

Are biological networks special?

Are biological networks special?

If so, why?

If not, how come?

How to define special?

- Compare with our expectation
 - Subjective
 - Example: Humans have very few genes
 - (only five to six times as many as E.coli)
 - Interesting (perhaps it shows our feeling for how complexity grows with the number of genes is not correct or humans are not as complex as we tend to think)
 - Compare with other networks
 - Other fields
 - social networks
 - biological
 - protein interaction versus protein regulation
 - Different species (Topic of evolution talk)

Today

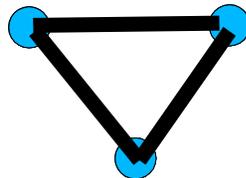
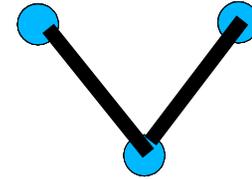
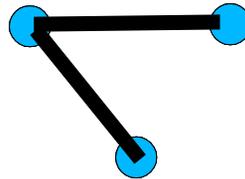
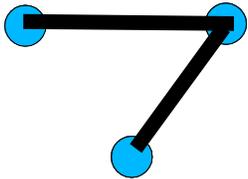
- Compare with random artificial networks
- How to create random networks
 - Special also depends on the context and how much you already know.
- How to compare networks
 - degree distribution
 - motifs
 - Next nearest neighbours

Today

- Compare with random artificial networks
- How to create random networks (1)
 - Special also depends on the context and how much you already know. (3)
- How to compare networks
 - degree distribution (2)
 - motifs (4)
 - Next nearest neighbours (perhaps)

Example:

Connected networks with three nodes.

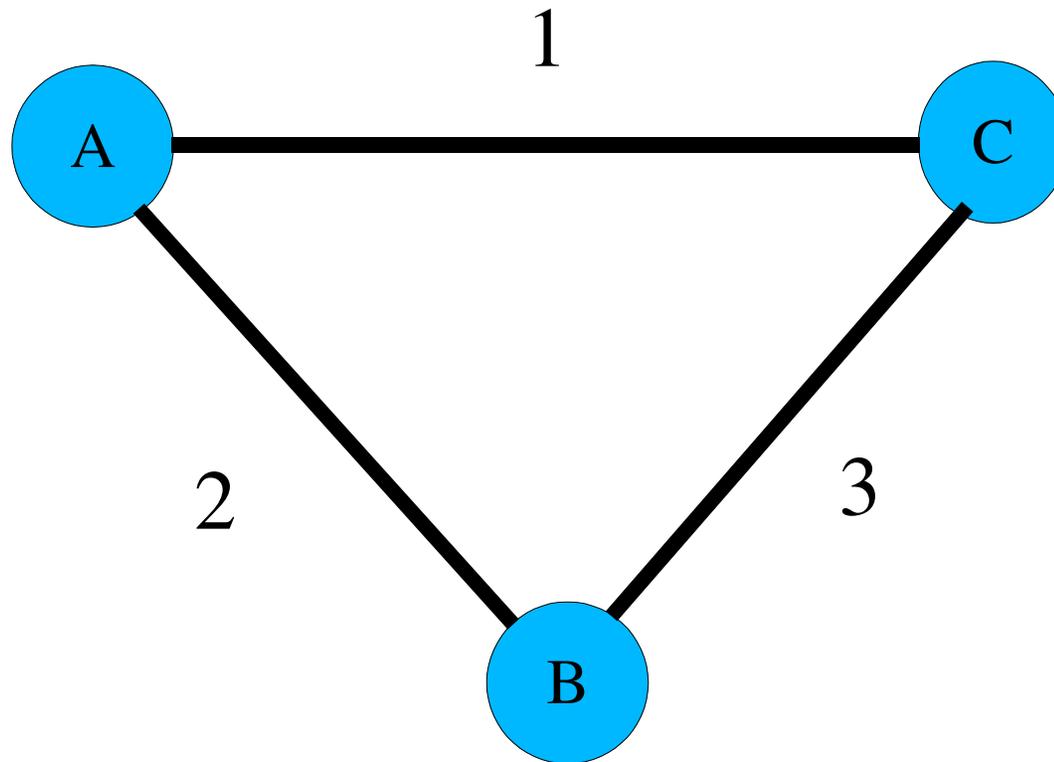


I observe 5 out of 10 three node networks are fully connected.

Is this unusual?

- It depends on how probable (p) a link is.
 - If almost all nodes are connected
 - 5 is quite few
 - If almost no nodes are connected
 - 5 is very high
- Perhaps you perform the pairwise experiment many times and in this way you measure p .

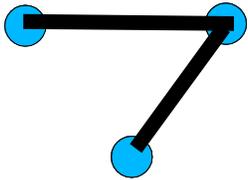
Student simulation



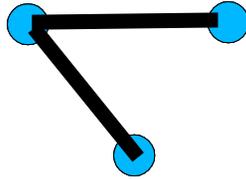
Three groups: $p=1/3$, $p=1/2$, $p=2/3$.

Assume the probability for a link between two nodes p is known.

