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

LaBRI - 7 Novembre 2017

Formalization of biological problems into general computer science problems

- help with automated reasoning on biological knowledge/models
- challenging: bring classes of difficult problems
  $\Rightarrow$ motivation for new theoretical/technical developments
- try to attract formal computer scientists

Outline

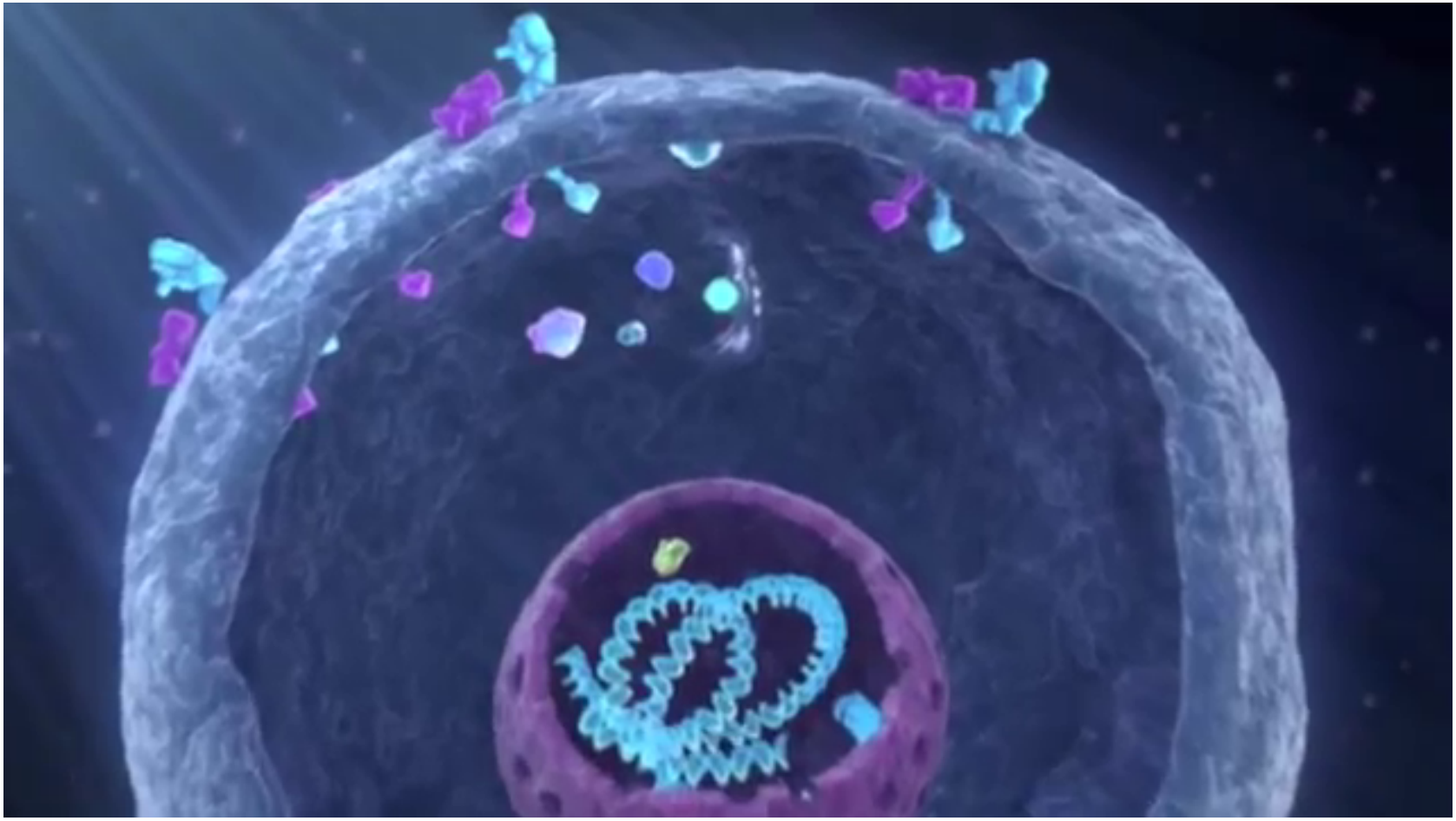
1. Computational models for biological processes
   - Networks, dynamics
2. Boolean networks
   - Definition
   - Properties of interest
3. Static analysis of Boolean networks by abstract interpretation
   - Main principle
   - Over- and under-approximation of trajectories
   - Applications
4. Perspectives

Outline

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

Biological processes

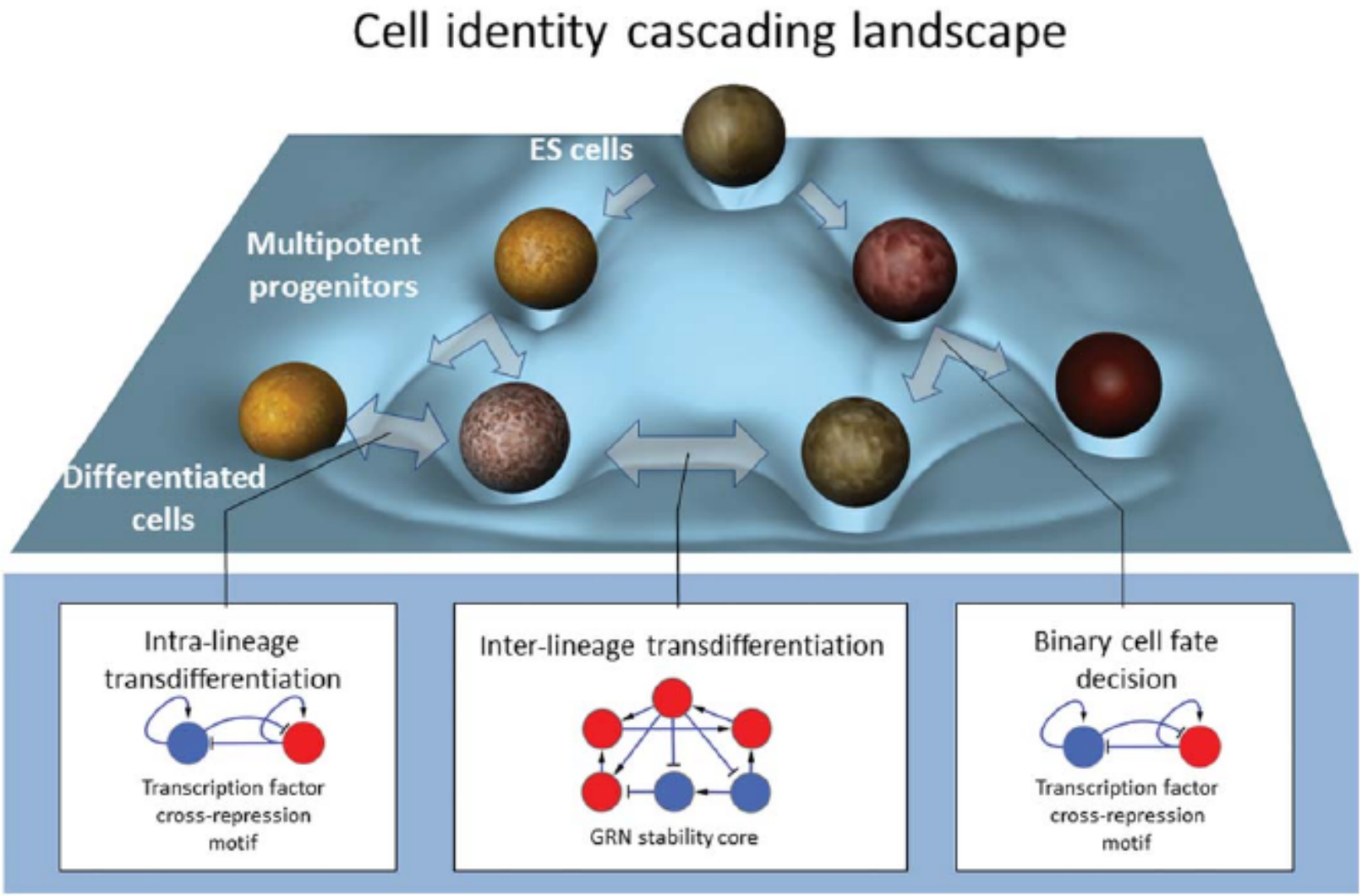
Cell division



(source: Genentech, https://www.youtube.com/watch?v=oDjDUUhGVsI)

