Capturing Dynamics of Large Interaction Networks using Causality Analysis

Loïc Paulevé

CNRS / LRI, Université Paris-Sud, France (Bioinfo team)
loic.pauleve@lri.fr
http://loicpauleve.name

June 16, 2014 – IGM, Université Paris-Sud, France

My Research Topics

Main: **Formal Characterisation of Dynamics of Biological Networks**
Automatic reasoning on dynamics of models of interactions

- Focus on very large interaction networks.
- PhD thesis *(O. Roux and M. Magnin, IRCCyN, Nantes).*

Link between Topology and Dynamics of Biological Networks
How the static network constrains the dynamics?

- Boolean networks, discrete Bio-chemical Reaction Networks.
- *A. Richard (CNRS/I3S, Nice), G. Craciun (Univ. Wisconsin).*

Stochastic and Spatial Simulation
Efficient execution of models

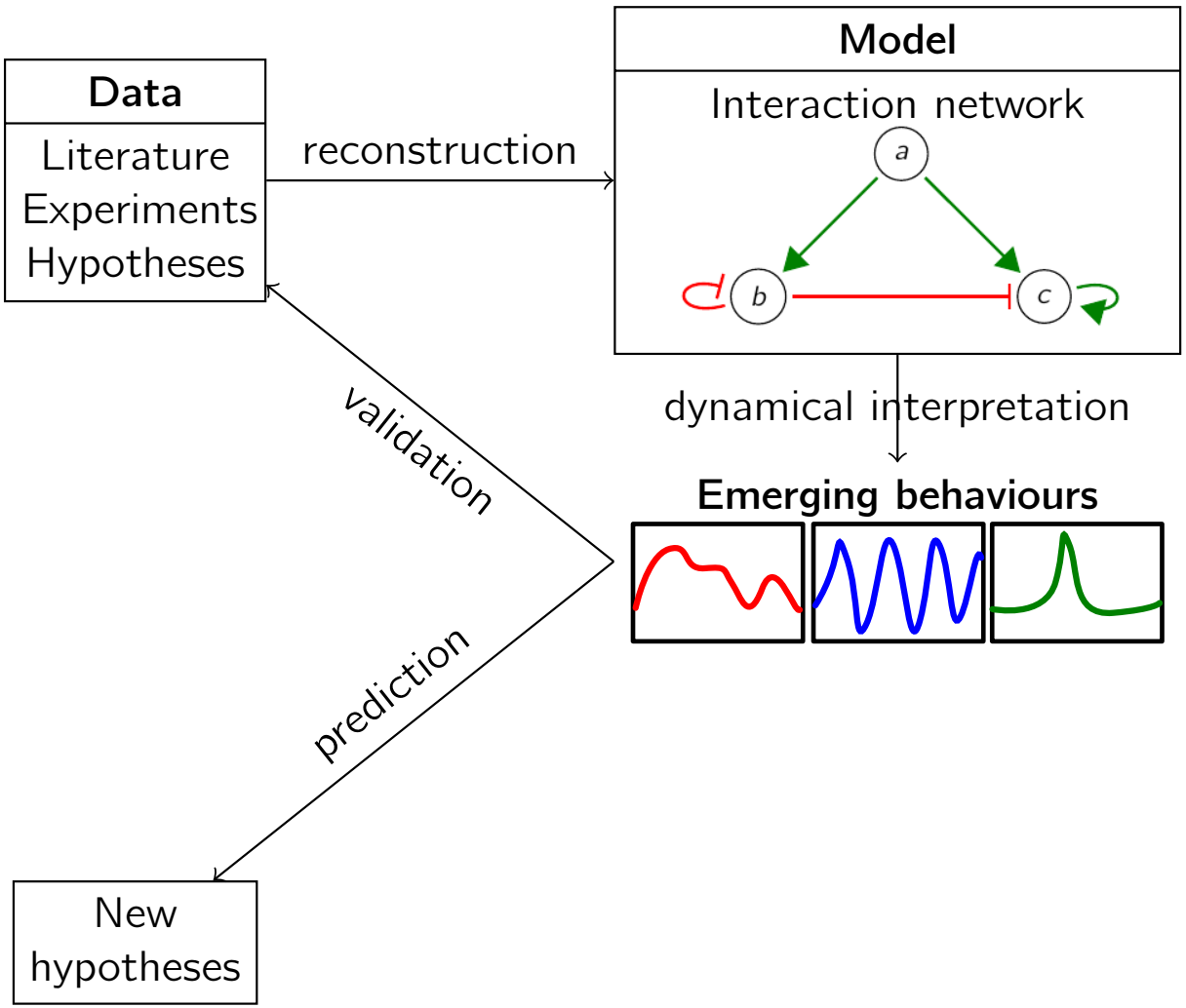
- Formal languages, approximate algorithms, spatial rule-based models.
- *A. Phillips (MSR Cambridge), P. Amar (LRI), M. Klann, T. Petrov and H. Koeppl (ETHZ).*

Quantitative Modelling
Encoding biological interactions

- Parameters inference from single-cell data; stochastic and continuous models.
- *L. Schwartz (Garches hospital) + Bio-info@LIX;*
  C. Zechner, H. Koeppl (ETHZ) and D. Larson (CCR, USA).

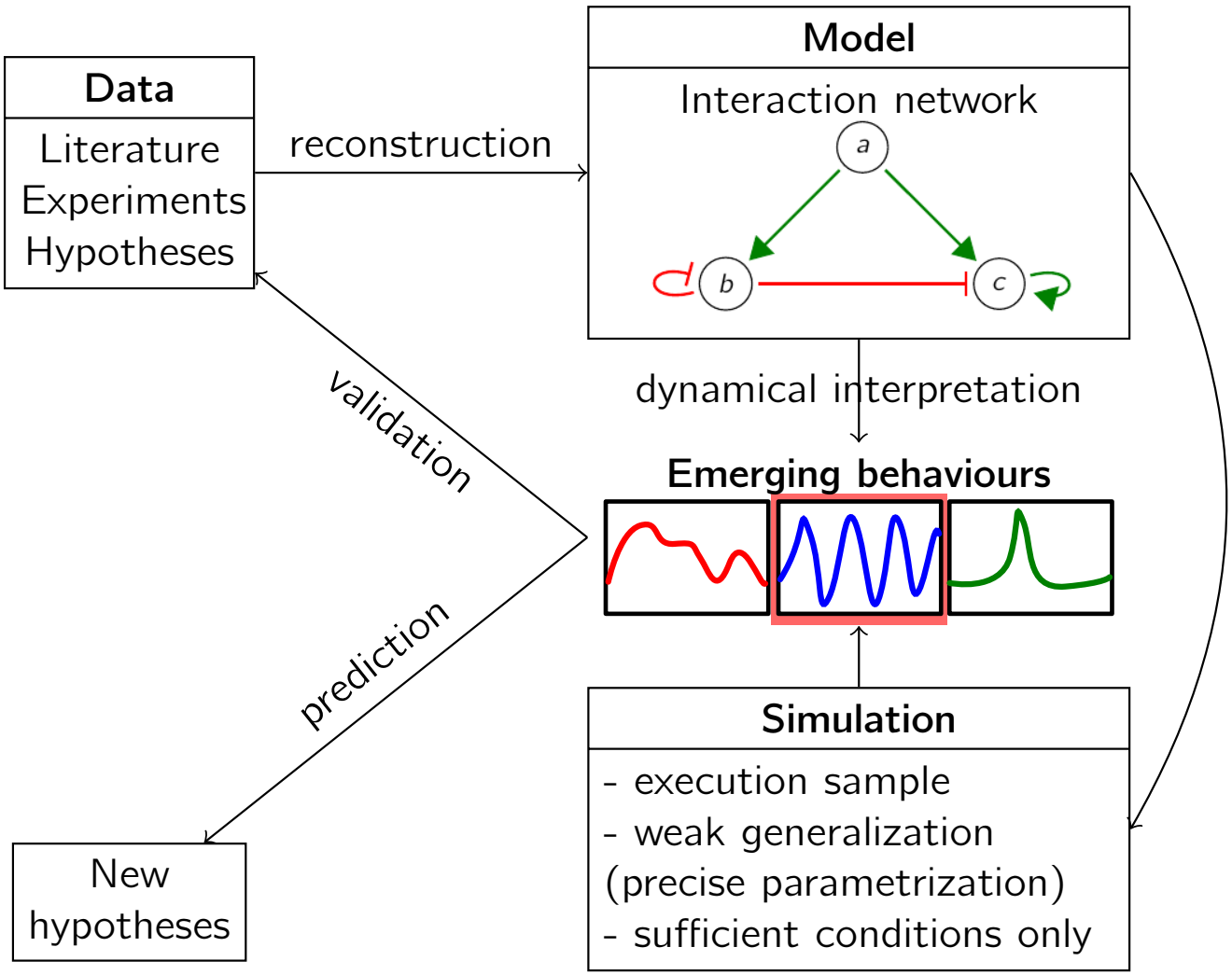
Formal Methods for Systems Biology

Aim: **understand, analyse, control** emerging dynamics.



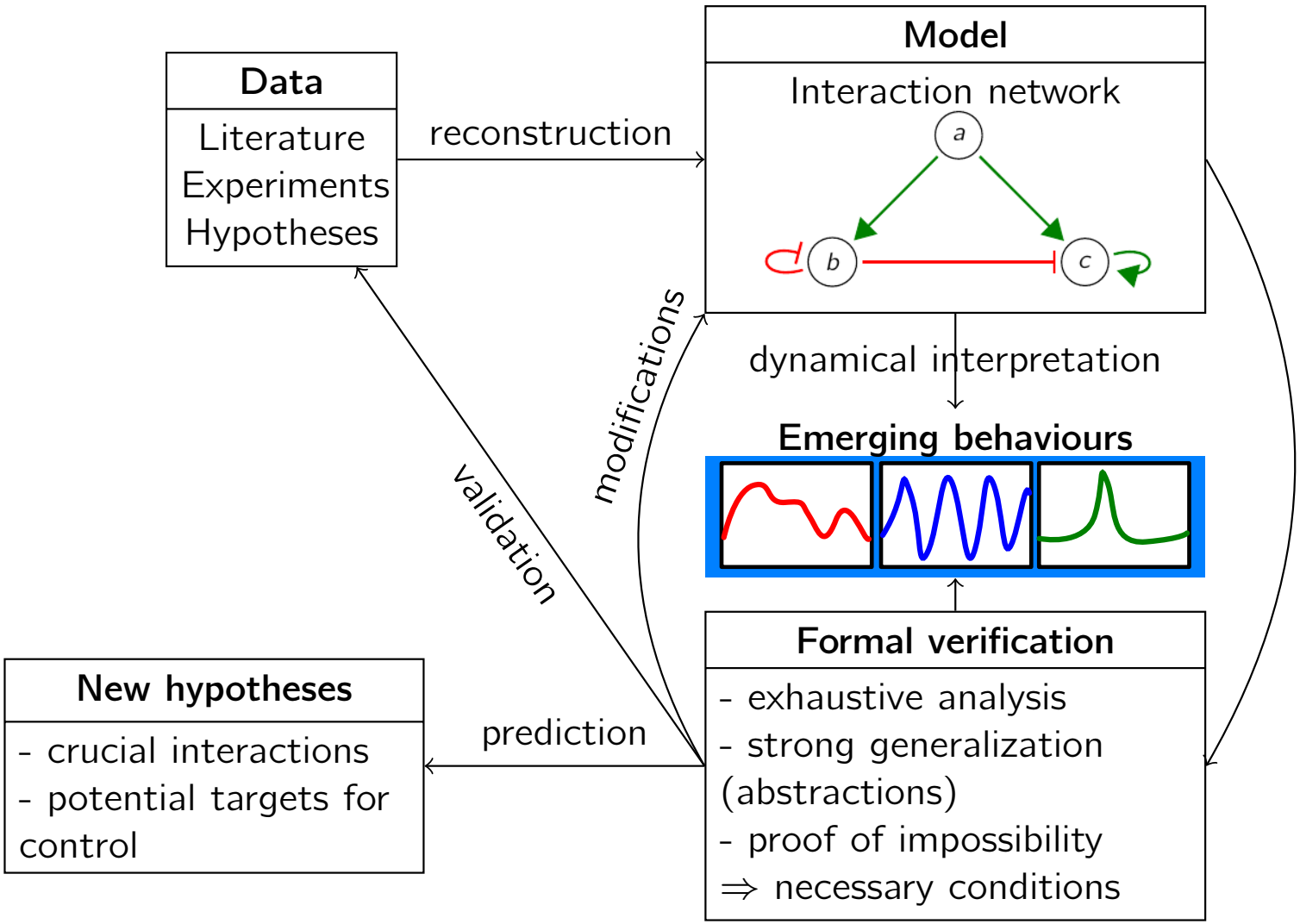
Formal Methods for Systems Biology

Aim: **understand, analyse, control** emerging dynamics.



Formal Methods for Systems Biology

Aim: **understand, analyse, control** emerging dynamics.

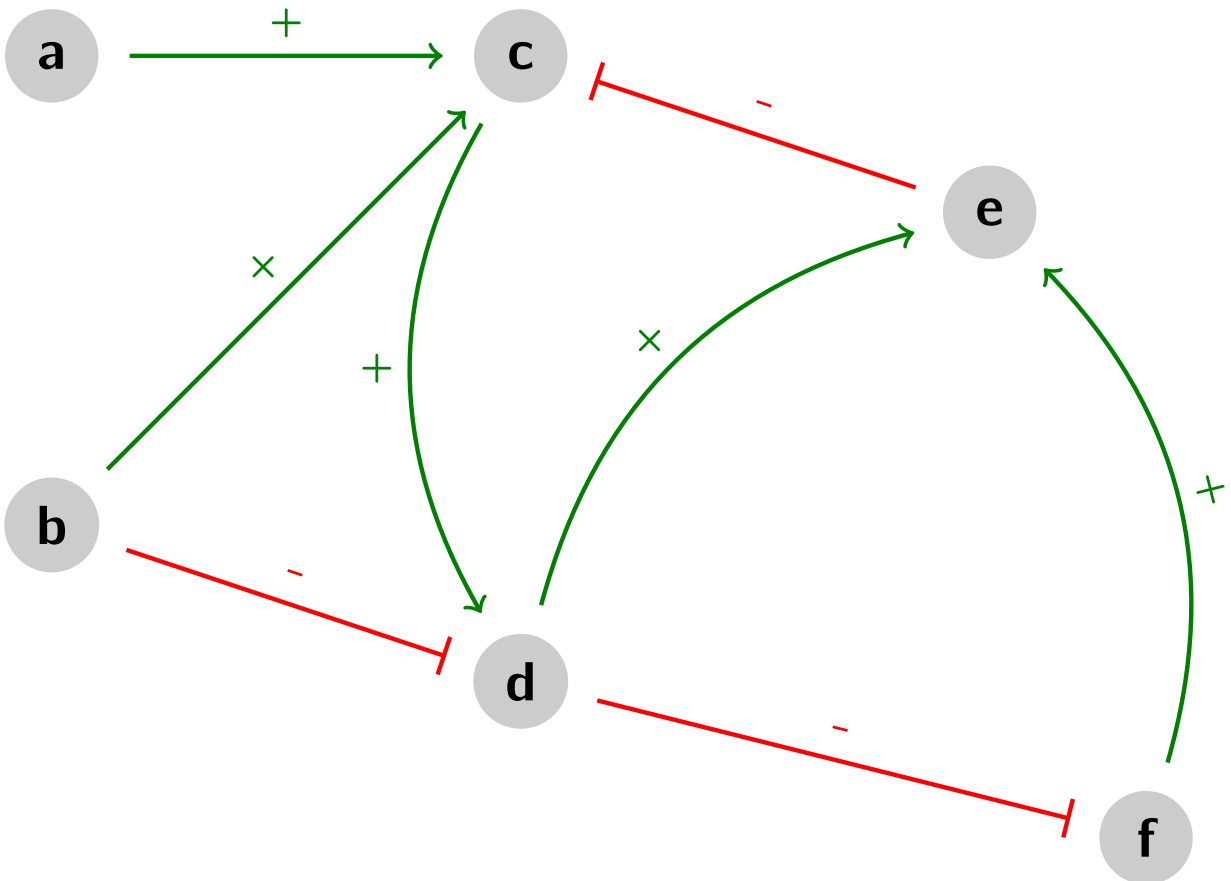


Outline

1. Interaction Networks
   - Qualitative Dynamics
   - Topology vs Dynamics
   - Automata Networks
2. Causality Analysis for Dynamics of Automata Networks
   - Abstract Interpretation
   - Graph of Local Causality
   - Reachability
   - Cut Sets
3. Discussion
   - Software
   - Conclusion, Future Work

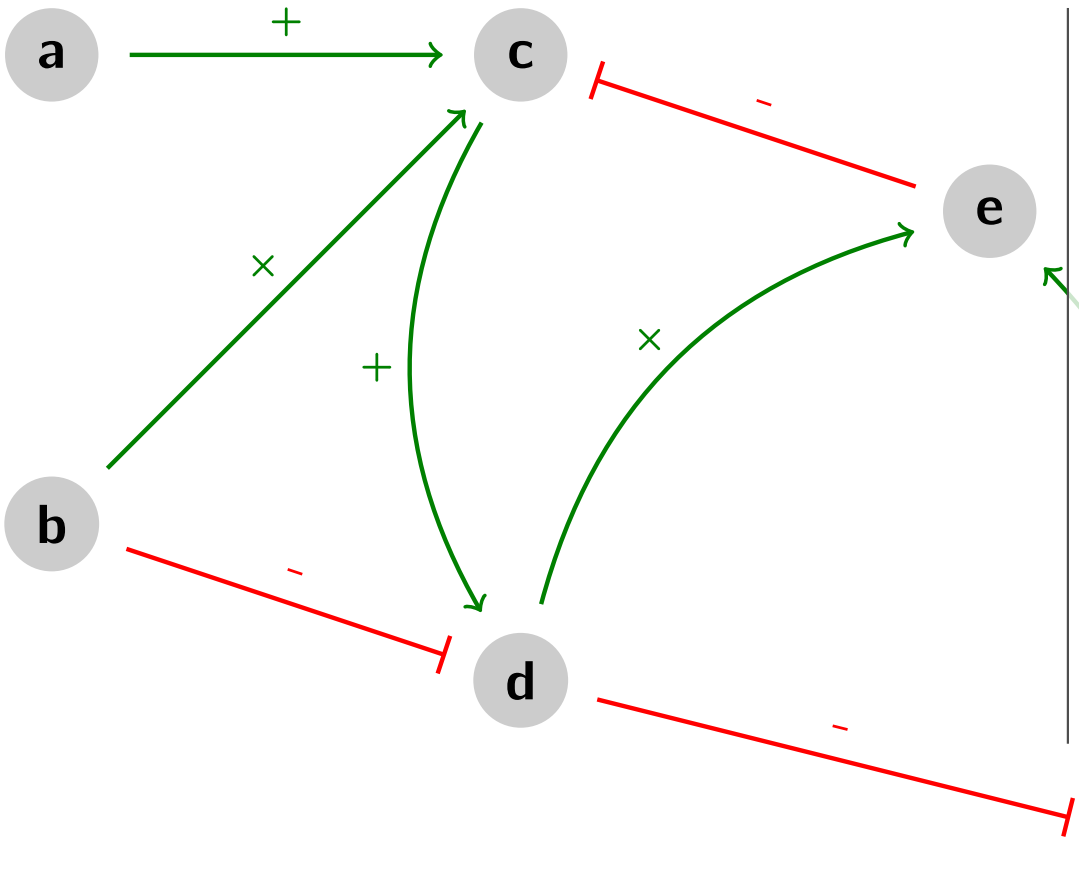
Interaction Networks

E.g., Signalling Networks, Gene Regulatory Networks



Interaction Networks

E.g., Signalling Networks, Gene Regulatory Networks



Components

Genes, proteins, complexes, ...

Activations

Positive influence (*a* increase may promote *c* increase).

Inhibitions

Negative influence (*b* increase may promote *d* decrease).

