Modelling dynamic networks

Regularization of non-homogeneous dynamic Bayesian network models by coupling interaction parameters

Marco Grzegorczyk

Johann Bernoulli Institute (JBI)

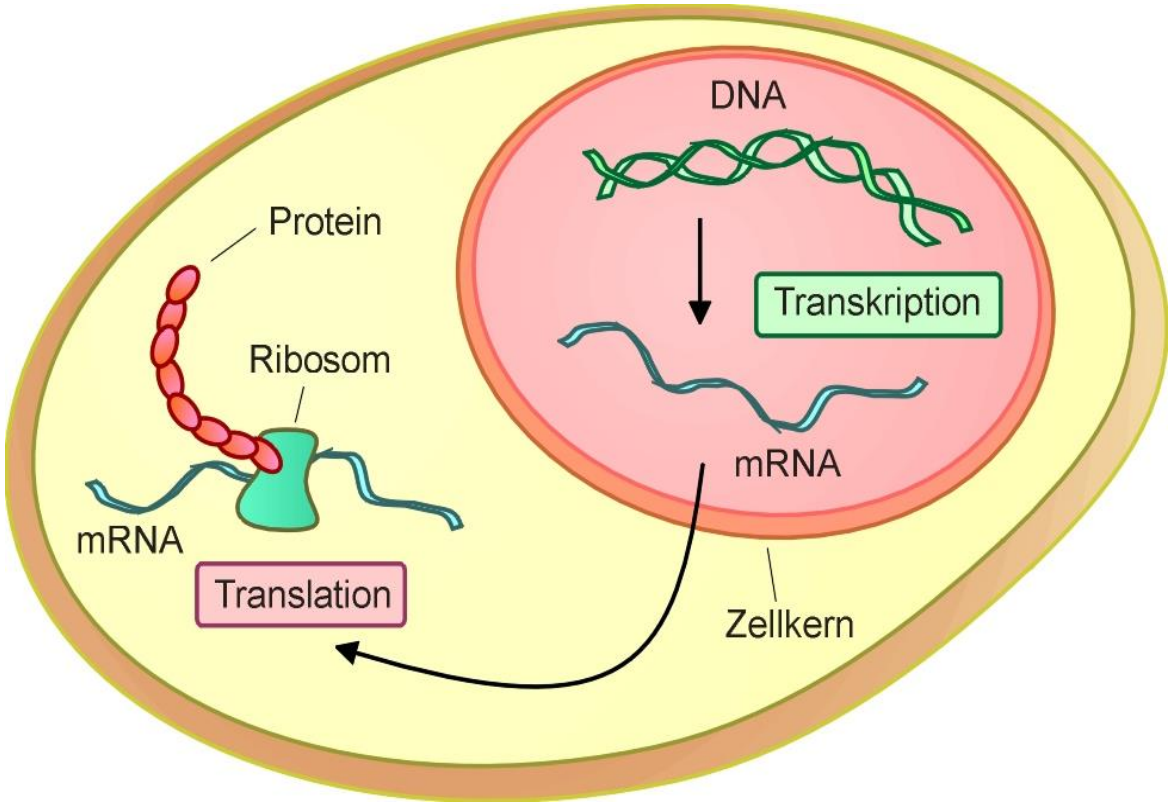
Rijksuniversiteit Groningen

Presentation at the Van Dantzig Seminar

VU University Amsterdam

9-Oct-2014

Cell Biology



Very brief introduction:

Each gene is the code for the synthesis of a specific protein.

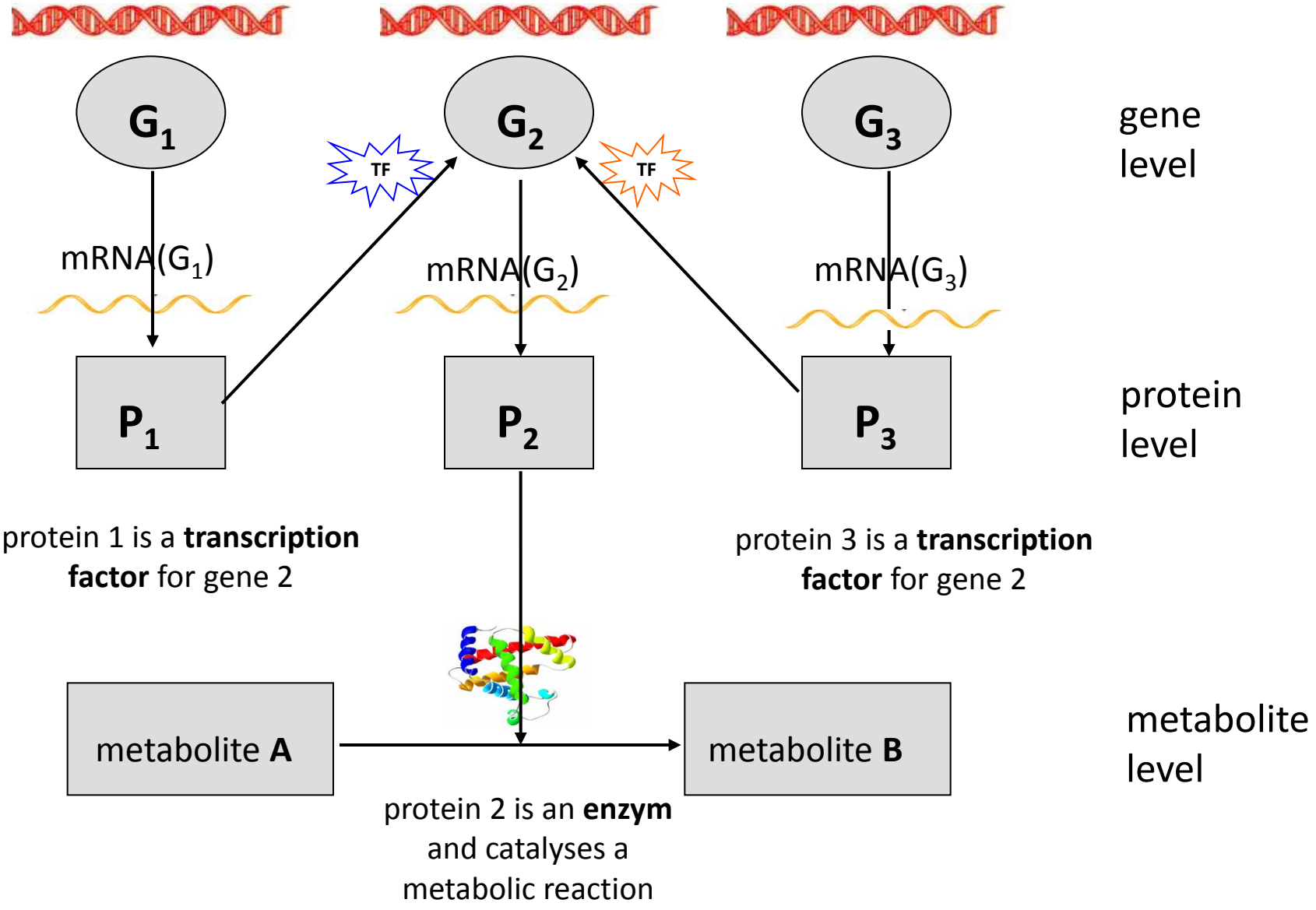
Transcription: gene → mRNA.

Translation: mRNA → protein.

Proteins are the „**functional units**“ of the cell.

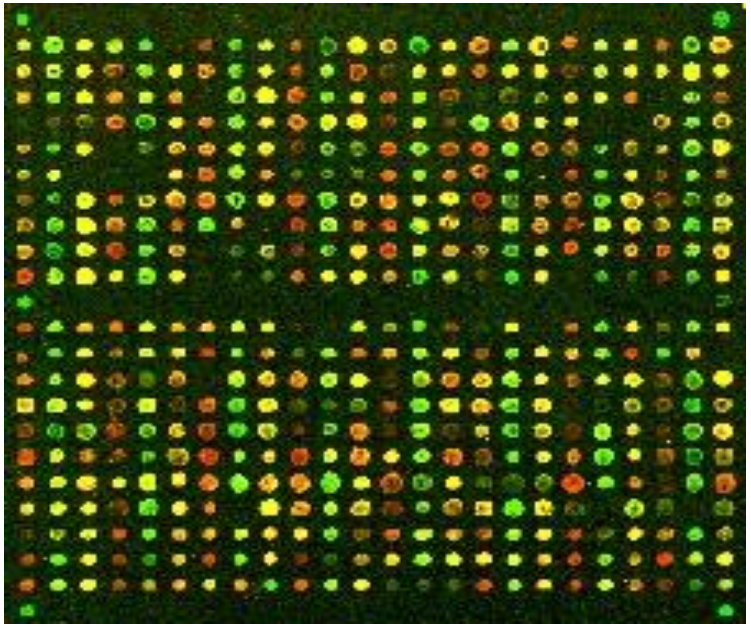
Proteins are enzymes, transription factors, etc.

Regulatory Network

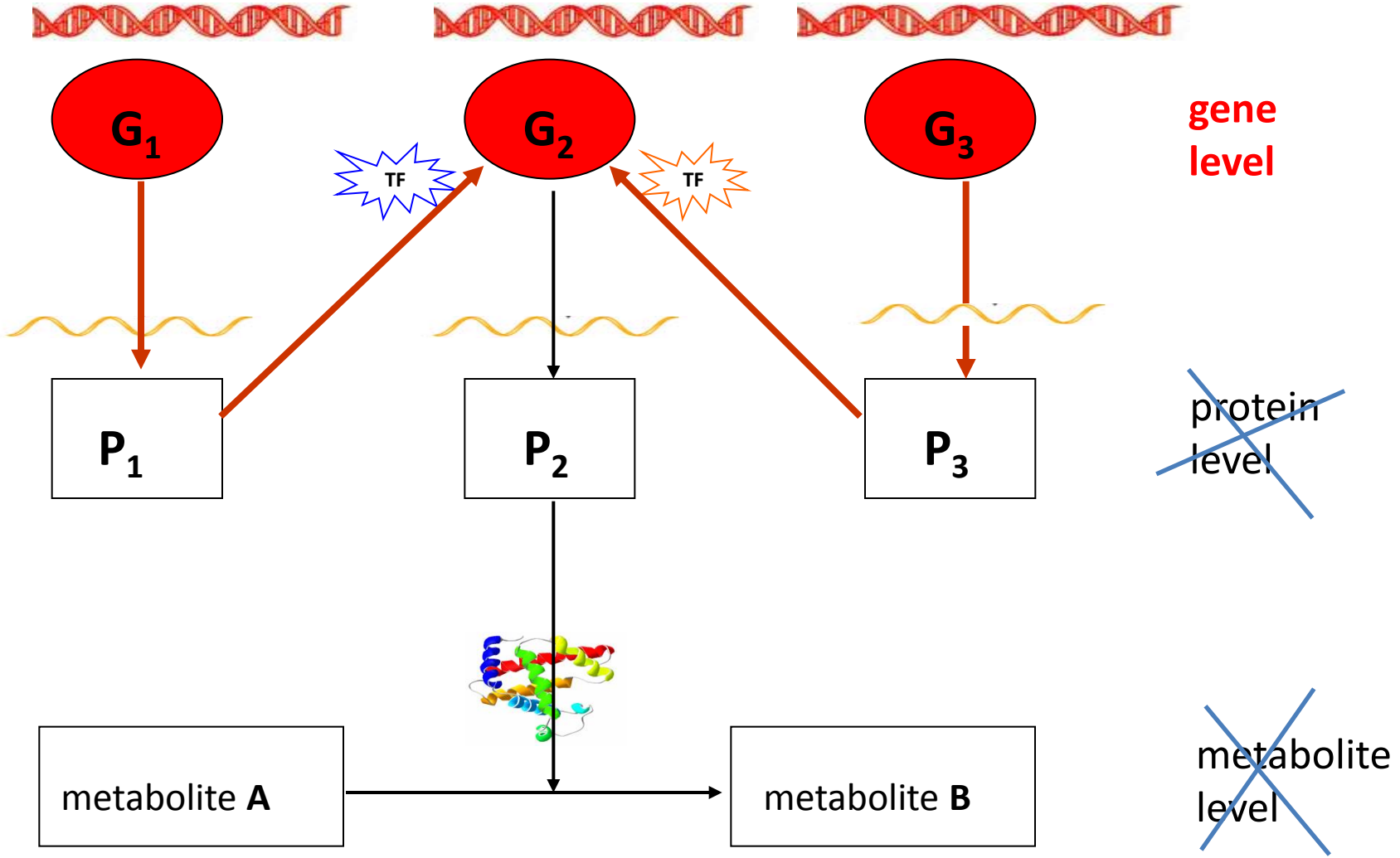


Microarray Chips

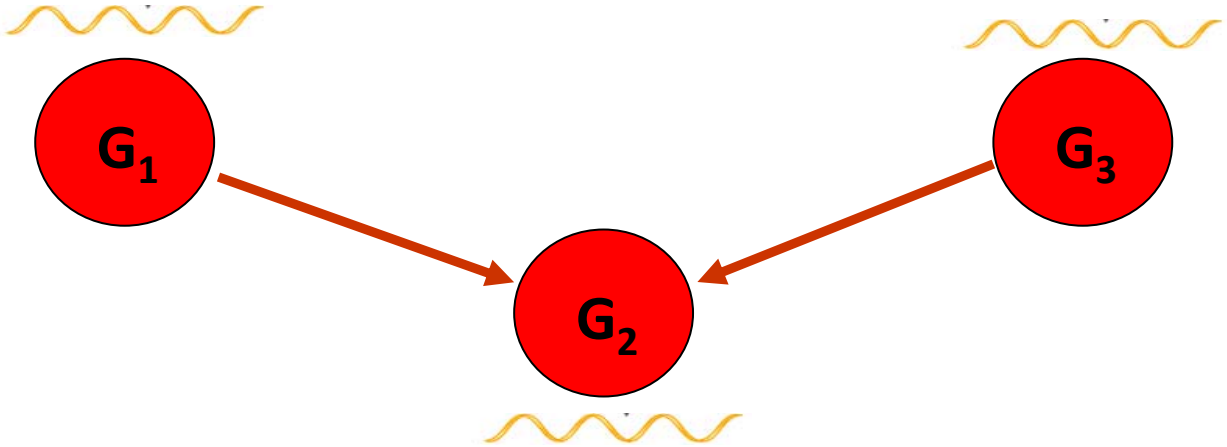
Expressions (activities) of thousands of genes in an experimental cell can be measured with Microarray Chips.



(Gen-)Regulatory Network



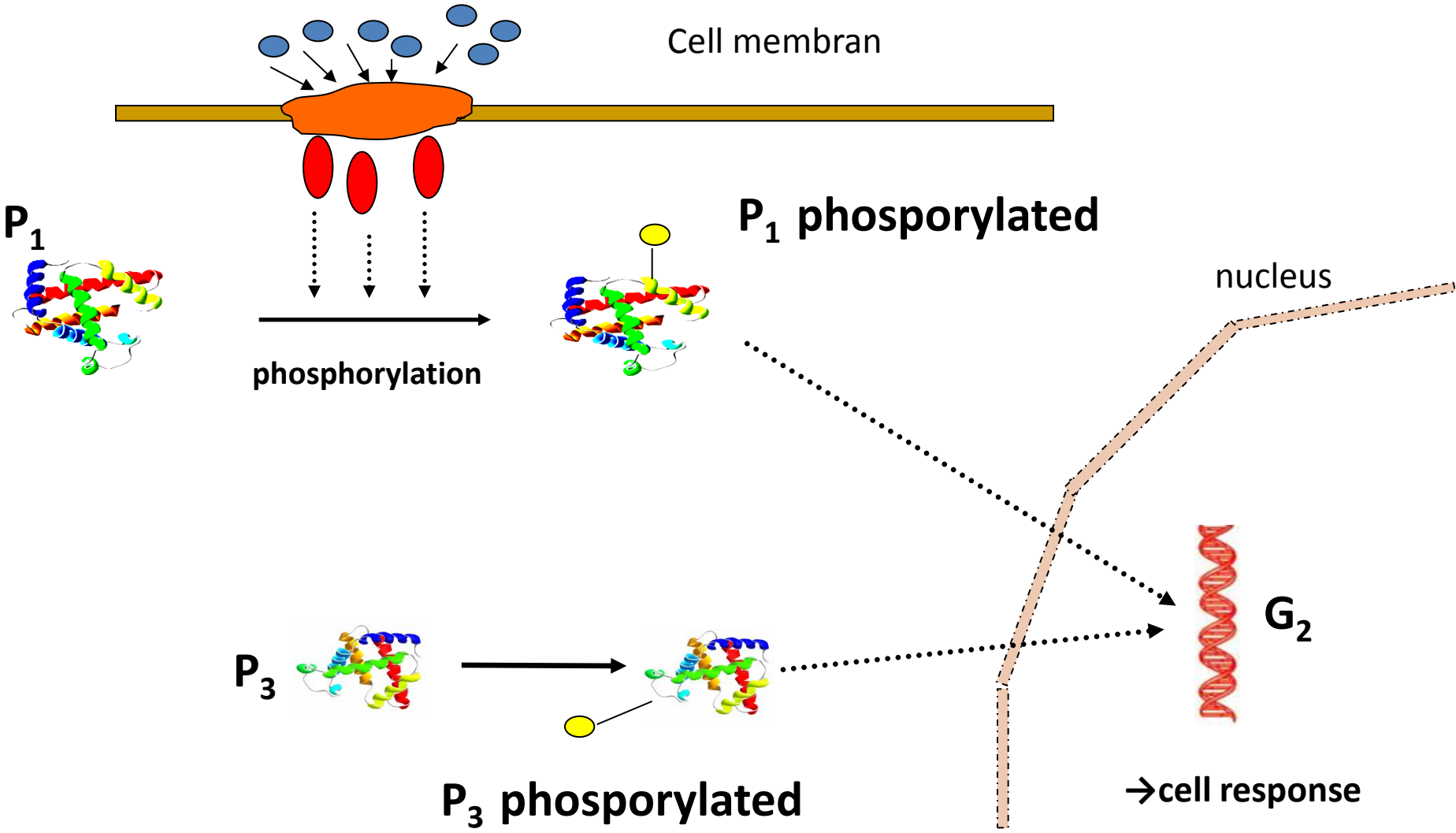
Gen-Regulatory Network



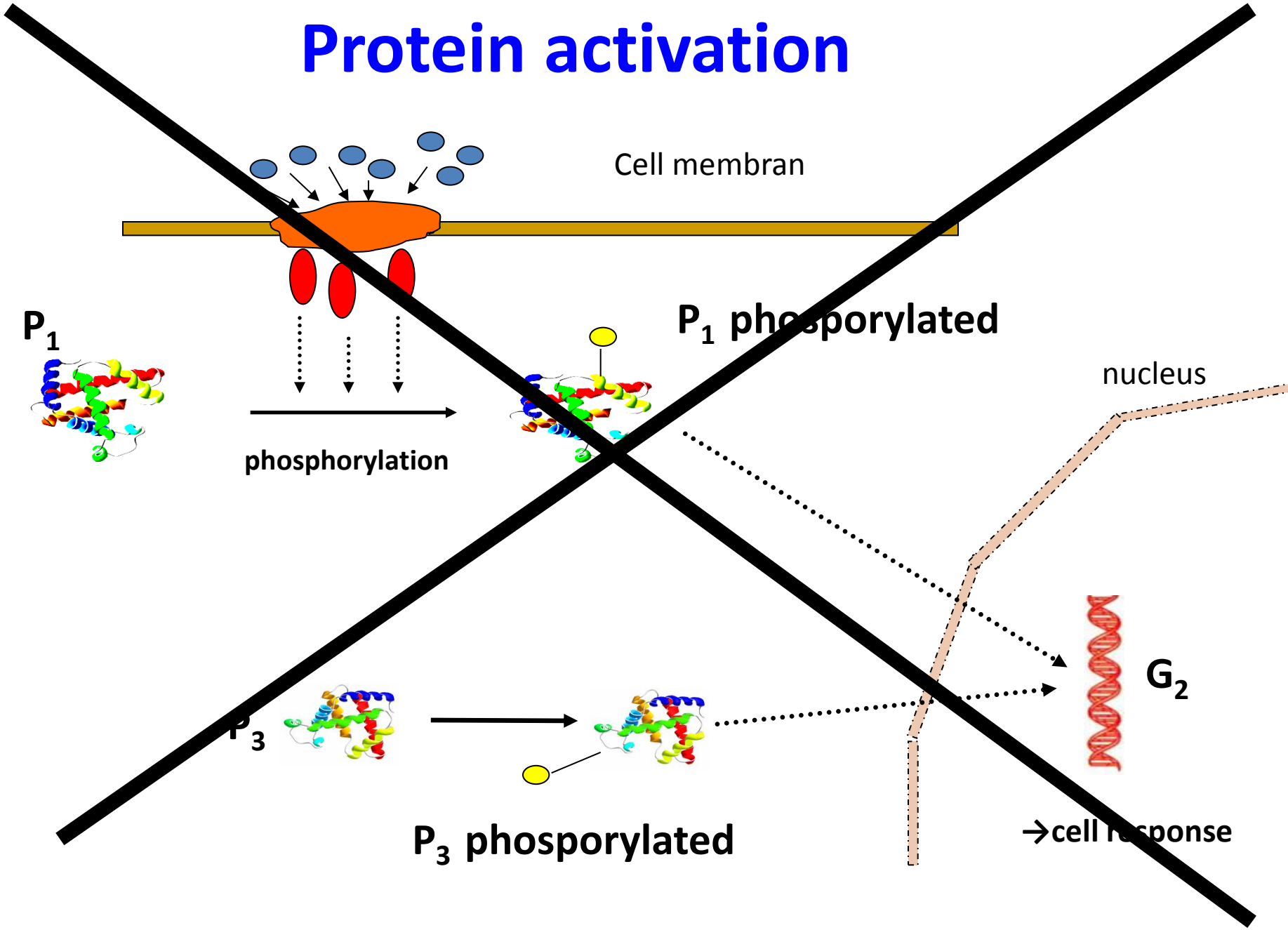
Goal: Learn from gene expression data that
gene 1 and gene 3 co-regulate gene 2

Remark: In gene regulatory networks **the protein level is ignored.**
That is, proteins may build complexes with each other or may have to be activated (e.g. phosphorylated) before they can bind to binding sites of genes**.**

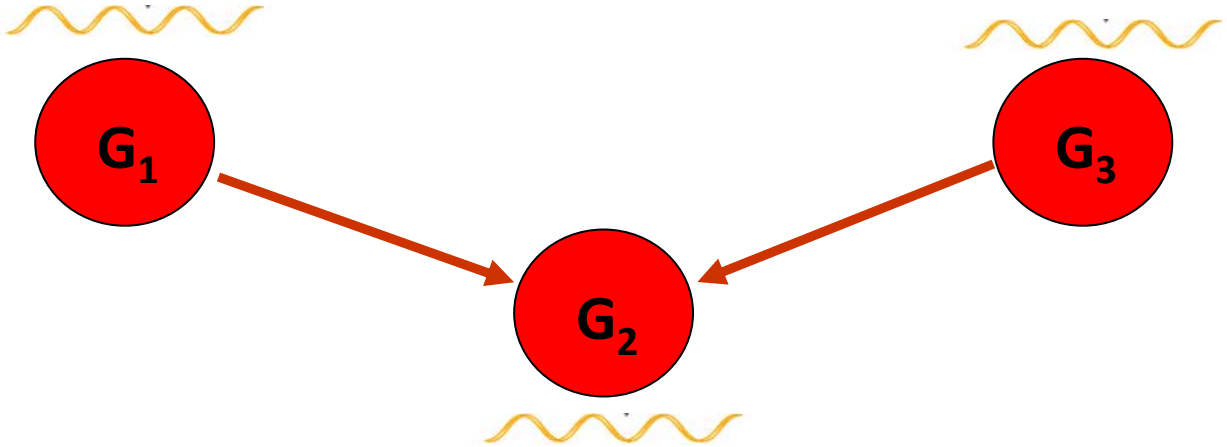
Protein activation



Protein activation



Gen-Regulatory Network



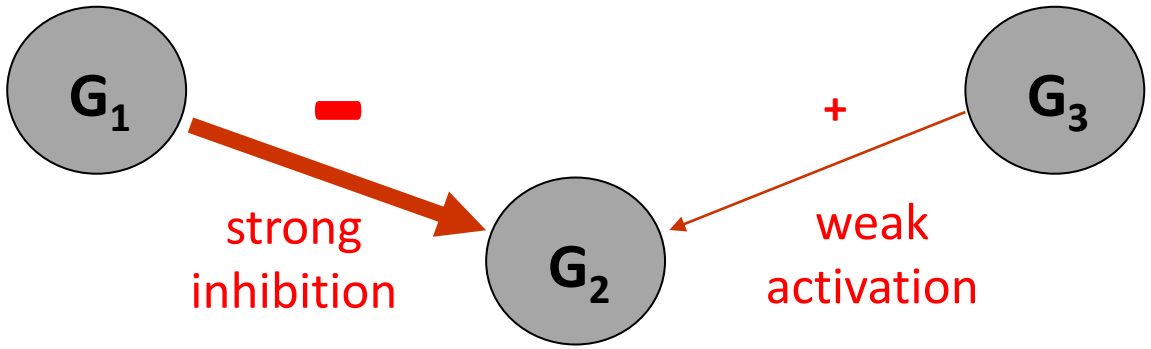
Goal: Learn from gene expression data that
gene 1 and gene 3 co-regulate gene 2

Remark: In gene regulatory networks **the protein level is ignored.**
That is, proteins may build complexes with each other or may have to be activated (e.g. phosphorylated) before they can bind to binding sites of genes**.**

Medical relevance
e.g. for tumour development
-- simplified example --

gene1 may be a
tumour suppressor gene

gene 3 may be an
oncogene

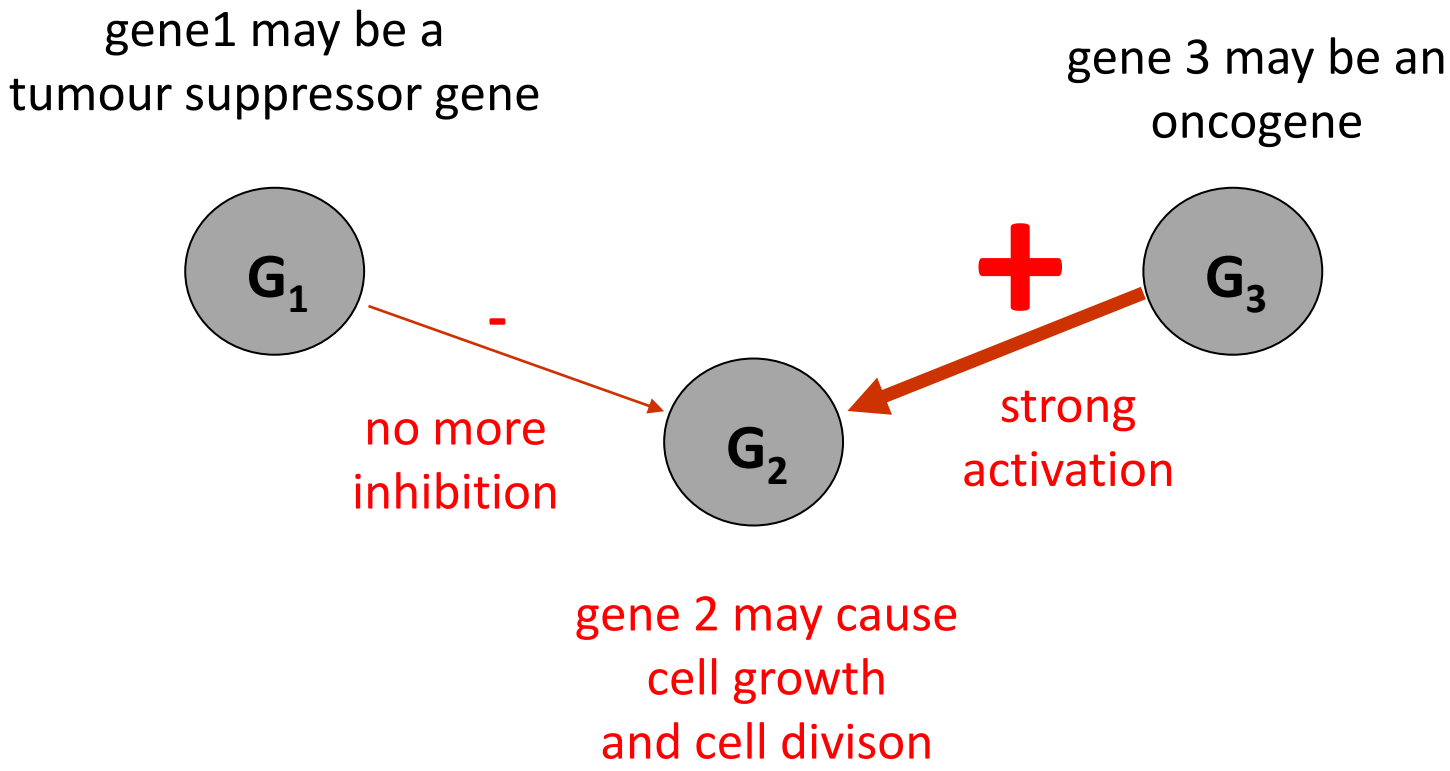


gene 2 may cause
cell growth
and cell divison

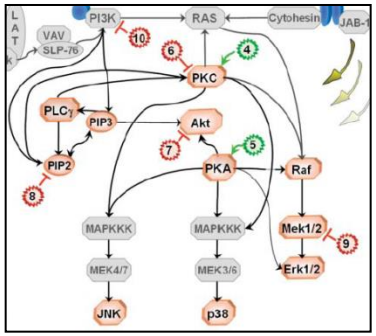
Healthy condition

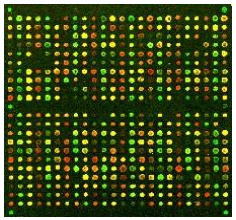
cell division is under control

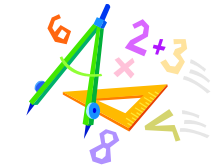
Medical relevance
e.g. for tumour development
-- simplified example --

