Reconstructing gene regulatory networks with Bayesian networks

Dirk Husmeier
GENOME
gene regulation

PROTEOME
protein-protein interactions

signal transduction

METABOLISM
Bio-chemical reactions
Protein signalling pathway

From Sachs et al Science 2005
Regulatory network
Network
unknown
Methodology

- Mechanistic models
- Bayesian networks
- Integration of biological prior knowledge
- Non-homogeneous Bayesian network for non-stationary processes
- Node-specific change-points and avoiding spurious feedback loops
Biological applications

• *Raf-Mek-Erk* protein signalling pathway
• Morphogenesis in *Drosophila melanogaster*
• Viral challenge and immune activation of macrophages
• Circadian regulation in *Arabidopsis thaliana*
Methodology

- Mechanistic models
- Bayesian networks
- Integration of biological prior knowledge
- Non-homogeneous Bayesian network for non-stationary processes
- Node-specific change-points and avoiding spurious feedback loops
Regulatory network
Elementary molecular components

DNA: Promoter

mRNA

Protein
Elementary molecular biological processes

- Binding / unbinding
- Transcription factors
- Promoter
- Transcription
- Degradation
- mRNA
- Translation
- Dimerization
- Undimerization
- Protein
Description with differential equations

\[
\frac{d}{dt}[a_2.rC] = \lambda^+_{a_2.rC}[a_2][rC] - \lambda^-_{a_2.rC}[a_2.rC]
\]

\[
\frac{d}{dt}[C] = \lambda_{rC}[rC] + \lambda_{a_2.rC}[a_2.rC] + \lambda_{b_2.rC}[b_2.rC] - \lambda_C[C]
\]

\[
\frac{d}{dt}[c] = \lambda_{Cc}[C] - \lambda_c[c]
\]

\[
\frac{d}{dt}[c_2] = \lambda^+_{cc}[c]^2 - \lambda^-_{cc}[c_2]
\]
Description with differential equations

\[
\frac{d}{dt}[a_{2,rC}] = \lambda^+_{a_{2,rC}}[a_2][rC] - \lambda^-_{a_{2,rC}}[a_2,rC]
\]

\[
\frac{d}{dt}[C] = \lambda_{rC}[rC] + \lambda_{a_{2,rC}}[a_2,rC] + \lambda_{b_{2,rC}}[b_2,rC] - \lambda_C[C]
\]

\[
\frac{d}{dt}[c] = \lambda_{C_c}[C] - \lambda_c[c]
\]

\[
\frac{d}{dt}[c_2] = \lambda^+_{cc}[c]^2 - \lambda^-_{cc}[c_2]
\]
\[
\frac{d[C]}{dt} = \lambda_r C [rC] + \lambda_{a2} C [a_2 rC] + \lambda_{b2} C [b_2 rC] - \lambda_C C
\]
Description with differential equations

\[
\frac{d}{dt}[a_{2,rC}] = \lambda_{a_{2,rC}}^+[a_2][rC] - \lambda_{a_{2,rC}}^-[a_2,rC]
\]

\[
\frac{d}{dt}[C] = \lambda_{rC}[rC] + \lambda_{a_{2,rC}}[a_2,rC] + \lambda_{b_{2,rC}}[b_2,rC] - \lambda_C[C]
\]

\[
\frac{d}{dt}[c] = \lambda_C[c] - \lambda_c[c]
\]

\[
\frac{d}{dt}[c_2] = \lambda_{cc}^+[c]^2 - \lambda_{cc}^-[c_2]
\]

Concentrations

Kinetic parameters \( q \)

Rates
Parameters $q$ known: Numerically integrate the differential equations for different hypothetical networks
Experiment:
Gene expression time series

Can we infer the correct gene regulatory network?
Model selection for known parameters $q$

Measured gene expression time series

Gene expression time series predicted with different models

Compare

Highest likelihood: best model

$p(D|q, M)$
Model selection for **unknown** parameters $q$

**Gene expression time series predicted with different models**

- $H_1$
- $H_2$
- $H_3$
- $H_4$

**Measured gene expression time series**

Highest likelihood: **over-fitting**

$$P(D|q, M)$$
Regularization

E.g.: BIC

\[ \log P(D|\hat{\mathbf{q}}, \mathcal{M}) - \frac{k}{2} \log N \]

- Maximum likelihood parameters
- Number of parameters
- Number of data points
Model selection: find the best pathway

Select the model $\mathcal{M}$ with the highest posterior probability:

$$P(\mathcal{M}|\mathcal{D}) \propto P(\mathcal{D}|\mathcal{M})P(\mathcal{M})$$

This requires an integration over the whole parameter space:

$$P(\mathcal{D}|\mathcal{M}) = \int P(\mathcal{D}|\mathbf{q}, \mathcal{M})P(\mathbf{q}|\mathcal{M})d\mathbf{q}$$
Model selection: find the best pathway

Select the model $\mathcal{M}$ with the highest posterior probability:

$$P(\mathcal{M}|\mathcal{D}) \propto P(\mathcal{D}|\mathcal{M})P(\mathcal{M})$$

This requires an integration over the whole parameter space:

$$P(\mathcal{D}|\mathcal{M}) = \int P(\mathcal{D}|q, \mathcal{M})P(q|\mathcal{M})dq$$

This integral is usually analytically intractable
Numerical integration by sampling from the prior

Model: $S$  Parameters: $\phi$

$$P(\mathcal{D}|S) = \int P(\mathcal{D}|\phi, S)P(\phi|S)d\phi$$

$$P(\mathcal{D}|S) \approx \frac{1}{N} \sum_{t=1}^{N} P(\mathcal{D}|\phi_t, S)$$

where $\{\phi_t\}$ is a sample from the prior distribution $P(\phi|S)$
Problem: Extremely poor convergence in high dimensions

Prior distribution
\[ P(\phi | S) \]

Likelihood function
\[ P(\mathcal{D} | \phi, S) \]

Taken from the MSc thesis by Ben Calderhead
Numerical integration by sampling from the posterior

Model: \( S \)  
Parameters: \( \phi \)

\[
P(\mathcal{D} | \phi, S) P(\phi | S) = P(\phi | \mathcal{D}, S) P(\mathcal{D} | S)
\]

\[
\int \frac{P(\phi | S)}{P(\mathcal{D} | S)} d\phi = \int \frac{P(\phi | \mathcal{D}, S)}{P(\mathcal{D} | \phi, S)} d\phi
\]

\[
\frac{1}{P(\mathcal{D} | S)} = \int \frac{P(\phi | \mathcal{D}, S)}{P(\mathcal{D} | \phi, S)} d\phi
\]

\[
\frac{1}{P(\mathcal{D} | S)} \approx \frac{1}{N} \sum_{t=1}^{N} \frac{1}{P(\mathcal{D} | \phi_t, S)}
\]

where \( \{\phi_t\} \) is a sample from the posterior distribution \( P(\phi | \mathcal{D}, S) \)
Problem: Poor convergence in high dimensions and instability

Prior distribution

\[ P(\phi|S) \]

Likelihood function

\[ P(\mathcal{D}|\phi, S) \]

\[ \approx \]

Posterior distribution

\[ P(\phi|\mathcal{D}, S) \]

Taken from the MSc thesis by Ben Calderhead
Importance sampling

\[ P(\mathcal{D}|S) = \int P(\mathcal{D}|\phi, S) P(\phi|S) d\phi \]