Biological Processes

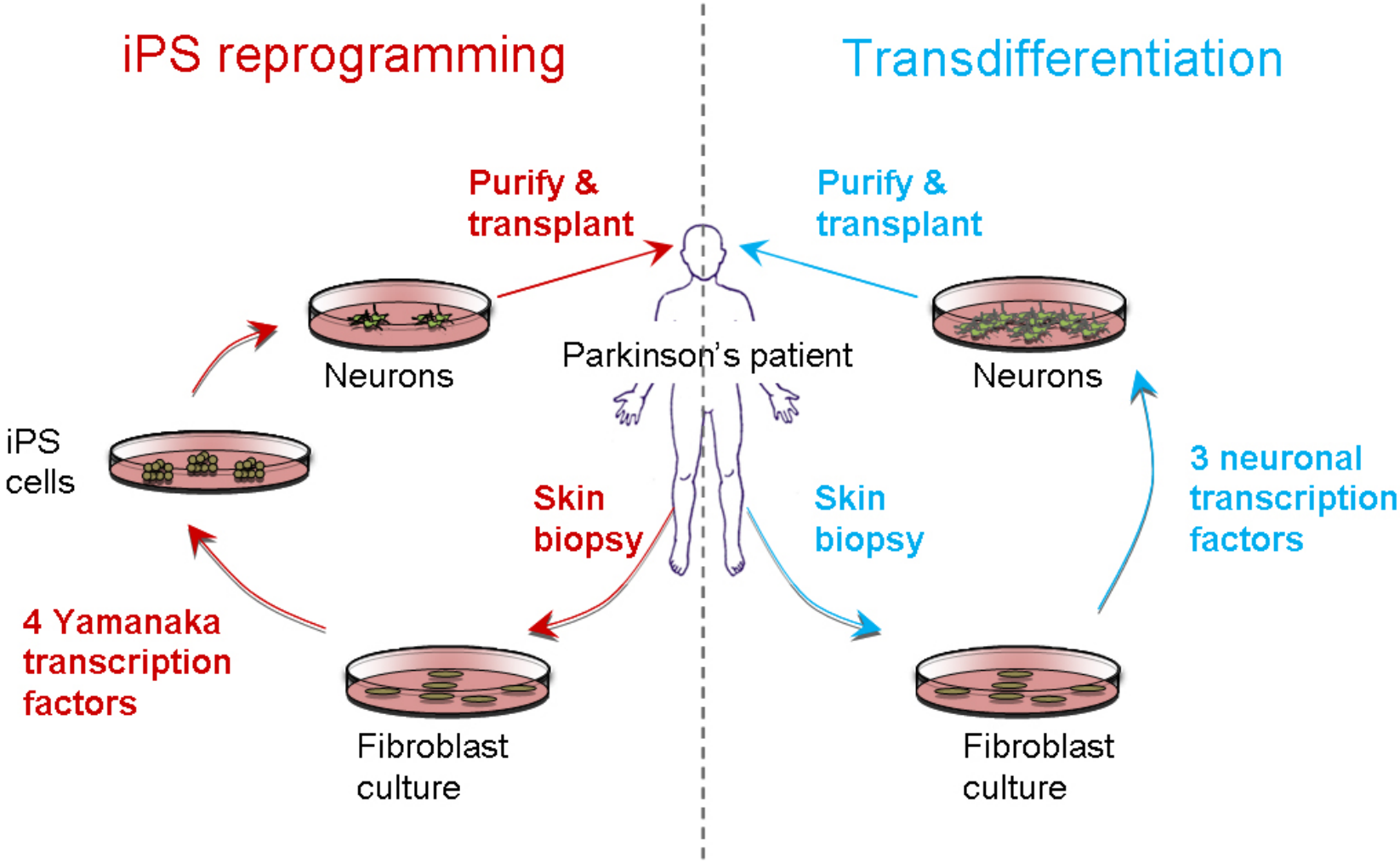
Cell differentiation



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

Biological Processes

Cell reprogramming



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

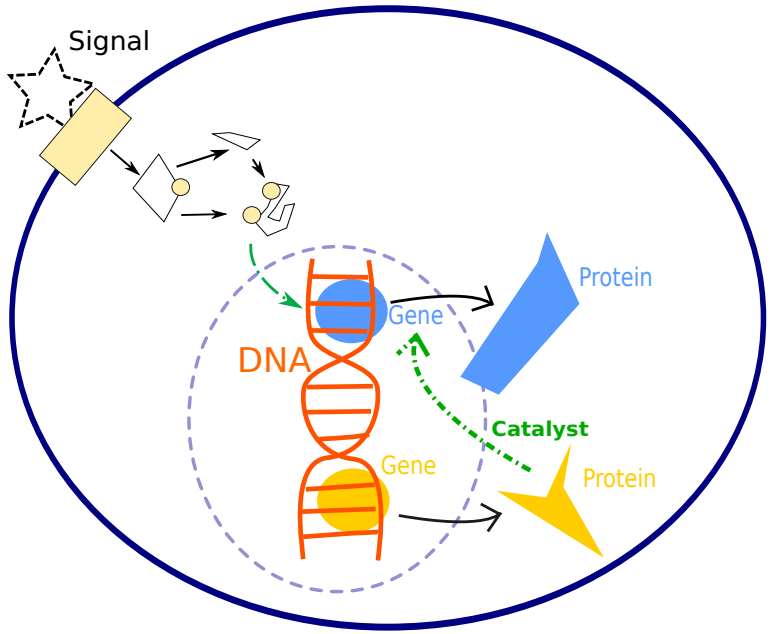
Biological processes

Numerous modelling approaches

- Structure of molecules (RNA, DNA, proteins)
  ⇒ predict dockings, change of conformation/function, ...
- Quantitative models (ODEs, stochastic population models, ...)
  ⇒ track evolution of concentrations/copy number of molecules
  ⇒ requires a huge amount of precise paramaters
- Qualitative models (Boolean networks, threshold networks, ...)
  ⇒ focus on causal processes
  ⇒ abstract/generic view of the system, requires few parameters
- Multi-cellular spatial models, organs, individuals, ecology, ...

No "true" model, each modelling approach is justified on its own.

Signalling and gene networks



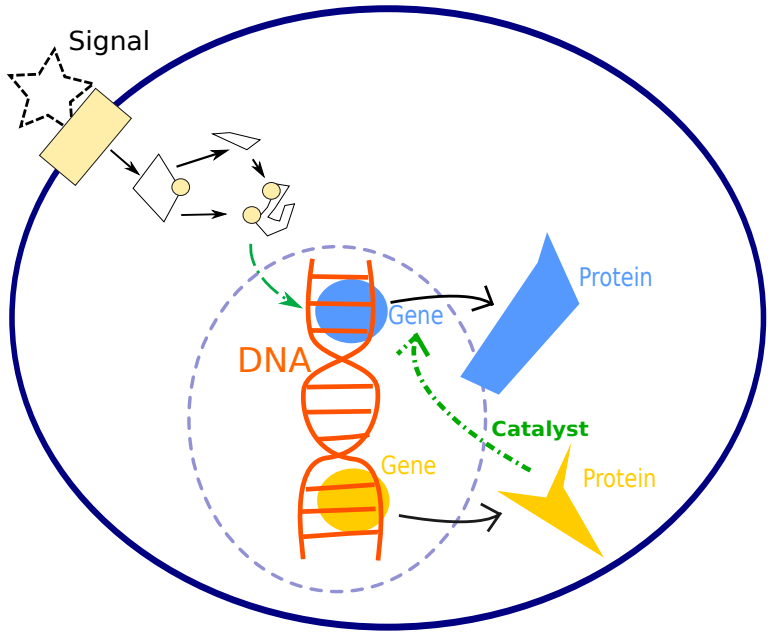
Prediction

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

Control

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

Signalling and gene networks



Prediction

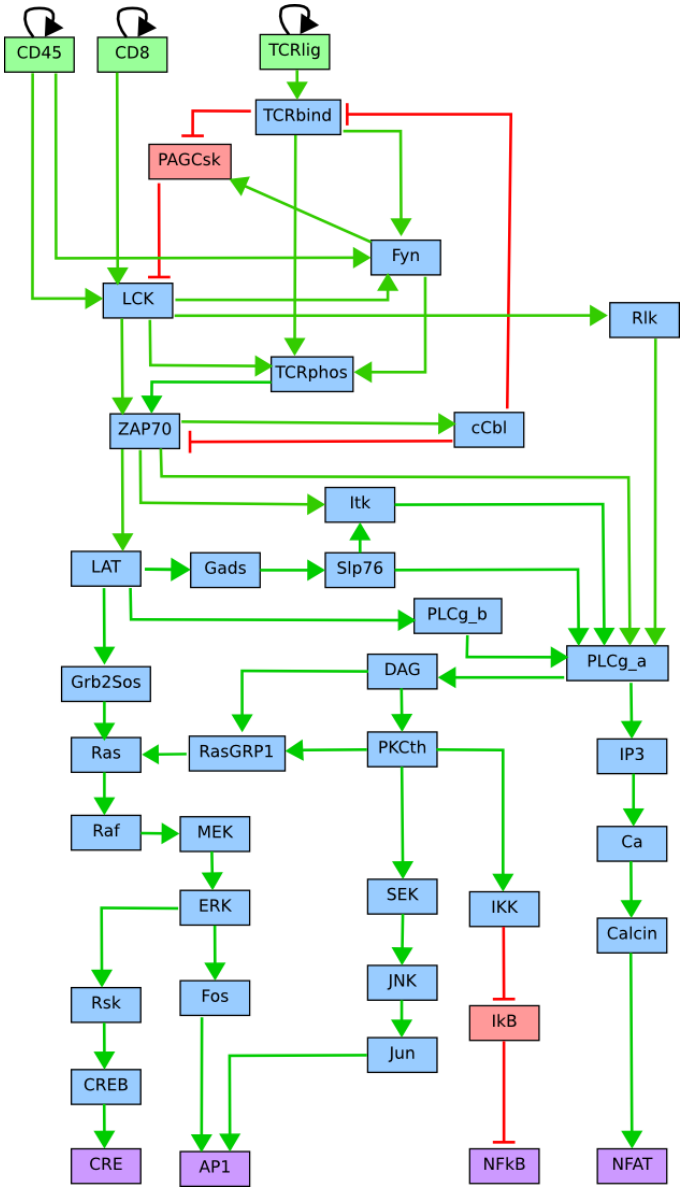
- Cell response w.r.t. signal+environment
- Long-term behaviours (differentiation)

Control

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

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

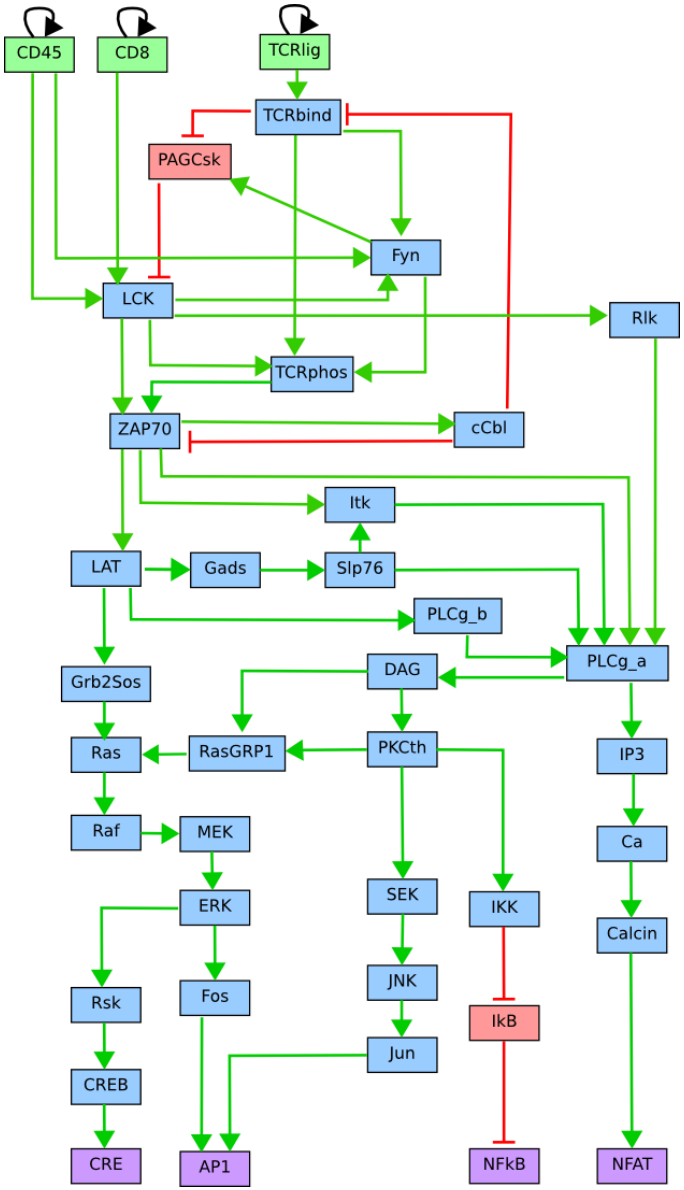
Computational models of biological networks



