

MOVEP 2014

Efficient analysis on very large models

Maxime FOLSCHETTE

MeForBio / IRCCyN / École Centrale de Nantes (Nantes, France)

`maxime.folschette@irccyn.ec-nantes.fr`
`http://maxime.folschette.name/`

Context and Aims

MeForBio team:
Qualitative modelling to study
large dynamical biological systems

Context and Aims

MeForBio team:
Qualitative modelling to study
large dynamical biological systems

- 1) The object: **Gene regulations**
Large discrete models to study gene interactions

Context and Aims

MeForBio team:

Qualitative modelling to study
large dynamical biological systems

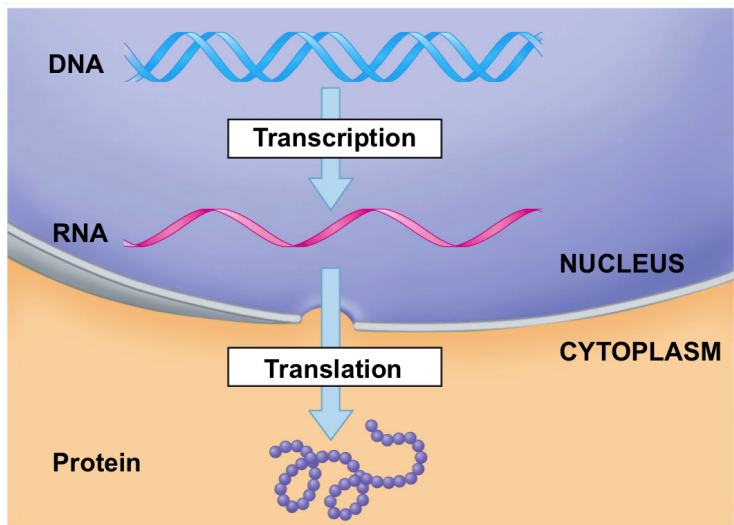
- 1) The object: **Gene regulations**
Large discrete models to study gene interactions
- 2) The method: **Static analysis**
Efficient methods thanks to the **Process Hitting** framework

Context and Aims

MeForBio team:
Qualitative modelling to study
large dynamical biological systems

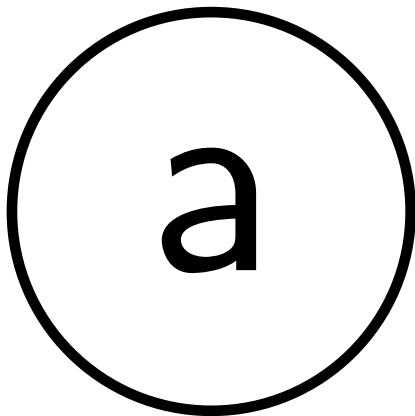
- 1) The object: **Gene regulations**
Large discrete models to study gene interactions
- 2) The method: **Static analysis**
Efficient methods thanks to the **Process Hitting** framework
- 3) The result: **Applications**
The example of **gene therapies**

Gene regulations



© 2012 Pearson Education, Inc.

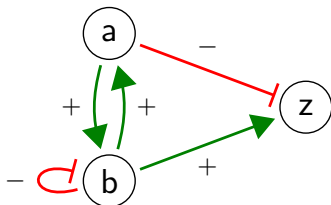
Gene regulations



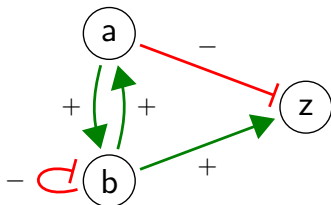
Usual biological algebraic models

[De Jong, *Journal of Computational Biology*, 2002]

Modelling interacting genes/proteins: **Boolean Networks**



Usual biological algebraic models

[De Jong, *Journal of Computational Biology*, 2002]Modelling interacting genes/proteins: **Boolean Networks**

Questions:

- How does **z** behave?
- Is it **possible** to make **a** inactive?
- If I **knock-out** **b**, what changes?


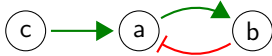
The combinatorial explosion

→ Problem: easy to understand but hard to study
exponential number of states





The combinatorial explosion

→ Problem: easy to understand but hard to study
exponential number of states

Model	Possible states
 <pre>graph LR; a((a)) -- green --> b((b)); b -- red --> a;</pre>	4
 <pre>graph LR; c((c)) -- green --> a((a)); a -- green --> b((b)); b -- red --> a;</pre>	8



The combinatorial explosion

→ Problem: easy to understand but hard to study
exponential number of states

Model	Possible states
	4
	8
⋮	⋮
(10)	1024


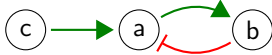
The combinatorial explosion

→ Problem: easy to understand but hard to study
exponential number of states

Model	Possible states
	4
	8
⋮	⋮
(10)	1024
(20)	1048576

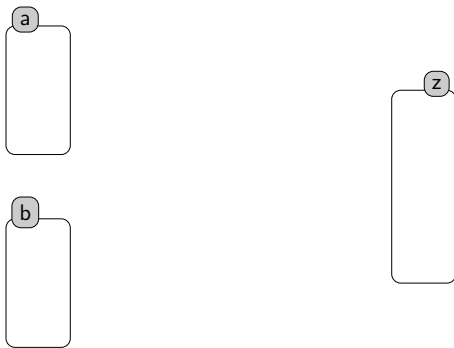
The combinatorial explosion

→ Problem: easy to understand but hard to study
exponential number of states

Model	Possible states
	4
	8
⋮	⋮
(10)	1024
(20)	1048576
(100)	12676506000000000000000000000000

The Process Hitting modelling

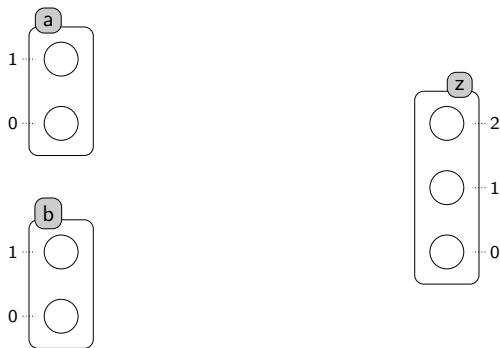
[Paulevé et al., *Transactions on Computational Systems Biology*, 2011]



Sorts: components a, b, z

The Process Hitting modelling

[Paulevé et al., *Transactions on Computational Systems Biology*, 2011]

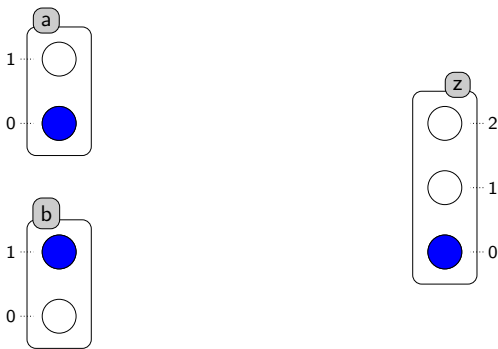


Sorts: components a, b, z

Processes: local states / levels of expression z_0, z_1, z_2

The Process Hitting modelling

[Paulevé et al., *Transactions on Computational Systems Biology*, 2011]



