# Parameter Identification and Model Ranking of Thomas Networks

CMSB 2012
The Royal Society, London


**David Šafránek**

Adam Streck, Juraj Kolčák
*Masaryk University Brno*

Hannes Klarner, Heike Siebert
*Freie Universität Berlin*


5th October, 2012

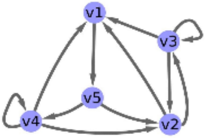
# Outline

1. Motivation and Background
2. Proposed Methodology and its Implementation
3. Performance Evaluation and Case Studies

# Outline

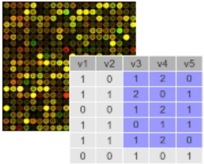
# Motivation: Learn More about Regulatory Networks

**Gene Regulatory Network**




Predicted structure
(databases, literature, ...)

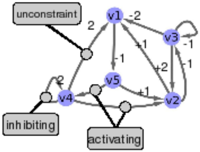
**Observations**



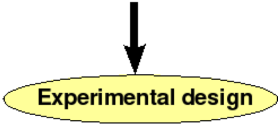
| v1 | v2 | v3 | v4 | v5 |
|---|---|---|---|---|
| 1 | 0 | 1 | 2 | 0 |
| 1 | 1 | 2 | 0 | 1 |
| 0 | 0 | 1 | 2 | 1 |
| 1 | 1 | 0 | 1 | 1 |
| 1 | 1 | 1 | 2 | 0 |
| 0 | 0 | 1 | 0 | 1 |

Discrete time series

**Hypothesis**

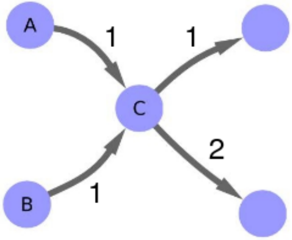



What kind of interactions?

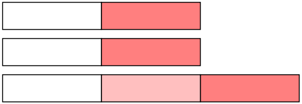


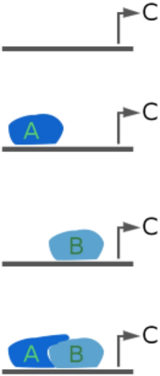

Modeling tools: C. Chaouiya, et al. 2003, GINsim., H. de Jong et al. 2002, GNA.
Data processing: I. Shmulevich, et al. 2002. Binary analysis and optimization-based normalization of gene expression data.; E. Dimitrova, et al. 2010. Discretization of time series data.

# From Structure to Dynamics




A:
B:
C:






$$\emptyset \rightarrow 0$$

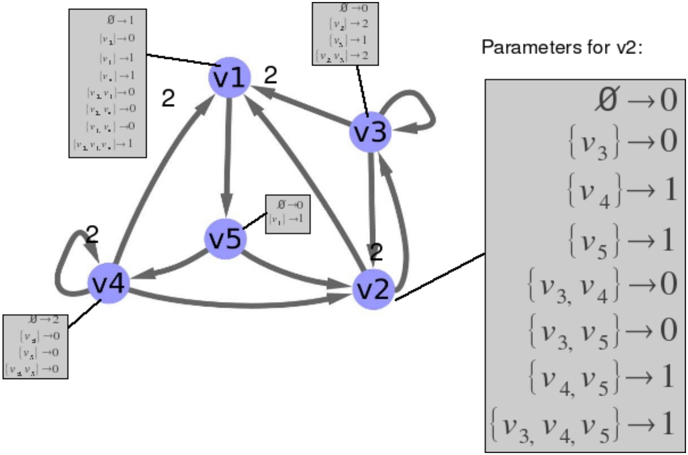
$$\{A\} \rightarrow 2$$

$$\{B\} \rightarrow 2$$

$$\{A, B\} \rightarrow 1$$

R. Thomas and R. d’Ari, CRC Press 1990. Biological feedback.

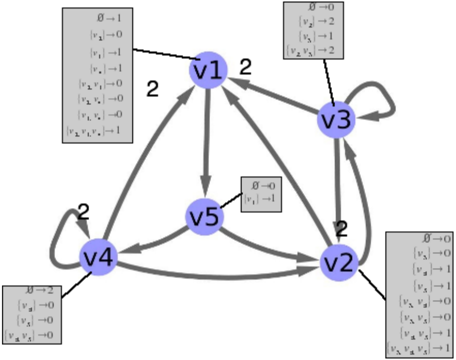
# Parameterization of Regulatory Networks

