Formal methods for capturing dynamics of biological networks

Loïc Paulevé

CNRS/LRI, Univ. Paris-Sud, Univ. Paris-Saclay – BioInfo team
`loic.pauleve@lri.fr`
`http://loicpauleve.name`

MEMICS, 15 October 2017, Telč, Czech Republic

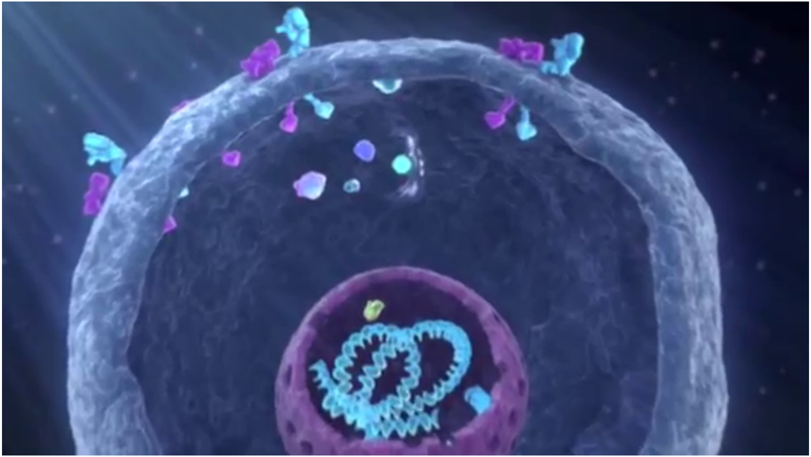
Outline

1. Computational models for biological processes
    - Networks, dynamics
    - Boolean networks
2. Computational challenges
    - Inference
    - Validation
    - Control
3. Static analysis of Boolean networks by abstract interpretation
    - Main principle
    - Over- and under-approximation of trajectories
    - Applications
4. Perspectives

Outline

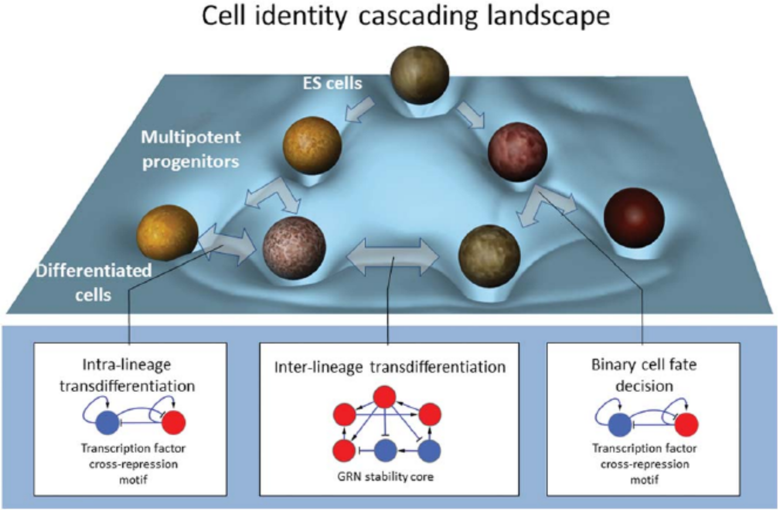
Biological processes

Cell division



Biological Processes

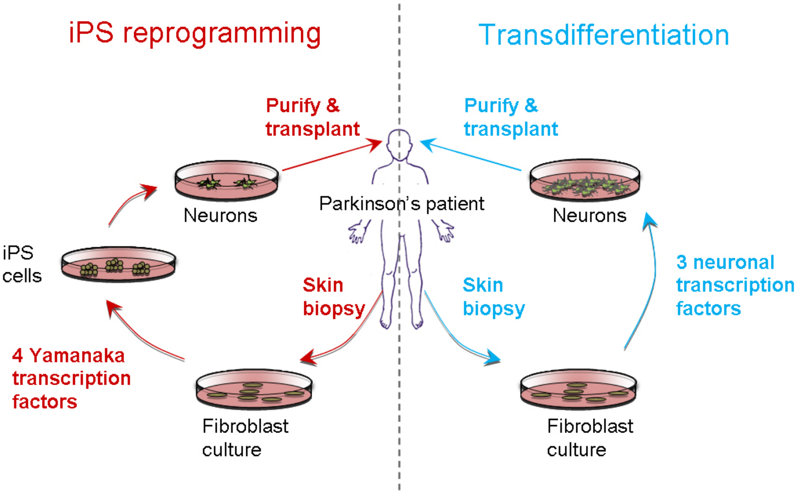
Cell differentiation



(source: Crespo et al. Stem cells 2013; 31:2127-2135)

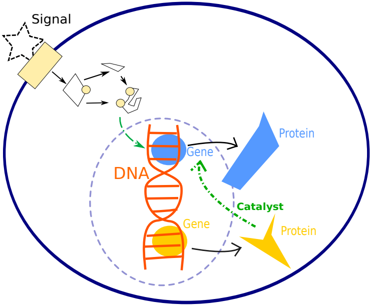
Biological Processes

Cell reprogramming



(credits: Thomas Graf, Centre for Genomic Regulation (Spain))

Biological networks



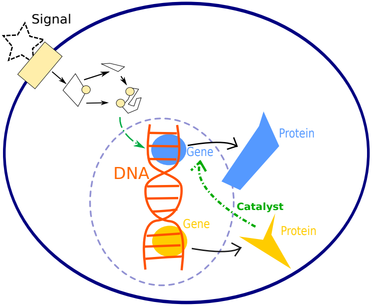
Prediction

- Cell response w.r.t. signal+environment
- Long-term behaviours (differentiation)

Control

- Mutations/Perturbations for modifying cell behaviour
- Trans/De-differentiation

Biological networks



Prediction

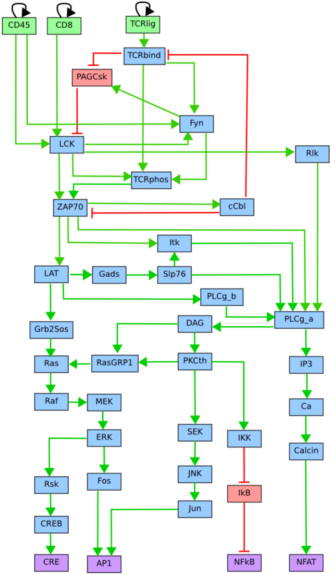
- Cell response w.r.t. signal+environment
- Long-term behaviours (differentiation)

Control

- Mutations/Perturbations for modifying cell behaviour
- Trans/De-differentiation

$$\Rightarrow \begin{cases} \textbf{Computational models of dynamics} \\ -\text{Formal verification} \\ -\text{Automatic reasoning} \end{cases}$$