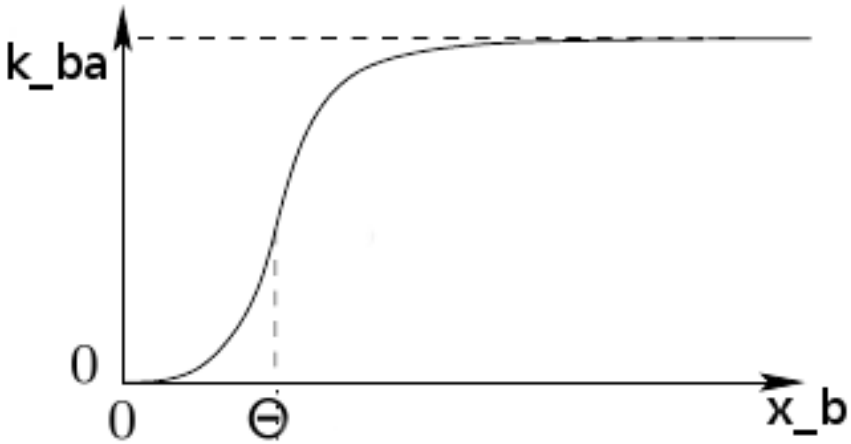
Typical settings

100 to +10,000 components

Few information on kinetics

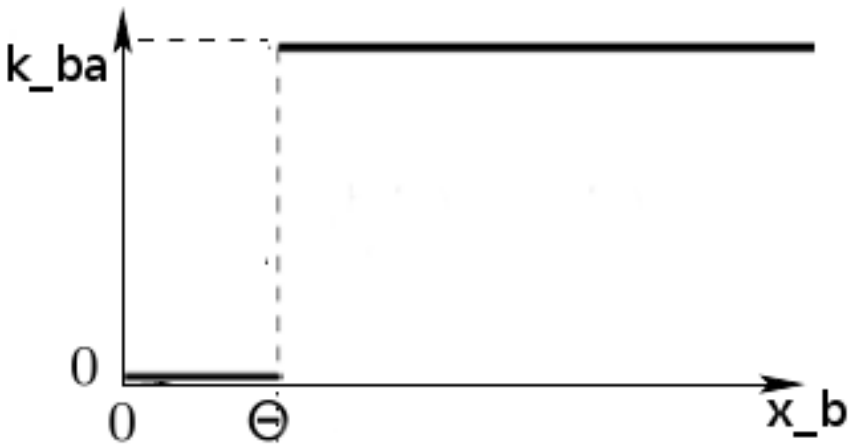
Classical Settings for Qualitative Dynamics

Continuous settings



$$\dot{x_a} = \cdots + k_{ba} s^{\alpha_{ba}}(x_b, \theta_{ba}) - \cdots$$

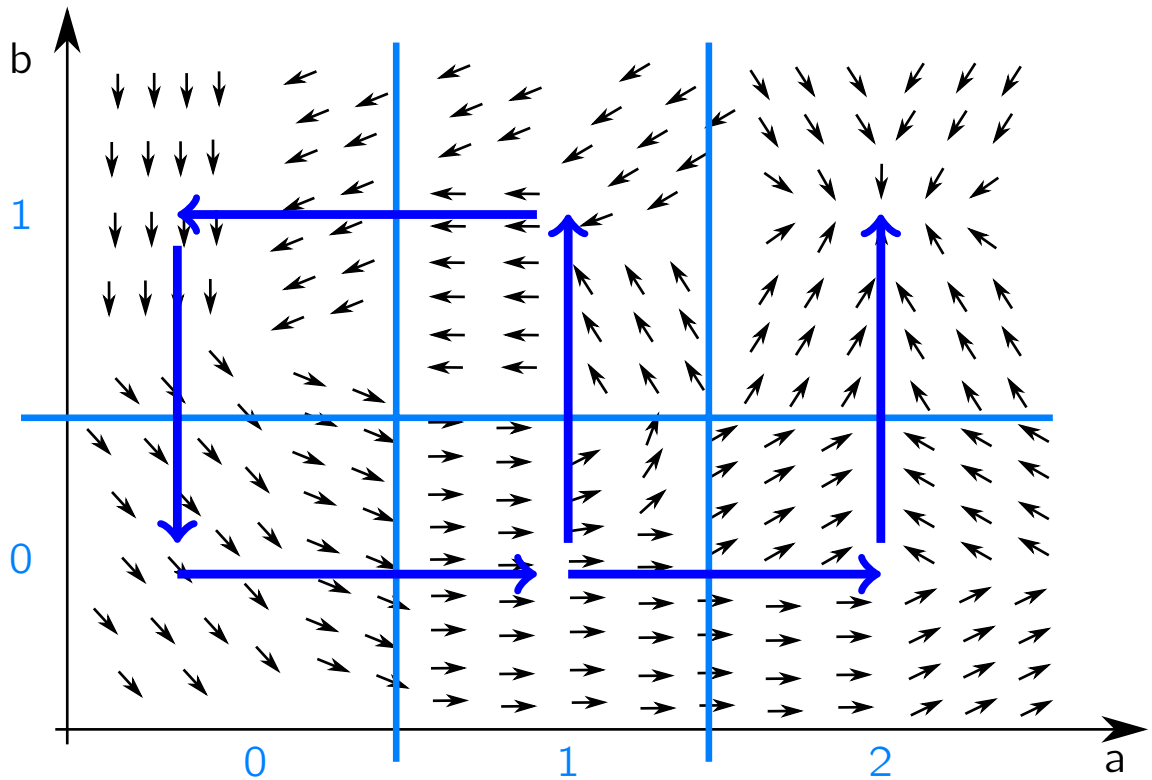
Qualitative settings



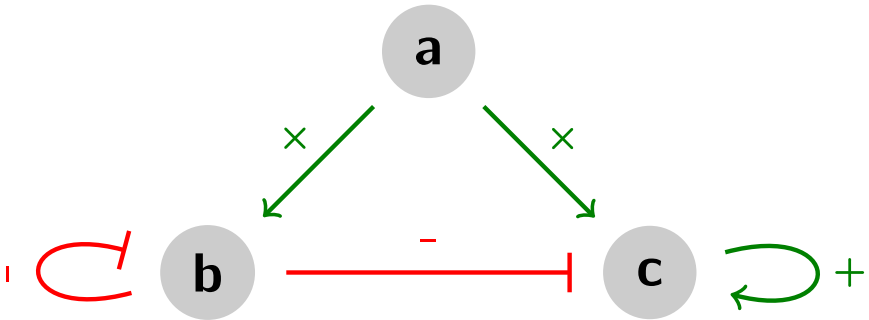
$$f_a(x) = \begin{cases} 1 & \text{if } x_b = 1 \\ 0 & \text{if } x_b = 0 \end{cases}$$

Qualitative Networks

- Assume a quantization of the species population/concentration.
- Have a finite discrete state space (typically 2^n states).
- Non-deterministic dynamics.



Dynamics of Qualitative Networks

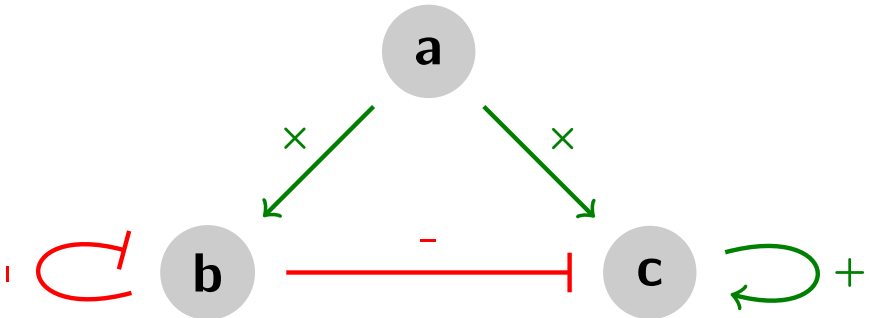


$$f^a(x) = 0$$
$$f^b(x) = x[a] \wedge \neg x[b]$$
$$f^c(x) = \neg x[b] \wedge (x[a] \vee x[c])$$

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

[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

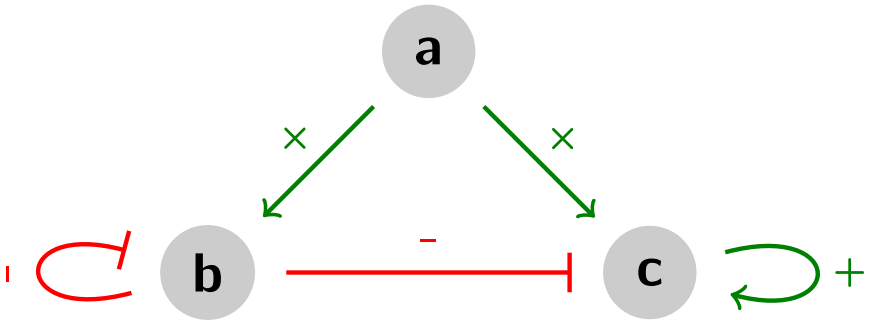


$$f^a(x) = 0$$
$$f^b(x) = x[a] \wedge \neg x[b]$$
$$f^c(x) = \neg x[b] \wedge (x[a] \vee x[c])$$

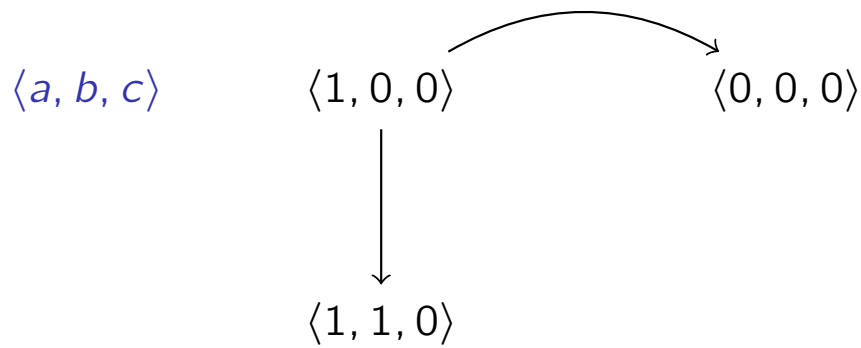
$\langle a, b, c \rangle$ $\quad\quad$ $\langle 1, 0, 0 \rangle$ $\quad\quad$ $\langle 0, 0, 0 \rangle$

[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

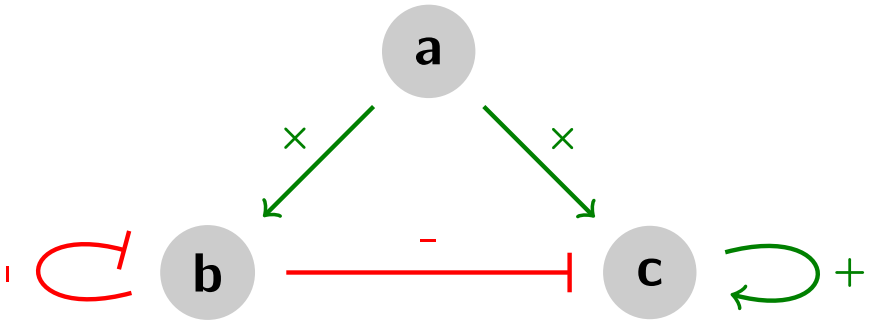


$$f^a(x) = 0$$
$$f^b(x) = x[a] \wedge \neg x[b]$$
$$f^c(x) = \neg x[b] \wedge (x[a] \vee x[c])$$

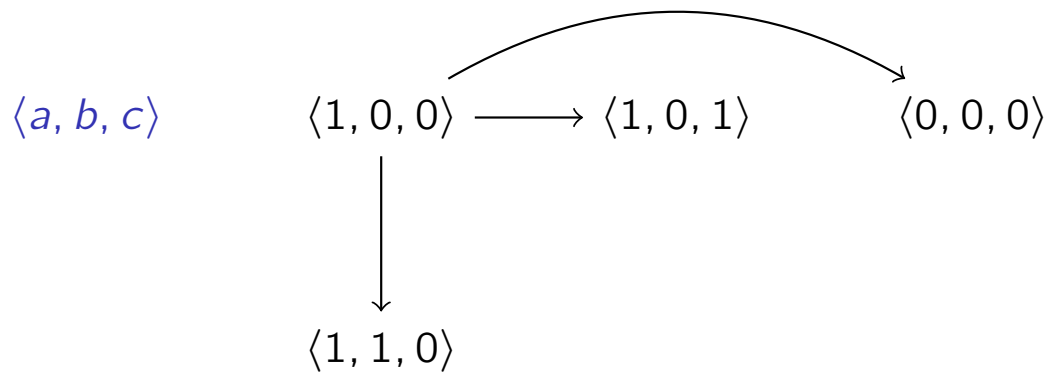


[René Thomas in Journal of Theoritical Biology, 1973] [A. Richard, J.-P. Comet, G. Bernot in Modern Formal Methods and Applications, 2006]

Dynamics of Qualitative Networks

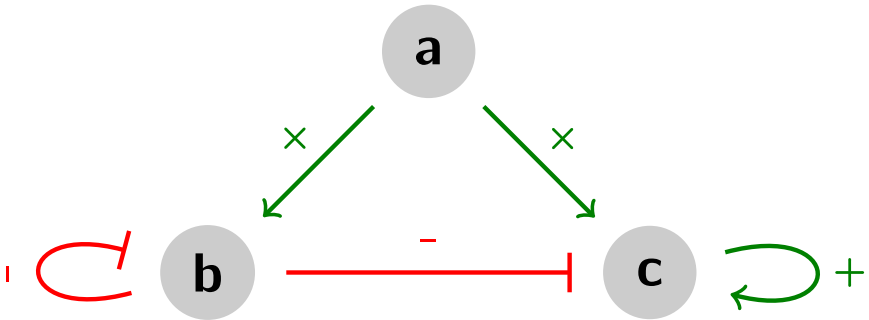


$$f^a(x) = 0$$
$$f^b(x) = x[a] \wedge \neg x[b]$$
$$f^c(x) = \neg x[b] \wedge (x[a] \vee x[c])$$

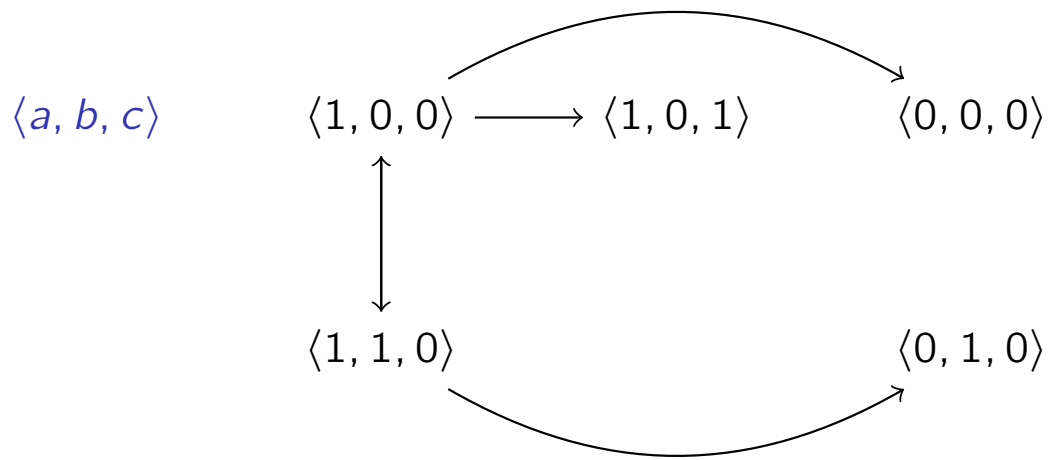


[René Thomas in Journal of Theoritical Biology, 1973] [A. Richard, J.-P. Comet, G. Bernot in Modern Formal Methods and Applications, 2006]

Dynamics of Qualitative Networks

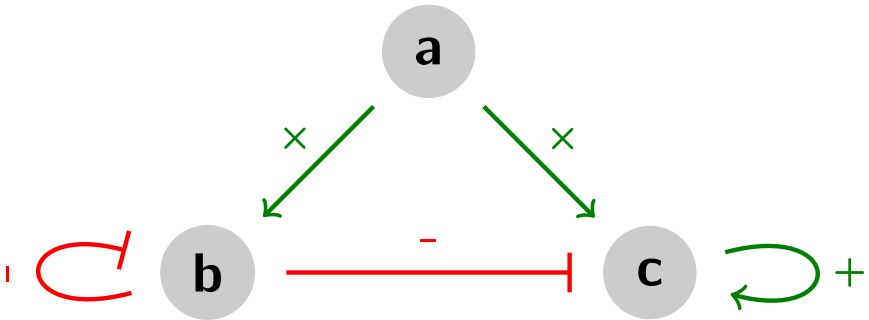


$$f^a(x) = 0$$
$$f^b(x) = x[a] \wedge \neg x[b]$$
$$f^c(x) = \neg x[b] \wedge (x[a] \vee x[c])$$

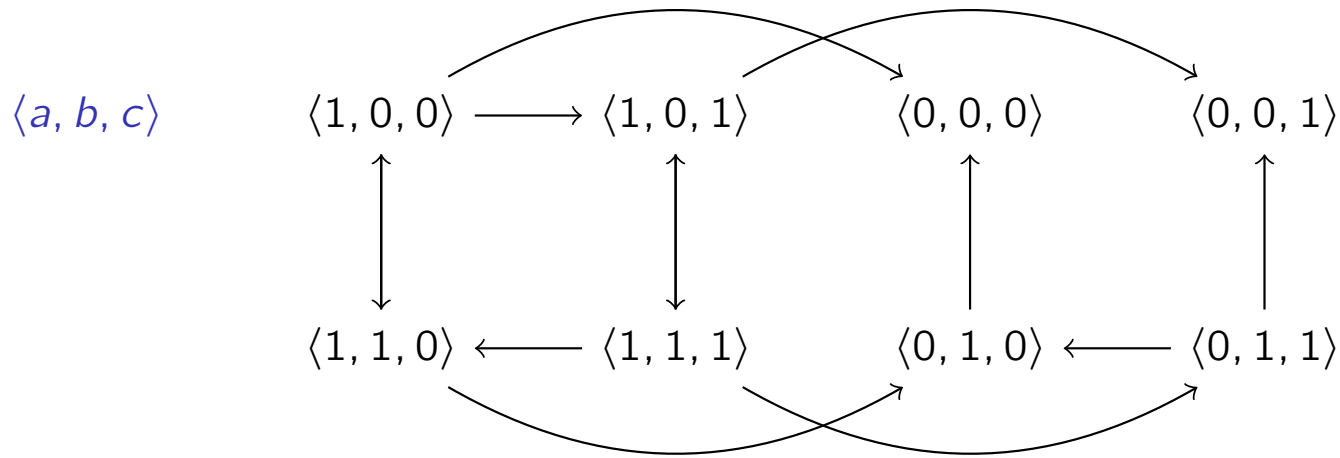


[René Thomas in Journal of Theoritical Biology, 1973] [A. Richard, J.-P. Comet, G. Bernot in Modern Formal Methods and Applications, 2006]

Dynamics of Qualitative Networks

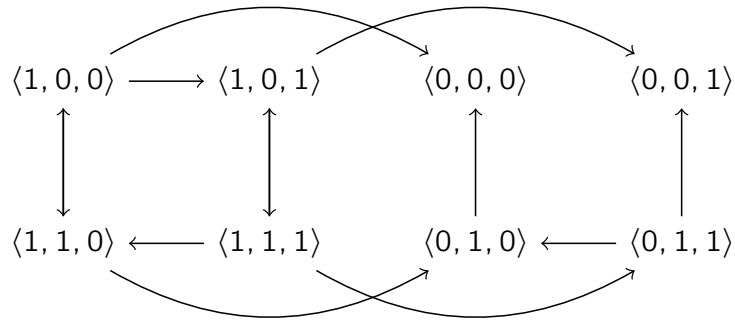


$$f^a(x) = 0$$
$$f^b(x) = x[a] \wedge \neg x[b]$$
$$f^c(x) = \neg x[b] \wedge (x[a] \vee x[c])$$



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

Formal Analysis of Dynamics



Reachability

- From given initial condition(s) (e.g. $\langle 1, 0, 0 \rangle$),
- is it possible to activate component z?

Control

- From given initial condition(s) (e.g. $\langle 1, 0, 0 \rangle$),
- how to prevent the activation of component z?