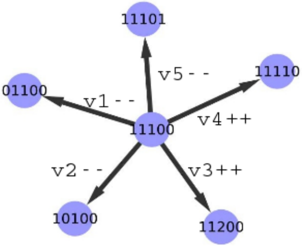



Target values assigned to regulatory contexts for all nodes make a **PARAMETER SET (parameterization).**

R. Thomas and R. d'Ari, CRC Press 1990. Biological feedback.

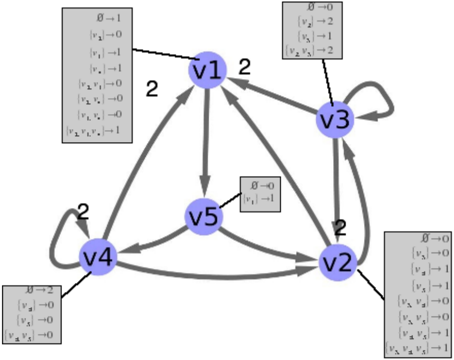
# Dynamics as a State Transition Graph

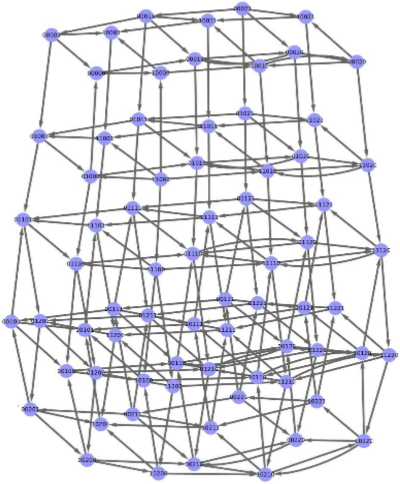



R. Thomas and R. d'Ari, CRC Press 1990. Biological feedback.

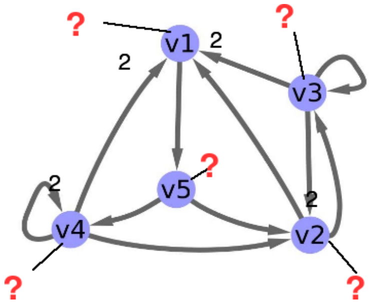
# Dynamics as a State Transition Graph



R. Thomas and R. d'Ari, CRC Press 1990. Biological feedback.

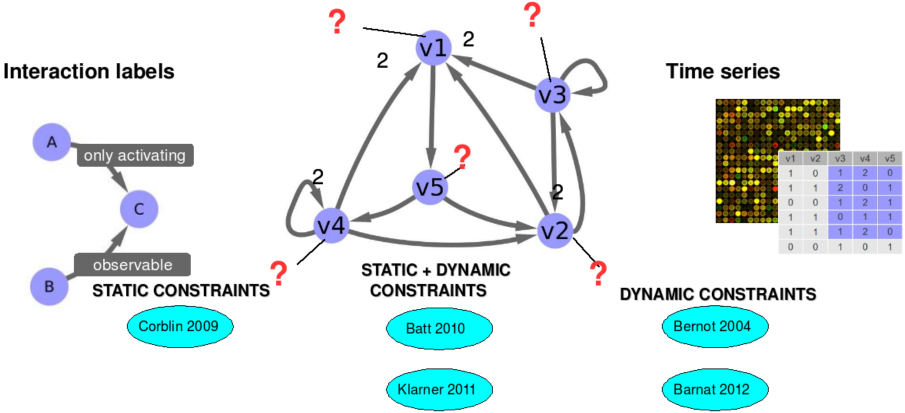
# Parameter Identification Problem




Number of possible parameterizations of a single node is **exponential** w.r.t. the node's in-degree.

(more precisely w.r.t. the number of regulatory contexts)

# Parameter Identification Problem: Solutions




| v1 | v2 | v3 | v4 | v5 |
|---|---|---|---|---|
| 1 | 0 | 1 | 2 | 0 |
| 1 | 1 | 2 | 0 | 1 |
| 0 | 0 | 1 | 2 | 1 |
| 1 | 1 | 0 | 1 | 1 |
| 1 | 1 | 1 | 2 | 0 |
| 0 | 0 | 1 | 0 | 1 |


G. Bernot et al. in JTB 2004: Application of formal methods to biological regulatory networks: extending Thomas asynchronous logical approach with temporal logic.