Arbitrary (possibly unnormalized) distribution \( Q(\phi) \)

\[ \frac{P(\mathcal{D}|S)}{Z_Q} = \int \frac{P(\mathcal{D}|\phi, S) P(\phi|S) Q(\phi)}{Z_Q} d\phi \]

\[ \frac{P(\mathcal{D}|S)}{Z_Q} \xlongequal{\text{sampled from}} \frac{1}{N} \sum_{t=1}^{N} c_t \]

\[ c_t = \frac{P(\mathcal{D}|\phi_t, S) P(\phi_t|S)}{Q(\phi_t)} \]
Annealed importance sampling

RADFORD M. NEAL*

Department of Statistics and Department of Computer Science, University of Toronto, Toronto, Ontario, Canada
radford@stat.utoronto.ca

Received March 1998 and accepted February 2000
Illustration of annealed importance sampling

Prior distribution

Posterior distribution

Taken from the MSc thesis by Ben Calderhead,
Outer loop:
Annealing scheme

Centre loop:
MCMC

Inner loop:
Numerical solution of differential equations
Systems biology

Bayesian ranking of biochemical system models

Vladislav Vyshemirsky* and Mark A. Girolami

Department of Computing Science, University of Glasgow, Glasgow, G12 8QQ, UK

Received on August 28, 2007; revised on October 26, 2007; accepted on December 3, 2007

Advance Access publication December 5, 2007

Associate Editor: Limsoon Wong
Outer loop:
Annealing scheme

Centre loop:
MCMC

Inner loop:
Numerical solution of differential equations

Accelerating Bayesian Inference over Nonlinear Differential Equations with Gaussian Processes

Ben Calderhead
Dept. of Computing Sci.
University of Glasgow
bc@cs.gla.ac.uk

Mark Girolami
Dept. of Computing Sci.
University of Glasgow
girolami@cs.gla.ac.uk

Neil D. Lawrence
School of Computer Sci.
University of Manchester
neil1@cs.man.ac.uk

NIPS 09
Objective: Reconstruction of regulatory networks \textit{ab initio}

Higher level of abstraction:
Bayesian networks
Methodology

• Mechanistic models
• **Bayesian networks**
• Integration of biological prior knowledge
• Non-homogeneous Bayesian network for non-stationary processes
• Node-specific change-points and avoiding spurious feedback loops
A Primer on Learning in Bayesian Networks for Computational Biology

Chris J. Needham*, James R. Bradford, Andrew J. Bulpitt, David R. Westhead

August 2007
Volume 3
Issue 8

Marriage between
graph theory
and
probability theory

Friedman et al. (2000), J. Comp. Biol. 7, 601-620
Bayesian network

Low osmolarity response genes

SLT2

MAP kinase

Rlm1p

Transcription factors

Swi4/6

Low osmolarity response genes

ODE model
Bayesian networks

• Marriage between graph theory and probability theory.

• Directed acyclic graph (DAG) representing conditional independence relations.

• It is possible to score a network in light of the data: $P(D|M)$, $D$: data, $M$: network structure.

• We can infer how well a particular network explains the observed data.

$$P(A, B, C, D, E, F) = P(A) \cdot P(B \mid A) \cdot P(C \mid A) \cdot P(D \mid B, C) \cdot P(E \mid D) \cdot P(F \mid C, D)$$
Linear model

\[ [A] = w_1[P_1] + w_2[P_2] + w_3[P_3] + w_4[P_4] + \text{noise} \]

\[ P(A|P_1, \ldots, P_n) = N \left( w_0 + \sum_{i=1}^{n} w_i P_i, \sigma^2 \right) \]
Nonlinear discretized model

- **Activator**: $P_1$ (blue)
- **Repressor**: $P_2$ (red)

**Activation**

- From $P_1$ to Activator
- From $P_2$ to Activator

**Inhibition**

- From $P_1$ to Repressor
- From $P_2$ to Repressor

**Allow for noise: probabilities**

- Conditional multinomial distribution

**Graphical Representation**

- Two diagrams showing the interactions between $P_1$, $P_2$, Activator, and Repressor.
Model $\mathcal{M}$

Parameters $\mathbf{q}$

$$P(\mathcal{D}|\mathcal{M}) = \int P(\mathcal{D}|\mathbf{q},\mathcal{M}) P(\mathbf{q}|\mathcal{M}) d\mathbf{q}$$

Integral \textit{analytically tractable}!
**BDe:**
Multinomial with a Dirichlet prior
Heckerman, Geiger, Chickering (1995)
Learning Bayesian Networks:
The Combination of Knowledge and Statistical Data
*Machine learning* 20, 245-274

**BGe:**
Linear Gaussian with a normal-gamma prior
Geiger and Heckerman (1994)
Learning Gaussian networks
*Proceedings of the Tenth Conference on Uncertainty in Artificial Intelligence*
Morgan Kaufmann publisher, San Francisco, 235-243
Example: 2 genes $\rightarrow$ 16 different network structures

Best network: maximum score $P(D|M)$
Identify the best network structure

Ideal scenario: Large data sets, low noise

$P(\mathcal{D}|\mathcal{M})$
Uncertainty about the best network structure

Limited number of experimental replications, high noise

$P(D|M)$
Sample of high-scoring networks

\[ P(D|M) \]
Sample of high-scoring networks

Feature extraction, e.g. marginal posterior probabilities of the edges
Sample of high-scoring networks

Feature extraction, e.g. marginal posterior probabilities of the edges

High-confident edge

High-confident non-edge

Uncertainty about edges
Can we generalize this scheme to more than 2 genes?

In principle yes.

However …
Number of structures

Number of nodes
Sampling from the posterior distribution

$$P(M|D) \propto P(D|M)P(M)$$

Configuration space of network structures $\mathcal{M}$

Find the high-scoring structures

Taken from the MSc thesis by Ben Calderhead
MCMC

Local change $\mathcal{M}_{old} \rightarrow \mathcal{M}_{new}$

If $P(\mathcal{D}|\mathcal{M}_{new}) > P(\mathcal{D}|\mathcal{M}_{old})$ accept

If $P(\mathcal{D}|\mathcal{M}_{new}) < P(\mathcal{D}|\mathcal{M}_{old})$ accept with probability

\[
\frac{P(\mathcal{D}|\mathcal{M}_{new})}{P(\mathcal{D}|\mathcal{M}_{old})}
\]

Configuration space of network structures $\mathcal{M}$

Taken from the MSc thesis by Ben Calderhead
MCMC moves

- Delete edge
- Reverse edge
- Create edge
Problem: Local changes $\rightarrow$ small steps $\rightarrow$ slow convergence, difficult to cross valleys.

Configuration space of network structures $\mathcal{M}$.
Problem: Global changes $\rightarrow$ large steps $\rightarrow$ low acceptance $\rightarrow$ slow convergence.
Can we make global changes that jump onto other peaks and are likely to be accepted?