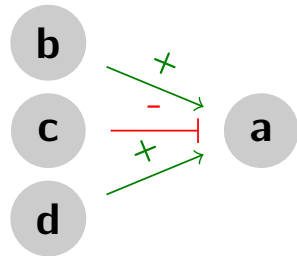
Attractors

- What are the reachable long-term behaviours?
- How to jump from one attractor to another?

Issues with Large Interaction Networks

Modelling issues

- Partially-specified interactions.
- Boolean networks need to be fully specified (deterministic Boolean function f_a).
- Intractable enumeration of all models.

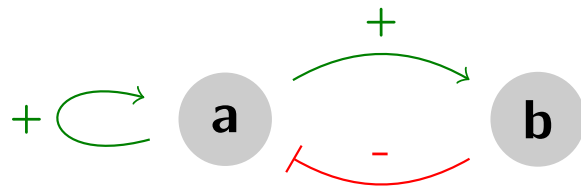


Analysis issues

- Combinatorial explosion of behaviours (e.g. 2^{100} - 10^{30} to $2^{10\,000}$ - 10^{3000} states).
- Large range of initial conditions to consider.
- Difficult to extract comprehensive proofs of (im)possibility.

Failure of classical model-checking techniques,
Need **new formal approaches** to capture dynamics of large networks

Static Analysis based on Interaction Graph

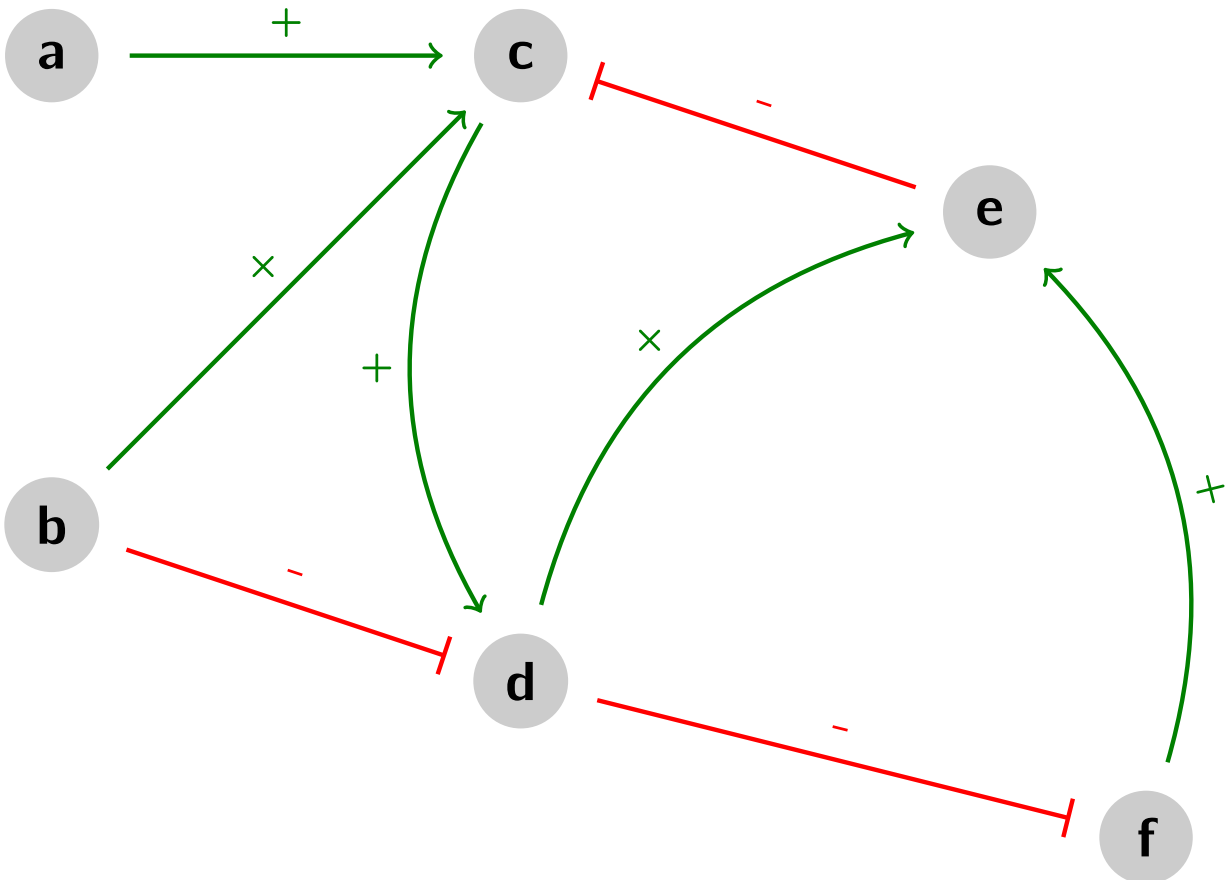


Relationships between the interaction graph and dynamical properties:

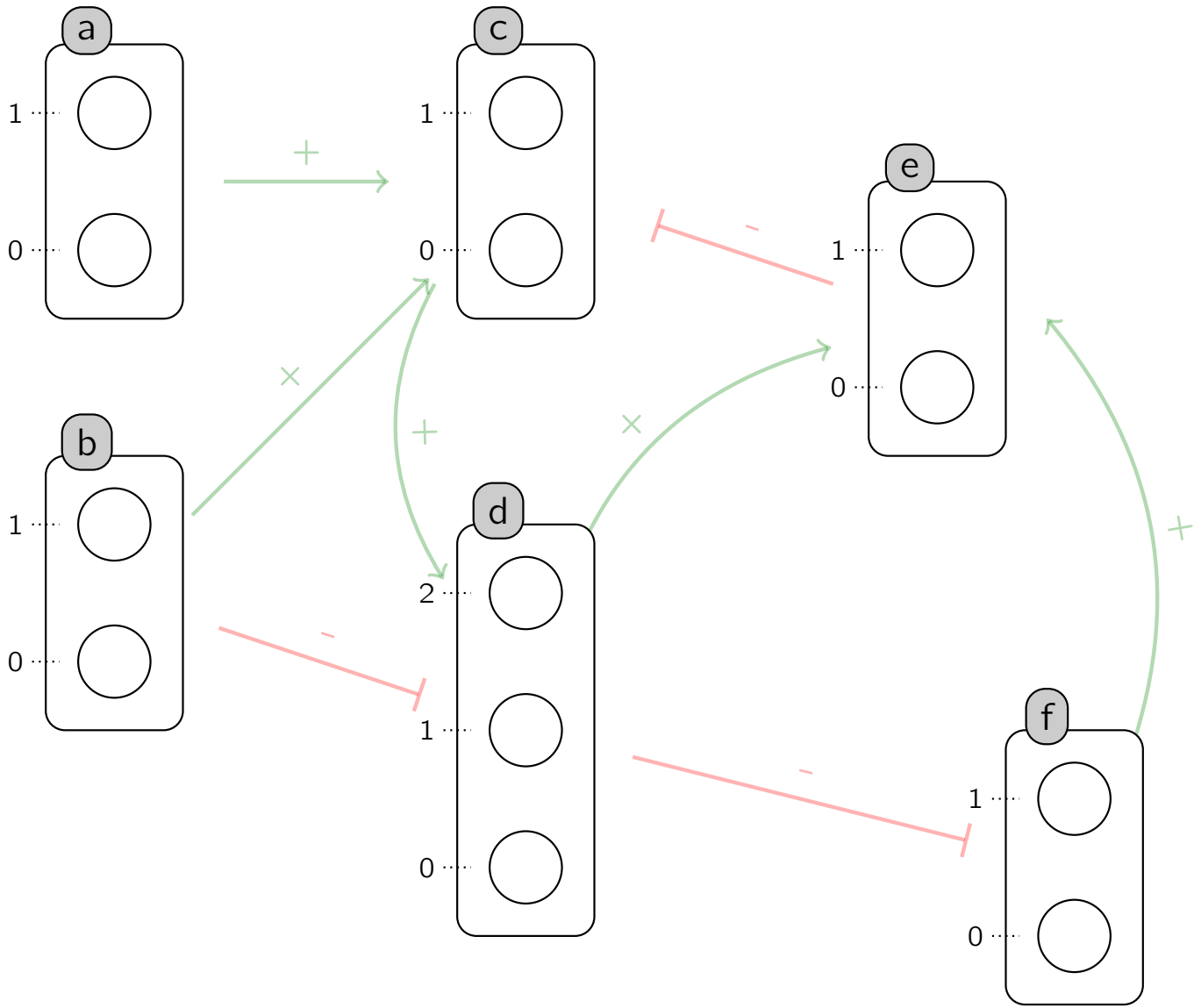
- Multi-stationnarity requires a positive circuit (René Thomas conjecture) [Soule in ComPlexUs, 2003] [Richard, Comet in Discrete Appl. Math., 2007].
- Sustained oscillations require a negative circuit (René Thomas conjecture) [Remy, *et al.* in Adv. Appl. Math., 2008] [Richard in Adv. Appl. Math., 2010].
- The maximum number of fixed points can be characterized [Aracena in Bul. of Mathematical Biology, 2008]; [Richard in Discrete Appl. Math., 2009].
- Topological Fixed Points [Paulevé, Richard in CRAS 2010].
- Difference between synchronous/asynchronous update [Noual, Regnault, Sené]
- etc.

(See [Paulevé, Richard at SASB'11] for a short survey).

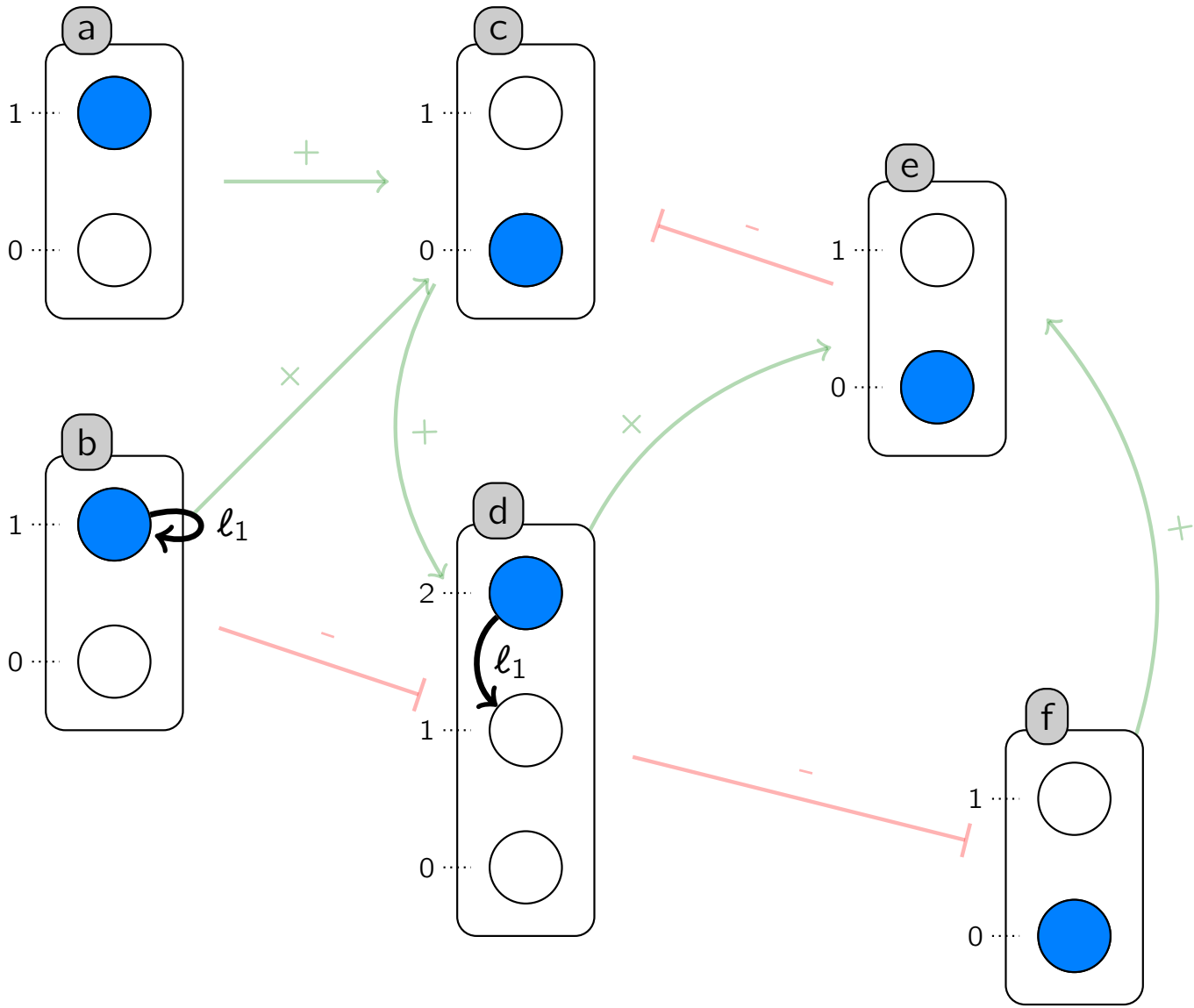
Interaction Networks with Automata Networks



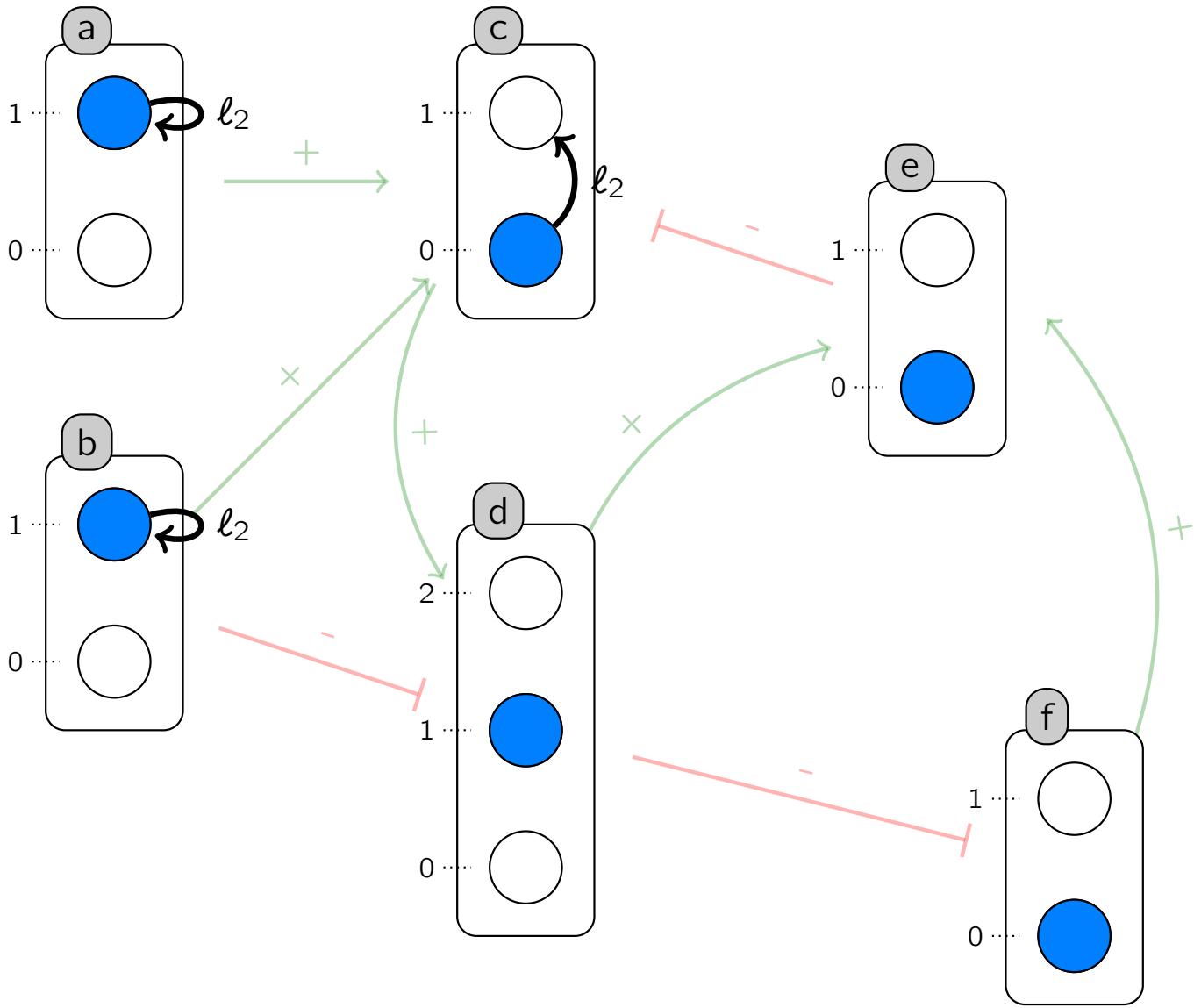
Interaction Networks with Automata Networks



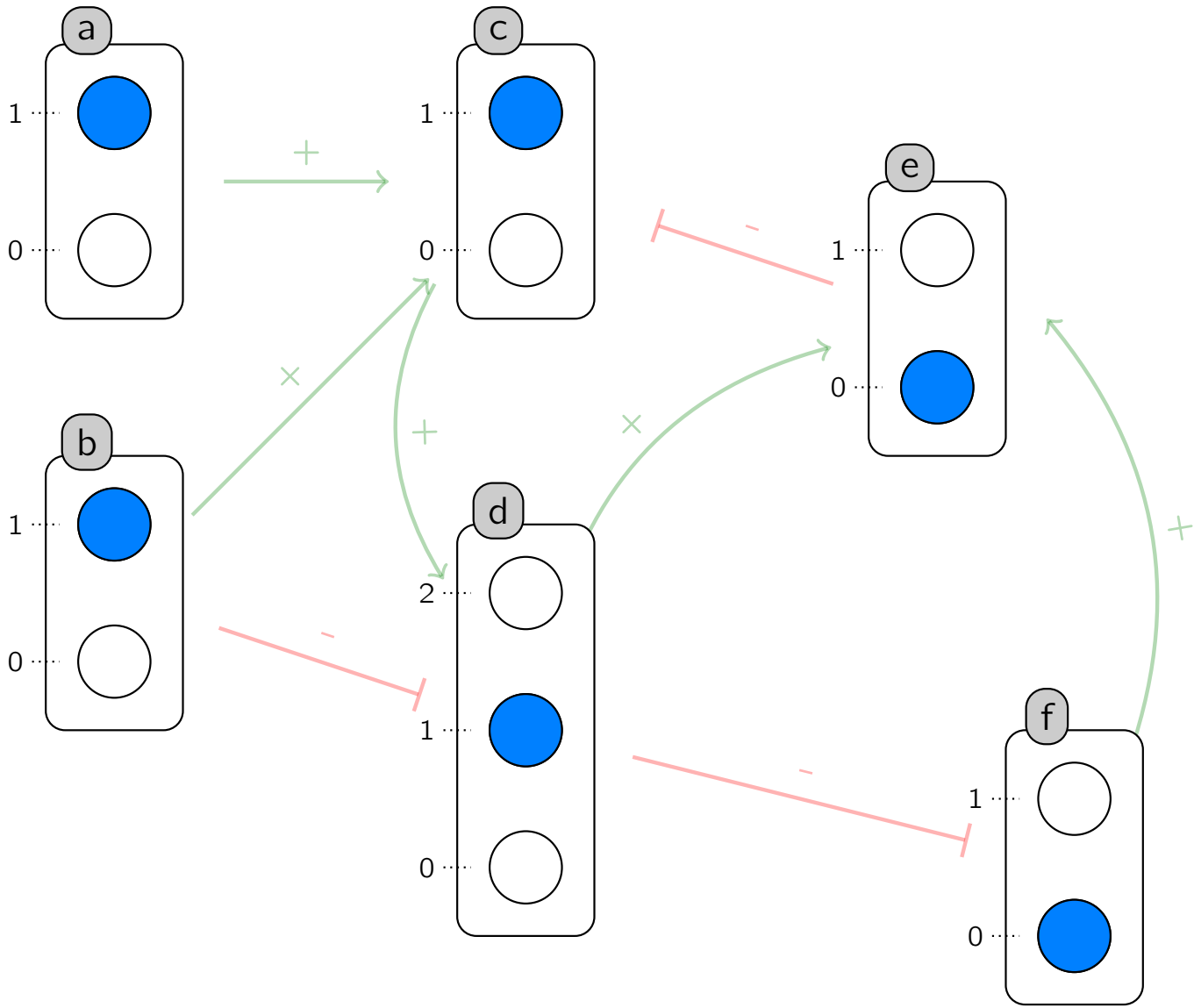
Interaction Networks with Automata Networks



