
ydeC
ydeB
ydeA
ydcX
ydcW
ydcV
ydcU
ydcT
ydcS
ydcR
ydcQ
ydcP
ydcO
ydcN
ydcM
ydcL
ydcK
ydcJ
ydcI
ydcH
ydcG
ydcF
ydcE
ydcD
ydcC
ydcB
ydcA
ydbX
ydbW
ydbV
ydbU
ydbT
ydbS
ydbR
ydbQ
ydbP
ydbO
ydbN
ydbM
ydbL
ydbK
ydbJ
ydbI
ydbH
ydbG
ydbF
ydbE
ydbD
ydbC
ydbB
ydbA
ydaX
ydaW
ydaV
ydaU
ydaT
ydaS
ydaR
ydaQ
ydaP
ydaO
ydaN
ydaM
ydaL
ydaK
ydaJ
ydaI
ydaH
ydaG
ydaF
ydaE
ydaD
ydaC
ydaB
yda

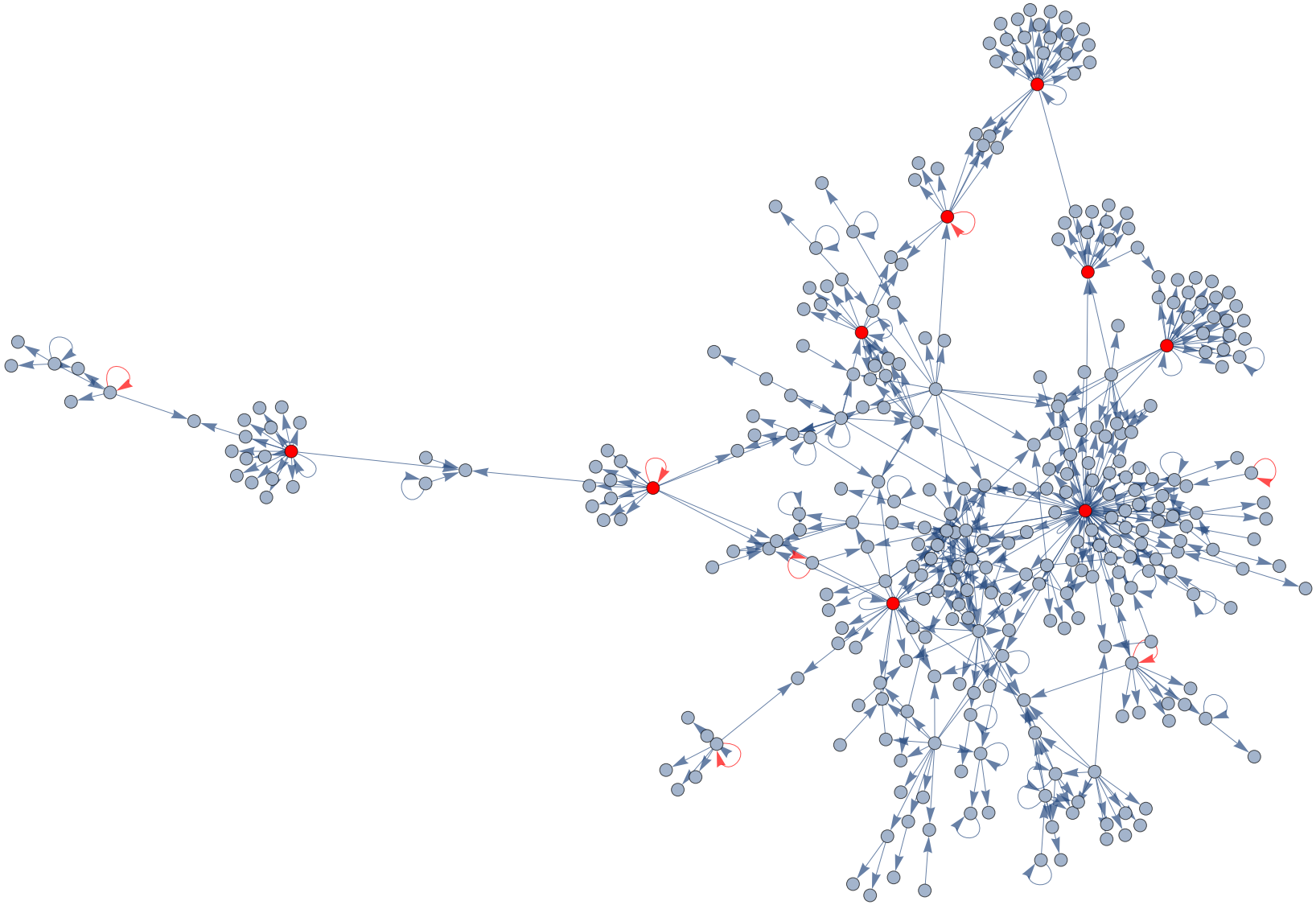
fraction of transcription network of E. coli



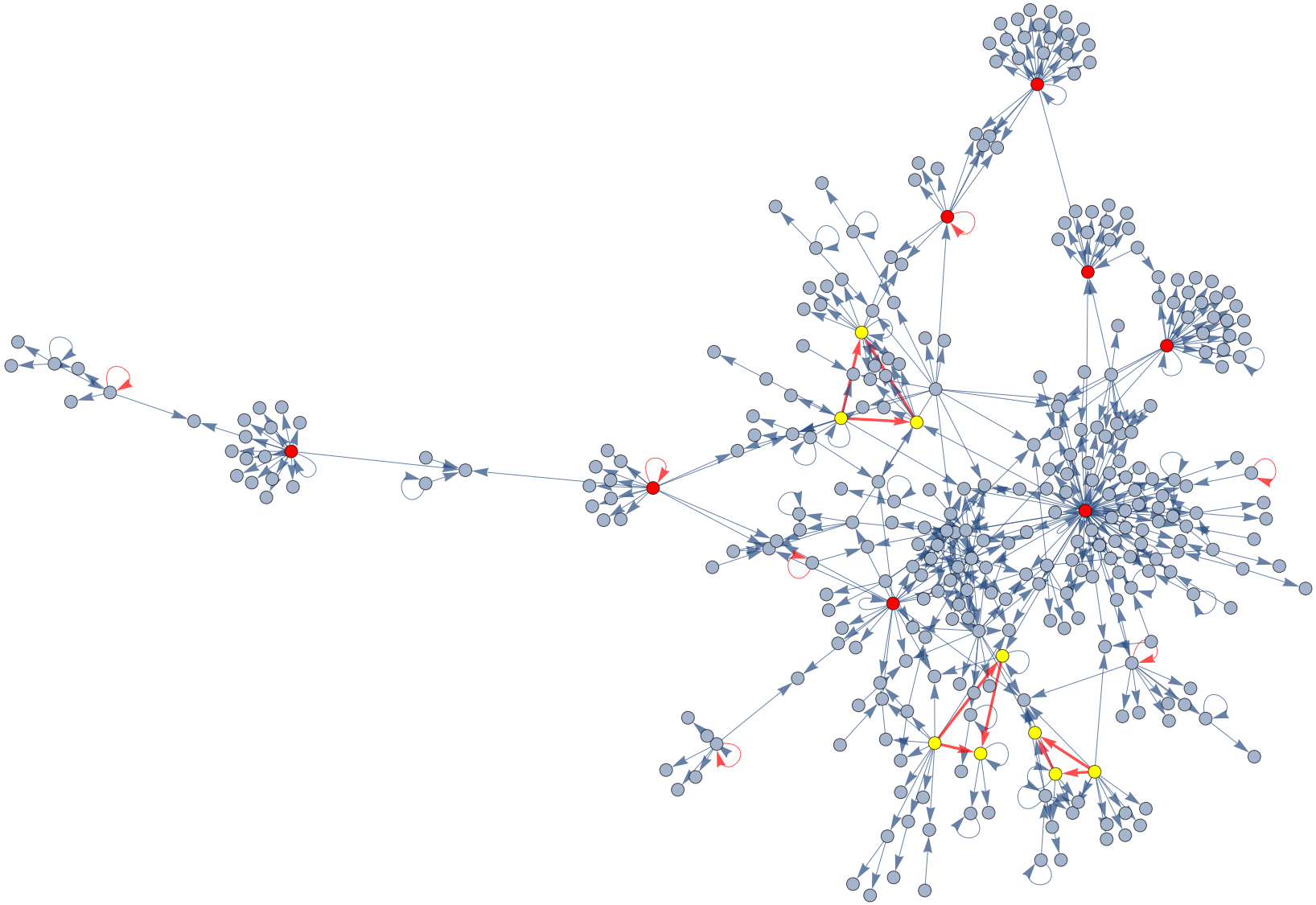
transcription network - autoregulation



transcription network - outstanding nodes



transcription network - loops



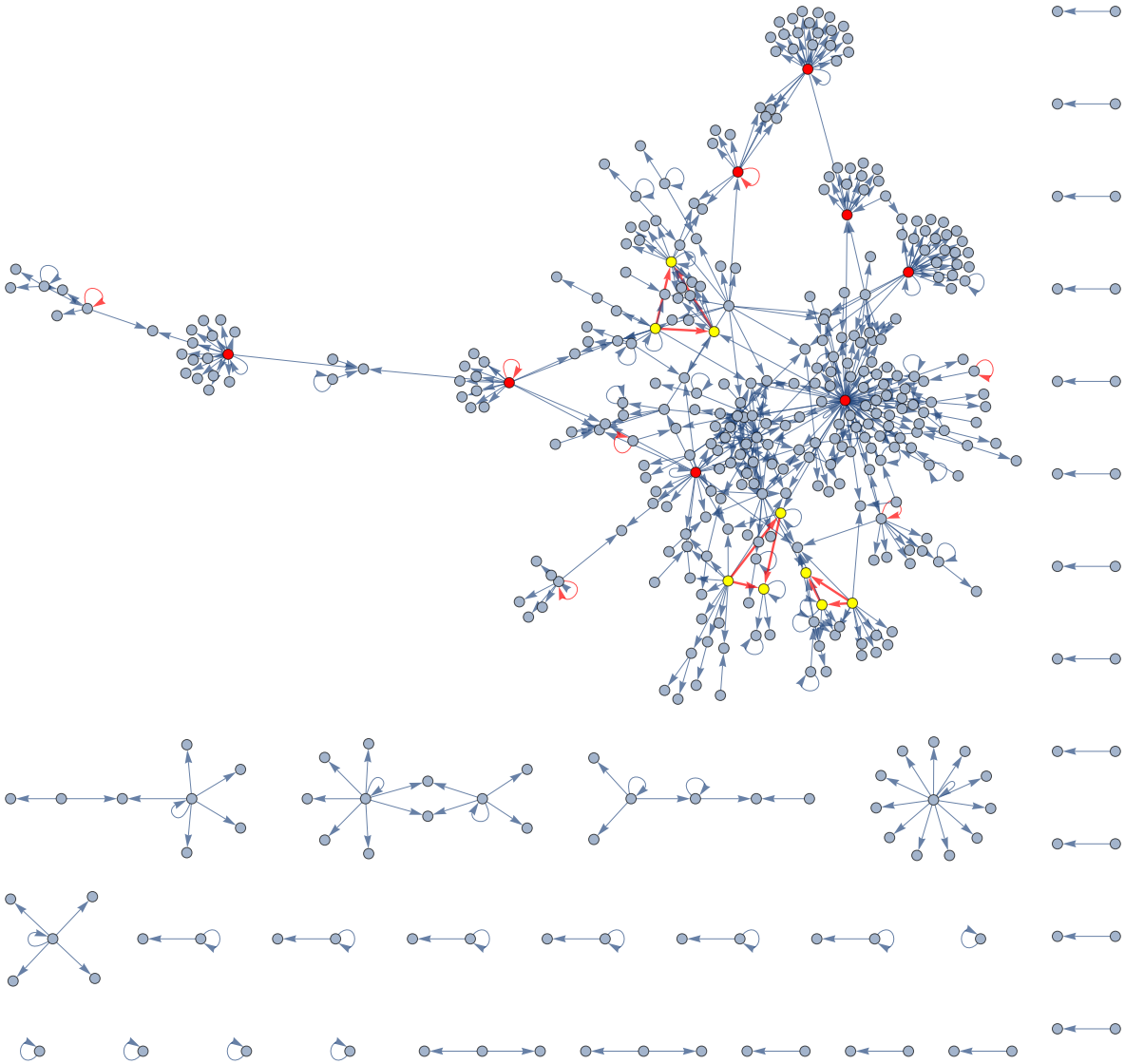
Is there a statistically significant aggregation of local patterns or "motifs" in the network?