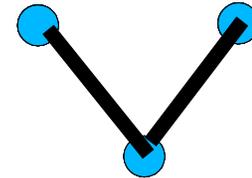
Probabilities for the different graphs



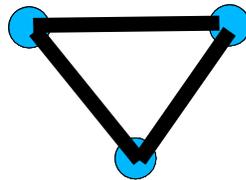
$$p^2(1-p)$$



$$p^2(1-p)$$

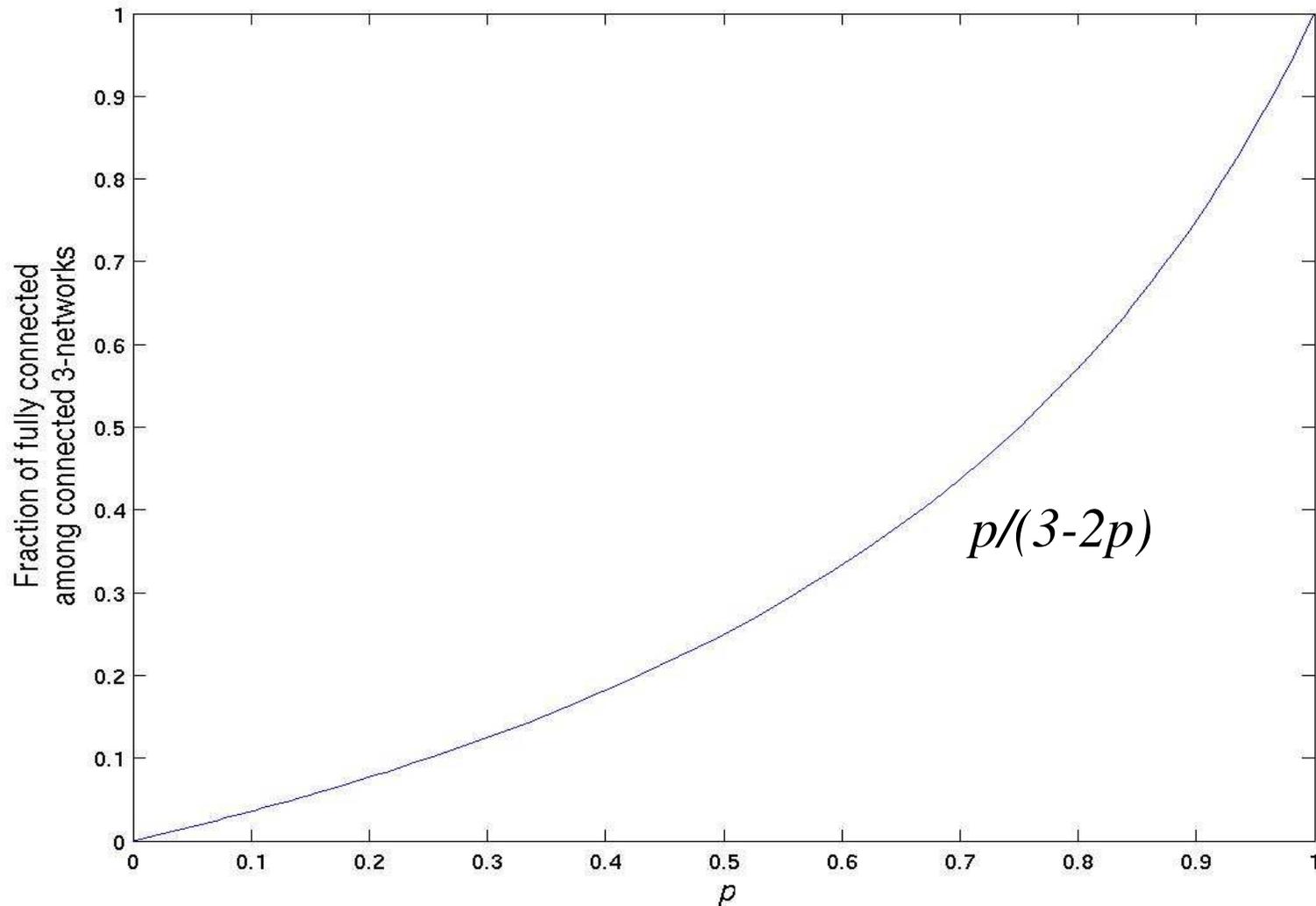


$$p^2(1-p)$$



$$p^3$$

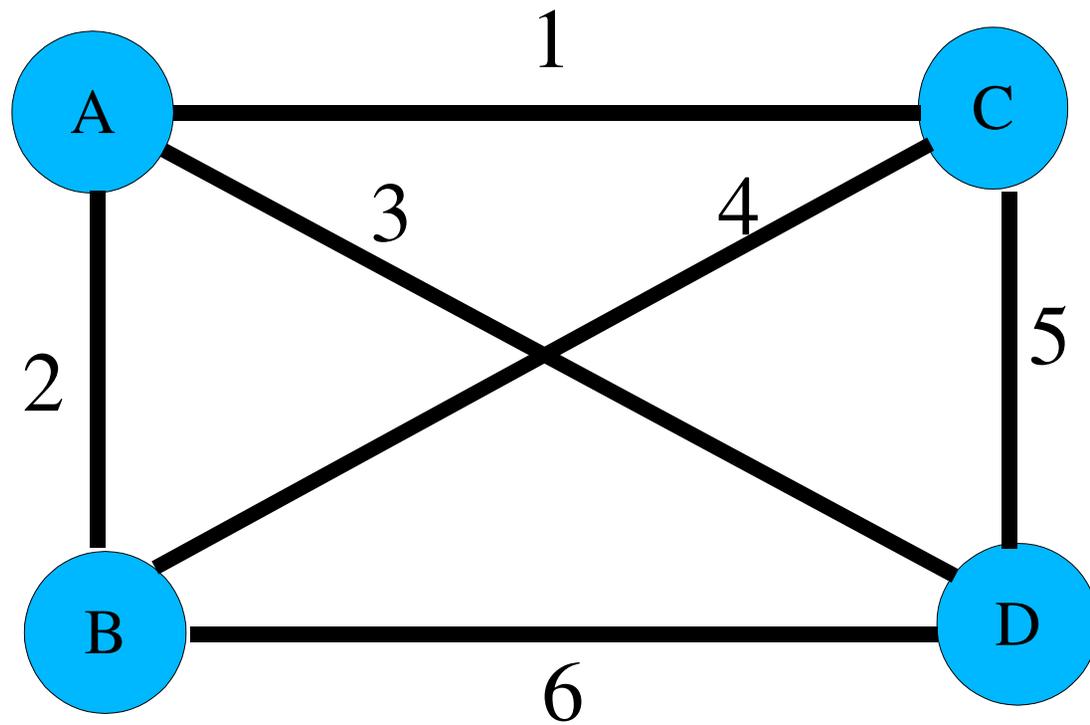
Fraction of fully connected networks among connected 3-networks



We need p-values!

- Probability to get 5 or more fully connected
 - $p=1/3$: 0.008
 - $p=1/2$: 0.08
 - $p=2/3$: 0.4
- Probability to get a more unlikely configuration
(Given we have overrepresentation, what then)
 - $p=1/3$: 0.008 (0.02)
 - $p=1/2$: 0.1 (0.2)
 - $p=2/3$: 0.5 (0.6)

Student simulation

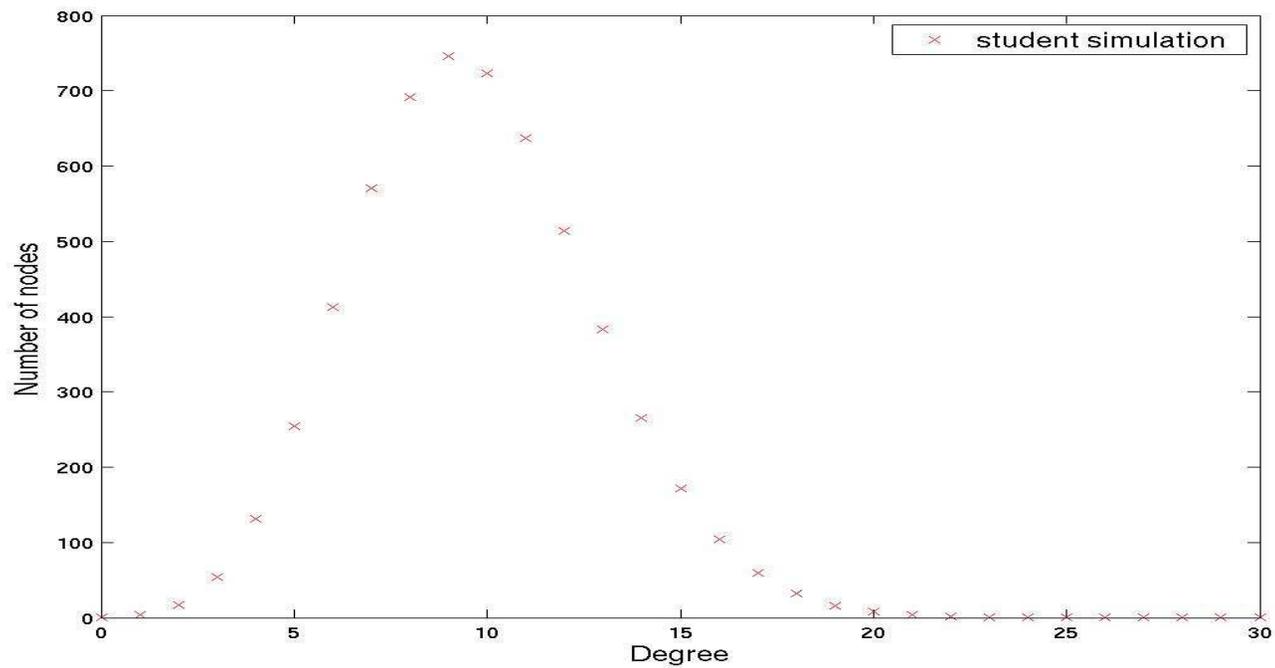


Degree distribution

- N nodes in the network.
- A node has $N-1$ potential links.
- Each link is picked with probability p
 - $p=L/N/(N-1)$ if there are L links in the real network.
- This implies the probability to have l links is given by the binomial distribution.
 - $\text{bino}(l, N-1, p)$
- We could also have got it by a huge student simulation (Generally we simulate)