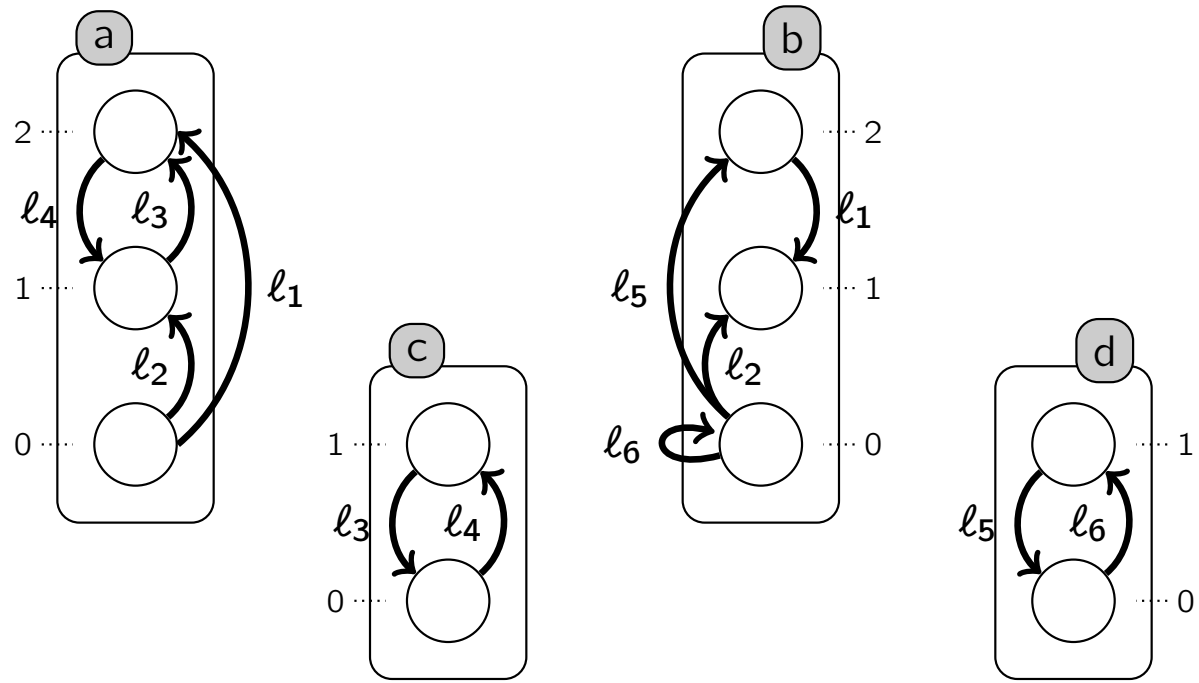
Interaction Networks with Automata Networks



Interaction Networks with Automata Networks



Indeterministic Finite Automata Networks

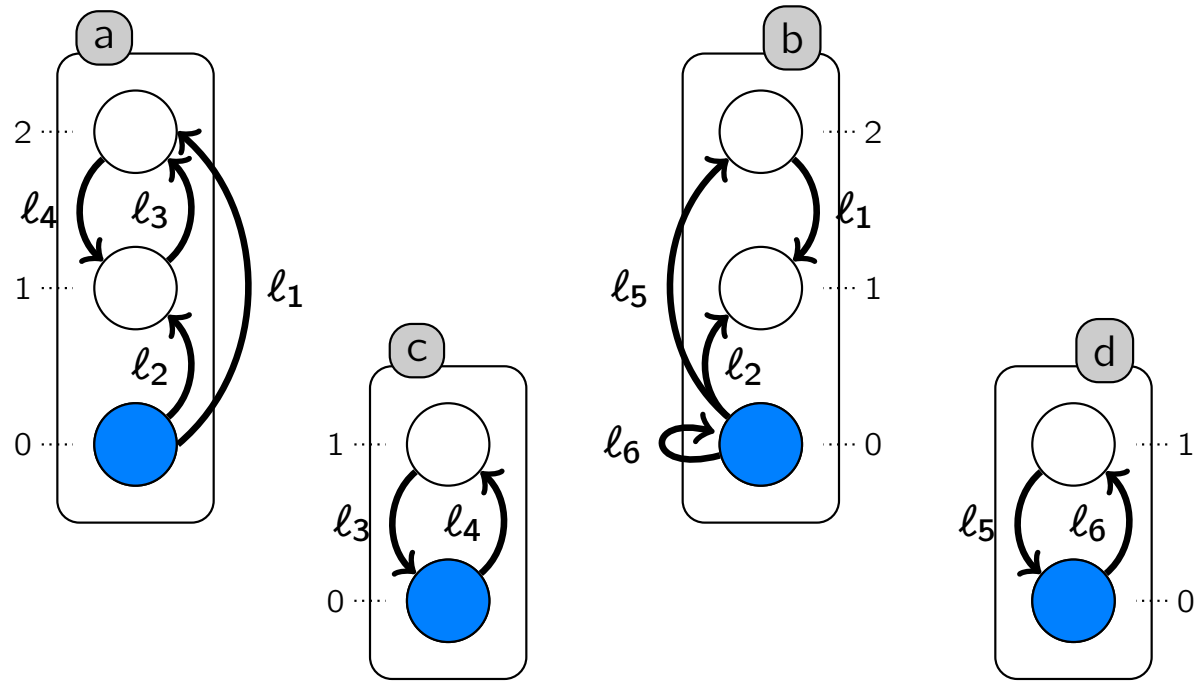


- Transition-centered specification
- Can model indeterministic discrete function:

$$f^a(x) = \begin{cases} 1 & \text{if } x[b] \geq 1 \vee x[c] \geq 1 \\ 0 & \text{if } x[b] = 0 \vee x[c] = 0 \end{cases}$$

- Can model any discrete network async/sync update.

Indeterministic Finite Automata Networks

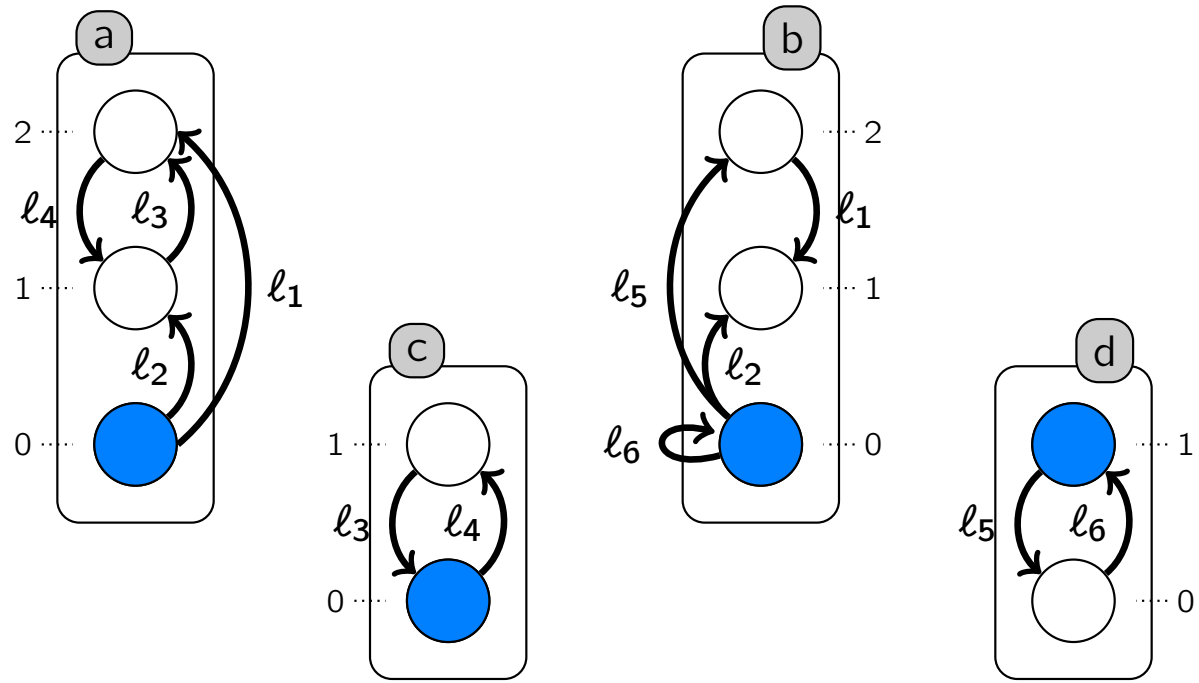


- Transition-centered specification
- Can model indeterministic discrete function:

$$f^a(x) = \begin{cases} 1 & \text{if } x[b] \geq 1 \vee x[c] \geq 1 \\ 0 & \text{if } x[b] = 0 \vee x[c] = 0 \end{cases}$$

- Can model any discrete network async/sync update.

Indeterministic Finite Automata Networks

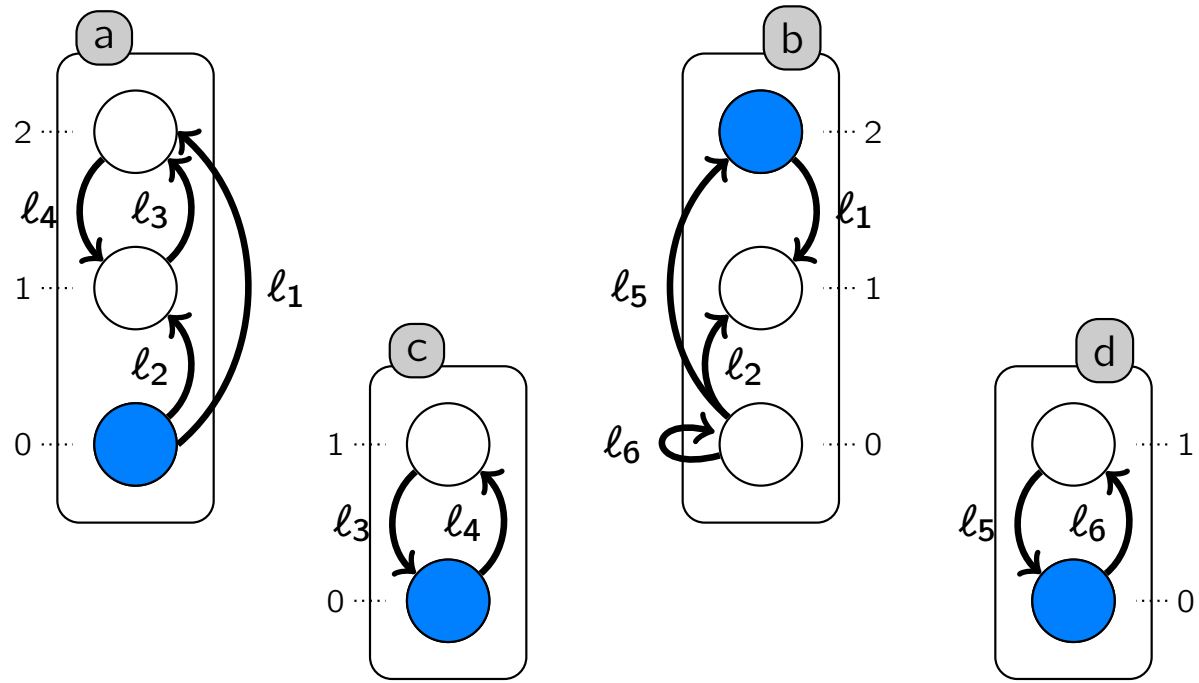


- Transition-centered specification
- Can model indeterministic discrete function:

$$f^a(x) = \begin{cases} 1 & \text{if } x[b] \geq 1 \vee x[c] \geq 1 \\ 0 & \text{if } x[b] = 0 \vee x[c] = 0 \end{cases}$$

- Can model any discrete network async/sync update.

Indeterministic Finite Automata Networks

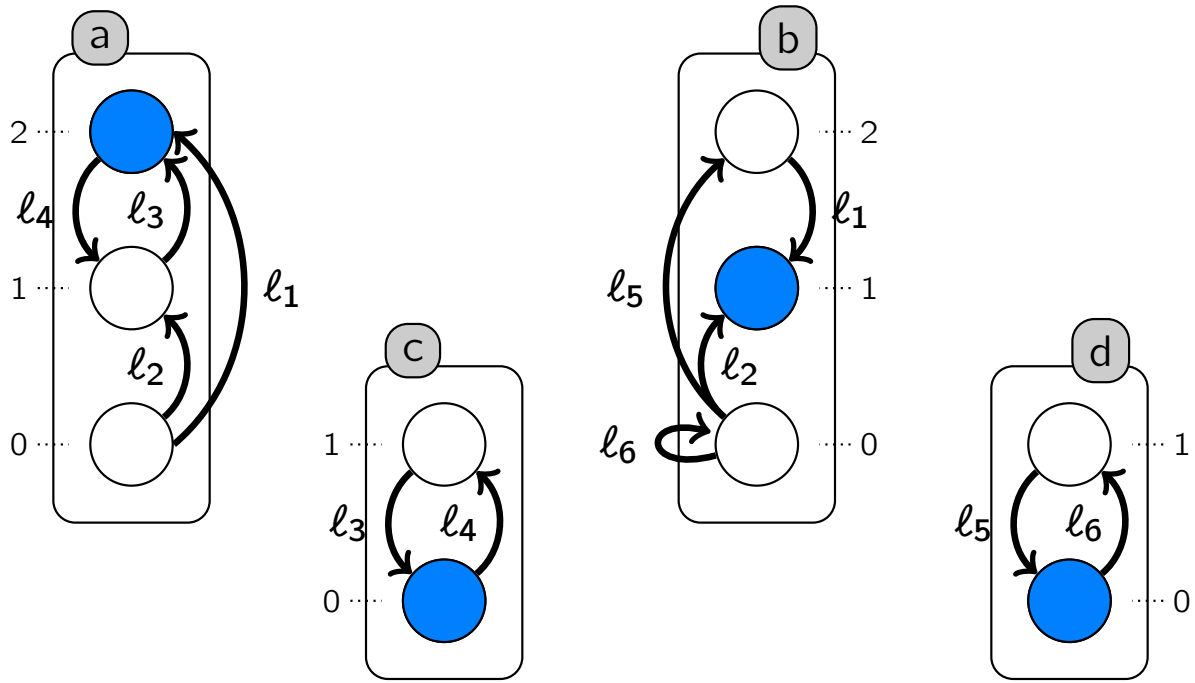


- Transition-centered specification
- Can model indeterministic discrete function:

$$f^a(x) = \begin{cases} 1 & \text{if } x[b] \geq 1 \vee x[c] \geq 1 \\ 0 & \text{if } x[b] = 0 \vee x[c] = 0 \end{cases}$$

- Can model any discrete network async/sync update.

Indeterministic Finite Automata Networks

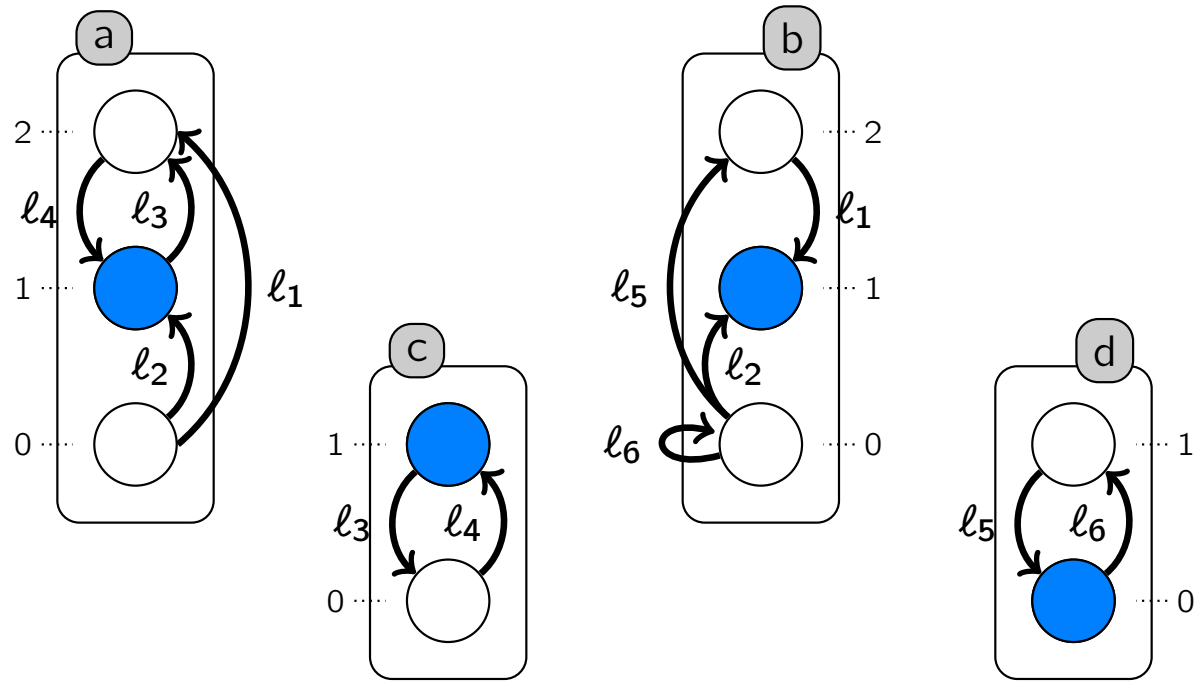


- Transition-centered specification
- Can model indeterministic discrete function:

$$f^a(x) = \begin{cases} 1 & \text{if } x[b] \geq 1 \vee x[c] \geq 1 \\ 0 & \text{if } x[b] = 0 \vee x[c] = 0 \end{cases}$$

- Can model any discrete network async/sync update.

Indeterministic Finite Automata Networks



- Transition-centered specification
- Can model indeterministic discrete function:

$$f^a(x) = \begin{cases} 1 & \text{if } x[b] \geq 1 \vee x[c] \geq 1 \\ 0 & \text{if } x[b] = 0 \vee x[c] = 0 \end{cases}$$

- Can model any discrete network async/sync update.