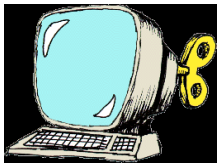


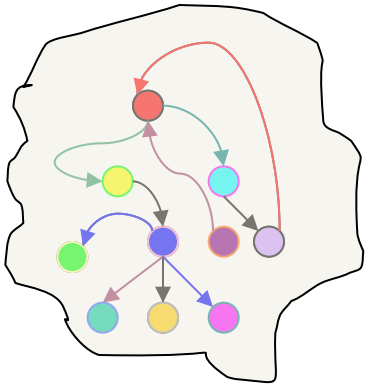
Tumour cell  **Altered pathway leads to uncontrolled cell division**


PI3K
RAS
Cytohesin
JAB-1
VAV
SLP-76
10
6
4
PKC
PLCγ
PIP3
Akt
7
5
PIP2
PKA
Raf
8
MAPKKK
MAPKKK
Mek1/2
9
MEK4/7
MEK3/6
Erk1/2
JNK
p38

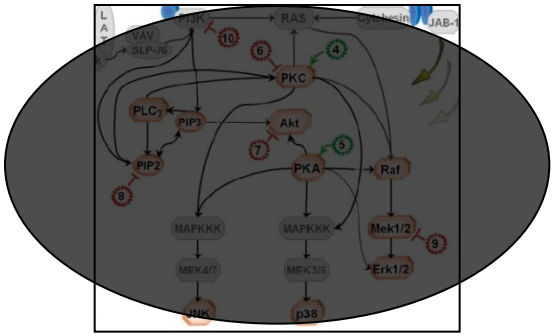


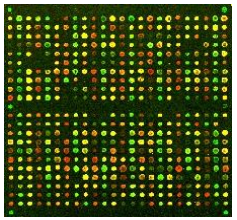


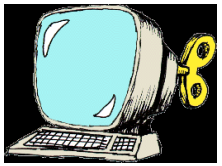


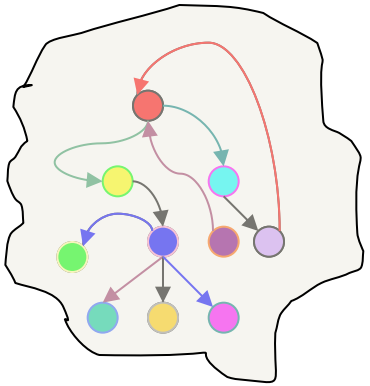


possibly completely unknown

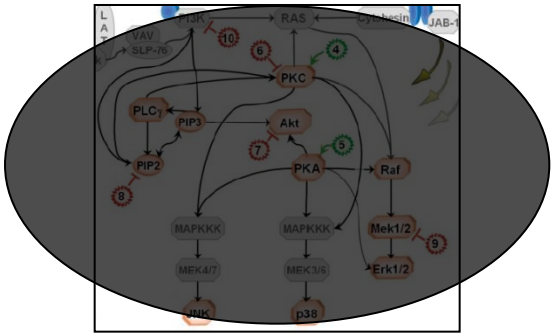




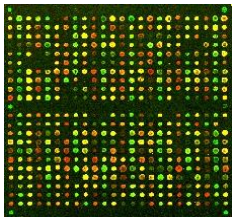




possibly completely unknown

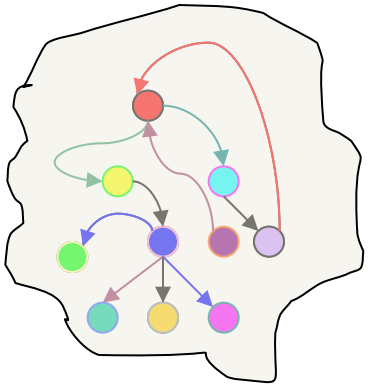


E.g.: Gene-Microarry experiments

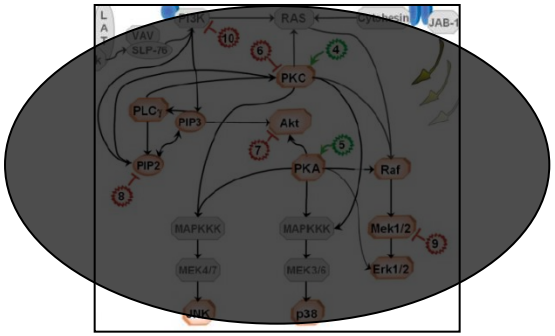


data

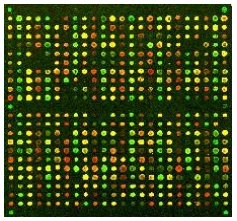
(expressions of genes)



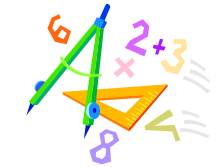
possibly completely unknown

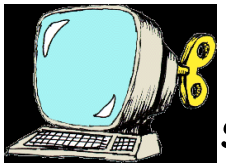


E.g.: Gene-Microarry experiments



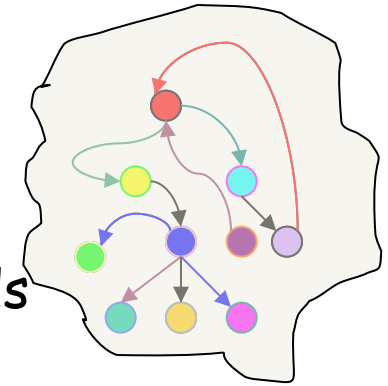
data **data**





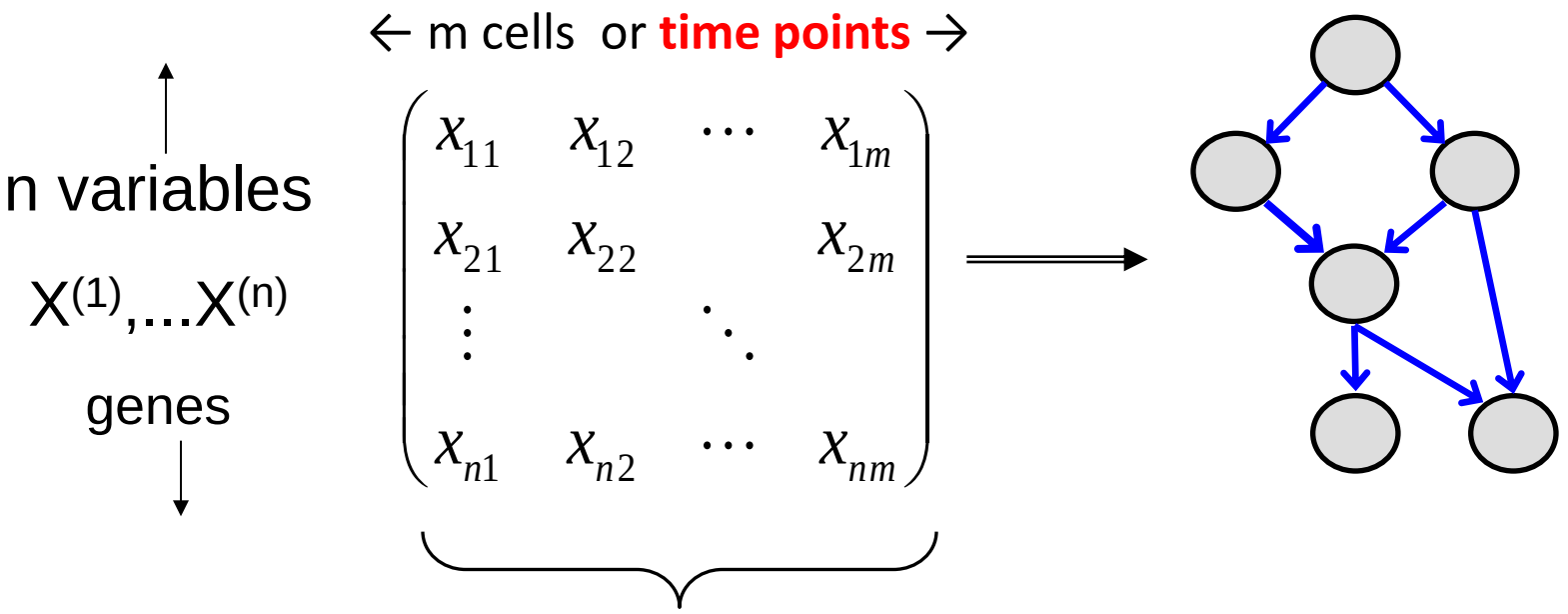
Machine Learning

statistical methods



Statistical Task

Extract a network from an n-by-m data matrix



Either m independent (steady-state) observations

of the system $X^{(1)},\ldots,X^{(n)}$

Or **time series** of the system of length m: $(X^{(1)},\ldots,X^{(n)})_{t=1,\ldots,m}$

Dynamic Bayesian networks

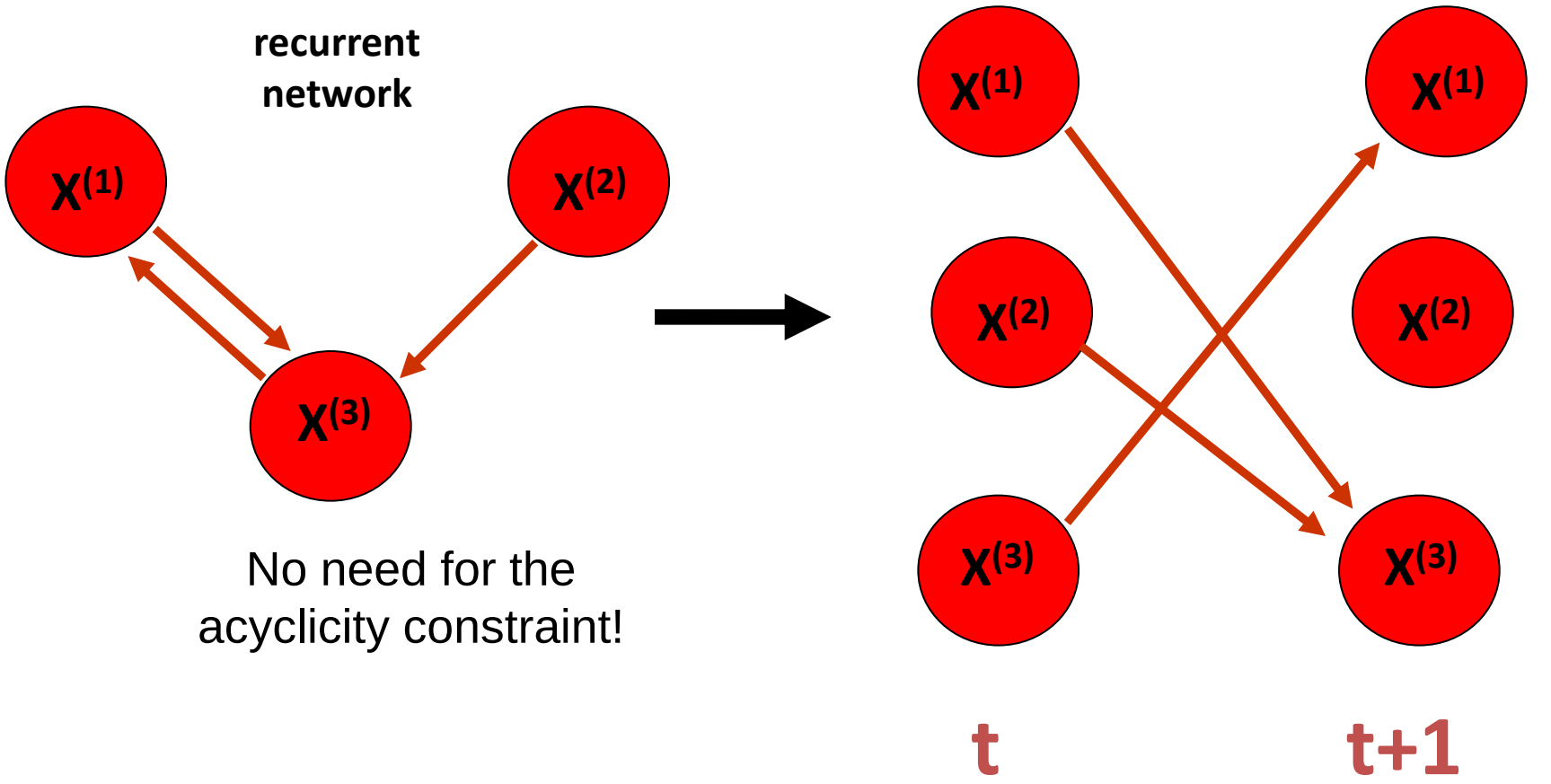
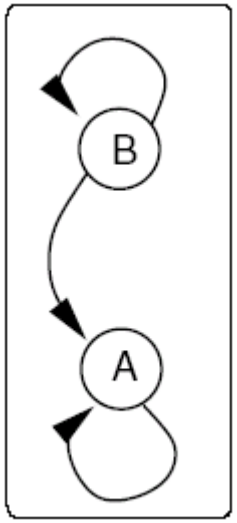


Illustration: Simple dynamic Bayesian network (DBN) with three nodes.
All interactions are subject to a time delay.

Static/dynamic Bayesian networks



Static Bayesian networks

Important feature: Network has to be acyclic

Implied factorisation:

~~$P(A,B) = P(B|B) \cdot P(A|A,B)$~~

Dynamic Bayesian networks

Network does not have to be acyclic

Implied factorisation:

$$P(A(t),B(t)|A(t-1),B(t-1)) = P(B(t)|B(t-1)) \cdot P(A(t)|A(t-1),B(t-1))$$

$(t=2,\ldots,m)$

→ cycles cannot make sense

Model assumption: **Homogeneous** Markov chain

Example: 4 genes, 10 time points

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}
$X^{(1)}$	$X_{1,1}$	$X_{1,2}$	$X_{1,3}$	$X_{1,4}$	$X_{1,5}$	$X_{1,6}$	$X_{1,7}$	$X_{1,8}$	$X_{1,9}$	$X_{1,10}$
$X^{(2)}$	$X_{2,1}$	$X_{2,2}$	$X_{2,3}$	$X_{2,4}$	$X_{2,5}$	$X_{2,6}$	$X_{2,7}$	$X_{2,8}$	$X_{2,9}$	$X_{2,10}$
$X^{(3)}$	$X_{3,1}$	$X_{3,2}$	$X_{3,3}$	$X_{3,4}$	$X_{3,5}$	$X_{3,6}$	$X_{3,7}$	$X_{3,8}$	$X_{3,9}$	$X_{3,10}$
$X^{(4)}$	$X_{4,1}$	$X_{4,2}$	$X_{4,3}$	$X_{4,4}$	$X_{4,5}$	$X_{4,6}$	$X_{4,7}$	$X_{4,8}$	$X_{4,9}$	$X_{4,10}$

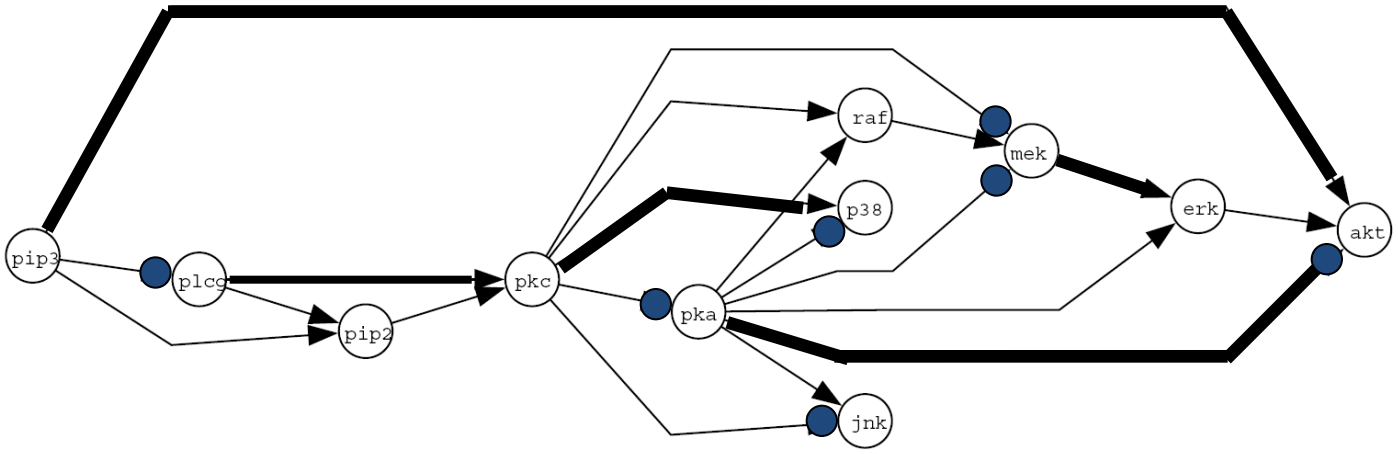
Impose changepoints to model non-homogeneous processes

changepoint ↓

		FIRST SEGMENT					SECOND SEGMENT			
$X^{(1)}$	$X_{1,1}$	$X_{1,2}$	$X_{1,3}$	$X_{1,4}$	$X_{1,5}$	$X_{1,6}$	$X_{1,7}$	$X_{1,8}$	$X_{1,9}$	$X_{1,10}$
$X^{(2)}$	$X_{2,1}$	$X_{2,2}$	$X_{2,3}$	$X_{2,4}$	$X_{2,5}$	$X_{2,6}$	$X_{2,7}$	$X_{2,8}$	$X_{2,9}$	$X_{2,10}$
$X^{(3)}$	$X_{3,1}$	$X_{3,2}$	$X_{3,3}$	$X_{3,4}$	$X_{3,5}$	$X_{3,6}$	$X_{3,7}$	$X_{3,8}$	$X_{3,9}$	$X_{3,10}$
$X^{(4)}$	$X_{4,1}$	$X_{4,2}$	$X_{4,3}$	$X_{4,4}$	$X_{4,5}$	$X_{4,6}$	$X_{4,7}$	$X_{4,8}$	$X_{4,9}$	$X_{4,10}$

Changepoint model

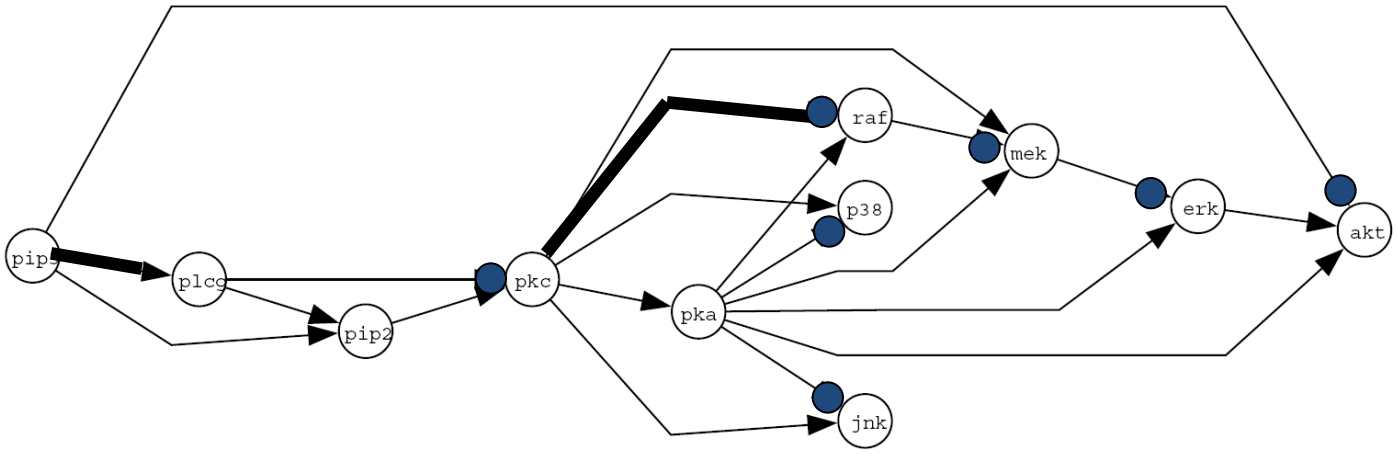
Our paradigm: Keep the network topology fixed but the interaction parameters can change with time.



Interaction parameters in the first segment

Changepoint model

Our paradigm: Keep the network topology fixed but the interaction parameters can change with time.



interaction parameters in the second segment

Introduce gene-specific changepoints to increase flexibility of the models

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}
$X^{(1)}$	$X_{1,1}$	$X_{1,2}$	$X_{1,3}$	$X_{1,4}$	$X_{1,5}$	$X_{1,6}$	$X_{1,7}$	$X_{1,8}$	$X_{1,9}$	$X_{1,10}$
$X^{(2)}$	$X_{2,1}$	$X_{2,2}$	$X_{2,3}$	$X_{2,4}$	$X_{2,5}$	$X_{2,6}$	$X_{2,7}$	$X_{2,8}$	$X_{2,9}$	$X_{2,10}$
$X^{(3)}$	$X_{3,1}$	$X_{3,2}$	$X_{3,3}$	$X_{3,4}$	$X_{3,5}$	$X_{3,6}$	$X_{3,7}$	$X_{3,8}$	$X_{3,9}$	$X_{3,10}$
$X^{(4)}$	$X_{4,1}$	$X_{4,2}$	$X_{4,3}$	$X_{4,4}$	$X_{4,5}$	$X_{4,6}$	$X_{4,7}$	$X_{4,8}$	$X_{4,9}$	$X_{4,10}$

Non-Homogeneous Dynamic Bayesian Networks (NH-DBN)

Idea: Combine a standard DBN with a node-specific multiple changepoint process.

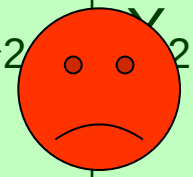
Lèbre, Becq, Devaux, Lelandais, Stumpf (2010)
Statistical inference of the time-varying structure of gene regulation networks
BMC Systems Biology

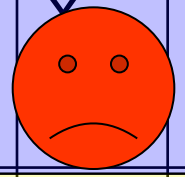
Robinson & Hartemink (2010)
Learning non-stationary dynamic Bayesian networks
Journal of Machine Learning Research

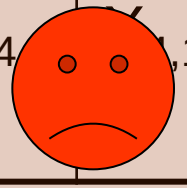
What is the problem with these approaches?

Practical problem: inference uncertainty in short time series segments

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}
$X^{(1)}$	$X_{1,1}$	$X_{1,2}$	$X_{1,3}$	$X_{1,4}$	$X_{1,5}$	$X_{1,6}$	$X_{1,7}$	$X_{1,8}$	$X_{1,9}$	$X_{1,10}$
$X^{(2)}$	$X_{2,1}$	$X_{2,2}$	$X_{2,3}$	$X_{2,4}$	$X_{2,5}$	$X_{2,6}$	$X_{2,7}$	$X_{2,8}$	$X_{2,9}$	$X_{2,10}$
$X^{(3)}$	$X_{3,1}$	$X_{3,2}$	$X_{3,3}$	$X_{3,4}$	$X_{3,5}$	$X_{3,6}$	$X_{3,7}$	$X_{3,8}$	$X_{3,9}$	$X_{3,10}$
$X^{(4)}$	$X_{4,1}$	$X_{4,2}$	$X_{4,3}$	$X_{4,4}$	$X_{4,5}$	$X_{4,6}$	$X_{4,7}$	$X_{4,8}$	$X_{4,9}$	$X_{4,10}$







Shortcomings

1. Practical problem

Short time series

inference uncertainty

2. Methodological problem

Prior independence is

biologically implausible

Is it plausible to assume a priori that the segment-specific interaction parameters are independent?

