Stéphanie Chevalier,
PhD student
@ Université Paris-Saclay, LRI (BioInfo)
& Institut Curie, U900 (SysBio)

supervisors:
Loïc Paulevé, LaBRI
Andrei Zinovyev, Institut Curie
Christine Froidevaux, LRI

Synthesis of Boolean Networks from Single Cell Trajectory-based Constraints

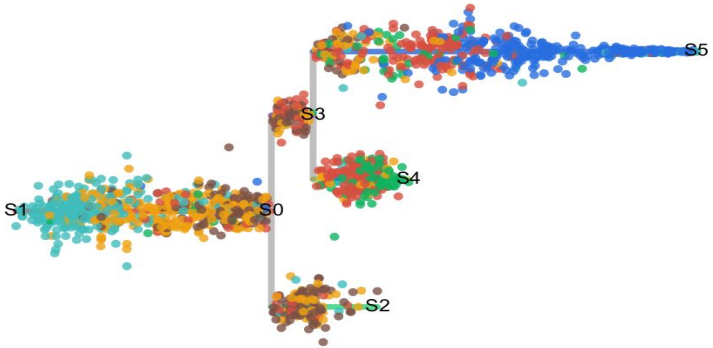
by an automatic inference of Boolean networks from static and dynamical knowledge on a system

Automatically design **models** from **knowledge** on a system

(BOOLEAN NETWORKS) (STRUCTURE & BEHAVIORS)

➔ *The aim :*

Be able to model <u>divergent processes</u> (cell differentiation, perturbations, mutants…)



Automatically design **models** from **knowledge** on a system

(BOOLEAN NETWORKS) (STRUCTURE & BEHAVIORS)

➔ *The aim :*

Be able to model divergent processes (cell differentiation, perturbations, mutants…)

Offer an <u>exhaustive enumeration</u> (to study motifs / gene importance in the process)

- ✗ evolutionary optimization algorithms: *limited access to the space of solutions*
- ✓ satisfiability problems

Automatically design **models** from **knowledge** on a system

(BOOLEAN NETWORKS) (STRUCTURE & BEHAVIORS)

➔ *The aim :*

Be able to model divergent processes (cell differentiation, perturbations, mutants…)

Offer an exhaustive enumeration (to study motifs / gene importance in the process)

Be <u>scalable</u> for networks of more than 100 nodes, with non-deterministic dynamics

Automatically design **models** from **knowledge** on a system

(BOOLEAN NETWORKS) (STRUCTURE & BEHAVIORS)

➔ *The issue :*

Direct enumeration of the BNs compatible with the input data (static and dynamical knowledge)

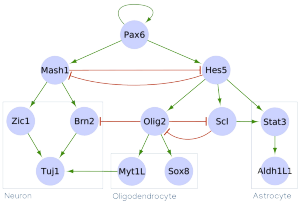
➔ *The methodology :*

Logical inference of a Boolean network (satisfiability problem) with constraints on:

the domain of its Boolean functions

to respect

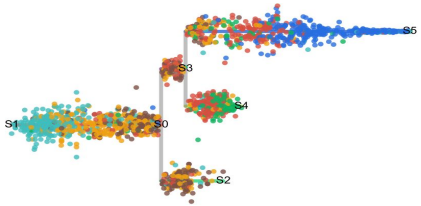
the knowledge about the structure



its dynamics

to respect

the observations

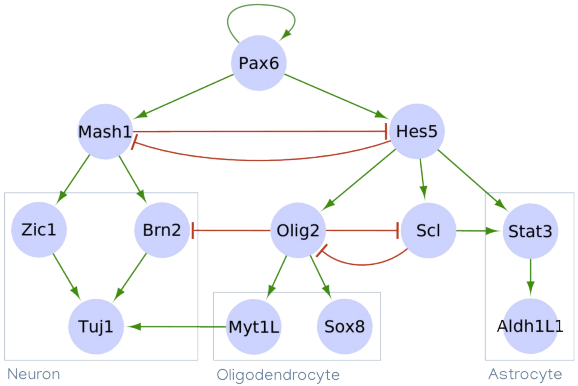


Automatically design **models** from **knowledge** on a system

(BOOLEAN NETWORKS) (STRUCTURE & BEHAVIORS)