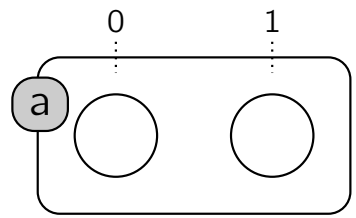
Interaction Networks with Automata Networks

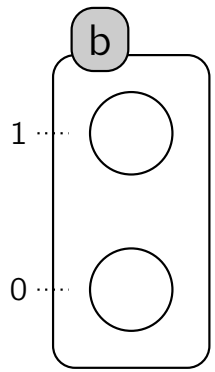
$f^a(x) = x[b] \wedge x[c]$

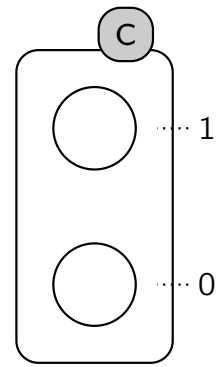
as transitions:

$a_0 \to a_1 \colon\ b_1 \wedge c_1$

$a_1 \to a_0 \colon\ b_1 \vee c_1$







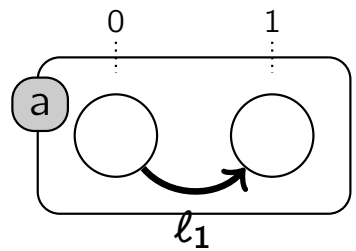
Interaction Networks with Automata Networks

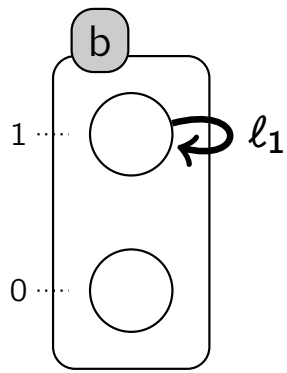
$f^a(x) = x[b] \wedge x[c]$

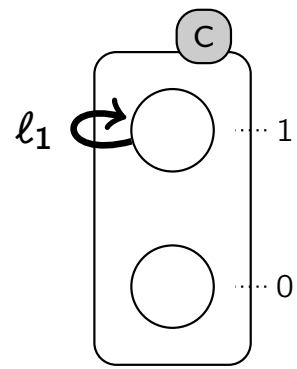
as transitions:

$a_0 \to a_1 \colon\ b_1 \wedge c_1$

$a_1 \to a_0 \colon\ b_1 \vee c_1$







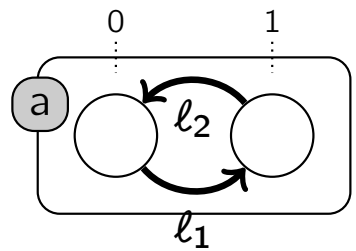
Interaction Networks with Automata Networks

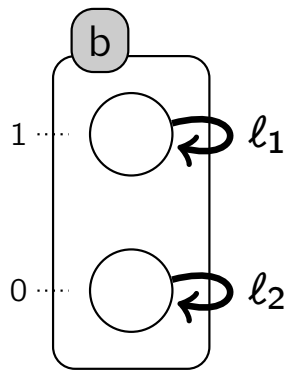
$f^a(x) = x[b] \wedge x[c]$

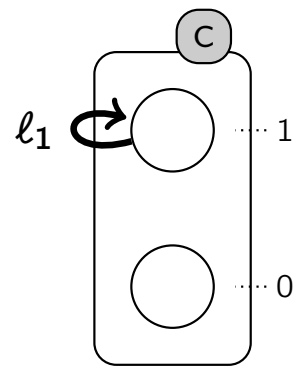
as transitions:

$a_0 \to a_1 \colon\ b_1 \wedge c_1$

$a_1 \to a_0 \colon\ b_1 \vee c_1$







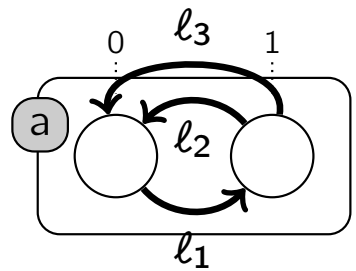
Interaction Networks with Automata Networks

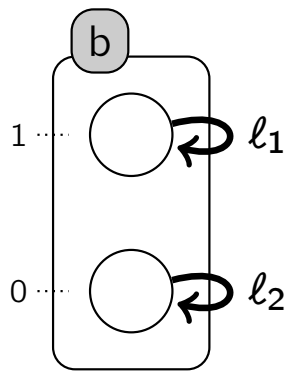
$f^a(x) = x[b] \wedge x[c]$

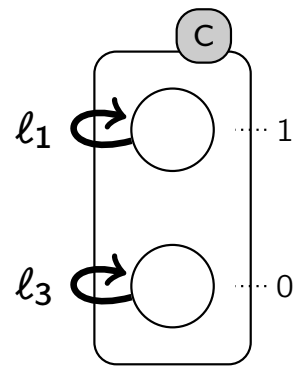
as transitions:

$a_0 \to a_1 \colon\ b_1 \wedge c_1$

$a_1 \to a_0 \colon\ b_1 \vee c_1$







Outline

1 Interaction Networks
- Qualitative Dynamics
- Topology vs Dynamics
- Automata Networks

2 Causality Analysis for Dynamics of Automata Networks
- Abstract Interpretation
- Graph of Local Causality
- Reachability
- Cut Sets

3 Discussion
- Software
- Conclusion, Future Work

Abstract Interpretation

Model $\longrightarrow$ Abstraction $\longrightarrow$ Decision (possibly incomplete)

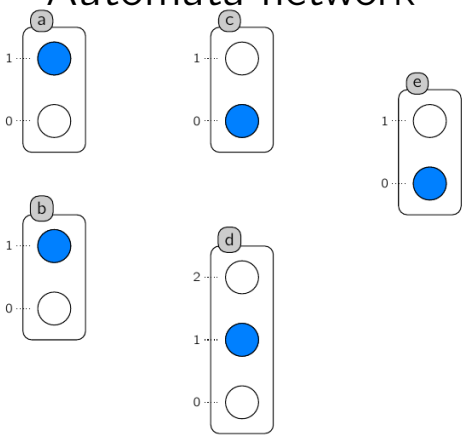
Abstract Interpretation

Model $\longrightarrow$ Abstraction $\longrightarrow$ Decision (possibly incomplete)

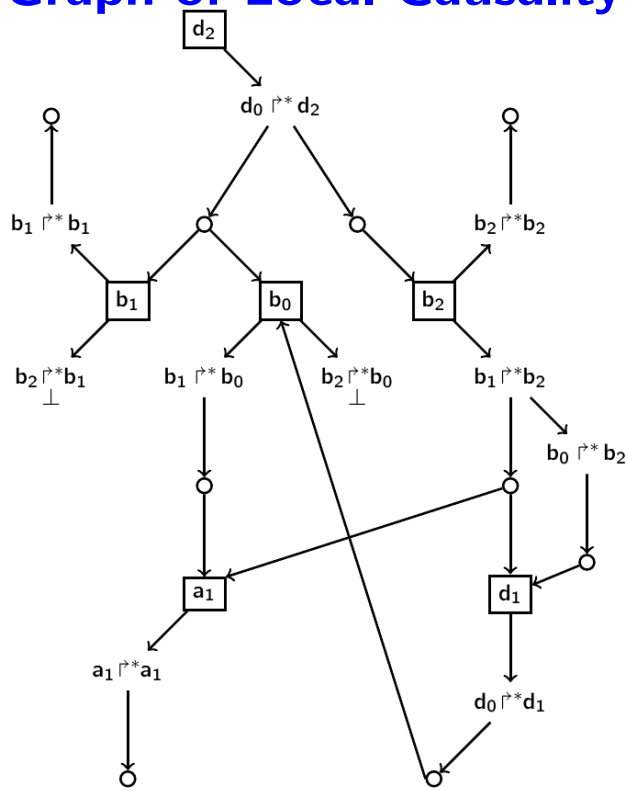
Automata Network $\longrightarrow$ Graph of Local Causality $\longrightarrow$ Incomplete decision

Overview

Automata network



Graph of Local Causality

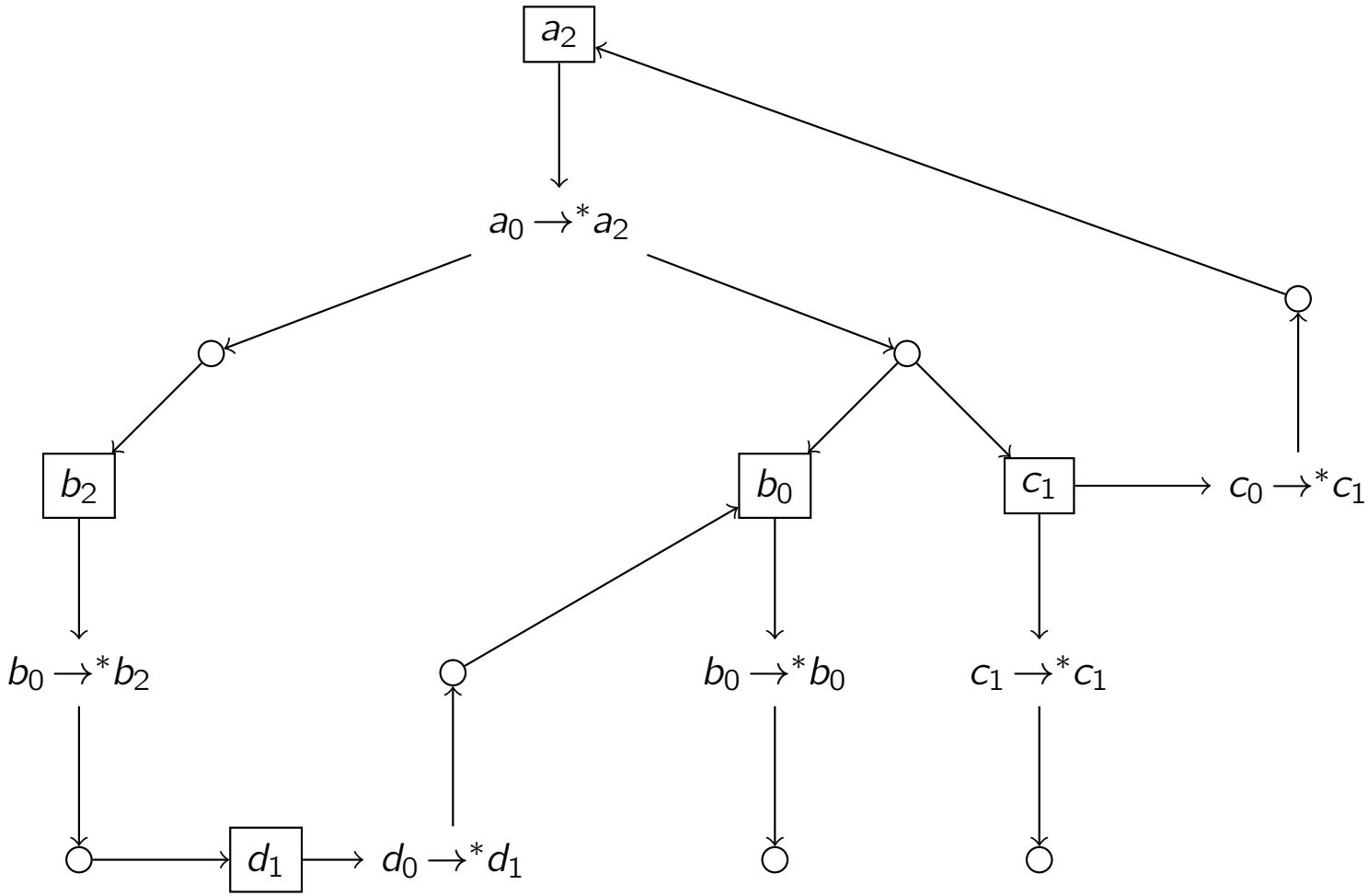


Over-approximation of reachability
Under-approximation of reachability
Under-approximation of cut sets

[Paulevé et al. in Math. Struct. in Comp. Sci. 2012]
[Paulevé et al. at CAV 2013]

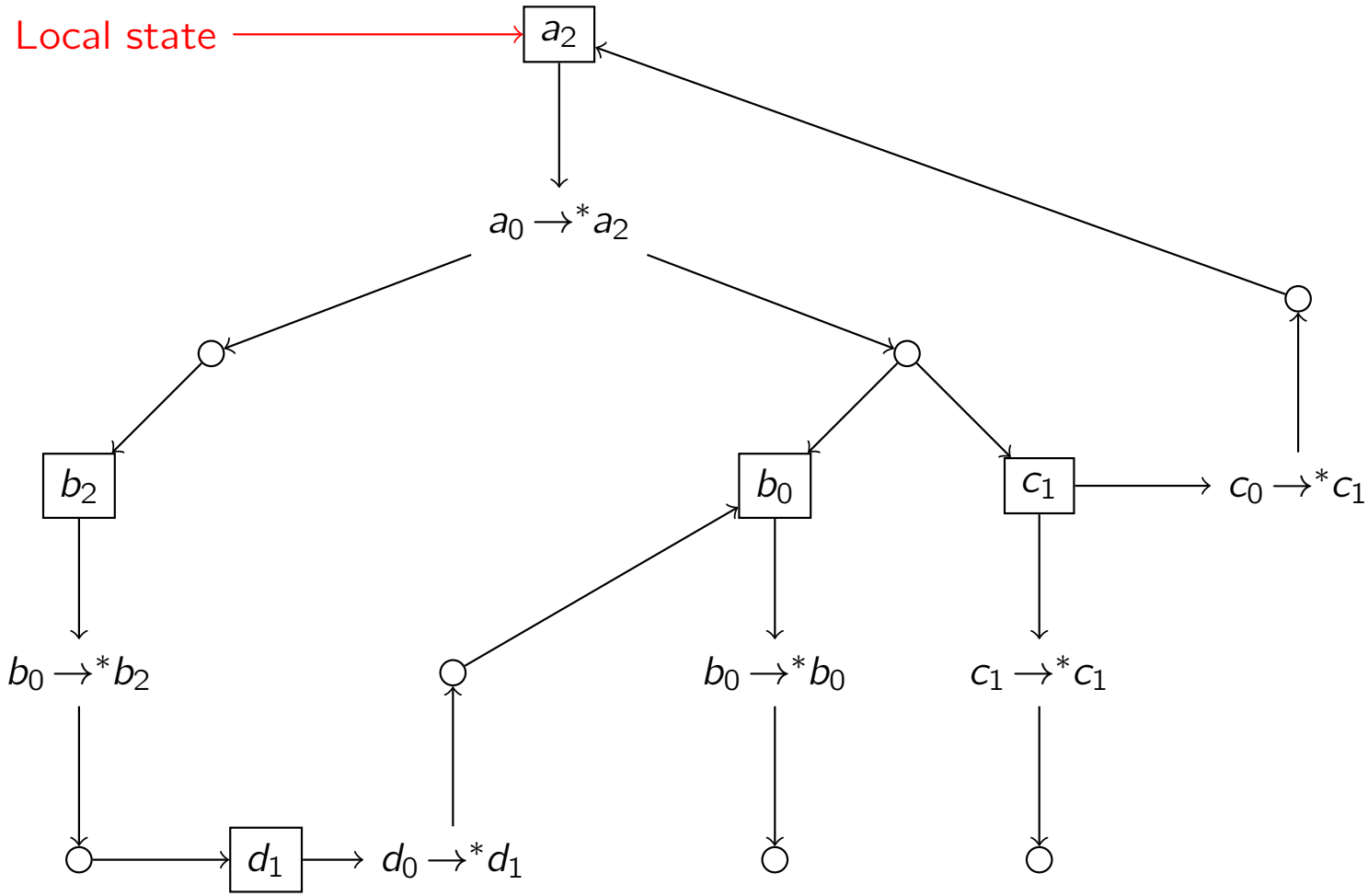
Graph of Local Causality

- Causality of a_2.
- Initial context $\varsigma = \{a \mapsto \{0\}; b \mapsto \{0\}; c \mapsto \{0,1\}; d \mapsto \{1\}\}$.



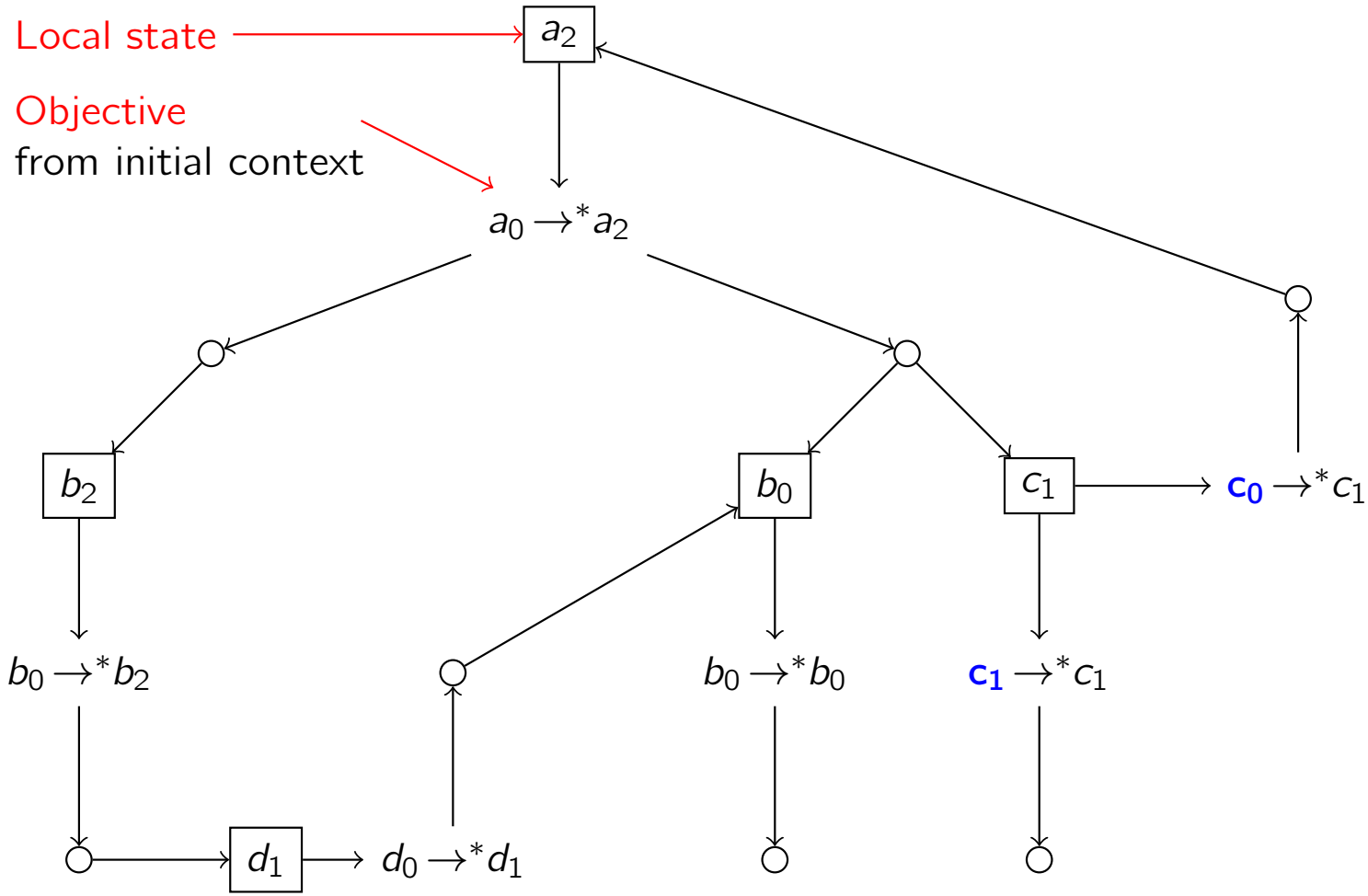
Graph of Local Causality

- Causality of a_2.
- Initial context $\varsigma = \{a \mapsto \{0\}; b \mapsto \{0\}; c \mapsto \{0,1\}; d \mapsto \{1\}\}$.