Idea: Information coupling among segments

Non-homogeneous DBN
(uncoupled NH-DBN)

Information coupling with respect to the interaction parameters (coupled NH-DBN)

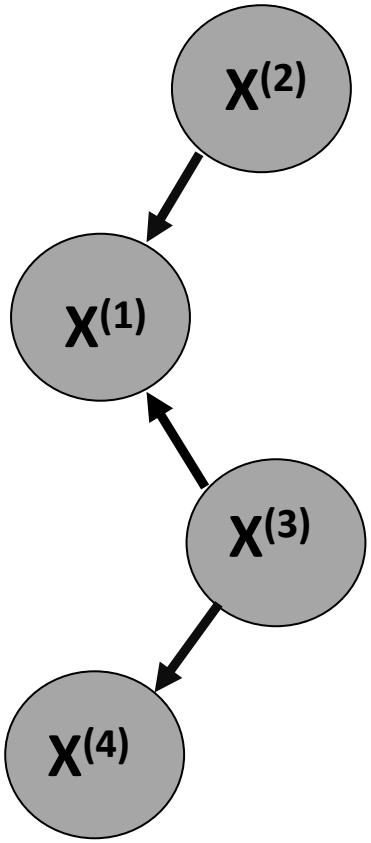
Grzegorczyk and Husmeier (2012a)
A non-homogeneous dynamic Bayesian network model with **sequentially** coupled interaction parameters for applications in systems and synthetic biology.
SAGMB

Grzegorczyk and Husmeier (2012b)
Bayesian regularization of non-homogeneous dynamic Bayesian networks by **globally** coupling interaction parameters.
AISTATS

Grzegorczyk and Husmeier (2013)
Regularization of Non-Homogeneous Dynamic Bayesian Networks with **Global** Information-Coupling based on Hierarchical Bayesian models.
Machine Learning

Bayesian regression models

complete network

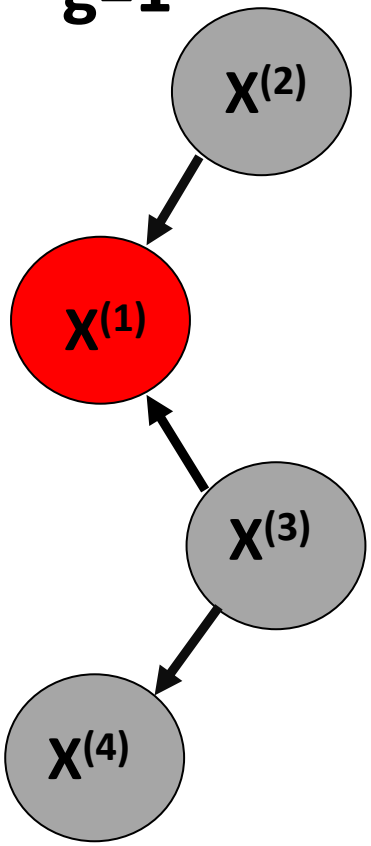


complete segmentation matrix

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}
$X^{(1)}$	$X_{1,1}$	$X_{1,2}$	$X_{1,3}$	$X_{1,4}$	$X_{1,5}$	$X_{1,6}$	$X_{1,7}$	$X_{1,8}$	$X_{1,9}$	$X_{1,10}$
$X^{(2)}$	$X_{2,1}$	$X_{2,2}$	$X_{2,3}$	$X_{2,4}$	$X_{2,5}$	$X_{2,6}$	$X_{2,7}$	$X_{2,8}$	$X_{2,9}$	$X_{2,10}$
$X^{(3)}$	$X_{3,1}$	$X_{3,2}$	$X_{3,3}$	$X_{3,4}$	$X_{3,5}$	$X_{3,6}$	$X_{3,7}$	$X_{3,8}$	$X_{3,9}$	$X_{3,10}$
$X^{(4)}$	$X_{4,1}$	$X_{4,2}$	$X_{4,3}$	$X_{4,4}$	$X_{4,5}$	$X_{4,6}$	$X_{4,7}$	$X_{4,8}$	$X_{4,9}$	$X_{4,10}$

Bayesian regression models

first gene
g=1

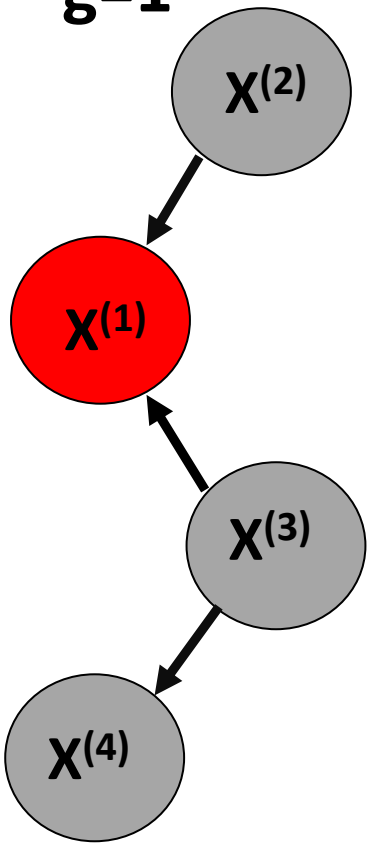


segmentation of node g=1

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}
$X^{(1)}$	$X_{1,1}$	$X_{1,2}$	$X_{1,3}$	$X_{1,4}$	$X_{1,5}$	$X_{1,6}$	$X_{1,7}$	$X_{1,8}$	$X_{1,9}$	$X_{1,10}$
$X^{(2)}$	$X_{2,1}$	$X_{2,2}$	$X_{2,3}$	$X_{2,4}$	$X_{2,5}$	$X_{2,6}$	$X_{2,7}$	$X_{2,8}$	$X_{2,9}$	$X_{2,10}$
$X^{(3)}$	$X_{3,1}$	$X_{3,2}$	$X_{3,3}$	$X_{3,4}$	$X_{3,5}$	$X_{3,6}$	$X_{3,7}$	$X_{3,8}$	$X_{3,9}$	$X_{3,10}$
$X^{(4)}$	$X_{4,1}$	$X_{4,2}$	$X_{4,3}$	$X_{4,4}$	$X_{4,5}$	$X_{4,6}$	$X_{4,7}$	$X_{4,8}$	$X_{4,9}$	$X_{4,10}$

Bayesian regression models

first gene
g=1



$y_{g=1,h=1} = (X_{1,2},\ldots, X_{1,6})$

$y_{g=1,h=2} = (X_{1,7},\ldots, X_{1,10})$

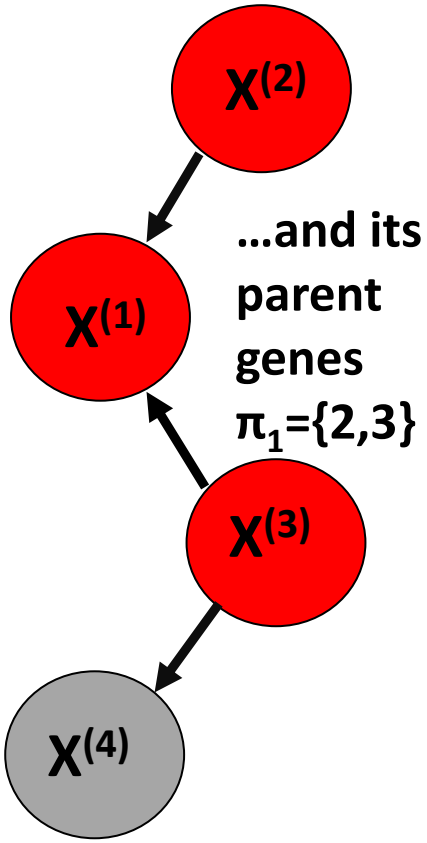
	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}
$X^{(1)}$	$X_{1,1}$	h=1					h=2			
$X^{(2)}$	$X_{2,1}$	$X_{2,2}$	$X_{2,3}$	$X_{2,4}$	$X_{2,5}$	$X_{2,6}$	$X_{2,7}$	$X_{2,8}$	$X_{2,9}$	$X_{2,10}$
$X^{(3)}$	$X_{3,1}$	$X_{3,2}$	$X_{3,3}$	$X_{3,4}$	$X_{3,5}$	$X_{3,6}$	$X_{3,7}$	$X_{3,8}$	$X_{3,9}$	$X_{3,10}$
$X^{(4)}$	$X_{4,1}$	$X_{4,2}$	$X_{4,3}$	$X_{4,4}$	$X_{4,5}$	$X_{4,6}$	$X_{4,7}$	$X_{4,8}$	$X_{4,9}$	$X_{4,10}$

changepoint $\tau_{g=1,1} = 6$

This changepoint divides the observations of node $X^{(1)}$ into $\mathbf{K_{g=1}=2}$ disjunct segments.

Bayesian regression models

first gene
g=1



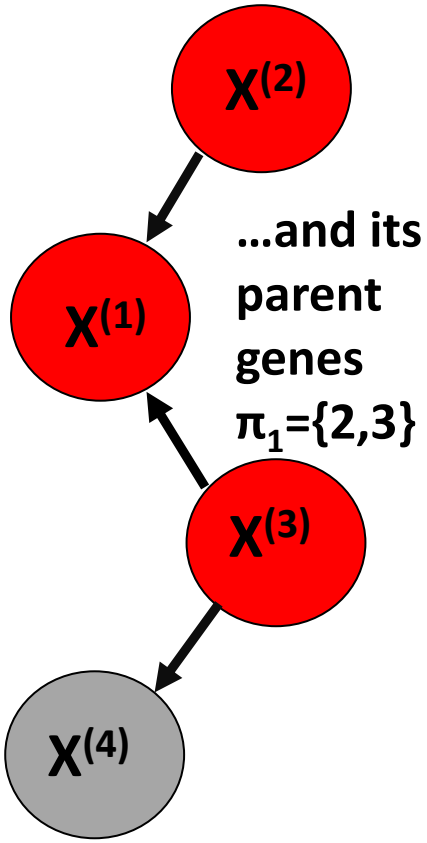
$$y_{g=1,h=1} = (X_{1,2},\ldots, X_{1,6})^T \quad y_{g=1,h=2} = (X_{1,7},\ldots, X_{1,10})^T$$

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}
$X^{(1)}$	$X_{1,1}$	$X_{1,2}$	$X_{1,3}$	$X_{1,4}$	$X_{1,5}$	$X_{1,6}$	$X_{1,7}$	$X_{1,8}$	$X_{1,9}$	$X_{1,10}$
$X^{(2)}$	$X_{2,1}$	$X_{2,2}$	$X_{2,3}$	$X_{2,4}$	$X_{2,5}$	$X_{2,6}$	$X_{2,7}$	$X_{2,8}$	$X_{2,9}$	$X_{2,10}$
$X^{(3)}$	$X_{3,1}$	$X_{3,2}$	$X_{3,3}$	$X_{3,4}$	$X_{3,5}$	$X_{3,6}$	$X_{3,7}$	$X_{3,8}$	$X_{3,9}$	$X_{3,10}$
$X^{(4)}$	$X_{4,1}$	$X_{4,2}$	$X_{4,3}$	$X_{4,4}$	$X_{4,5}$	$X_{4,6}$	$X_{4,7}$	$X_{4,8}$	$X_{4,9}$	$X_{4,10}$

For both segments h=1 and h=2 determine the observations which belong to the parent nodes of $X^{(1)}$.
Note that all interactions are subject to a **time lag of size 1**.

Bayesian regression models

first gene
g=1



…and its parent genes π_1={2,3}

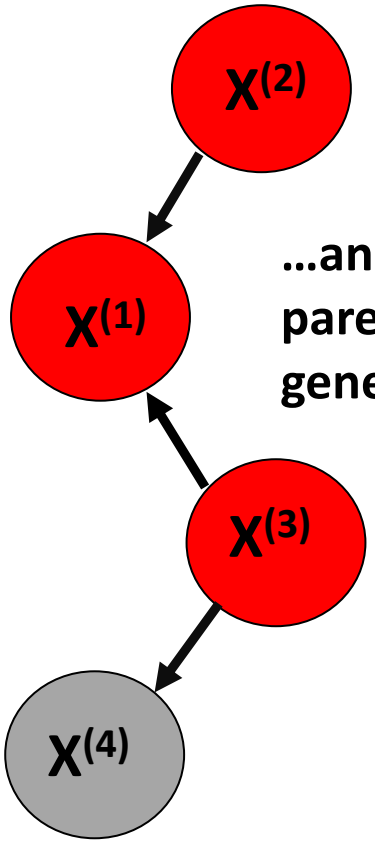
$$y_{g=1,h=1} = (X_{1,2},\ldots,X_{1,6})^T \quad y_{g=1,h=2} = (X_{1,7},\ldots,X_{1,10})^T$$

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}
$X^{(1)}$	$X_{1,1}$	$X_{1,2}$	$X_{1,3}$	$X_{1,4}$	$X_{1,5}$	$X_{1,6}$	$X_{1,7}$	$X_{1,8}$	$X_{1,9}$	$X_{1,10}$
$X^{(2)}$	$X_{2,1}$	$X_{2,2}$	$X_{2,3}$	$X_{2,4}$	$X_{2,5}$	$X_{2,6}$	$X_{2,7}$	$X_{2,8}$	$X_{2,9}$	$X_{2,10}$
$X^{(3)}$	$X_{3,1}$	$X_{3,2}$	$X_{3,3}$	$X_{3,4}$	$X_{3,5}$	$X_{3,6}$	$X_{3,7}$	$X_{3,8}$	$X_{3,9}$	$X_{3,10}$
$X^{(4)}$	$X_{4,1}$	$X_{4,2}$	$X_{4,3}$	$X_{4,4}$	$X_{4,5}$	$X_{4,6}$	$X_{4,7}$	$X_{4,8}$	$X_{4,9}$	$X_{4,10}$

For both segments h=1 and h=2 determine the observations which belong to the parent nodes of $X^{(1)}$.
Note that all interactions are subject to a **time lag of size 1**.

Bayesian regression models

first gene g=1…



…and its parent genes

$$y_{g=1,h=1} = (X_{1,2},\ldots, X_{1,6})^T \quad y_{g=1,h=2} = (X_{1,7},\ldots, X_{1,10})^T$$

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}
$X^{(1)}$	$X_{1,1}$	$X_{1,2}$	$X_{1,3}$	$X_{1,4}$	$X_{1,5}$	$X_{1,6}$	$X_{1,7}$	$X_{1,8}$	$X_{1,9}$	$X_{1,10}$
$X^{(2)}$	$X_{2,1}$	$X_{2,2}$	$X_{2,3}$	$X_{2,4}$	$X_{2,5}$	$X_{2,6}$	$X_{2,7}$	$X_{2,8}$	$X_{2,9}$	$X_{2,10}$
$X^{(3)}$	$X_{3,1}$	$X_{3,2}$	$X_{3,3}$	$X_{3,4}$	$X_{3,5}$	$X_{3,6}$	$X_{3,7}$	$X_{3,8}$	$X_{3,9}$	$X_{3,10}$
$X^{(4)}$	$X_{4,1}$	$X_{4,2}$	$X_{4,3}$	$X_{4,4}$	$X_{4,5}$	$X_{4,6}$	$X_{4,7}$	$X_{4,8}$	$X_{4,9}$	$X_{4,10}$

$$X_{\pi_1=\{2,3\},h=1} = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ X_{2,1} & X_{2,2} & \cdots & X_{2,5} \\ X_{3,1} & X_{3,2} & \cdots & X_{3,5} \end{pmatrix}$$

$$X_{\pi_1=\{2,3\},h=1} = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ X_{2,6} & X_{2,7} & \cdots & X_{2,9} \\ X_{3,6} & X_{3,7} & \cdots & X_{3,9} \end{pmatrix}$$

For each gene g=1,….,G
and each gene-specific segment h=1,…,K_g:

Likelihood model:

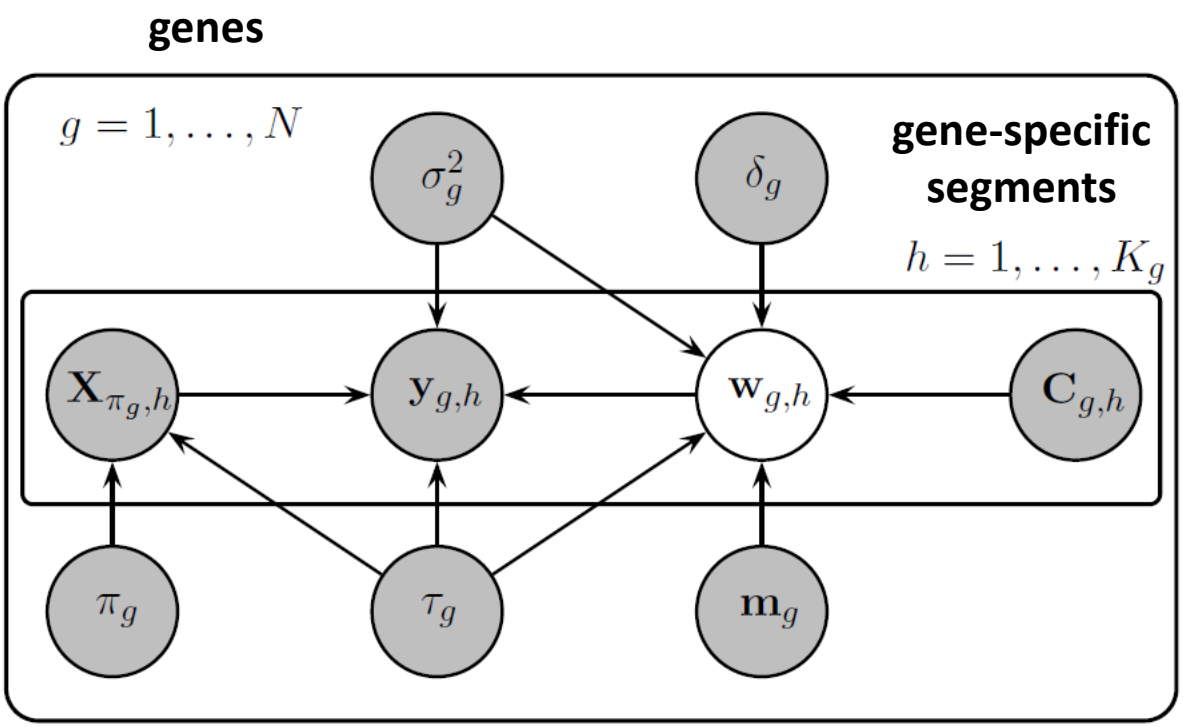
$$\underbrace{\mathbf{y}_{g,h}}_{\text{target observations}} \sim \mathcal{N}(\underbrace{\mathbf{X}^T_{\pi_g,h}}_{\text{regressor matrix}} \underbrace{\mathbf{w}_{g,h}}_{\text{regression coefficients}}, \underbrace{\sigma^2_g}_{\text{noise variance}} \mathbf{I})$$

Prior on the regression coefficients $w_{g,h}$:

$$\mathbf{w}_{g,h} \sim \mathcal{N}(\mathbf{m}_g, \underbrace{\sigma^2_g}_{\text{noise variance}} \underbrace{\delta_g}_{\text{SNR hyperparameter}} \mathbf{C}_{g,h})$$

Note that the explicit dependence on the noise variance leads to a fully conjugate prior.

Graphical representation of the regression models



For $g = 1, \ldots, N$**:**

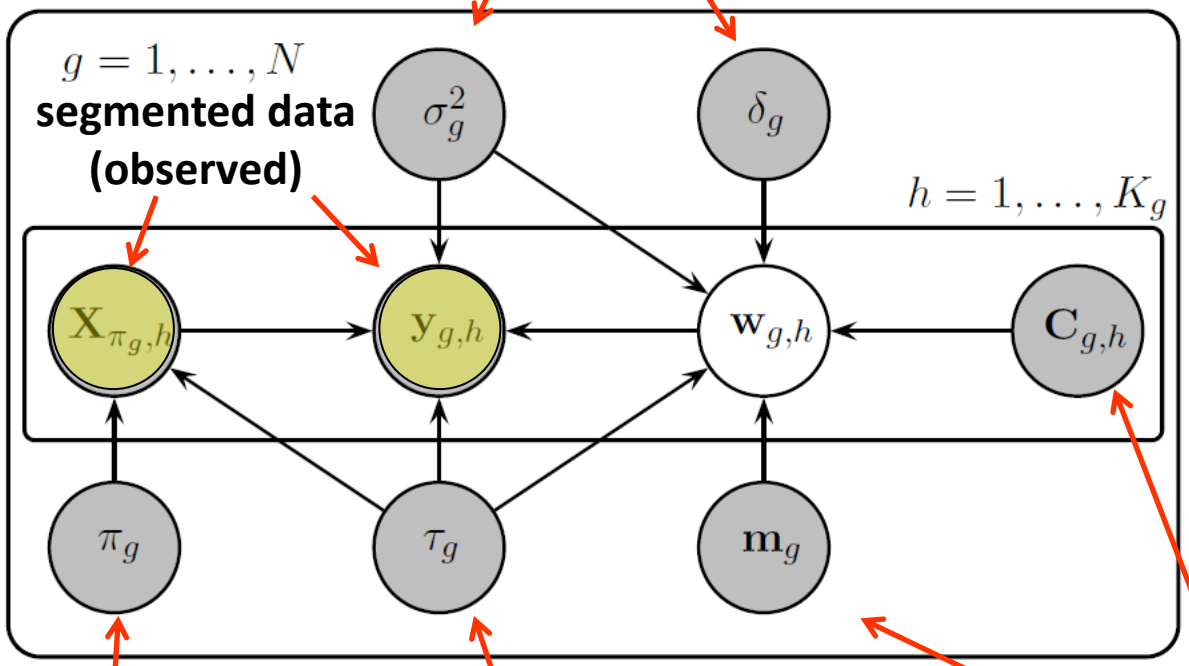
For $h = 1, \ldots, K_g$**:**

$$\mathbf{w}_{g,h} \sim \mathcal{N}(\mathbf{m}_g, \sigma_g^2 \delta_g \mathbf{C}_{g,h})$$

$$\mathbf{y}_{g,h} \sim \mathcal{N}(\mathbf{X}_{\pi_g,h}^T \mathbf{w}_{g,h}, \sigma_g^2 \mathbf{I})$$

$$\tau_g = \{\tau_{g,1}, \ldots, \tau_{g,K_g-1}\}$$

Graphical representation of the regression models



For $g = 1, \ldots, N$**:**

For $h = 1, \ldots, K_g$**:**

$$\mathbf{w}_{g,h} \sim \mathcal{N}(\mathbf{m}_g, \sigma_g^2 \delta_g \mathbf{C}_{g,h})$$

$$\mathbf{y}_{g,h} \sim \mathcal{N}(\mathbf{X}_{\pi_g,h}^T \mathbf{w}_{g,h}, \sigma_g^2 \mathbf{I})$$

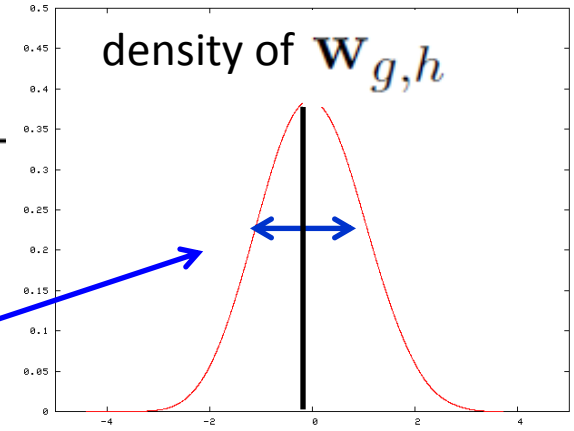
parent sets implied by the network

changepoint set (segmentation)

$$\tau_g = \{\tau_{g,1}, \ldots, \tau_{g,K_g-1}\}$$

fixed hyperparameters

In the absence of any genuine prior knowledge : $\mathbf{m_g=0}$ and $\mathbf{C_{g,h}=I}$



fixed hyperparameters

SNR parameter

$g = 1, \ldots, N$
segmented data (observed)

$h = 1, \ldots, K_g$

σ_g^2 δ_g $\mathbf{X}_{\pi_g,h}$ $\mathbf{y}_{g,h}$ $\mathbf{w}_{g,h}$ $\mathbf{C}_{g,h}$ π_g τ_g $\mathbf{m}_g$

For $g = 1, \ldots, N$**:**

For $h = 1, \ldots, K_g$**:**

$\mathbf{w}_{g,h} \sim \mathcal{N}(\mathbf{m}_g, \sigma_g^2 \delta_g \mathbf{C}_{g,h})$

$\mathbf{y}_{g,h} \sim \mathcal{N}(\mathbf{X}_{\pi_g,h}^T \mathbf{w}_{g,h}, \sigma_g^2 \mathbf{I})$

parent sets implied by the network

changepoint set (segmentation)

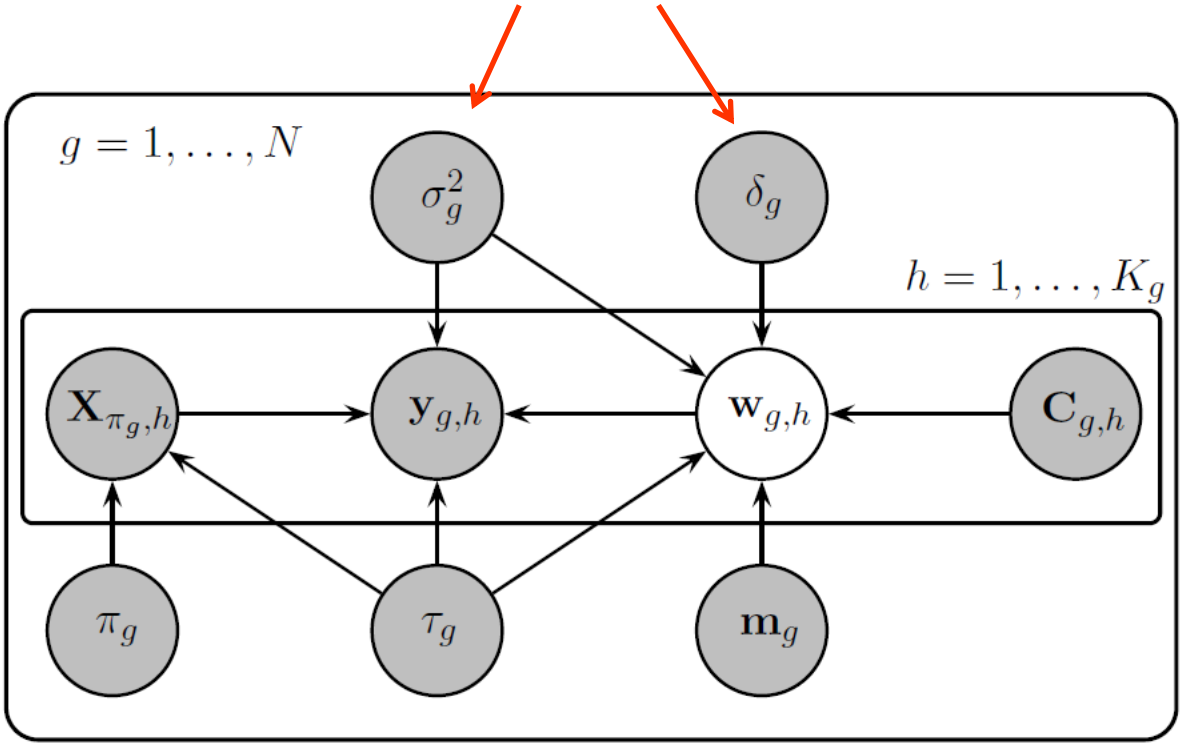
$\tau_g = \{\tau_{g,1}, \ldots, \tau_{g,K_g-1}\}$

fixed hyperparameters

In the absence of any genuine prior knowledge : $\mathbf{m_g=0}$ and $\mathbf{C_{g,h}=I}$

Graphical representation of the regression models

Are these hyperparameters actually known?