Degree distribution

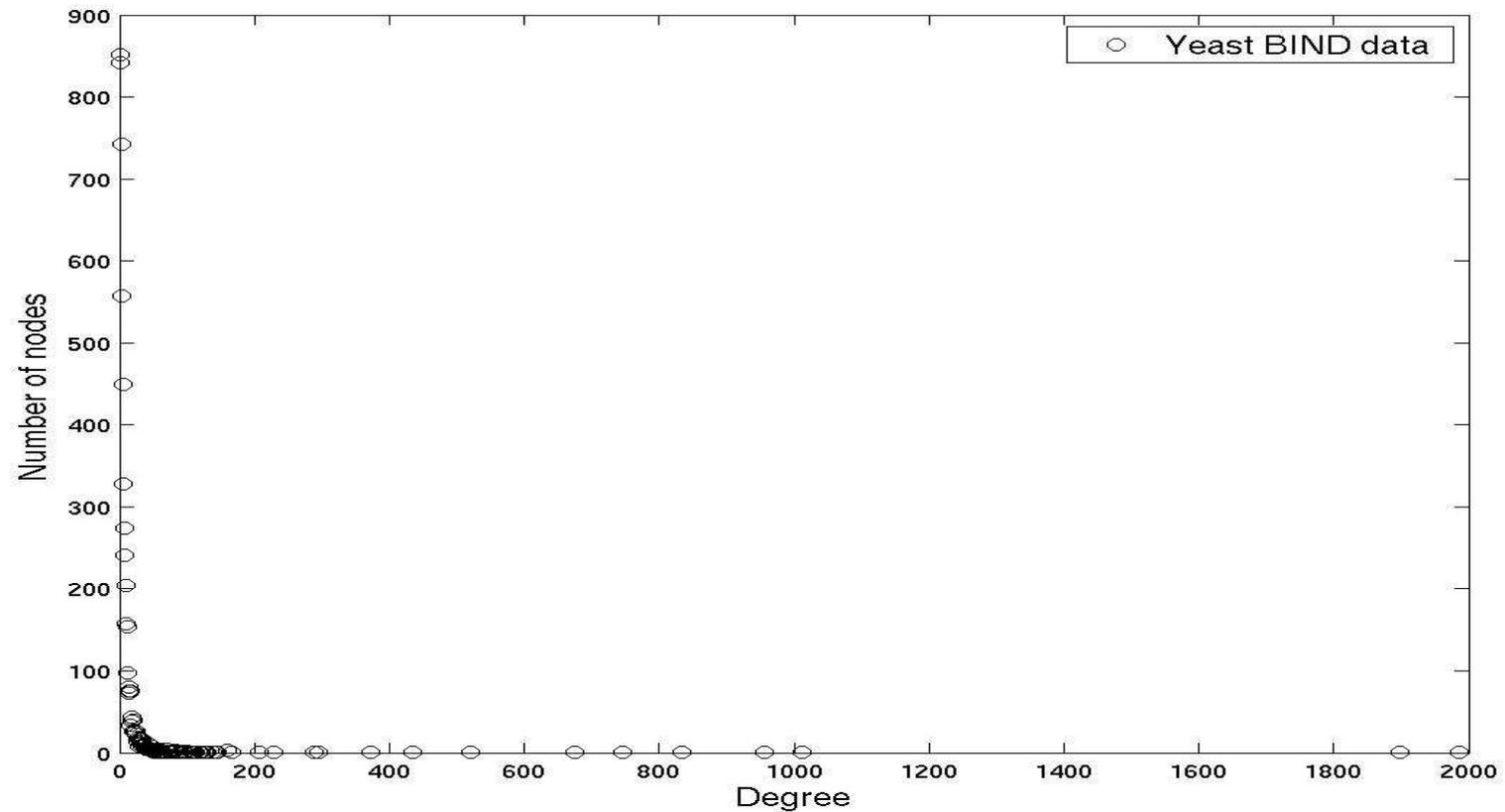
Number of nodes=5800, number of links=28110



Note the network is not single connected, but almost

Degree distribution for *S. cerevisiae*

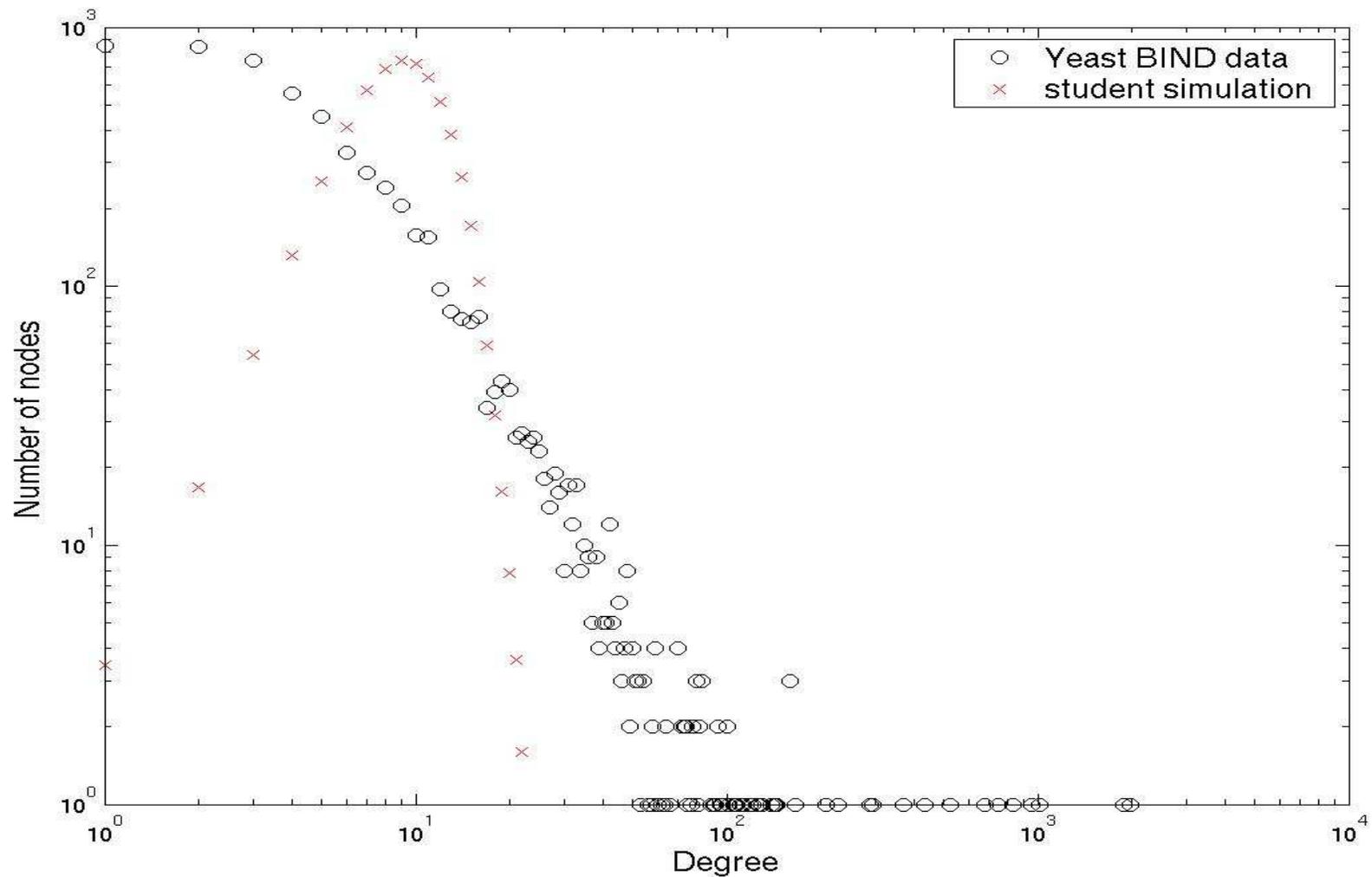
Number of nodes=5800, number of links=28110