[Naldi et al, PLOS Comput Biol 2010]

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

Computational models of biological networks



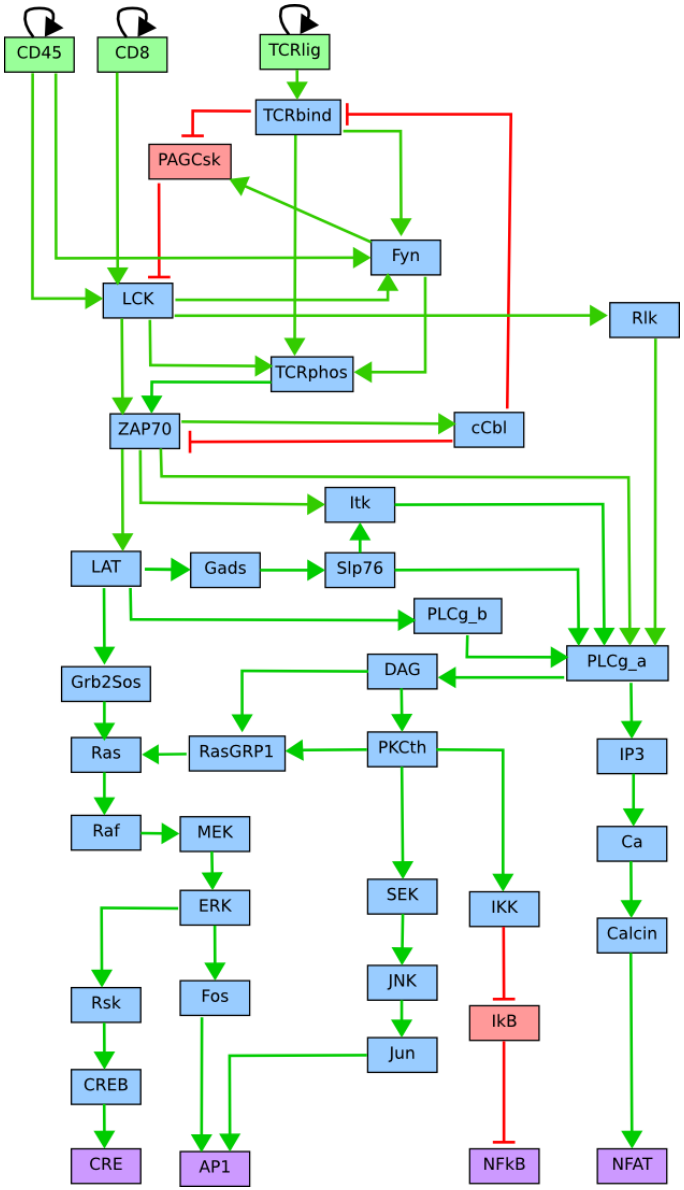
[Naldi et al, PLOS Comput Biol 2010]

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]

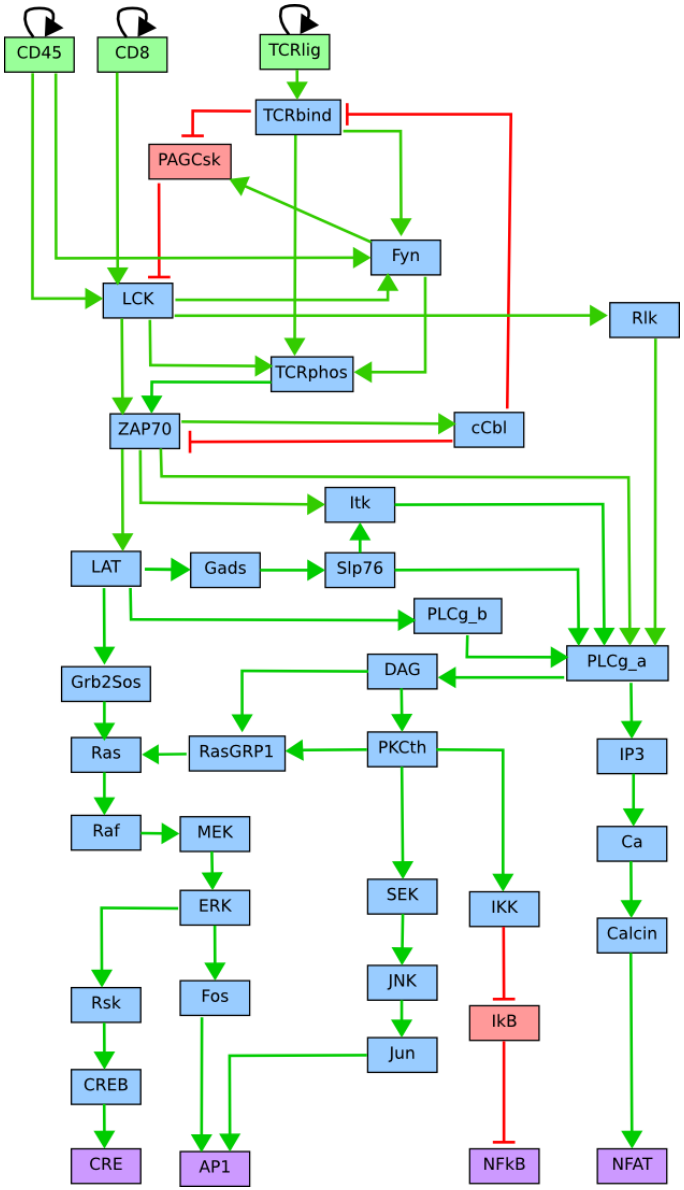
Network: account for **indirect influences** between **entities** of a system

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

Outline

1 Computational models for biological processes
Networks, dynamics

2 Boolean networks
Definition
Properties of interest

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:

$$f_1(x) = 0$$
$$f_2(x) = x_1 \wedge \neg x_2$$
$$f_3(x) = \neg x_2 \wedge (x_1 \vee x_3)$$

Definition (Asynchronous transition)

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

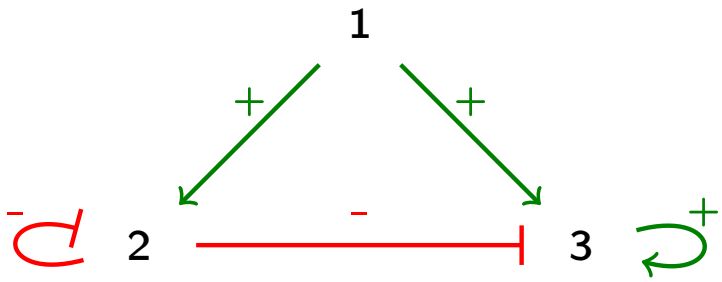
$$x \to y \iff \Delta(x,y) = \{i\} \wedge y_i = f_i(x)$$

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

(non-deterministic semantics)

Example

Influence graph (knowledge)



Boolean network

$$f_1(x) = 0$$
$$f_2(x) = x_1 \wedge \neg x_2$$
$$f_3(x) = \neg x_2 \wedge (x_1 \vee x_3)$$

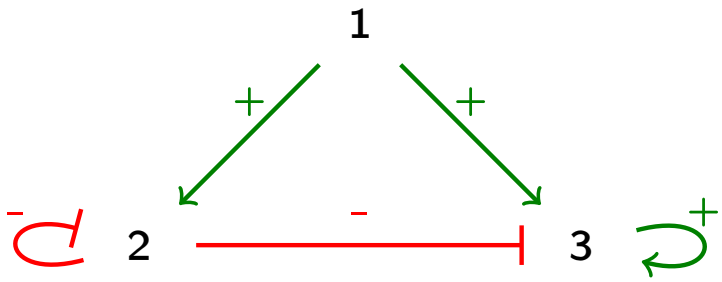
State transition graph with asynchronous updating mode

$\langle 1, 0, 0 \rangle$

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

Example

Influence graph (knowledge)



Boolean network

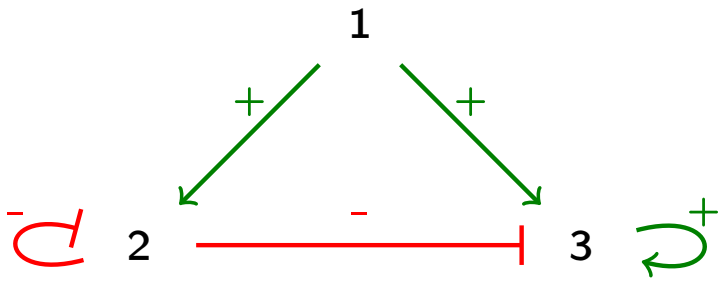
$$f_1(x) = 0$$
$$f_2(x) = x_1 \wedge \neg x_2$$
$$f_3(x) = \neg x_2 \wedge (x_1 \vee x_3)$$

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]

Example

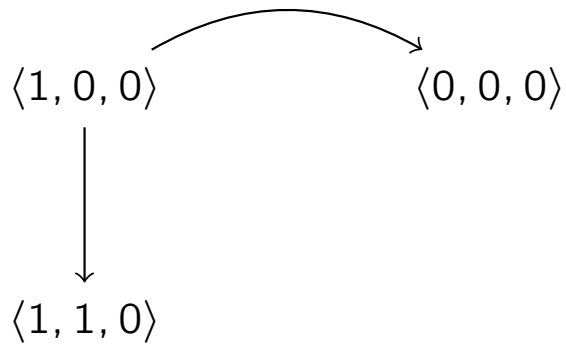
Influence graph (knowledge)