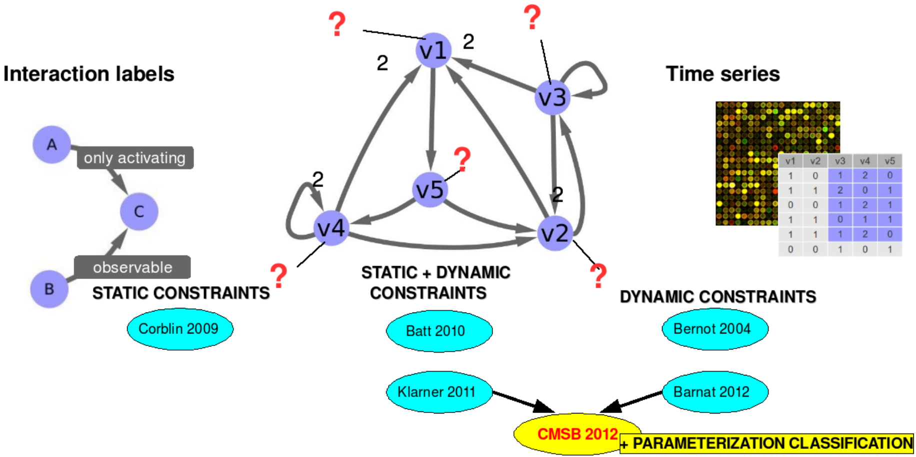
Corblin et al. in BioSystems 2009: A declarative constraint-based method for analyzing discrete gene regulation networks.

Batt et al. in Bioinf. 2010: Efficient parameter search for qualitative models of regulatory networks using symbolic model checking.

Klarner et al. in CMSB 2011: Parameter inference for asynchronous logical networks using discrete time series.

Barnat et al. in IEEE/ACM TCBB 2012: On Parameter Synthesis by Parallel Model Checking.

# Parameter Identification Problem: Solutions




G. Bernot et al. in JTB 2004: Application of formal methods to biological regulatory networks: extending Thomas asynchronous logical approach with temporal logic.

Corblin et al. in BioSystems 2009: A declarative constraint-based method for analyzing discrete gene regulation networks.

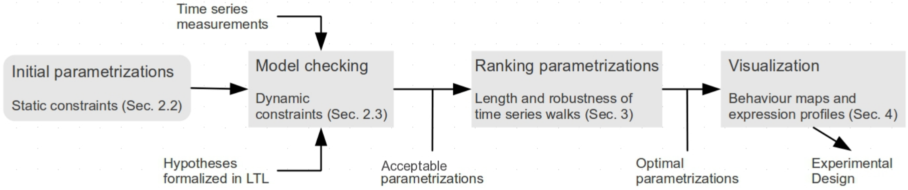
Batt et al. in Bioinf. 2010: Efficient parameter search for qualitative models of regulatory networks using symbolic model checking.

Klarner et al. in CMSB 2011: Parameter inference for asynchronous logical networks using discrete time series.

Barnat et al. in IEEE/ACM TCBB 2012: On Parameter Synthesis by Parallel Model Checking.

# Outline

# Our Contribution




- a prototype tool chain:

  Parsybone – `https://github.com/sybila/Parsybone.git`

  ParameterFilter – `https://github.com/sybila/ParameterFilter.git`

- distributed computation of acceptable parameterizations
- employing witnesses (counterexamples) to rank obtained parameterizations
- visualization of the results (export to Cytoscape)

# Time-series Measurement as a Dynamic Constraint

**Time-series measurement**

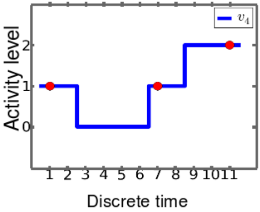
| v1 | v2 | v3 | v4 | v5 |
|---|---|---|---|---|
| 1 | 1 | 1 | 1 | 1 |
| 1 | 0 | 1 | 1 | 0 |
| 1 | 1 | 2 | 2 | 1 |

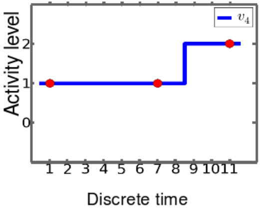
Encoded in LTL:

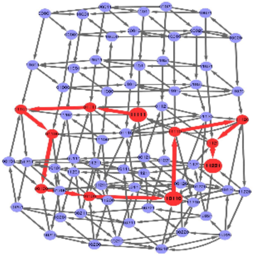
$\sigma(1) = \bigwedge_{i=1}^{5} v_i = 1$
$\sigma(2) = \bigwedge_{i\in\{1,2,4\}} v_i = 1 \wedge \bigwedge_{i\in\{2,5\}} v_i = 0$
$\sigma(3) = \bigwedge_{i\in\{1,2,5\}} v_i = 1 \wedge \bigwedge_{i\in\{3,4\}} v_i = 2$

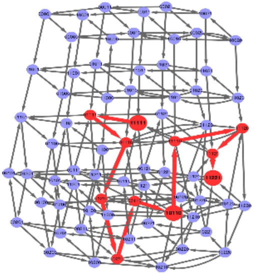
$\varphi = \sigma(1) \wedge \mathbf{F}(\sigma(2) \wedge \mathbf{F}(\sigma(3)))$

**Expression of *v4* along red path**





# Time-series Measurement as a Dynamic Constraint

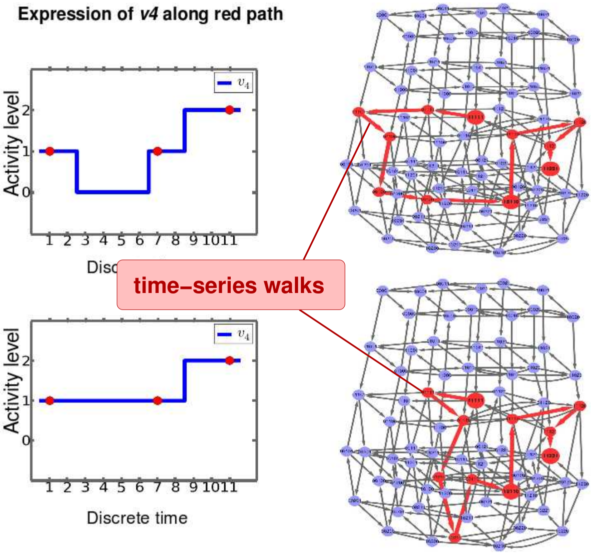
**Time-series measurement**

| v1 | v2 | v3 | v4 | v5 |
|---|---|---|---|---|
| 1 | 1 | 1 | 1 | 1 |
| 1 | 0 | 1 | 1 | 0 |
| 1 | 1 | 2 | 2 | 1 |

Encoded in LTL:

$\sigma(1) = \bigwedge_{i=1}^{5} v_i = 1$
$\sigma(2) = \bigwedge_{i \in \{1,2,4\}} v_i = 1 \wedge \bigwedge_{i \in \{2,5\}} v_i = 0$
$\sigma(3) = \bigwedge_{i \in \{1,2,5\}} v_i = 1 \wedge \bigwedge_{i \in \{3,4\}} v_i = 2$

$\varphi = \sigma(1) \wedge \mathbf{F}(\sigma(2) \wedge \mathbf{F}(\sigma(3)))$



# Time-series Measurement as a Dynamic Constraint

**Time-series measurement**

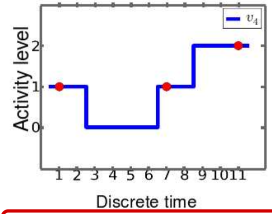
| v1 | v2 | v3 | v4 | v5 |
|---|---|---|---|---|
| 1 | 1 | 1 | 1 | 1 |
| 1 | 0 | 1 | 1 | 0 |
| 1 | 1 | 2 | 2 | 1 |

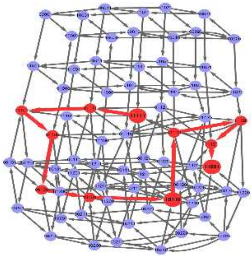
Encoded in LTL:

$$\sigma(1) = \bigwedge_{i=1}^{5} v_i = 1$$
$$\sigma(2) = \bigwedge_{i\in\{1,2,4\}} v_i = 1 \wedge \bigwedge_{i\in\{2,5\}} v_i = 0$$
$$\sigma(3) = \bigwedge_{i\in\{1,2,5\}} v_i = 1 \wedge \bigwedge_{i\in\{3,4\}} v_i = 2$$

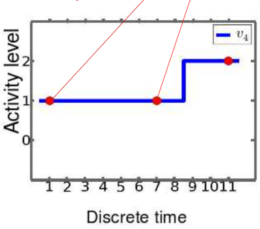
$$\varphi = \sigma(1) \wedge (\sigma(1)\mathbf{U}(\sigma(2) \wedge \mathbf{F}(\sigma(3))))$$