1) **The data**

STRUCTURE: **known and putative interactions between components**

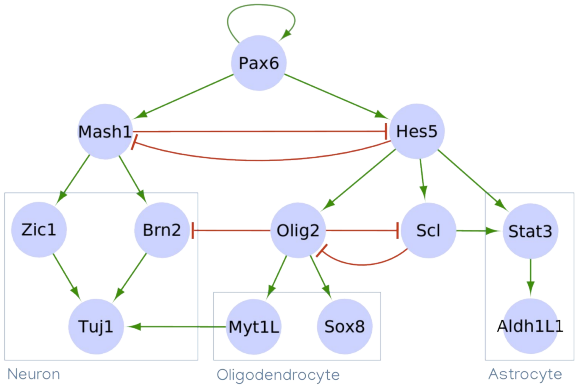


Automatically design **models** from **knowledge** on a system

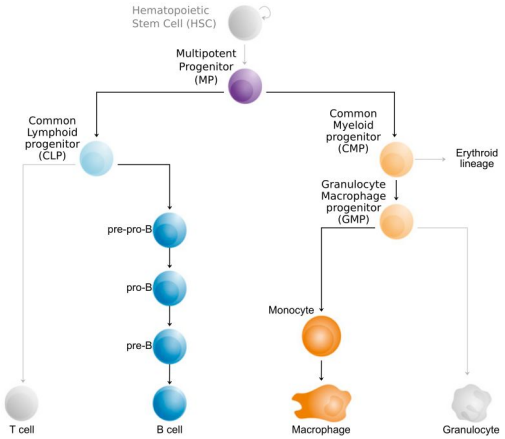
(BOOLEAN NETWORKS) (STRUCTURE & BEHAVIORS)

1) **The data**

STRUCTURE: **known and putative interactions between components**



BEHAVIORS: **dynamics of biological observations along processes** which are (most of the time) ***partial observations*** of the system

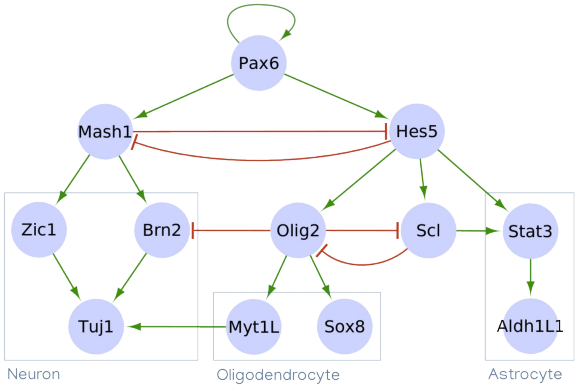


Automatically design **models** from **knowledge** on a system

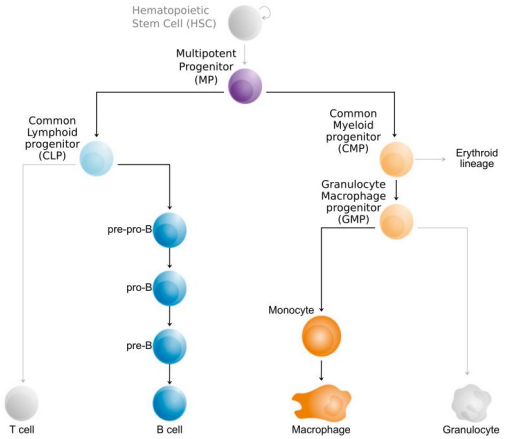
(BOOLEAN NETWORKS) (STRUCTURE & BEHAVIORS)

1) The data

STRUCTURE: **known and putative interactions between components**



BEHAVIORS: **dynamics of biological observations along processes** which are (most of the time) ***partial observations*** of the system



example:

B

gene expr. in CMP:

Flt3 = 1
Gfi1 = 0
...

F

gene expr. in macrophage:

Flt3 = 1
Gfi1 = 1
...

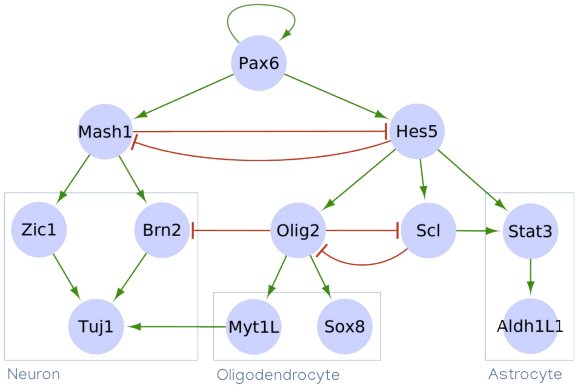
...