Boolean network

$$f_1(x) = 0$$
$$f_2(x) = x_1 \wedge \neg x_2$$
$$f_3(x) = \neg x_2 \wedge (x_1 \vee x_3)$$

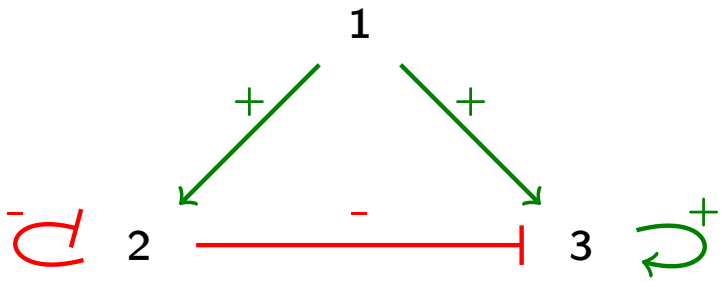
State transition graph with asynchronous updating mode



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

Example

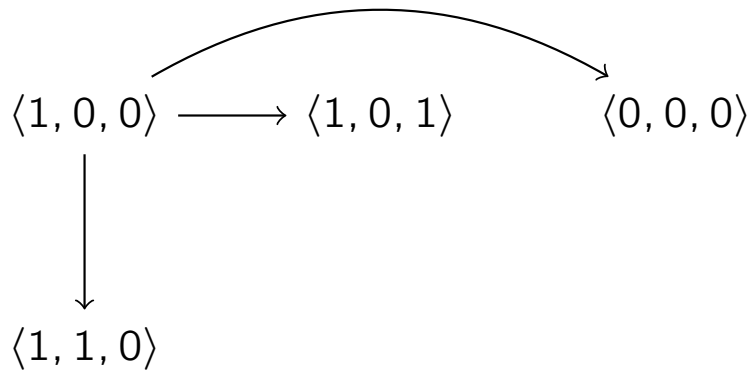
Influence graph (knowledge)



Boolean network

$$f_1(x) = 0$$
$$f_2(x) = x_1 \wedge \neg x_2$$
$$f_3(x) = \neg x_2 \wedge (x_1 \vee x_3)$$

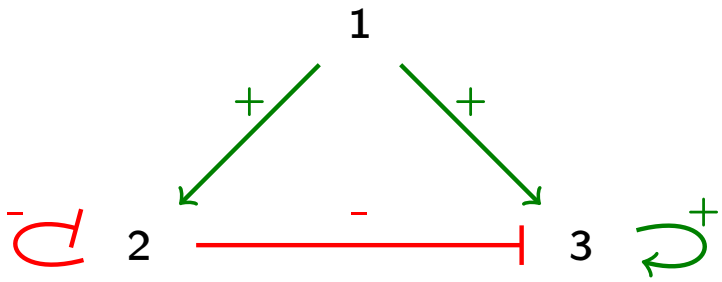
State transition graph with asynchronous updating mode



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

Example

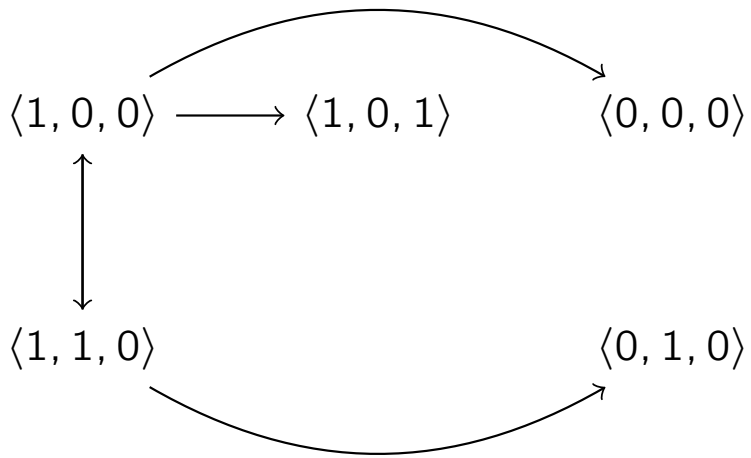
Influence graph (knowledge)



Boolean network

$$f_1(x) = 0$$
$$f_2(x) = x_1 \wedge \neg x_2$$
$$f_3(x) = \neg x_2 \wedge (x_1 \vee x_3)$$

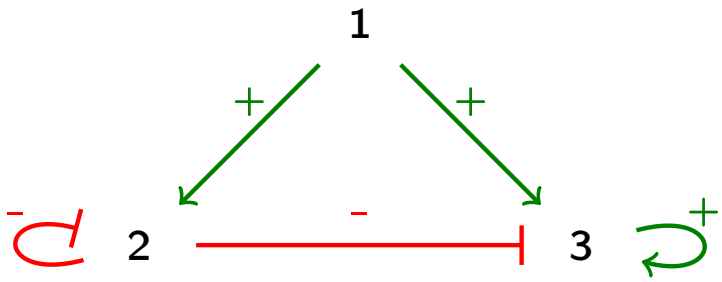
State transition graph with asynchronous updating mode



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

Example

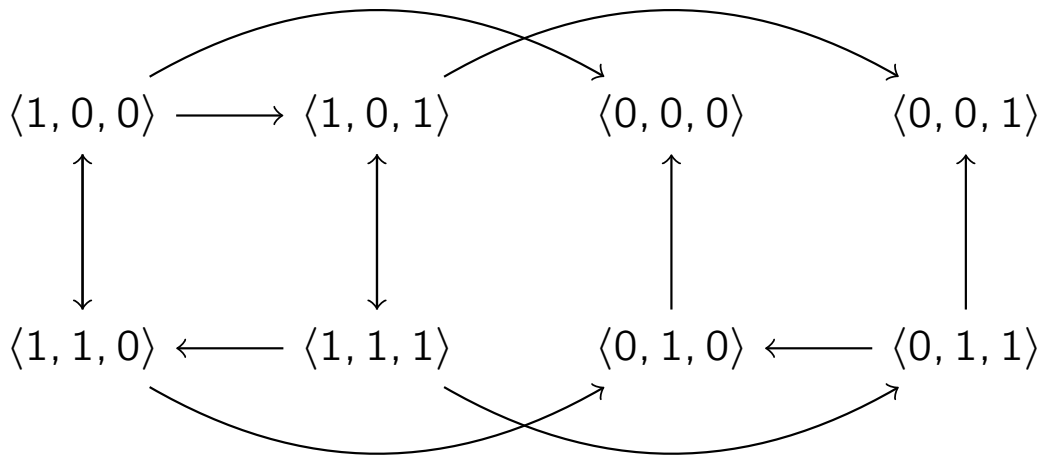
Influence graph (knowledge)



Boolean network

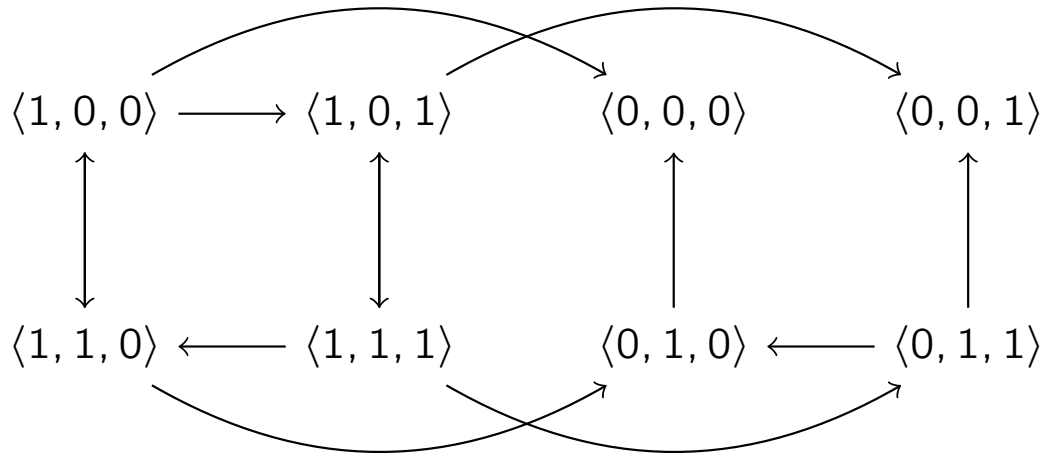
$$f_1(x) = 0$$
$$f_2(x) = x_1 \wedge \neg x_2$$
$$f_3(x) = \neg x_2 \wedge (x_1 \vee x_3)$$

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]

Properties of interest



- Fixpoints: $f(x) = x$
- Attractors: smallest set of states closed by $\rightarrow$ (terminal strongly connected components)
- Reachability: there is a path from x to y

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
- BDDs/BMC/... have a hard time on biological networks...
- Difficult to extract comprehensive proofs of (im)possibility.

(Reachability is PSPACE-complete)

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
- BDDs/BMC/... have a hard time on biological networks...
- Difficult to extract comprehensive proofs of (im)possibility.

(Reachability is PSPACE-complete)

Prediction for control (e.g., perturbations to prevent/enforce behaviour)

- Model checking complexity
- + too many candidates for brute-force screening $\Rightarrow$ 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
- BDDs/BMC/... have a hard time on biological networks...
- Difficult to extract comprehensive proofs of (im)possibility.

(Reachability is PSPACE-complete)

Prediction for control (e.g., perturbations to prevent/enforce behaviour)

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

Network inference (find Boolean networks satisfying reachability constraints)

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

Outline

1 Computational models for biological processes
Networks, dynamics

2 Boolean networks
Definition
Properties of interest

3 Static analysis of Boolean networks by abstract interpretation
Main principle
Over- and under-approximation of trajectories
Applications

4 Perspectives

Abstractions for transient dynamics of Boolean Networks

Intuition: exploit the low scope of transitions (concurrency)

- Static analysis by abstract interpretation [Cousot and Cousot 77]
- Intermediate representation (Local Causality Graph) to reason on necessary/sufficient conditions for transitions
- Implementation mixes algorithms on graphs and SAT (ASP).

