# A logical approach to identify Boolean networks modeling cell differentiation

Stéphanie Chevalier[1], Mihaly Koltai[2], Andrei Zinovyev[2], Christine Froidevaux[1], Loïc Paulevé[1]

(1) Laboratoire de Recherche en Informatique (LRI), Université Paris Sud - Paris Saclay, CNRS

(2) U900 Cancer et génome : bioinformatique, biostatistiques et épidémiologie, Institut Curie, Inserm, MINES ParisTech

# Qualitative method

**Boolean network (BN) with n genes :**

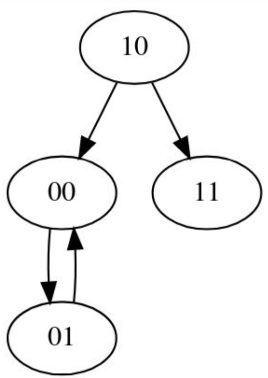
$$F = (f_1, ..., f_n),\ \forall i \in \{1, ..., n\},\ f_i : \mathbb{B}^n \rightarrow \mathbb{B} \quad (\mathbb{B} = \{0, 1\}),$$

with $f_i$ the target value of the $i^{th}$ gene

**Boolean network dynamic : transition graph**

Example :

$f_A = A \land B$
$f_B = A \lor \neg B$




Nodes : **states** (active/inactive status of all the genes)

Edges : **asynchronous updates** (one gene is updated)

# What are the genes interactions that allow the differentiation?

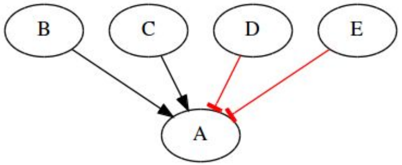
## A priori knowledge

*Influence graph* from databases and expert knowledge :
**Prior Knowledge Network (PKN)**

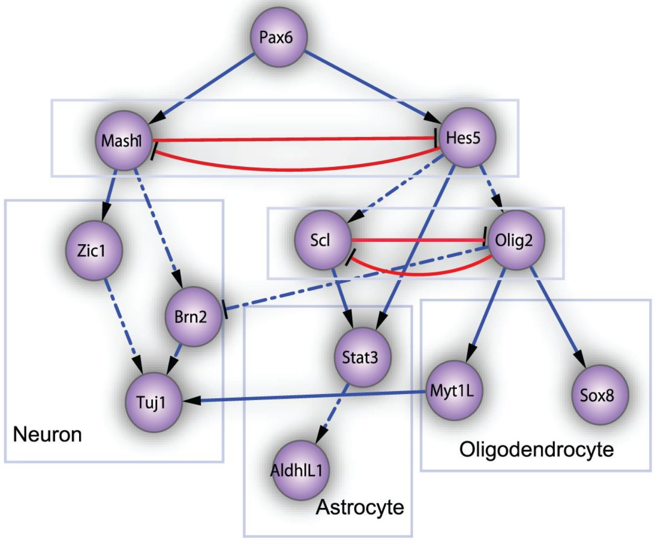
Complex biological interactions :

- **Non contextualized network**
- **Genes cooperation needed for the influence?**

Example for A activation : $B \lor \neg D \lor (C \land \neg E)$



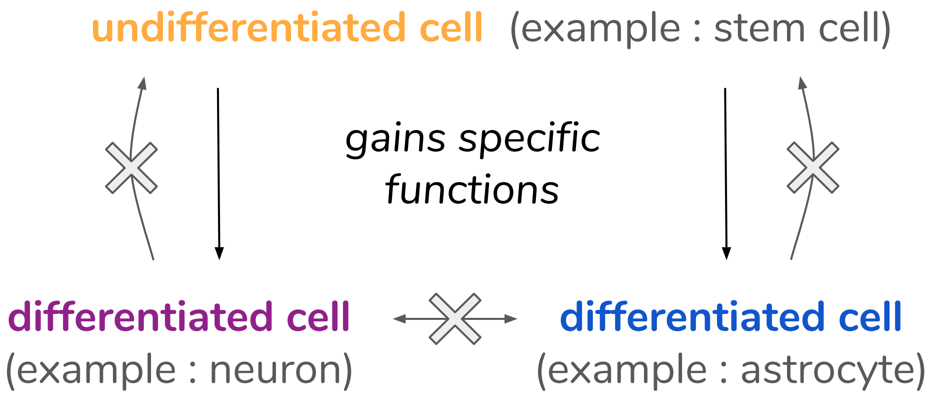

But 168 possibilities for this node !