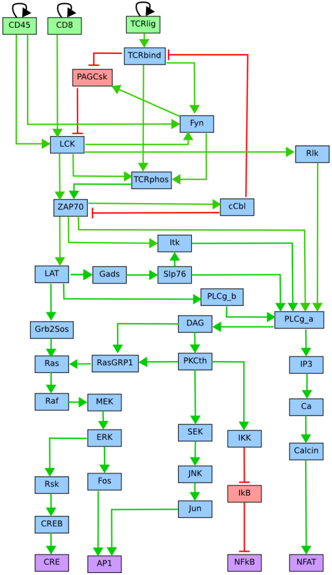
Computational models of biological networks



[Naldi et al, PLOS Comput Biol 2010]

Network: account for **indirect influences** between **entities** of a system

Computational models of biological networks



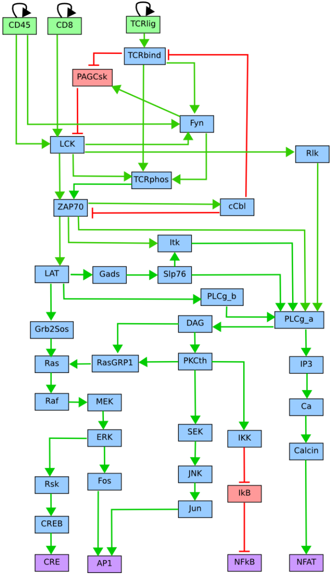
[Naldi et al, PLOS Comput Biol 2010]

Network: account for **indirect influences** between **entities** of a system

A biological model is typically built from

- literature (tedious)
- (curated) databases: pull interactions discovered in very different experimental settings
- network inference from data: prune networks to fit with data; identify new interactions
- expert knowledge (people)

Computational models of biological networks



[Naldi et al, PLOS Comput Biol 2010]

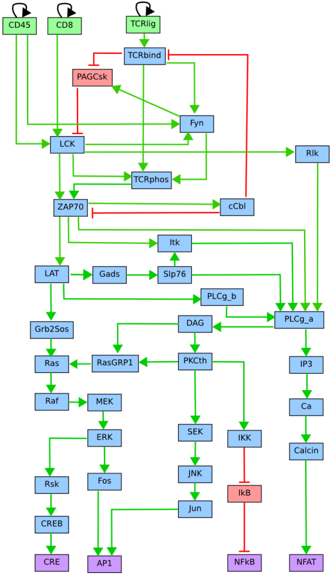
Network: account for **indirect influences** between **entities** of a system

A biological model is typically built from

- literature (tedious)
- (curated) databases: pull interactions discovered in very different experimental settings
- network inference from data: prune networks to fit with data; identify new interactions
- expert knowledge (people)

⇒ **uncertainties / hypotheses remain**
⇒ **set of candidate models**

Computational models of biological networks



[Naldi et al, PLOS Comput Biol 2010]

Network: account for **indirect influences** between **entities** of a system

A biological model is typically built from

- literature (tedious)
- (curated) databases: pull interactions discovered in very different experimental settings
- network inference from data: prune networks to fit with data; identify new interactions
- expert knowledge (people)

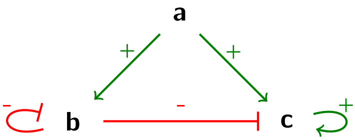
⇒ **uncertainties / hypotheses remain**
⇒ **set of candidate models**

Need for efficient methods to

- validate, refine candidate models
- make predictions robust to uncertainties

Computational models of biological networks

Knowledge



\+ Semantics

Ordinary differential equations

$$\frac{da}{dt} = -k_{da}a$$
$$\frac{db}{dt} = \frac{k_{ab}a}{1+k_{ab}a}\frac{1}{1+k_{bb}b} - k_{db}b$$
$$\frac{dc}{dt} = (\frac{k_{ac}a}{1+k_{ac}a}\frac{k_{cc}c}{1+k_{cc}c}+)\frac{1}{1+k_{bc}b} - k_{dc}c$$

Boolean network

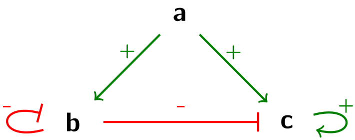
$$f_a(a,b,c) = 0$$
$$f_b(a,b,c) = a \wedge \neg b$$
$$f_c(a,b,c) = \neg b \wedge (a \vee c)$$

Semantics

- Mathematically defines what a state is,
- and how it evolves with time (sequences or chronometry)
- Requires additional parameters, usually not in knowledge

Computational models of biological networks

Knowledge



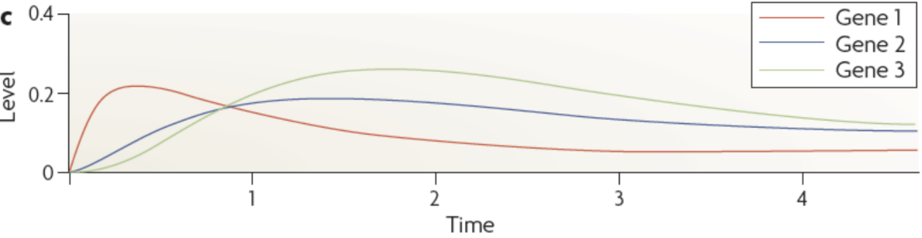
+ Semantics

Ordinary differential equations

$$\frac{da}{dt} = -k_{da}a$$

$$\frac{db}{dt} = \frac{k_{ab}a}{1+k_{ab}a}\frac{1}{1+k_{bb}b} - k_{db}b$$

$$\frac{dc}{dt} = (\frac{k_{ac}a}{1+k_{ac}a}\frac{k_{cc}c}{1+k_{cc}c}+)\frac{1}{1+k_{bc}b} - k_{dc}c$$



Motivating question

Given a computational model of a network,
how to prove that a behaviour is impossible?

Example: it is impossible to reach the state of interest in the current condition

This question is key for:

- Model verification: do we miss something?
- Control prediction: perturbations which makes a behaviour impossible

Motivating question

Given a computational model of a network,
how to prove that a behaviour is impossible?

Example: it is impossible to reach the state of interest in the current condition

This question is key for:

- Model verification: do we miss something?
- Control prediction: perturbations which makes a behaviour impossible

Requires a complete assessment of model capabilities:

- ~~simulation~~
- **formal verification**

Motivating question

Given a computational model of a network,
how to prove that a behaviour is impossible?