Basically:
Approx. of PSPACE problems with P.e^d or NP.e^d problems
where d is the in-degree of nodes in the Boolean network

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 $\mathrm{v}_i : a \rightsquigarrow b$

Definition

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

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

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

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

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

Transitions Prime Implicants

Illustration

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

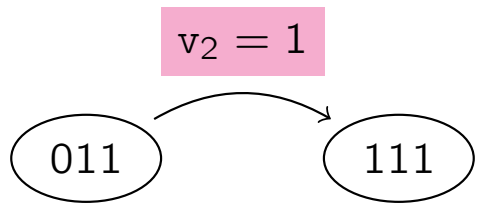
Implicants for the transition $\langle 011 \rangle \rightarrow \langle 111 \rangle$

Transitions Prime Implicants

Illustration

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

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

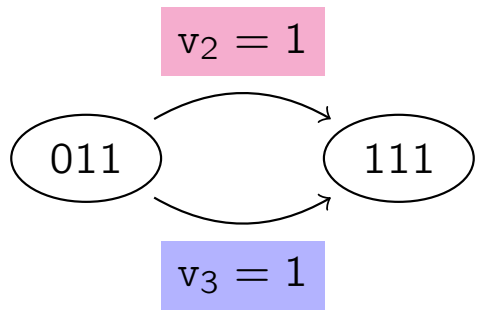


Transitions Prime Implicants

Illustration

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

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

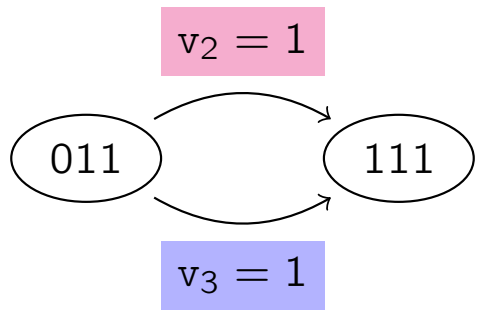


Transitions Prime Implicants

Illustration

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

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



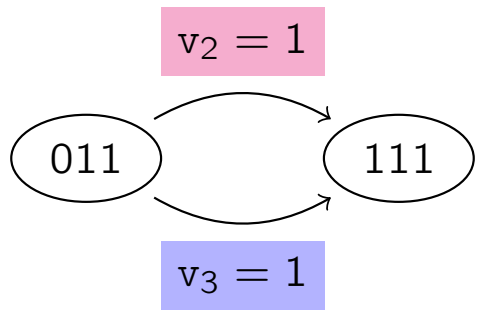
$\Rightarrow$ 2 prime implicants: $\langle v_1 : 0 \rightsquigarrow 1, [\ v_2 = 1\ ] \rangle$; $\langle v_1 : 0 \rightsquigarrow 1, [\ v_3 = 1\ ] \rangle$

Transitions Prime Implicants

Illustration

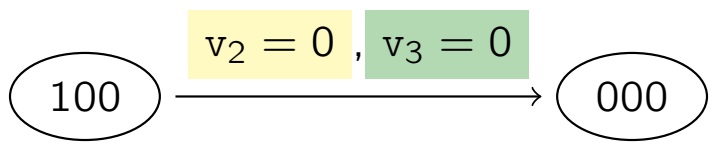
$$f_1(x) = x_2 \vee x_3$$

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



$\Rightarrow$ 2 prime implicants: $\langle v_1 : 0 \rightsquigarrow 1, [\ v_2 = 1\ ] \rangle$; $\langle v_1 : 0 \rightsquigarrow 1, [\ v_3 = 1\ ] \rangle$

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

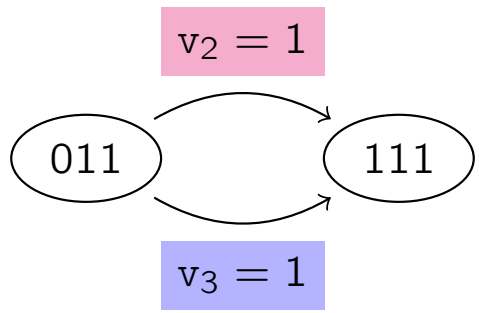


Transitions Prime Implicants

Illustration

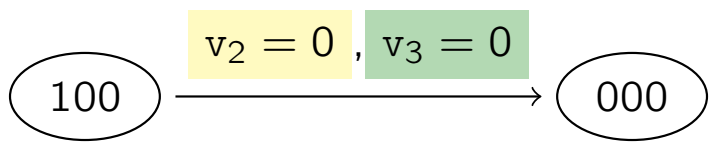
$$f_1(x) = x_2 \vee x_3$$

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



$\Rightarrow$ 2 prime implicants: $\langle v_1 : 0 \rightsquigarrow 1, [\ v_2 = 1\ ] \rangle$; $\langle v_1 : 0 \rightsquigarrow 1, [\ v_3 = 1\ ] \rangle$

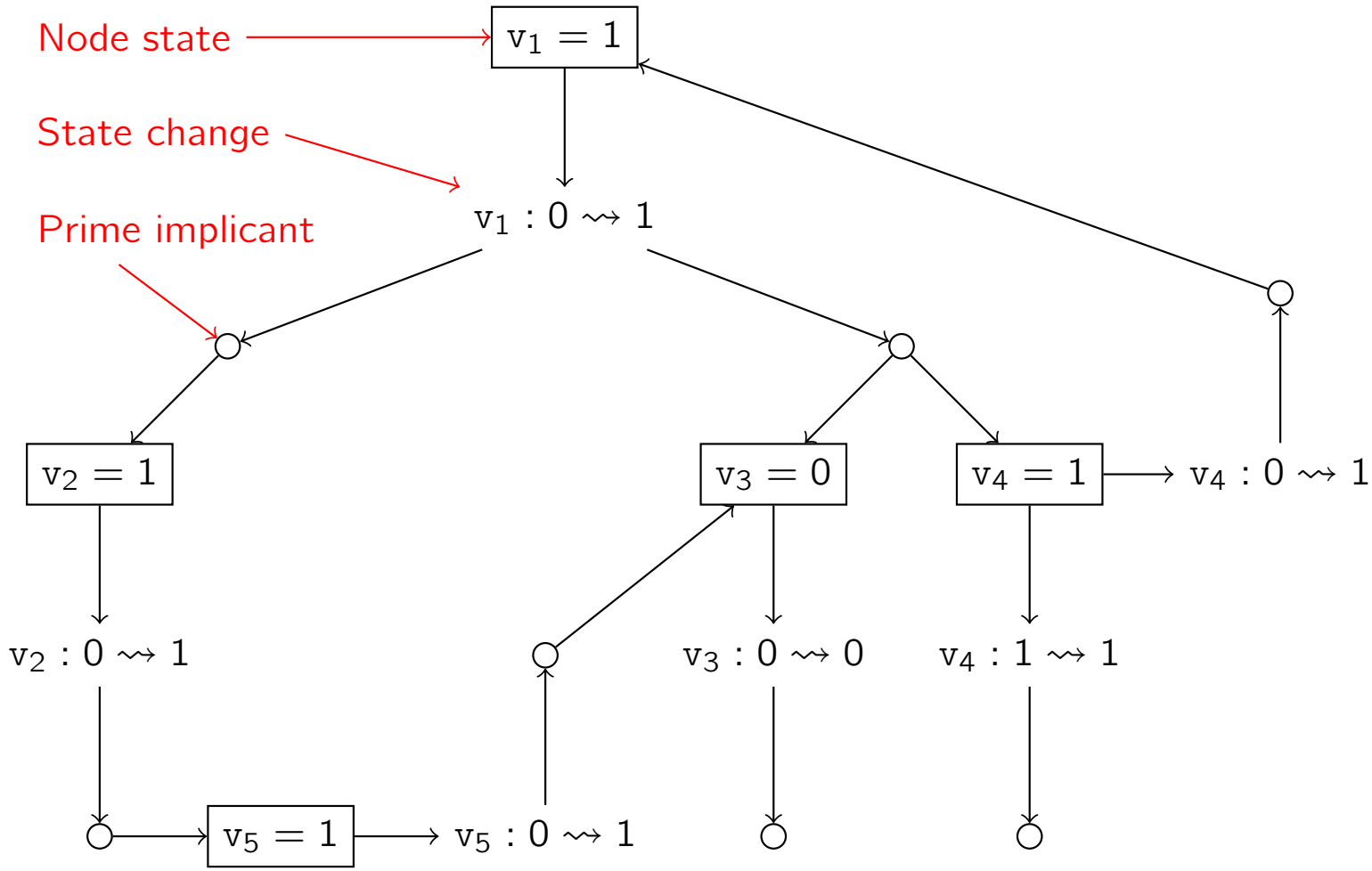
Implicants for the transition $\langle 100 \rangle \to \langle 000 \rangle$



$\Rightarrow$ 1 prime implicant: $\langle v_1 : 1 \rightsquigarrow 0, [\ v_2 = 0 \ \wedge\ v_3 = 0\ ] \rangle$

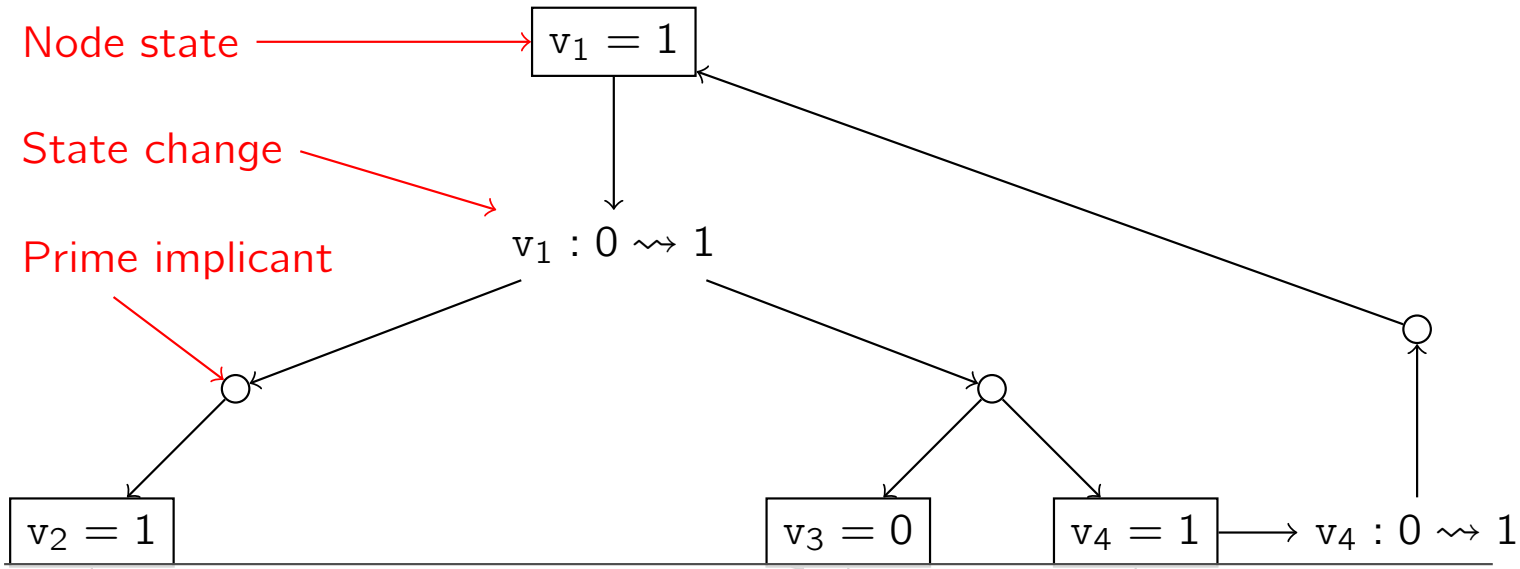
Local Causality Graph (LCG)