Compare to a randomized network
$\rightarrow$ YES!

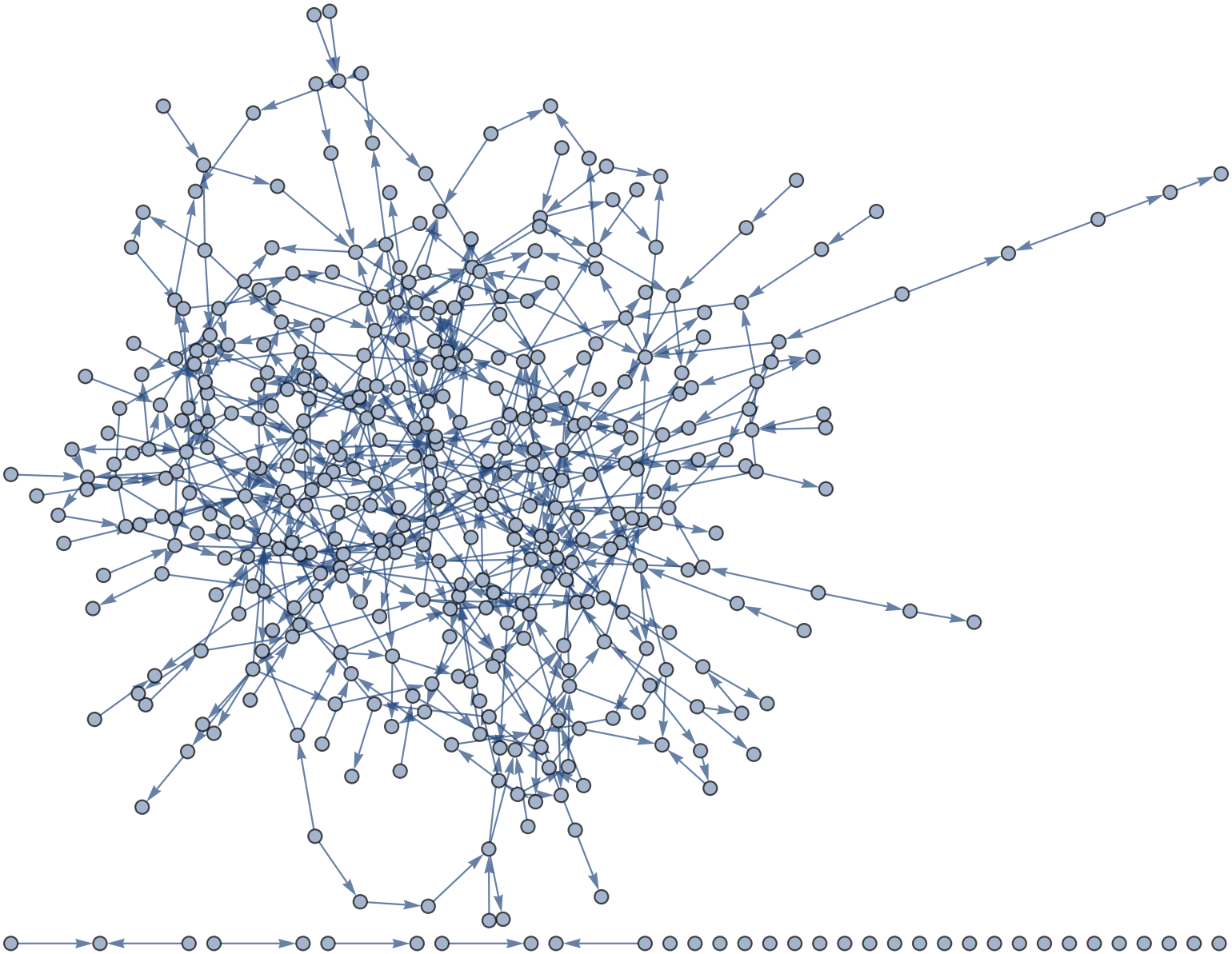
Which are the motifs that evolution prefers?

What are their functions in the network?

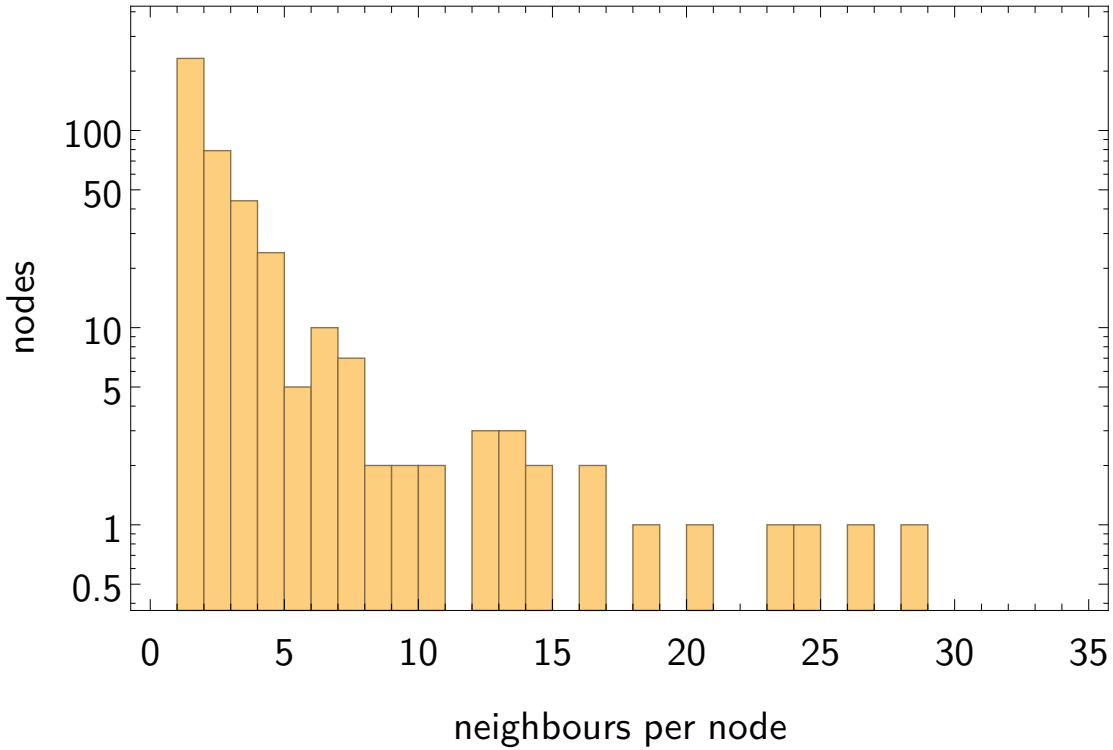
fraction of transcription network of E. coli



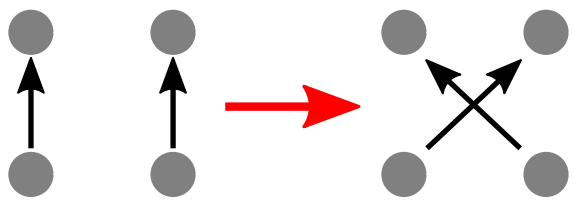
random network



histogram of E.coli transcription network



Randomization: Choose two random edges, swap target nodes
$\rightarrow$ the histogram is preserved