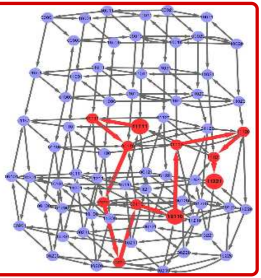
**Expression of *v4* along red path**



monotonicity between 1st and 2nd measurement



# Time-series Measurement as a Dynamic Constraint

**Time-series measurement**

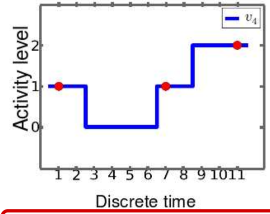
| v1 | v2 | v3 | v4 | v5 |
|---|---|---|---|---|
| **?** | 1 | 1 | 1 | **?** |
| 1 | 0 | 1 | 1 | 0 |
| 1 | 1 | 2 | 2 | 1 |

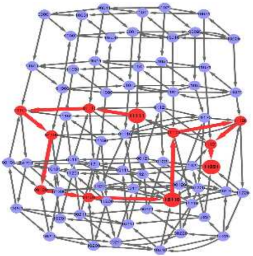
Encoded in LTL:

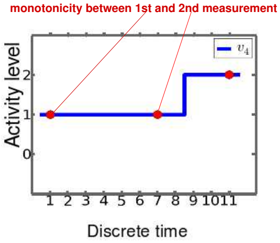
$$\sigma(1) = \bigwedge_{i=2}^{4} v_i = 1$$
$$\sigma(2) = \bigwedge_{i\in\{1,2,4\}} v_i = 1 \wedge \bigwedge_{i\in\{2,5\}} v_i = 0$$
$$\sigma(3) = \bigwedge_{i\in\{1,2,5\}} v_i = 1 \wedge \bigwedge_{i\in\{3,4\}} v_i = 2$$

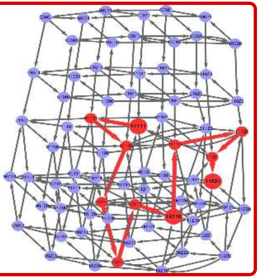
$$\varphi = \sigma(1) \wedge (\sigma(1)\mathbf{U}(\sigma(2) \wedge \mathbf{F}(\sigma(3))))$$