Note the huge range along the x-axis

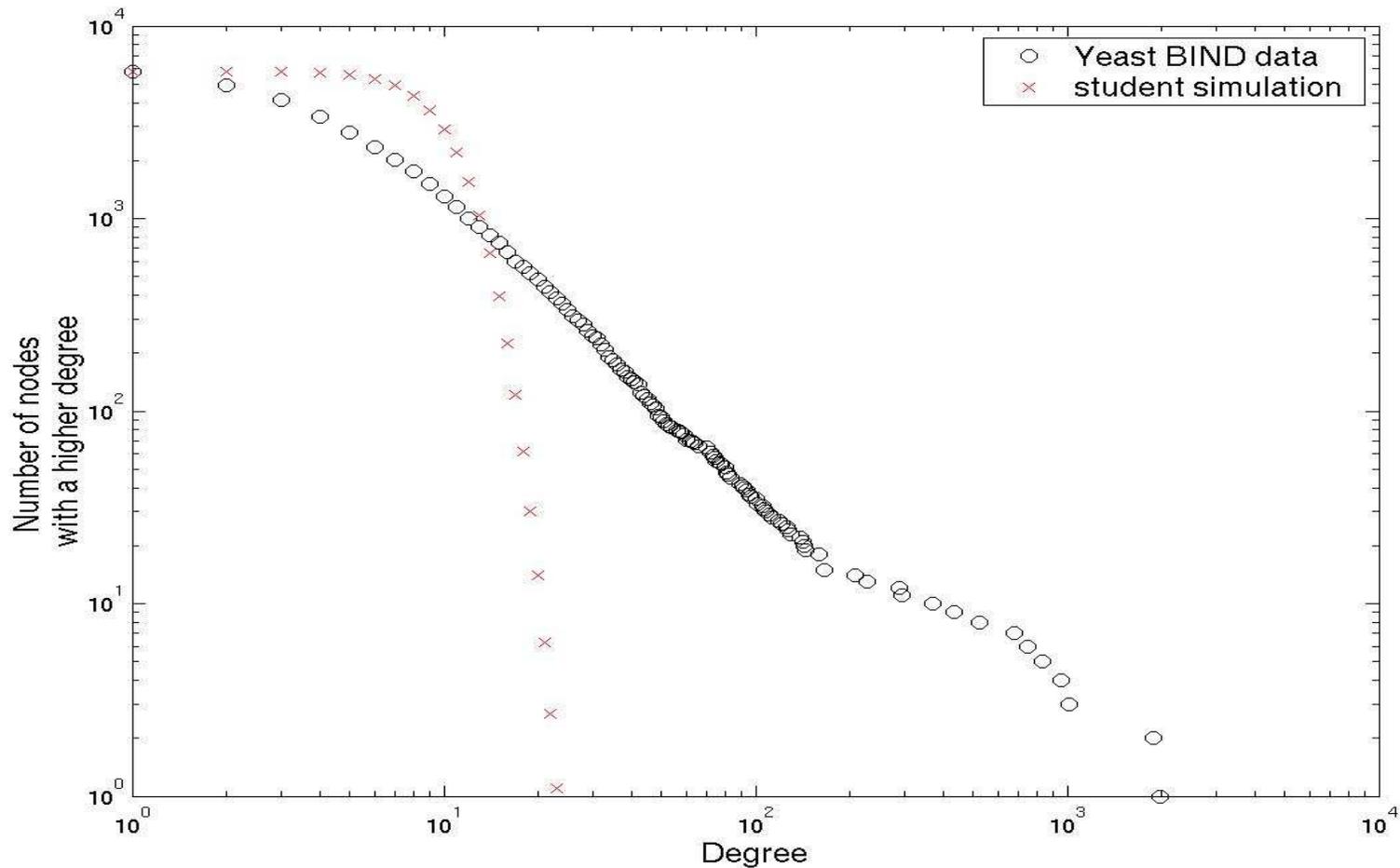
Degree distribution

Number of nodes=5800, number of links=28110



Degree distribution

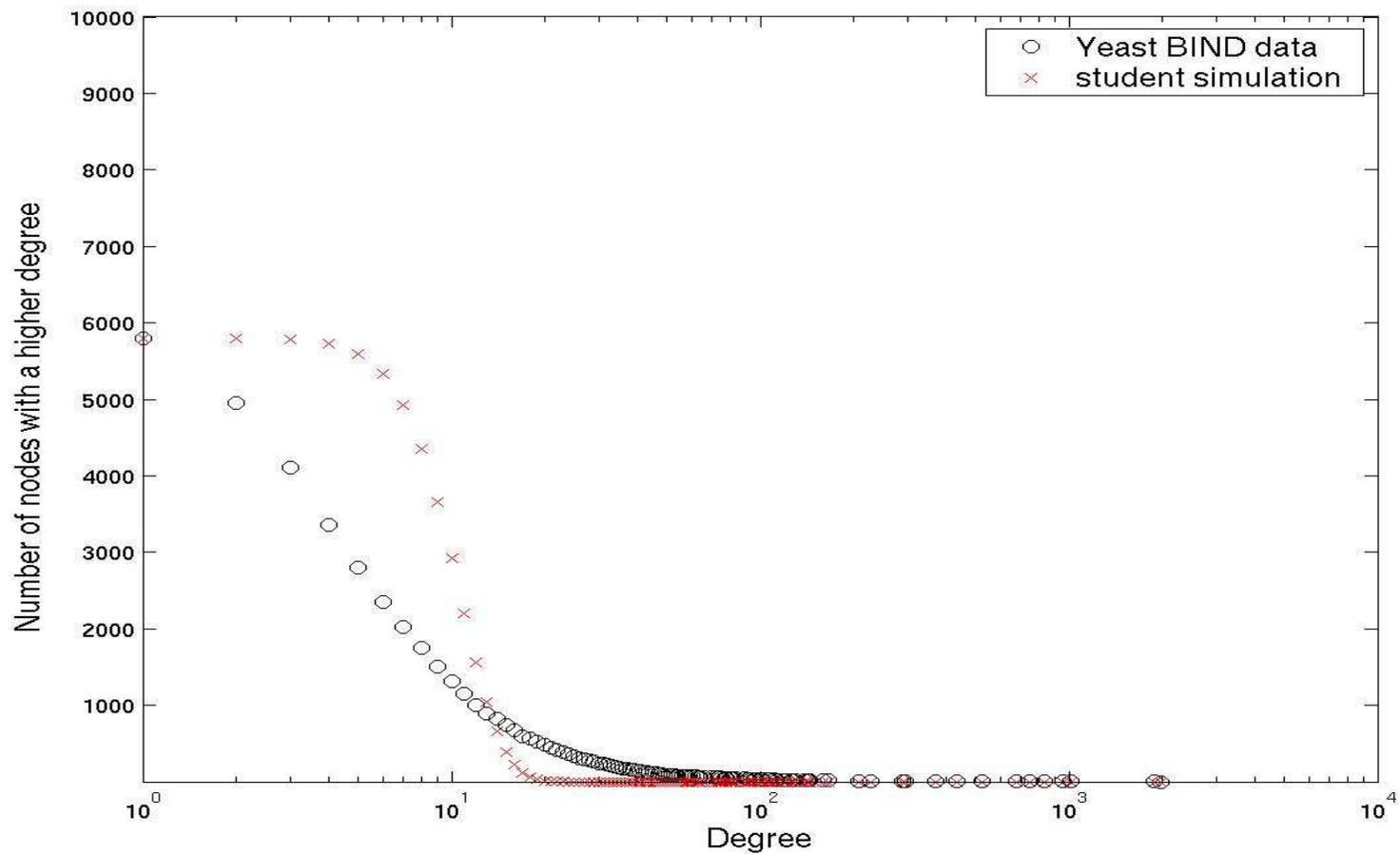
Number of nodes=5800, number of links=28110



The p-value of getting the yeast distribution by chance is 0.0000000000

Degree distribution

Number of nodes=5800, number of links=28110



The degree distributions differ

- Can we trust the student simulation for other network properties?
- For instance we might look for triangles in the big network as you did in the small 3-network.
- Is this legitimate???

No!!

- Probability of a triangle depends on the probability for a link to be present
- The probability for a link depends on the degree of the node in question.

Context dependence

- Example
 - Blue eyes are rare on a world wide scale
 - Blue eyes are common in Sweden
 - If you find a person with blue eyes it is unusual
 - However if you know he/she is swedish blue eyes are not unusual.
- Have to ensure our random networks have the right degree distribution.

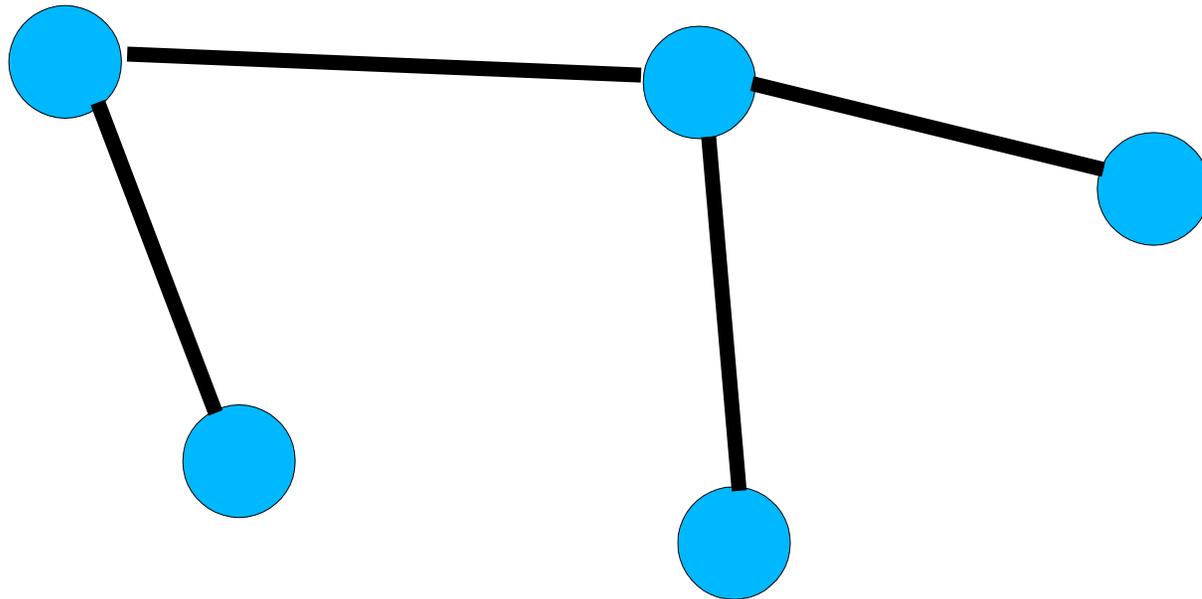
How to make a random network with the right distribution

- Too many networks exist, so it is not possible to try out all
- Exists algorithms like the student simulation that generate scale-free networks. However the generated networks are only a small subset of all scale-free networks.
 - Have some 'other' properties that are not like real networks. (Algorithmic artefacts).