[1] Qiu X, Ding S, Shi T. From Understanding the Development Landscape of the Canonical Fate-Switch Pair to Constructing a Dynamic Landscape for Two-Step Neural Differentiation. PLOS ONE 7(12):e49271, 2012

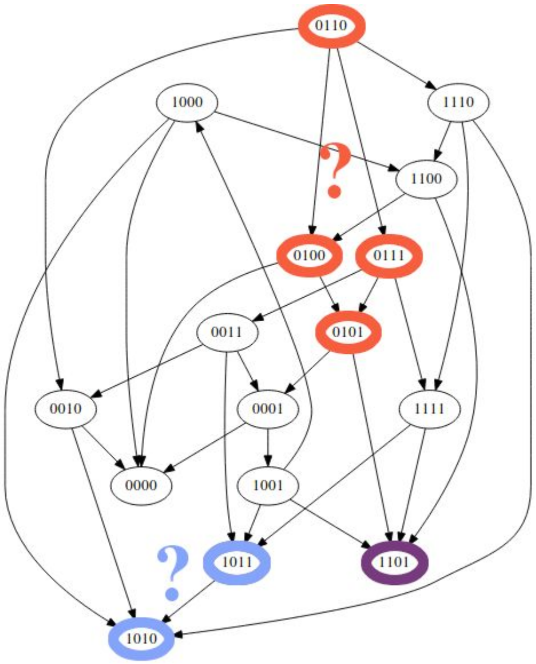
**Which are the relevant relationships involved in a biological process?**

## Differentiation process



# Inference of Boolean networks compatible with differentiation data

## Differentiation process




*Transition graph of a 4 genes BN*

Experimental measurements :

- **Time series through a differentiation process**
- **Measurements for only a few genes**

*Example :*

T0 : undifferentiated cell

- **gene1 inactive**
- **gene2 active**

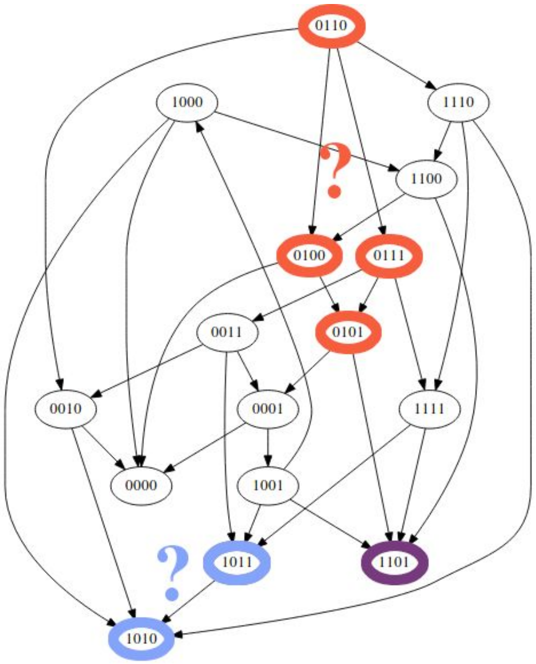
T1>T0 : differentiated cell type 1

- **gene1 active**
- **gene2 active**

T1>T0 : differentiated cell type 2

- **gene1 active**
- **gene2 inactive**

## Differentiation process




*Transition graph of a 4 genes BN*

Differentiation data interpretation (hypotheses) :

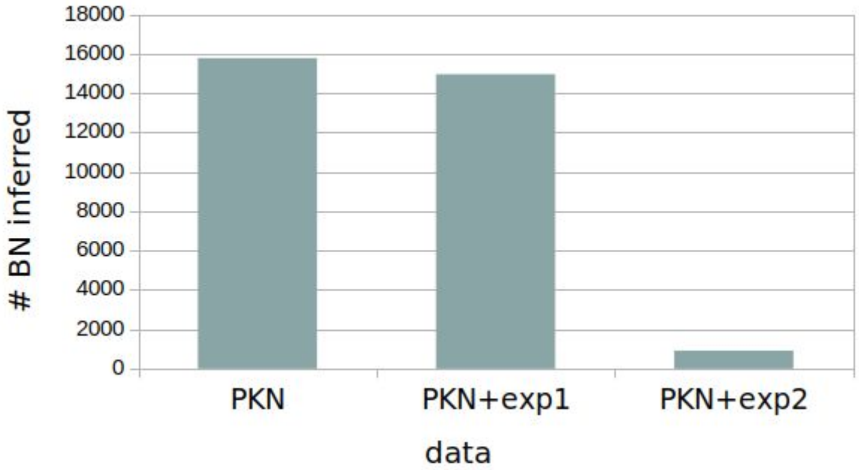
- **Reachability** : state $T_{i+1}$ reachable from $T_i$
- **Specialization** : differently differentiated states have no common descendant
- **Stability** : stable differentiated states are in attractors
- **Representativeness** : proportion of states compatible with the observations