Automatically design **models** from **knowledge** on a system

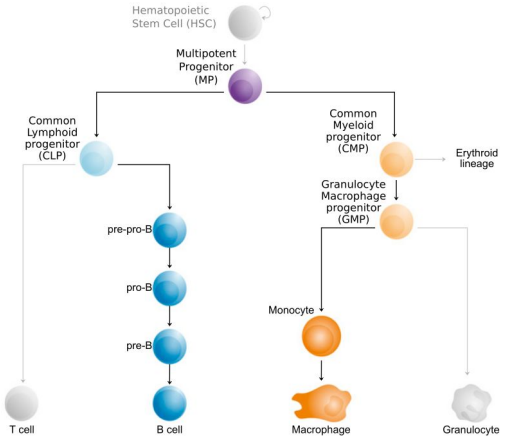
(BOOLEAN NETWORKS) (STRUCTURE & BEHAVIORS)

1) **The data**

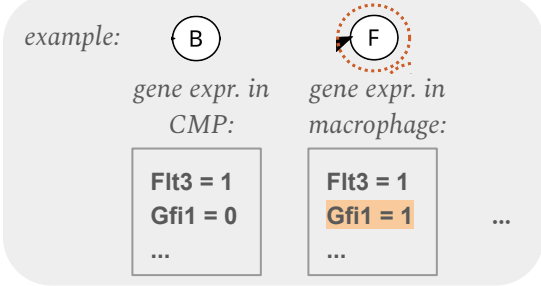
STRUCTURE: **known and putative interactions between components**

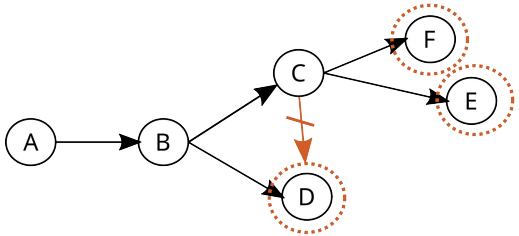


BEHAVIORS: **dynamics of biological observations along processes** which are (most of the time) ***partial observations*** of the system



example:



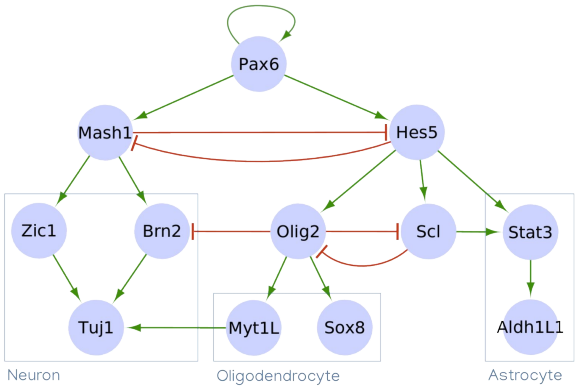


Automatically design **models** from **knowledge** on a system

(BOOLEAN NETWORKS) (STRUCTURE & BEHAVIORS)

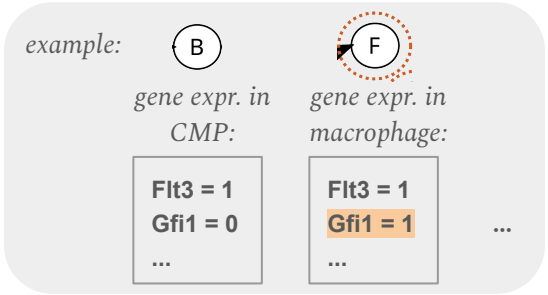
main point: **in input, the data are**

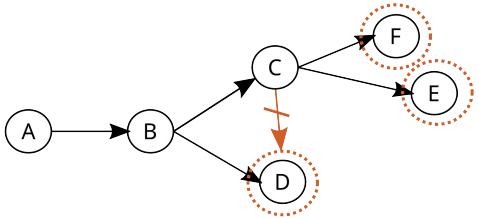
1) static knowledge (PKN)



constrains the domain of the Boolean functions of the models

2) dynamical knowledge (observations)





constrains the dynamics of the models

2) Boolean network

Discrete dynamical system

A **configuration** is a vector $x \in \{0, 1\}^n$

An **observation** is a vector $a \in \{0, 1, \text{'NA'}\}^n$

A configuration x is compatible with an observation a if $\forall i \in [n], a_i=1 \Rightarrow x_i=1$ et $a_i=0 \Rightarrow x_i=0$

Example for a BN with 3 nodes:

➔ the configuration 011 means:

- gene 1 is silenced
- genes 2 & 3 are expressed

A Boolean network of dimension *n*

is a function $f : \{0, 1\}^n \rightarrow \{0, 1\}^n$

$\forall i \in [n], f_i : \{0, 1\}^n \rightarrow \{0, 1\}$

Automatically design models from knowledge on a system

(BOOLEAN NETWORKS) (STRUCTURE & BEHAVIORS)

2) Boolean network

Discrete dynamical system

A **configuration** is a vector $x \in \{0, 1\}^n$

An **observation** is a vector $a \in \{0, 1, \text{'NA'}\}^n$

A configuration x is compatible with an observation a if $\forall i \in [n]$, $a_i=1 \Rightarrow x_i=1$ et $a_i=0 \Rightarrow x_i=0$