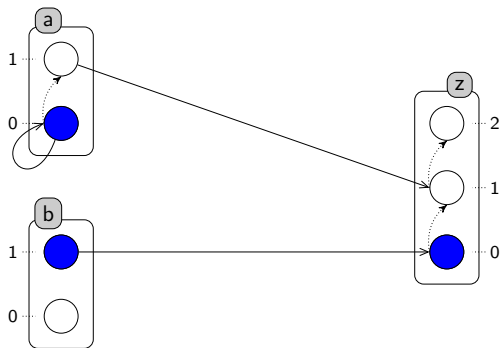
Sorts: components a, b, z

Processes: local states / levels of expression z_0, z_1, z_2

States: sets of active processes $\langle a_0, b_1, z_0 \rangle$

The Process Hitting modelling

[Paulevé et al., *Transactions on Computational Systems Biology*, 2011]



Sorts: components a, b, z

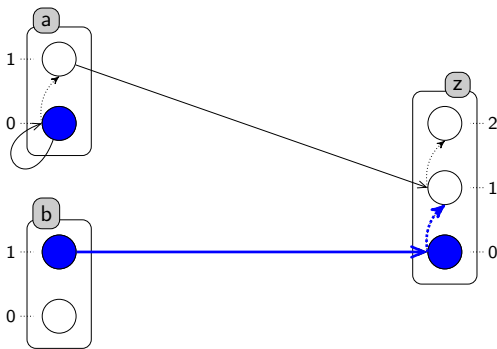
Processes: local states / levels of expression z_0, z_1, z_2

States: sets of active processes $\langle a_0, b_1, z_0 \rangle$

Actions: dynamics $b_1 \rightarrow z_0 \uparrow z_1, a_0 \rightarrow a_0 \uparrow a_1, a_1 \rightarrow z_1 \uparrow z_2$

The Process Hitting modelling

[Paulevé et al., *Transactions on Computational Systems Biology*, 2011]



Sorts: components a, b, z

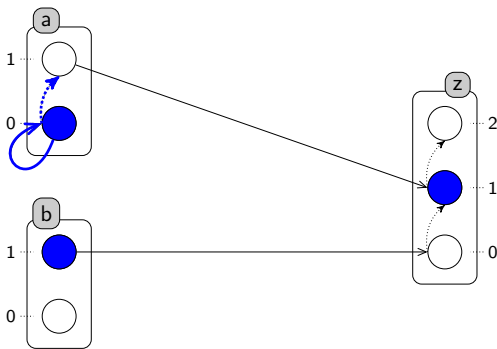
Processes: local states / levels of expression z_0, z_1, z_2

States: sets of active processes $\langle a_0, b_1, z_0 \rangle$

Actions: dynamics $\underline{b_1 \rightarrow z_0} \uparrow z_1, a_0 \rightarrow a_0 \uparrow a_1, a_1 \rightarrow z_1 \uparrow z_2$

The Process Hitting modelling

[Paulevé et al., *Transactions on Computational Systems Biology*, 2011]



Sorts: components a, b, z

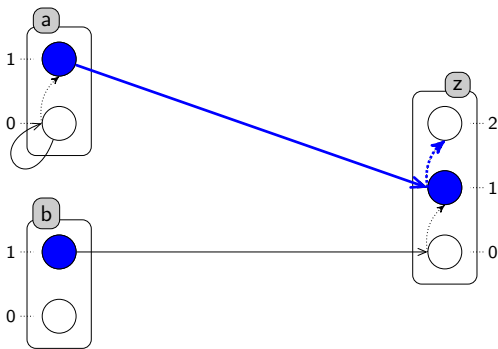
Processes: local states / levels of expression z_0, z_1, z_2

States: sets of active processes $\langle a_0, b_1, z_1 \rangle$

Actions: dynamics $b_1 \rightarrow z_0 \uparrow z_1, \underline{a_0 \rightarrow a_0 \uparrow a_1}, a_1 \rightarrow z_1 \uparrow z_2$

The Process Hitting modelling

[Paulevé et al., *Transactions on Computational Systems Biology*, 2011]



Sorts: components a, b, z

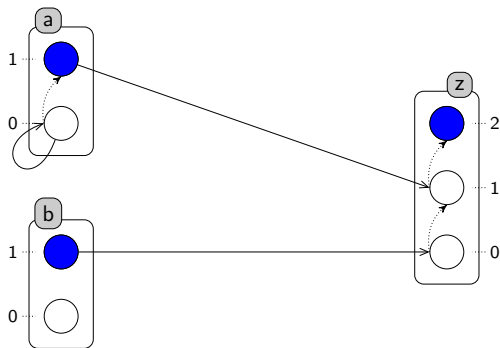
Processes: local states / levels of expression z_0, z_1, z_2

States: sets of active processes $\langle a_1, b_1, z_1 \rangle$

Actions: dynamics $b_1 \rightarrow z_0 \uparrow z_1, a_0 \rightarrow a_0 \uparrow a_1, \underline{a_1 \rightarrow z_1 \uparrow z_2}$

The Process Hitting modelling

[Paulevé et al., *Transactions on Computational Systems Biology*, 2011]



Sorts: components a, b, z

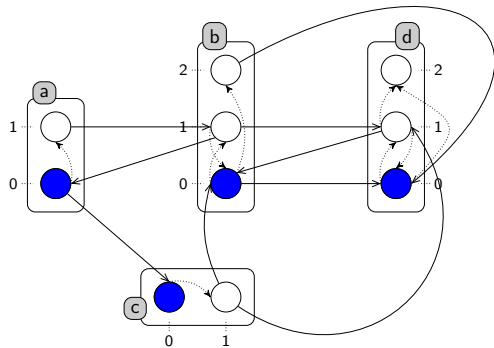
Processes: local states / levels of expression z_0, z_1, z_2

States: sets of active processes $\langle a_1, b_1, z_2 \rangle$

Actions: dynamics $b_1 \rightarrow z_0 \uparrow z_1, a_0 \rightarrow a_0 \uparrow a_1, a_1 \rightarrow z_1 \uparrow z_2$

Static analysis: successive reachability

[Paulevé *et al.*, *Mathematical Structures in Computer Science*, 2012]

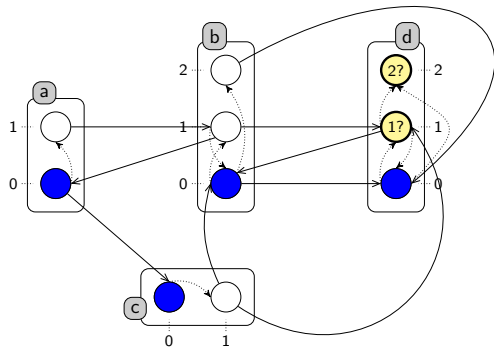


• Initial state

$\langle a_1, b_0, c_0, d_0 \rangle$

Static analysis: successive reachability

[Paulevé *et al.*, *Mathematical Structures in Computer Science*, 2012]