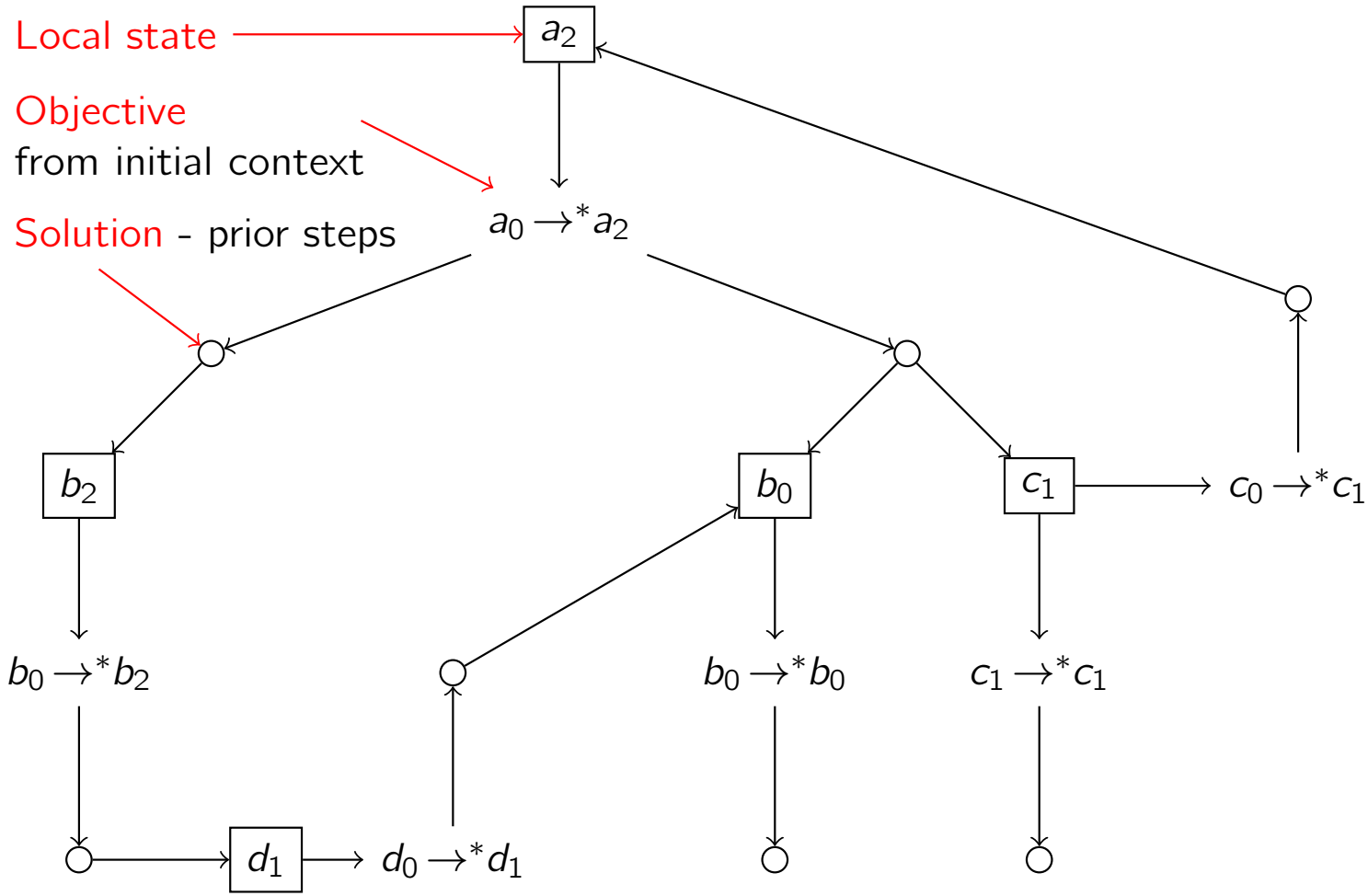
Graph of Local Causality

- Causality of a_2.
- Initial context $\varsigma = \{a \mapsto \{0\}; b \mapsto \{0\}; c \mapsto \{\mathbf{0}, \mathbf{1}\}; d \mapsto \{1\}\}$.

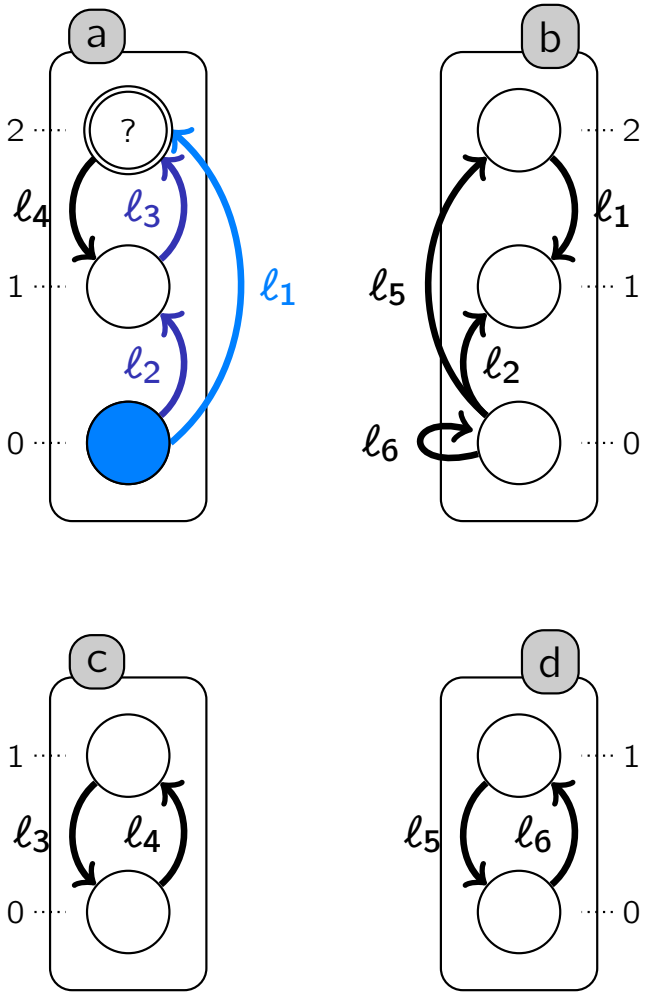


Graph of Local Causality

- Causality of a_2.
- Initial context $\varsigma = \{a \mapsto \{0\}; b \mapsto \{0\}; c \mapsto \{0,1\}; d \mapsto \{1\}\}$.

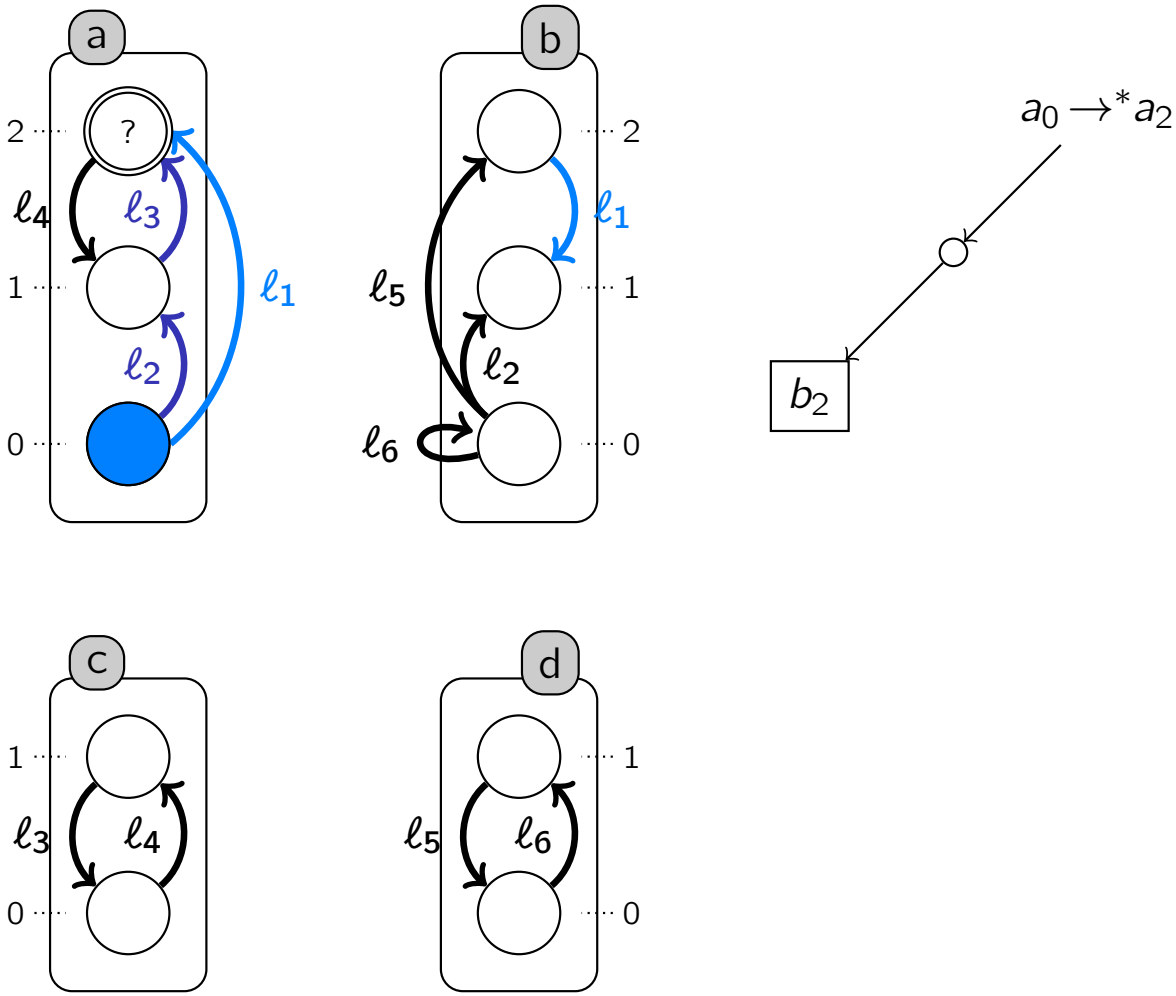


Computing GLC for Automata Networks

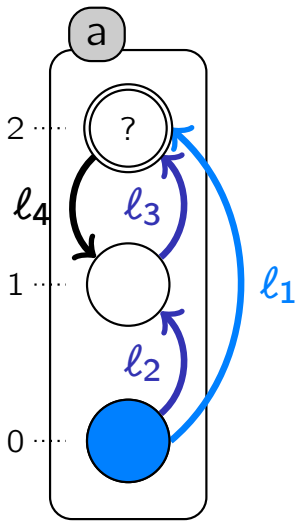


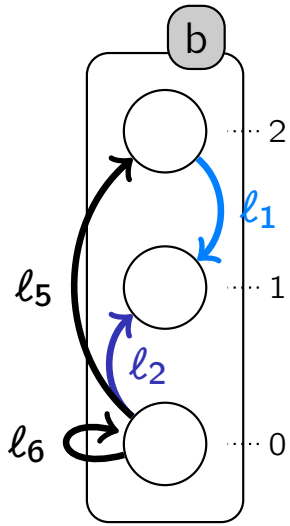
$a_0 \rightarrow^* a_2$

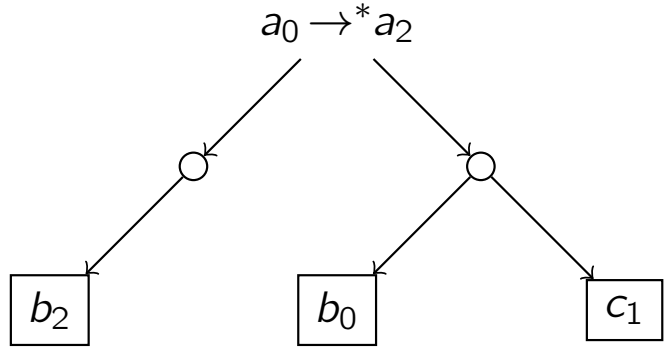
Computing GLC for Automata Networks



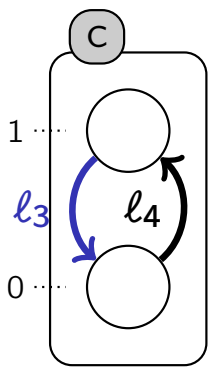
Computing GLC for Automata Networks

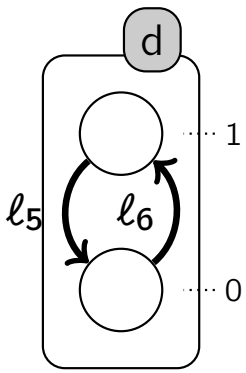




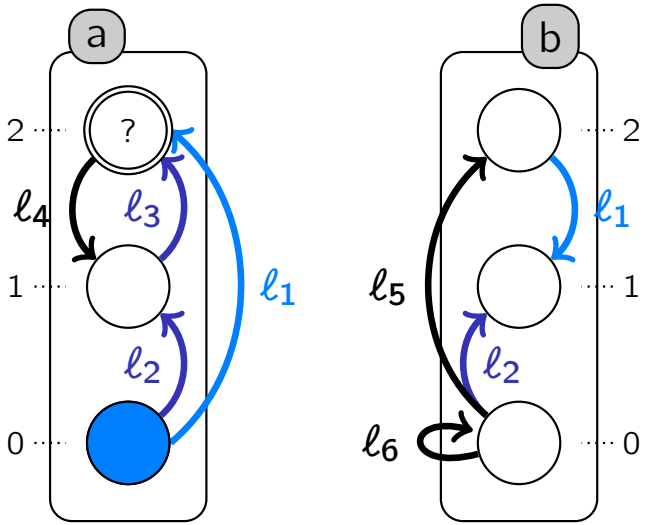


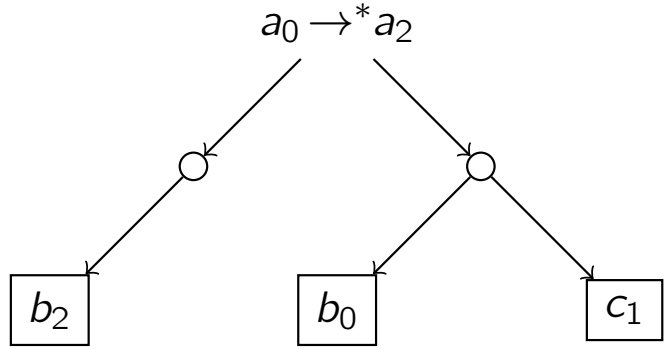
(ignore order, count, synchronism)



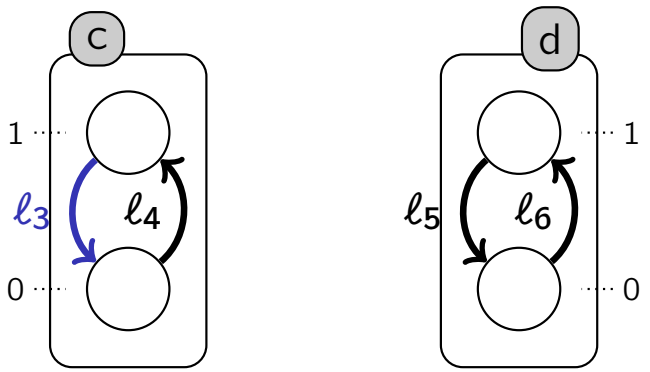


Computing GLC for Automata Networks





(ignore order, count, synchronism)

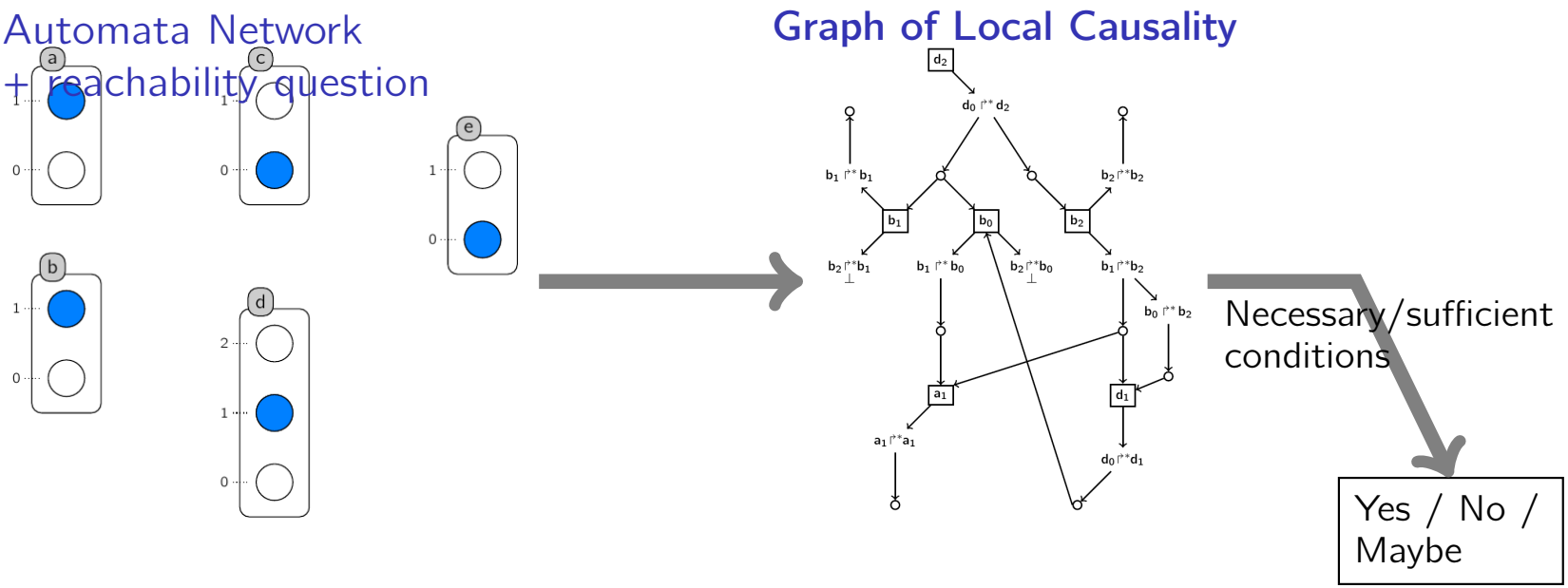


Complexity of GLC (construction + size of GLC)

- polynomial in the total number of local states;
- exponential in the number of local transitions within one automaton.

Efficient Reachability Analysis

Abstract interpretation of Automata Networks dynamics



Reach a_i, then b_j, etc.

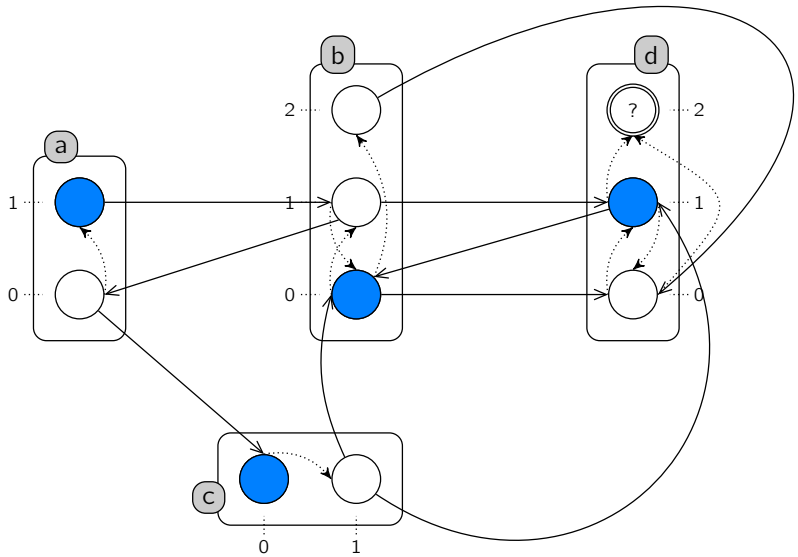
- Over- and under-approximations of local rechability properties.
- Low complexity: poly(nb. automata) $\times$ exp(nb of procs in one automaton)

$\Longrightarrow$ efficient with a small number of processes per automaton, while a very large number of automata can be handled.

[**Mathematical Structures in Computer Science** (2012); workshop **SASB**'10]

Over-approximation of Reachability

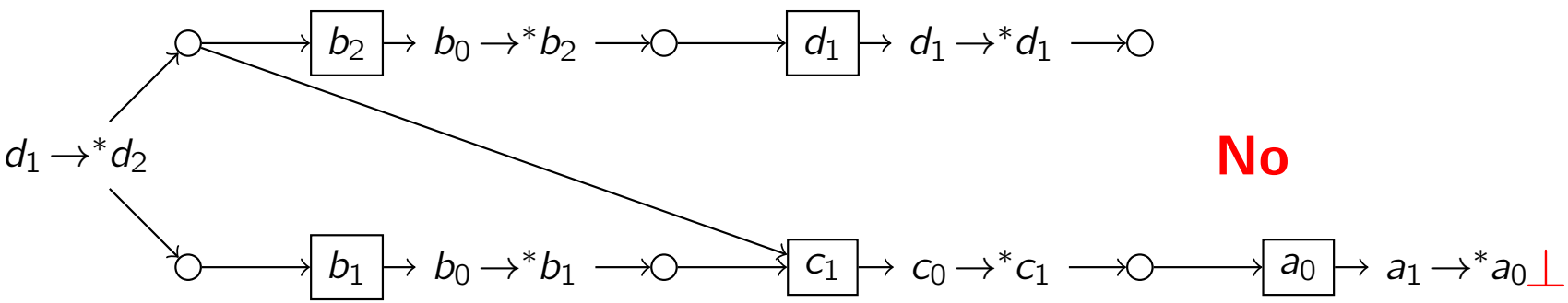
Example



Necessary condition for reaching d_2:

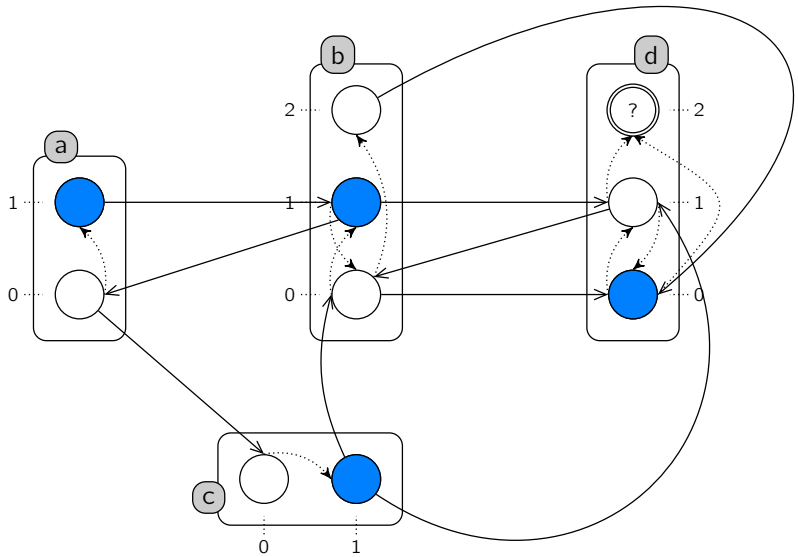
There exists a traversal of the GLC s.t.:

- objective $\rightarrow$ follow at least one solution;
- process $\rightarrow$ follow all objectives;
- no cycle.