For $g = 1, \ldots, N$**:**

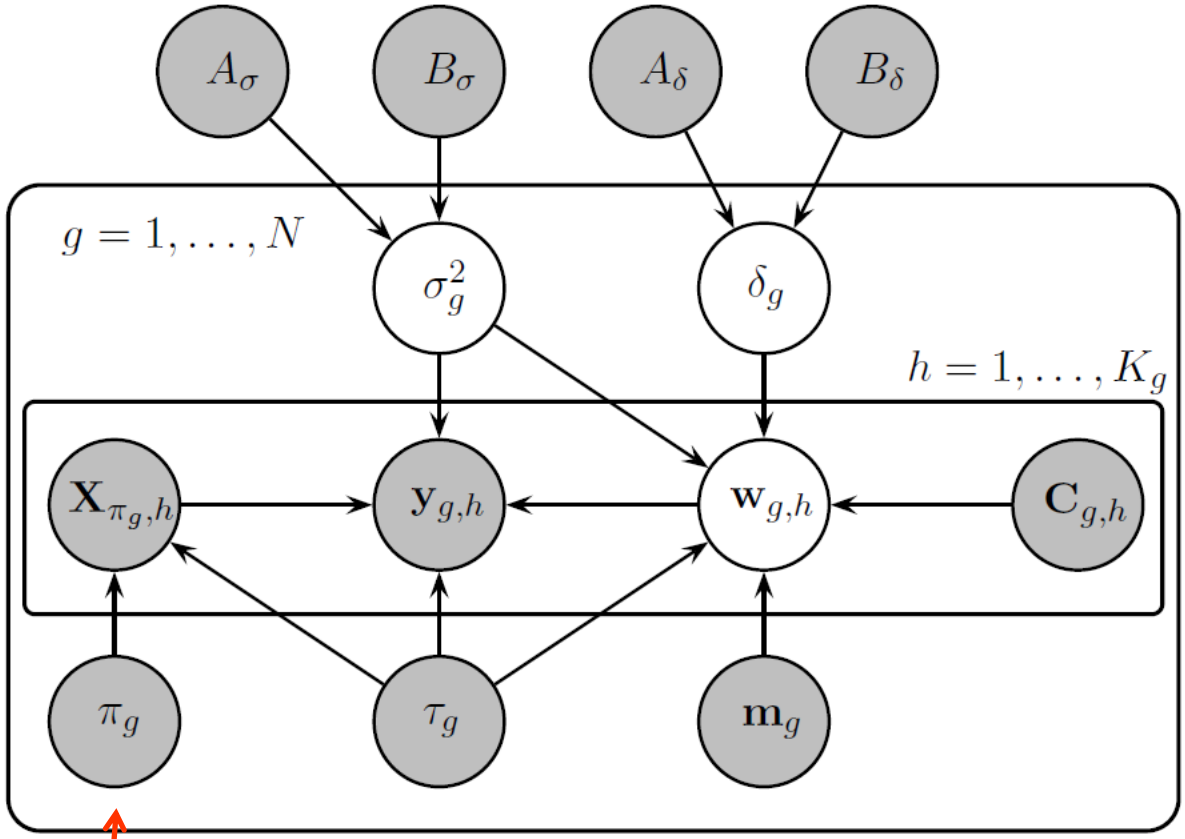
For $h = 1, \ldots, K_g$**:**

$$\mathbf{w}_{g,h} \sim \mathcal{N}(\mathbf{m}_g, \sigma_g^2 \delta_g \mathbf{C}_{g,h})$$

$$\mathbf{y}_{g,h} \sim \mathcal{N}(\mathbf{X}_{\pi_g,h}^T \mathbf{w}_{g,h}, \sigma_g^2 \mathbf{I})$$

$$\tau_g = \{\tau_{g,1}, \ldots, \tau_{g,K_g-1}\}$$

Graphical representation of the regression models



For $g = 1, \ldots, N$**:**

$\sigma_g^{-2} \sim Gam(A_\sigma, B_\sigma)$

$\delta_g^{-1} \sim Gam(A_\delta, B_\delta)$

For $h = 1, \ldots, K_g$**:**

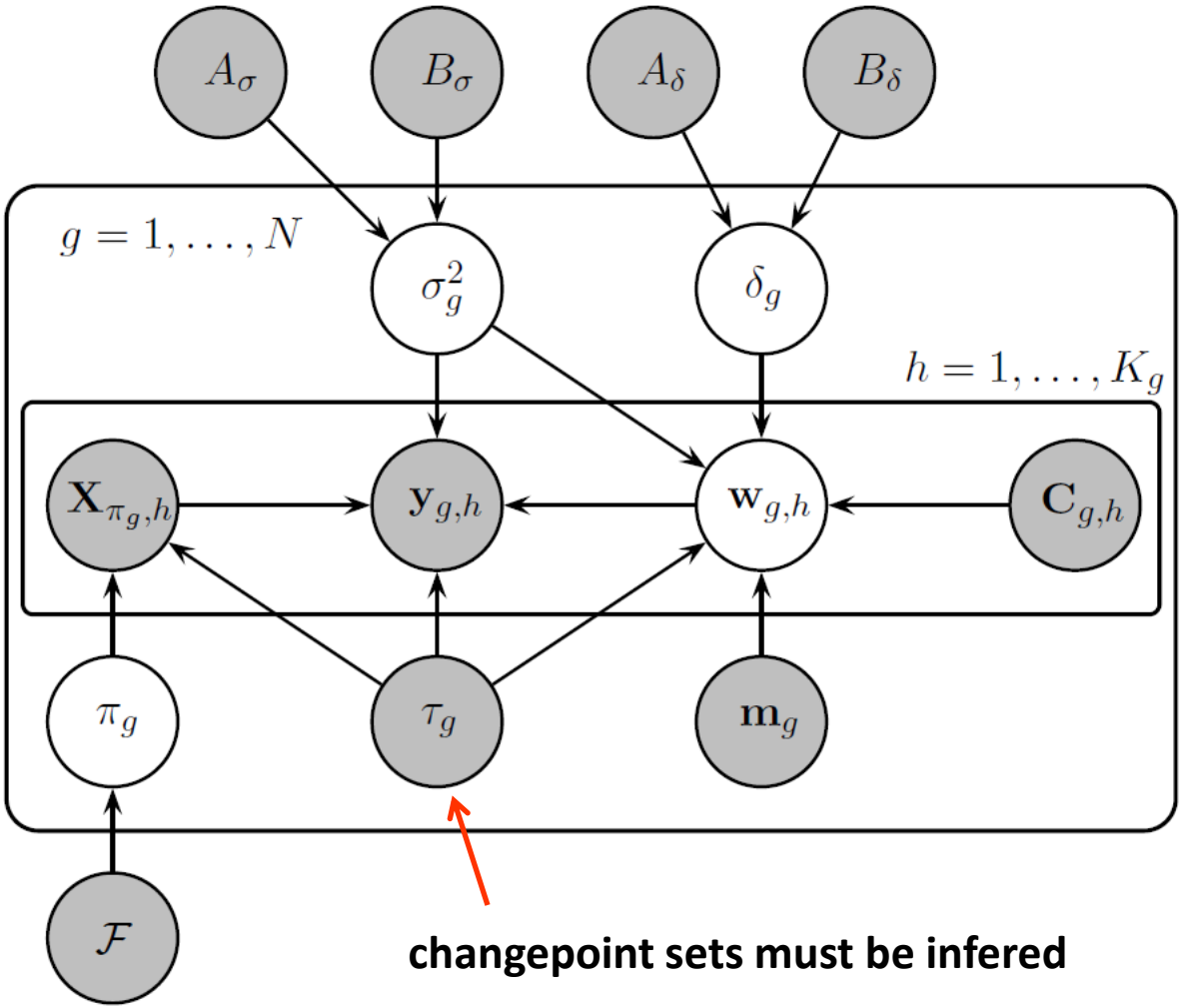
$\mathbf{w}_{g,h} \sim \mathcal{N}(\mathbf{m}_g, \sigma_g^2 \delta_g \mathbf{C}_{g,h})$

$\mathbf{y}_{g,h} \sim \mathcal{N}(\mathbf{X}_{\pi_g,h}^T \mathbf{w}_{g,h}, \sigma_g^2 \mathbf{I})$

parent sets (networks) must be infered

$\tau_g = \{\tau_{g,1}, \ldots, \tau_{g,K_g-1}\}$

Graphical model representation



changepoint sets must be infered

$\pi_g \sim Uni$
$|\pi_g| \leq \mathcal{F}$

$\tau_g = \{\tau_{g,1}, \ldots, \tau_{g,K_g-1}\}$

For $g = 1, \ldots, N$**:**

$\sigma_g^{-2} \sim Gam(A_\sigma, B_\sigma)$

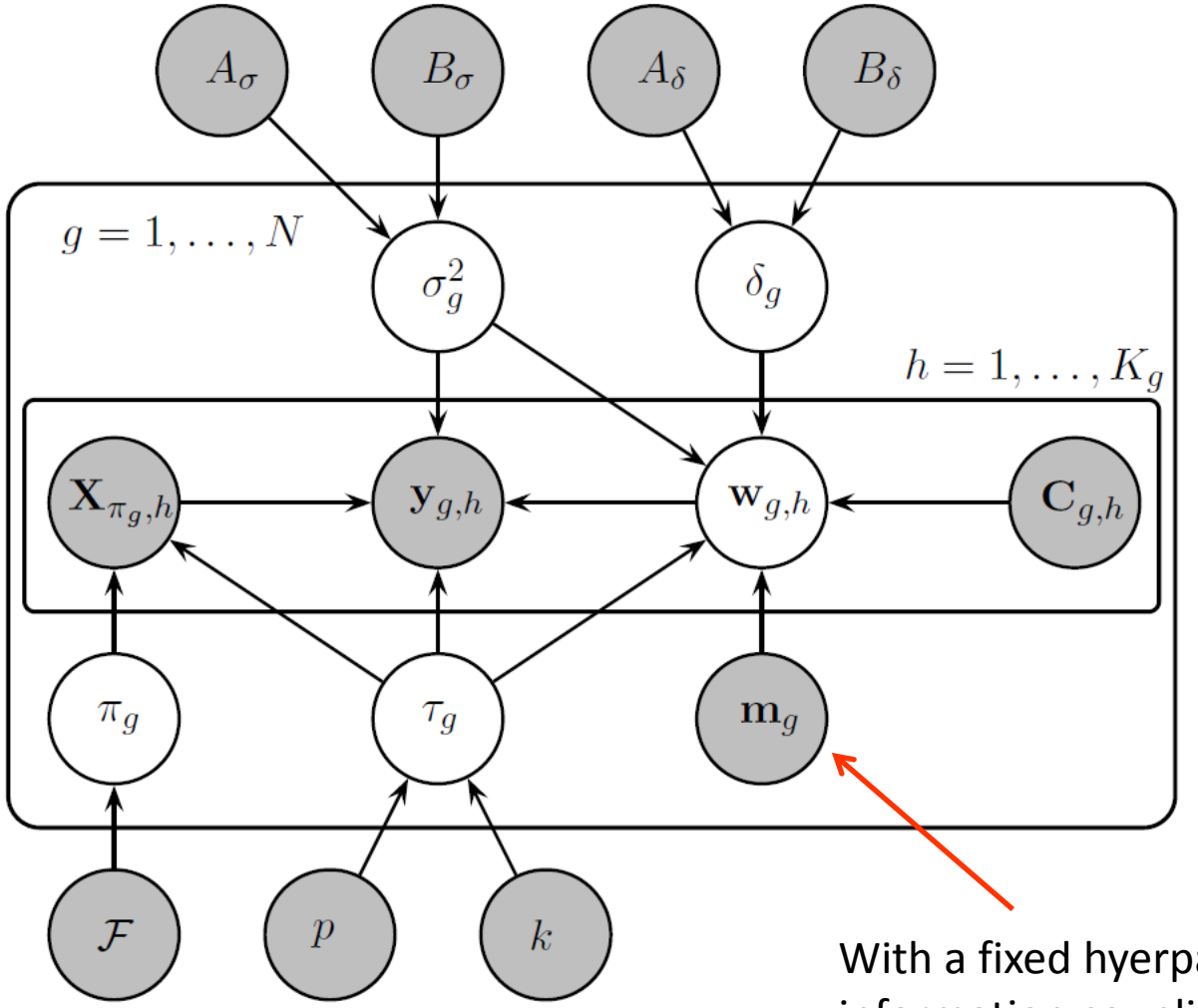
$\delta_g^{-1} \sim Gam(A_\delta, B_\delta)$

For $h = 1, \ldots, K_g$**:**

$\mathbf{w}_{g,h} \sim \mathcal{N}(\mathbf{m}_g, \sigma_g^2 \delta_g \mathbf{C}_{g,h})$

$\mathbf{y}_{g,h} \sim \mathcal{N}(\mathbf{X}_{\pi_g,h}^T \mathbf{w}_{g,h}, \sigma_g^2 \mathbf{I})$

Graphical model representation



For $g = 1, \ldots, N$**:**

$\sigma_g^{-2} \sim Gam(A_\sigma, B_\sigma)$

$\delta_g^{-1} \sim Gam(A_\delta, B_\delta)$

For $h = 1, \ldots, K_g$**:**

$\mathbf{w}_{g,h} \sim \mathcal{N}(\mathbf{m}_g, \sigma_g^2 \delta_g \mathbf{C}_{g,h})$

$\mathbf{y}_{g,h} \sim \mathcal{N}(\mathbf{X}_{\pi_g,h}^T \mathbf{w}_{g,h}, \sigma_g^2 \mathbf{I})$

$\pi_g \sim Uni$

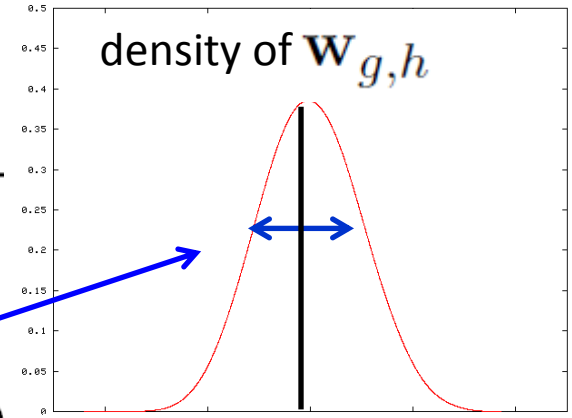
$|\pi_g| \leq \mathcal{F}$

$\tau_g = \{\tau_{g,1}, \ldots, \tau_{g,K_g-1}\}$

$T_{g,h} := \tau_{g,h} - \tau_{g,h-1}$

$T_{g,h} \sim NBIN(p, k)$

With a fixed hyerparameter $\mathbf{m}_g$ there is **no** information coupling between the segment-specific regression coefficients.



$\mathbf{m}_g$ fixed

A_σ B_σ A_δ B_δ

$g = 1, \ldots, N$

σ_g^2 δ_g

$h = 1, \ldots, K_g$

$\mathbf{X}_{\pi_g,h}$ $\mathbf{y}_{g,h}$ $\mathbf{w}_{g,h}$ $\mathbf{C}_{g,h}$

π_g τ_g $\mathbf{m}_g$

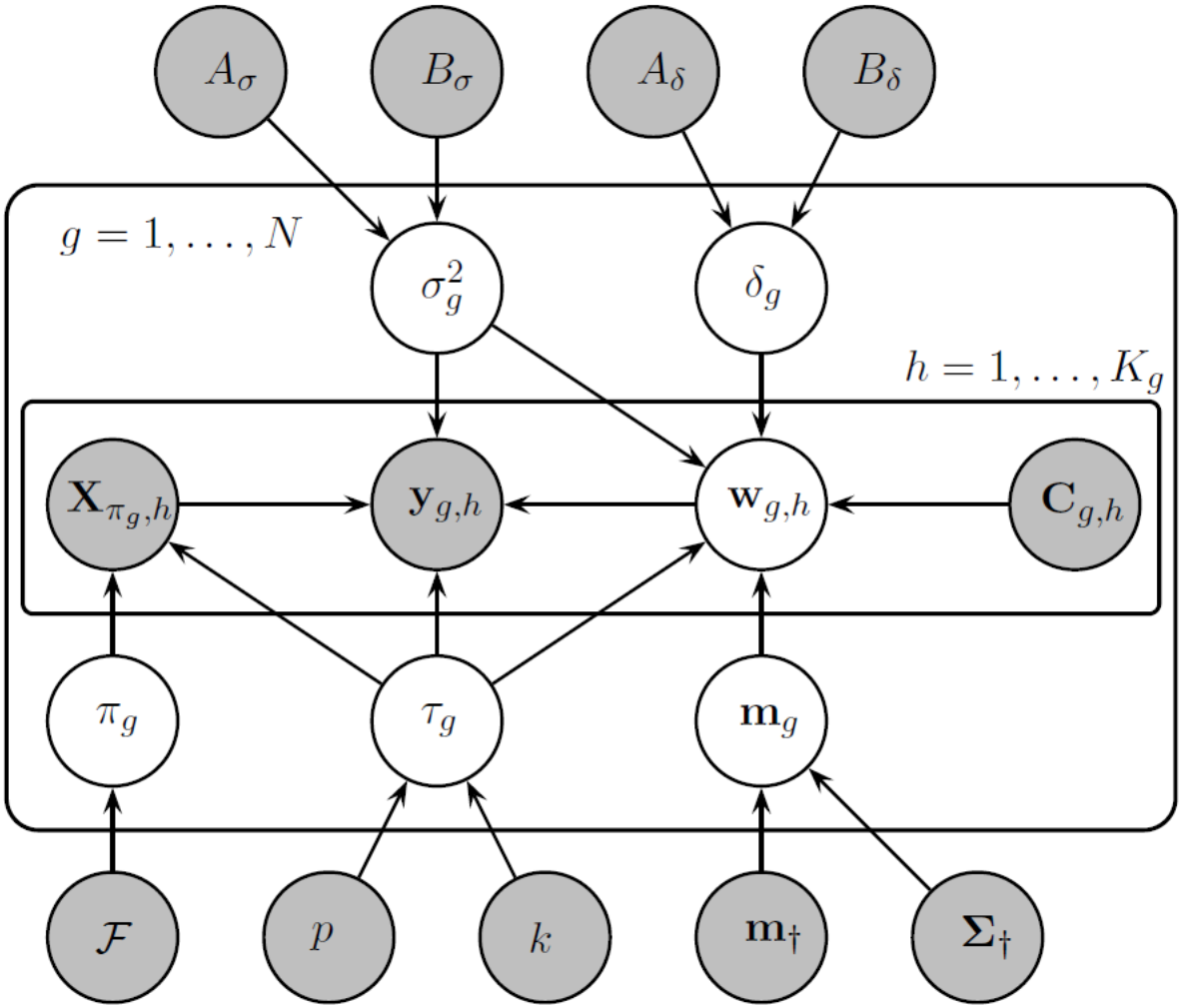
$\mathcal{F}$ p k

$\pi_g \sim Uni$
$|\pi_g| \leq \mathcal{F}$

$\tau_g = \{\tau_{g,1}, \ldots, \tau_{g,K_g-1}\}$
$T_{g,h} := \tau_{g,h} - \tau_{g,h-1}$
$T_{g,h} \sim NBIN(p, k)$

With a fixed hyerparameter $\mathbf{m}_g$ there is **no** information coupling between the segment-specific regression coefficients.

Graphical model representation



$\pi_g \sim Uni$
$|\pi_g| \leq \mathcal{F}$

$\tau_g = \{\tau_{g,1}, \ldots, \tau_{g,K_g-1}\}$
$T_{g,h} := \tau_{g,h} - \tau_{g,h-1}$
$T_{g,h} \sim NBIN(p, k)$

$\mathbf{m}_g \sim \mathcal{N}(\mathbf{m}_\dagger, \mathbf{\Sigma}_\dagger)$

For $g = 1, \ldots, N$**:**

$\sigma_g^{-2} \sim Gam(A_\sigma, B_\sigma)$

$\delta_g^{-1} \sim Gam(A_\delta, B_\delta)$

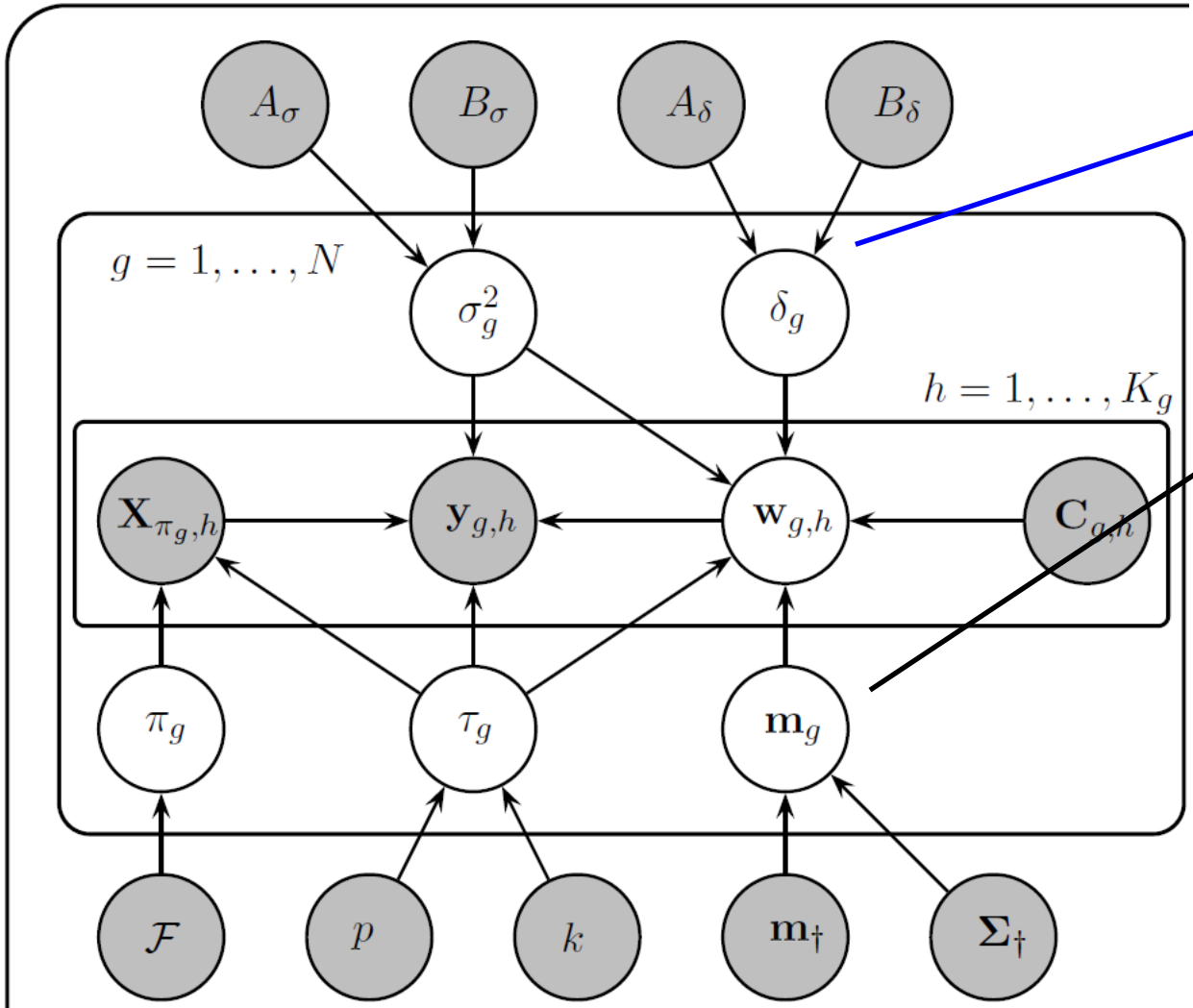
For $h = 1, \ldots, K_g$**:**

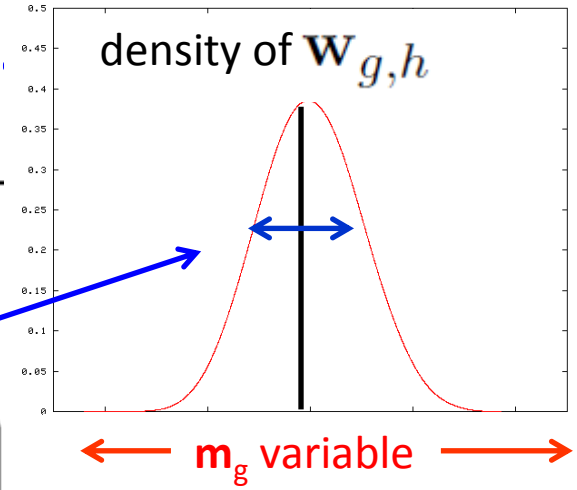
$\mathbf{w}_{g,h} \sim \mathcal{N}(\mathbf{m}_g, \sigma_g^2 \delta_g \mathbf{C}_{g,h})$

$\mathbf{y}_{g,h} \sim \mathcal{N}(\mathbf{X}_{\pi_g,h}^T \mathbf{w}_{g,h}, \sigma_g^2 \mathbf{I})$

Main idea from:
Grzegorczyk and Husmeier (2012b)
Bayesian regularization of non-homogeneous dynamic Bayesian networks by **globally** coupling interaction parameters.
AISTATS

Graphical model represen





so that
the segment-specific
regression coefficients
are coupled

→ information exchange
among segments

Main idea from:
Grzegorczyk and Husmeier (2012b)
Bayesian regularization of non-homogeneous dynamic Bayesian networks by **globally** coupling interaction parameters.
AISTATS

RJMCMC inference Part 1 of 3

1. Noise variances:

$$\sigma_g^{-2} | (\mathbf{y}_{g,.}, \mathbf{X}_{\pi_g,.}, \delta_g)$$

2. Regression coefficients:

$$P(\mathbf{w}_{g,h} | \mathbf{y}_{g,h}, \mathbf{X}_{\pi_g,h}, \sigma_g)$$

3. Coupling hyperparameters:

$$P(\delta_g^{-1} | \mathbf{y}_{g,.}, \mathbf{w}_{g,.}, \boldsymbol{\sigma}_{g,.}^2, \mathbf{X}_{\pi_g,.})$$

can be sampled with standard collapsed and uncollapsed Gibbs sampling steps

That is, sample each variable from the **conditional distribution**, conditional on its **Markov blanket.**

Conjugate prior distributions: sampling from standard distributions

Collapsing: integrate some variables in the Markov blanket out analytically

RJMCMC inference Part 2 of 3

4. Network inference by a Metropolis Hastings sampling scheme, which changes the network by **adding** and **removing** individual edges:

$$P(\mathcal{M}|\mathcal{D},\{\boldsymbol{\tau}_g\},\boldsymbol{\delta}) \propto \underbrace{P(\mathcal{M})}_{\text{network prior}} \prod_g \prod_h \underbrace{P(\mathbf{y}_{g,h}|\mathbf{X}_{\pi_g,h},\delta_g)}_{\text{marginal likelihoods can be computed in closed form:}}$$

5. Changepoint inference by a Metropolis Hastings sampling scheme, which changes the segmentation by **adding** and **removing** gene-specific **changepoints**:

$$P(\{\boldsymbol{\tau}_g\}|\mathcal{D},\boldsymbol{\delta},\mathcal{M}) \propto \prod_g \underbrace{P(\boldsymbol{\tau}_g)}_{\text{changepoint prior}} \prod_h \underbrace{P(\mathbf{y}_{g,h}|\mathbf{X}_{\pi_g,h},\delta_g)}_{\text{marginal likelihoods can be computed in closed form:}}$$

RJMCMC inference Part 3 of 3

6. The global mean vector $\mathbf{m}_g$ can be sampled with a collapsed Gibbs sampling steps:

$$\mathbf{m}_g | (\mathbf{w}_{g,1}, \ldots, \mathbf{w}_{g,K_g}) \sim \mathcal{N}(\mathbf{m}_{\star,g}, \boldsymbol{\Sigma}_{\star,g})$$

with the sufficient statistics:

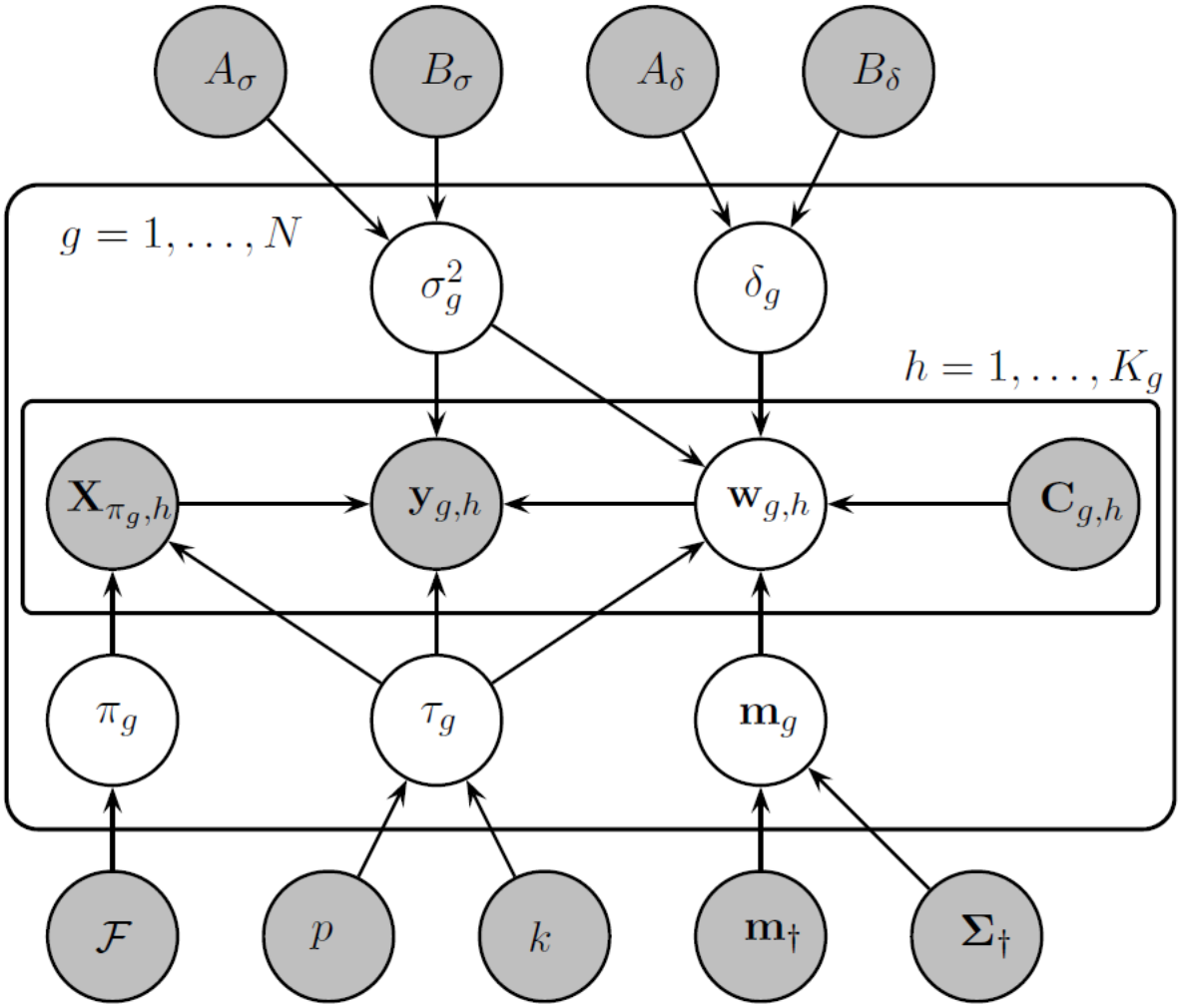
$$\boldsymbol{\Sigma}_{\star,g} := (\boldsymbol{\Sigma}_{\dagger}^{-1} + K_g \boldsymbol{\Sigma}_0^{-1})^{-1}$$

$$\mathbf{m}_{\star,g} := \boldsymbol{\Sigma}_{\star,g}(\boldsymbol{\Sigma}_{\dagger}^{-1}\mathbf{m}_{\dagger} + \boldsymbol{\Sigma}_0^{-1}[\sum_{h=1}^{K_g} \mathbf{w}_{g,h}])$$