- Initial state

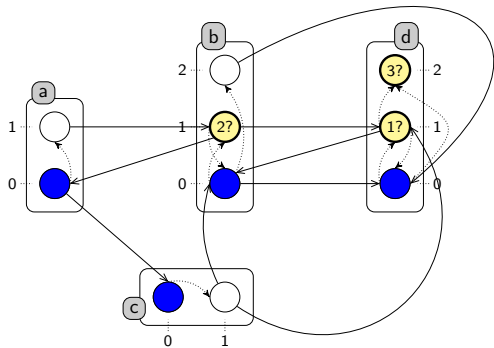
$\langle a_1, b_0, c_0, d_0 \rangle$

- Objectives

$[\uparrow d_1 :: \uparrow d_2]$

Static analysis: successive reachability

[Paulevé *et al.*, *Mathematical Structures in Computer Science*, 2012]



- Initial state

$\langle a_1, b_0, c_0, d_0 \rangle$

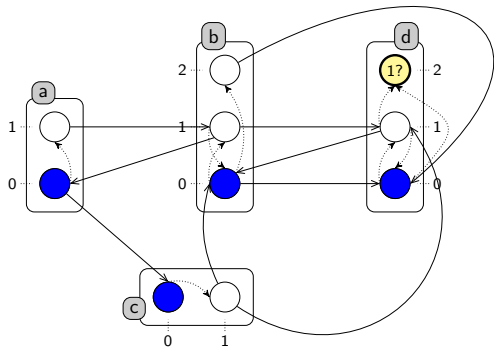
- Objectives

$[\uparrow d_1 :: \uparrow d_2]$

$[\uparrow d_1 :: \uparrow b_1 :: \uparrow d_2]$

Static analysis: successive reachability

[Paulevé *et al.*, *Mathematical Structures in Computer Science*, 2012]



- Initial state

$\langle a_1, b_0, c_0, d_0 \rangle$

- Objectives

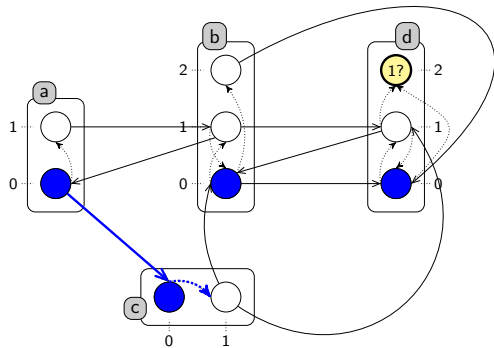
$[\uparrow d_1 :: \uparrow d_2]$

$[\uparrow d_1 :: \uparrow b_1 :: \uparrow d_2]$

$[\uparrow d_2]$

Static analysis: successive reachability

[Paulevé *et al.*, *Mathematical Structures in Computer Science*, 2012]



- Initial state

$\langle a_1, b_0, c_0, d_0 \rangle$

- Objectives

$[\uparrow d_1 :: \uparrow d_2]$

$[\uparrow d_1 :: \uparrow b_1 :: \uparrow d_2]$

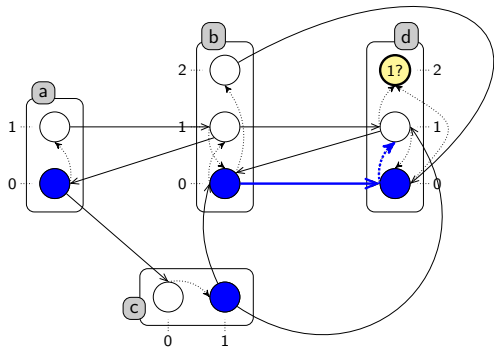
$[\uparrow d_2]$

→ Concretization of the objective = scenario

$\underline{a_0 \rightarrow c_0} \uparrow c_1 :: b_0 \rightarrow d_0 \uparrow d_1 :: c_1 \rightarrow b_0 \uparrow b_1 :: b_1 \rightarrow d_1 \uparrow d_2$

Static analysis: successive reachability

[Paulevé *et al.*, *Mathematical Structures in Computer Science*, 2012]



- Initial state

$\langle a_1, b_0, c_0, d_0 \rangle$

- Objectives

$[\uparrow d_1 :: \uparrow d_2]$

$[\uparrow d_1 :: \uparrow b_1 :: \uparrow d_2]$

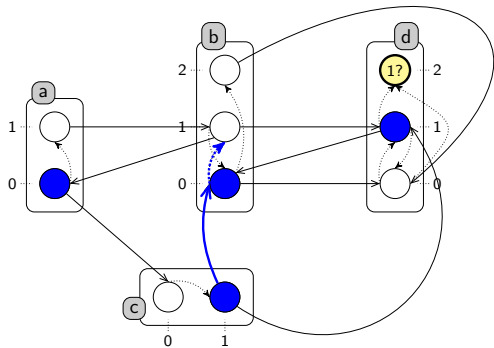
$[\uparrow d_2]$

→ Concretization of the objective = scenario

$a_0 \rightarrow c_0 \uparrow c_1 :: \underline{b_0 \rightarrow d_0 \uparrow d_1} :: c_1 \rightarrow b_0 \uparrow b_1 :: b_1 \rightarrow d_1 \uparrow d_2$

Static analysis: successive reachability

[Paulevé *et al.*, *Mathematical Structures in Computer Science*, 2012]



- Initial state

$\langle a_1, b_0, c_0, d_0 \rangle$

- Objectives

$[\uparrow d_1 :: \uparrow d_2]$

$[\uparrow d_1 :: \uparrow b_1 :: \uparrow d_2]$

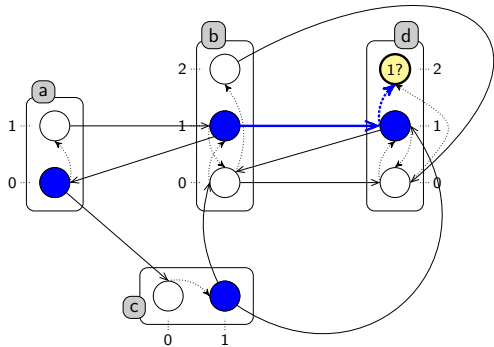
$[\uparrow d_2]$

→ Concretization of the objective = scenario

$a_0 \rightarrow c_0 \uparrow c_1 :: b_0 \rightarrow d_0 \uparrow d_1 :: \underline{c_1 \rightarrow b_0 \uparrow b_1} :: b_1 \rightarrow d_1 \uparrow d_2$

Static analysis: successive reachability

[Paulevé *et al.*, *Mathematical Structures in Computer Science*, 2012]