Over-approximation of Reachability

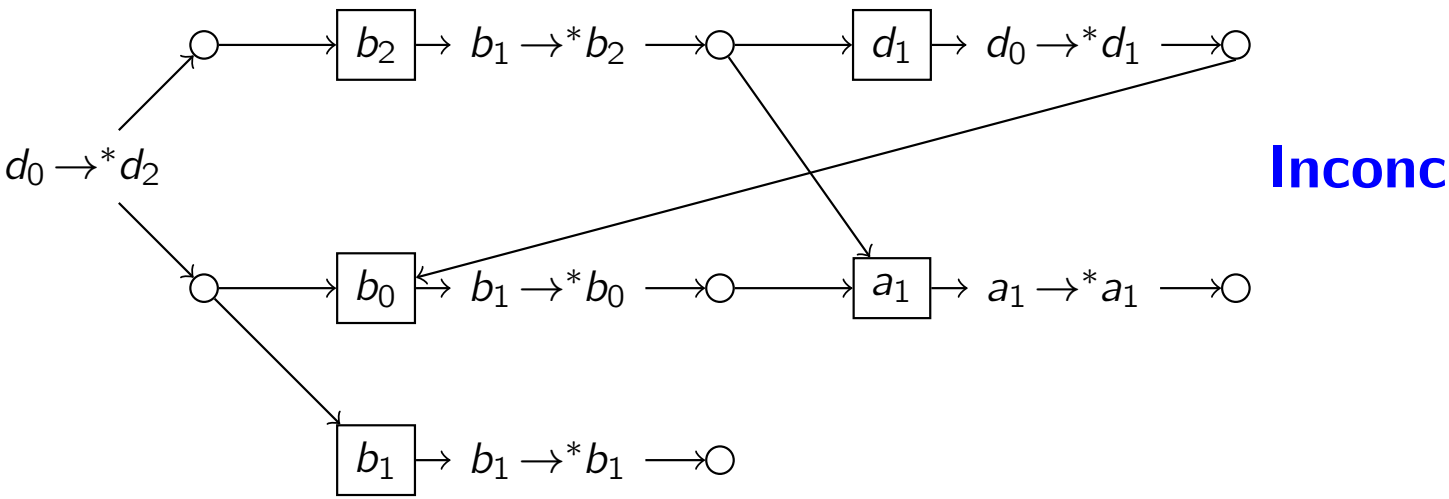
Example



Necessary condition for reaching d_2:

There exists a traversal of the GLC s.t.:

- objective $\rightarrow$ follow at least one solution;
- process $\rightarrow$ follow all objectives;
- no cycle.

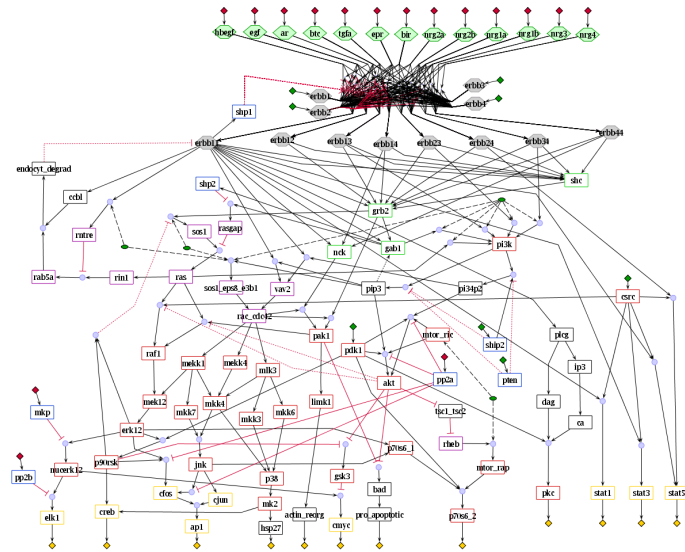


Inconc

Applications

- Signalling networks.
- Wide-range of biological/arbitrary reachability analysis.
- Always conclusive.

Model	Biocham[1]	libDDD[2]	PINT[3]
EGFR 20	[3s-KO]	[1s-150s]	0.007s
TCR 40	[1s-KO]	[0.6s-KO]	0.004s
TCR 94	KO	KO	0.030s
EGFR 104	KO	KO	0.050s



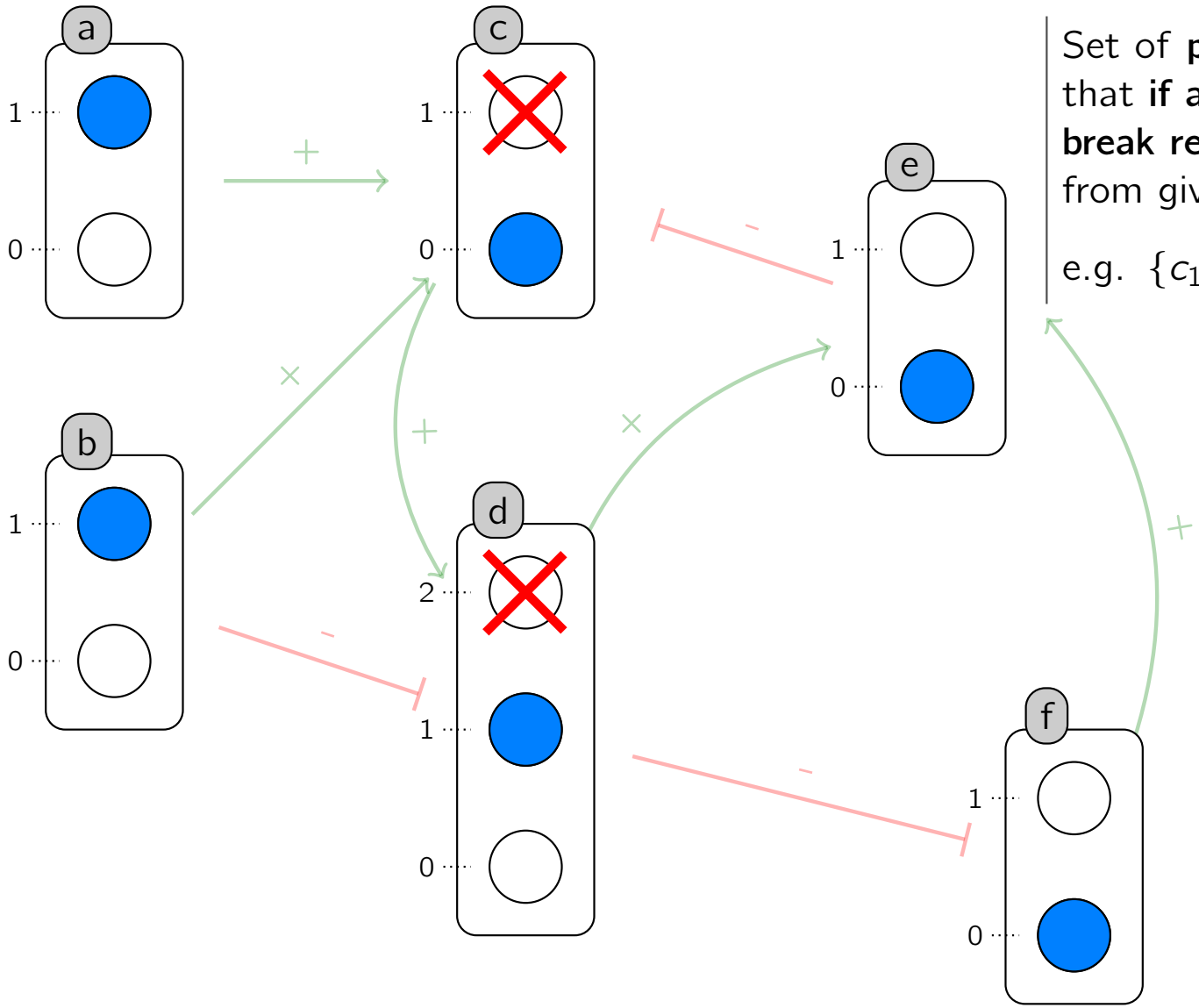
[1] `http://contraintes.inria.fr/biocham` (using NuSMV2)
[2] `http://move.lip6.fr/software/DDD`
[3] `http://loicpauleve.name/pint`

Cut Sets for Reachability

[Paulevé et al. at CAV'13]

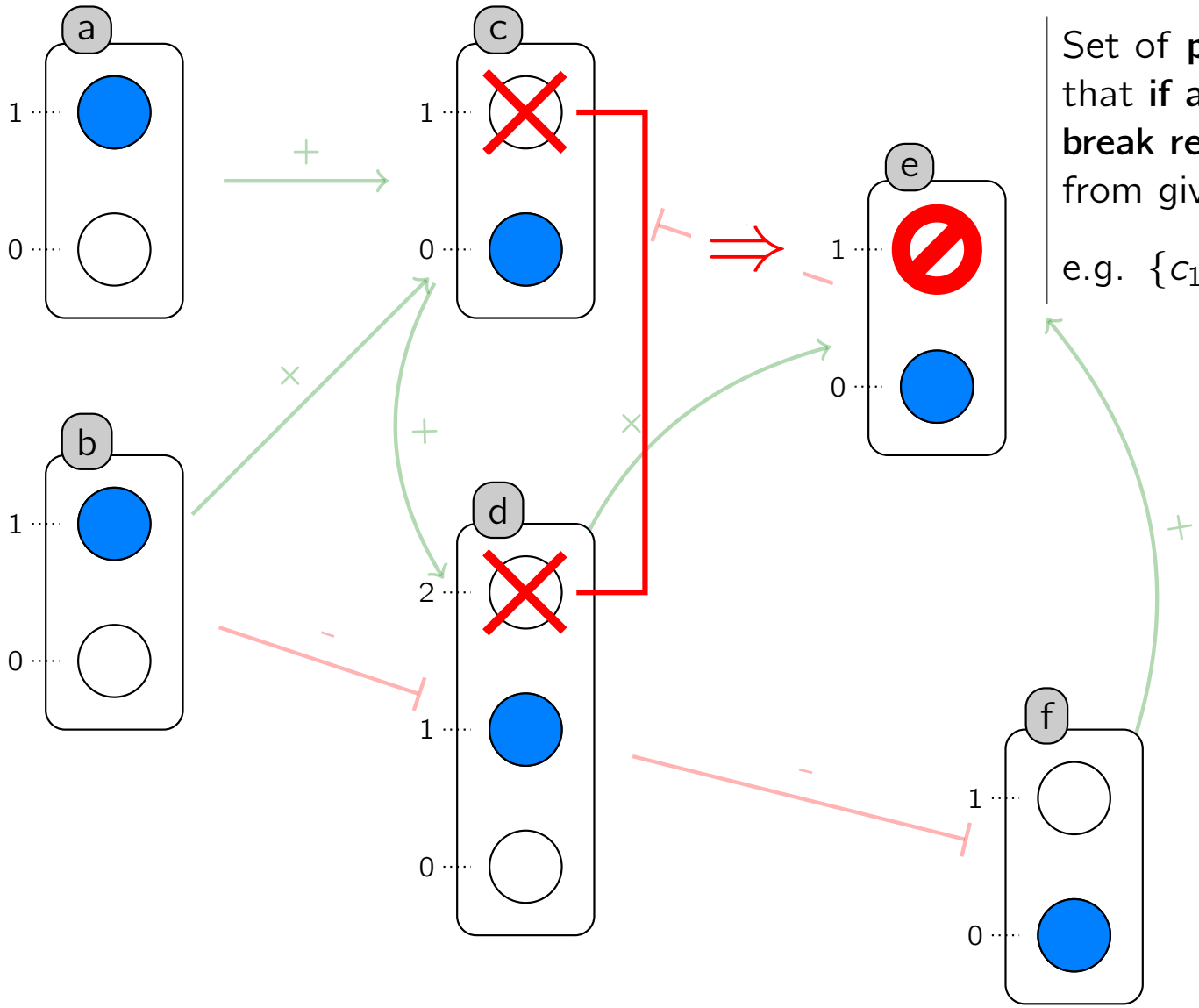
Set of **processes** that **if all disabled break reachability** from given initial states

e.g. $\{c_1, d_2\}$



Cut Sets for Reachability

[Paulevé et al. at CAV'13]

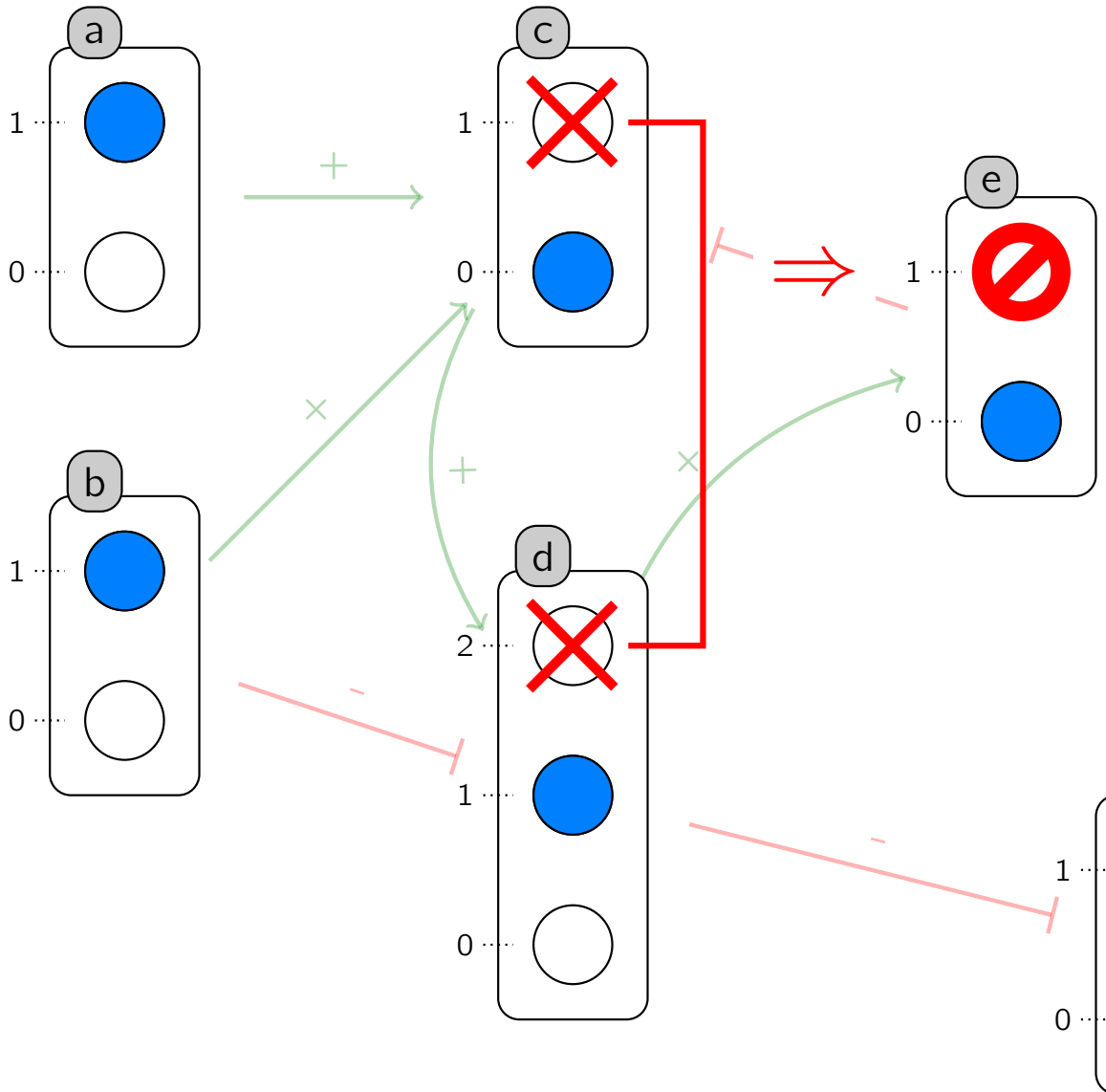


Set of **processes** that **if all disabled break reachability** from given initial states

e.g. $\{c_1, d_2\}$

Cut Sets for Reachability

[Paulevé et al. at CAV'13]



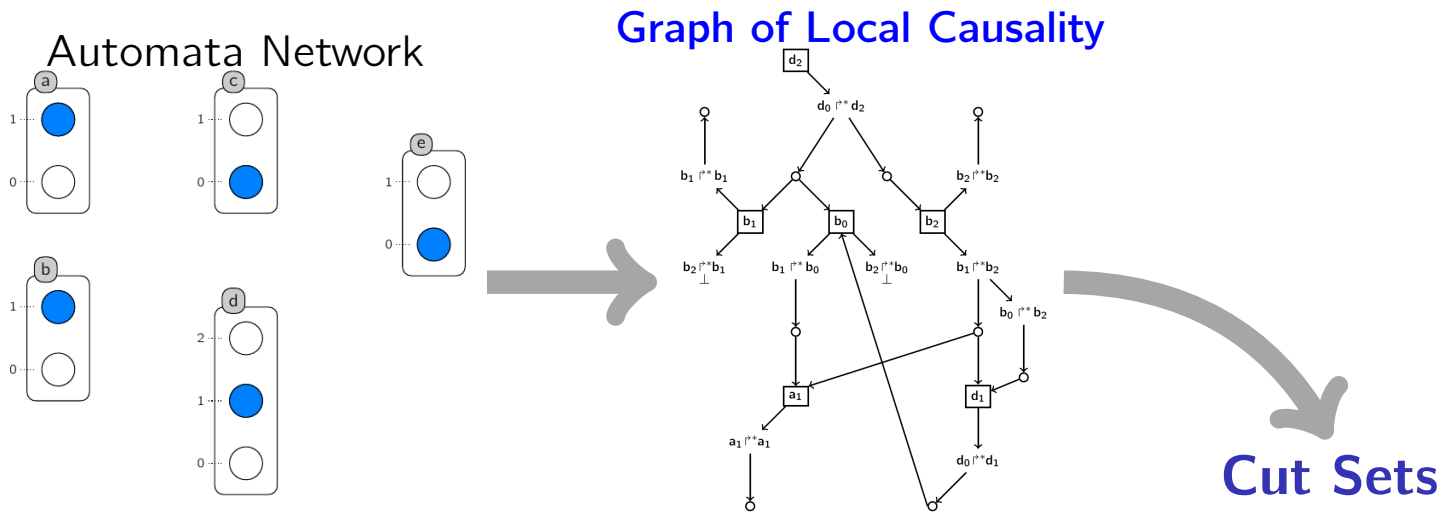
Set of **processes** that **if all disabled break reachability** from given initial states

e.g. $\{c_1, d_2\}$

Applications

- Potential **control targets**
- Refute model if reachability still occurs in the modified (real) system

Cut Sets for Reachability



Algorithm

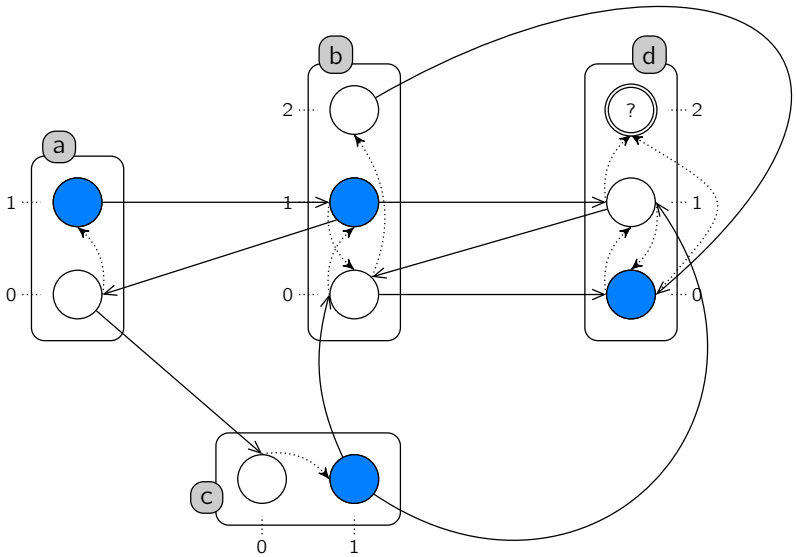
- Graph flooding algorithm.
- Computes all cut sets at once: no enumeration of candidates.
- Very efficient with large networks.