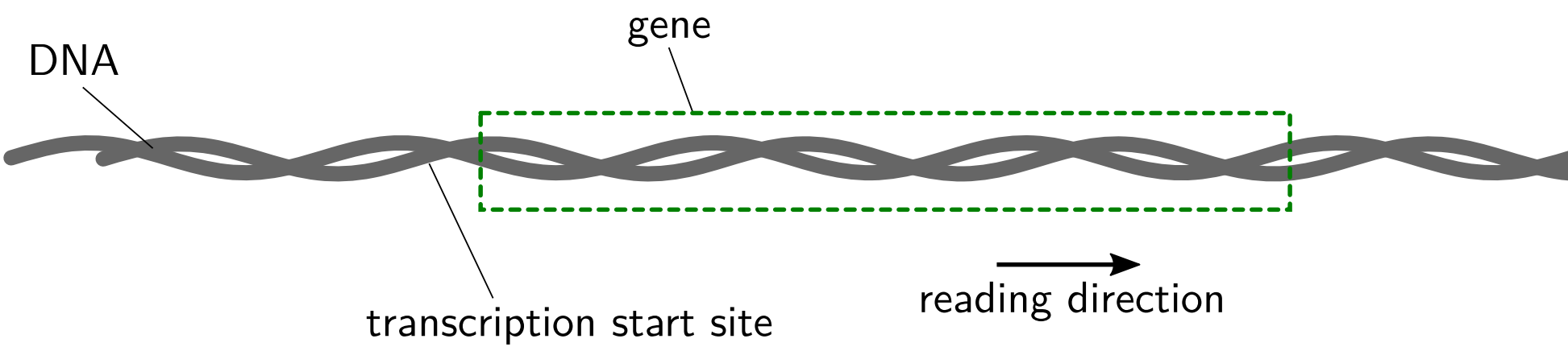
introduction

genetics
- transcription
- translation
- dynamics

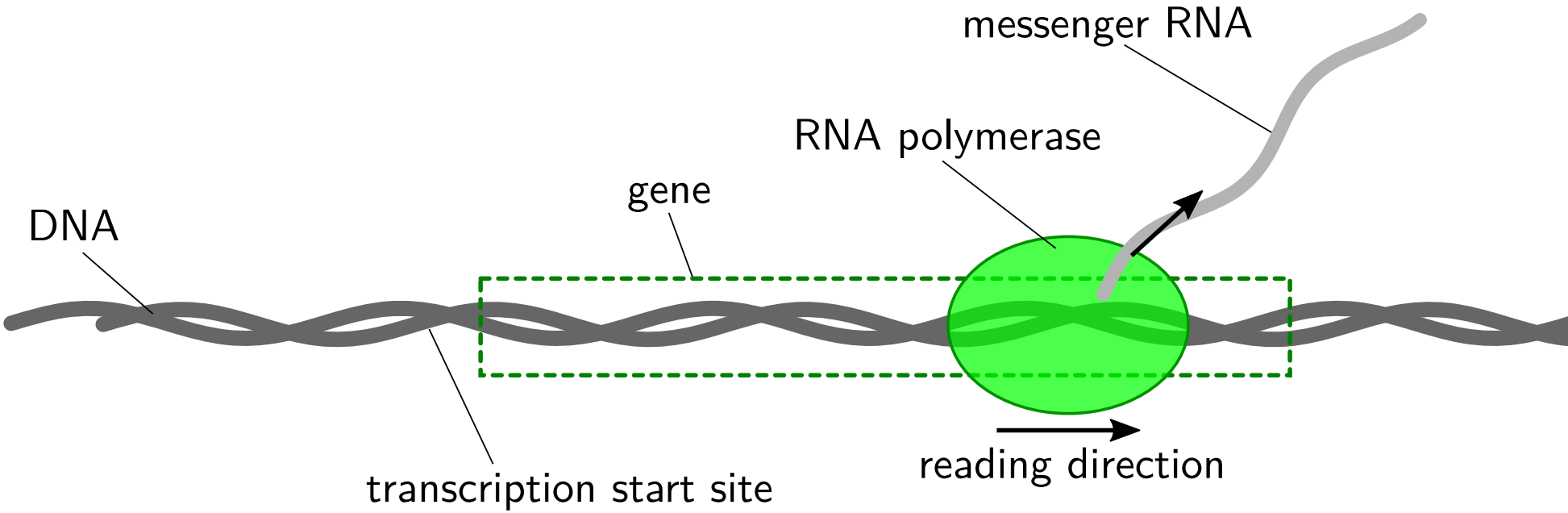
motifs

gene expression - transcription

gene expression - transcription

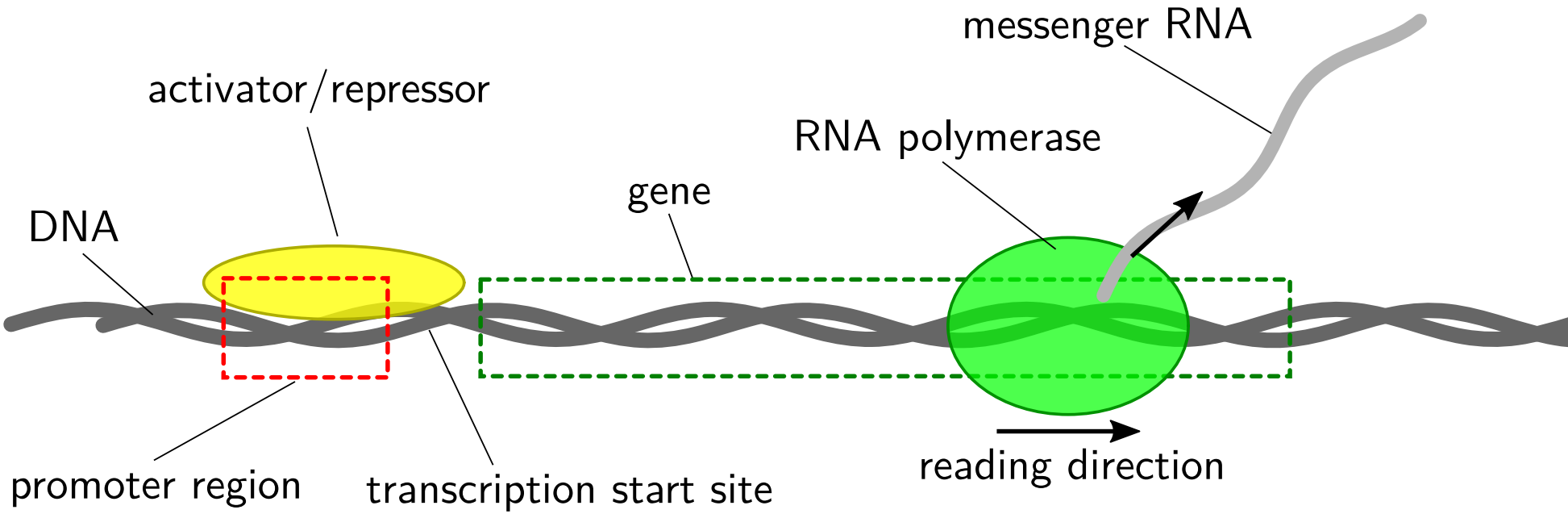


gene expression - transcription



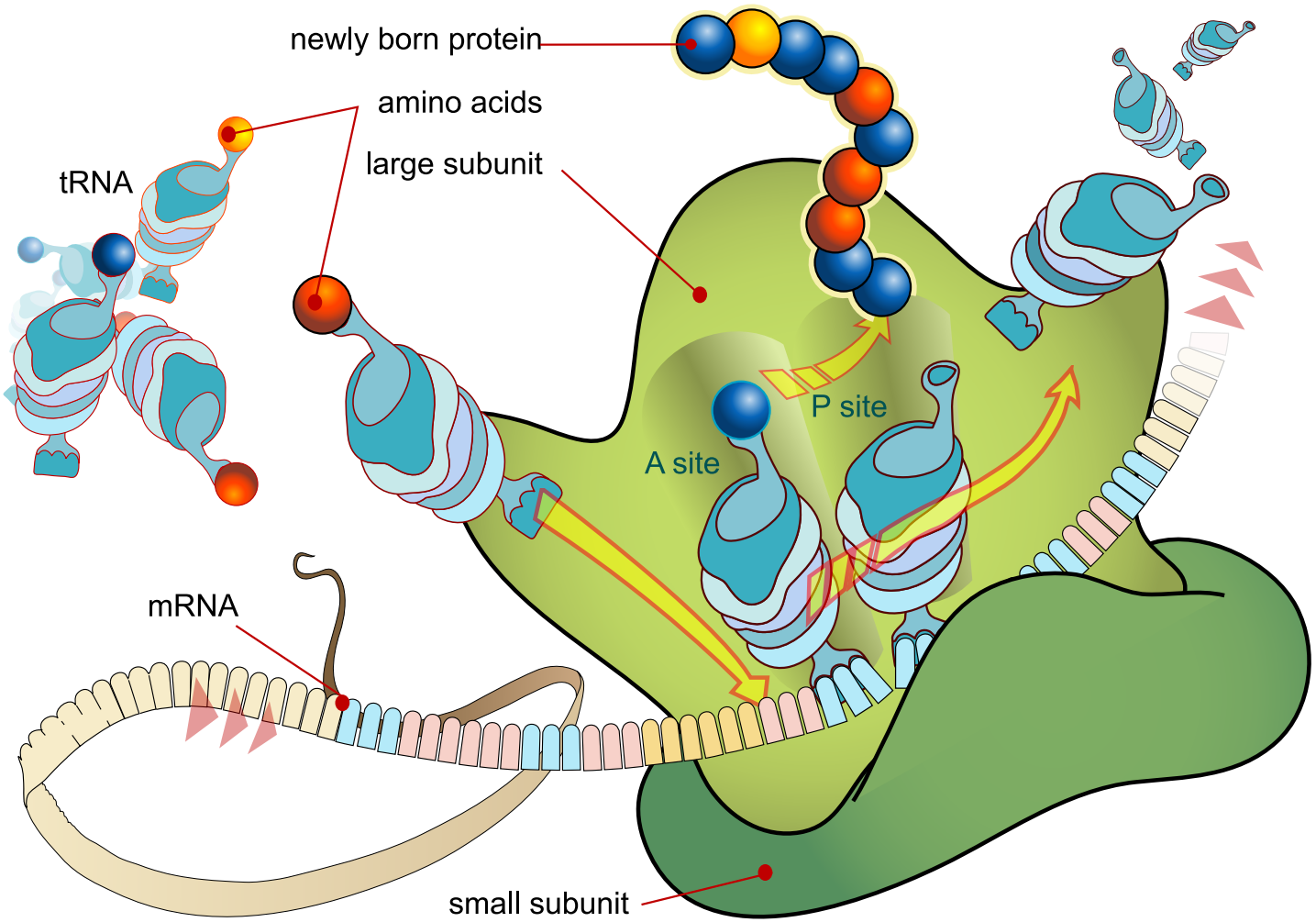
- RNAp transcribes DNA to mRNA

gene expression - transcription



- RNAp transcribes DNA to mRNA
- Activator/repressor interacts with transcription start site
  $\rightarrow$ Enhances/inhibits attachment of RNAp

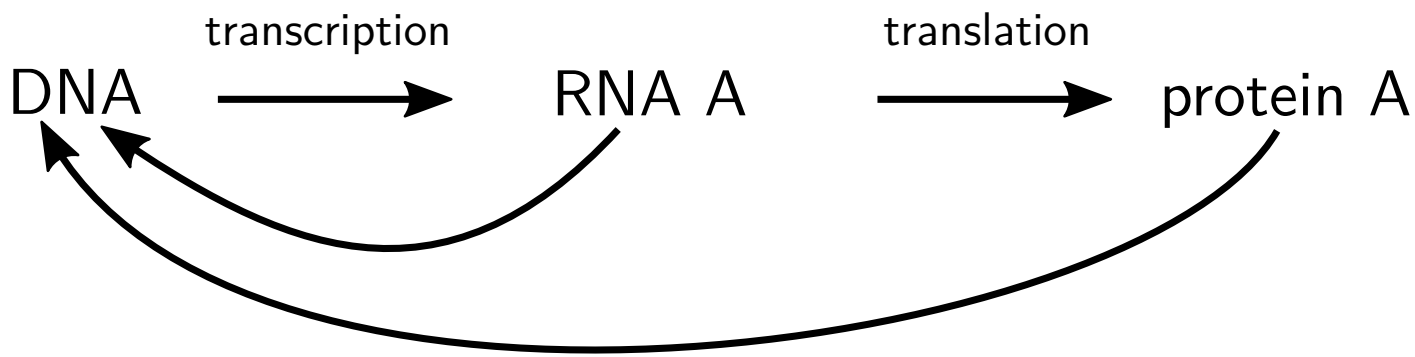
gene expression - translation of mRNA to protein