$P(M|D)$

Configuration space of network structures $\mathcal{M}$
Improving the structure MCMC sampler for Bayesian networks by introducing a new edge reversal move

Marco Grzegorczyk • Dirk Husmeier

Received: 27 March 2007 / Revised: 11 January 2008 / Accepted: 28 March 2008 / Published online: 17 April 2008
Springer Science+Business Media, LLC 2008
Methodology

• Mechanistic models
• Bayesian networks
• Integration of biological prior knowledge
• Non-homogeneous Bayesian network for non-stationary processes
• Node-specific change-points and avoiding spurious feedback loops
Reconstructing Gene Regulatory Networks with Bayesian Networks by Combining Expression Data with Multiple Sources of Prior Knowledge

Adriano V. Werhli* Dirk Husmeier†
Bayesian inference

Select the model $\mathcal{M}$ based on the posterior probability:

$$P(\mathcal{M}|D) \propto P(D|\mathcal{M})P(\mathcal{M})$$

This requires an integration over the whole parameter space:

$$P(D|\mathcal{M}) = \int P(D|q, \mathcal{M})P(q|\mathcal{M})dq$$
Uncertainty about the best network structure

Limited number of experimental replications,
high noise
Reduced uncertainty by using prior knowledge

\[ P(M|D) \propto P(D|M)P(M) \]
Bayesian analysis: integration of prior knowledge

Hyperparameter $\beta$ trades off data versus prior knowledge

Microarray data

KEGG pathway
Hyperparameter $\beta$ trades off data versus prior knowledge.
Hyperparameter $\beta$ trades off data versus prior knowledge

Microarray data

KEGG pathway
Input:

$D$

MCMC

Learn:

$\mathcal{M}$

$\beta, \mathcal{M} \sim \mathcal{P}(\beta, \mathcal{M} | D, \mathcal{B})$
Deviation between the network $G$ and the prior knowledge $B$:

$$E(G) = \sum_{i,j=1}^{N} |B_{i,j} - G_{i,j}|$$

Prior distribution over networks:

$$P(G|\beta) = \frac{e^{-\beta E(G)}}{Z(\beta)}$$

$$Z(\beta) = \sum_{G \in \mathcal{G}} e^{-\beta E(G)}$$
Sample networks and hyperparameters from the posterior distribution with MCMC

\[ P(G, \beta_1, \beta_2 | D) \]

Proposal probabilities

\[ Q(G_{\text{new}} | G_{\text{old}}) \]
\[ R(\beta_{1\text{new}} | \beta_{1\text{old}}) \]
\[ R(\beta_{2\text{new}} | \beta_{2\text{old}}) \]

Metropolis-Hastings scheme

\[ A = \min \left\{ \frac{P(D, G_{\text{new}}, \beta_{1\text{new}}, \beta_{2\text{new}})Q(G_{\text{old}} | G_{\text{new}})R(\beta_{1\text{old}} | \beta_{1\text{new}})R(\beta_{2\text{old}} | \beta_{2\text{new}})}{P(D, G_{\text{old}}, \beta_{1\text{old}}, \beta_{2\text{old}})Q(G_{\text{new}} | G_{\text{old}})R(\beta_{1\text{new}} | \beta_{1\text{old}})R(\beta_{2\text{new}} | \beta_{2\text{old}})}, 1 \right\} \]
Biological applications

- **Raf-Mek-Erk** protein signalling pathway
- Morphogenesis in *Drosophila melanogaster*
- Viral challenge and immune activation of macrophages
- Circadian regulation in *Arabidopsis thaliana*
Example: Protein signalling pathway

Cell membran

phosphorylation

nucleus

-> cell response
Raf signalling pathway

From Sachs et al Science 2005

Activators
1. α-CD3
2. α-CD28
3. ICAM-2
4. PMA
5. β2cAMP

Inhibitors
6. G06976
7. AKT inh
8. Psitect
9. U0126
10. LY294002
Flow cytometry data

Causal Protein-Signaling Networks Derived from Multiparameter Single-Cell Data
Karen Sachs,1* Omar Perez,2* Dana Pe’er,3* Douglas A. Lauffenburger,1† Garry P. Nolan2†

- Intracellular multicolour flow cytometry experiments: concentrations of 11 proteins
- 5400 cells have been measured under 9 different cellular conditions (cues)
- Downsampling to 100 instances (5 separate subsets): indicative of microarray experiments
Gold-standard network

Inferred network

Deterministic inference
Sample of high-scoring networks

\[ P(\mathcal{M} | \mathcal{D}) \]
Sample of high-scoring networks

Feature extraction, e.g. marginal posterior probabilities of the edges

High-confident edge

High-confident non-edge

Uncertainty about edges
Gold standard network

Inferred network distribution

Data

Probabilistic inference
Gold-standard network

Thresholding

True positives False positives
Prior knowledge from KEGG

Data: protein concentrations from flow cytometry experiments
Protein signalling network from the literature

ROC Curve

AUC
Predicted network

11 nodes, 20 edges, 90 non-edges

20 top-scoring edges: 15/20 correct, 5/90 false

75% 94%
Methodology

- Mechanistic models
- Bayesian networks
- Integration of biological prior knowledge
- **Non-homogeneous Bayesian network for non-stationary processes**
- Node-specific change-points and avoiding spurious feedback loops
Systems biology

Modelling non-stationary gene regulatory processes with a non-homogeneous Bayesian network and the allocation sampler

Marco Grzegorczyk¹,²,*, Dirk Husmeier²,³,*, Kieron D. Edwards⁴, Peter Ghazal²,⁵ and Andrew J. Millar¹,²

¹School of Biological Sciences, The University of Edinburgh, Swann Building, King's Buildings, Edinburgh EH9 3JR, ²Centre for Systems Biology at Edinburgh (CSBE), Darwin Building, King's Buildings, Edinburgh EH9 3JU, ³Biomathematics and Statistics Scotland (BioSS), JCMB, King's Buildings, Edinburgh EH9 3JZ, ⁴Advanced Technologies (Cambridge) Ltd, Cambridge CB4 0WA and ⁵Division of Pathway Medicine (DPM), Medical School, The University of Edinburgh, Chancellor's Buildings, Edinburgh EH16 4SB, UK

Received on May 17, 2008; revised on July 9, 2009; accepted on July 14, 2008

Advance Access publication July 26, 2008

Associate Editor: Trey Ideker
Dynamic Bayesian network
Example: 4 genes, 10 time points

<table>
<thead>
<tr>
<th></th>
<th>t₁</th>
<th>t₂</th>
<th>t₃</th>
<th>t₄</th>
<th>t₅</th>
<th>t₆</th>
<th>t₇</th>
<th>t₈</th>
<th>t₉</th>
<th>t₁₀</th>
</tr>
</thead>
<tbody>
<tr>
<td>(X^{(1)})</td>
<td>(X_{1,1})</td>
<td>(X_{1,2})</td>
<td>(X_{1,3})</td>
<td>(X_{1,4})</td>
<td>(X_{1,5})</td>
<td>(X_{1,6})</td>
<td>(X_{1,7})</td>
<td>(X_{1,8})</td>
<td>(X_{1,9})</td>
<td>(X_{1,10})</td>
</tr>
<tr>
<td>(X^{(2)})</td>
<td>(X_{2,1})</td>
<td>(X_{2,2})</td>
<td>(X_{2,3})</td>
<td>(X_{2,4})</td>
<td>(X_{2,5})</td>
<td>(X_{2,6})</td>
<td>(X_{2,7})</td>
<td>(X_{2,8})</td>
<td>(X_{2,9})</td>
<td>(X_{2,10})</td>
</tr>
<tr>
<td>(X^{(3)})</td>
<td>(X_{3,1})</td>
<td>(X_{3,2})</td>
<td>(X_{3,3})</td>
<td>(X_{3,4})</td>
<td>(X_{3,5})</td>
<td>(X_{3,6})</td>
<td>(X_{3,7})</td>
<td>(X_{3,8})</td>
<td>(X_{3,9})</td>
<td>(X_{3,10})</td>
</tr>
<tr>
<td>(X^{(4)})</td>
<td>(X_{4,1})</td>
<td>(X_{4,2})</td>
<td>(X_{4,3})</td>
<td>(X_{4,4})</td>
<td>(X_{4,5})</td>
<td>(X_{4,6})</td>
<td>(X_{4,7})</td>
<td>(X_{4,8})</td>
<td>(X_{4,9})</td>
<td>(X_{4,10})</td>
</tr>
</tbody>
</table>
### Standard dynamic Bayesian network: homogeneous model