Example: it is impossible to reach the state of interest in the current condition

This question is key for:

- Model verification: do we miss something?
- Control prediction: perturbations which makes a behaviour impossible

Requires a complete assessment of model capabilities:

- ~~simulation~~
- **formal verification**

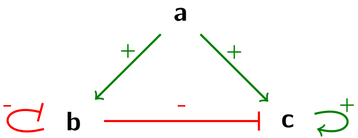
Same principle to prove absence of bugs in computer programs
⇒ similar technologies, very different models.

Outline

1. Computational models for biological processes
   - Networks, dynamics
   - Boolean networks
2. Computational challenges
   - Inference
   - Validation
   - Control
3. Static analysis of Boolean networks by abstract interpretation
   - Main principle
   - Over- and under-approximation of trajectories
   - Applications
4. Perspectives

Dynamics of Qualitative Networks

Example in Boolean case



$f_a(a,b,c) = 0$

$f_b(a,b,c) = a \land \neg b$

$f_c(a,b,c) = \neg b \land (a \lor c)$

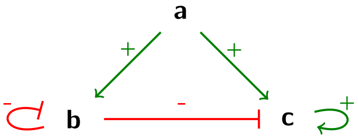
State transition graph with asynchronous updating mode

$\langle a, b, c \rangle$ $\langle 1, 0, 0 \rangle$

[René Thomas in Journal of Theoritical Biology, 1973] [A. Richard, J.-P. Comet, G. Bernot in Modern Formal Methods and Applications, 2006]

Dynamics of Qualitative Networks

Example in Boolean case



$f_a(a, b, c) = 0$

$f_b(a, b, c) = a \wedge \neg b$

$f_c(a, b, c) = \neg b \wedge (a \vee c)$

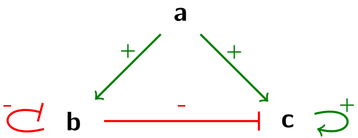
State transition graph with asynchronous updating mode

$$\begin{tikzcd} \langle a, b, c\rangle & \langle 1, 0, 0\rangle \arrow[r, bend left] & \langle 0, 0, 0\rangle \end{tikzcd}$$

[René Thomas in Journal of Theoritical Biology, 1973] [A. Richard, J.-P. Comet, G. Bernot in Modern Formal Methods and Applications, 2006]

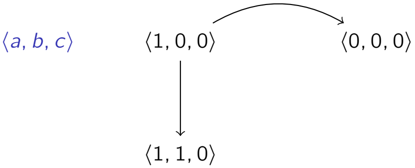
Dynamics of Qualitative Networks

Example in Boolean case



$$f_a(a, b, c) = 0$$
$$f_b(a, b, c) = a \wedge \neg b$$
$$f_c(a, b, c) = \neg b \wedge (a \vee c)$$

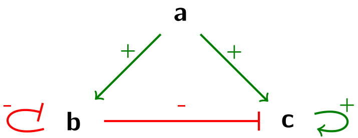
State transition graph with asynchronous updating mode



[René Thomas in Journal of Theoritical Biology, 1973] [A. Richard, J.-P. Comet, G. Bernot in Modern Formal Methods and Applications, 2006]

Dynamics of Qualitative Networks

Example in Boolean case

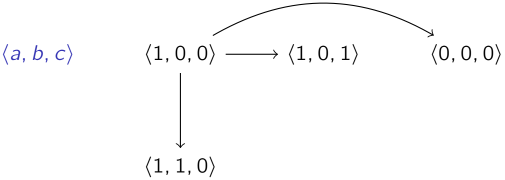


$f_a(a, b, c) = 0$

$f_b(a, b, c) = a \wedge \neg b$

$f_c(a, b, c) = \neg b \wedge (a \vee c)$

State transition graph with asynchronous updating mode



$\langle a, b, c \rangle$

$\langle 1, 0, 0 \rangle$ → $\langle 1, 0, 1 \rangle$

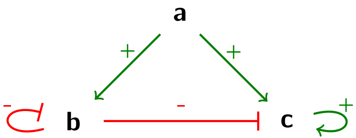
$\langle 1, 0, 0 \rangle$ → $\langle 0, 0, 0 \rangle$

$\langle 1, 0, 0 \rangle$ → $\langle 1, 1, 0 \rangle$

[René Thomas in Journal of Theoritical Biology, 1973] [A. Richard, J.-P. Comet, G. Bernot in Modern Formal Methods and Applications, 2006]

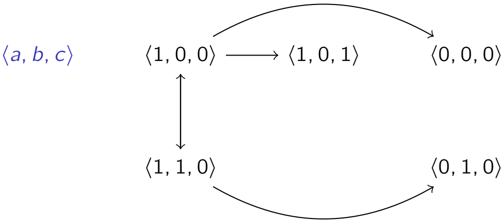
Dynamics of Qualitative Networks

Example in Boolean case



$$f_a(a, b, c) = 0$$
$$f_b(a, b, c) = a \land \neg b$$
$$f_c(a, b, c) = \neg b \land (a \lor c)$$

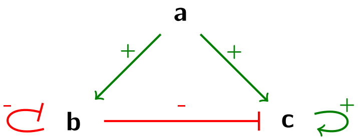
State transition graph with asynchronous updating mode



[René Thomas in Journal of Theoritical Biology, 1973] [A. Richard, J.-P. Comet, G. Bernot in Modern Formal Methods and Applications, 2006]

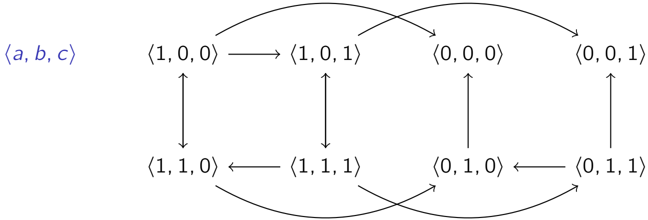
Dynamics of Qualitative Networks

Example in Boolean case



$$f_a(a, b, c) = 0$$
$$f_b(a, b, c) = a \wedge \neg b$$
$$f_c(a, b, c) = \neg b \wedge (a \vee c)$$

State transition graph with asynchronous updating mode



$\langle a, b, c \rangle$

$\langle 1, 0, 0 \rangle$ $\langle 1, 0, 1 \rangle$ $\langle 0, 0, 0 \rangle$ $\langle 0, 0, 1 \rangle$

$\langle 1, 1, 0 \rangle$ $\langle 1, 1, 1 \rangle$ $\langle 0, 1, 0 \rangle$ $\langle 0, 1, 1 \rangle$

[René Thomas in Journal of Theoritical Biology, 1973] [A. Richard, J.-P. Comet, G. Bernot in Modern Formal Methods and Applications, 2006]

Formal Verification of Qualitative Networks

Qualitative models

- Few parameters compared to quantitative models:
  ⇒ quite direct translation from knowledge to computational model
  ⇒ results are general (independant of speed of reactions, precise quantities. . .)
- Not suited for quantitative predictions.