- Initial state $\langle 00010 \rangle$; Goal $[v_1 = 1]$



Local Causality Graph (LCG)

- Initial state $\langle 00010 \rangle$; Goal $[v_1 = 1]$



Necessary condition for reachability

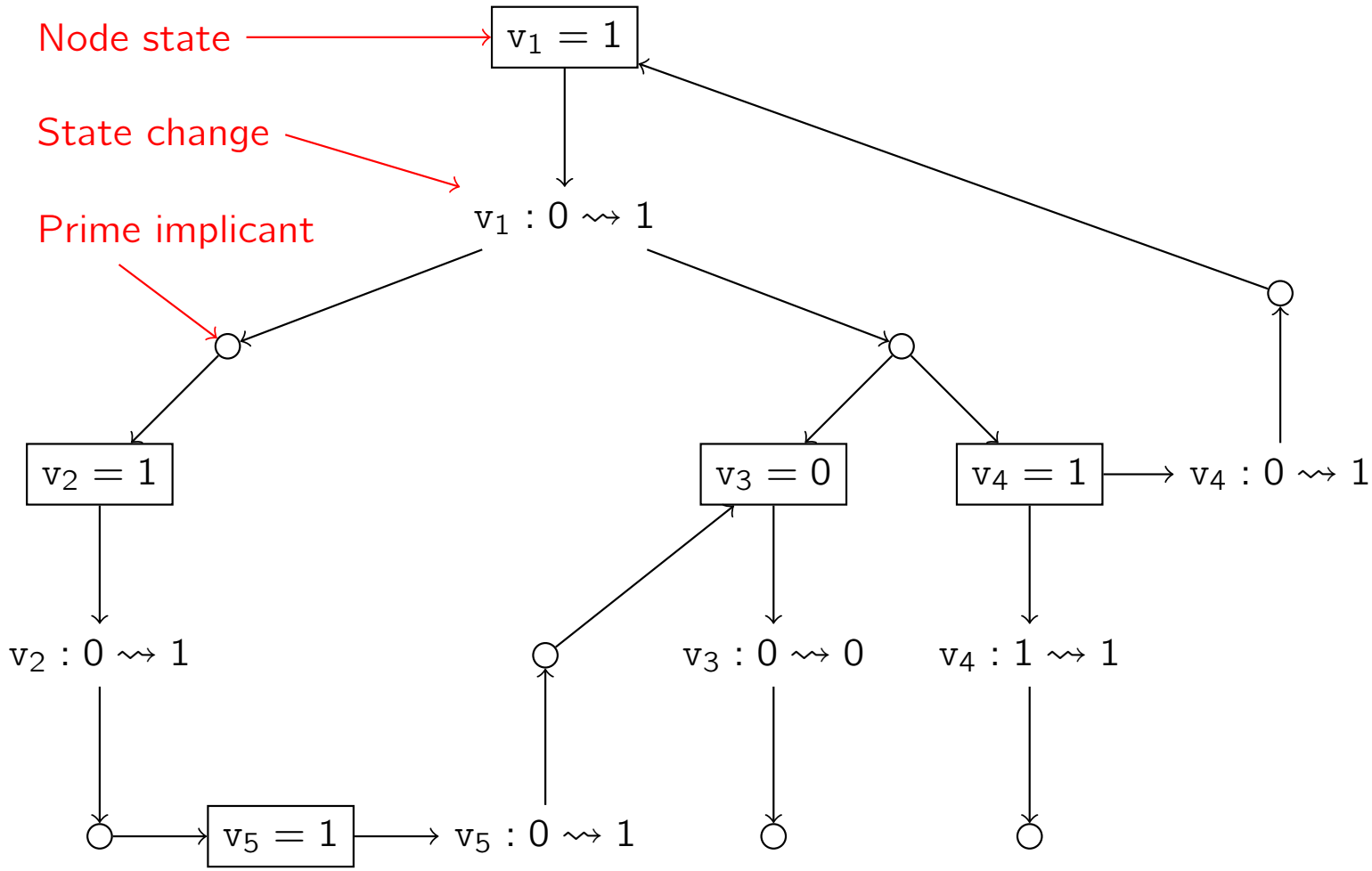
OA($x \rightarrow^* [v_1 = 1]$) $\equiv$ there is an acyclic traversal from $v_1 = 1$ s.t.

- node state change $\rightarrow$ follow at least one child;
- other nodes $\rightarrow$ follow all children;
- terminates on empty "local cause" (leafs).

(can be verified linearly in the size of the LCG).

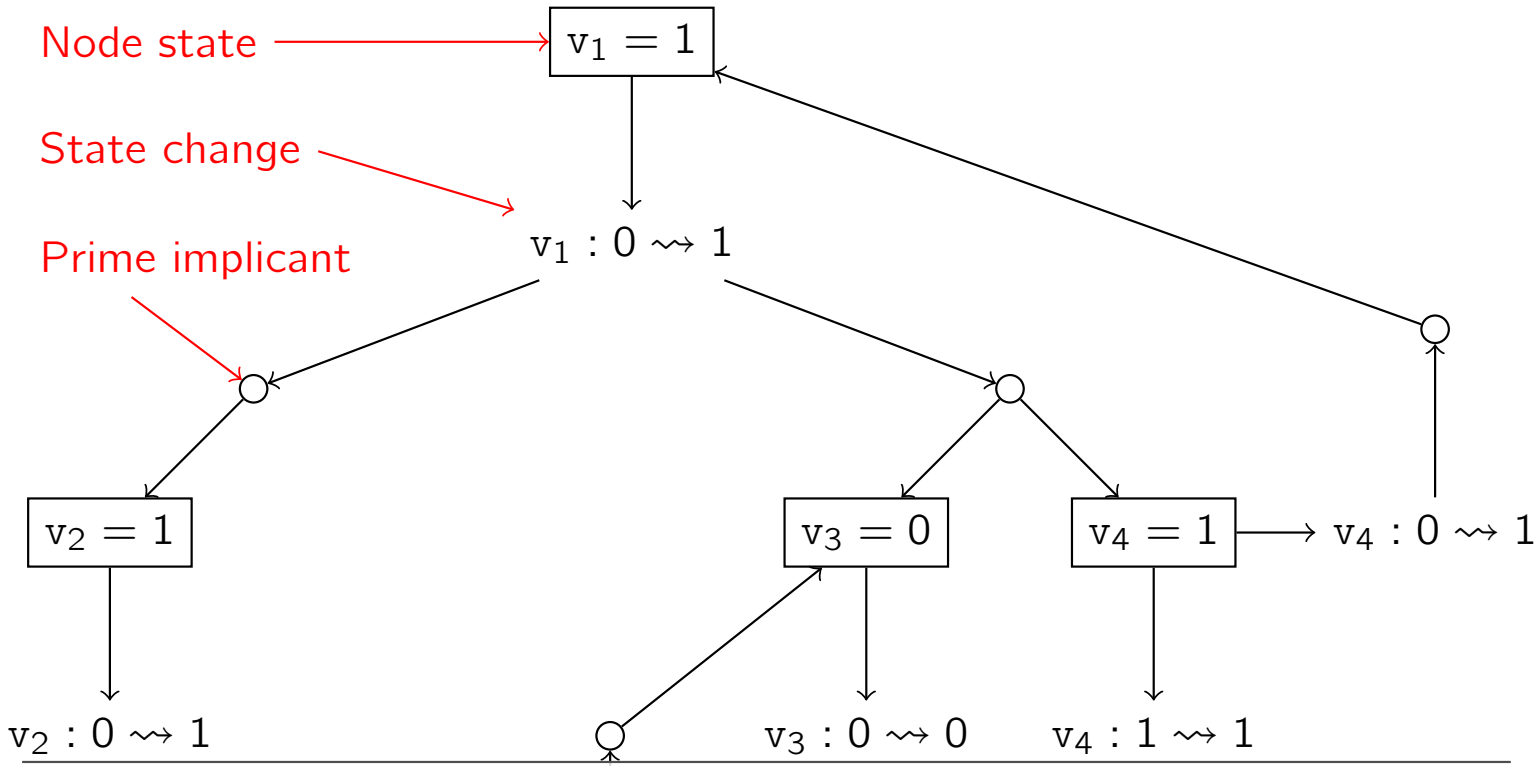
Local Causality Graph (LCG)

- Initial state $\langle 00010 \rangle$; Goal $[v_1 = 1]$



Local Causality Graph (LCG)