Overall sampling scheme:

"Metropolis-Hastings-RJMCMC scheme within a partially collapsed Gibbs sampler"

Empirical comparison: (1) globally coupled NH-DBN



For $g = 1, \ldots, N$**:**

$\sigma_g^{-2} \sim Gam(A_\sigma, B_\sigma)$

$\delta_g^{-1} \sim Gam(A_\delta, B_\delta)$

For $h = 1, \ldots, K_g$**:**

$\mathbf{w}_{g,h} \sim \mathcal{N}(\mathbf{m}_g, \sigma_g^2 \delta_g \mathbf{C}_{g,h})$

$\mathbf{y}_{g,h} \sim \mathcal{N}(\mathbf{X}_{\pi_g,h}^T \mathbf{w}_{g,h}, \sigma_g^2 \mathbf{I})$

$\pi_g \sim Uni$

$|\pi_g| \leq \mathcal{F}$

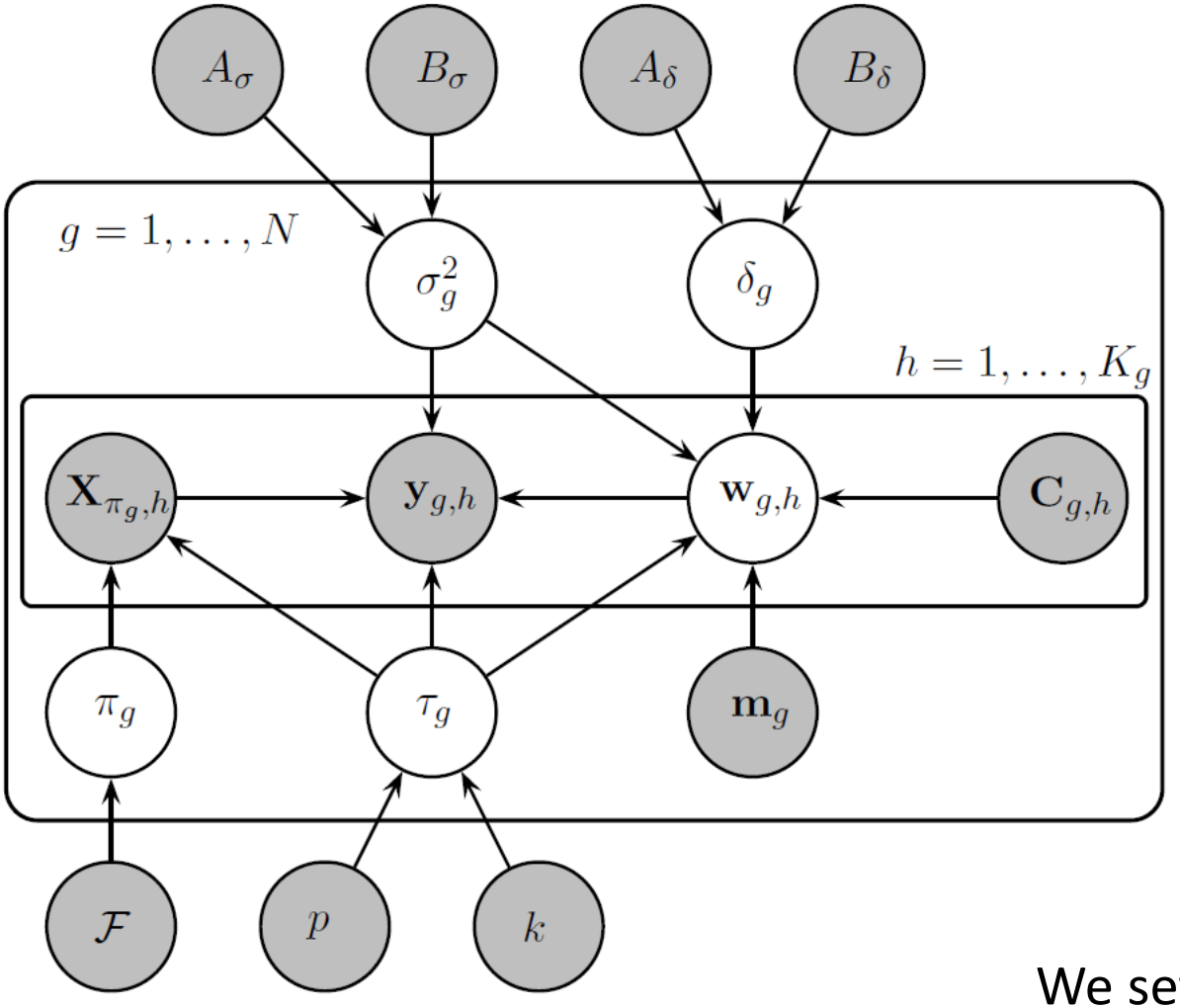
$\tau_g = \{\tau_{g,1}, \ldots, \tau_{g,K_g-1}\}$

$T_{g,h} := \tau_{g,h} - \tau_{g,h-1}$

$T_{g,h} \sim NBIN(p, k)$

$\mathbf{m}_g \sim \mathcal{N}(\mathbf{m}_\dagger, \mathbf{\Sigma}_\dagger)$

Empirical comparison: (2) uncoupled NH-DBN



$\pi_g \sim Uni$
$|\pi_g| \leq \mathcal{F}$

$\tau_g = \{\tau_{g,1}, \ldots, \tau_{g,K_g-1}\}$
$T_{g,h} := \tau_{g,h} - \tau_{g,h-1}$
$T_{g,h} \sim NBIN(p, k)$

For $g = 1, \ldots, N$**:**

$\sigma_g^{-2} \sim Gam(A_\sigma, B_\sigma)$

$\delta_g^{-1} \sim Gam(A_\delta, B_\delta)$

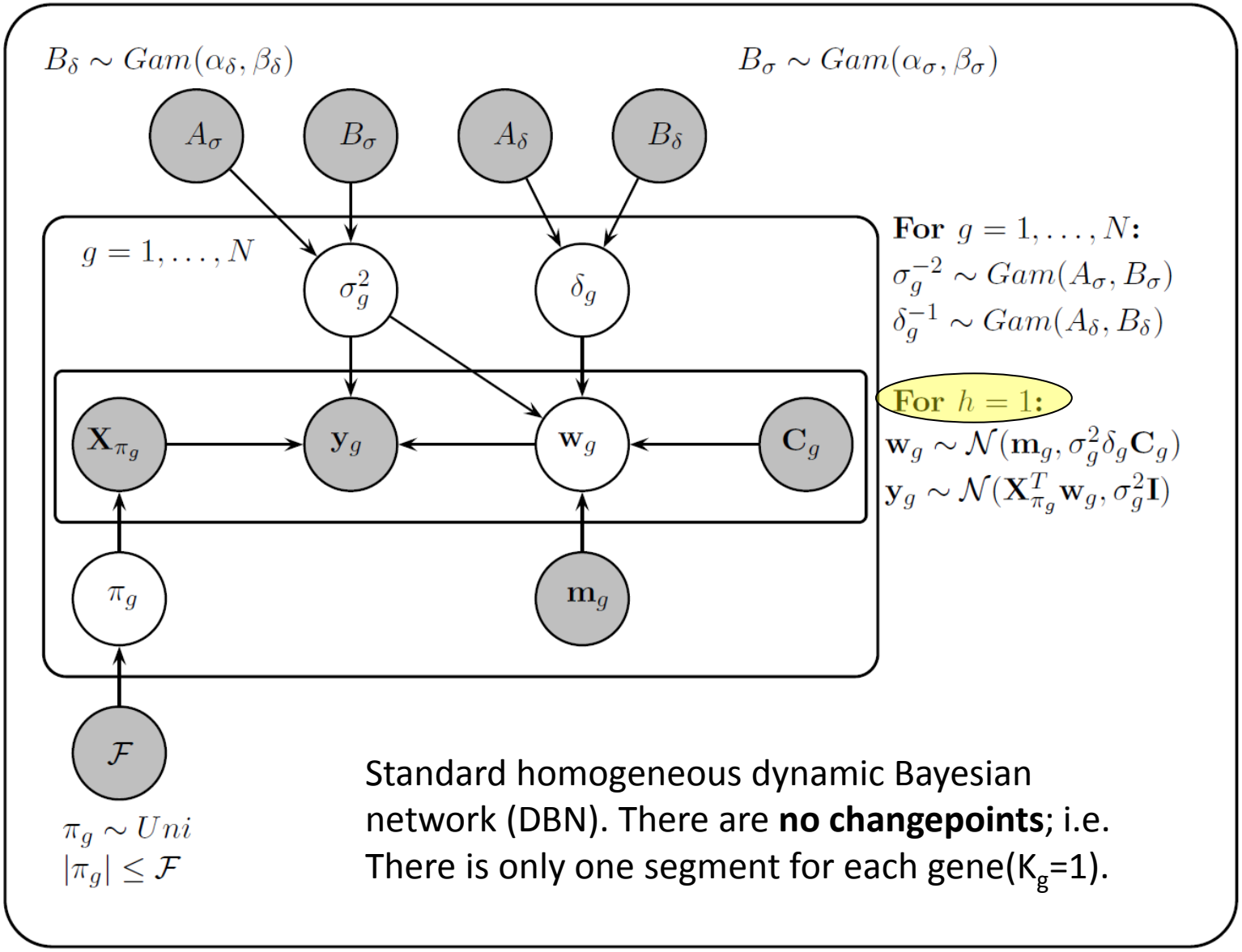
For $h = 1, \ldots, K_g$**:**

$\mathbf{w}_{g,h} \sim \mathcal{N}(\mathbf{m}_g, \sigma_g^2 \delta_g \mathbf{C}_{g,h})$

$\mathbf{y}_{g,h} \sim \mathcal{N}(\mathbf{X}_{\pi_g,h}^T \mathbf{w}_{g,h}, \sigma_g^2 \mathbf{I})$

We set: $\mathbf{m_g}$**=0**.

Empirical comparison: (3) Homogeneous DBN



Standard homogeneous dynamic Bayesian network (DBN). There are **no changepoints**; i.e. There is only one segment for each gene(K_g=1).

Empirical comparison: (4) Sequentially coupled NH-DBN

$$P(\mathbf{w}_{g,h}|\mathbf{m}_{g,h-1},\sigma^2_{g,h},\delta_g,\lambda_g) = \begin{cases} \mathcal{N}(\mathbf{w}_{g,1}|\mathbf{m}_{g,0}=\mathbf{0},\delta_g\sigma^2_{g,h}\mathbf{C}_{g,h}), & h=1 \\ \mathcal{N}(\mathbf{w}_{g,h}|\mathbf{m}_{g,h-1},\lambda_g\sigma^2_{g,h}\mathbf{C}_{g,h}), & h\geq 2 \end{cases} \quad (1)$$

where $\mathbf{m}_{g,h-1}$ $(h \geq 2)$ depends on the preceding segment:

$$\mathbf{m}_{g,h} = \mathbf{\Sigma}_{g,h}([\lambda_g\mathbf{C}_{g,h}]^{-1}\mathbf{m}_{g,(h-1)} + \mathbf{X}_{\pi_g,h}\mathbf{y}_{g,h}) \quad (2)$$

For h≥2:

The prior expectation of the regression coefficients for segment h+1, $\mathbf{m}_{g,h}$, depends on the posterior distribution of the regression coefficients $\mathbf{w}_{g,h}$ for segment h.

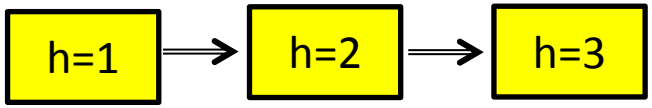
The coupling strength depends on the hyperparameter λ_g.

Main idea from: Grzegorczyk and Husmeier (2012a)
A non-homogeneous dynamic Bayesian network model with **sequentially** coupled interaction parameters for applications in systems and synthetic biology.
SAGMB

Information coupling

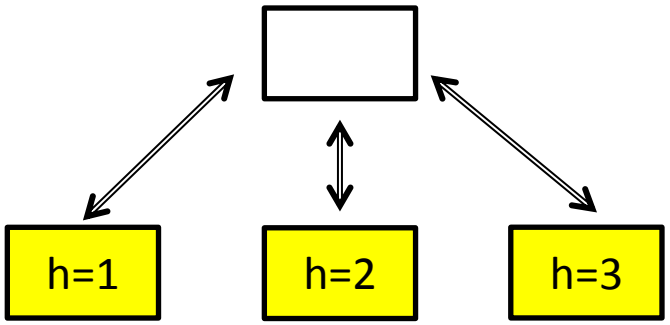
Sequential coupling

- Information is shared between neighbouring segments
- **For example:** morphogenesis



Global coupling

- Segments are treated as interchangeable and information is shared globally
- **For example:** different experimental scenarios or environmental conditions



Empirical evaluation

1. Simulated data

2. Data from synthetic biology

3. Data from a real application

Empirical evaluation

1. Simulated data

Known gold standard 

Simulation process does not reflect real biology 

2. Data from synthetic biology

Known gold standard 

Real wet lab data 

Regulatory network small 

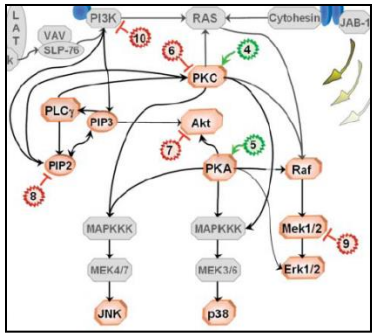
3. Data from a real application

Real wet lab data 

No gold standard 

Reconstruction Accuracy

true network



biological knowledge

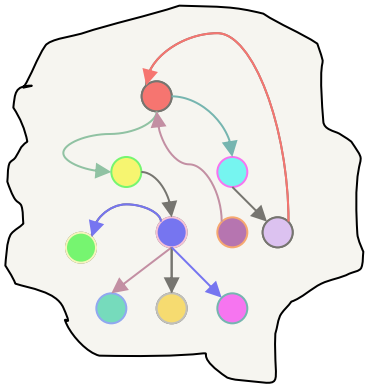
(gold standard network)

Evaluation of

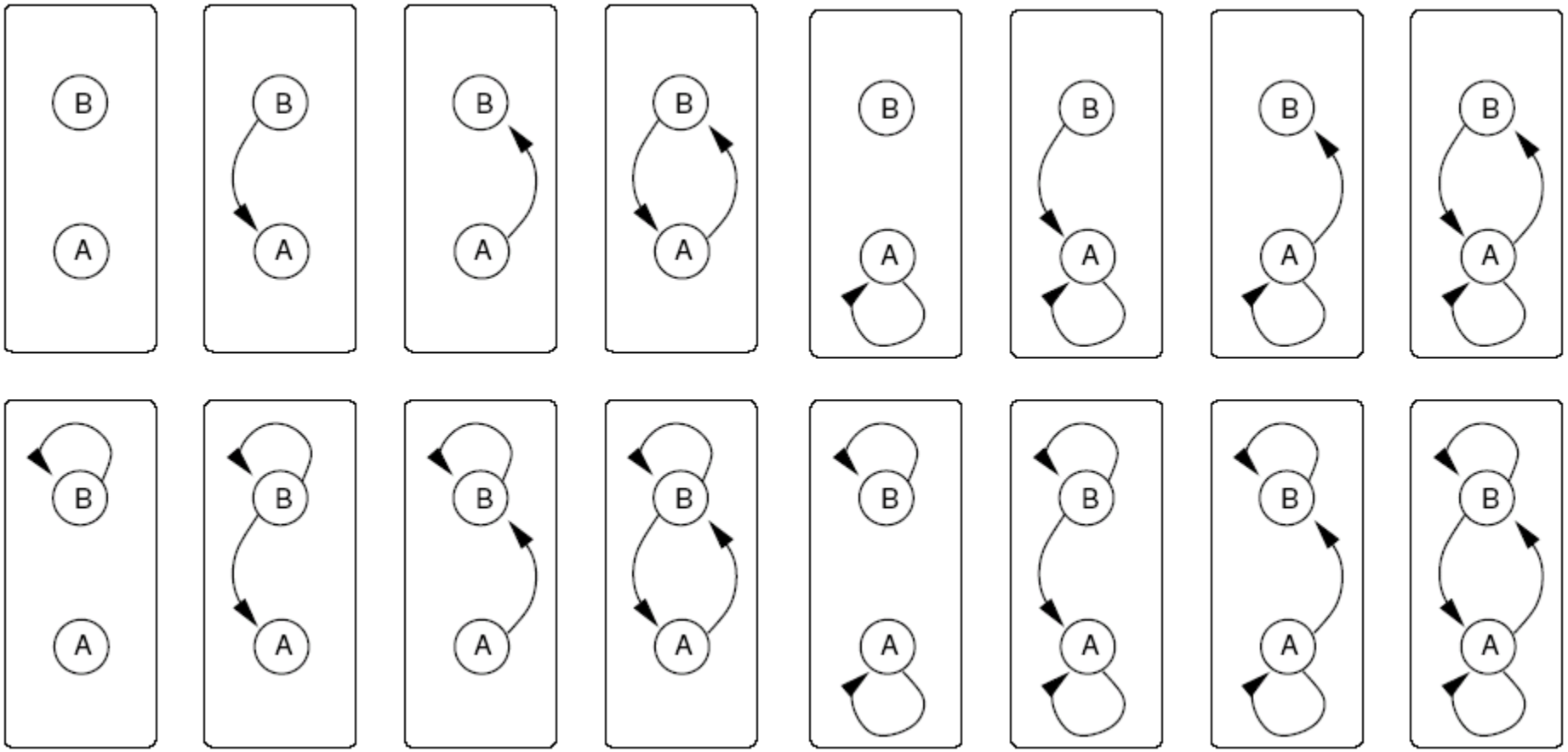
learning

performance

extracted network



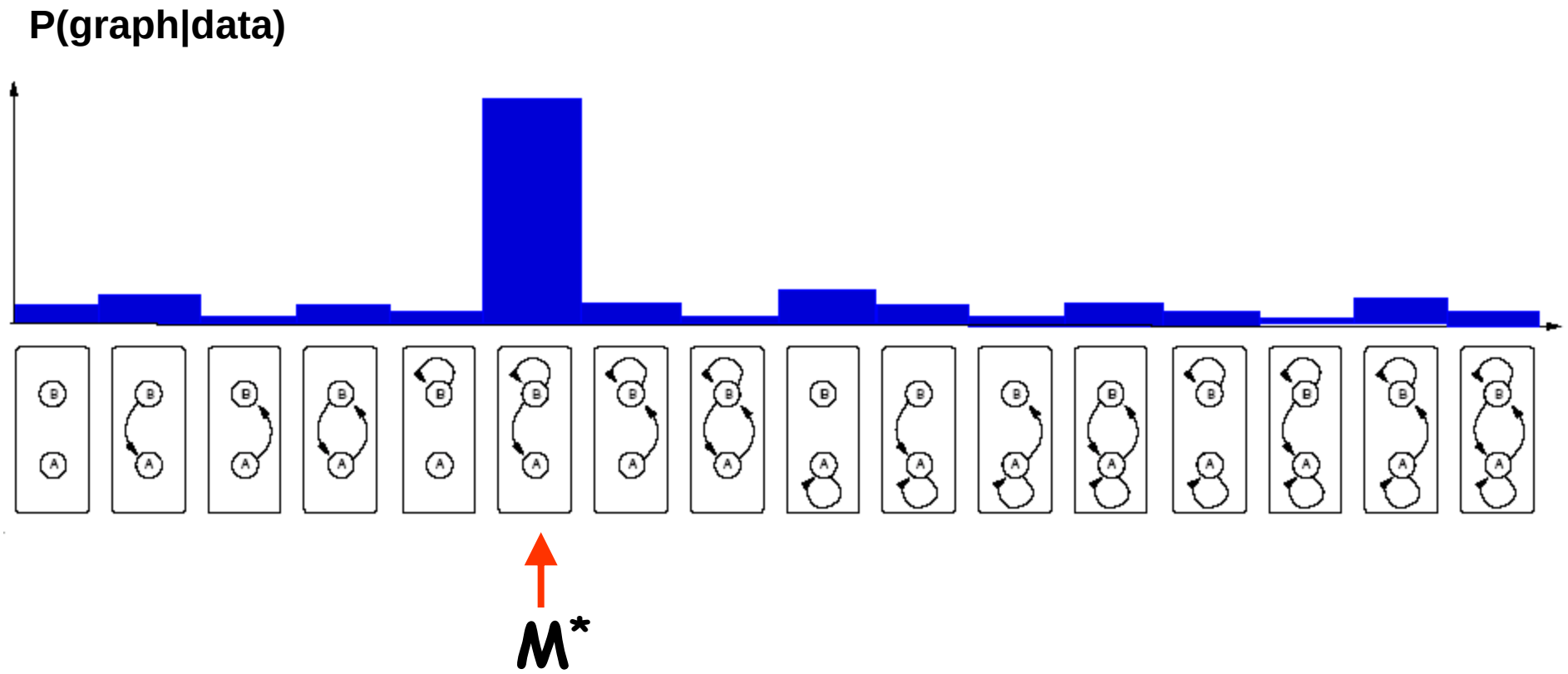
Example: 2 genes → 16 different (dynamic) network structures



Best network: maximum score $P(\mathcal{D}|\mathcal{M})$

Ideal scenario: Large data sets, low noise

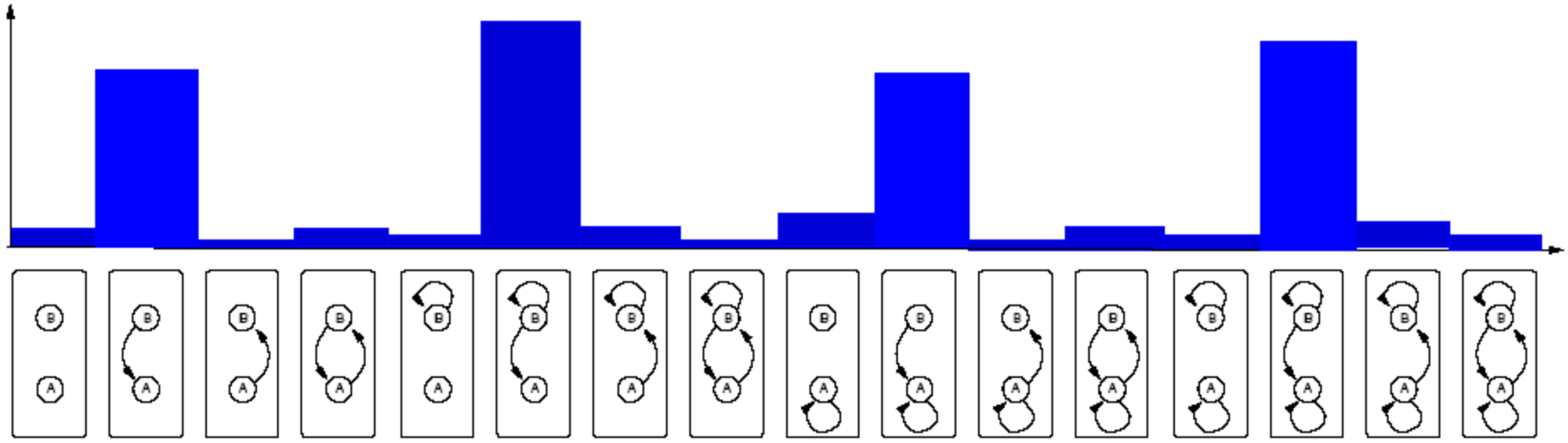
Identify the best network structure



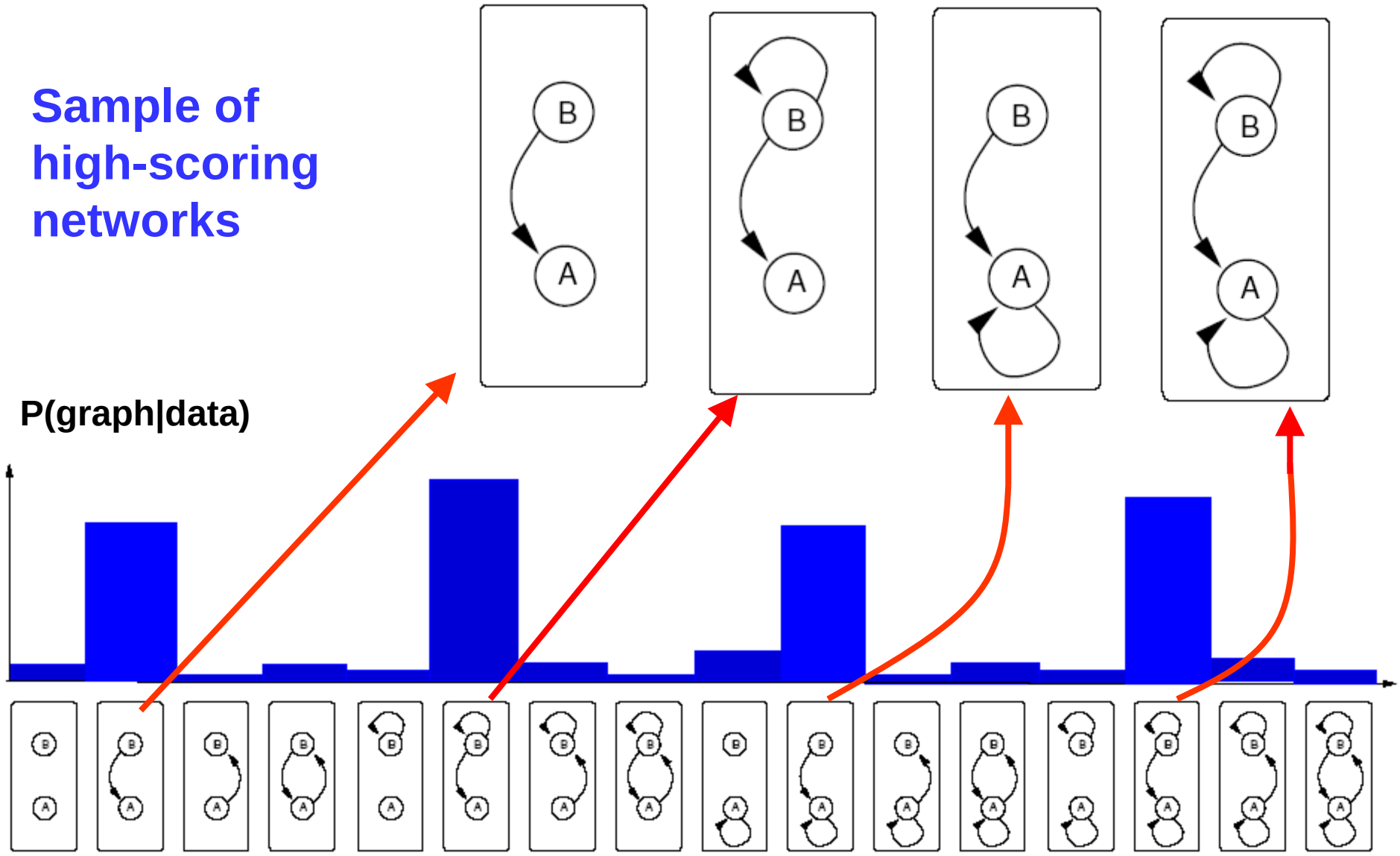
Realistic: Limited number of experimental replications, high noise