Formal Verification of Qualitative Networks

Qualitative models

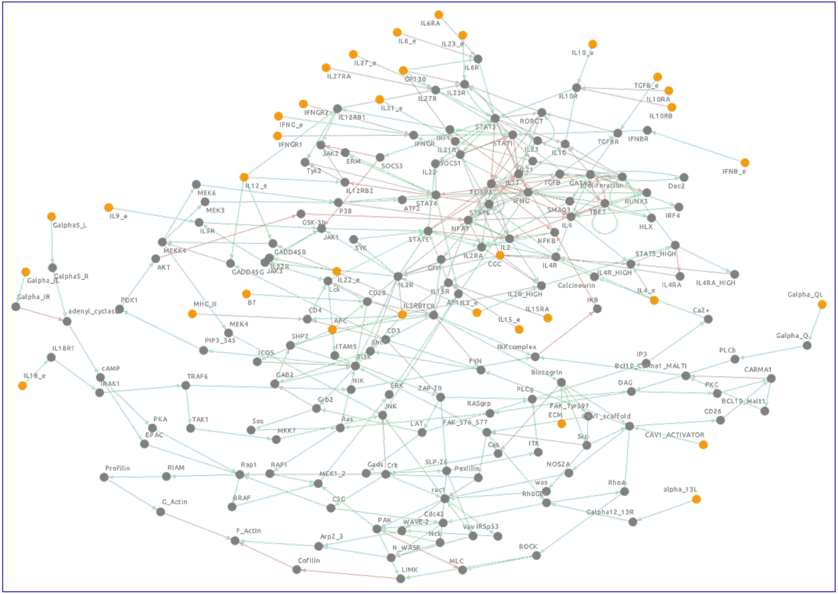
- Few parameters compared to quantitative models:
  ⇒ quite direct translation from knowledge to computational model
  ⇒ results are general (independant of speed of reactions, precise quantities. . .)
- Not suited for quantitative predictions.

State transition graph

- Allows an exhaustive view of model capabilities;
- Automatic "model checking" w.r.t. specifications.

but. . .

Formal Verification of Qualitative Networks



Tractability issues

Model validation (model checking)

- Combinatorial explosion of behaviours
  networks with 100 to 1,000 nodes: 2^{100} - 10^{30} to $2^{1\,000}$ - 10^{300} states
- Large range of initial conditions to consider.
- Difficult to extract comprehensive proofs of (im)possibility.

Tractability issues

Model validation (model checking)

- Combinatorial explosion of behaviours
  networks with 100 to 1,000 nodes: 2^{100} - 10^{30} to $2^{1\,000}$ - 10^{300} states
- Large range of initial conditions to consider.
- Difficult to extract comprehensive proofs of (im)possibility.

Prediction for control (e.g., mutations to block/enforce behaviour)

- Model checking complexity
- + too many candidates for brute-force screening ⇒ deduction/abduction

Tractability issues

Model validation (model checking)

- Combinatorial explosion of behaviours
  networks with 100 to 1,000 nodes: 2^{100} - 10^{30} to $2^{1\,000}$ - 10^{300} states
- Large range of initial conditions to consider.
- Difficult to extract comprehensive proofs of (im)possibility.

Prediction for control (e.g., mutations to block/enforce behaviour)

- Model checking complexity
- + too many candidates for brute-force screening $\Rightarrow$ deduction/abduction

Network inference

- Combinatorial explosion of model parameters
- Data involve time series: reachability checking complexity..

Tractability issues

Model validation (model checking)

- Combinatorial explosion of behaviours
  networks with 100 to 1,000 nodes: 2^{100} - 10^{30} to $2^{1\,000}$ - 10^{300} states
- Large range of initial conditions to consider.
- Difficult to extract comprehensive proofs of (im)possibility.

Prediction for control (e.g., mutations to block/enforce behaviour)

- Model checking complexity
- + too many candidates for brute-force screening ⇒ deduction/abduction

Network inference

- Combinatorial explosion of model parameters
- Data involve time series: reachability checking complexity..

⇒ avoid building the state graph! compute something else (abstraction)

Outline

1. Computational models for biological processes
   - Networks, dynamics
   - Boolean networks
2. Computational challenges
   - Inference
   - Validation
   - Control
3. Static analysis of Boolean networks by abstract interpretation
   - Main principle
   - Over- and under-approximation of trajectories
   - Applications
4. Perspectives

Semantics of Boolean networks

Definition (Boolean network)

$f = \langle f_1, \cdots, f_n \rangle$ with $f_i : \{0,1\}^n \to \{0,1\}$

Example:

$$\begin{aligned} f_1(x) &= 0 \\ f_2(x) &= x_1 \wedge \neg x_2 \\ f_3(x) &= \neg x_2 \wedge (x_1 \vee x_3) \end{aligned}$$

Definition (Asynchronous transition)

Irreflexixe relation $\to \subseteq \{0,1\}^n \times \{0,1\}^n$ such that

$$x \to y \iff \boldsymbol{\Delta}(x,y) = \{i\} \wedge y_i = f_i(x)$$

where $\boldsymbol{\Delta}(x,y) = \{i \in \{1, \cdots, n\} \mid x_i \neq y_i\}$

Transition Prime Implicants

Consider a Boolean network f where

$$f_1(x) = x_2$$

there exists 2^{n-2} transitions of the form

$$01x_2 \cdots x_n \to 11x_2 \cdots x_n$$

Transition Prime Implicants

Consider a Boolean network f where

$$f_1(x) = x_2$$

there exists 2^{n-2} transitions of the form

$$01x_2 \cdots x_n \to 11x_2 \cdots x_n$$

Prime implicants of transitions

- Express the minimal cause of a node value change $\mathtt{v}_i : a \rightsquigarrow b$

Definition

C is a prime implicants for a node value change $\mathtt{v}_i : a \rightsquigarrow b$ iff its a conjunction of literals of the form $[\mathtt{v}_j = d]$ such that

$$[\mathtt{v}_i = a \wedge C] \text{ is a prime implicant of } [f_i(\mathtt{v}) = b]$$