- Initial state $\langle 00010 \rangle$; Goal $[v_1 = 1]$



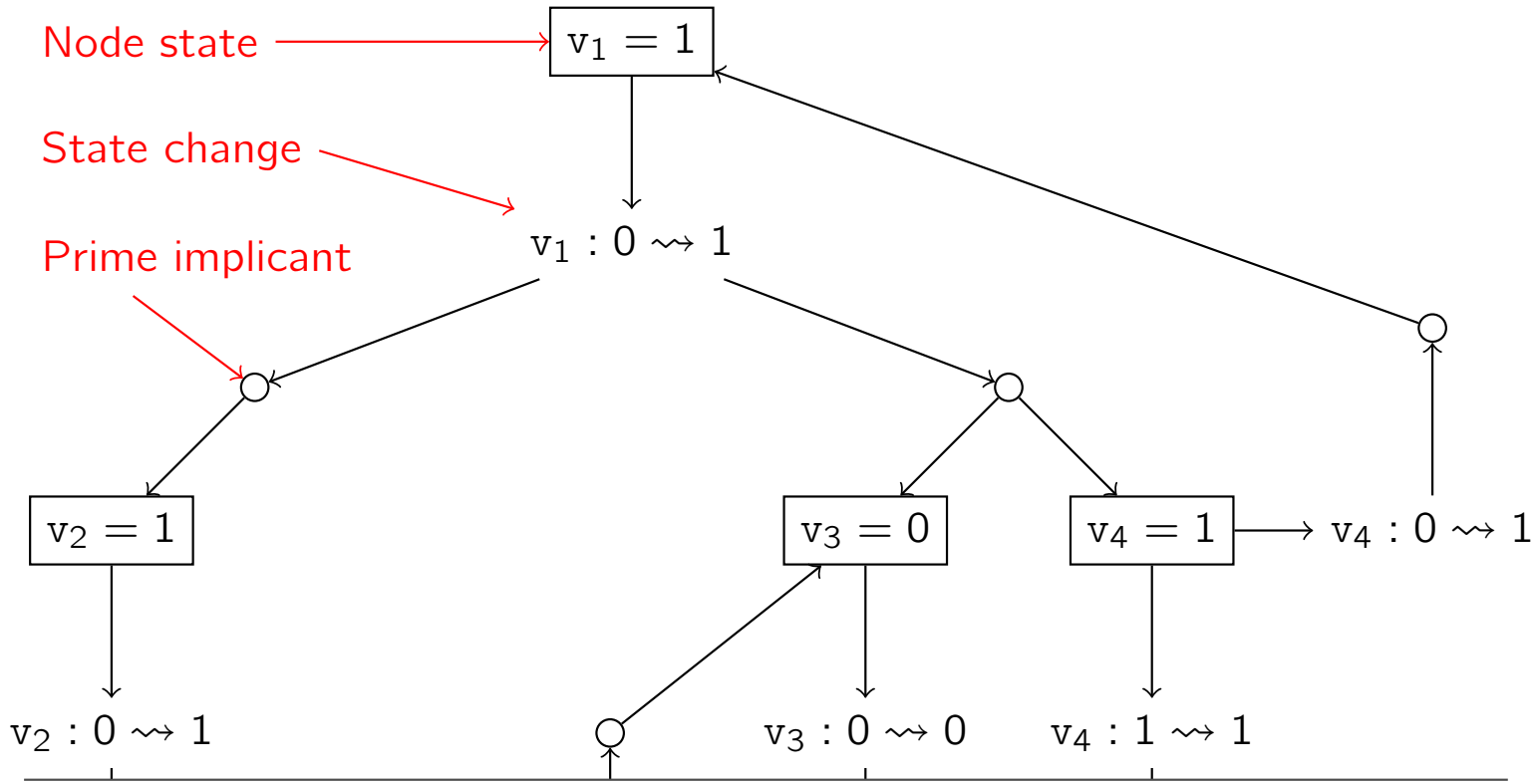
Sufficient condition for reachability

UA($x \rightarrow^* [v_1 = 1]$) $\equiv \exists$ particular acyclic sub-LCG

NP formulation (find the right combination of prime implicants).

Local Causality Graph (LCG)

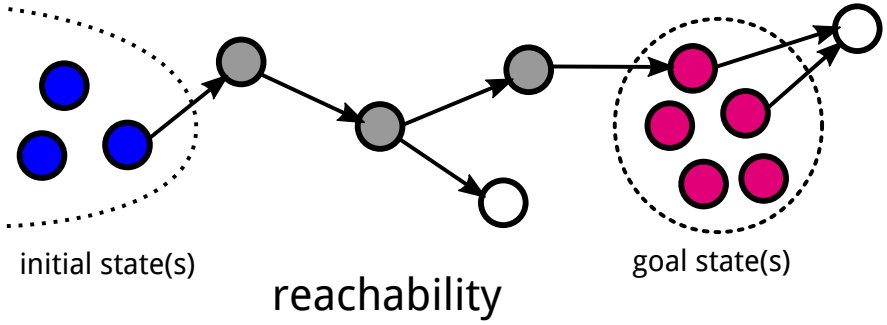
- Initial state $\langle 00010 \rangle$; Goal $[v_1 = 1]$



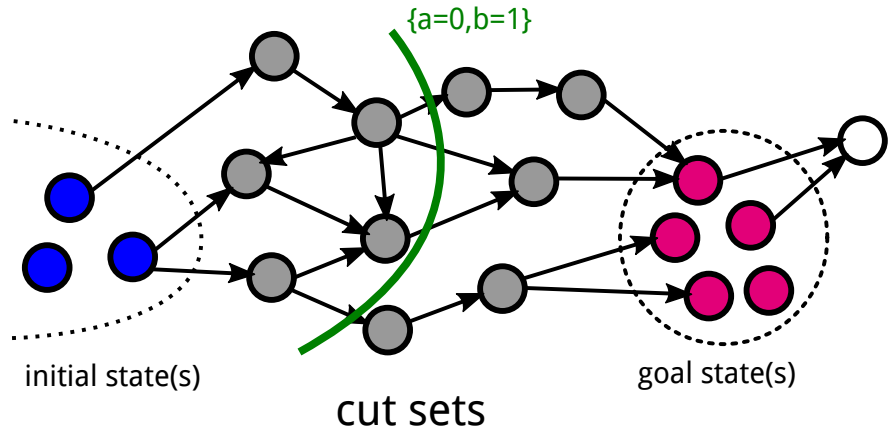
Formal approximations of reachability

$$\mathrm{UA}(x \rightarrow^* [v_1 = 1]) \Rightarrow x \rightarrow^* [v_1 = 1] \Rightarrow \mathrm{OA}(x \rightarrow^* [v_1 = 1])$$