Uncertainty about the best network

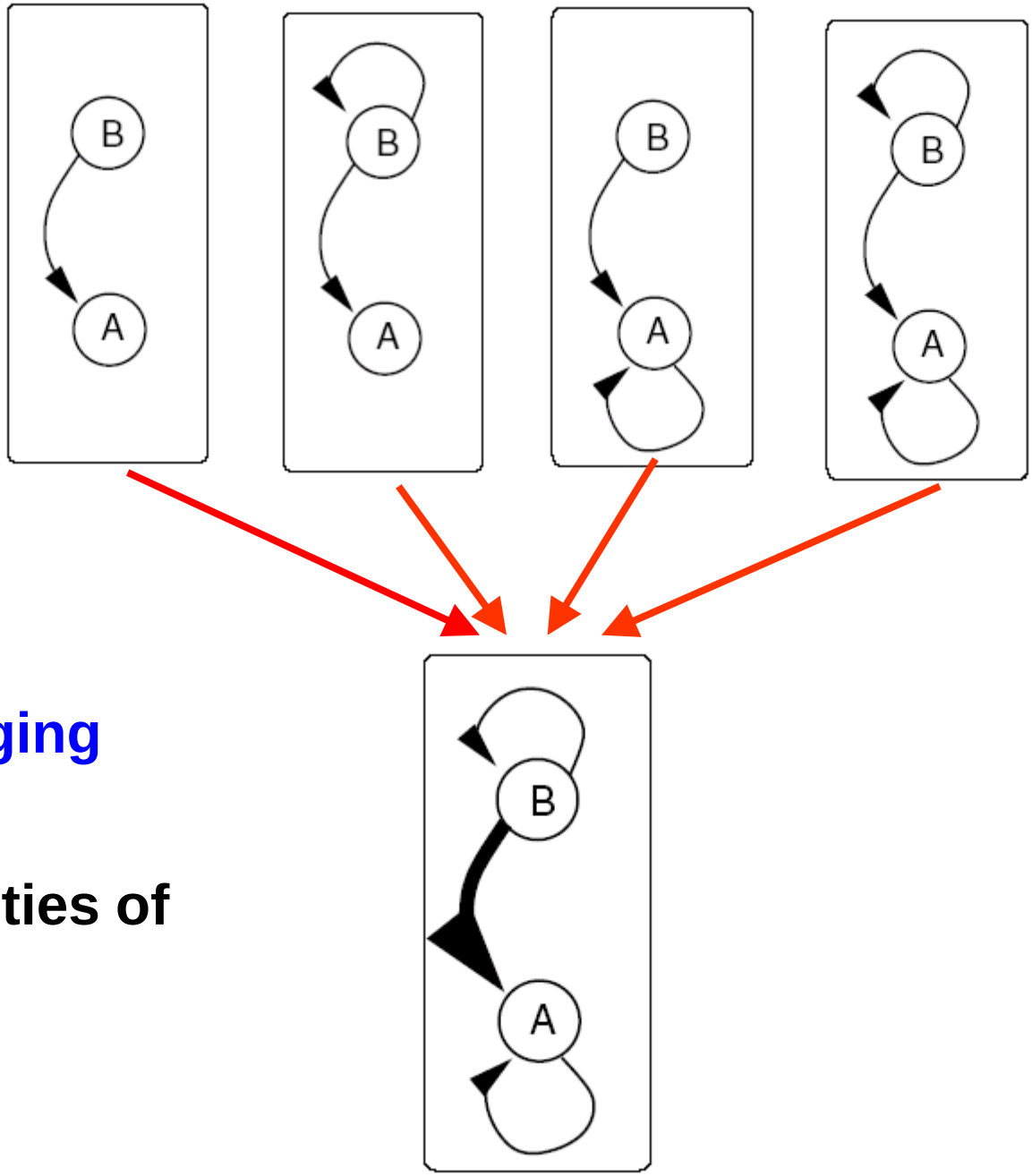
P(graph|data)



Sample of high-scoring networks



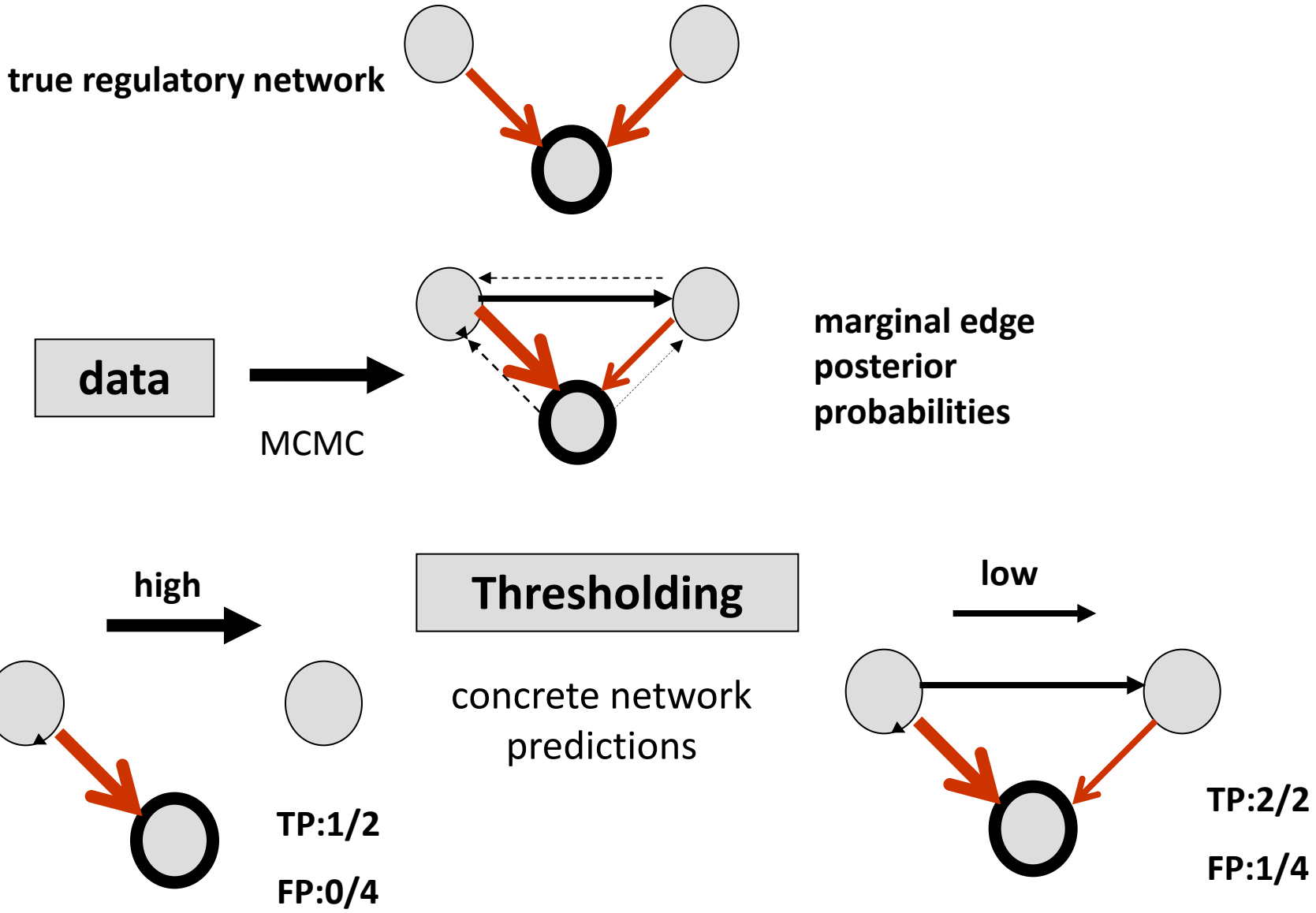
MCMC sample of high-scoring networks

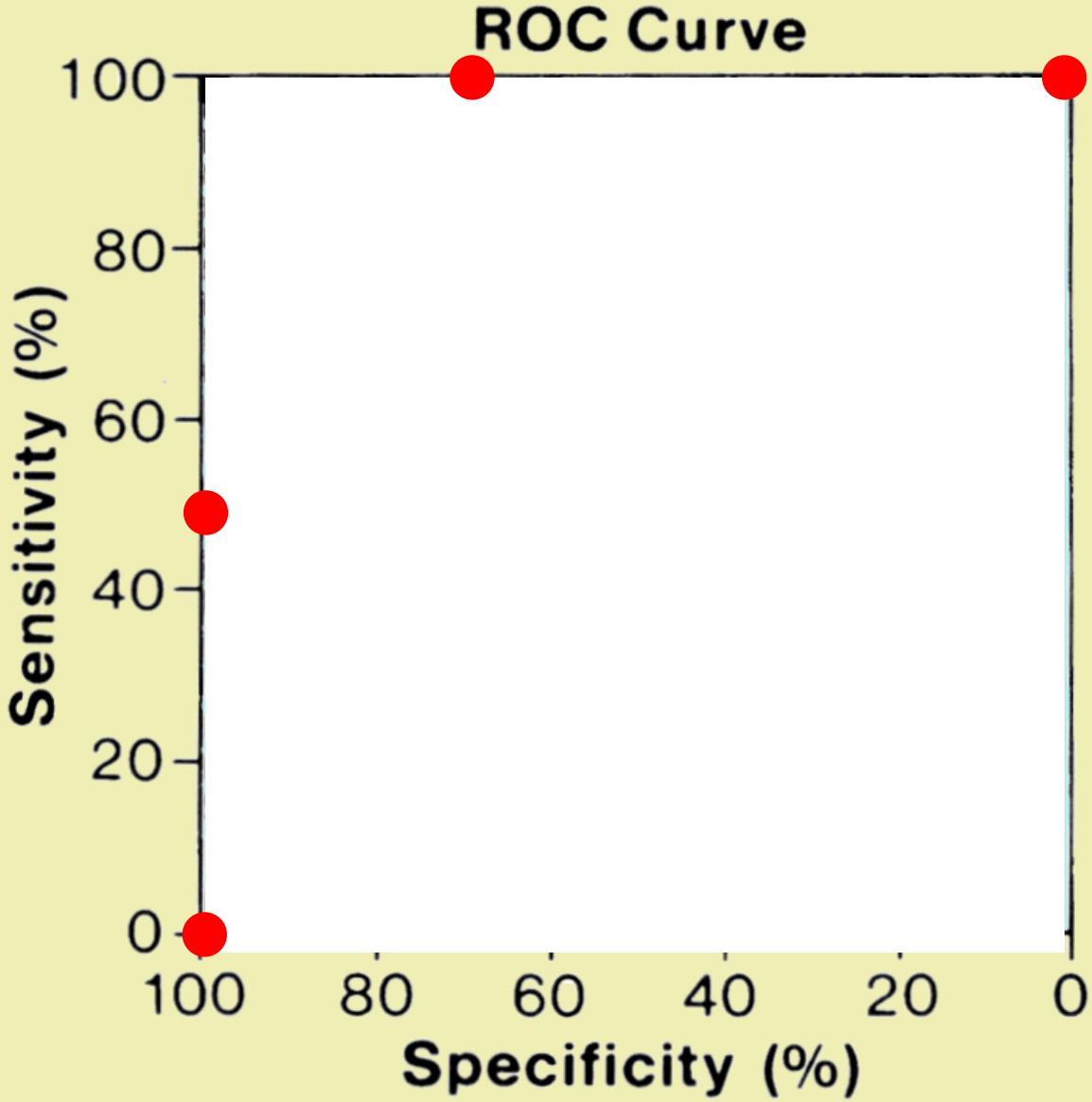


Idea: Model Averaging

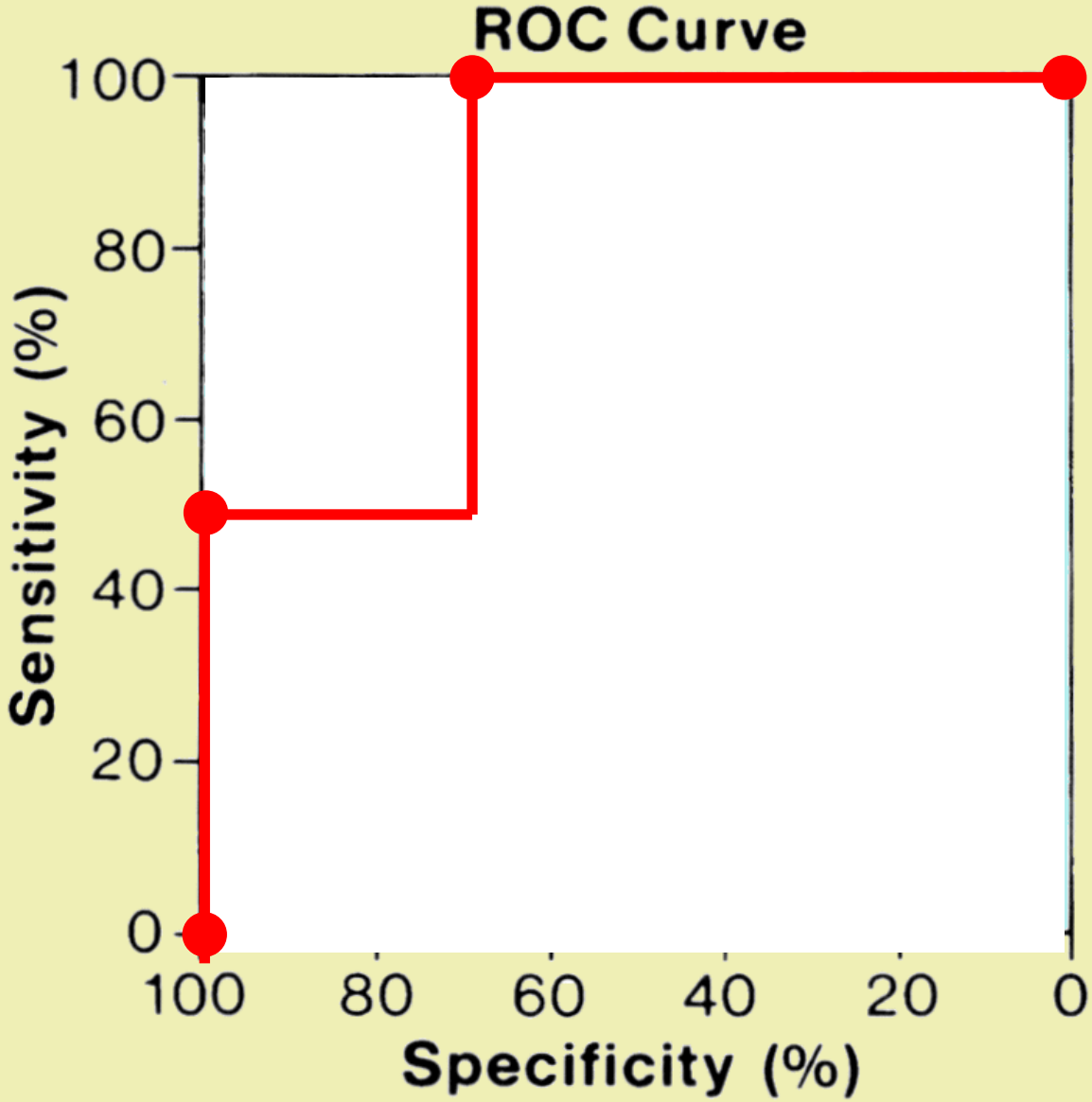
Compute marginal posterior probabilities of the edges

Probabilistic inference

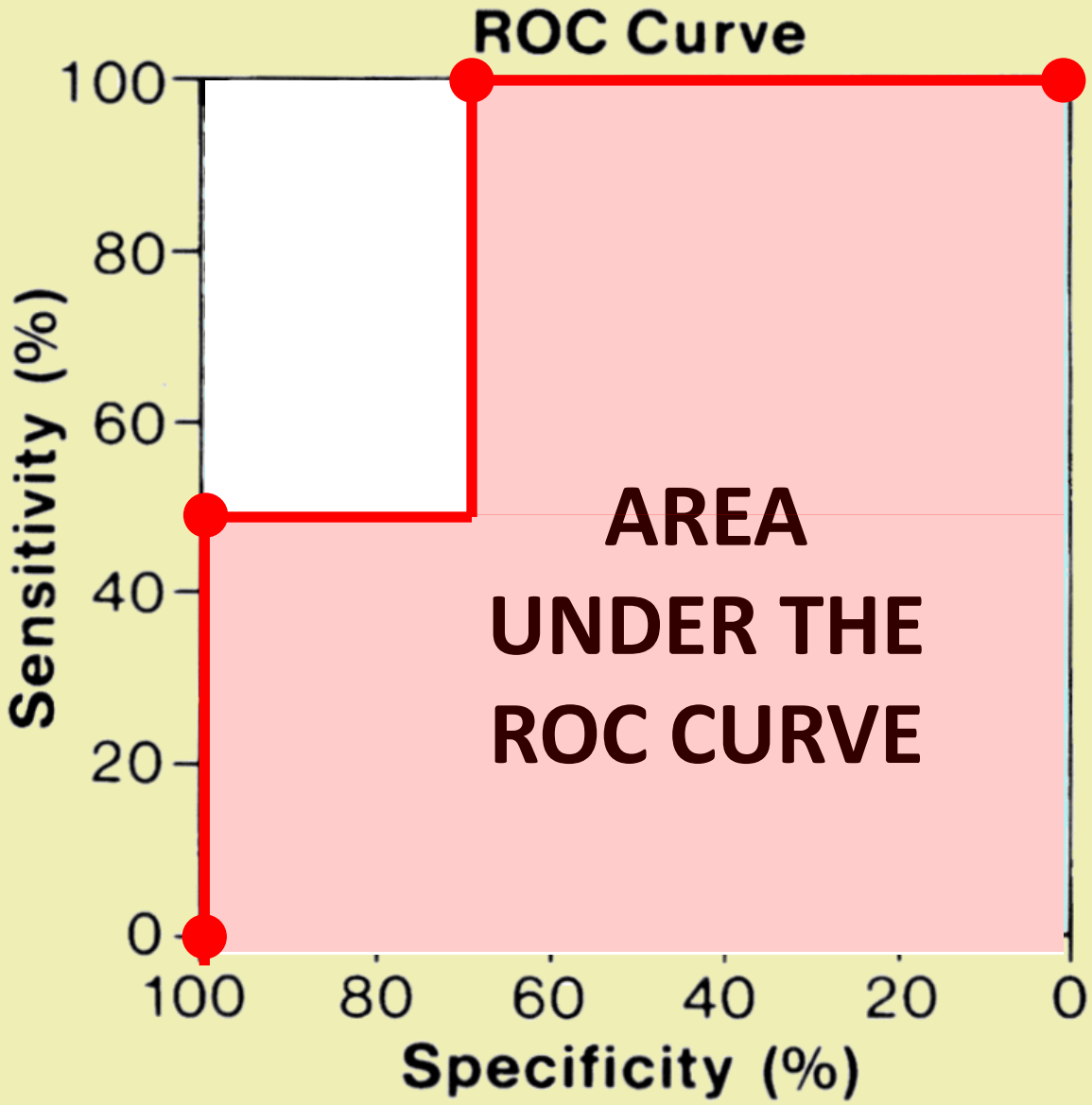




From Perry Sprawls



From Perry Sprawls



From Perry Sprawls

1. Simulated data

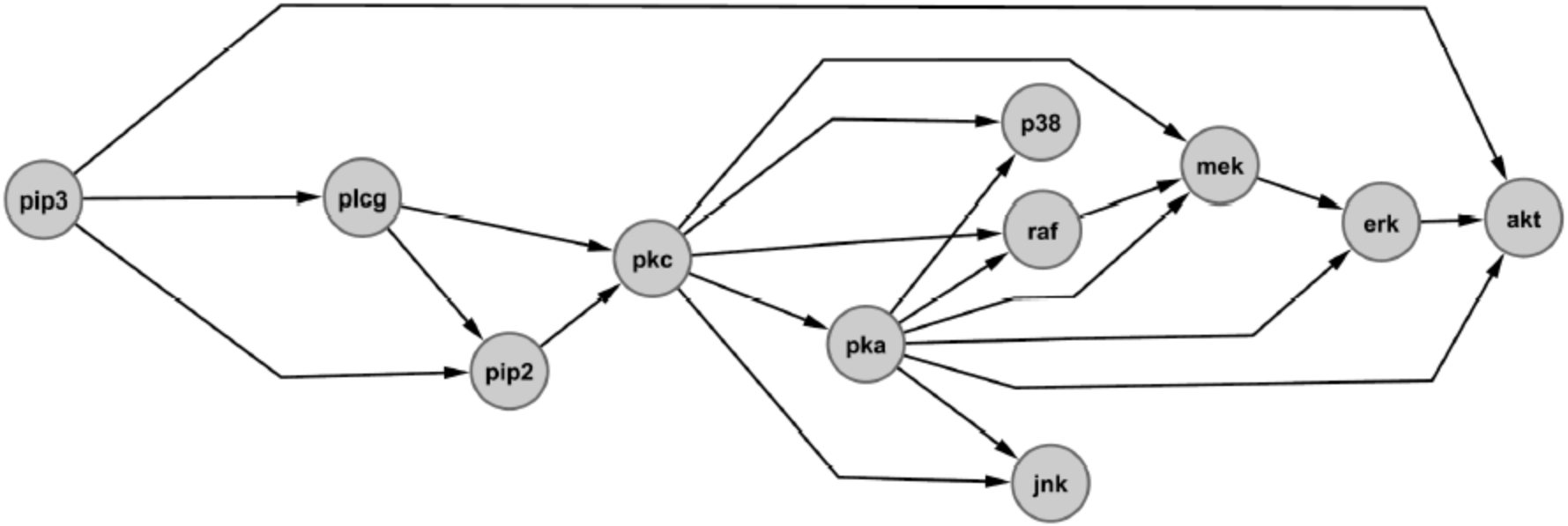


Figure: The RAF protein signalling pathway as reported in Sachs et al. (Science, 2005). The RAF network consists of 11 nodes (proteins) and 20 directed edges.

$$\mathbf{y}_{g,h} \sim \mathcal{N}(\mathbf{X}^T_{\pi_g,h}\mathbf{w}_{g,h}, \sigma_g^2\mathbf{I})$$

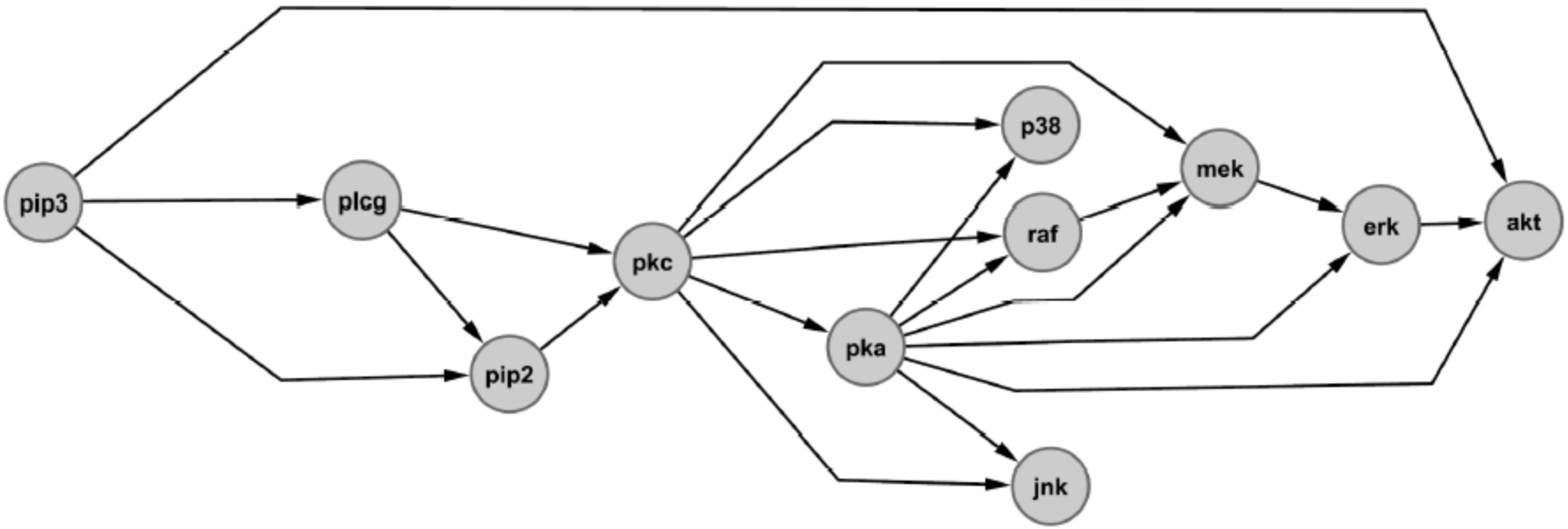


Figure: The RAF protein signalling pathway as reported in Sachs et al. (Science, 2005). The RAF network consists of 11 nodes (proteins) and 20 directed edges.

$$\mathbf{y}_{g,h} \sim \mathcal{N}(\mathbf{X}^T_{\pi_{g,h}} \mathbf{w}_{g,h}, \sigma_g^2 \mathbf{I})$$

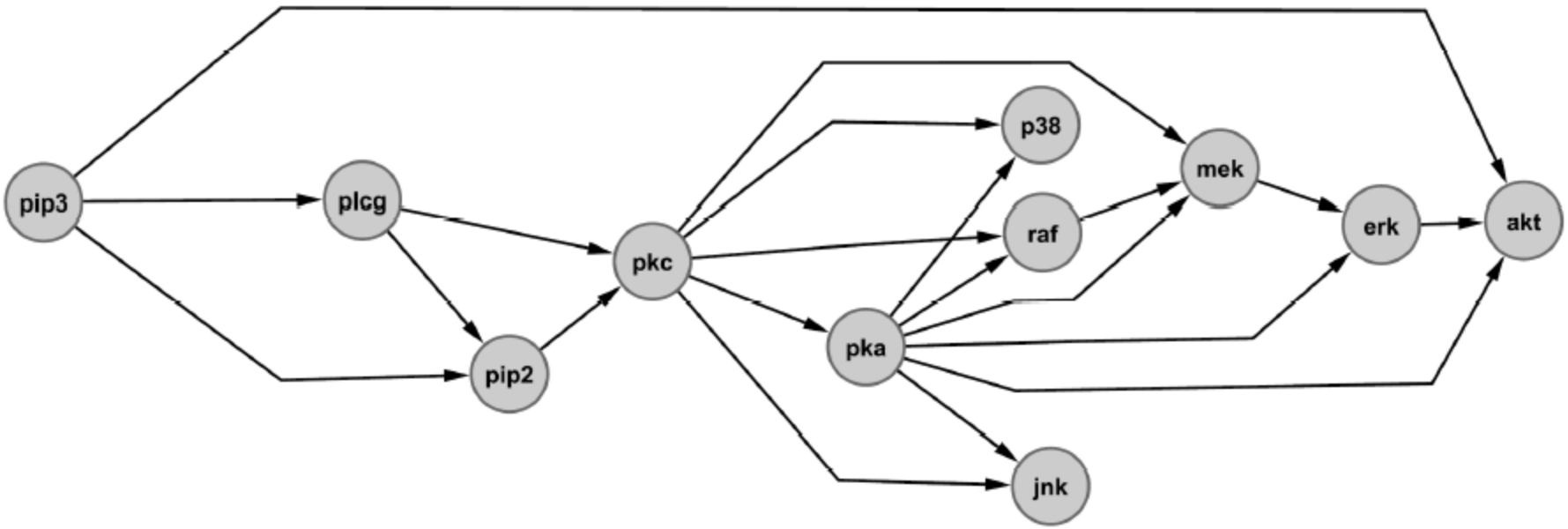


Figure: The RAF protein signalling pathway as reported in Sachs et al. (Science, 2005). The RAF network consists of 11 nodes (proteins) and 20 directed edges.