<table>
<thead>
<tr>
<th></th>
<th>$t_1$</th>
<th>$t_2$</th>
<th>$t_3$</th>
<th>$t_4$</th>
<th>$t_5$</th>
<th>$t_6$</th>
<th>$t_7$</th>
<th>$t_8$</th>
<th>$t_9$</th>
<th>$t_{10}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X^{(1)}$</td>
<td>$X_{1,1}$</td>
<td>$X_{1,2}$</td>
<td>$X_{1,3}$</td>
<td>$X_{1,4}$</td>
<td>$X_{1,5}$</td>
<td>$X_{1,6}$</td>
<td>$X_{1,7}$</td>
<td>$X_{1,8}$</td>
<td>$X_{1,9}$</td>
<td>$X_{1,10}$</td>
</tr>
<tr>
<td>$X^{(2)}$</td>
<td>$X_{2,1}$</td>
<td>$X_{2,2}$</td>
<td>$X_{2,3}$</td>
<td>$X_{2,4}$</td>
<td>$X_{2,5}$</td>
<td>$X_{2,6}$</td>
<td>$X_{2,7}$</td>
<td>$X_{2,8}$</td>
<td>$X_{2,9}$</td>
<td>$X_{2,10}$</td>
</tr>
<tr>
<td>$X^{(3)}$</td>
<td>$X_{3,1}$</td>
<td>$X_{3,2}$</td>
<td>$X_{3,3}$</td>
<td>$X_{3,4}$</td>
<td>$X_{3,5}$</td>
<td>$X_{3,6}$</td>
<td>$X_{3,7}$</td>
<td>$X_{3,8}$</td>
<td>$X_{3,9}$</td>
<td>$X_{3,10}$</td>
</tr>
<tr>
<td>$X^{(4)}$</td>
<td>$X_{4,1}$</td>
<td>$X_{4,2}$</td>
<td>$X_{4,3}$</td>
<td>$X_{4,4}$</td>
<td>$X_{4,5}$</td>
<td>$X_{4,6}$</td>
<td>$X_{4,7}$</td>
<td>$X_{4,8}$</td>
<td>$X_{4,9}$</td>
<td>$X_{4,10}$</td>
</tr>
</tbody>
</table>
Our new model: heterogeneous dynamic Bayesian network. Here: 2 components

<table>
<thead>
<tr>
<th></th>
<th>$t_1$</th>
<th>$t_2$</th>
<th>$t_3$</th>
<th>$t_4$</th>
<th>$t_5$</th>
<th>$t_6$</th>
<th>$t_7$</th>
<th>$t_8$</th>
<th>$t_9$</th>
<th>$t_{10}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X^{(1)}$</td>
<td></td>
<td>$X_{1,1}$</td>
<td>$X_{1,2}$</td>
<td>$X_{1,3}$</td>
<td>$X_{1,4}$</td>
<td>$X_{1,5}$</td>
<td>$X_{1,6}$</td>
<td>$X_{1,7}$</td>
<td>$X_{1,8}$</td>
<td>$X_{1,9}$</td>
</tr>
<tr>
<td>$X^{(2)}$</td>
<td>$X_{2,1}$</td>
<td></td>
<td>$X_{2,2}$</td>
<td>$X_{2,3}$</td>
<td>$X_{2,4}$</td>
<td>$X_{2,5}$</td>
<td>$X_{2,6}$</td>
<td>$X_{2,7}$</td>
<td>$X_{2,8}$</td>
<td>$X_{2,9}$</td>
</tr>
<tr>
<td>$X^{(3)}$</td>
<td>$X_{3,1}$</td>
<td></td>
<td>$X_{3,2}$</td>
<td>$X_{3,3}$</td>
<td>$X_{3,4}$</td>
<td>$X_{3,5}$</td>
<td>$X_{3,6}$</td>
<td>$X_{3,7}$</td>
<td>$X_{3,8}$</td>
<td>$X_{3,9}$</td>
</tr>
<tr>
<td>$X^{(4)}$</td>
<td>$X_{4,1}$</td>
<td></td>
<td>$X_{4,2}$</td>
<td>$X_{4,3}$</td>
<td>$X_{4,4}$</td>
<td>$X_{4,5}$</td>
<td>$X_{4,6}$</td>
<td>$X_{4,7}$</td>
<td>$X_{4,8}$</td>
<td>$X_{4,9}$</td>
</tr>
</tbody>
</table>
Our new model: heterogeneous dynamic Bayesian network. Here: 3 components

<table>
<thead>
<tr>
<th></th>
<th>$t_1$</th>
<th>$t_2$</th>
<th>$t_3$</th>
<th>$t_4$</th>
<th>$t_5$</th>
<th>$t_6$</th>
<th>$t_7$</th>
<th>$t_8$</th>
<th>$t_9$</th>
<th>$t_{10}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X^{(1)}$</td>
<td>$X_{1,1}$</td>
<td>$X_{1,2}$</td>
<td>$X_{1,3}$</td>
<td>$X_{1,4}$</td>
<td>$X_{1,5}$</td>
<td>$X_{1,6}$</td>
<td>$X_{1,7}$</td>
<td>$X_{1,8}$</td>
<td>$X_{1,9}$</td>
<td>$X_{1,10}$</td>
</tr>
<tr>
<td>$X^{(2)}$</td>
<td>$X_{2,1}$</td>
<td>$X_{2,2}$</td>
<td>$X_{2,3}$</td>
<td>$X_{2,4}$</td>
<td>$X_{2,5}$</td>
<td>$X_{2,6}$</td>
<td>$X_{2,7}$</td>
<td>$X_{2,8}$</td>
<td>$X_{2,9}$</td>
<td>$X_{2,10}$</td>
</tr>
<tr>
<td>$X^{(3)}$</td>
<td>$X_{3,1}$</td>
<td>$X_{3,2}$</td>
<td>$X_{3,3}$</td>
<td>$X_{3,4}$</td>
<td>$X_{3,5}$</td>
<td>$X_{3,6}$</td>
<td>$X_{3,7}$</td>
<td>$X_{3,8}$</td>
<td>$X_{3,9}$</td>
<td>$X_{3,10}$</td>
</tr>
<tr>
<td>$X^{(4)}$</td>
<td>$X_{4,1}$</td>
<td>$X_{4,2}$</td>
<td>$X_{4,3}$</td>
<td>$X_{4,4}$</td>
<td>$X_{4,5}$</td>
<td>$X_{4,6}$</td>
<td>$X_{4,7}$</td>
<td>$X_{4,8}$</td>
<td>$X_{4,9}$</td>
<td>$X_{4,10}$</td>
</tr>
</tbody>
</table>
Change-point process