**How infer Boolean network that respect these properties ?**

# BN exhaustive inference

Exhaustiveness importance :

- **study the inferred BN variability**
  (graph influence topology & nodes importance)
- **quantify the data informativeness :**
  (data relevance to infer BN)



# Inference of Boolean networks compatible with differentiation data

## BN exhaustive inference

**1) Direct approach :**

PKN — *Enumeration* → $BN_1, BN_2 \dots BN_k$

**combinatorial explosion**

(Dedekind numbers)

$BN_1, BN_2 \dots BN_k$, Experimental data, Properties — *Model checking* → Compatible BN

***PSPACE complete***

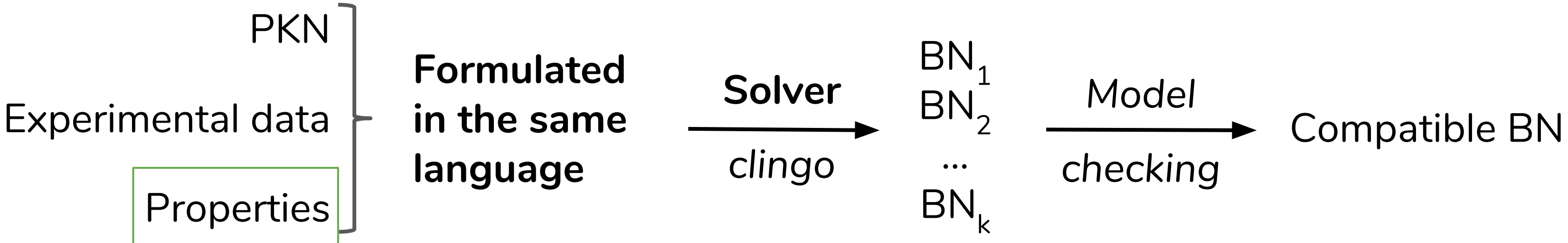
*(reachability)*

| indegree | # monotonous Boolean functions per node |
|---|---|
| 0 | 2 |
| 2 | 6 |
| 4 | 168 |
| 6 | 7 828 354 |
| 8 | 56 130 437 228 687 557 907 788 |

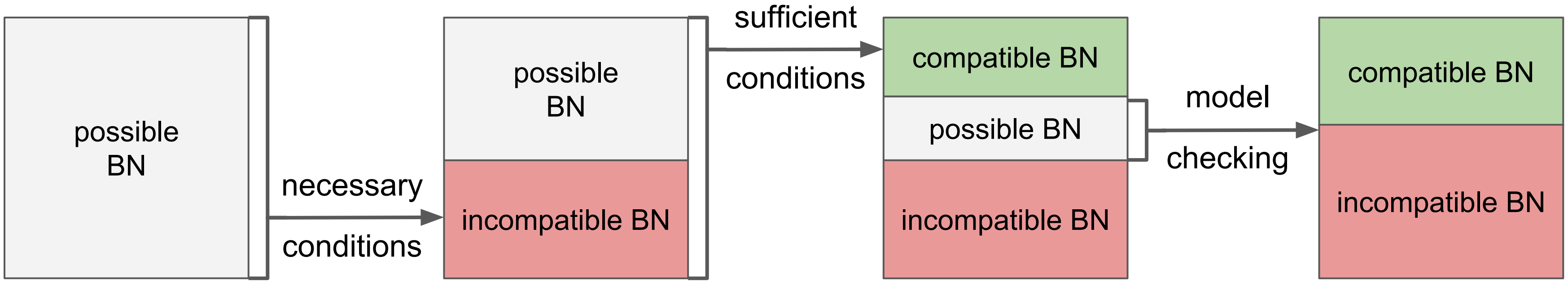
# Inference of Boolean networks compatible with differentiation data

## BN exhaustive inference

**2) Logical approach :**
Answer-Set Programming (ASP)




constraints (necessary and/or sufficient conditions)



# BN exhaustive inference

**2) Logical approach :**
Answer-Set Programming (ASP)

Necessary condition on **reachability** (*Caspots*[2])

Caspots extension :