$$\mathbf{w}_{g,\star} \sim \mathcal{N}(0,1), \quad \tilde{\mathbf{w}}_{g,h} \sim \mathcal{N}(0,1),$$

$$\mathbf{w}_{g,h} = \frac{\frac{\mathbf{w}_{g,\star}}{|\mathbf{w}_{g,\star}|_2} + \varepsilon \frac{\tilde{\mathbf{w}}_{g,h}}{|\mathbf{w}_{g,h}|_2}}{|\frac{\mathbf{w}_{g,\star}}{|\mathbf{w}_{g,\star}|_2} + \varepsilon \frac{\tilde{\mathbf{w}}_{g,h}}{|\mathbf{w}_{g,h}|_2}|_2}$$

$$\mathbf{y}_{g,h} \sim \mathcal{N}(\mathbf{X}^T_{\pi_{g,h}} \mathbf{w}_{g,h}, \sigma^2_g \mathbf{I})$$

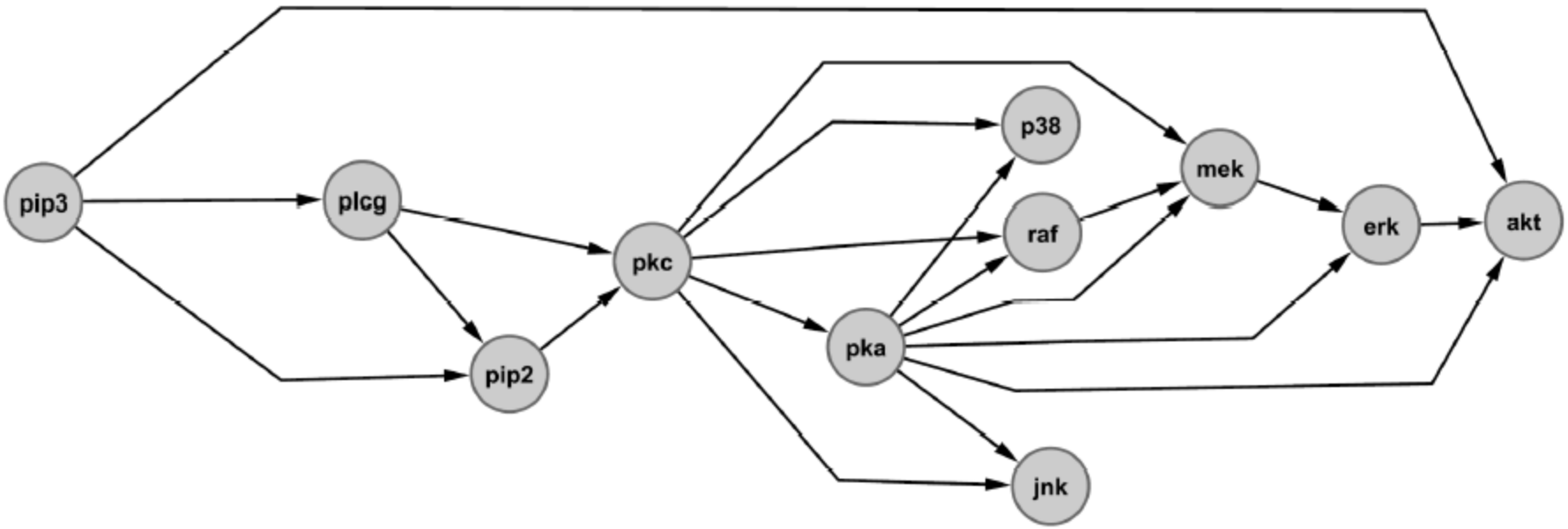


Figure: The RAF protein signalling pathway as reported in Sachs et al. (Science, 2005). The RAF network consists of 11 nodes (proteins) and 20 directed edges.

Generate data sets with **4 segments h=1,…,4** and **10 observations** per segment.

Use **three noise levels** (SNR=10, 3, and 1)

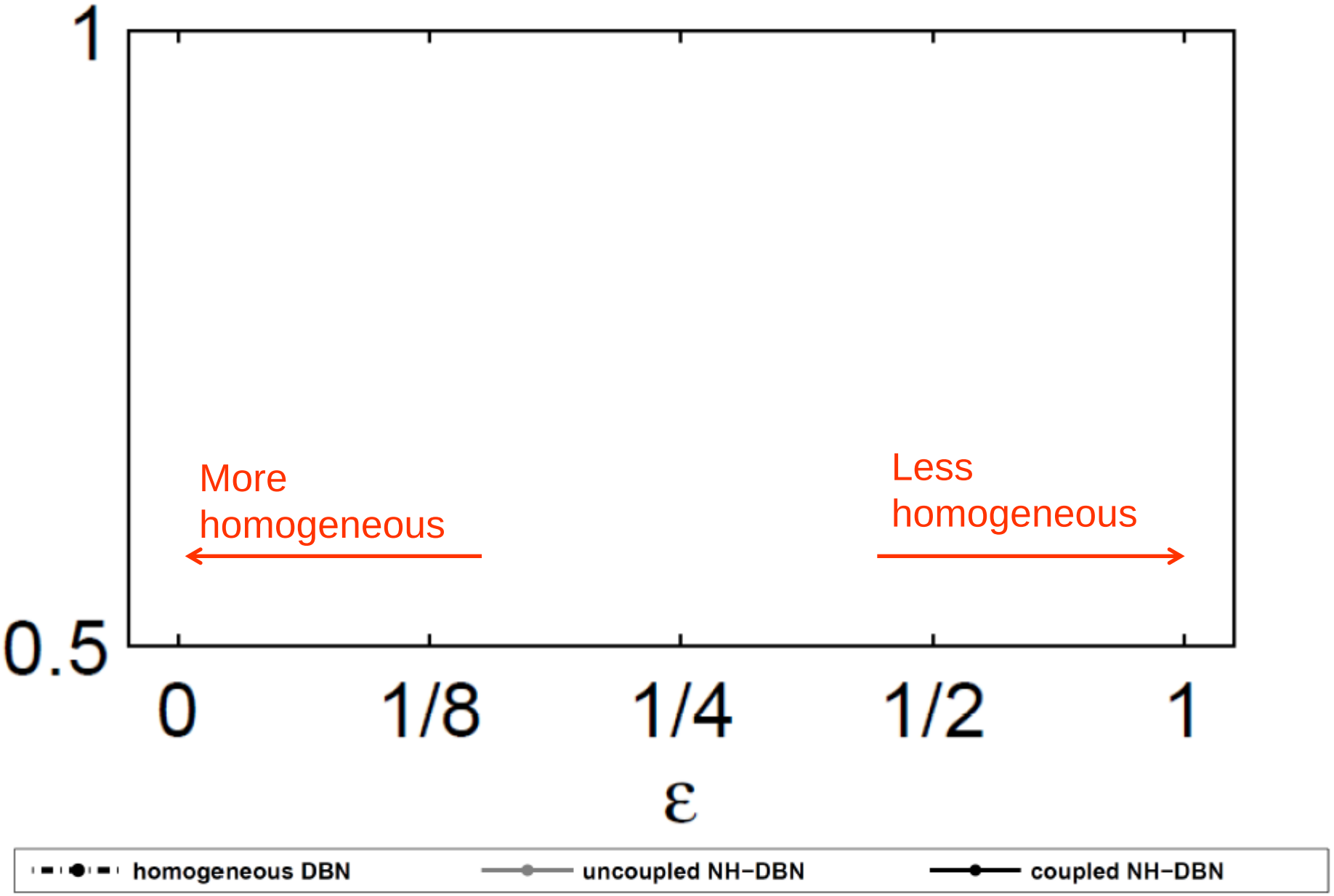
Use the parameter **ε to vary the similarity** of the segment-specific interaction parameters.

ε=0 -> homogeneous data

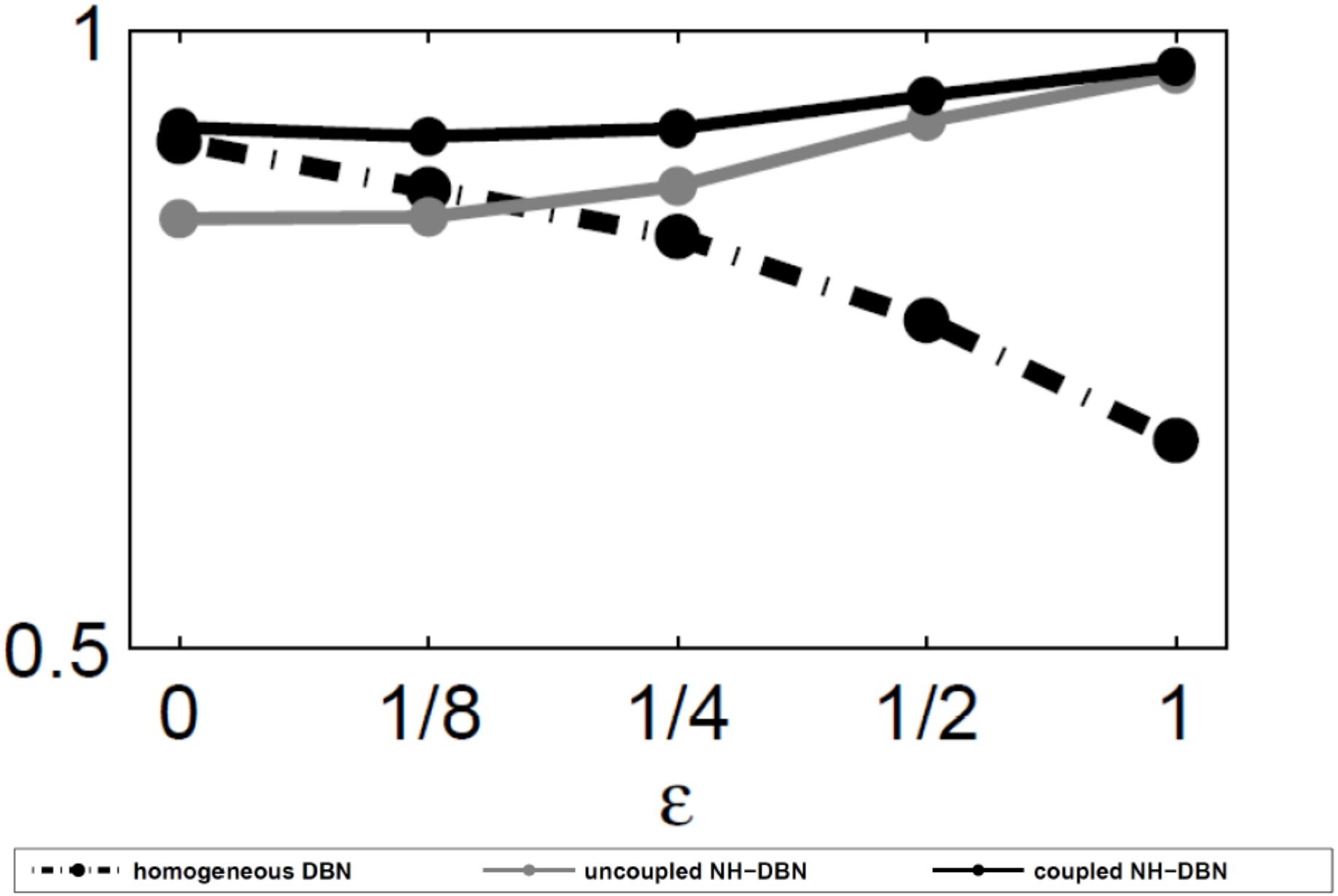
…

ε=1 -> non-homogeneous data

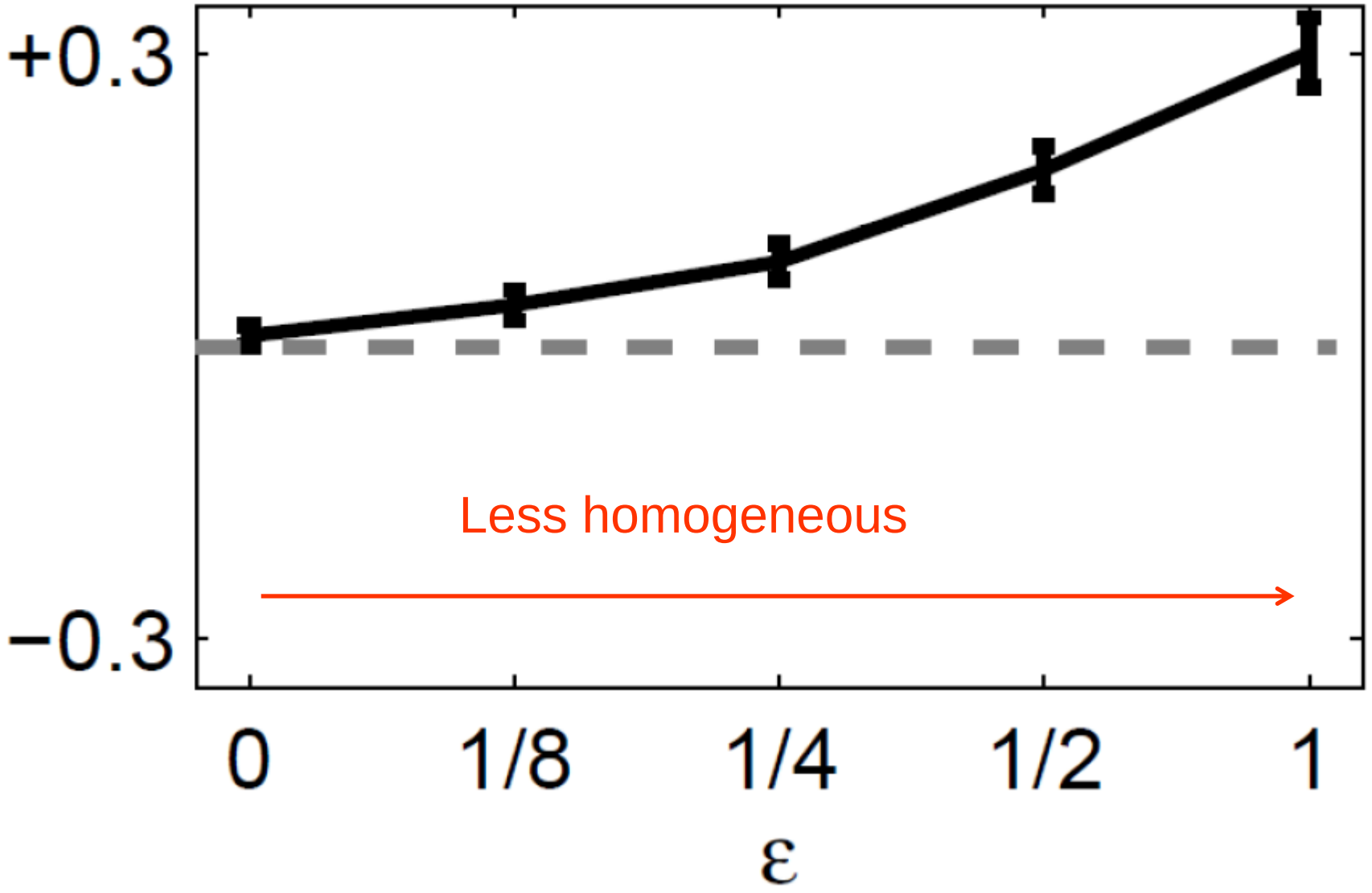
AUC for SNR=3


1
0.5
0
1/8
1/4
1/2
1
ε
More homogeneous
Less homogeneous
homogeneous DBN
uncoupled NH−DBN
coupled NH−DBN

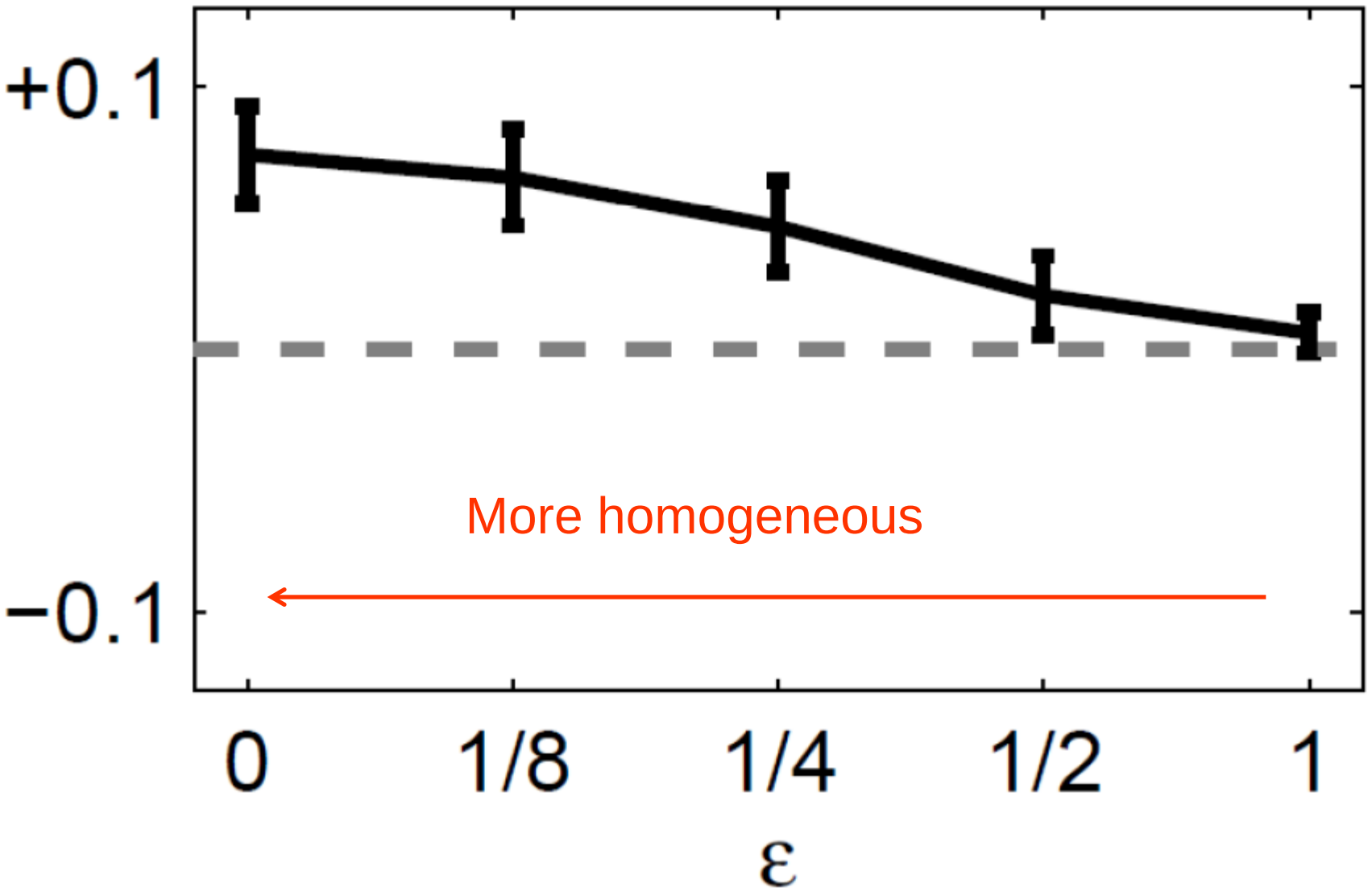
AUC for SNR=3

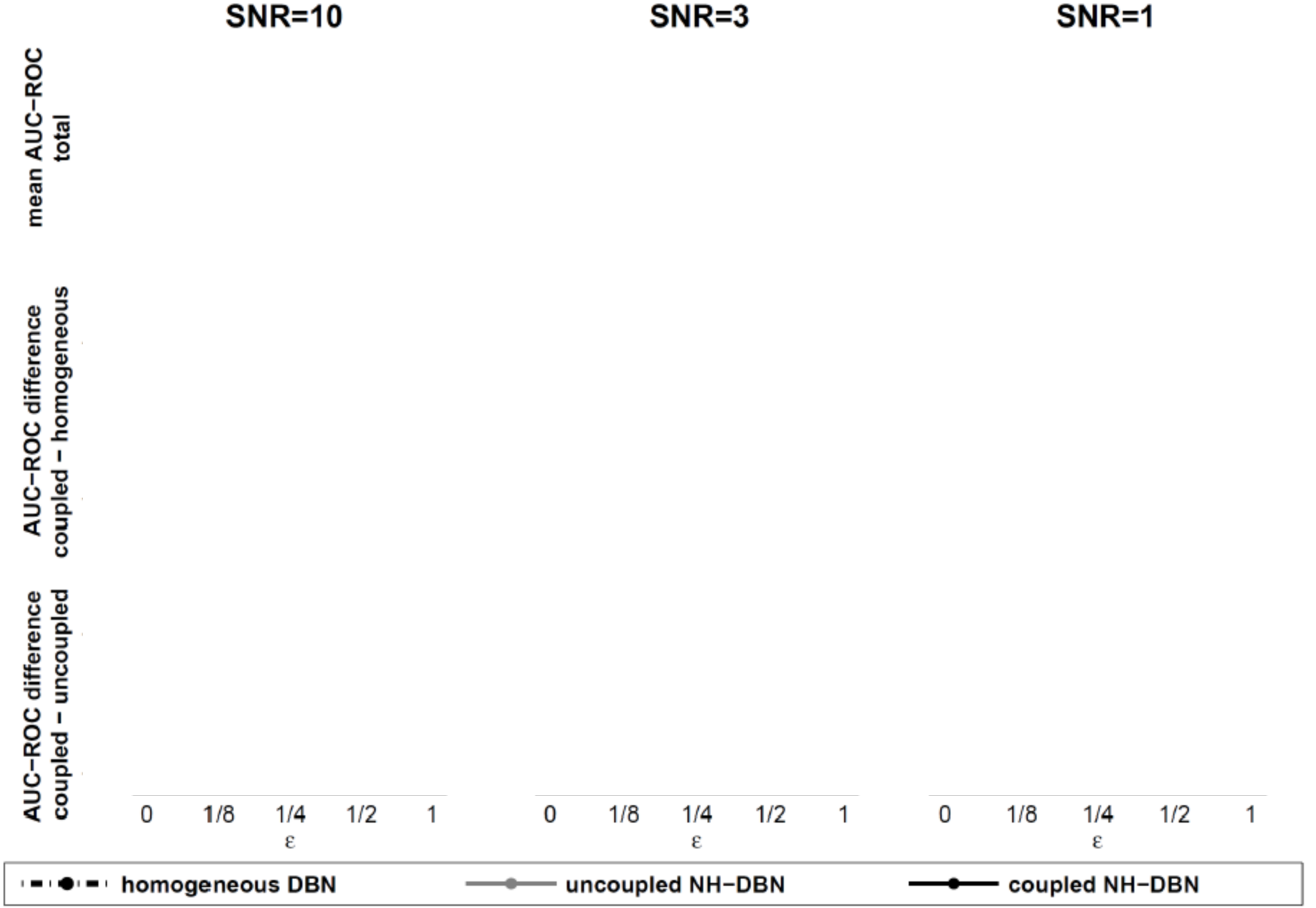

1
0.5
0
1/8
1/4
1/2
1
ε
homogeneous DBN
uncoupled NH−DBN
coupled NH−DBN

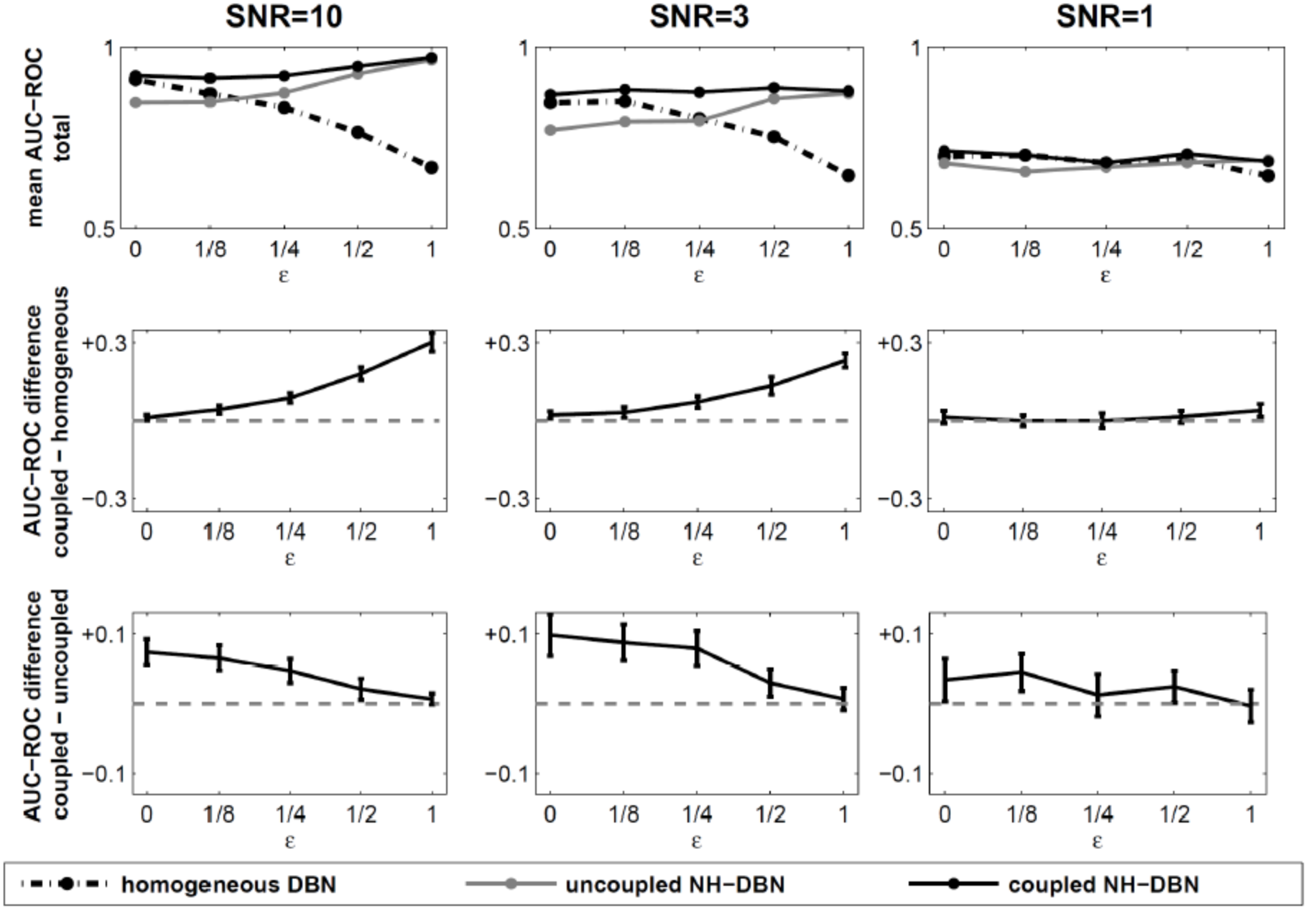
AUC difference: coupled NH-DBN – homogeneous DBN