Free allocation
Learning with MCMC

Number of components (here: 3)

Allocation vector

<table>
<thead>
<tr>
<th></th>
<th>$t_1$</th>
<th>$t_2$</th>
<th>$t_3$</th>
<th>$t_4$</th>
<th>$t_5$</th>
<th>$t_6$</th>
<th>$t_7$</th>
<th>$t_8$</th>
<th>$t_9$</th>
<th>$t_{10}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X^{(1)}$</td>
<td>$X_{1,1}$</td>
<td>$X_{1,2}$</td>
<td>$X_{1,3}$</td>
<td>$X_{1,4}$</td>
<td>$X_{1,5}$</td>
<td>$X_{1,6}$</td>
<td>$X_{1,7}$</td>
<td>$X_{1,8}$</td>
<td>$X_{1,9}$</td>
<td>$X_{1,10}$</td>
</tr>
<tr>
<td>$X^{(2)}$</td>
<td>$X_{2,1}$</td>
<td>$X_{2,2}$</td>
<td>$X_{2,3}$</td>
<td>$X_{2,4}$</td>
<td>$X_{2,5}$</td>
<td>$X_{2,6}$</td>
<td>$X_{2,7}$</td>
<td>$X_{2,8}$</td>
<td>$X_{2,9}$</td>
<td>$X_{2,10}$</td>
</tr>
<tr>
<td>$X^{(3)}$</td>
<td>$X_{3,1}$</td>
<td>$X_{3,2}$</td>
<td>$X_{3,3}$</td>
<td>$X_{3,4}$</td>
<td>$X_{3,5}$</td>
<td>$X_{3,6}$</td>
<td>$X_{3,7}$</td>
<td>$X_{3,8}$</td>
<td>$X_{3,9}$</td>
<td>$X_{3,10}$</td>
</tr>
<tr>
<td>$X^{(4)}$</td>
<td>$X_{4,1}$</td>
<td>$X_{4,2}$</td>
<td>$X_{4,3}$</td>
<td>$X_{4,4}$</td>
<td>$X_{4,5}$</td>
<td>$X_{4,6}$</td>
<td>$X_{4,7}$</td>
<td>$X_{4,8}$</td>
<td>$X_{4,9}$</td>
<td>$X_{4,10}$</td>
</tr>
</tbody>
</table>
Learning with MCMC

\[ \mathcal{M} \rightarrow q \]

\[ D \]

\begin{tabular}{cccccccccc}
  \hline
  & \( t_1 \) & \( t_2 \) & \( t_3 \) & \( t_4 \) & \( t_5 \) & \( t_6 \) & \( t_7 \) & \( t_8 \) & \( t_9 \) & \( t_{10} \) \\
  \( X^{(1)} \) & \( X_{1,1} \) & \( X_{1,2} \) & \( X_{1,3} \) & \( X_{1,4} \) & \( X_{1,5} \) & \( X_{1,6} \) & \( X_{1,7} \) & \( X_{1,8} \) & \( X_{1,9} \) & \( X_{1,10} \) \\
  \( X^{(2)} \) & \( X_{2,1} \) & \( X_{2,2} \) & \( X_{2,3} \) & \( X_{2,4} \) & \( X_{2,5} \) & \( X_{2,6} \) & \( X_{2,7} \) & \( X_{2,8} \) & \( X_{2,9} \) & \( X_{2,10} \) \\
  \( X^{(3)} \) & \( X_{3,1} \) & \( X_{3,2} \) & \( X_{3,3} \) & \( X_{3,4} \) & \( X_{3,5} \) & \( X_{3,6} \) & \( X_{3,7} \) & \( X_{3,8} \) & \( X_{3,9} \) & \( X_{3,10} \) \\
  \( X^{(4)} \) & \( X_{4,1} \) & \( X_{4,2} \) & \( X_{4,3} \) & \( X_{4,4} \) & \( X_{4,5} \) & \( X_{4,6} \) & \( X_{4,7} \) & \( X_{4,8} \) & \( X_{4,9} \) & \( X_{4,10} \) \\
  \hline
\end{tabular}
Learning with MCMC

- Parameters fixed
- Complexity of marginalization: \( k \times m \)

\[
\begin{array}{cccccccccccc}
    & t_1 & t_2 & t_3 & t_4 & t_5 & t_6 & t_7 & t_8 & t_9 & t_{10} \\
X^{(1)} & X_{1,1} & X_{1,2} & X_{1,3} & X_{1,4} & X_{1,5} & X_{1,6} & X_{1,7} & X_{1,8} & X_{1,9} & X_{1,10} \\
X^{(2)} & X_{2,1} & X_{2,2} & X_{2,3} & X_{2,4} & X_{2,5} & X_{2,6} & X_{2,7} & X_{2,8} & X_{2,9} & X_{2,10} \\
X^{(3)} & X_{3,1} & X_{3,2} & X_{3,3} & X_{3,4} & X_{3,5} & X_{3,6} & X_{3,7} & X_{3,8} & X_{3,9} & X_{3,10} \\
X^{(4)} & X_{4,1} & X_{4,2} & X_{4,3} & X_{4,4} & X_{4,5} & X_{4,6} & X_{4,7} & X_{4,8} & X_{4,9} & X_{4,10} \\
\end{array}
\]
Learning with MCMC

Parameters not fixed

Complexity of marginalization: $k^m$
Learning with MCMC

Parameters can be integrated out

Allocations fixed
Bayesian finite mixtures with an unknown number of components: The allocation sampler

Agostino Nobile • Alastair T. Fearnside

Received: August 2005 / Accepted: December 2006 / Published online: 7 February 2007
© Springer Science + Business Media, LLC 2007
Allocation sampler versus change-point process

- More flexibility, unrestricted mixture model.
- Not restricted to time series
- Higher computational costs

- Incorporates plausible prior knowledge for time series.
- Reduced complexity
- Less universal, not applicable to static data
Prior probability of assigning two time points to the same component. White=1. Black=0.

Allocation sampler  Change-point process
Change-point model versus free allocation: *Arabidopsis thaliana* (13 time points)

![Figure 4](image)
Synthetic study: posterior probability of the number of components

(a) $K_{TRUE} = 1$

(b) $K_{TRUE} = 2$

(c) $K_{TRUE} = 3$

(d) $K_{TRUE} = 4$

(e) $K_{TRUE} = 5$
Biological applications

- "Raf-Mek-Erk" protein signalling pathway
- **Morphogenesis in *Drosophila melanogaster***
- Viral challenge and immune activation of macrophages
- Circadian regulation in *Arabidopsis thaliana*
Morphogenesis in *Drosophila melanogaster*

- Gene expression measurements over 66 time steps of 4028 genes (Arbeitman et al., Science, 2002).
- Selection of 11 genes involved in muscle development.

Zhao et al. (2006), Bioinformatics 22
Heterogeneous dynamic Bayesian network: Plausible segmentation?