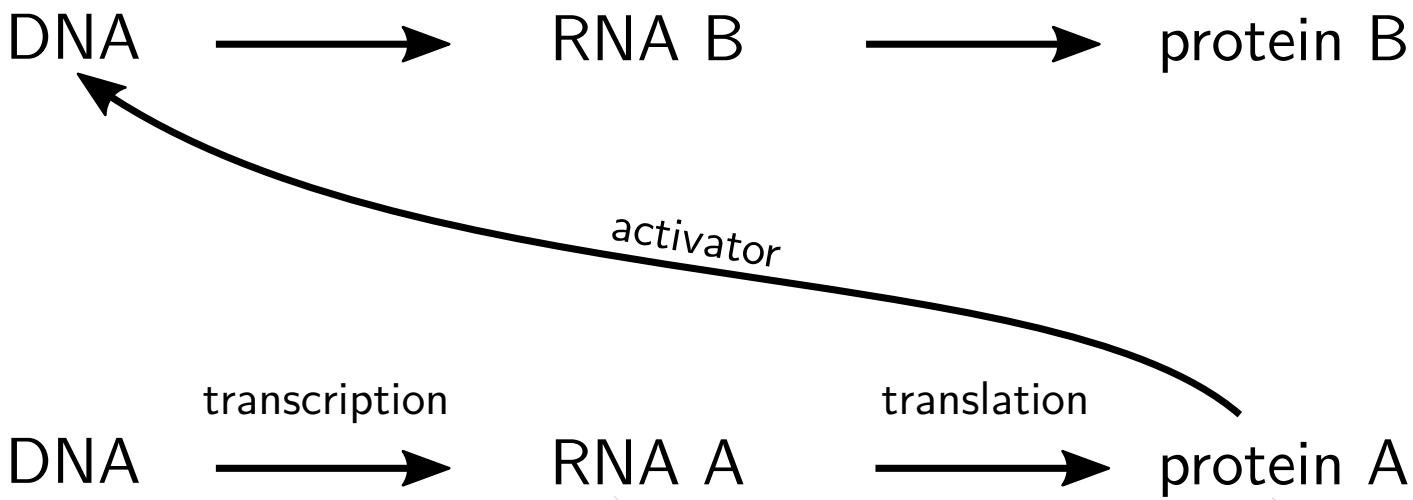
en.wikipedia.org/wiki/Translation_(biology)

regulation of gene expression

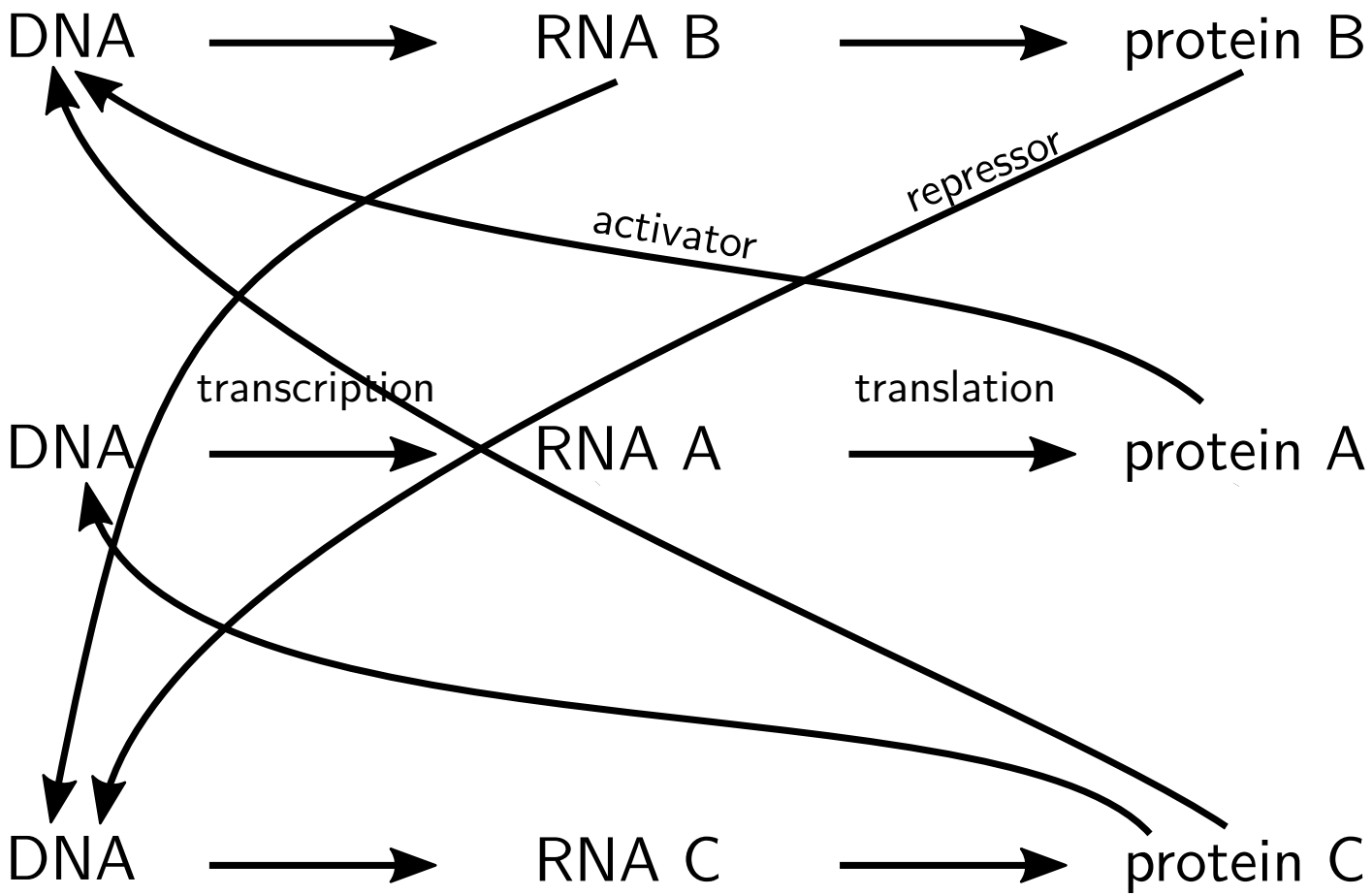
regulation of gene expression



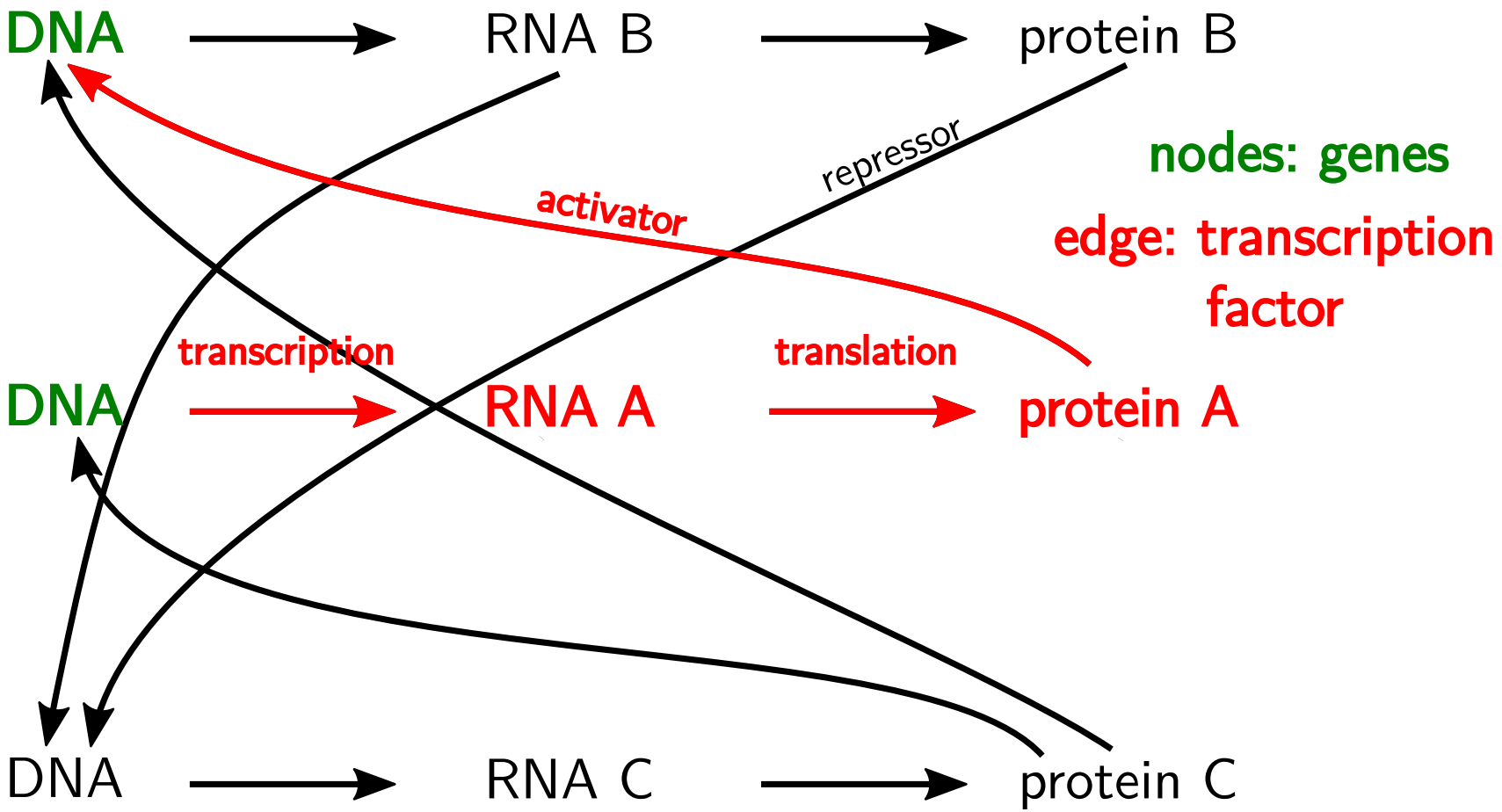
regulation of gene expression



regulation of gene expression



regulation of gene expression



dynamics of simple gene regulation

Transcription factor X regulates expression of protein Y: $X \rightarrow Y$

Simple assumptions:

- If X is in active form X^*, Y is produced at a constant rate β
- Degradation rate of Y is α

$$\Rightarrow \frac{dY}{dt} = \beta - \alpha Y$$

$$\text{Stable state:} \quad \frac{dY}{dt} \stackrel{!}{=} 0 \qquad \Rightarrow \qquad Y_{st} = \frac{\beta}{\alpha}$$

dynamics of simple gene regulation

Transcription factor X regulates expression of protein Y: $X \rightarrow Y$

Simple assumptions:

- If X is in active form X^*, Y is produced at a constant rate β
- Degradation rate of Y is α

$$\Rightarrow \frac{dY}{dt} = \beta - \alpha Y$$

$$\text{Stable state:} \quad \frac{dY}{dt} \stackrel{!}{=} 0 \qquad \Rightarrow \qquad Y_{st} = \frac{\beta}{\alpha}$$

dynamics of simple gene regulation

Transcription factor X regulates expression of protein Y: $X \rightarrow Y$

Simple assumptions:

- If X is in active form X^*, Y is produced at a constant rate β
- Degradation rate of Y is α

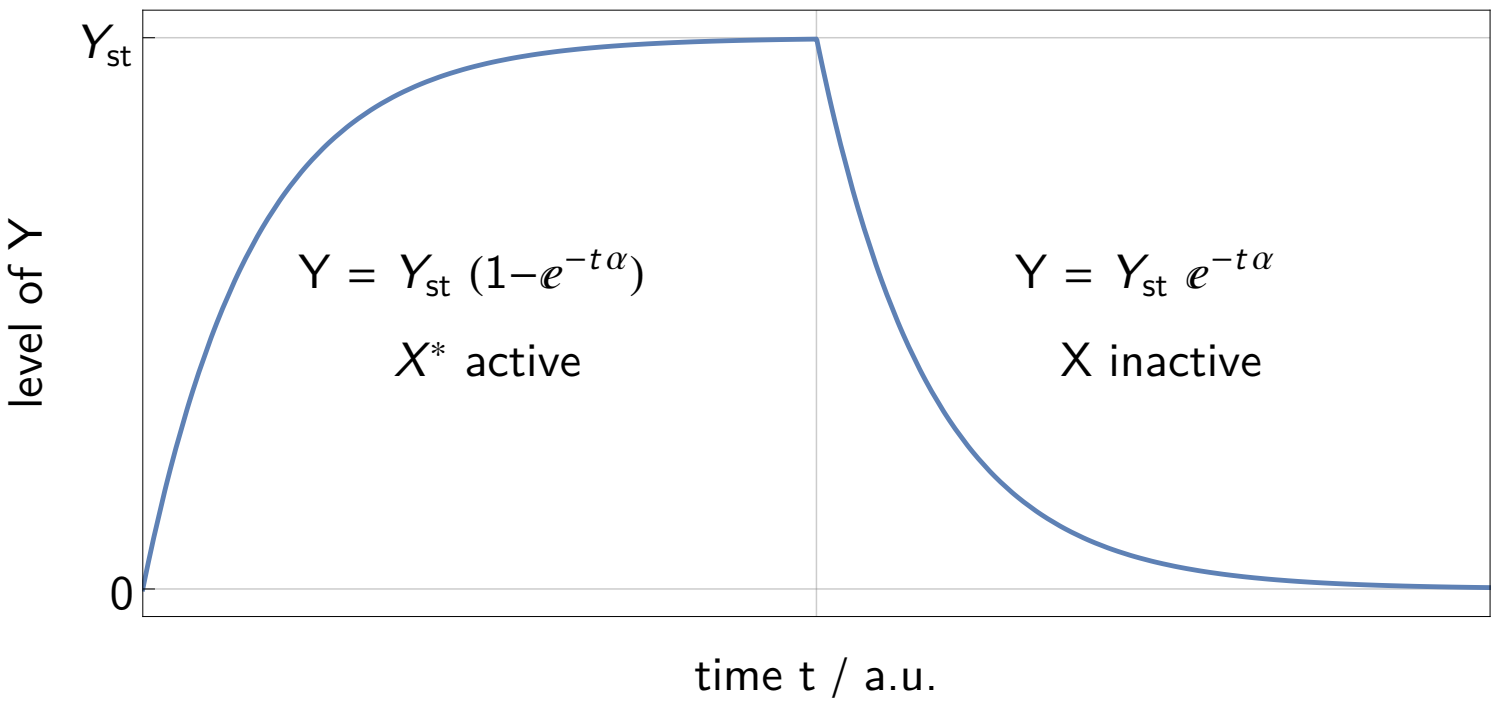
$$\Rightarrow \frac{dY}{dt} = \beta - \alpha Y$$

Stable state: $\frac{dY}{dt} \stackrel{!}{=} 0 \quad \Rightarrow \quad Y_{st} = \frac{\beta}{\alpha}$

dynamics of simple gene regulation

$$\frac{\mathrm{dY}}{\mathrm{d}t} = \beta - \alpha \mathrm{Y}$$

$$\mathrm{Y_{st}} = \frac{\beta}{\alpha}$$



Hill-function

More realistic model: Rate of production of Y is a function of X^*

f[X^*] has three parameters $\rightarrow$ each edge carries three numbers

- K: Level of X^* to significantly activate expression
- β: Maximal expression level
- n: Cooperativity: Number of molecules needed for activation

Rate of production of Y: $f[X^*] = \beta \frac{{X^*}^n}{K^n + {X^*}^n}$

Hill-function

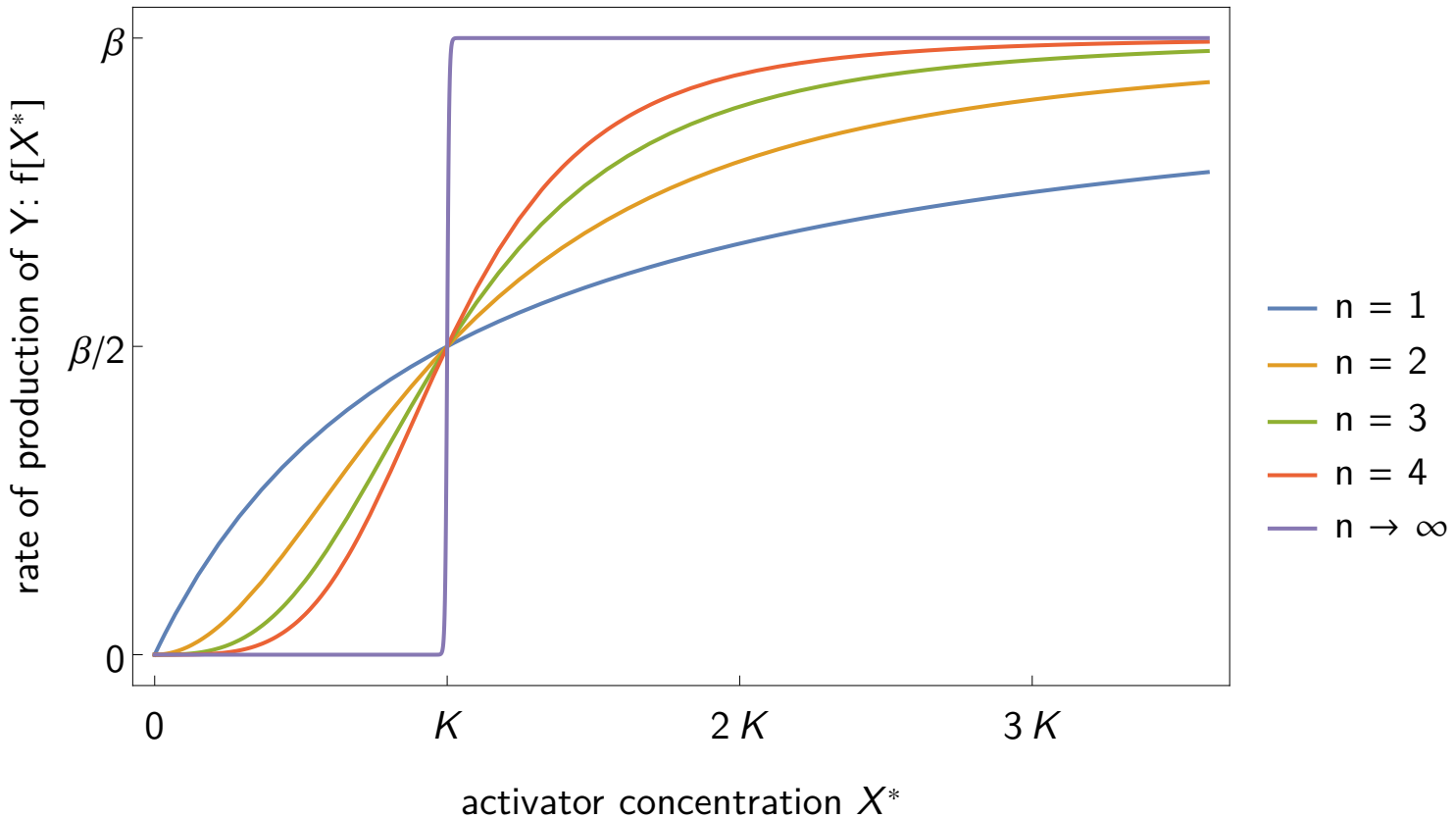
More realistic model: Rate of production of Y is a function of X^*

f[X^*] has three parameters $\rightarrow$ each edge carries three numbers

- K: Level of X^* to significantly activate expression
- β: Maximal expression level
- n: Cooperativity: Number of molecules needed for activation

$$\text{Rate of production of Y: } f[X^*] = \beta \frac{{X^*}^n}{K^n + {X^*}^n}$$

Hill-function



$$f[X^*] = \beta \frac{{X^*}^n}{K^n + {X^*}^n}$$

Hill-function

Parameters can be tuned during evolution:

K: Mutations of binding site in the promoter area

β: Mutations in the RNAp binding site

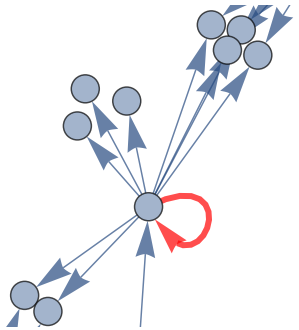
$$\text{Rate of production of Y: } f[X^*] = \beta \frac{{X^*}^n}{K^n + {X^*}^n}$$

introduction

genetics

motifs
- autoregulation
- feed forward loop
- single-input module
- dense overlapping regulon
- occurrence of motifs in various networks

autoregulation



Probability of a self-edge in random network with N nodes:

$$p_{\text{self}} = \frac{1}{N}$$

Probability of k self-edges in a random network with N nodes and E edges:

$$P(k) = \binom{E}{k} \cdot p_{\text{self}}^{k} \cdot (1 - p_{\text{self}})^{E-k}$$

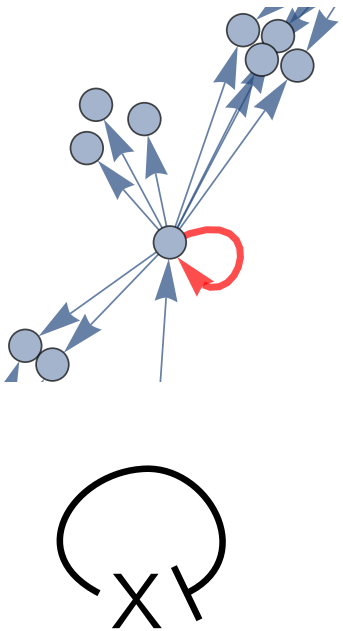
The expectation value for the number of self-edges is