➔ necessary and sufficient condition on a stability hypothesis : **fixed point**

ASP file for fixed point

```
eval(E,clause,N,C,-1) :- clause(N,C,L,-V) ; fp(E,L,V) ; not clamped(E,N,_).
eval(E,clause,N,C,1) :- fp(E,L,V) : clause(N,C,L,V) ; clause(N,C,_,_) ; fp(E,_,_) ; not clamped(E,N,_).

eval(E,N,V) :- clamped(E,N,V).
eval(E,N,V) :- constant(N); fp(E,N,V) ; not clamped(E,N,_).
eval(E,N,1) :- eval(E,clause,N,C,1) ; clause(N,C,_,_) ; not clamped(E,N,_).
eval(E,N,-1) :- eval(E,clause,N,C,-1) : clause(N,C,_,_) ; fp(E,_,_); node(N); not constant(N) ; not clamped(E,N,_).

:- node(N) ; fp(E,N,V) ; eval(E,N,-V).
:- constant(N) ; fp(E1,N,V) ; fp(E2,N,-V) ; not clamped(E1,N,V) ; not clamped(E2,N,-V).

fp(E,N,V) :- guessed(E,T,N,V) ; is_fixpoint(E,T).
```

[2] Max Ostrowski, Loïc Paulevé, Torsten Schaub, Anne Siegel, Carito Guziolowski. Boolean Network Identification from Perturbation Time Series Data combining Dynamics Abstraction and Logic Programming. BioSystems, Elsevier, 2016

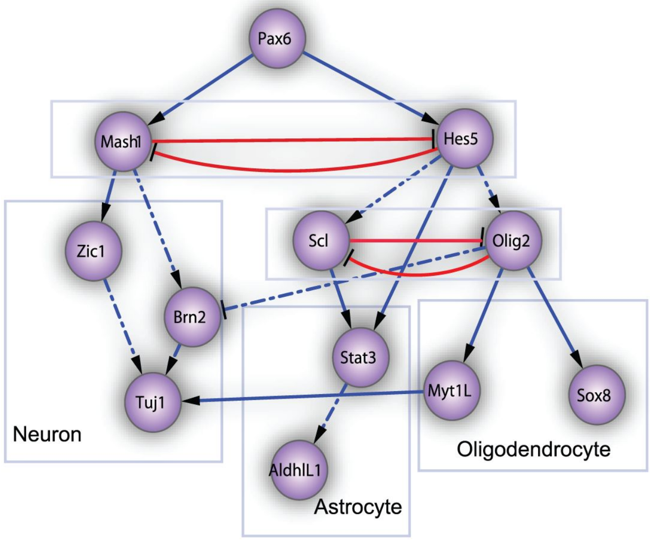
# Inference of Boolean networks compatible with differentiation data

## BN exhaustive inference

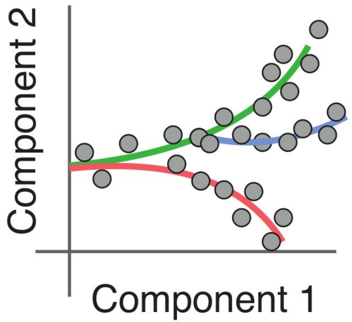
**2) Logical approach :**
Answer-Set Programming (ASP)

Small application on :

- *PKN* : CNS development[1]



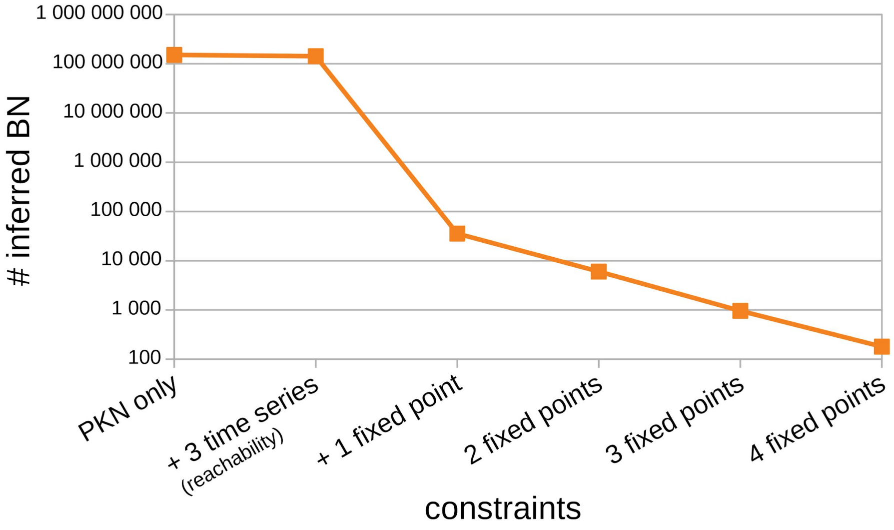

- *Data* : 4 possible fates (3 specializations + 1 steady state) → 4 fixed points