**Expression of *v4* along red path**

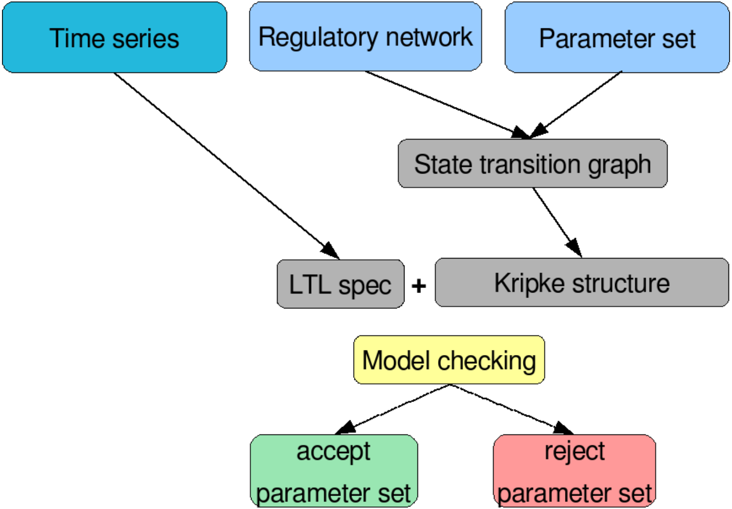




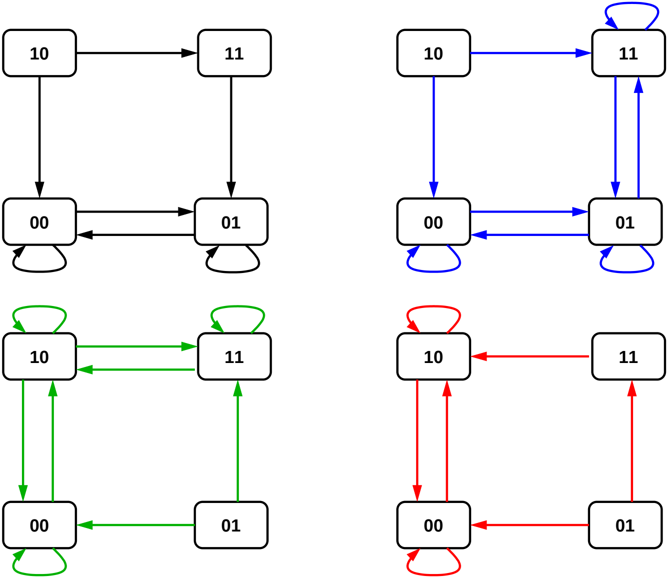



# Identifying Parameters by Model Checking

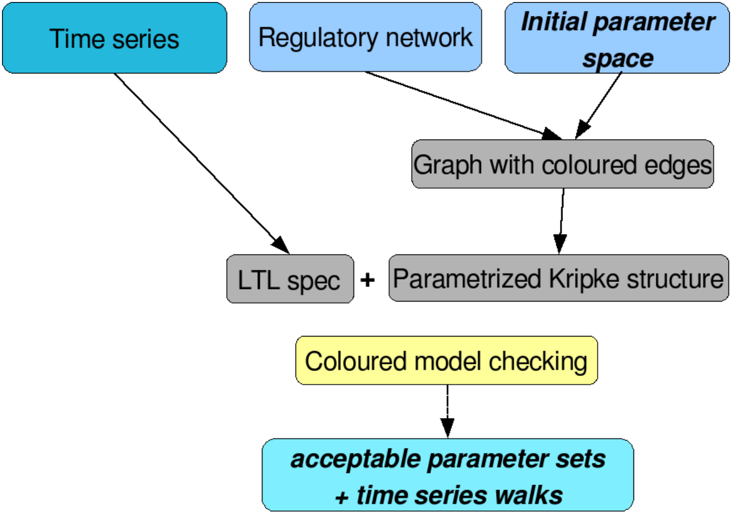



Naive approaches: G. Bernot et al. in JTB 2004; H. Klarner et al. in CMSB 2011

# Effect of Parameters on State Transition Graph



# Identifying Parameters by Coloured Model Checking

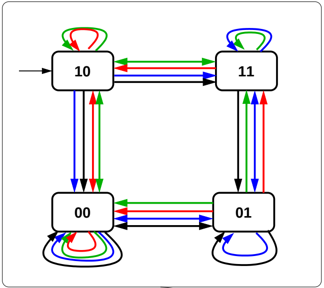


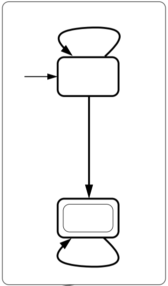

Heuristics: J. Barnat et al. in TCBB 2012; refined in the presented paper

# Identifying Parameters by Coloured Model Checking

**parameterized Kripke structure of the model**

**never claim Buchi automaton**



**[A=1,B=0] & F([A=0,B=0])**

**return accepting paths of the product automaton**
*respective parameter sets are acceptable*
**we decide on all parameterizations at once**

**parameter sets acceptable for the dynamic constraint**

**time−series walks of acceptable parametrizations**

# Model Checking on Coloured Graphs

## Idea

- represent each parameterization by a distinct colour
- find accepting cycles and get colours enabling accepting paths

## Procedure

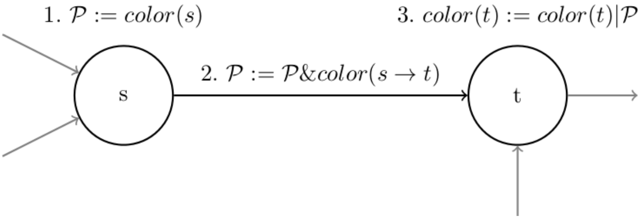
1. compute initial mapping of colours to states
   $\Rightarrow$ propagate colours through the entire graph (BFS reachability)
   $\Rightarrow$ accepting states know all colours by which they are reached
2. for each reachable accepting cycle aggregate the valid colours

# Model Checking on Coloured Graphs

## Implementation

- explicit representation of indexed parameter sets (ordered bit vectors)
- parameter space split to exclusive blocks equal to size of integer type
- each block contains “close” parameter sets
- data-parallel distribution: blocks evenly distributed over the cluster

| … | $P_{i-1}$ | $P_i$ | $P_{i+1}$ | … |
|---|---|---|---|---|
| | | ⋮ | | |
| … | 1 | 0 | 0 | … |
| … | 0 | 1 | 0 | … |
| … | 0 | 0 | 1 | … |
| | | ⋮ | | |