We write it $\langle \mathtt{v}_i : a \rightsquigarrow b, C \rangle$

In our case, only one transition prime implicant for $\mathtt{v}_1 : 0 \rightsquigarrow 1$:

$$\langle \mathtt{v}_1 : 0 \rightsquigarrow 1, [\mathtt{v}_2 = 1] \rangle$$

Transitions Prime Implicants

Illustration

$$f_1(x) = x_2 \vee x_3$$

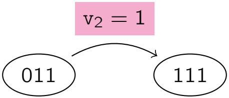
Implicants for the transition $\langle 011 \rangle \rightarrow \langle 111 \rangle$

Transitions Prime Implicants

Illustration

$$f_1(x) = x_2 \vee x_3$$

Implicants for the transition $\langle 011\rangle \rightarrow \langle 111\rangle$

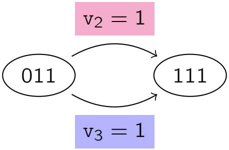


Transitions Prime Implicants

Illustration

$$f_1(x) = x_2 \vee x_3$$

Implicants for the transition $\langle 011\rangle \rightarrow \langle 111\rangle$

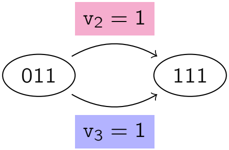


Transitions Prime Implicants

Illustration

$$f_1(x) = x_2 \vee x_3$$

Implicants for the transition $\langle 011 \rangle \rightarrow \langle 111 \rangle$



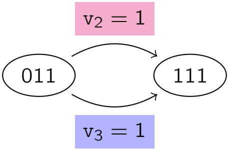
$\Rightarrow$ 2 prime implicants: $\langle \mathrm{v}_1 : 0 \rightsquigarrow 1, [\, \mathrm{v}_2 = 1 \,] \rangle$; $\langle \mathrm{v}_1 : 0 \rightsquigarrow 1, [\, \mathrm{v}_3 = 1 \,] \rangle$

Transitions Prime Implicants

Illustration

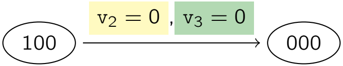
$$f_1(x) = x_2 \vee x_3$$

Implicants for the transition $\langle 011\rangle \rightarrow \langle 111\rangle$



$\Rightarrow$ 2 prime implicants: $\langle \mathrm{v}_1 : 0 \rightsquigarrow 1, [\,\mathrm{v}_2 = 1\,]\rangle$; $\langle \mathrm{v}_1 : 0 \rightsquigarrow 1, [\,\mathrm{v}_3 = 1\,]\rangle$

Implicants for the transition $\langle 100\rangle \rightarrow \langle 000\rangle$

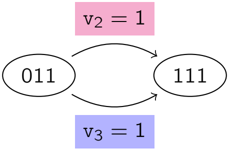


Transitions Prime Implicants

Illustration

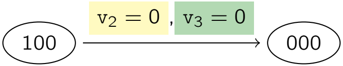
$$f_1(x) = x_2 \vee x_3$$

Implicants for the transition $\langle 011\rangle \rightarrow \langle 111\rangle$



⇒ 2 prime implicants: $\langle \mathrm{v}_1 : 0 \rightsquigarrow 1, [\,\mathrm{v}_2 = 1\,]\rangle$; $\langle \mathrm{v}_1 : 0 \rightsquigarrow 1, [\,\mathrm{v}_3 = 1\,]\rangle$

Implicants for the transition $\langle 100\rangle \rightarrow \langle 000\rangle$



⇒ 1 prime implicant: $\langle \mathrm{v}_1 : 1 \rightsquigarrow 0, [\,\mathrm{v}_2 = 0 \wedge \mathrm{v}_3 = 0\,]\rangle$

Galois connections

A Galois connection $(C, \leq) \underset{\alpha}{\overset{\gamma}{\leftrightarrows}} (A, \preceq)$ ensures that

- $\forall c \in C, c \leq \gamma(\alpha(c))$
- $\forall a \in A, \alpha(\gamma(a)) \preceq a$

Concrete transitions vs abstract prime implicants

$$(\wp(\rightarrow), \subseteq) \underset{\alpha}{\overset{\gamma}{\leftrightarrows}} (\wp(\text{transitions prime implicants}), \subseteq)$$

where $| \rightarrow | \in O(n \cdot 2^n)$; $|\text{transition prime implicants}| \in O(n \cdot 2^d)$
d is the number of nodes a node's f_i depends on (nb of direct regulators)

Galois connections

A Galois connection $(\mathrm{C}, \leq) \underset{\alpha}{\overset{\gamma}{\leftrightarrows}} (\mathrm{A}, \preceq)$ ensures that

- $\forall c \in C, c \leq \gamma(\alpha(c))$
- $\forall a \in A, \alpha(\gamma(a)) \preceq a$

Concrete transitions vs abstract prime implicants

$$(\wp(\to), \subseteq) \underset{\alpha}{\overset{\gamma}{\leftrightarrows}} (\wp(\text{transitions prime implicants}), \subseteq)$$

where $|\to| \in O(n \cdot 2^n)$; $|\text{transition prime implicants}| \in O(n \cdot 2^d)$
d is the number of nodes a node's f_i depends on (nb of direct regulators)

Concrete states vs abstract literals $L = \{[\mathrm{v}_i = n] \mid i \in \{1, \cdots, n\}, b \in \{0, 1\}\}$

$$(\wp(\{0,1\}^n), \subseteq) \underset{\mathring{\alpha}}{\overset{\mathring{\gamma}}{\leftrightarrows}} (\wp(L), \subseteq)$$

where $|\{0,1\}^n| = 2^n$; $|L| = 2 \cdot n$

Over-approximation of reachability

Concrete reach

$S = \{x \qquad\quad \}$

$T = \{$

$\}$

Over-approximation of reachability

Concrete reach

$S = \{x \qquad\quad \}$

$T = \{x \rightarrow y$

$\}$

Over-approximation of reachability

Concrete reach

$S = \{x, y \quad \}$

$T = \{x \rightarrow y$

$\}$

Over-approximation of reachability

Concrete reach

$$S = \{x, y \qquad \}$$
$$T = \{x \rightarrow y,$$
$$y \rightarrow z, y \rightarrow z'\}$$

Over-approximation of reachability

Concrete reach

$$S = \{x, y, z, z'\}$$

$$T = \{x \rightarrow y,$$

$$y \rightarrow z, y \rightarrow z'\}$$

Over-approximation of reachability

Concrete reach

$$S = \{x, y, z, z'\}$$
$$T = \{x \to y,$$
$$y \to z, y \to z'\}$$

Reachable transitions from x

=

lfp reach

Over-approximation of reachability

Concrete reach

$$S = \{x, y, z, z'\}$$
$$T = \{x \to y,$$
$$y \to z, y \to z'\}$$

Reachable transitions from x

=

lfp reach

Abstract $\text{reach}^{\#}$

Over-approximation of reachability

Concrete reach

$$S = \{x, y, z, z'\}$$
$$T = \{x \to y,\quad y \to z, y \to z'\}$$

Reachable transitions from x
=
lfp reach

Abstract reach$^{\#}$

$$S^{\#} = \{[\mathrm{v}_1 = 0], [\mathrm{v}_2 = 0], [\mathrm{v}_3 = 0]\quad\}$$
$$T^{\#} = \{\quad\}$$