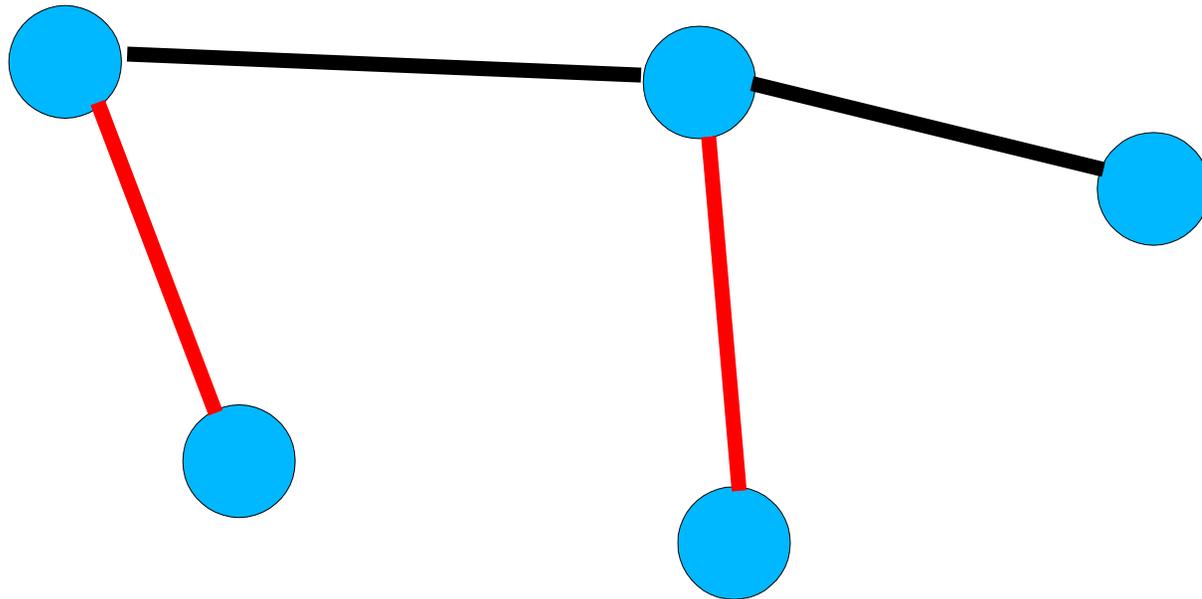
Heuristic method

- Randomize the given network
- The routine is quite flexible preserve 'other' known features.
- It is believed that it samples a large subset of networks. (Not been proven to my knowledge)
- De facto standard
- Conserves the degree of all nodes.
 - Keep map to other node properties.