<table>
<thead>
<tr>
<th></th>
<th>(t_1)</th>
<th>(t_2)</th>
<th>(t_3)</th>
<th>(t_4)</th>
<th>(t_5)</th>
<th>(t_6)</th>
<th>(t_7)</th>
<th>(t_8)</th>
<th>(t_9)</th>
<th>(t_{10})</th>
</tr>
</thead>
<tbody>
<tr>
<td>(X^{(1)})</td>
<td>(X_{1,1})</td>
<td>(X_{1,2})</td>
<td>(X_{1,3})</td>
<td>(X_{1,4})</td>
<td>(X_{1,5})</td>
<td>(X_{1,6})</td>
<td>(X_{1,7})</td>
<td>(X_{1,8})</td>
<td>(X_{1,9})</td>
<td>(X_{1,10})</td>
</tr>
<tr>
<td>(X^{(2)})</td>
<td>(X_{2,1})</td>
<td>(X_{2,2})</td>
<td>(X_{2,3})</td>
<td>(X_{2,4})</td>
<td>(X_{2,5})</td>
<td>(X_{2,6})</td>
<td>(X_{2,7})</td>
<td>(X_{2,8})</td>
<td>(X_{2,9})</td>
<td>(X_{2,10})</td>
</tr>
<tr>
<td>(X^{(3)})</td>
<td>(X_{3,1})</td>
<td>(X_{3,2})</td>
<td>(X_{3,3})</td>
<td>(X_{3,4})</td>
<td>(X_{3,5})</td>
<td>(X_{3,6})</td>
<td>(X_{3,7})</td>
<td>(X_{3,8})</td>
<td>(X_{3,9})</td>
<td>(X_{3,10})</td>
</tr>
<tr>
<td>(X^{(4)})</td>
<td>(X_{4,1})</td>
<td>(X_{4,2})</td>
<td>(X_{4,3})</td>
<td>(X_{4,4})</td>
<td>(X_{4,5})</td>
<td>(X_{4,6})</td>
<td>(X_{4,7})</td>
<td>(X_{4,8})</td>
<td>(X_{4,9})</td>
<td>(X_{4,10})</td>
</tr>
</tbody>
</table>
Posterior probability

Number of components
Four stages of the *Drosophila* life cycle:

embryo $\rightarrow$ larva $\rightarrow$ pupa $\rightarrow$ adult
Transition Probabilities

time
Morphogenetic transitions: Embryo $\rightarrow$ larva $\rightarrow$ pupa $\rightarrow$ adult

Gene expression program governing the transition to adult morphology active well before the fly emerges from the pupa.
Biological applications

- Raf-Mek-Erk protein signalling pathway
- Morphogenesis in *Drosophila melanogaster*
- Viral challenge and immune activation of macrophages
- Circadian regulation in *Arabidopsis thaliana*
- Ecological networks
Collaboration with DPM (Division of Pathway Medicine, Edinburgh University)
12 hour time course measuring total RNA

30 min sampling

25 samples per group:
- Infection with CMV
- Pre-treatment with IFNγ
- IFNγ + CMV
Posterior probability of the number of components (top) and co-allocation of two time points to the same component (bottom)

Infection  Treatment  Infection+treatment

White=1  Black=0
Literature $\rightarrow$ “Known” interactions between three cytokines: IRF1, IRF2 and IRF3

Evaluation:
Average marginal posterior probabilities of the edges versus non-edges
Sample of high-scoring networks

$P(M|D)$
Sample of high-scoring networks

Feature extraction, e.g. marginal posterior probabilities of the edges

High-confident edge

High-confident non-edge

Uncertainty about edges
IRF1
IRF2
IRF3

Average edge score
Average non-edge score
IRF1
IRF2
IRF3

Average edge score
Average non-edge score
Gold standard known \rightarrow Posterior probabilities of true interactions

New method
Gold standard known → Posterior probabilities of true interactions

(a) CMV

(b) IFNg

(c) CMV + IFNγ

Homogeneous model
Biological applications

- *Raf-Mek-Erk* protein signalling pathway
- Morphogenesis in *Drosophila melanogaster*
- Viral challenge and immune activation of macrophages
- Circadian regulation in *Arabidopsis thaliana*
Circadian rhythms in *Arabidopsis thaliana*

Collaboration with the Institute of Molecular Plant Sciences at Edinburgh University (Andrew Miller’s group)

2 time series $T_{20}$ and $T_{28}$ of microarray gene expression data from *Arabidopsis thaliana*.

- **Focus on**: 9 circadian genes: LHY, CCA1, TOC1, ELF4, ELF3, GI, PRR9, PRR5, and PRR3

- Both **time series** measured under constant light condition at **13 time points**: 0h, 2h, ..., 24h, 26h

- Plants entrained with different **light:dark cycles**
  - 10h:10h ($T_{20}$) and 14h:14h ($T_{28}$)
Gene expression time series plots (Arabidopsis data $T_{20}$ and $T_{28}$)
Posterior probability of the number of components

(a) series $T_{20}$

(b) series $T_{28}$
Posterior probability of the number of components

(a) series $T_{20}$

(b) series $T_{28}$
Posterior Probability of co-assignment of two time points to the same component

Biologically plausible phase shift
Predicted network

Blue – activation
Red – inhibition
Black – mixture

Three different line widths:
- thin = PP>0.5
- medium = PP>0.75
- fat = PP>0.9
Two major gene classes…

Morning genes
e.g. LHY, CCA1
... repress evening genes
e.g. TOC1, ELF3, ELF4, GI, LUX
... which activate LHY and CCA1
Literature vs. inferred network

False negatives
False positives
True positives (TP) = 8
False positives (FP) = 13
False negatives (FN) = 5
True negatives (TN) = 9²-8-13-5= 55

Sensitivity = TP/[TP+FN] = 62%
Specificity = TN/[TN+FP] = 81%
Overview of the plant clock model

Morning

\[
\text{LHY/CCA1} \rightarrow \text{PRR9/PRR7} \rightarrow \text{Y (GI)} \rightarrow \text{TOC1} \rightarrow \text{ZTL}
\]

Evening

Unknown component X allows > 8h delay between TOC1 and LHY/CCA1 expression

Locke et al.
Mol. Syst. Biol. 2006
Methodology

• Mechanistic models
• Bayesian networks
• Integration of biological prior knowledge
• Non-homogeneous Bayesian network for non-stationary processes
• Node-specific change-points and avoiding spurious feedback loops
## Standard dynamic Bayesian network: homogeneous model

<table>
<thead>
<tr>
<th></th>
<th>$t_1$</th>
<th>$t_2$</th>
<th>$t_3$</th>
<th>$t_4$</th>
<th>$t_5$</th>
<th>$t_6$</th>
<th>$t_7$</th>
<th>$t_8$</th>
<th>$t_9$</th>
<th>$t_{10}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X^{(1)}$</td>
<td>$X_{1,1}$</td>
<td>$X_{1,2}$</td>
<td>$X_{1,3}$</td>
<td>$X_{1,4}$</td>
<td>$X_{1,5}$</td>
<td>$X_{1,6}$</td>
<td>$X_{1,7}$</td>
<td>$X_{1,8}$</td>
<td>$X_{1,9}$</td>
<td>$X_{1,10}$</td>
</tr>
<tr>
<td>$X^{(2)}$</td>
<td>$X_{2,1}$</td>
<td>$X_{2,2}$</td>
<td>$X_{2,3}$</td>
<td>$X_{2,4}$</td>
<td>$X_{2,5}$</td>
<td>$X_{2,6}$</td>
<td>$X_{2,7}$</td>
<td>$X_{2,8}$</td>
<td>$X_{2,9}$</td>
<td>$X_{2,10}$</td>
</tr>
<tr>
<td>$X^{(3)}$</td>
<td>$X_{3,1}$</td>
<td>$X_{3,2}$</td>
<td>$X_{3,3}$</td>
<td>$X_{3,4}$</td>
<td>$X_{3,5}$</td>
<td>$X_{3,6}$</td>
<td>$X_{3,7}$</td>
<td>$X_{3,8}$</td>
<td>$X_{3,9}$</td>
<td>$X_{3,10}$</td>
</tr>
<tr>
<td>$X^{(4)}$</td>
<td>$X_{4,1}$</td>
<td>$X_{4,2}$</td>
<td>$X_{4,3}$</td>
<td>$X_{4,4}$</td>
<td>$X_{4,5}$</td>
<td>$X_{4,6}$</td>
<td>$X_{4,7}$</td>
<td>$X_{4,8}$</td>
<td>$X_{4,9}$</td>
<td>$X_{4,10}$</td>
</tr>
</tbody>
</table>
### Heterogeneous dynamic Bayesian network