Returned cut sets

- All valid (break the concerned reachability).
- Some may be missed, some may be non-minimal.

Identification of Cut Sets

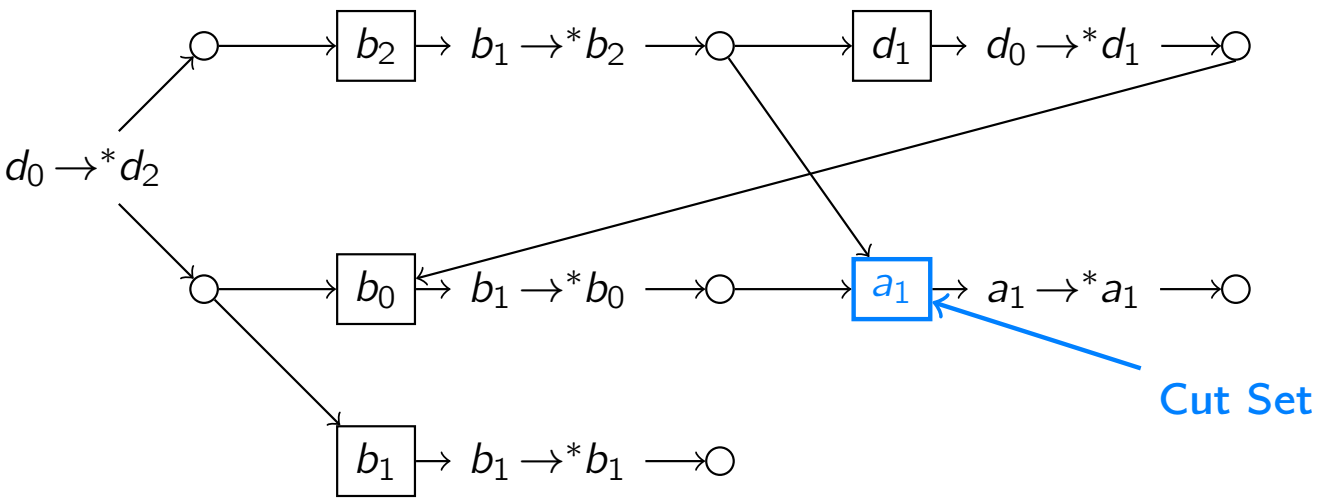
[Paulevé, Andrieux, Koeppl at Computer Aided Verification 2013]



Necessary condition for reaching d_2:

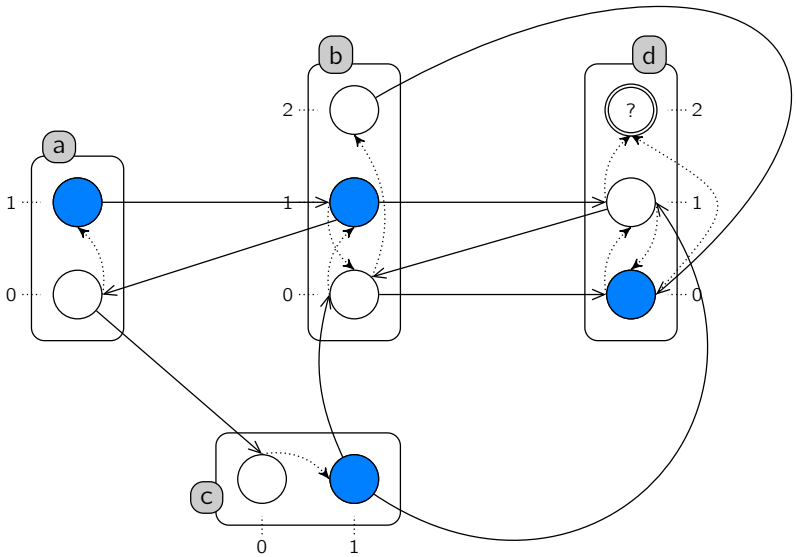
There exists a traversal of the GLC s.t.:

- objective $\rightarrow$ follow at least one solution;
- process $\rightarrow$ follow all objectives;
- no cycle.



Identification of Cut Sets

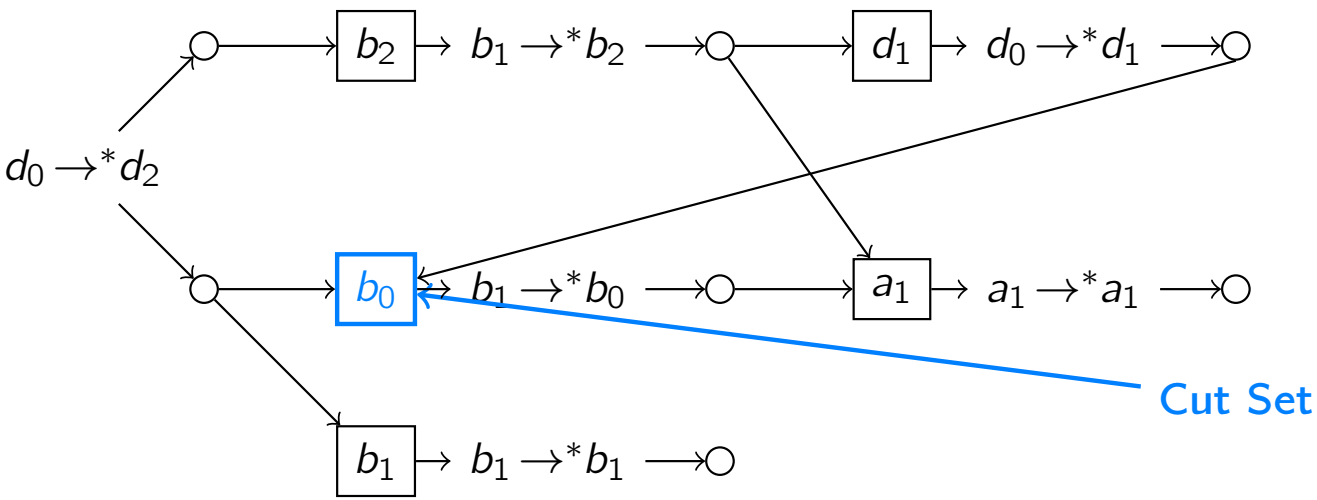
[Paulevé, Andrieux, Koeppl at Computer Aided Verification 2013]



Necessary condition for reaching d_2:

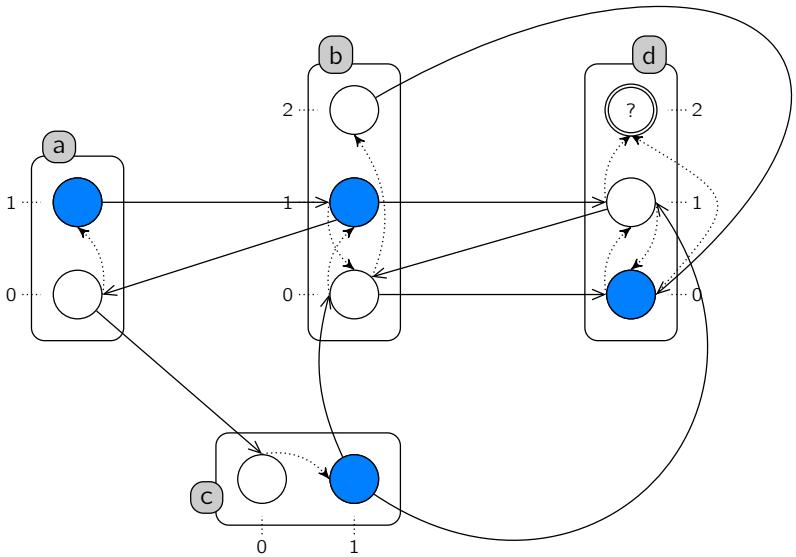
There exists a traversal of the GLC s.t.:

- objective $\rightarrow$ follow at least one solution;
- process $\rightarrow$ follow all objectives;
- no cycle.



Identification of Cut Sets

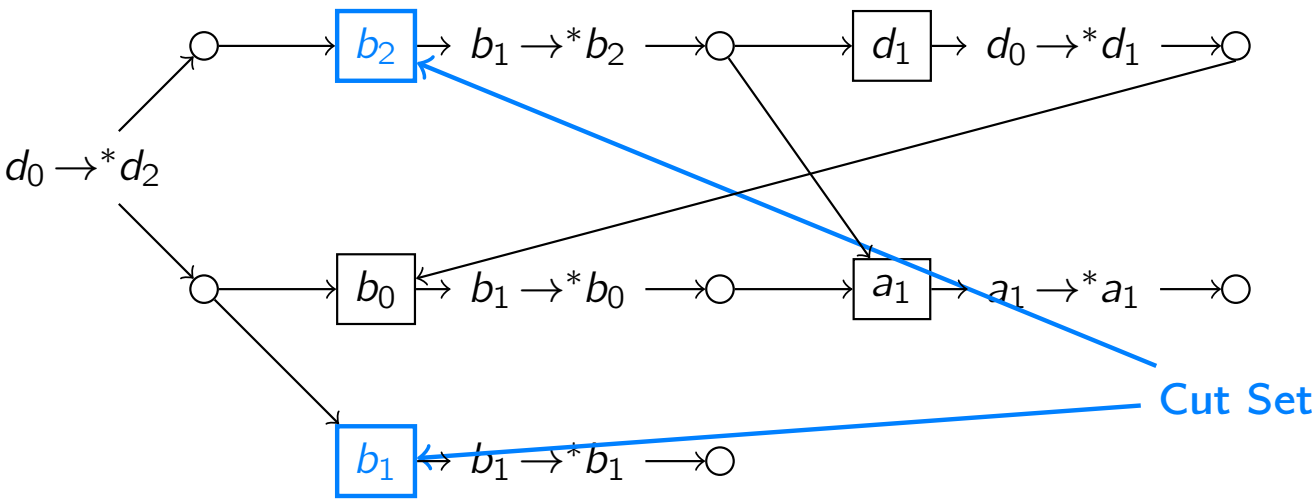
[Paulevé, Andrieux, Koeppl at Computer Aided Verification 2013]



Necessary condition for reaching d_2:

There exists a traversal of the GLC s.t.:

- objective $\rightarrow$ follow at least one solution;
- process $\rightarrow$ follow all objectives;
- no cycle.



Formal analysis of the whole PID

Pathway Interaction Database

- Inductions, inhibitions, transcriptional regulation, complex formations, ...
- More than 9000 interacting components.
- Large environment (3000 entry-points).

Graph of Local Causality

- From Process Hitting model (boolean interpretation).
- (Independent) reachability of active SNAIL, active p15INK4b.
- 20 000 nodes, including 5600 processes (biological or cooperative).

Cut *N*-sets computed

N	Exec. time	$SNAIL_1$	$p15INK4b_1$
1	0.9s	1	1
2	1.6s	+6	+6
3	5.4s	+0	+92
4	39s	+30	+60
5	8.3m	+90	+80
6	2.6h	+930	+208

Outline

1 Interaction Networks
- Qualitative Dynamics
- Topology vs Dynamics
- Automata Networks

2 Causality Analysis for Dynamics of Automata Networks
- Abstract Interpretation
- Graph of Local Causality
- Reachability
- Cut Sets

3 Discussion
- Software
- Conclusion, Future Work

Pint Software

http://loicpauleve.name/pint

Pint

- Textual language for Process Hitting (Automata Networks subclass)
- Command line utilities for analysis.

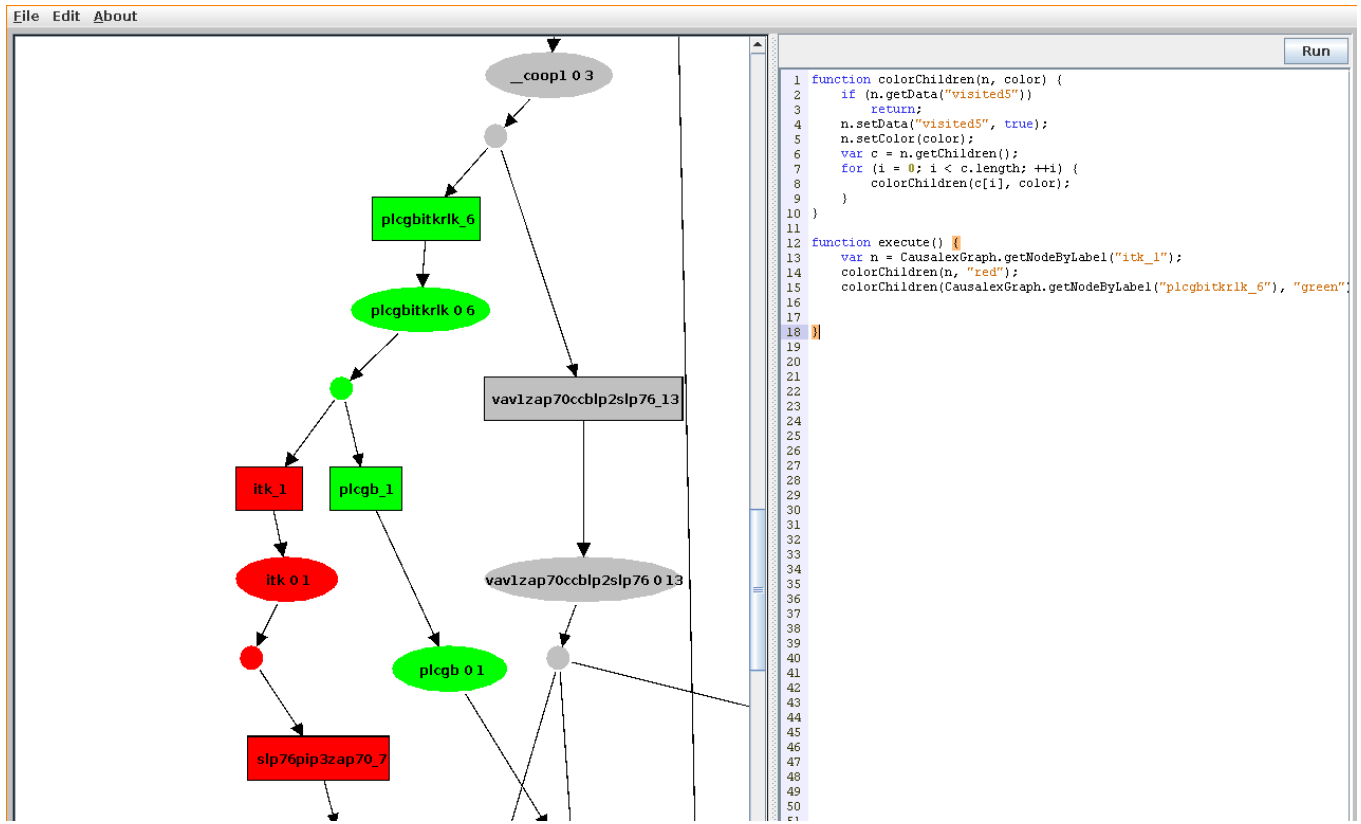
Main features

- Reachability analysis.
- Cut set analysis.
- Listing of fixed points (steady states).
- Non-markovian simulator for stochasticity absorption.
- Importation from various formats (CellNetAnalyser, SIF, ginML (partial), etc.)
- Exportation to various formats (PRISM, Biocham, Boolean networks, etc.)

Graphical interface in progress. . .

CausalEx Software

Coming soon...



Graphical interface for exploring Graphs of Local Causality

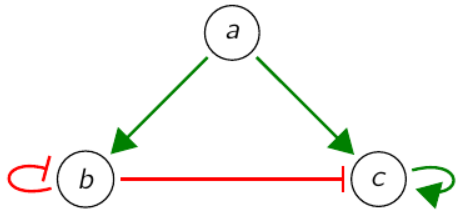
- Navigation
- Interactive scripting (javascript)
- Algorithm visualization

ACK: Fabienne Hirwa and Jean-Christophe Souplet from the software development team/LRI

Summary

Model

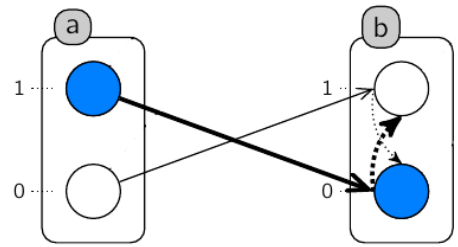
Interaction network



Dynamical interpretation

Automata Networks

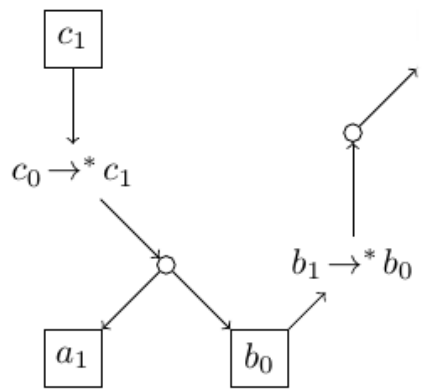
- qualitative model
- general framework
- transition centered



Static analysis

Graph of Local Causality

- necessary steps for reachability
- compact representation of dynamics



Motifs

Local Reachability

(reach a_1 then b_0, etc.)

- low complexity
- over-approximation (No/?)
- under-approximation (Yes/?)

Necessary motifs

Necessary Processes

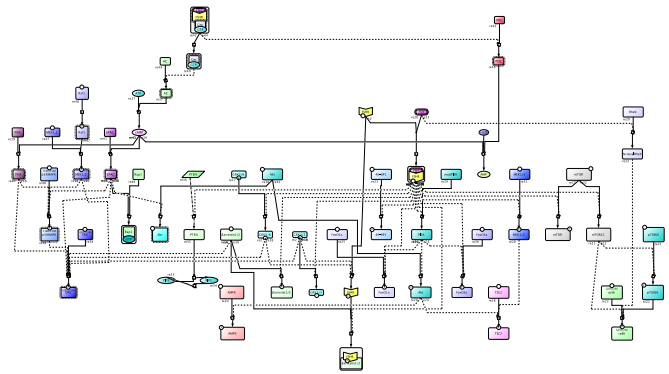
- Potential therapeutic targets
- very large networks

[conf. **Computer Aided Verification** 2013]
[journal **Math. Structures in Comp. Sci.** 2012]
[journal **IEEE Trans. on Soft. Engineering** 2011]
[journal **Trans. on Comput. Systems Bio.** 2011]

Future work

Dynamics of G protein-coupled receptors
w/ Christine Froidevaux, Adrien Rougny + INRA Tours

- Models of FSH and EGFR from litterature + experiments
- Large networks: around 200 components
- Dynamical analysis: key components for cell response
- Control: prevent particular activations while preserving others.



Modelling and Analysis

- Time scales: prioritized transitions.
- Characterize attractors and their basins.
- Model reduction using GLC.

Thank you for your attention.