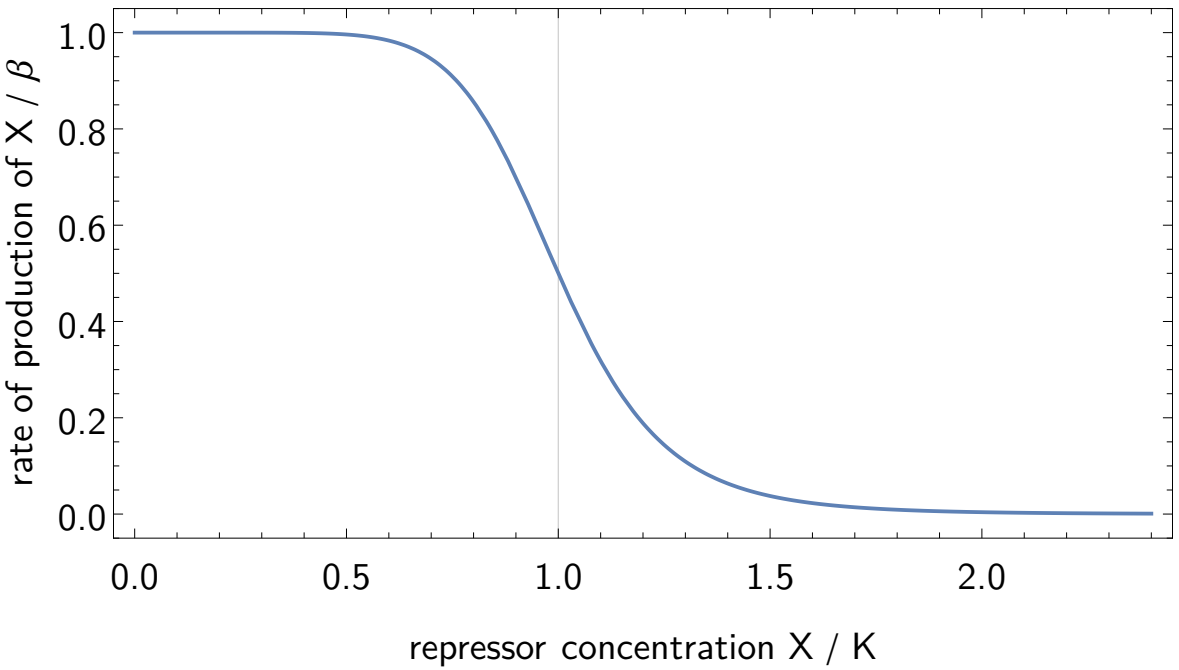
$$< N_{\text{self}} >= E \cdot p_{\text{self}} = \frac{E}{N}$$

For the E.coli network this yields $\frac{E}{N} = \frac{519}{424} = 1.2$ self-edges

The real network has 40 self-edges $\rightarrow$ high statistical significance

negative autoregulation



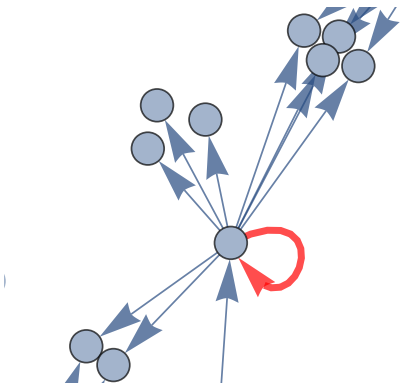


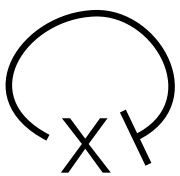
- System with simple regulation: $Y_{st} = \frac{\beta}{\alpha}$
$\rightarrow$ unstable, many factors influence β and α

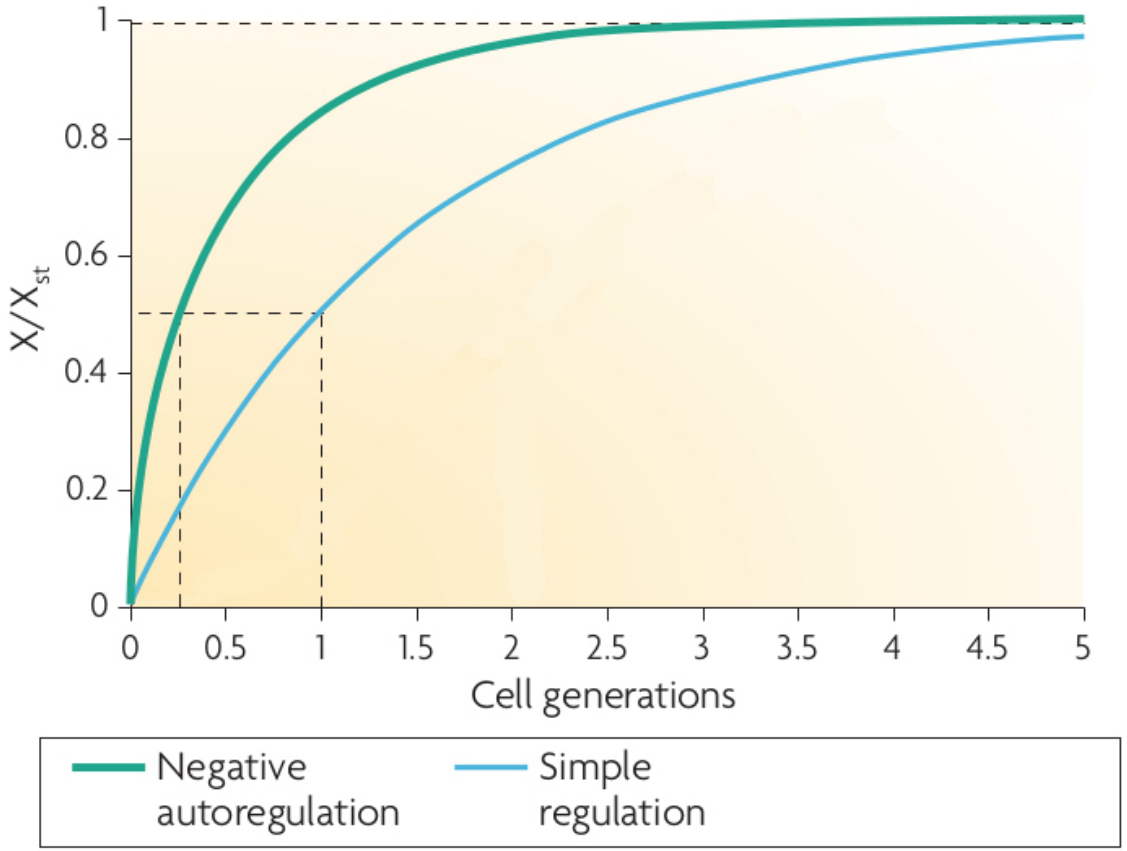
- System with negative autoregulation: $Y_{st} \approx K$
$\rightarrow$ stable, K is specified by strength of chemical bonds

$\Rightarrow$ Noise suppression

negative autoregulation



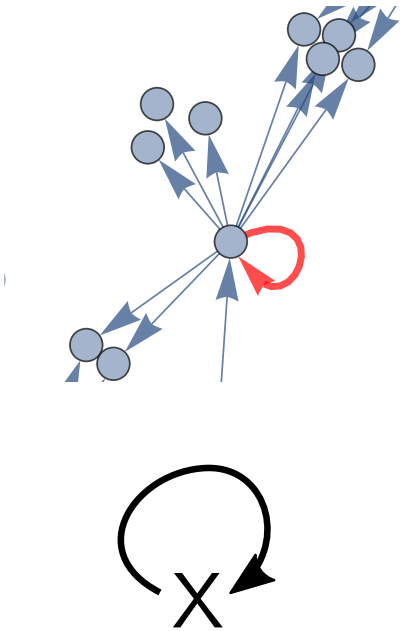


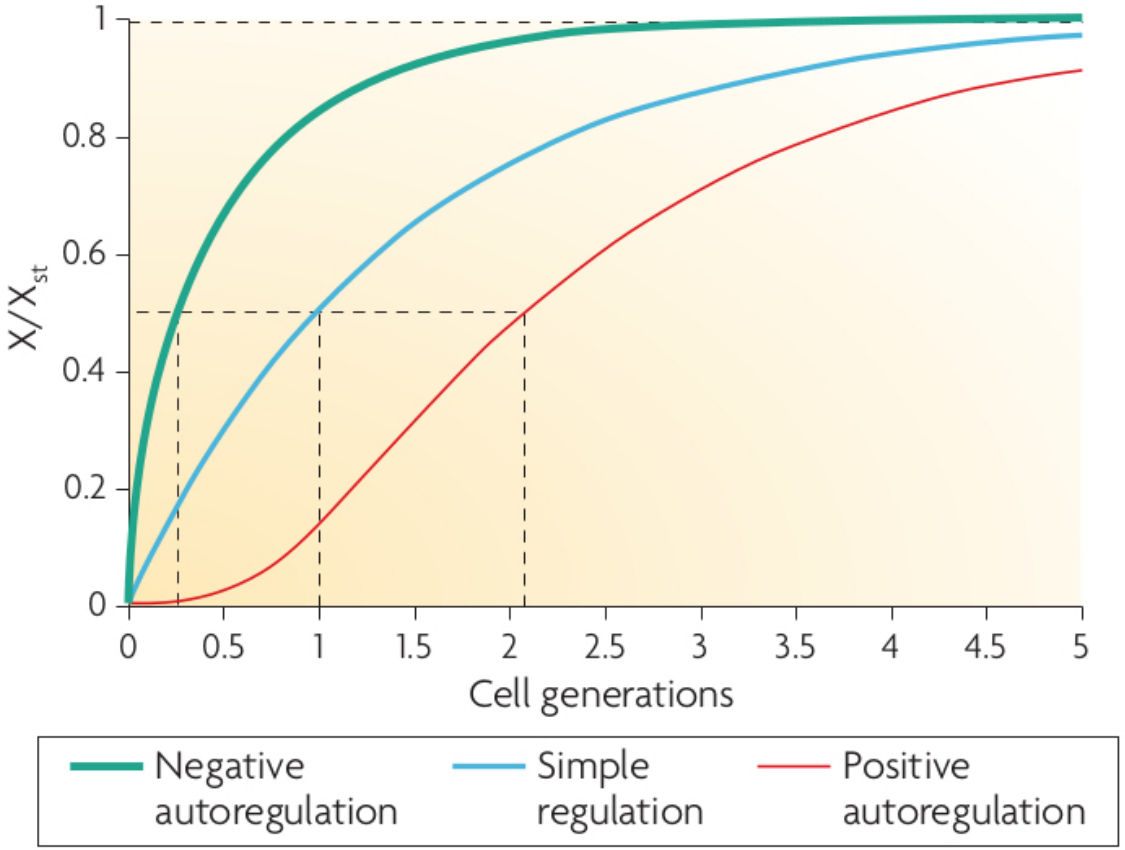


- High production rate β can cause strong initial rise
- High autorepression leads to saturation at stable level K

$\Rightarrow$ response acceleration

positive autoregulation





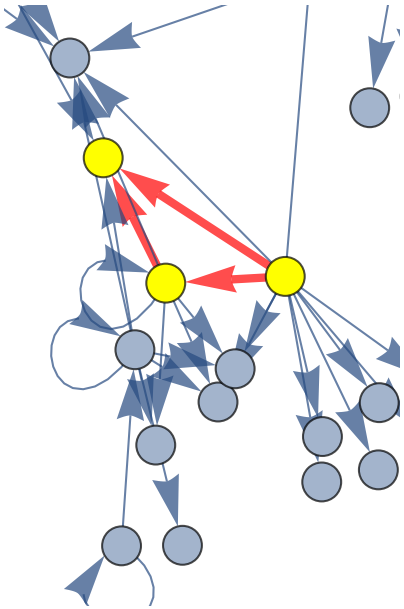
$\Rightarrow$ higher response time

$\Rightarrow$ noise amplifying

Why is noise good for a cell?