Over-approximation of reachability

Concrete reach

$$S = \{x, y, z, z'\}$$
$$T = \{x \rightarrow y, \quad y \rightarrow z, y \rightarrow z'\}$$

Reachable transitions from x

=

lfp reach

Abstract reach$^{\#}$

$$S^{\#} = \{[\mathrm{v}_1 = 0], [\mathrm{v}_2 = 0], [\mathrm{v}_3 = 0] \quad \}$$
$$T^{\#} = \{\langle \mathrm{v}_2 : 0 \rightsquigarrow 1, [\mathrm{v}_1 = 0] \rangle \quad \}$$

Over-approximation of reachability

Concrete reach

$$S = \{x, y, z, z'\}$$
$$T = \{x \to y,$$
$$y \to z, y \to z'\}$$

Reachable transitions from x

=

lfp reach

Abstract reach$^{\#}$

$$S^{\#} = \{[\mathrm{v}_1 = 0], [\mathrm{v}_2 = 0], [\mathrm{v}_3 = 0]$$
$$\}$$
$$T^{\#} = \{\langle \mathrm{v}_{2} : 0 \rightsquigarrow \mathbf{1}, [\mathrm{v}_1 = 0]\rangle$$

$$\}$$

Over-approximation of reachability

Concrete reach

$$S = \{x, y, z, z'\}$$
$$T = \{x \to y,\; y \to z, y \to z'\}$$

Reachable transitions from x

=

lfp reach

Abstract reach$^{\#}$

$$S^{\#} = \{[\mathrm{v}_1 = 0], [\mathrm{v}_2 = 0], [\mathrm{v}_3 = 0], [\mathrm{v}_2 = 1]\ \}$$
$$T^{\#} = \{\langle \mathrm{v}_{\mathbf{2}} : 0 \rightsquigarrow \mathbf{1}, [\mathrm{v}_1 = 0]\rangle$$
$$\}$$

Over-approximation of reachability

Concrete reach

$$S = \{x, y, z, z'\}$$
$$T = \{x \rightarrow y,\ y \rightarrow z, y \rightarrow z'\}$$

Reachable transitions from x

=

lfp reach

Abstract reach$^{\#}$

$$S^{\#} = \{[\mathrm{v}_1 = 0], [\mathrm{v}_2 = 0], [\mathrm{v}_3 = 0], [\mathrm{v}_2 = 1]\ \}$$
$$T^{\#} = \{\langle \mathrm{v}_2 : 0 \rightsquigarrow 1, [\mathrm{v}_1 = 0]\rangle, \langle \mathrm{v}_1 : 0 \rightsquigarrow 1, [\mathrm{v}_2 = 1]\rangle, \langle \mathrm{v}_3 : 0 \rightsquigarrow 1, [\mathrm{v}_1 = 0 \land \mathrm{v}_2 = 1]\rangle\}$$

Over-approximation of reachability

Concrete reach

$$S = \{x, y, z, z'\}$$

$$T = \{x \to y,\quad y \to z, y \to z'\}$$

Reachable transitions from x
=
lfp reach

Abstract reach$^\#$

$$S^\# = \{[\mathrm{v}_1 = 0], [\mathrm{v}_2 = 0], [\mathrm{v}_3 = 0], [\mathrm{v}_2 = 1]\ \}$$

$$T^\# = \{\langle \mathrm{v}_2 : 0 \rightsquigarrow 1, [\mathrm{v}_1 = 0]\rangle, \langle \mathrm{v}_1 : 0 \rightsquigarrow 1, [\mathrm{v}_2 = 1]\rangle, \langle \mathrm{v}_3 : 0 \rightsquigarrow 1, [\mathrm{v}_1 = 0 \wedge \mathrm{v}_2 = 1]\rangle\}$$

$$\alpha(\mathrm{lfp}\ \text{reach}) \subseteq \mathrm{lfp}\ \text{reach}^\#$$

Goal-oriented properties

Goal: reach a state verifying $[\mathrm{v}_g = 1]$

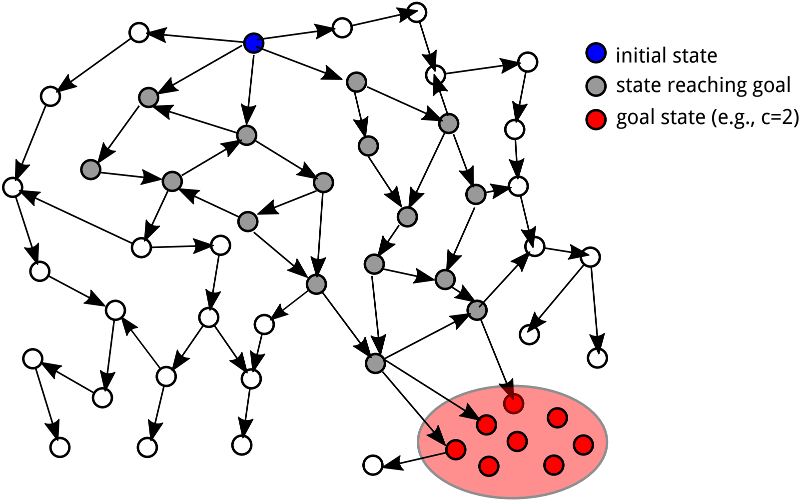
Goal-oriented properties

Goal: reach a state verifying $[\mathrm{v}_g = 1]$

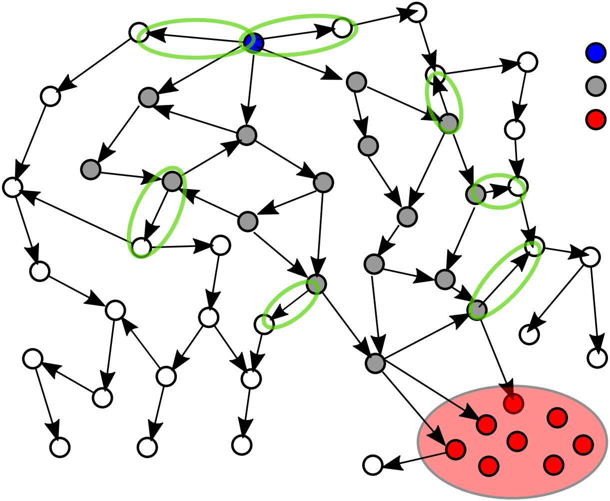
Over-approximation of direct paths [CMSB 2016; CONCUR 2017]