Randomization I

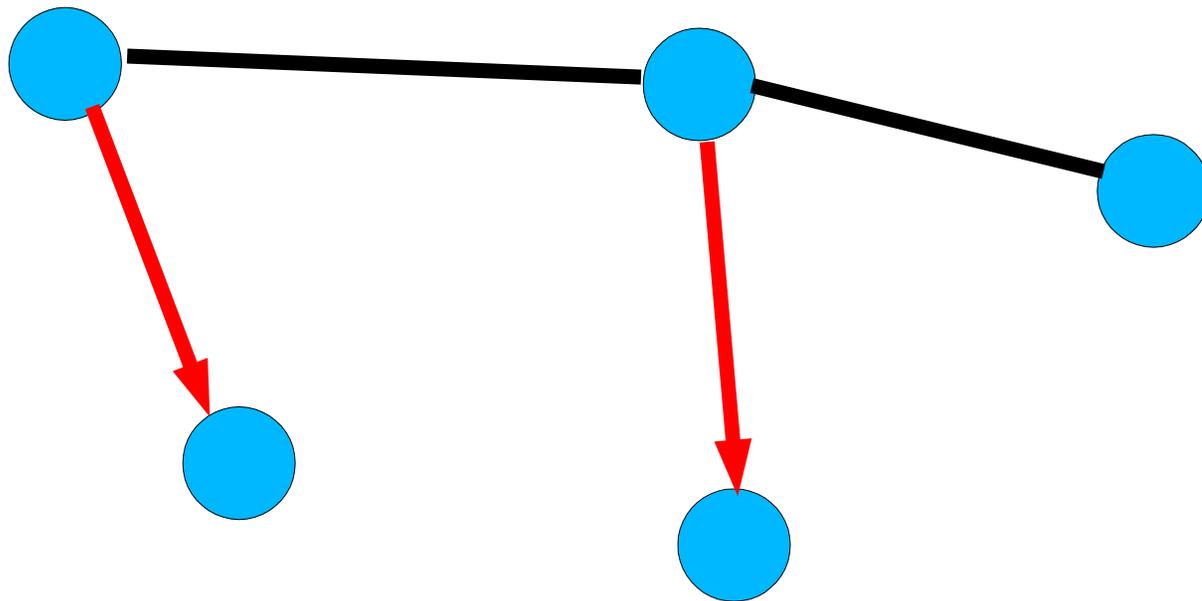


Randomization II



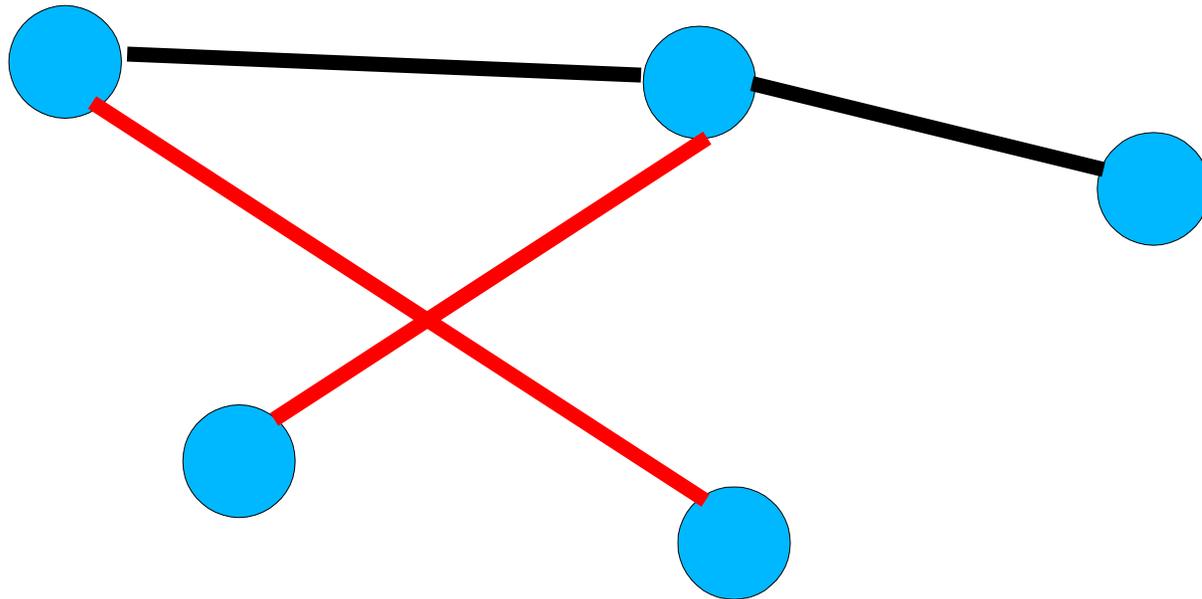
Pick two links

Randomization III



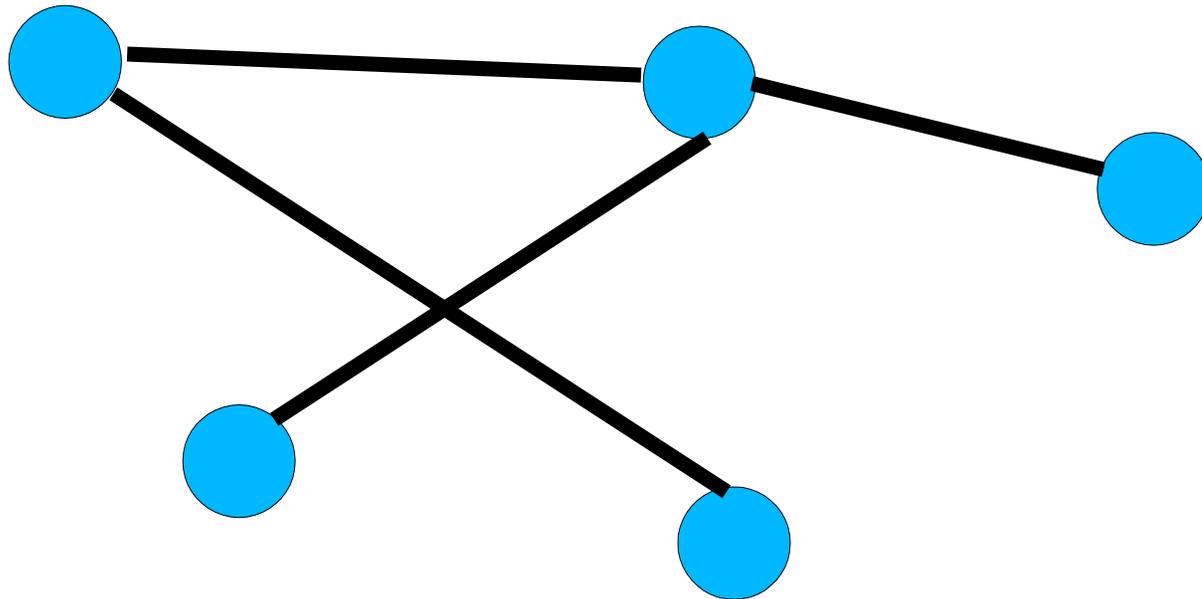
If undirected: Pick a direction

Randomization IV



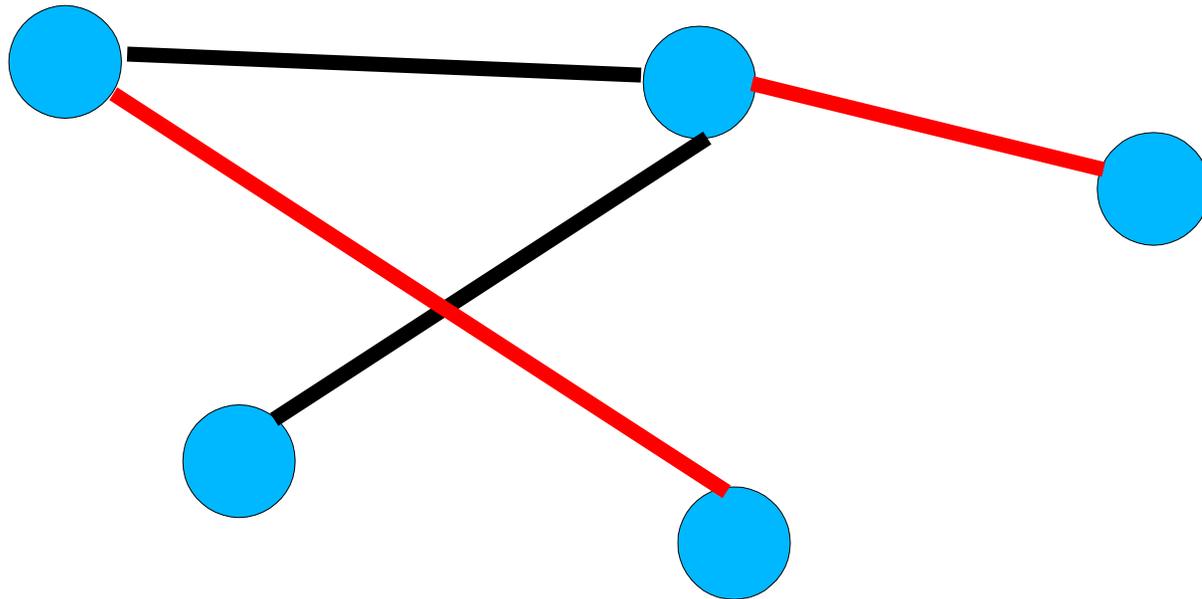
Swop arrow heads

Randomization V



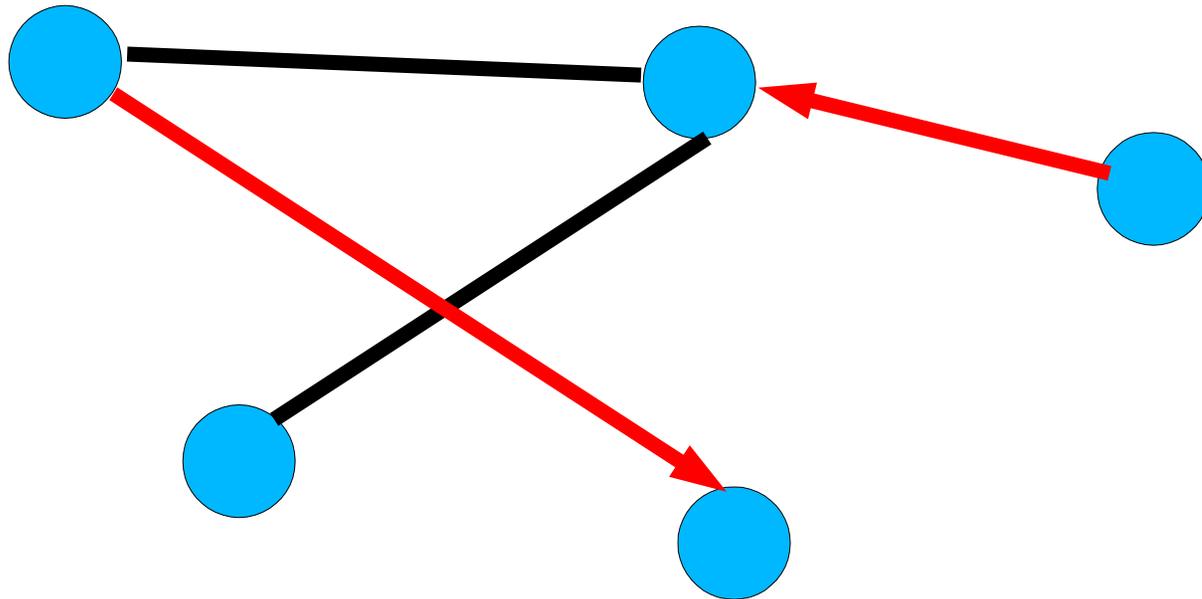
End of step 1

Randomization VI



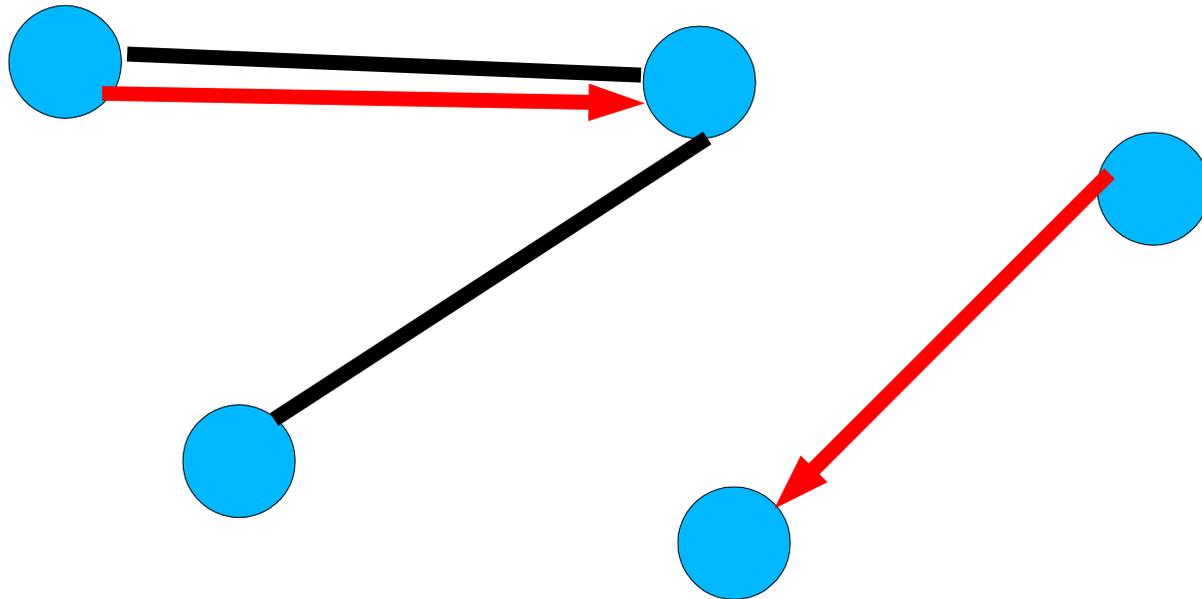
Pick two new links

Randomization VII



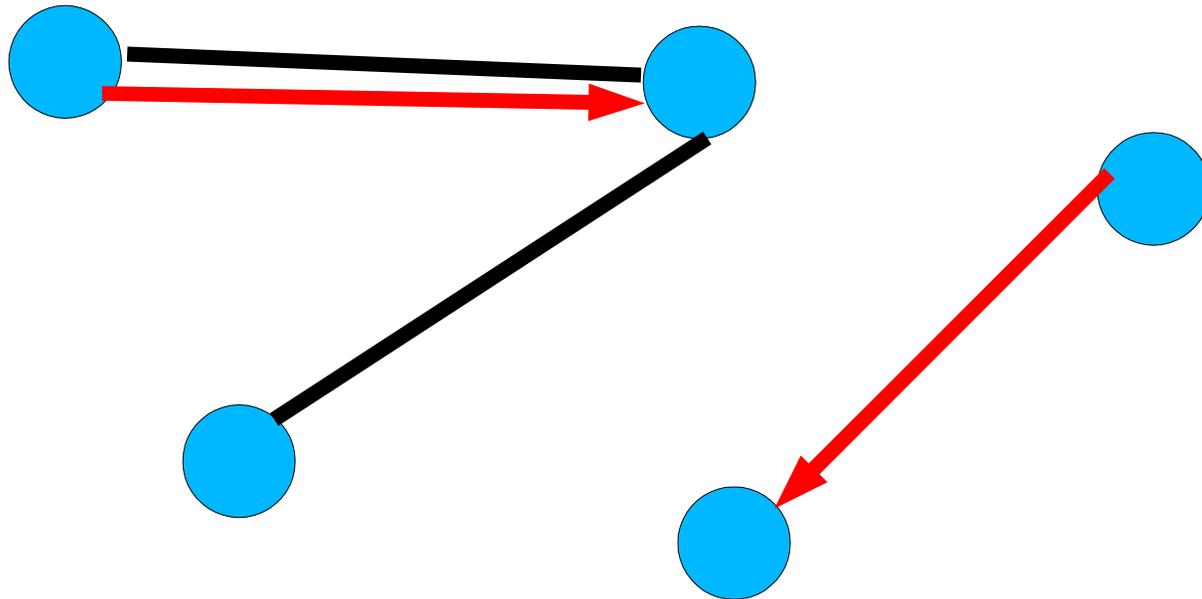
Pick directions

Randomization VIII



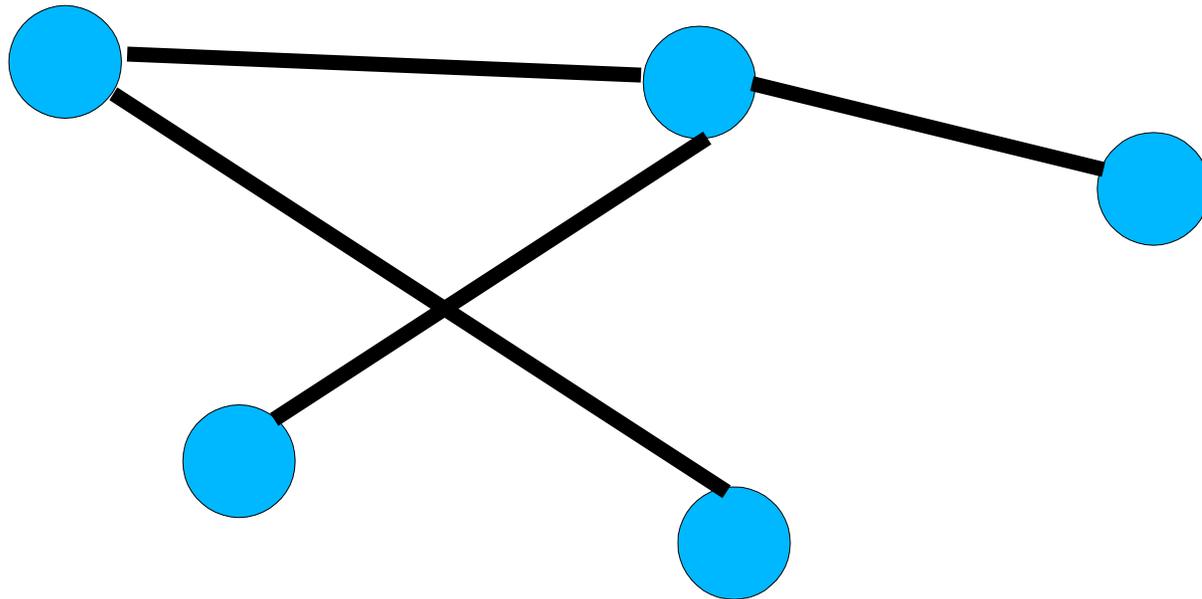
Swop arrowheads

Randomization IX



Created a double link (and a not single connected fraction)
Abandon move i.e. undo and start again at end of step 1.

Randomization $X=V$



End of step 1

Randomization

- Implemented for this course as a subroutine in Perl
- Two versions
 - one for directed networks
 - randomize_directed
 - one for undirected networks
 - randomize_symmetric

'Other features'

- Motifs
 - Example are the triangles from before.
- Degree distribution of the neighbors to the nodes with a fixed degree.
- Look at properties of the nodes
 - Functional classification of the gene
 - Is the gene lethal
 - Is it part of a gene family

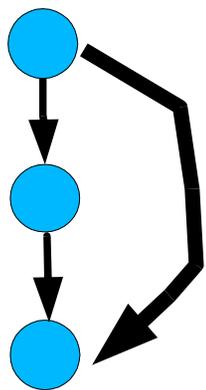
Motifs

- Big networks are never 100% identical
- However small subparts of networks might be.
- Such small sub networks are called motifs
- Simplest motif is a triangle.