$\rightarrow$ diversity

feed forward loops



Coherent FFL

Coherent type 1

Coherent type 2

Coherent type 3

Coherent type 4

Incoherent FFL

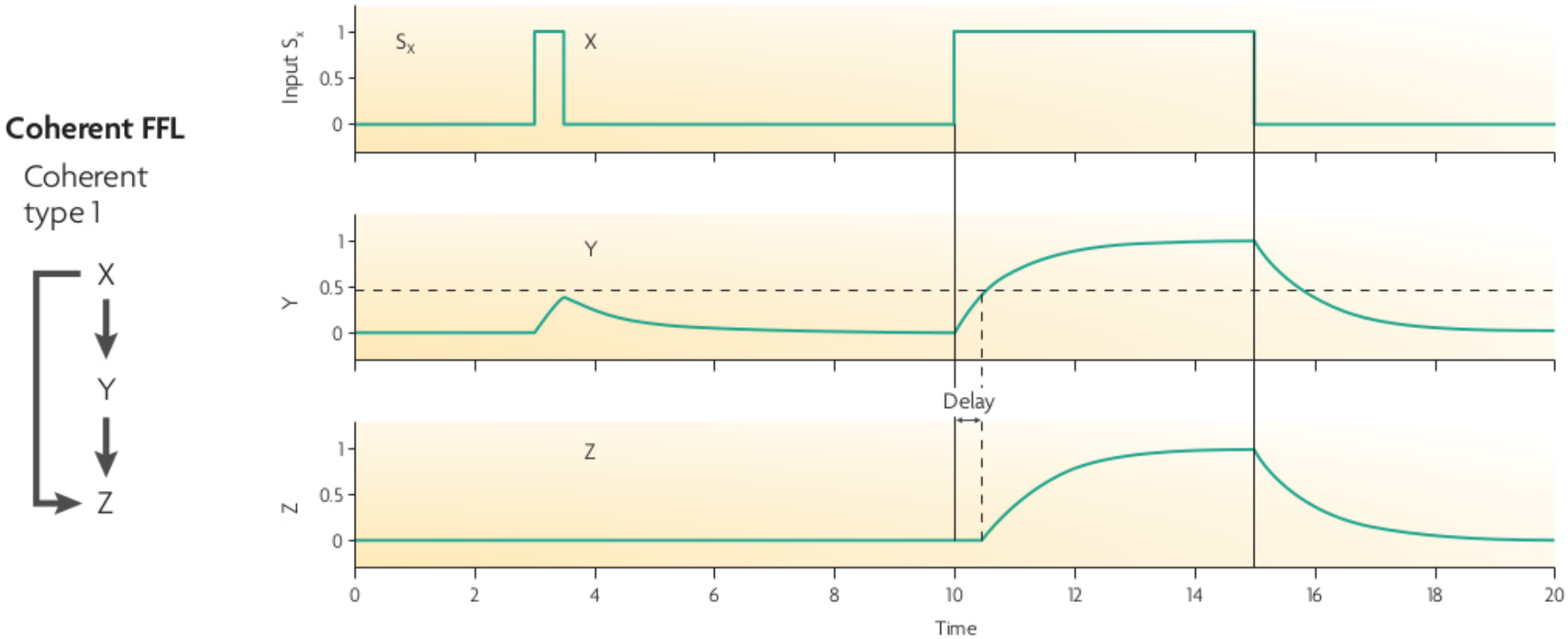
Incoherent type 1

Incoherent type 2

Incoherent type 3

Incoherent type 4

coherent type 1 feed forward loop



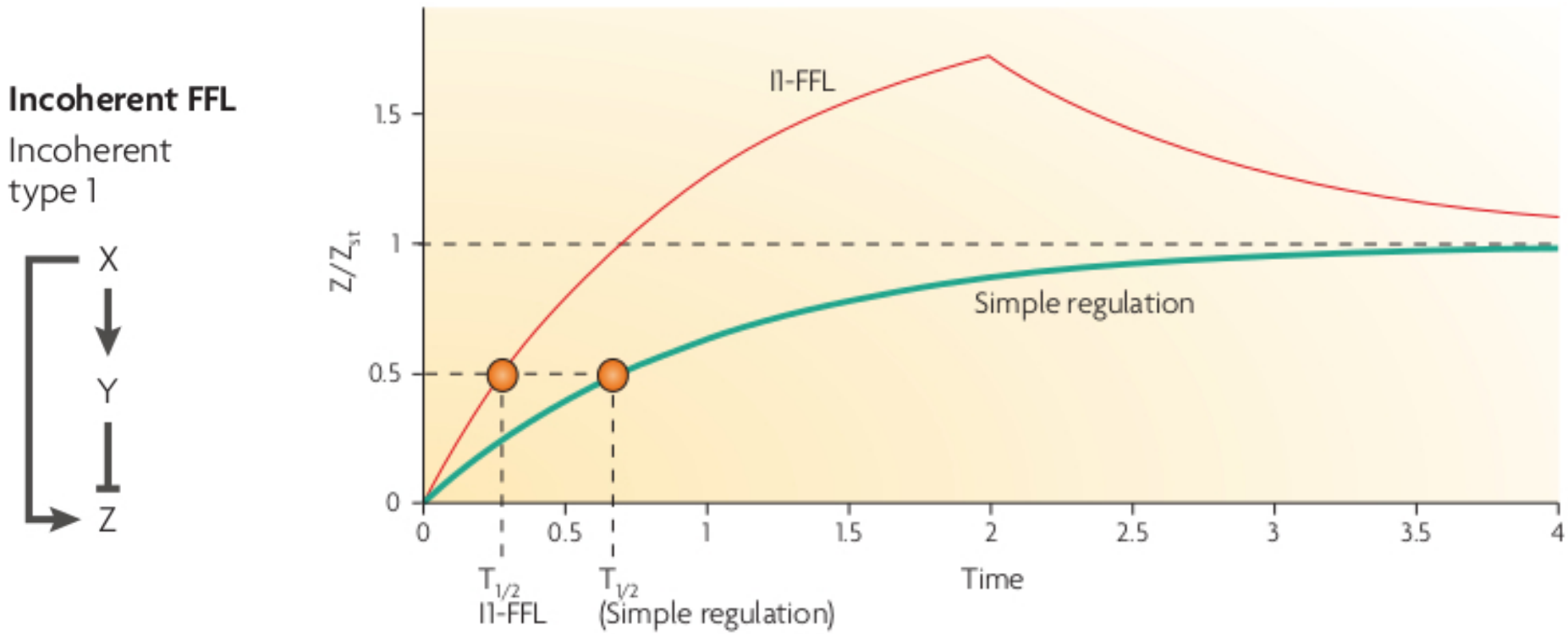
AND connection for Z

→ only respond to persistent stimulation, elevator door effect

OR connection for Z

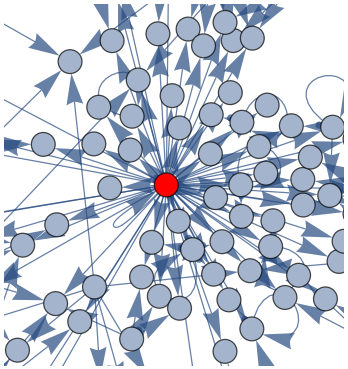
→ no delay after stimulation, but delay after stimulation stops