- Acyclic trajectories where all transitions are causally related to the goal
- Over-approximation in linear time with nb of transition prime implicants
- $\Rightarrow$ model reduction; goal-driven exploration of the state space

Goal-oriented properties

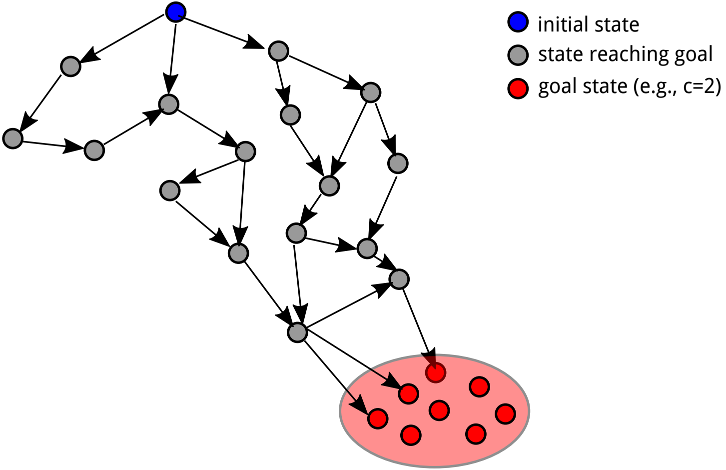


Goal-oriented properties



- 🔵 initial state
- ⚪ state reaching goal
- 🔴 goal state (e.g., c=2)

Goal-oriented properties



Goal-oriented properties

Goal: reach a state verifying $[\mathrm{v}_g = 1]$

Over-approximation of direct paths [CMSB 2016; CONCUR 2017]

- Acyclic trajectories where all transitions are causally related to the goal
- Over-approximation in linear time with nb of transition prime implicants
- ⇒ model reduction; goal-driven exploration of the state space

Goal-oriented properties

Goal: reach a state verifying $[\mathtt{v}_g = 1]$

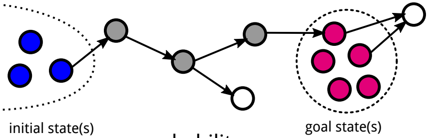
Over-approximation of direct paths [CMSB 2016; CONCUR 2017]

- Acyclic trajectories where all transitions are causally related to the goal
- Over-approximation in linear time with nb of transition prime implicants
- ⇒ model reduction; goal-driven exploration of the state space

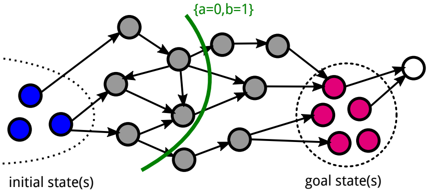
Under-approximation of reachability [TCS 2015]

- Sufficient condition on a set of prime implicants to ensure existence of a concrete trajectory from an initial state
- Deciding if a set satisfies sufficient condition: linear time
- Finding a set satisfying sufficient condition: NP
  (choices are the different implicants for a same $\mathtt{v}_i : a \rightsquigarrow b$)

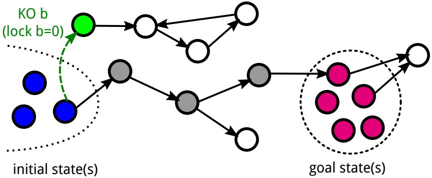
Applications



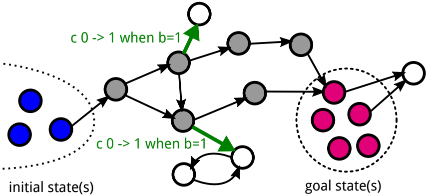
reachability



cut sets

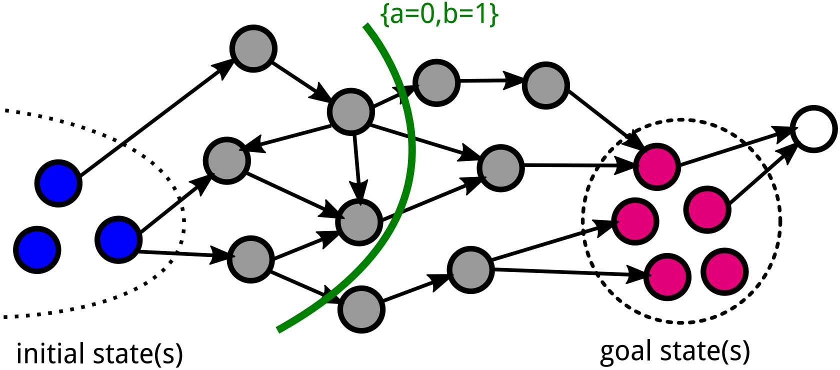


mutations



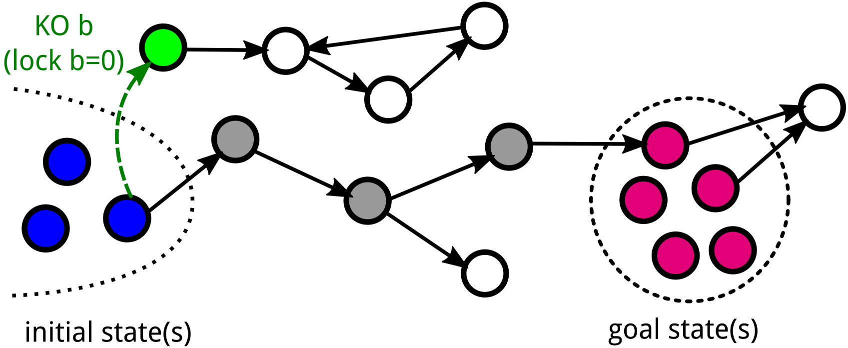
bifurcations

Applications



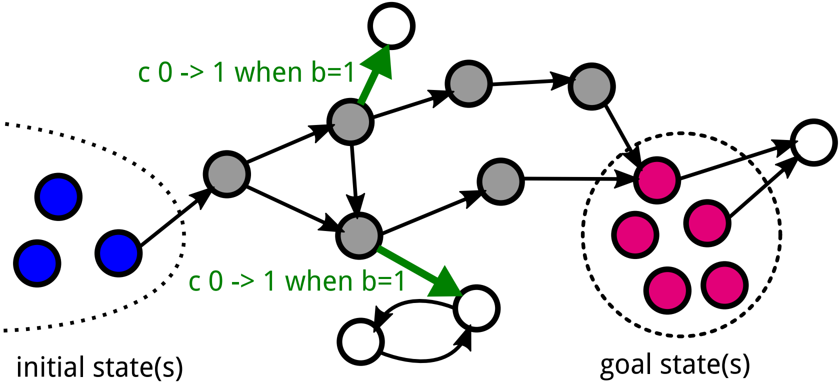
cut sets

Applications



mutations

Applications



bifurcations

Formal Approximations

Examples of implementations

Reachability [LP, M Magnin, O Roux in MSCS 2012; M Folschette, LP, M Magnin, O Roux in TCS 2015]