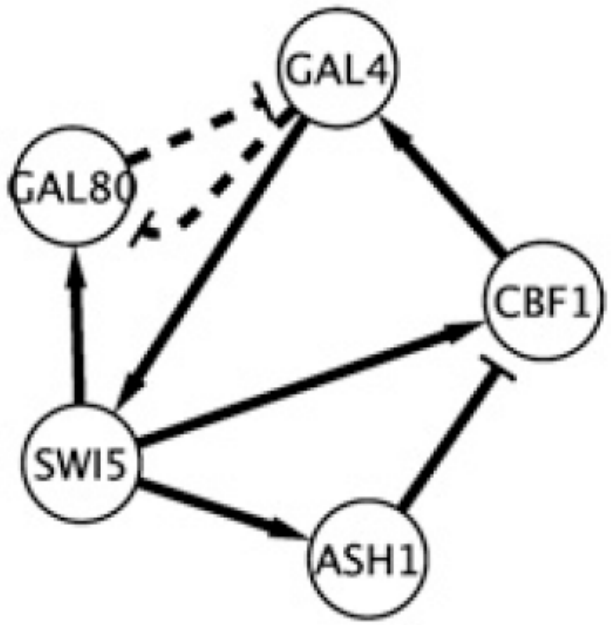
AUC difference: coupled NH-DBN – uncoupled NH-DBN




SNR=10
SNR=3
SNR=1
mean AUC−ROC total
AUC−ROC difference coupled − homogeneous
AUC−ROC difference coupled − uncoupled
0
1/8
1/4
1/2
1
ε
homogeneous DBN
uncoupled NH−DBN
coupled NH−DBN

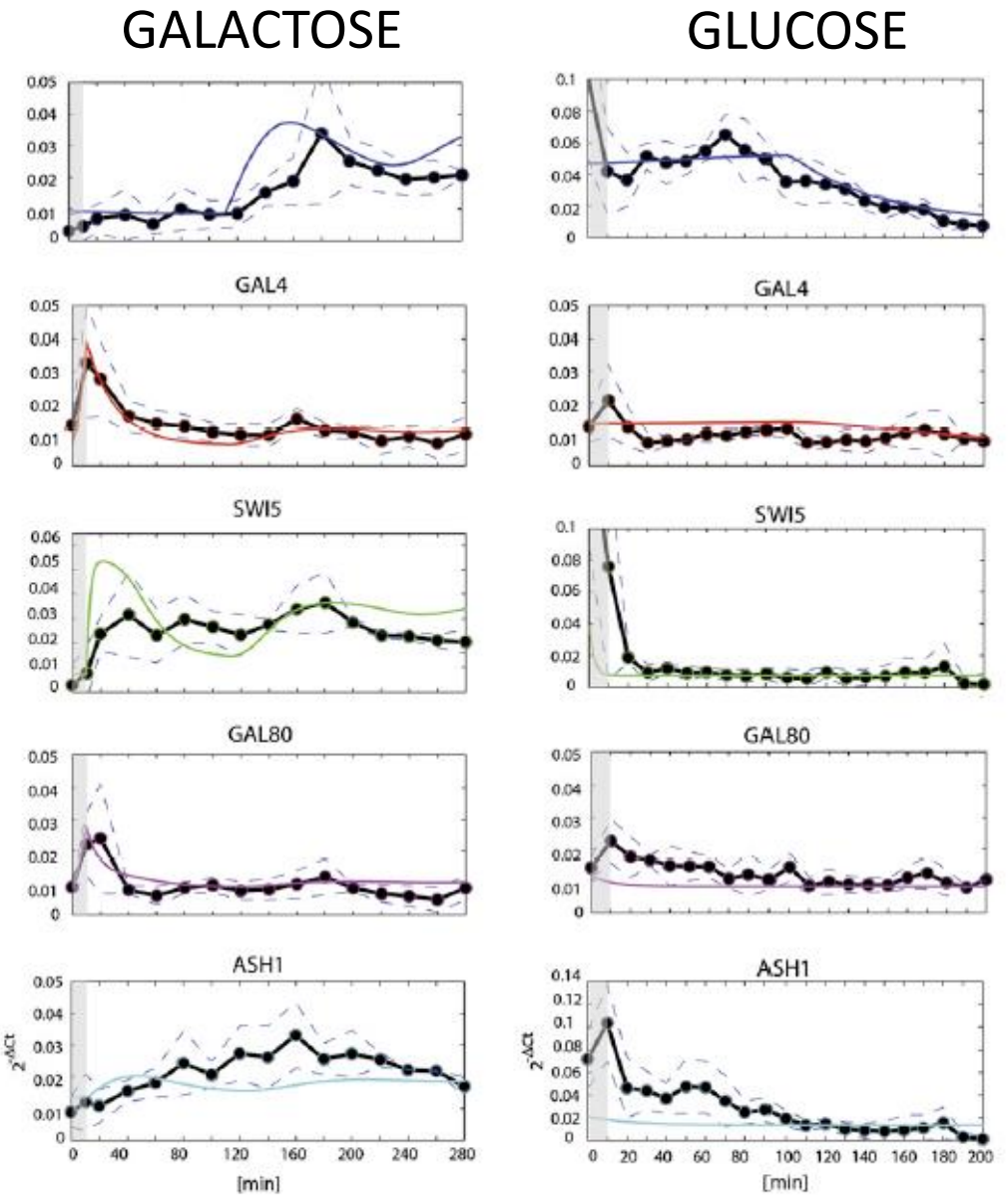

SNR=10
SNR=3
SNR=1
mean AUC−ROC total
AUC−ROC difference coupled − homogeneous
AUC−ROC difference coupled − uncoupled
1
0.5
+0.3
−0.3
+0.1
−0.1
0
1/8
1/4
1/2
1
ε
homogeneous DBN
uncoupled NH−DBN
coupled NH−DBN

2. Data from synthetic biology



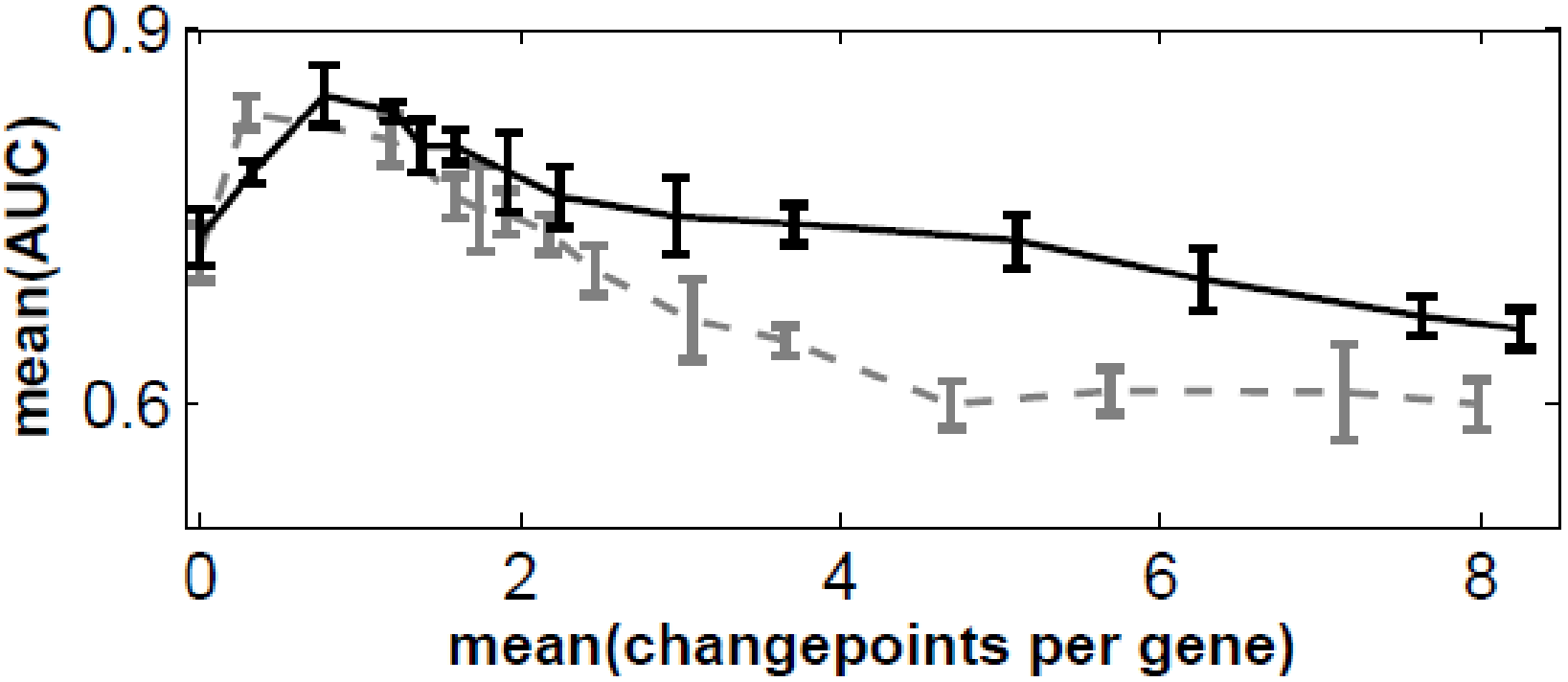
Synthetic network in yeast, **as designed in Cantone et al. (2009)**

Carbon-source switch from galactose to glucose

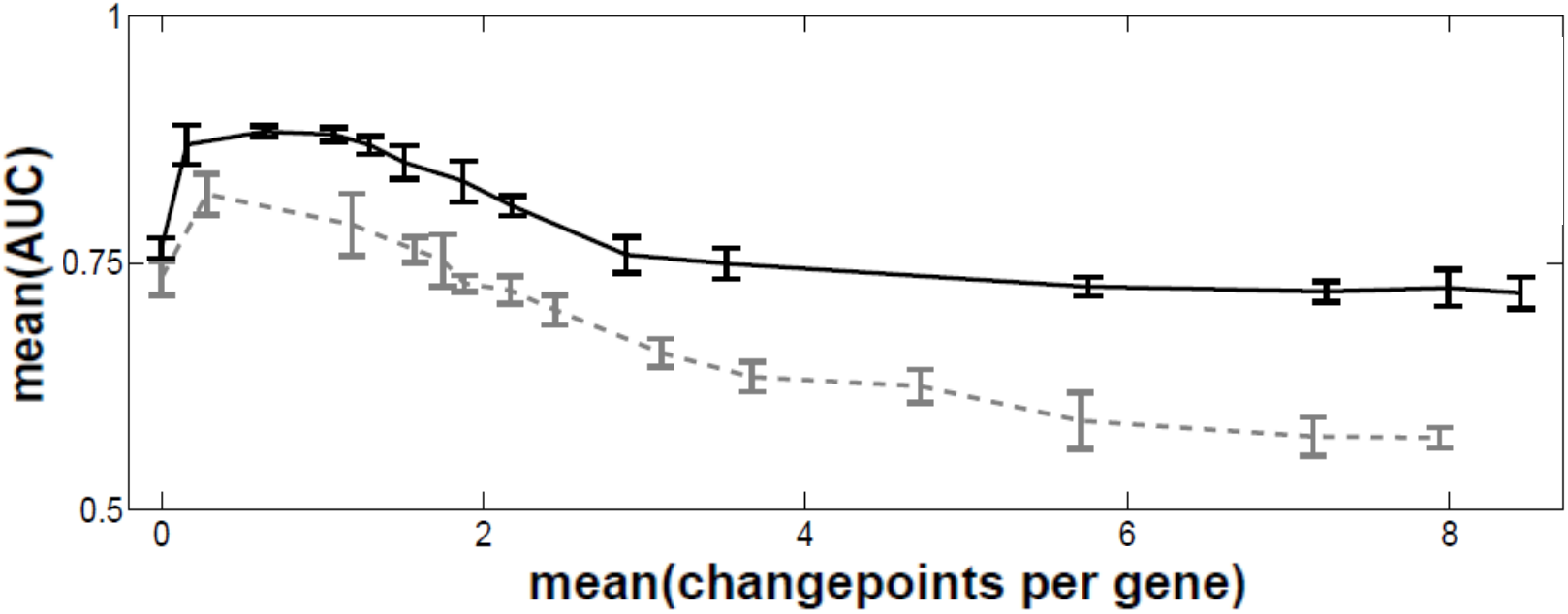


in vivo gene expression levels measured with RT-PCRat 37 time points (in two mediums)

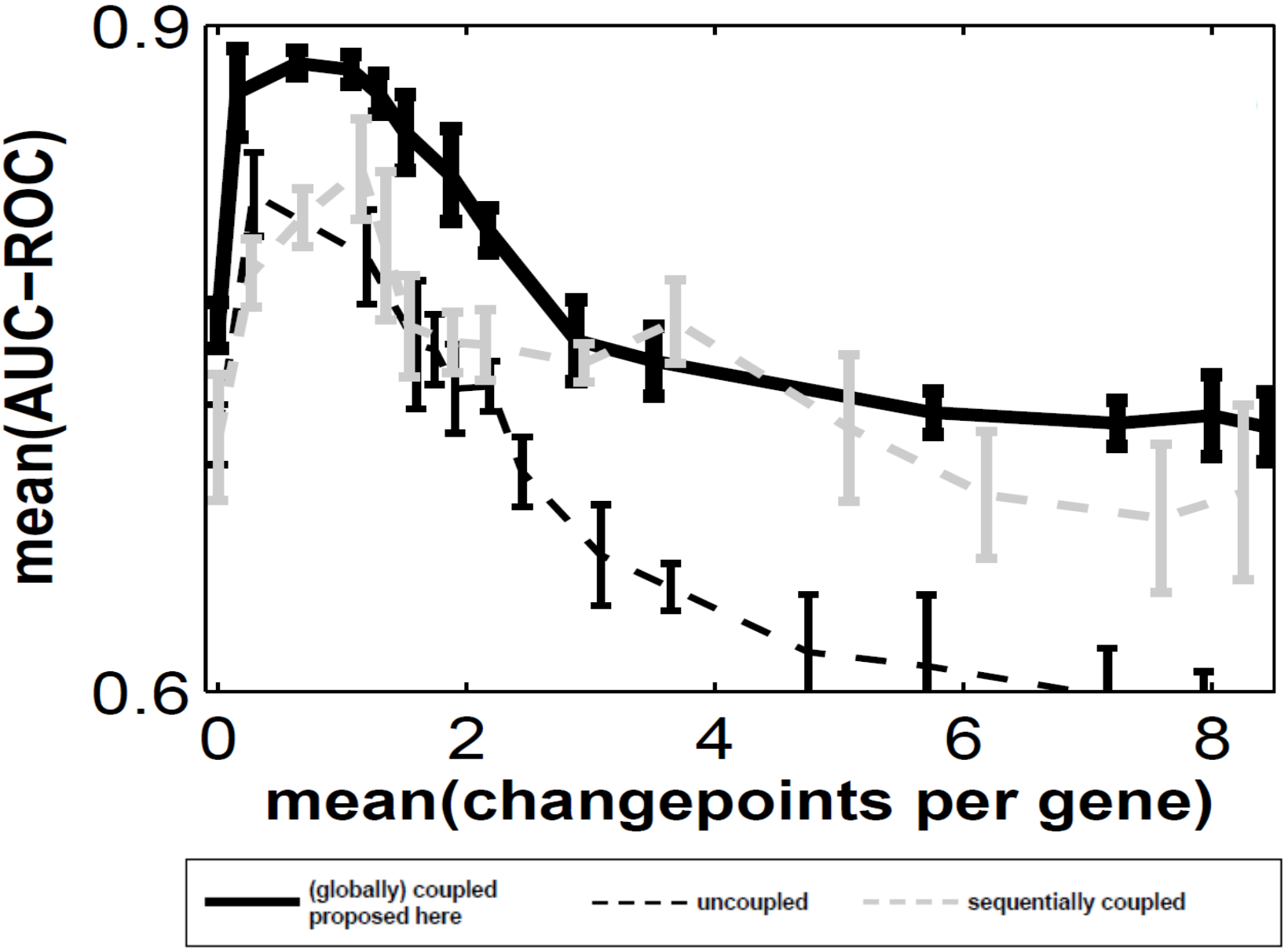
AUC score comparison
sequentially coupled NH-DBN versus
uncoupled NH-DBN
for different changepoint prior hyperparameters
(different numbers of changepoints per gene)



AUC score comparison
globally coupled NH-DBN versus
uncoupled NH-DBN
for different changepoint prior hyperparameters
(different numbers of changepoints per gene)

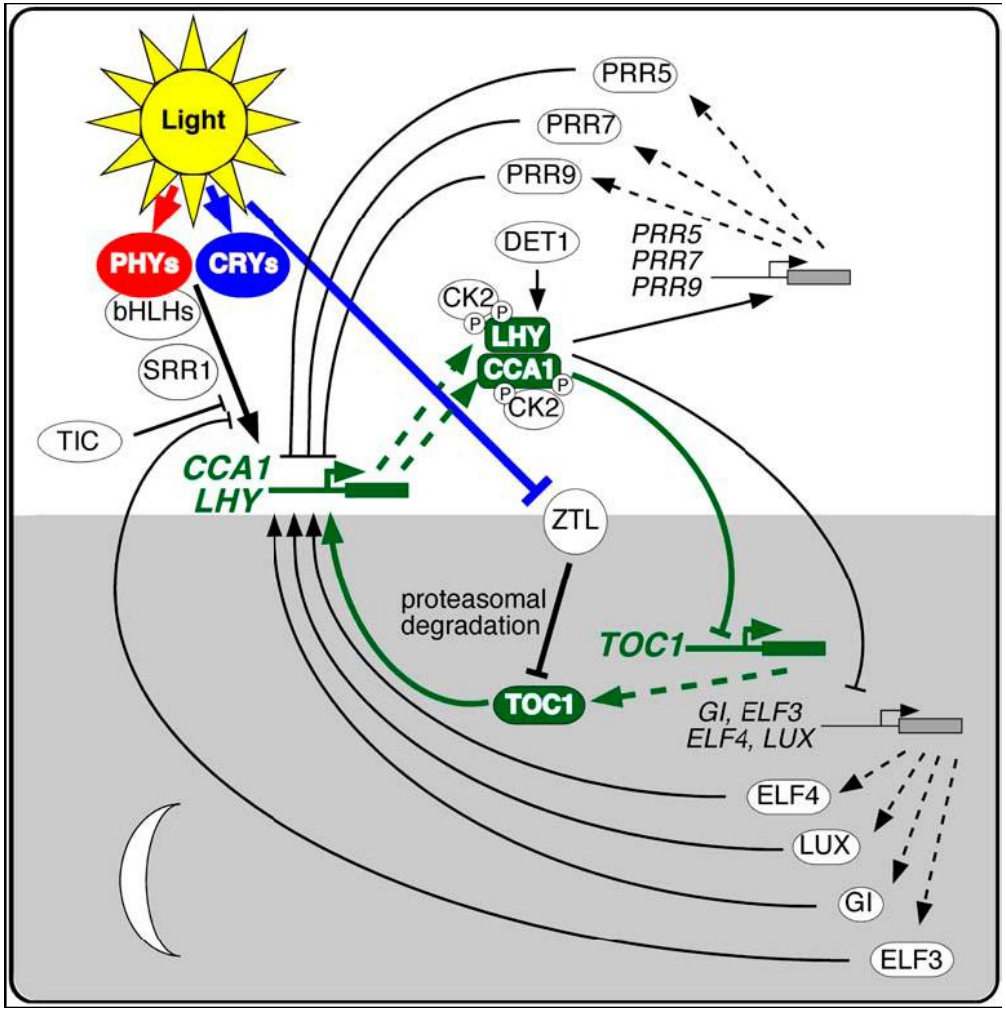


AUC score comparison of all three NH-DBNs



3. Data from a real application

Circadian regulation in Arabidopsis



Circadian rhythms in *Arabidopsis thaliana*

Collaboration with the Institute of Molecular Plant Sciences at Edinburgh University

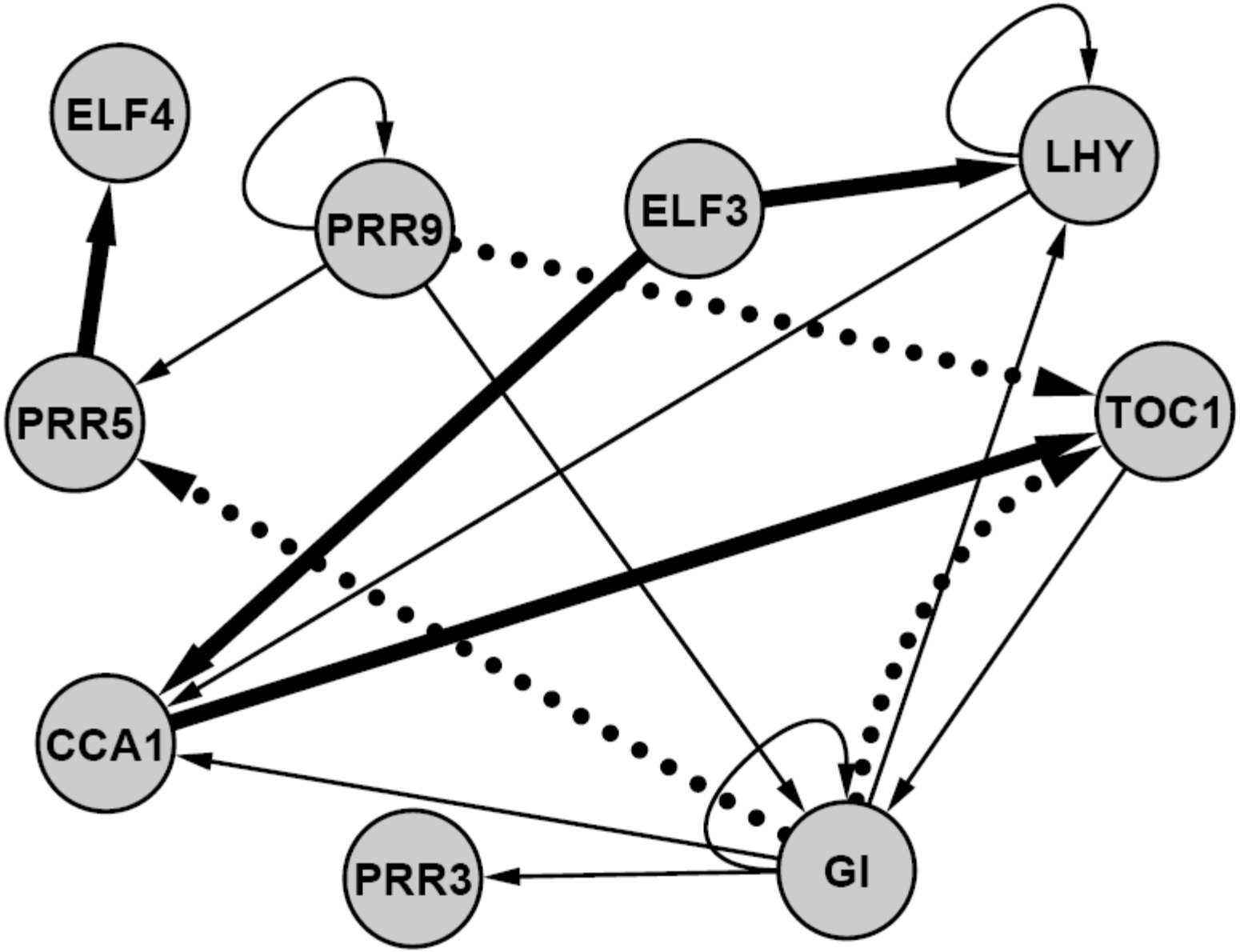
4 time series of microarray gene expression data from *Arabidopsis thaliana*.

- Focus on: 9 circadian genes:

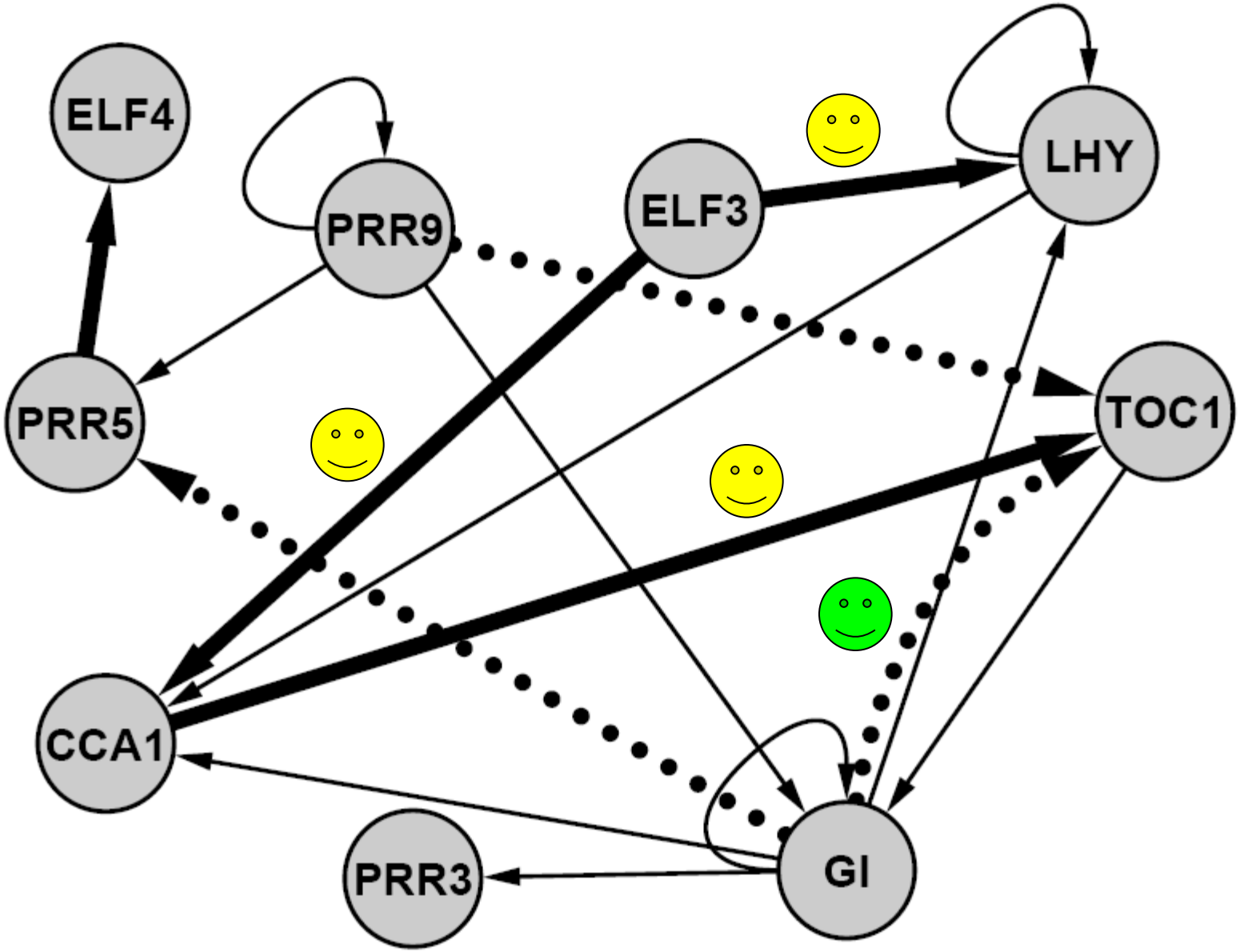
LHY, CCA1, TOC1, ELF4, ELF3, GI, PRR9, PRR5, and PRR3

- The four time series were measured under constant light condition at 13 time points: 0h, 2h,…, 24h, 26h

- Seedlings entrained with light:dark cycles of different periods



Thin black edges indicate interactions that are inferred with both NH-DBNs. Three edges (dotted) are inferred with the uncoupled NH-DBN only while four edges (bold) are inferred with the coupled NH-DBN only.



Thin black edges indicate interactions that are inferred with both NH-DBNs. Three edges (dotted) are inferred with the uncoupled NH-DBN only while four edges (bold) are inferred with the coupled NH-DBN only.

Thank you
for
your
attention!

Any
questions?