# Parameterization Ranking: Length Cost

- theoretically infinitely many time-series walks
- fix a dynamic constraint and focus on compatible **shortest walks**
  - penalize unnecessarily higher energy cost
  - avoid complex model realizations of the constraint
- assign each parameterization its **length cost** – the length of a shortest time-series walk
- consider parameterizations with minimum length cost

# Parameterization Ranking: Robustness

- non-deterministic dynamics caused by asynchronicity
- how can we interpret walks with less options to walk off the "optimal path" and miss the expected final state of the time-series?
- the property of the model, but...
  - another classification of parameterizations

- **local robustness**:
  property of a state – $\frac{\text{number of valid successors}}{\text{out degree}}$
- **global robustness**:
  property of a walk – product of local robustness over all states of the walk
- **model robustness**:
  property of a parameterization – average of global robustness over all time-series walks

# Parameterization Ranking: Robustness

- non-deterministic dynamics caused by asynchronicity
- how can we interpret walks with less options to walk off the "optimal path" and miss the expected final state of the time-series?
- the property of the model, but...
  - another classification of parameterizations

- **local robustness – approximated**:

$$\mathrm{Prob}(x) = \frac{1}{\mathrm{out\_degree}(x)}$$

- **global robustness**:

  property of a walk – product of local robustness over all states of the walk

- **model robustness**:

  property of a parameterization – average of global robustness over all time-series walks

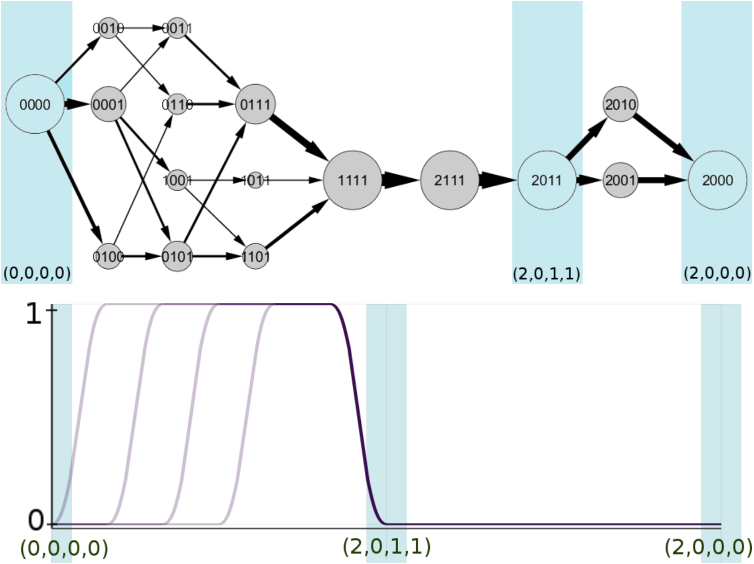
# Parameterization Ranking: Overall Procedure

**INPUT:** regulatory network, initial parameter space, static and dynamic constraints
**OUTPUT:** subset of the initial parameter space containing optimal parameterizations

1. Remove parametrizations violating static constraints
2. Compute parameterizations acceptable by dynamic constraints
3. Select parametrizations with minimal length cost
4. Select parametrizations with maximal robustness

# Visualising Results

## Behaviour Maps and Expression Profiles

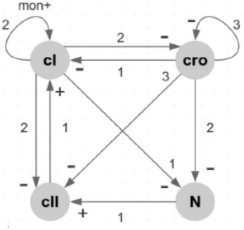



projection to the 2nd component (gene)

# Outline

# Scalability Evaluation

GRN of Bacteriophage $\lambda$




Lysogenic time-series

| cI | cII | cro | N |
|---|---|---|---|
| 0 | 0 | 0 | 0 |
| 2 | 1 | 0 | 1 |
| 2 | 0 | 0 | 0 |

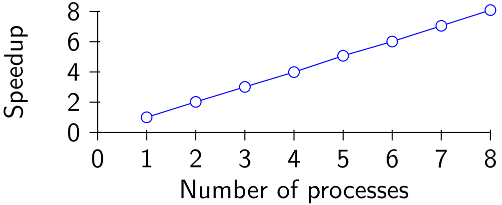
[Thieffry et al. 1995]