- Initial state

$\langle a_1, b_0, c_0, d_0 \rangle$

- Objectives

$[\uparrow d_1 :: \uparrow d_2]$

$[\uparrow d_1 :: \uparrow b_1 :: \uparrow d_2]$

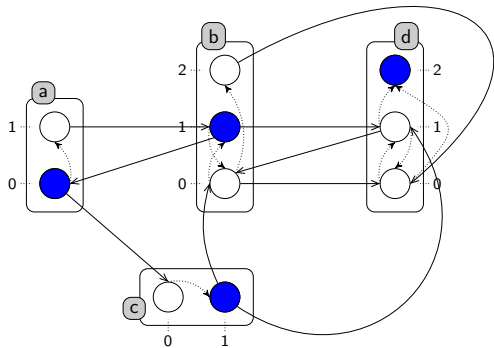
$[\uparrow d_2]$

→ Concretization of the objective = scenario

$a_0 \rightarrow c_0 \uparrow c_1 :: b_0 \rightarrow d_0 \uparrow d_1 :: c_1 \rightarrow b_0 \uparrow b_1 :: \underline{b_1 \rightarrow d_1 \uparrow d_2}$

Static analysis: successive reachability

[Paulevé *et al.*, *Mathematical Structures in Computer Science*, 2012]



- Initial state

$\langle a_1, b_0, c_0, d_0 \rangle$

- Objectives

$[\uparrow d_1 :: \uparrow d_2]$

$[\uparrow d_1 :: \uparrow b_1 :: \uparrow d_2]$

$[\uparrow d_2]$

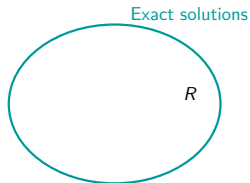
→ Concretization of the objective = scenario

$a_0 \rightarrow c_0 \uparrow c_1 :: b_0 \rightarrow d_0 \uparrow d_1 :: c_1 \rightarrow b_0 \uparrow b_1 :: b_1 \rightarrow d_1 \uparrow d_2$

Over- and Under-approximations

[Paulevé *et al.*, *Mathematical Structures in Computer Science*, 2012]

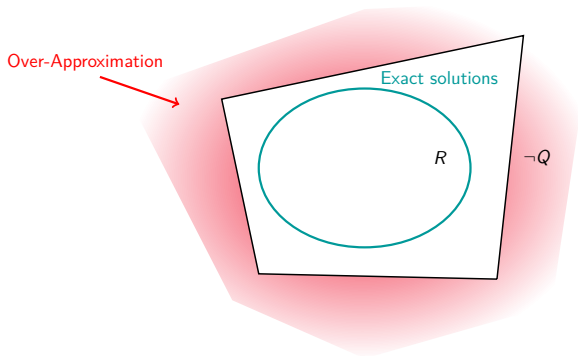
- Directly checking R is hard (**exponential**)
- Rather check **approximations** P and Q so that: $P \Rightarrow R \Rightarrow Q$



Over- and Under-approximations

[Paulevé *et al.*, *Mathematical Structures in Computer Science*, 2012]

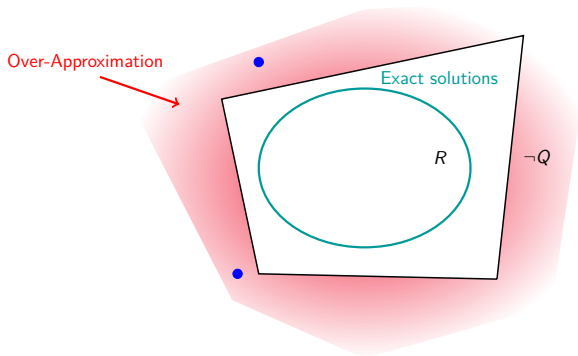
- Directly checking R is hard (**exponential**)
- Rather check **approximations** P and Q so that: $P \Rightarrow R \Rightarrow Q$



Over- and Under-approximations

[Paulevé *et al.*, *Mathematical Structures in Computer Science*, 2012]

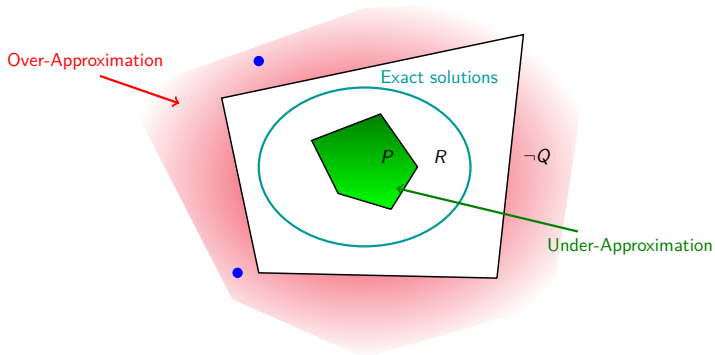
- Directly checking R is hard (**exponential**)
- Rather check **approximations** P and Q so that: $P \Rightarrow R \Rightarrow Q$



Over- and Under-approximations

[Paulevé *et al.*, *Mathematical Structures in Computer Science*, 2012]

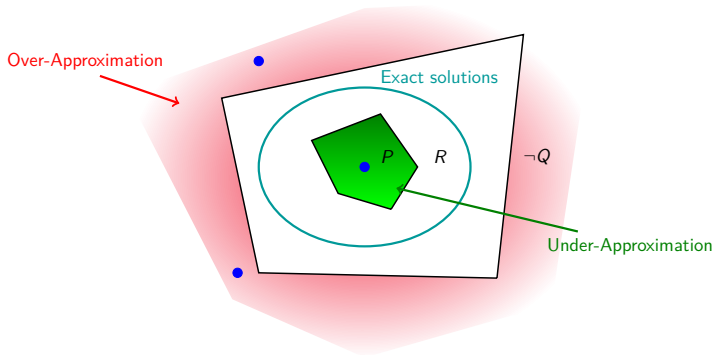
- Directly checking R is hard (**exponential**)
- Rather check **approximations** P and Q so that: $P \Rightarrow R \Rightarrow Q$



Over- and Under-approximations

[Paulevé *et al.*, *Mathematical Structures in Computer Science*, 2012]

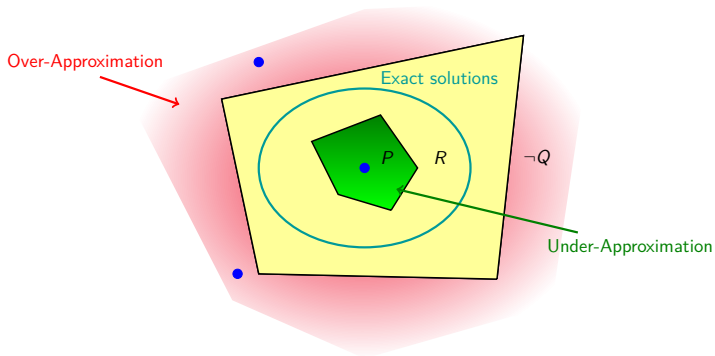
- Directly checking R is hard (**exponential**)
- Rather check **approximations** P and Q so that: $P \Rightarrow R \Rightarrow Q$