- Over-approximation (necessary condition): $\mathsf{OA}(s \rightarrow^* g)$
  P w/ # prime implicants
- Under-approximation (sufficient condition): $\mathsf{UA}(s \rightarrow^* g)$
  NP w/ # prime implicants

Cut-sets [LP, G Andrieux, H Koeppl at CAV 2013]

- UA: $a_i, b_j, \cdots : \mathbf{disable}(a_i, b_j, \cdots) \wedge \neg\, \mathsf{OA}(s \rightarrow^* g)$

Mutations for blocking g [LP at CMSB 2017]

- UA: $a_i, b_j, \cdots : \mathbf{lock}(a_i, b_j, \cdots) \wedge \neg\, \mathsf{OA}(s \rightarrow^* g)$

Bifurcations [L F Fitime, C Guziolowski, O Roux, LP in BMC Algorithms for Mol Bio, 2017]

- UA: $s_b, t_b : \mathsf{UA}(s \rightarrow^* s_b) \wedge \mathsf{UA}(s_b \rightarrow^* g) \wedge \neg\, \mathsf{OA}(s_b \cdot t_b \rightarrow^* g)$

Formal approximations

Summary of the approach

Abstract interpretation of Boolean networks

- Reduce complexity from exp with nb nodes to nb transition prime implicants
    - worst case: exp with in-degree d;
    - monotonic functions: $< \binom{d}{d/2}$;
    - biological networks: most of time linear/poly with nb of activators
- SAT implementation of formal approximations
- Low number of variables compared to nb reachable states
- Is generalized to muti-valued networks and automata networks/petri nets (NP approximation of PSPACE problems)

⇒ highly tractable for large biological networks

Formal approximations

Summary of the approach

Abstract interpretation of Boolean networks

- Reduce complexity from exp with nb nodes to nb transition prime implicants
    - worst case: exp with in-degree d;
    - monotonic functions: $< \binom{d}{d/2}$;
    - biological networks: most of time linear/poly with nb of activators
- SAT implementation of formal approximations
- Low number of variables compared to nb reachable states
- Is generalized to muti-valued networks and automata networks/petri nets (NP approximation of PSPACE problems)

⇒ highly tractable for large biological networks

Compared to Bounded Model-Checking (BMC):

- BMC is an under-approximation only, no necessary condition
- BMC can lead to a huge, when not intractable, number of variables (states reachable in less than n transitions)
- BMC gives incomplete capture of trajectories (crucial for control)

In practice

Software **Pint** - Static analyzer for dynamics of automata networks
`http://loicpauleve.name/pint` [CMSB 2017]

- Input: Boolean/discrete networks; automata networks; 1-bounded Petri nets
- Answer-Set Programming implementation for solution enumeration (clingo)
- Scalable to networks between 100 to 10,000 nodes

Tutorial on `http://tmpnb.loicpauleve.name`

Outline

1. Computational models for biological processes
    - Networks, dynamics
    - Boolean networks
2. Computational challenges
    - Inference
    - Validation
    - Control
3. Static analysis of Boolean networks by abstract interpretation
    - Main principle
    - Over- and under-approximation of trajectories
    - Applications
4. Perspectives

Related projects

- Abstract interpretation
- Concurrency; partial order approaches; unfoldings
- Formalization of biological problems into general computer science problems
  ⇒ necessary to attract formal computer science people

Related projects

- Abstract interpretation
- Concurrency; partial order approaches; unfoldings
- Formalization of biological problems into general computer science problems
  ⇒ necessary to attract formal computer science people

Parametric Boolean networks

- Boolean network identification from time series data [BioSystems 2016]
- Abstract interpretation with unfolding semantics [SASB 2016; PhD thesis of Juraj Kolčák]

Related projects

- Abstract interpretation
- Concurrency; partial order approaches; unfoldings
- Formalization of biological problems into general computer science problems
  $\Rightarrow$ necessary to attract formal computer science people

Parametric Boolean networks

- Boolean network identification from time series data [BioSystems 2016]
- Abstract interpretation with unfolding semantics [SASB 2016; PhD thesis of Juraj Kolčák]

Cellular reprogramming

- France-Luxembourg project "AlgoReCell" `http://algorecell.lri.fr` (2017)
- Teams: computer science; computational systems biology; experimental biologists
- Experimental goal: trans-differentiate adipocytes (fat) into osteoblasts (bone)

Outlook

In systems biology

- Need for detailed networks, but lack of precise knowledge
- ⇒ Boolean networks are more and more popular
- Success stories start to come out

Outlook

In systems biology

- Need for detailed networks, but lack of precise knowledge
- ⇒ Boolean networks are more and more popular
- Success stories start to come out

Computational methods allow to address larger and larger networks;
quite soon at similar scale as genome. . .

Outlook

In systems biology

- Need for detailed networks, but lack of precise knowledge
- ⇒ Boolean networks are more and more popular
- Success stories start to come out

Computational methods allow to address larger and larger networks; quite soon at similar scale as genome. . .

BUT

Outlook

In systems biology

- Need for detailed networks, but lack of precise knowledge
- ⇒ Boolean networks are more and more popular
- Success stories start to come out

Computational methods allow to address larger and larger networks;
quite soon at similar scale as genome. . .

BUT

Major remaining challenge: deal with huge number of candidate models

- network inference lead to many networks, equivalent w.r.t. available data
- most methods take as input a single model. . .
- how to make convincing predictions?

Acknowledgements

LS2N, Nantes MeForBio

- Olivier Roux
- Morgan Magnin
- Maxime Folschette

Institut Curie

- Laurence Calzone
- Andrei Zinovyev

LSV/Inria Saclay, ENS Cachan

- Stefan Haar
- Thomas Chatain
- Stefan Schwoon
- Hugues Mandon
- Juraj Kolčák

Masaryk University

- David Šafránek

Funding

cnrs

ANR

Luxembourg National
Research Fund