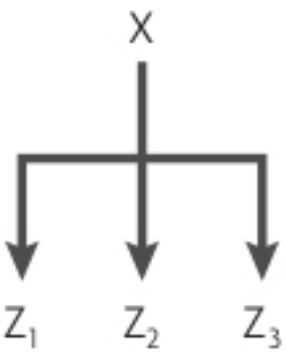
incoherent type 1 feed forward loop

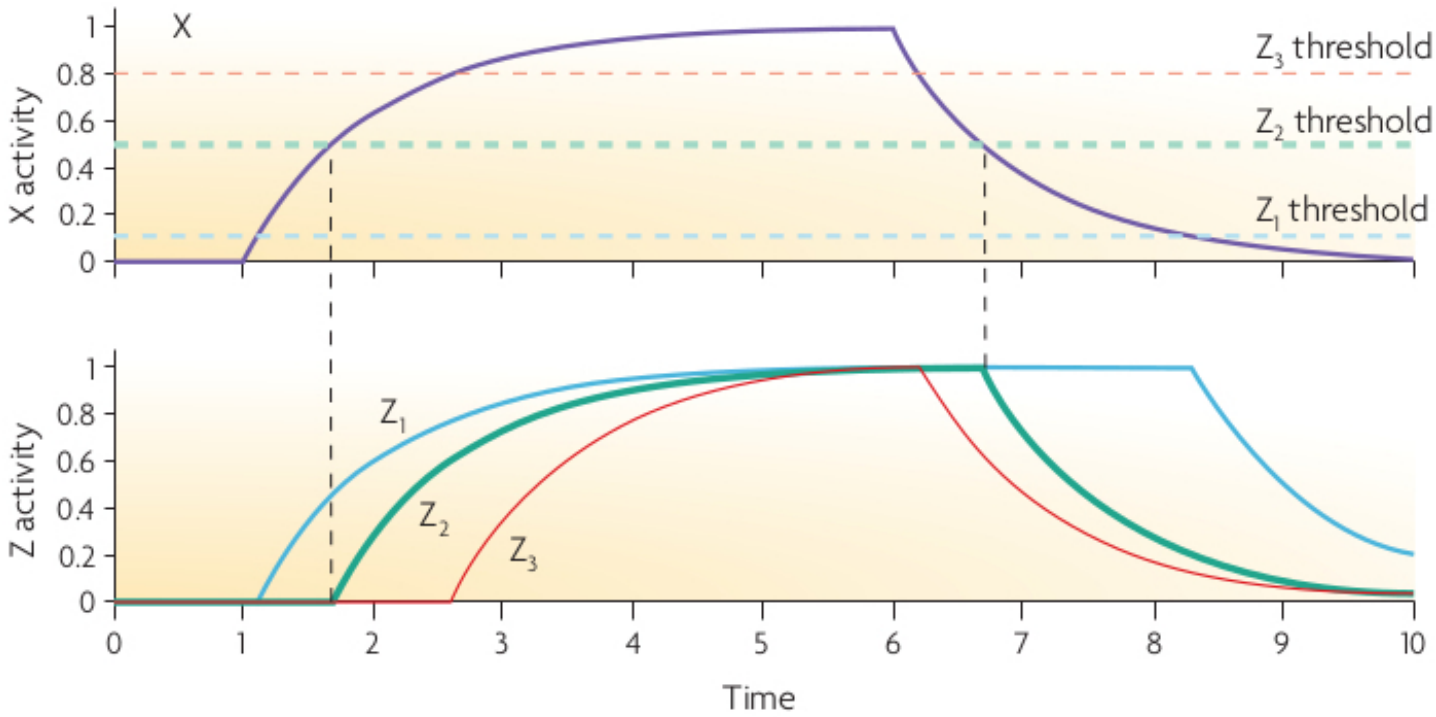


- Response acceleration
- Pulse generator

single-input module

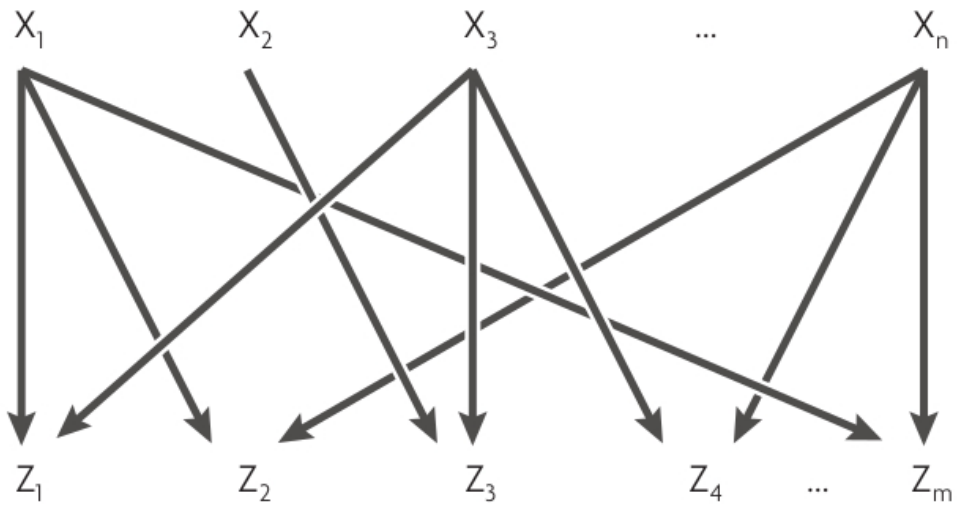






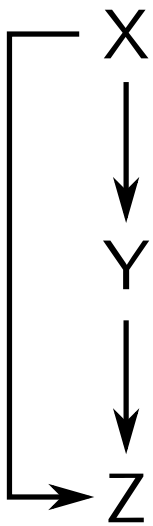
$\rightarrow$ timed expression of different genes

dense overlapping regulon

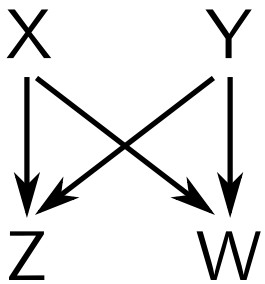


- Not yet very well understood
- Detailed information about connection strength is needed

motifs in the transcription network of E. coli



feed-forward loop

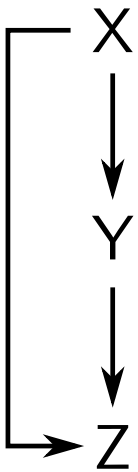


bi-fan

$N_{\text{real}} = 40 \quad N_{\text{rand}} = 7 \pm 3$

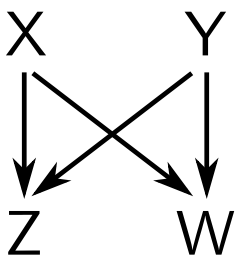
$N_{\text{real}} = 203 \quad N_{\text{rand}} = 47 \pm 12$

motifs in the neural network of C. elegans



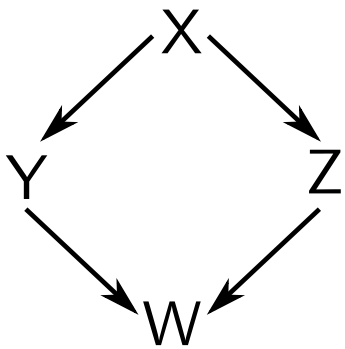
feed-forward loop

$N_{\text{real}} = 125$
$N_{\text{rand}} = 90 \pm 10$



bi-fan

$N_{\text{real}} = 127$
$N_{\text{rand}} = 55 \pm 13$

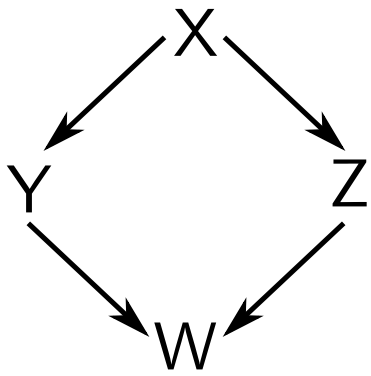


bi-parallel

$N_{\text{real}} = 227$
$N_{\text{rand}} = 35 \pm 10$

motifs in the food web of little rock

three chain

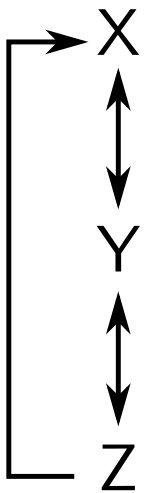


bi-parallel

$N_{\text{real}} = 3219 \quad N_{\text{rand}} = 3120 \pm 50$

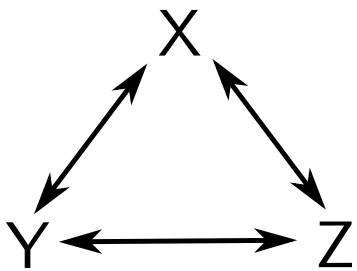
$N_{\text{real}} = 7295 \quad N_{\text{rand}} = 2220 \pm 210$

motifs in the world wide web



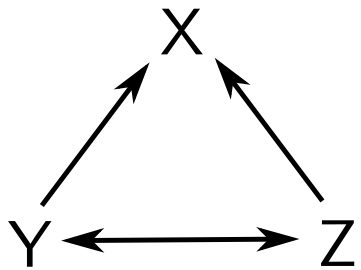
feedback with two mutual dyads

$N_{\text{real}} = 110000$
$N_{\text{rand}} = 2000 \pm 100$



fully connected triad

$N_{\text{real}} = 6800000$
$N_{\text{rand}} = 50000 \pm 400$



uplinked mutual dyad

$N_{\text{real}} = 1200000$
$N_{\text{rand}} = 10000 \pm 200$

motifs developmental networks

a careful look at the statistical methods

What has been done?

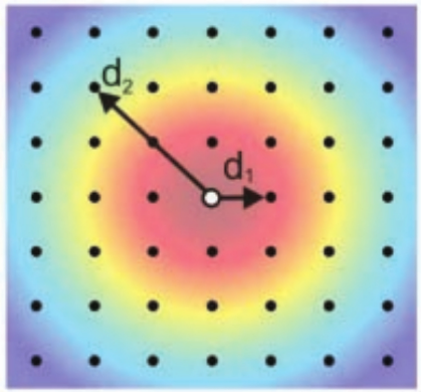
- Real network $\overset{\text{randomization}}{\longrightarrow}$ random network
- Posing of null hypothesis:
  "The occurrence of motifs is the same in both networks"
- The null hypothesis is rejected by a statistical test

$\rightarrow$ Conclusion:

Evolution prefers motifs that are overrepresented and disfavours others

a careful look at the statistical methods

- Create a random *toy model*
- Probability of connection between two nodes reduces with distance



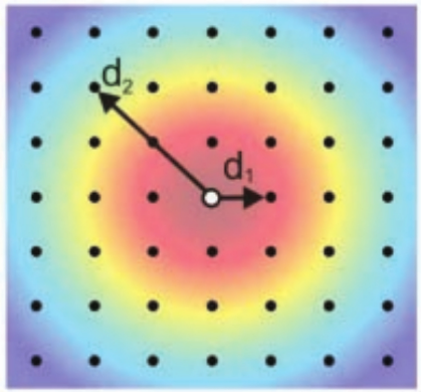
- Test model against random network with same # of nodes and edges
- There is a significant occurrence of motifs in the toy model!

  $\rightarrow$ It is clearly not evolution which prefers those motifs

  $\rightarrow$ Spatial clustering can create motifs

a careful look at the statistical methods

- Create a random *toy model*
- Probability of connection between two nodes reduces with distance



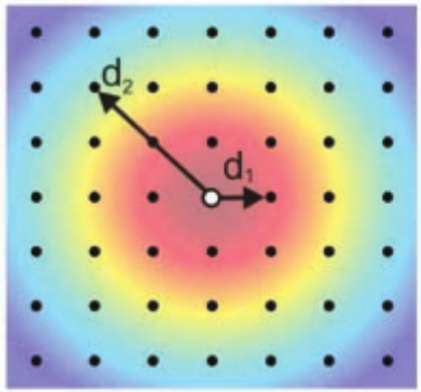
- Test model against random network with same # of nodes and edges
- There is a significant occurrence of motifs in the toy model!

  $\rightarrow$ It is clearly not evolution which prefers those motifs

  $\rightarrow$ Spatial clustering can create motifs

a careful look at the statistical methods

- Create a random *toy model*
- Probability of connection between two nodes reduces with distance



- Test model against random network with same # of nodes and edges
- There is a significant occurrence of motifs in the toy model!

  $\rightarrow$ It is clearly not evolution which prefers those motifs

  $\rightarrow$ Spatial clustering can create motifs

references

- Alon, U. *Network motifs: theory and experimental approaches*. Nature Rev. Genet. 8, 450-461 (2007).
- Shen-Orr, S. S., Milo, R., Mangan, S. & Alon, U. *Network motifs in the transcriptional regulation network of Escherichia coli.* Nature Genet. 31, 64–68 (2002).
- Milo, R. et al. *Network motifs: simple building blocks of complex networks.* Science 298, 824–827 (2002).
- Alon, U. *Introduction to Systems Biology: Design Principles Of Biological Circuits* (CRC, Boca Raton, 2007).
- Artzy-Randrup, Y. et al. *Comment on "Network Motifs: Simple Building Blocks of Complex Networks" and "Superfamilies of Evolved and Designed Networks"*. Science 305, 1107 (2004).

résumé

- We can identify repeated small patterns in real networks

- Comparison to randomized networks shows a significant accumulation of motifs in real networks

- We have to be careful which random network to use as null hypothesis

- We can describe the behaviour of the isolated motif