<table>
<thead>
<tr>
<th>(X^{(1)})</th>
<th>(t_1)</th>
<th>(t_2)</th>
<th>(t_3)</th>
<th>(t_4)</th>
<th>(t_5)</th>
<th>(t_6)</th>
<th>(t_7)</th>
<th>(t_8)</th>
<th>(t_9)</th>
<th>(t_{10})</th>
</tr>
</thead>
<tbody>
<tr>
<td>(X_{1,1})</td>
<td>(X_{1,2})</td>
<td>(X_{1,3})</td>
<td>(X_{1,4})</td>
<td>(X_{1,5})</td>
<td>(X_{1,6})</td>
<td>(X_{1,7})</td>
<td>(X_{1,8})</td>
<td>(X_{1,9})</td>
<td>(X_{1,10})</td>
<td></td>
</tr>
<tr>
<td>(X^{(2)})</td>
<td>(X_{2,1})</td>
<td>(X_{2,2})</td>
<td>(X_{2,3})</td>
<td>(X_{2,4})</td>
<td>(X_{2,5})</td>
<td>(X_{2,6})</td>
<td>(X_{2,7})</td>
<td>(X_{2,8})</td>
<td>(X_{2,9})</td>
<td>(X_{2,10})</td>
</tr>
<tr>
<td>(X^{(3)})</td>
<td>(X_{3,1})</td>
<td>(X_{3,2})</td>
<td>(X_{3,3})</td>
<td>(X_{3,4})</td>
<td>(X_{3,5})</td>
<td>(X_{3,6})</td>
<td>(X_{3,7})</td>
<td>(X_{3,8})</td>
<td>(X_{3,9})</td>
<td>(X_{3,10})</td>
</tr>
<tr>
<td>(X^{(4)})</td>
<td>(X_{4,1})</td>
<td>(X_{4,2})</td>
<td>(X_{4,3})</td>
<td>(X_{4,4})</td>
<td>(X_{4,5})</td>
<td>(X_{4,6})</td>
<td>(X_{4,7})</td>
<td>(X_{4,8})</td>
<td>(X_{4,9})</td>
<td>(X_{4,10})</td>
</tr>
</tbody>
</table>
Heterogenous dynamic Bayesian network with node-specific breakpoints

<table>
<thead>
<tr>
<th></th>
<th>$t_1$</th>
<th>$t_2$</th>
<th>$t_3$</th>
<th>$t_4$</th>
<th>$t_5$</th>
<th>$t_6$</th>
<th>$t_7$</th>
<th>$t_8$</th>
<th>$t_9$</th>
<th>$t_{10}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X^{(1)}$</td>
<td>$X_{1,1}$</td>
<td>$X_{1,2}$</td>
<td>$X_{1,3}$</td>
<td>$X_{1,4}$</td>
<td>$X_{1,5}$</td>
<td>$X_{1,6}$</td>
<td>$X_{1,7}$</td>
<td>$X_{1,8}$</td>
<td>$X_{1,9}$</td>
<td>$X_{1,10}$</td>
</tr>
<tr>
<td>$X^{(2)}$</td>
<td>$X_{2,1}$</td>
<td>$X_{2,2}$</td>
<td>$X_{2,3}$</td>
<td>$X_{2,4}$</td>
<td>$X_{2,5}$</td>
<td>$X_{2,6}$</td>
<td>$X_{2,7}$</td>
<td>$X_{2,8}$</td>
<td>$X_{2,9}$</td>
<td>$X_{2,10}$</td>
</tr>
<tr>
<td>$X^{(3)}$</td>
<td>$X_{3,1}$</td>
<td>$X_{3,2}$</td>
<td>$X_{3,3}$</td>
<td>$X_{3,4}$</td>
<td>$X_{3,5}$</td>
<td>$X_{3,6}$</td>
<td>$X_{3,7}$</td>
<td>$X_{3,8}$</td>
<td>$X_{3,9}$</td>
<td>$X_{3,10}$</td>
</tr>
<tr>
<td>$X^{(4)}$</td>
<td>$X_{4,1}$</td>
<td>$X_{4,2}$</td>
<td>$X_{4,3}$</td>
<td>$X_{4,4}$</td>
<td>$X_{4,5}$</td>
<td>$X_{4,6}$</td>
<td>$X_{4,7}$</td>
<td>$X_{4,8}$</td>
<td>$X_{4,9}$</td>
<td>$X_{4,10}$</td>
</tr>
</tbody>
</table>
The dynamic BGe network

\[ P(\mathcal{D}|\mathcal{G}, \theta) = \prod_{n=1}^{N} \prod_{t=2}^{m} P\left(X_n(t) = \mathcal{D}_{n,t} | \pi_n(t-1) = \mathcal{D}_{(\pi_n,t-1)}, \theta_n \right) \]

\[ P(\mathcal{D}|\mathcal{G}) = \int P(\mathcal{D}|\mathcal{G}, \theta) P(\theta|\mathcal{G}) d\theta = \prod_{n=1}^{N} \Psi(\mathcal{D}_{\pi_n}^{\pi_n}, \mathcal{G}) \]

\[ \Psi(\mathcal{D}_{\pi_n}^{\pi_n}, \mathcal{G}) = \int \prod_{t=2}^{m} P\left(X_n(t) = \mathcal{D}_{n,t} | \pi_n(t-1) = \mathcal{D}_{(\pi_n,t-1)}, \theta_n \right) P(\theta_n|\mathcal{G}) d\theta_n \]