Over- and Under-approximations

[Paulevé *et al.*, *Mathematical Structures in Computer Science*, 2012]

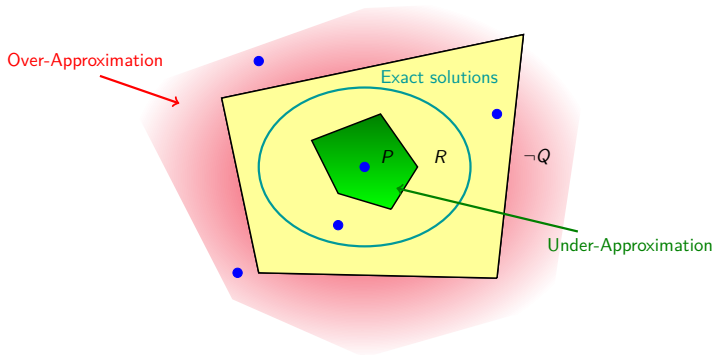
- Directly checking R is hard (**exponential**)
- Rather check **approximations** P and Q so that: $P \Rightarrow R \Rightarrow Q$



Over- and Under-approximations

[Paulevé *et al.*, *Mathematical Structures in Computer Science*, 2012]

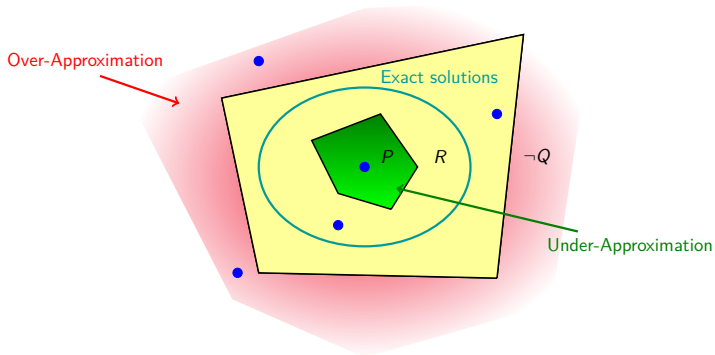
- Directly checking R is hard (**exponential**)
- Rather check **approximations** P and Q so that: $P \Rightarrow R \Rightarrow Q$



Over- and Under-approximations

[Paulevé *et al.*, *Mathematical Structures in Computer Science*, 2012]

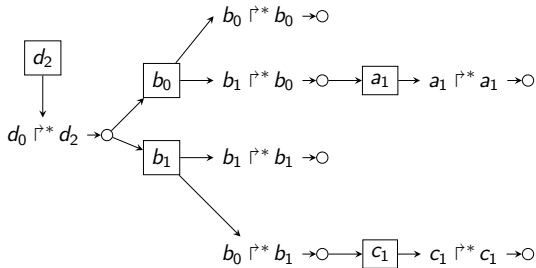
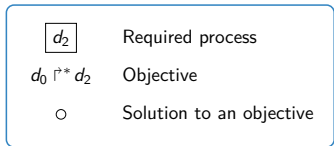
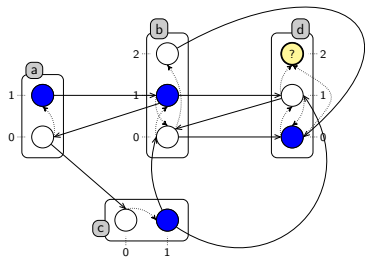
- Directly checking R is hard (**exponential**)
- Rather check **approximations** P and Q so that: $P \Rightarrow R \Rightarrow Q$



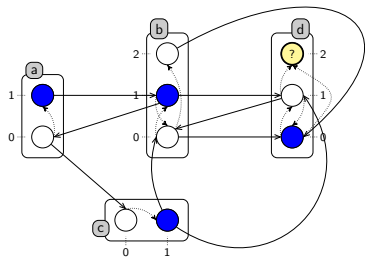
Computing P or Q is much simpler (roughly **polynomial**)

- Efficient for big models → **Hundredths of seconds**

Under-approximation

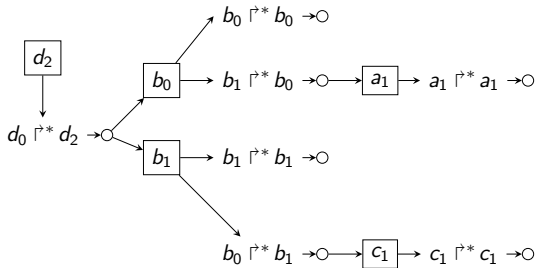
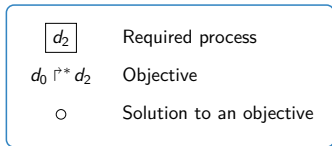


Under-approximation

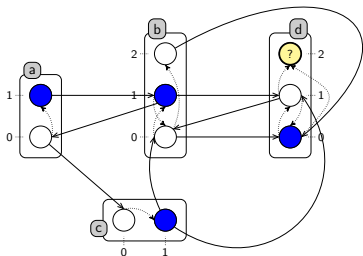


Sufficient condition:

- no cycle
- each objective has a solution



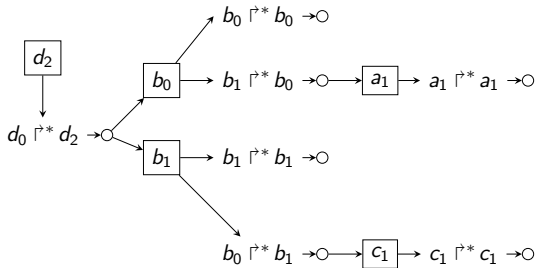
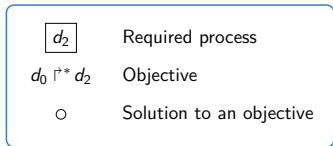
Under-approximation



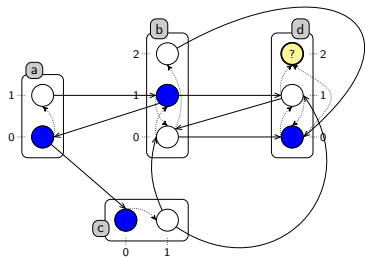
Sufficient condition:

- no cycle
- each objective has a solution

P is true $\Rightarrow R$ is true

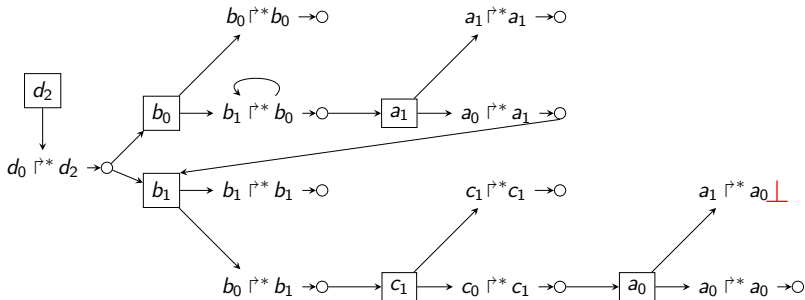


Under-approximation

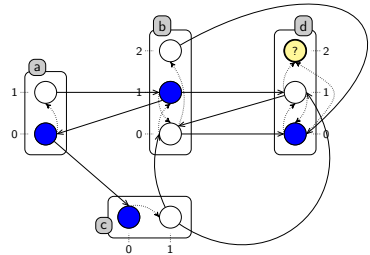


Sufficient condition:

- no cycle
- each objective has a solution



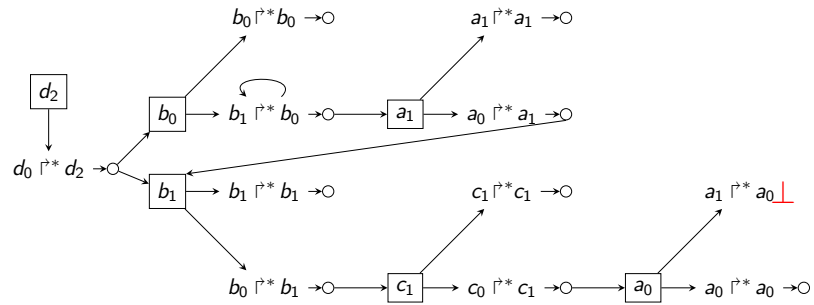
Under-approximation



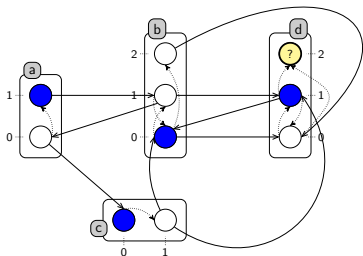
Sufficient condition:

- no cycle
- each objective has a solution

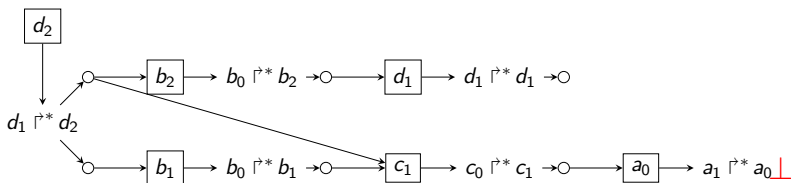
P is false \Rightarrow Inconclusive



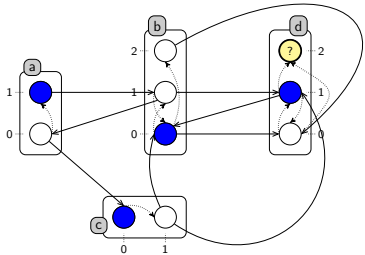
Over-approximation



Necessary condition:



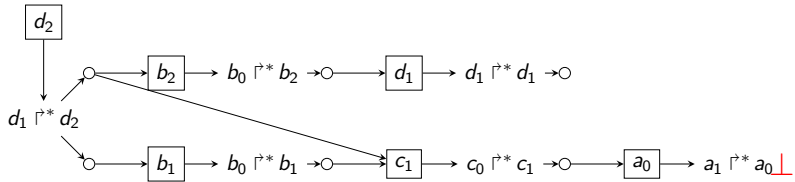
Over-approximation



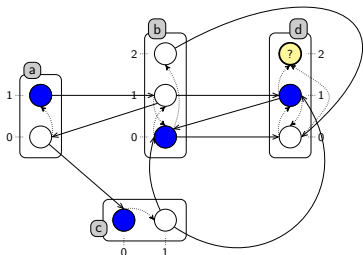
Necessary condition:

There exists a traversal with no cycle

- objective → follow **one** solution
- solution → follow **all** processes
- process → follow **all** objectives



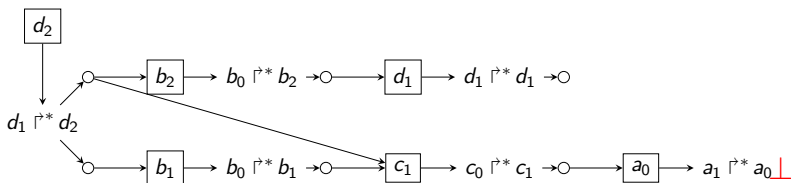
Over-approximation



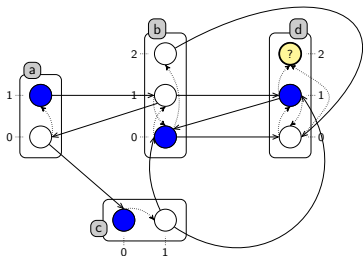
Necessary condition:

There exists a traversal with no cycle

- objective \rightarrow follow **one** solution
- solution \rightarrow follow **all** processes
- process \rightarrow follow **all** objectives



Over-approximation

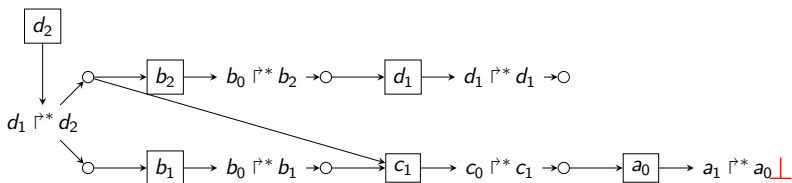


Necessary condition:

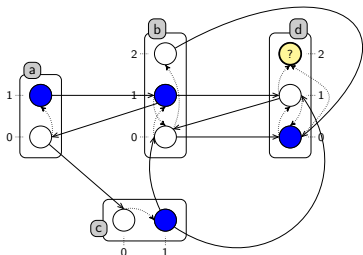
There exists a traversal with no cycle

- objective \rightarrow follow **one** solution
- solution \rightarrow follow **all** processes
- process \rightarrow follow **all** objectives

Q is false $\Rightarrow R$ is false



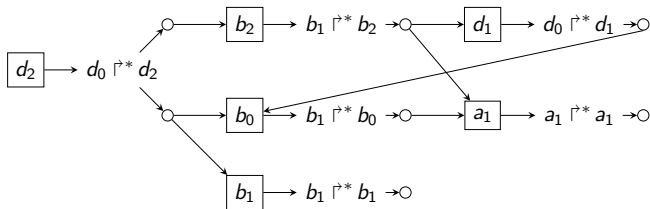
Over-approximation



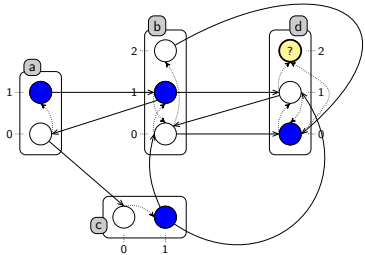
Necessary condition:

There exists a traversal with no cycle

- objective \rightarrow follow **one** solution
- solution \rightarrow follow **all** processes
- process \rightarrow follow **all** objectives