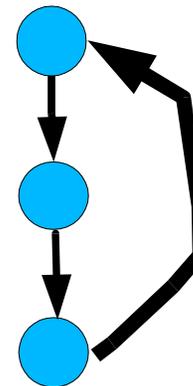
Transcriptional regulation in *E. coli*

▣(Nature Genetics 2003 Shen-Orr)

▣Triangles are now either feed forward loops or feed backward loops depending on the orientation of the regulation



Feed forward loop



Feed backwards loop

Transcriptional regulation

- Result in 2003 (Nature Genetics, Alon):
- Only feed forward loops are over-represented
 - 40 observed in real network
 - 7 observed on average in random networks
 - standard deviation is of order 4
 - z-score = $(40-7)/4=8$
 - Gauss p-value estimate is 10^{-17}

Transcriptional regulation

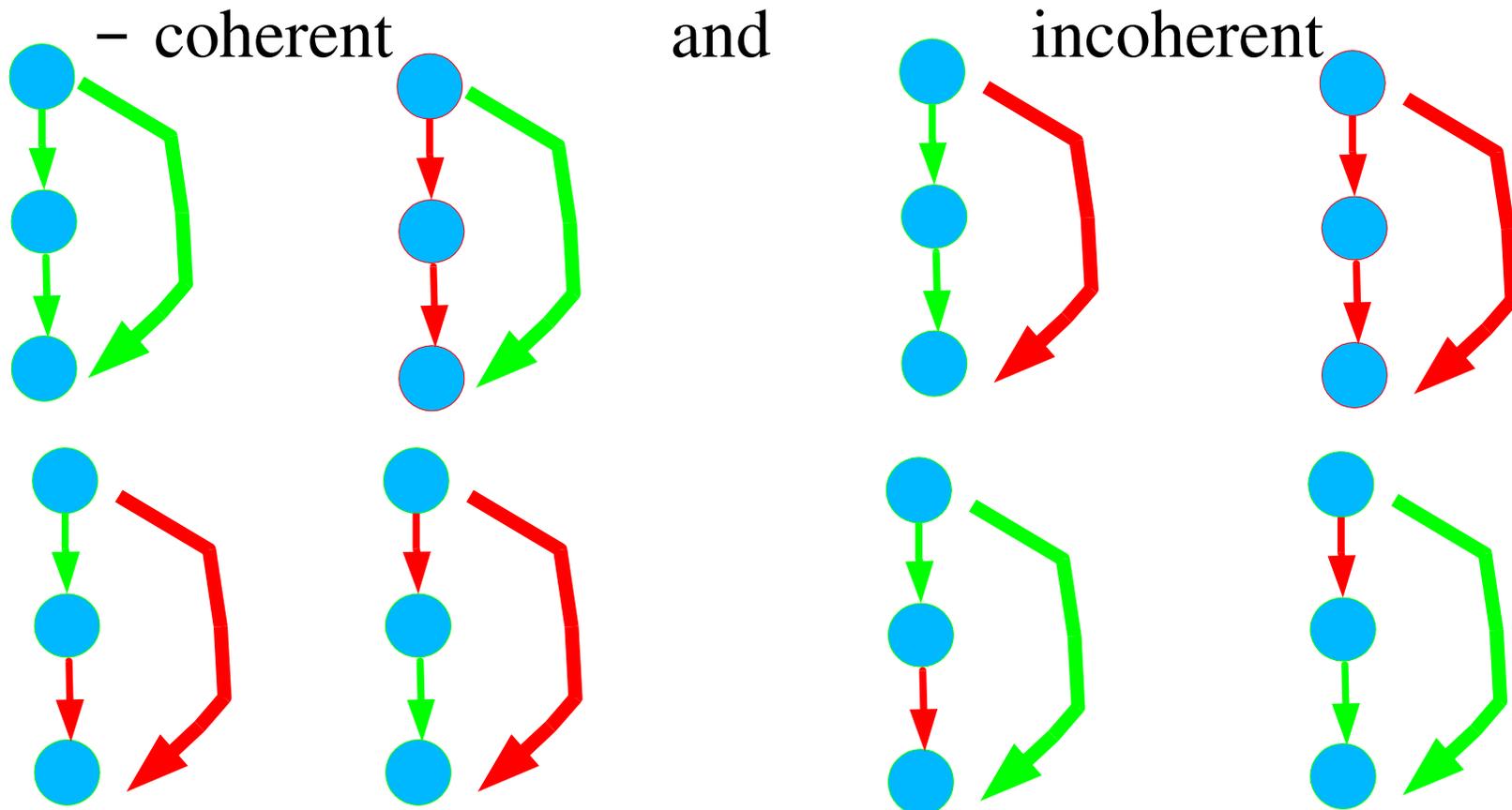
- Regulation (links) comes with a sign
- activate (336 links)
 - Turns on the gene
- repress (214 links)
 - Turns off the transcription of the gene
- Dual (29 links)
 - sometimes it activates and sometimes it repress
- One node might have links of several kinds/colors

Transcriptional regulation

- When randomizing the degree for every kind of link (color) for every node should be conserved.
 - the above routine should only swop links of the same kind (color)
- More positive regulation than negative
 - This might change in the future.
 - Experimental details might be important.
 - Might be different for eukariots.

Transcriptional regulation

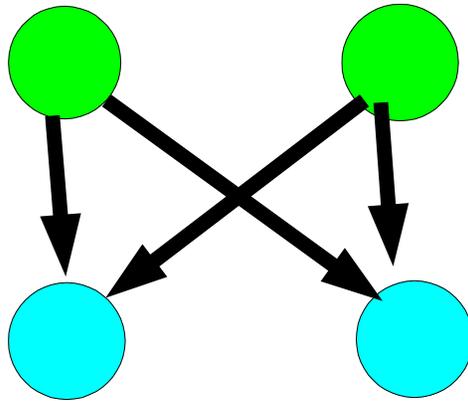
- Two kinds of feed forward loops, as regulation comes with a sign (**activate** and **repress**).



Transcriptional regulation

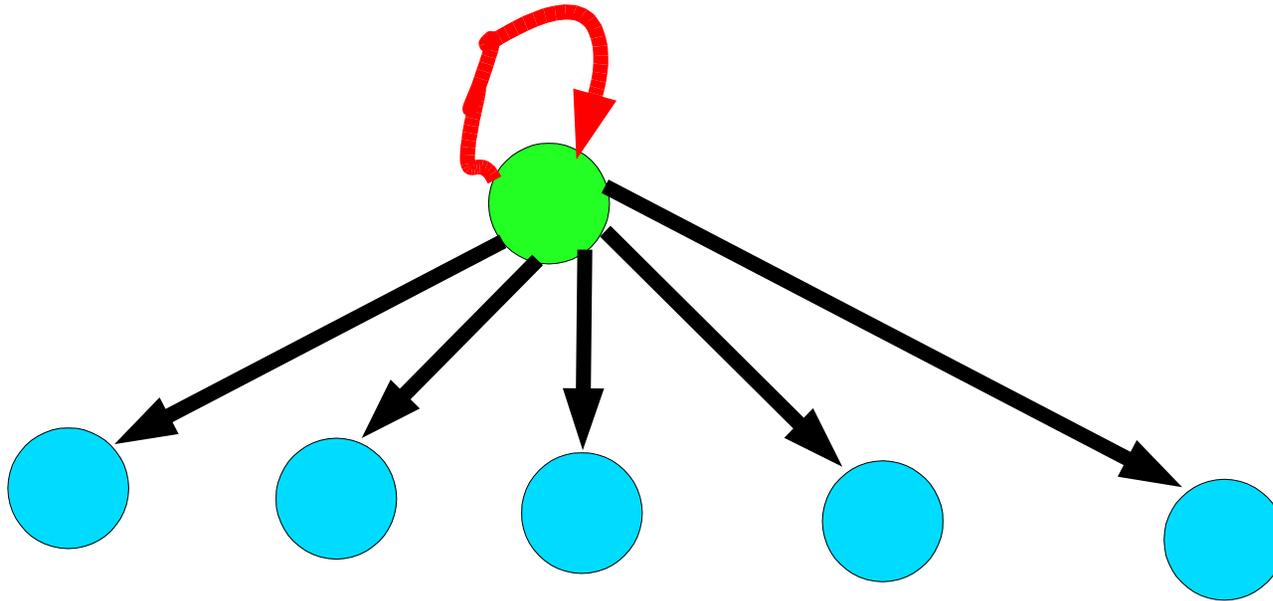
- Coherent feedforward loop
 - Real 34
 - Random 4.4, std. dev. 3
 - z-value= $(34-4.4)/3=9$ (p-value 10^{-22})
- Incoherent feedforward loop
 - Real 6
 - Random 2.5, std. dev. 2
 - z-value= $(6-2.5)/2=1.7$ (p-value 0.04)

Pairs of operons



- Real network 203
- Randomized 57 std. dev. 14
- Gauss z-score $(203-57)/14=10$
- p-value 10^{-25}

Single input motifs



- At least 13 **operons (targets)**.
- All the **operons** have degree one
- The control **gene** (hub) is often negatively autoregulated (70% are autoregulated versus 50%).

Single input motifs

- Real network
 - 68
- Randomized network
 - 28 std. dev. 7
 - Gauss z-score = $(68-28)/7=5.5$ (p-value 10^{-8})
 - p-value <0.01

Dense overlapping regions

- Part of the network with many links from one group of transcription factors to a group of operons.
- Generalization of the pair of operons motif.