The non-stationary dynamic change-point BGe model

\[ P(\mathcal{D}|\mathcal{G}, \mathbf{V}, \mathbf{K}, \theta) = \prod_{n=1}^{N} \prod_{t=2}^{m} \prod_{k=1}^{\mathcal{K}_n} P\left(X_n(t) = \mathcal{D}_{n,t} | \pi_n(t-1) = \mathcal{D}_{(\pi_n,t-1)}, \theta_n^k \right) \delta_{V_n(t), k} \]

\[ P(\mathcal{D}|\mathcal{G}, \mathbf{V}, \mathbf{K}) = \int P(\mathcal{D}|\mathcal{G}, \mathbf{V}, \mathbf{K}, \theta) P(\theta) d\theta = \prod_{n=1}^{N} \prod_{k=1}^{\mathcal{K}_n} \Psi(\mathcal{D}_{\pi_n}^{\pi_n}[k, \mathbf{V}_n], \mathcal{G}) \]

\[ \Psi(\mathcal{D}_{\pi_n}^{\pi_n}[k, \mathbf{V}_n], \mathcal{G}) = \int \prod_{t=2}^{m} P\left(X_n(t) = \mathcal{D}_{n,t} | \pi_n(t-1) = \mathcal{D}_{(\pi_n,t-1)}, \theta_n^k \right) \delta_{V_n(t), k} P(\theta_n^k|\mathcal{G}) d\theta_n^k \]
MCMC scheme

\[
P(\mathcal{G}, \mathcal{V}, \mathcal{K}, \mathcal{D}) = P(\mathcal{D}|\mathcal{G}, \mathcal{V}, \mathcal{K}) \cdot P(\mathcal{G}) \cdot P(\mathcal{V} | \mathcal{K}) \cdot P(\mathcal{K})
\]

\[
= P(\mathcal{G}) \cdot \prod_{n=1}^{N} \left\{ P(\mathcal{V}_n | \mathcal{K}_n) \cdot P(\mathcal{K}_n) \cdot \prod_{k=1}^{\mathcal{K}_n} \Psi(\mathcal{D}_n^{\pi_n}[k, \mathcal{V}_n], \mathcal{G}) \right\}
\]

**Moves that change the network structure**

\[
A(\mathcal{G}^{i+1}|\mathcal{G}^{i}) = \min \left\{ 1, \frac{P(\mathcal{D}|\mathcal{G}^{i+1}, \mathcal{V}^{i}, \mathcal{K}^{i}) \cdot P(\mathcal{G}^{i+1})}{P(\mathcal{D}|\mathcal{G}^{i}, \mathcal{V}^{i}, \mathcal{K}^{i}) \cdot P(\mathcal{G}^{i})} \frac{|\mathcal{N}(\mathcal{G}^{i})|}{|\mathcal{N}(\mathcal{G}^{i+1})|} \right\}
\]

**Moves that act on change-points: reallocation, birth and death moves**

\[
(\mathcal{K}^{i}_n, \mathcal{V}^{i}_n) \rightarrow (\mathcal{K}^{i+1}_n, \mathcal{V}^{i+1}_n)
\]

\[
R = \frac{\prod_{k=1}^{\mathcal{K}^{i+1}_n} \Psi(\mathcal{D}_n^{\pi_n}[k, \mathcal{V}^{i+1}_n], \mathcal{G})}{\prod_{k=1}^{\mathcal{K}^{i}_n} \Psi(\mathcal{D}_n^{\pi_n}[k, \mathcal{V}^{i}_n], \mathcal{G})} \times A \times B
\]

\[
A = \frac{P(\mathcal{V}_n^{i+1}|\mathcal{K}^{i+1}_n) \cdot P(\mathcal{K}^{i+1}_n)}{P(\mathcal{V}_n^{i}|\mathcal{K}^{i}_n) \cdot P(\mathcal{K}^{i}_n)}
\]

\[B\] is the inverse proposal probability ratio.
Avoiding spurious feedback loops

\[ X(t + 1) = X(t) + c + \sigma_x \cdot \phi_X(t) \]
\[ Y(t + 1) = f(X(t)) + \sigma_y \cdot \phi_Y(t) \]

\[ \phi(.) \] iid normally distributed random variables
Avoiding spurious feedback loops

Marginal posterior probability for edges

BGe

New model
BGe

New model

<table>
<thead>
<tr>
<th>$\sigma_y$</th>
<th>$\sigma_y=0.1$</th>
<th>$\sigma_y=0.25$</th>
<th>$\sigma_y=0.5$</th>
<th>$\sigma_y=1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma_x$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>0.25</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>0.5</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$\sigma_x$</th>
<th>$\sigma_x=0.1$</th>
<th>$\sigma_x=0.25$</th>
<th>$\sigma_x=0.5$</th>
<th>$\sigma_x=1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma_y$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>0.25</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>0.5</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

The diagram compares the BGe and New model for different values of $\sigma_x$ and $\sigma_y$. The x-axis and y-axis represent the varying parameters, while the bars indicate the model's performance or some metric of interest.
Application to macrophage gene expression data

\[ \xi = \frac{1}{N_{sl}} \sum_{sl=1}^{N_{sl}} P(e_{sl}|D) - \frac{1}{N_{nl}} \sum_{nl=1}^{N_{nl}} P(e_{nl}|D) \]

\( e_{sl} \) is an edge corresponding to a self-loop

\( e_{nl} \) is an edge corresponding to a non-self-loop

(a) CMV

(b) IFN\( \gamma \)

(c) CMV + IFN\( \gamma \)
Comparison with related methods

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Score</td>
<td>Marginal Likelihood</td>
<td>Marginal Likelihood</td>
<td>Marginal Likelihood</td>
<td>Marginal Likelihood</td>
<td>BIC</td>
</tr>
<tr>
<td>Node-specific change-points</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Structure constant</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Data format</td>
<td>Continuous</td>
<td>Discrete</td>
<td>Continuous</td>
<td>Continuous</td>
<td>Continuous</td>
</tr>
<tr>
<td>Latent variable format</td>
<td>Change-point process</td>
<td>Change-point process</td>
<td>Change-point process</td>
<td>Free allocation</td>
<td>Free allocation</td>
</tr>
<tr>
<td>Inference</td>
<td>MCMC</td>
<td>MCMC</td>
<td>MCMC</td>
<td>MCMC</td>
<td>EM</td>
</tr>
</tbody>
</table>
Networks for simulating data
Generating synthetic data

\[ X(t + 1) = \phi_X(t); \quad Y(t + 1) = \phi_Y(t) \]

\[ W(t + 1) = W(t) + \frac{2\pi}{m} + c_W \cdot \phi_W(t) \]

\[ Z(t + 1) = c_X \cdot X(t) + c_Y \cdot Y(t) + \sin(W(t)) + c_Z \cdot \phi_Z(t + 1) \]
Application to *Arabidopsis thaliana*

<table>
<thead>
<tr>
<th>Source</th>
<th>Segment 1</th>
<th>Segment 2</th>
<th>Segment 3</th>
<th>Segment 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time points</td>
<td>12</td>
<td>13</td>
<td>13</td>
<td>13</td>
</tr>
<tr>
<td>Time interval</td>
<td>4h</td>
<td>4h</td>
<td>2h</td>
<td>2h</td>
</tr>
<tr>
<td>Pretreatment</td>
<td>12h-light 12h-dark cycle</td>
<td>12h-light 12h-dark cycle</td>
<td>10h-light 10h-dark cycle</td>
<td>14h-light 14h-dark cycle</td>
</tr>
<tr>
<td>Measurement</td>
<td>Constant light</td>
<td>Constant light</td>
<td>Constant light</td>
<td>Constant light</td>
</tr>
<tr>
<td>Laboratory</td>
<td>Kay Lab</td>
<td>Miller Lab</td>
<td>Miller Lab</td>
<td>Miller Lab</td>
</tr>
</tbody>
</table>

Table 2: Overview of the gene expression time series segments for Arabidopsis.
Summary

- Mechanistic models
- Bayesian networks
- Integration of biological prior knowledge
- Non-homogeneous Bayesian network for non-stationary processes
- Avoiding spurious feedback
Acknowledgements

Adriano Werhli

Marco Grzegorczyk
Thank you!

Any questions?