Over-approximation

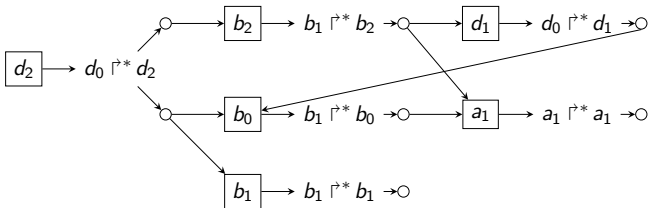


Necessary condition:

There exists a traversal with no cycle

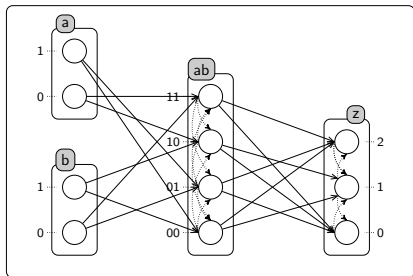
- objective → follow **one** solution
- solution → follow **all** processes
- process → follow **all** objectives

R is true ⇒ Inconclusive



Translation of PH models

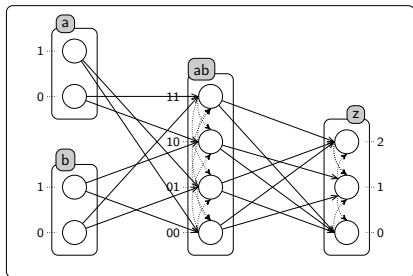
[Folschette *et al.*, *Computational Methods in Systems Biology*, 2012]



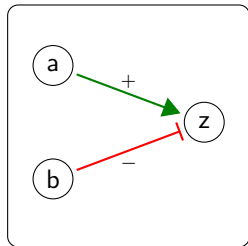
Process Hitting
Efficient but recent

Translation of PH models

[Folschette *et al.*, *Computational Methods in Systems Biology*, 2012]



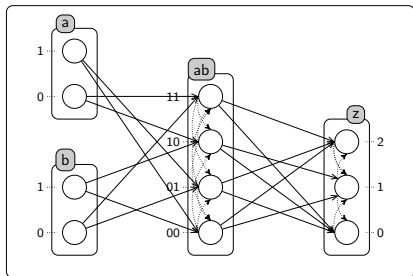
Process Hitting
Efficient but recent



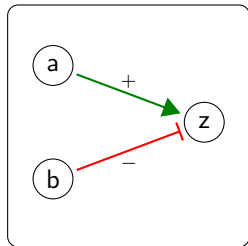
Boolean Networks
Widespread & readable

Translation of PH models

[Folschette *et al.*, *Computational Methods in Systems Biology*, 2012]



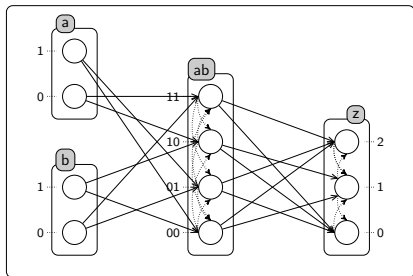
Process Hitting
Efficient but recent



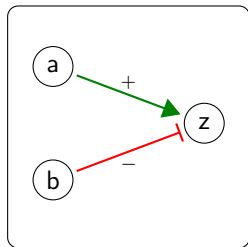
Boolean Networks
Widespread & readable

Translation of PH models

[Folschette *et al.*, *Computational Methods in Systems Biology*, 2012]



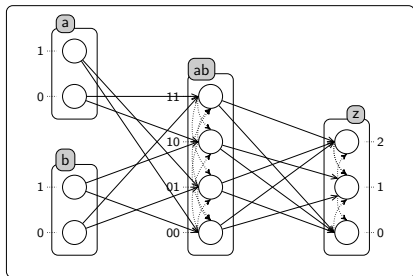
Process Hitting
Efficient but recent



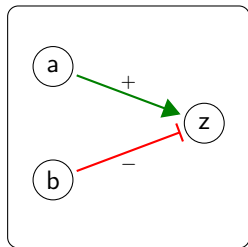
Boolean Networks
Widespread & readable

Enrichment of PH semantics

[Folschette *et al.*, CS2Bio'13, 2013]



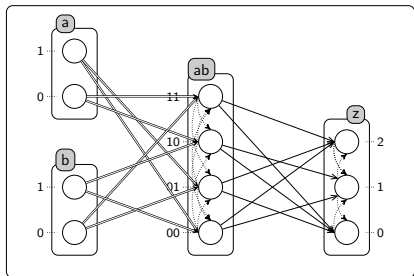
Process Hitting
Loose behaviour



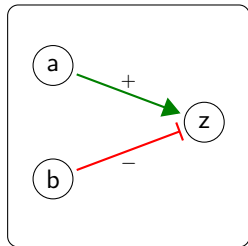
Boolean Networks
Accurate behaviour

Enrichment of PH semantics

[Folschette *et al.*, CS2Bio'13, 2013]



Process Hitting
Accurate behaviour

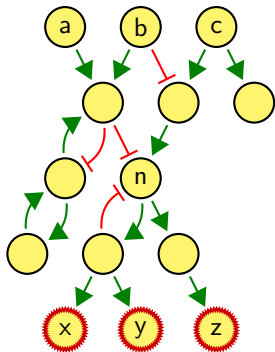


Boolean Networks
Accurate behaviour

Gene therapies

Modify DNA to cure a disease

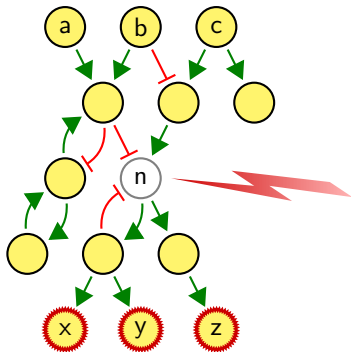
- Replace a mutated gene → remove a **harmful protein**
- Add a new gene → produce a **therapeutic protein**



Gene therapies

Modify DNA to cure a disease

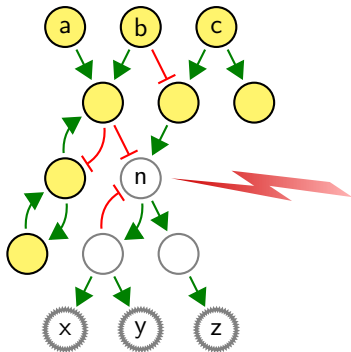
- Replace a mutated gene → remove a **harmful protein**
- Add a new gene → produce a **therapeutic protein**



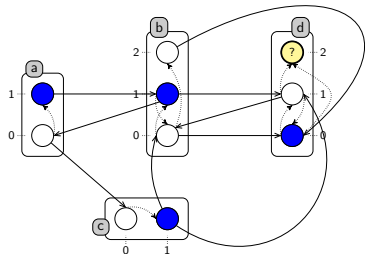
Gene therapies

Modify DNA to cure a disease

- Replace a mutated gene → remove a **harmful protein**
- Add a new gene → produce a **therapeutic protein**