Lytic time-series

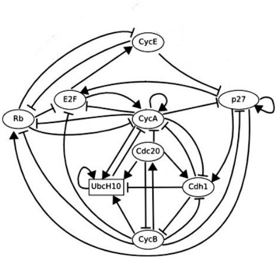
| cI | cII | cro | N |
|---|---|---|---|
| 0 | 0 | 0 | 0 |
| 0 | 0 | 2 | 1 |
| 0 | 0 | 2 | 0 |
| 0 | 0 | 3 | 0 |
| 0 | 0 | 2 | 0 |

- conjuction of both time-series lead to 537 parametrizations
- required memory: $\leq$ 3MB

| Process count | Average runtime |
|---|---|
| 1 | 5.315 s |
| 2 | 2.634 s |
| 3 | 1.767 s |
| 4 | 1.332 s |
| 5 | 1.048 s |
| 6 | 0.884 s |
| 7 | 0.754 s |
| 8 | 0.657 s |



# Performance Evaluation



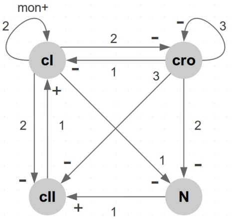
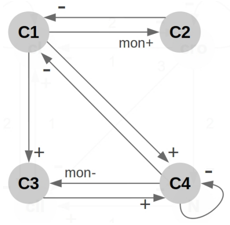

GRN of Mammalian Cell Cycle

[Fauré et al. 2006]

- a time-series with 8 measurements
- $6.8 \cdot 10^8$ initial parametrizations
- $3.1 \cdot 10^8$ acceptable parametrizations computed
- setup: 8 processes running on 2 CPUs with 4 cores each
- required memory: $\leq$ 15MB

| Process ID | Runtime | Result set size | Process ID | Runtime | Result set size |
|---|---|---|---|---|---|
| 1 | 29.07 h | 38,522,403 | 5 | 29.70 h | 38,523,691 |
| 2 | 31.08 h | 38,521,943 | 6 | 28.81 h | 38,523,255 |
| 3 | 27.22 h | 38,521,656 | 7 | 29.55 h | 38,522,328 |
| 4 | 32.32 h | 38,522,343 | 8 | 28.83 h | 38,523,020 |

# Case Studies

| | Bacteriophage $\lambda$[1] | Rat neural system[2] |
|---|---|---|
| |  |  |
| | [Thieffry et al. 1995] | [Wahde et al. 2001] |
| Init. Parameter Space | $6.9 \cdot 10^9$ | $2.6 \cdot 10^5$ |
| Static Constraints | $8.2 \cdot 10^4$ | 162 |
| Dynamic Constraints | 537 | 108 |
| Length Cost (min) | 28 (length 9) | 108 (length 5) |
| Robustness (max) | 3 (9.7%) | 4 (75%) |

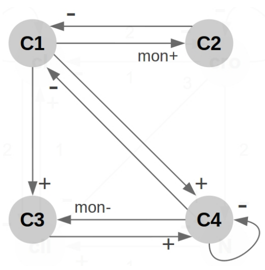




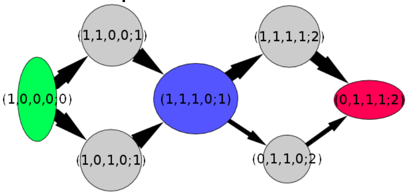
[1] CMSB 2012 Proceedings
[2] FI MU Technical Report

# Rat Neural System: Inferring New Hypothesis

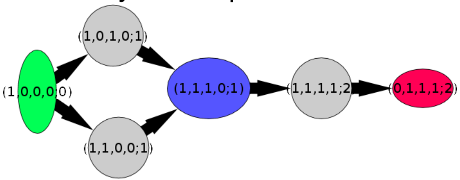
[Wan 1998, Wahde 2001]




Shortest paths




Maximally robust paths




## Predicted Hypothesis

Genes in cluster 4 express before the cluster 1 expression starts to degrade.

# Conclusions

## Achievements

- computational improvement in model checking-based parameter identification for Thomas networks
- ranking procedures for distinguishing the models

## Future work

- *vast amount of data generated...*what to do next?
  - more sophisticated model ranking (biologically relevant criteria)
  - finding commonalities in models, e.g., for refining static constraints (CMSB 2011)

Thank you for your attention!