# BN exhaustive inference

**2) Logical approach :**
Answer-Set Programming (ASP)

## Constraints impact : reachability & fixed point



# Conclusion & Perspectives

**Exhaustive model inference** for **differentiation data**

- thanks to **constraints** fitting the experimental conditions (fixed points already implemented)
- allows **data informativeness quantification**

*Coming work :*

- Implementation of **negative reachability conditions**
- Generalization of the method to **mutation data** (perturbed differentiation data) and **quantitative** mutation data (fates proportion w.r.t. mutation)
- **Predict mutations combinations** that trigger fates

# BN exhaustive inference

**2) Logical approach :**
Answer-Set Programming (ASP)

PKN
Experimental data
Properties

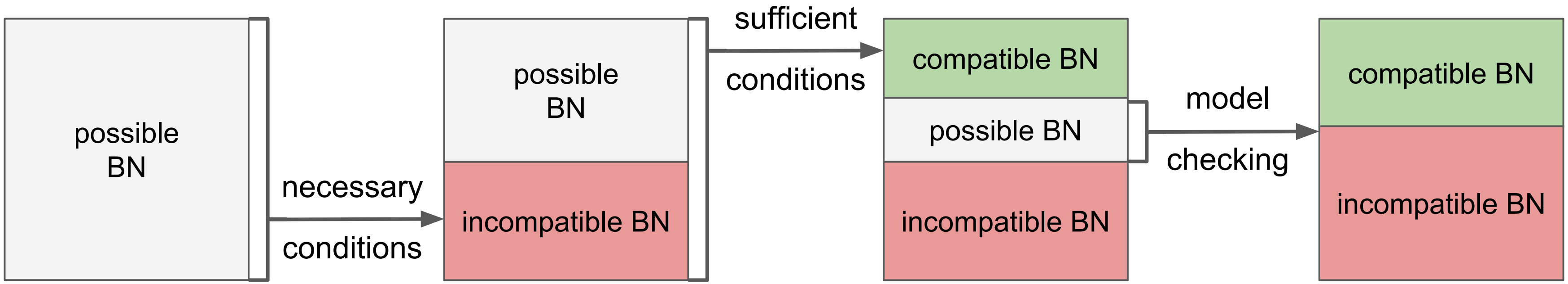
**Formulated in the same language** → **Solver** *clingo* → $BN_1$, $BN_2$, …, $BN_k$ → *Model checking* → Compatible BN

Properties → constraints (necessary and/or sufficient conditions)

## BN exhaustive inference

**2) Logical approach :**
Answer-Set Programming (ASP)

constraints (necessary and/or sufficient conditions)




- necessary condition on **reachability** (*Caspots*[2])
- necessary and sufficient condition on a stability hypothesis : **fixed point**

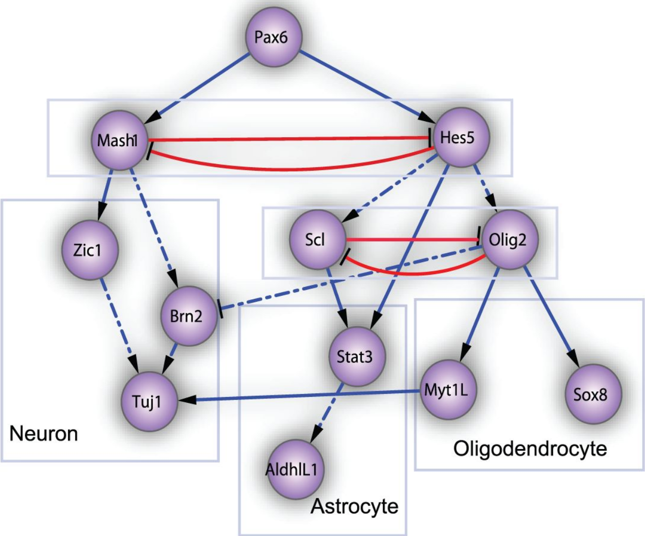
[2] Max Ostrowski, Loïc Paulevé, Torsten Schaub, Anne Siegel, Carito Guziolowski. Boolean Network Identification from Perturbation Time Series Data combining Dynamics Abstraction and Logic Programming. BioSystems, Elsevier, 2016

# Inference of Boolean networks compatible with differentiation data

## BN exhaustive inference

Small application on :

- PKN : *CNS development*[2]
- Data : 4 fates




**2) Logical approach :**
Answer-Set Programming (ASP)

**Constraints impact : reachability & fixed point**