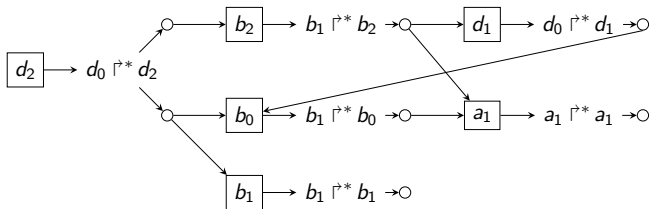
Back to the Over-approximation



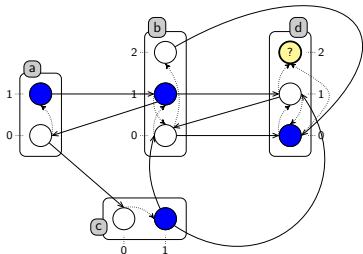
Necessary condition:

There exists a traversal with no cycle

- objective \rightarrow follow **one** solution
- solution \rightarrow follow **all** processes
- process \rightarrow follow **all** objectives



Back to the Over-approximation

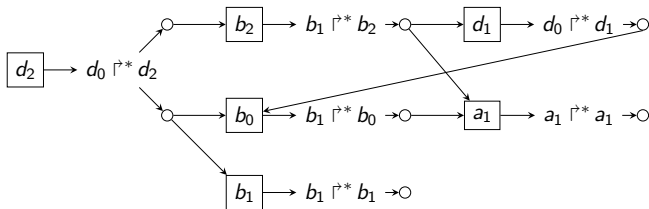


Necessary condition:

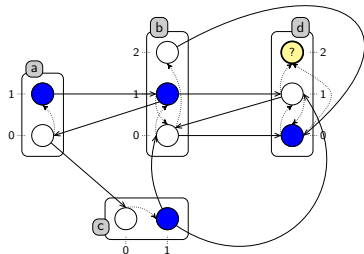
There exists a traversal with no cycle

- objective \rightarrow follow **one** solution
- solution \rightarrow follow **all** processes
- process \rightarrow follow **all** objectives

R is true \Rightarrow Inconclusive



Back to the Over-approximation

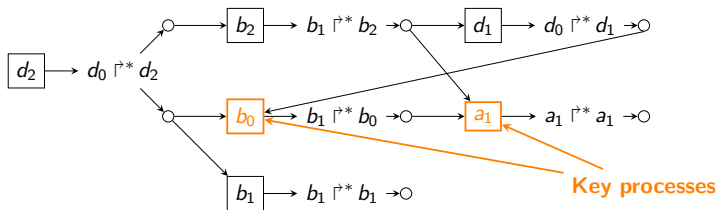


Necessary condition:

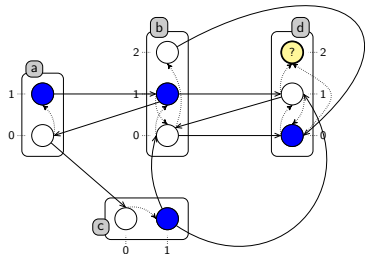
There exists a traversal with no cycle

- objective \rightarrow follow **one** solution
- solution \rightarrow follow **all** processes
- process \rightarrow follow **all** objectives

R is true \Rightarrow Inconclusive



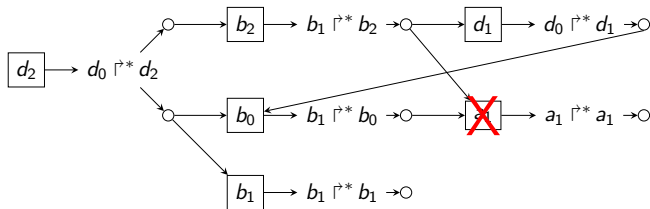
Back to the Over-approximation



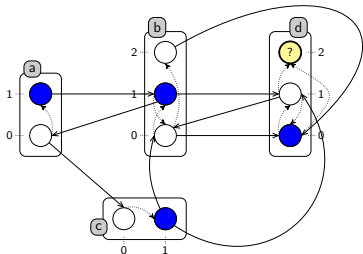
Necessary condition:

There exists a traversal with no cycle

- objective \rightarrow follow **one** solution
- solution \rightarrow follow **all** processes
- process \rightarrow follow **all** objectives



Back to the Over-approximation

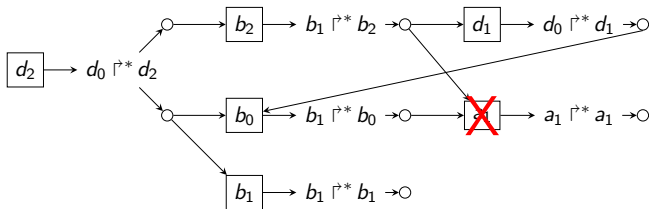


Necessary condition:

There exists a traversal with no cycle

- objective \rightarrow follow **one** solution
- solution \rightarrow follow **all** processes
- process \rightarrow follow **all** objectives

Q is false $\Rightarrow R$ is false



Summary & Conclusion

- What is Bio-informatics?
 - Qualitative modelling of **gene regulations**
 - Large models are hard to study (**exponential**)
- What do I do?
 - The **Process Hitting** modelling
 - Very efficient on large-scale models (**polynomial**)
 - My contribution: reach the expressivity of **boolean networks**
- What for?
 - **Validating & utilizing** biological models
 - Gene therapies

Bibliography

- Loïc Paulevé, Morgan Magnin, Olivier Roux. **Refining dynamics of gene regulatory networks in a stochastic π -calculus framework**. In Corrado Priami, Ralph-Johan Back, Ion Petre, and Erik de Vink, editors: *Transactions on Computational Systems Biology XIII*, Lecture Notes in Computer Science, 171-191. Springer Berlin Heidelberg, 2011.
- Loïc Paulevé, Morgan Magnin, Olivier Roux. **Static analysis of biological regulatory networks dynamics using abstract interpretation**. *Mathematical Structures in Computer Science*. 2012.
- Hidde de Jong. **Modeling and simulation of genetic regulatory systems: a literature review**, *Journal of Computational biology* 9(1), 67-103. 2002.
- Maxime Folschette, Loïc Paulevé, Katsumi Inoue, Morgan Magnin, Olivier Roux. **Concretizing the Process Hitting into Biological Regulatory Networks**. In David Gilbert and Monika Heiner, editors, *Computational Methods in Systems Biology X*, Lecture Notes in Computer Science, 166-186. Springer Berlin Heidelberg, 2012.
- Maxime Folschette, Loïc Paulevé, Morgan Magnin, Olivier Roux. **Under-approximation of Reachability in Multivalued Asynchronous Networks**. In E. Merelli and A. Troina, editors, *4th International Workshop on Interactions between Computer Science and Biology (CS2Bio'13)*, Electronic Notes in Theoretical Computer Science, Volume 299, 33-51. June 2013.

Thank you