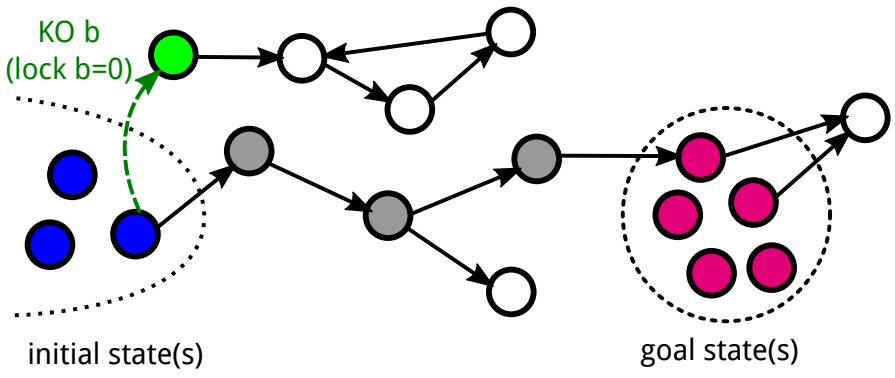
Applications



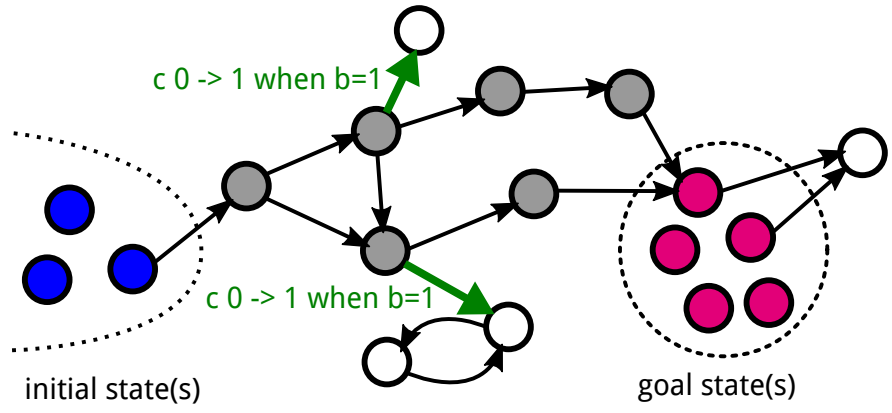
reachability



cut sets

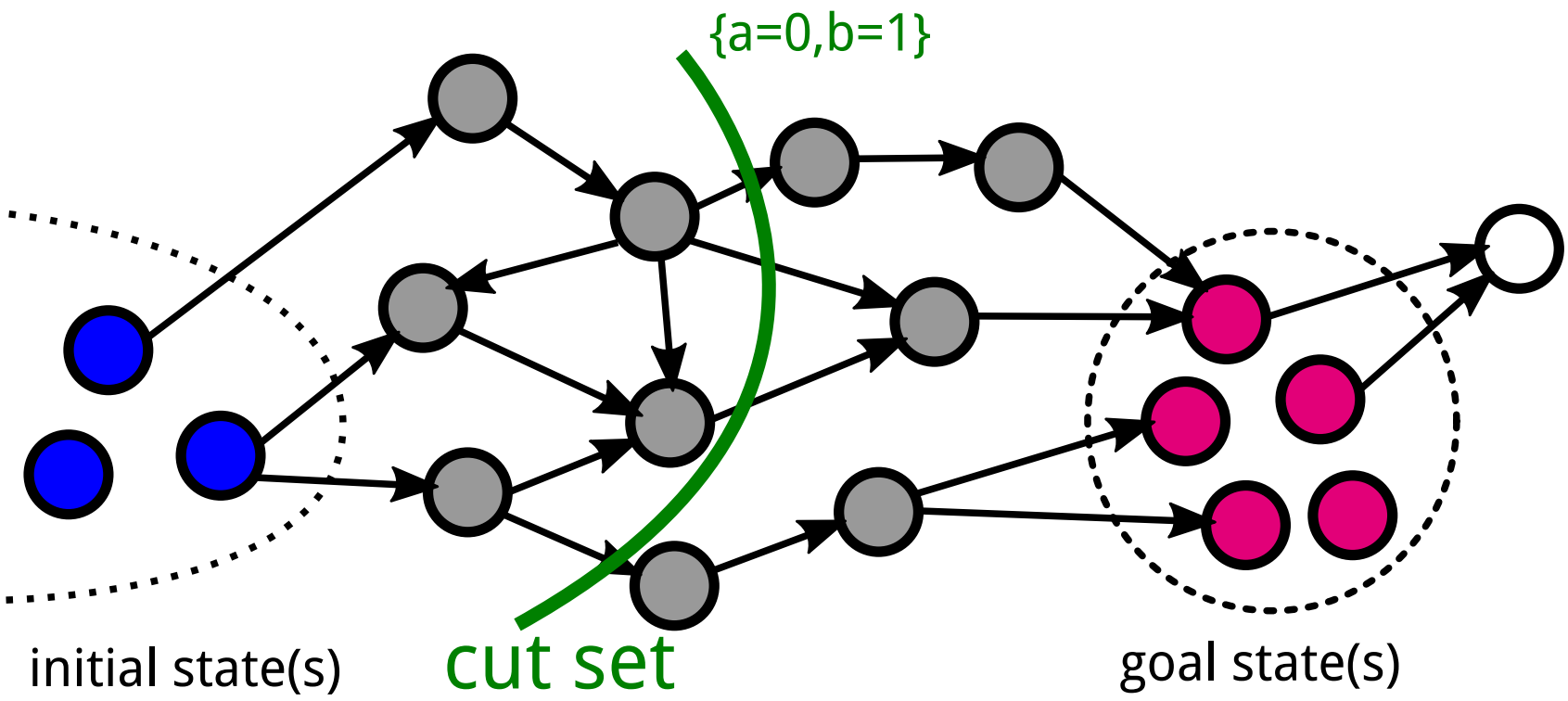


mutations



bifurcations

Applications

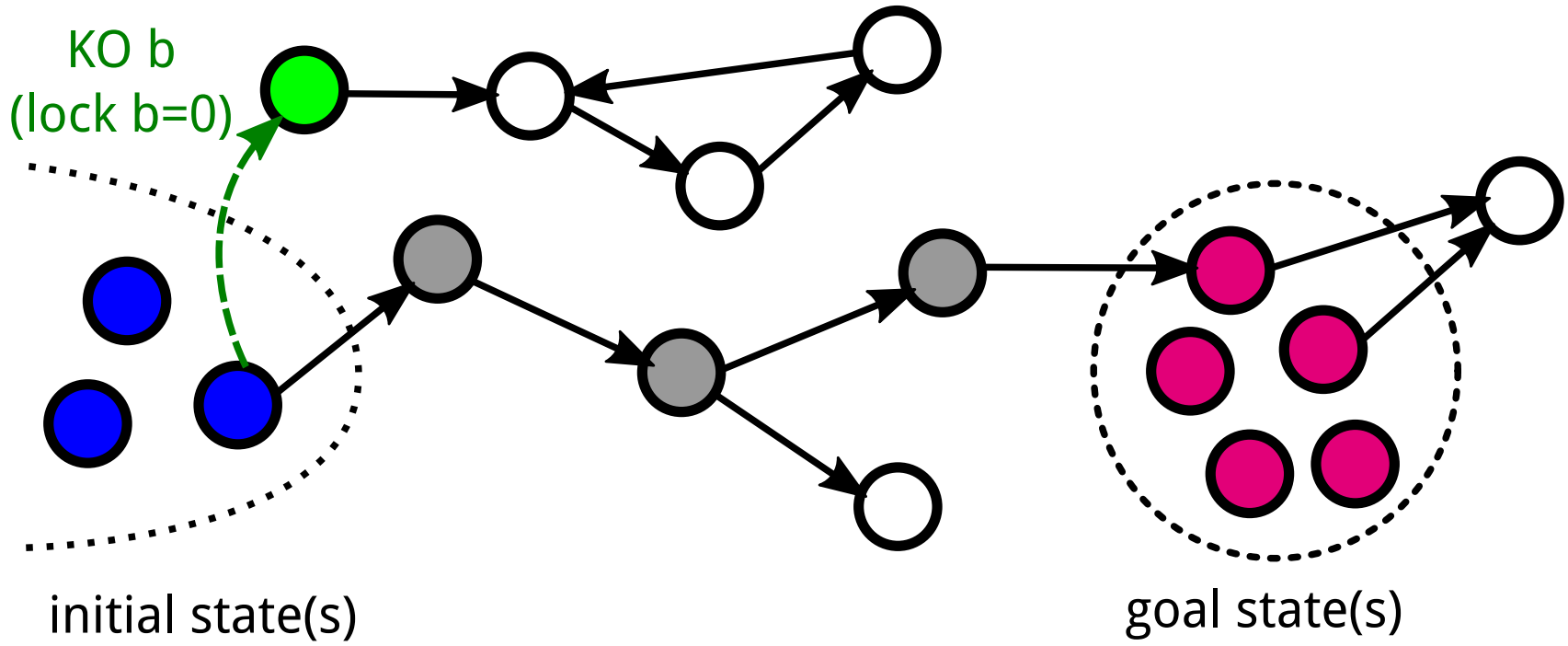


Common features (markers) of **all** trajectories (necessary steps)

Under-approximation:

$$a_i, b_j, \cdots : \textbf{disable}(a_i, b_j, \cdots) \wedge \neg\, \mathrm{OA}(s \rightarrow^* g)$$

Applications



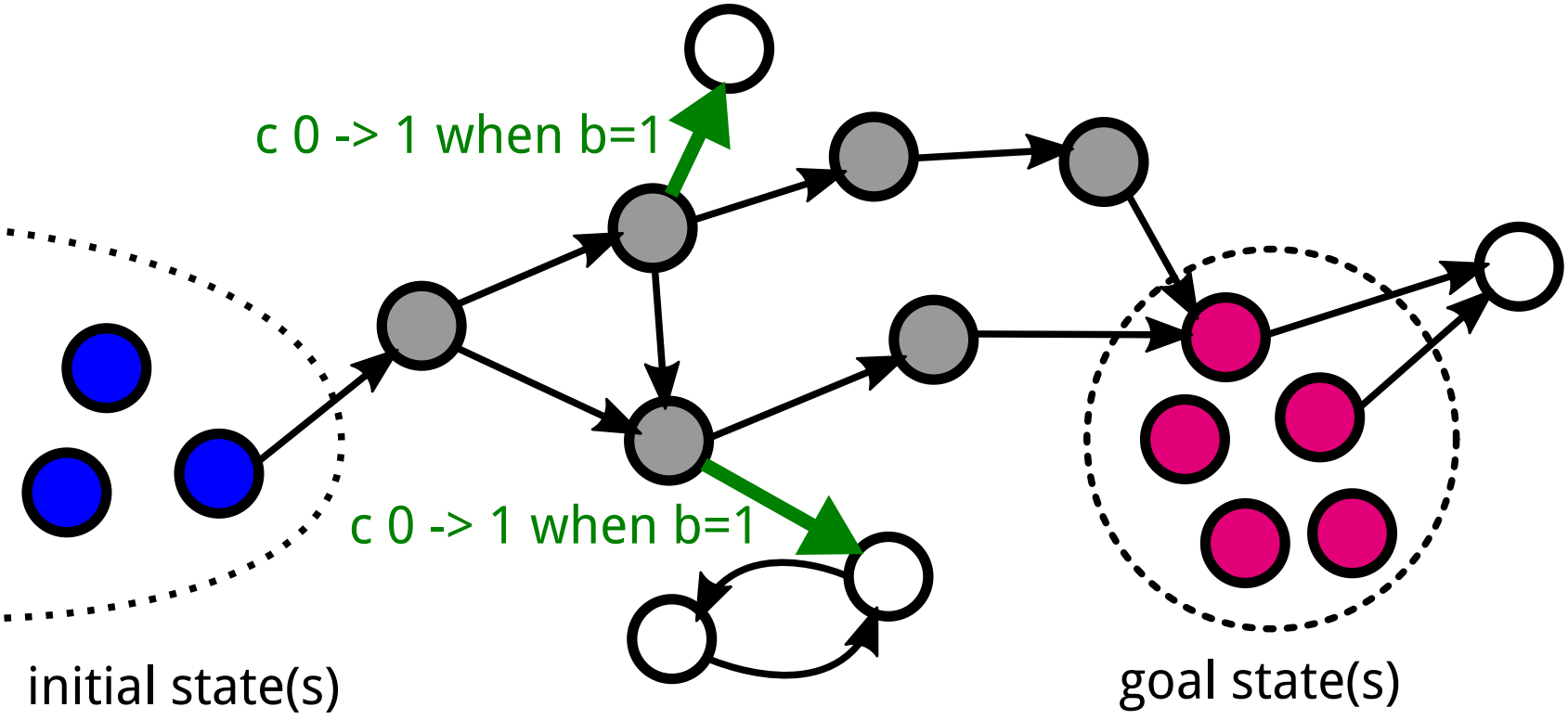
mutations

Control of reachability

Under-approximation:

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

Applications



bifurcations

Key transitions responsible for capability loss (differentiation)

Under-approximation:

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

Formal Approximations

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

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

(# prime implicants: e^d in general; $\binom{d}{d/2}$ monotonous functions; much less in practice)

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

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

Mutations for blocking g [LP at CMSB 2017]

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

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

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

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`

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`

CaspoTS - Boolean network identification from time series data
`https://github.com/pauleve/caspots` [BioSystems 2016]

- Input: influence graph + reachability constraints
- Ouput: all *minimal* Boolean networks that satisfy both constraints
- Answer-Set Programming implementation (over-approximation)
- Scalable to networks up to 50-100 nodes

Outline

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

Outlook

In systems biology

- Need for global networks, but lack of precise knowledge
- $\Rightarrow$ 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 genome scale...

Outlook

In systems biology

- Need for global networks, but lack of precise knowledge
- $\Rightarrow$ 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 genome scale...

BUT

Outlook

In systems biology

- Need for global networks, but lack of precise knowledge
- $\Rightarrow$ 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 genome scale...

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? (model counting...)

Acknowledgements

LSV/Inria Saclay, ENS Cachan

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

Institut Curie

- Laurence Calzone
- Andrei Zinovyev

Univ. Luxembourg

- Thomas Sauter
- Lasse Sinkonnen
- Julia Becker
- Jun Pang

LCSB, Luxembourg

- Antonio del Sol
- Andras Hartmann
- Andrzej Mizera
- Sacha Zickenrott

Masaryk University

- David Šafránek

IRISA, Rennes Dyliss

- Anne Siegel

LS2N, Nantes MeForBio

- Olivier Roux
- Morgan Magnin
- Maxime Folschette
- Louis Fippo Fitime

Funding

cnrs ANR Luxembourg National Research Fund