A Boolean network of dimension *n*

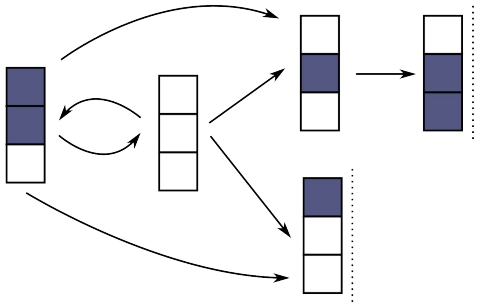
is a function $f : \{0, 1\}^n \to \{0, 1\}^n$

$\forall i \in [n]$, $f_i : \{0, 1\}^n \to \{0, 1\}$

Example of a BN with 3 nodes:

$$f_1(x) := \neg x_2$$
$$f_2(x) := \neg x_1$$
$$f_3(x) := \neg x_1 \wedge x_2$$

Asynchronous dynamics of f :



Automatically design **models** from **knowledge** on a system

(BOOLEAN NETWORKS) (STRUCTURE & BEHAVIORS)

2) Boolean network

Discrete dynamical system

A **configuration** is a vector $x \in \{0, 1\}^n$

An **observation** is a vector $a \in \{0, 1, \text{'NA'}\}^n$

A configuration x is compatible with an observation a if $\forall i \in [n]$, $a_i=1 \Rightarrow x_i=1$ et $a_i=0 \Rightarrow x_i=0$

A Boolean network of dimension *n*

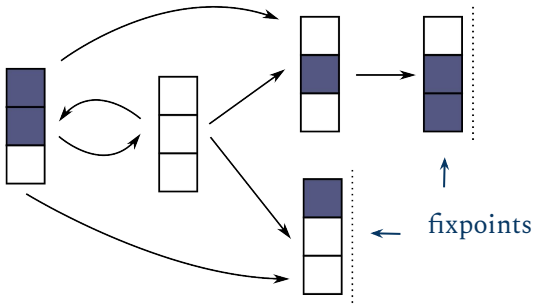
is a function $f : \{0, 1\}^n \to \{0, 1\}^n$

$\forall i \in [n]$, $f_i : \{0, 1\}^n \to \{0, 1\}$

Example of a BN with 3 nodes:

$$f_1(x) := \neg x_2$$
$$f_2(x) := \neg x_1$$
$$f_3(x) := \neg x_1 \wedge x_2$$

Asynchronous dynamics of f :



Automatically design **models** from **knowledge** on a system

(BOOLEAN NETWORKS) (STRUCTURE & BEHAVIORS)

2) Boolean network

Discrete dynamical system

A **configuration** is a vector $x \in \{0, 1\}^n$

An **observation** is a vector $a \in \{0, 1, \text{'NA'}\}^n$

A configuration x is compatible with an observation a if $\forall i \in [n]$, $a_i=1 \Rightarrow x_i=1$ et $a_i=0 \Rightarrow x_i=0$

A Boolean network of dimension *n*

is a function $f : \{0, 1\}^n \rightarrow \{0, 1\}^n$

$\forall i \in [n]$, $f_i : \{0, 1\}^n \rightarrow \{0, 1\}$

Semantics (synchronous, asynchronous, etc.) : strong impact on prediction of trajectories

➤ we rely on **Most Permissive Boolean Networks** (Paulevé et al, Nature Comm. 2020)

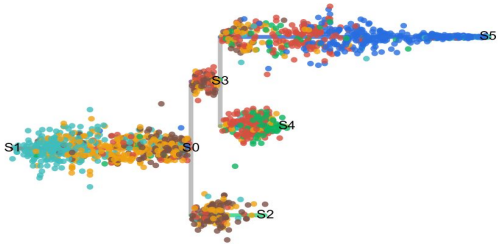
➪ brings stronger modelling guarantee w.r.t. to quantitative systems

➪ lower cost: avoid the state space explosion

Methodology to model from scRNA-seq

scRNA-seq differentiation data: gene measurements across cells at different stage of differentiation

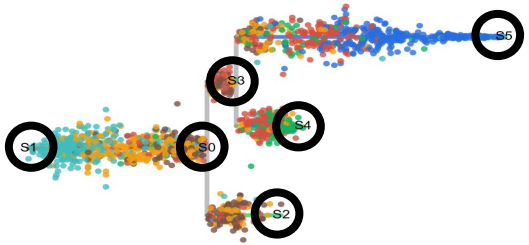
1) From data, we use **trajectory reconstruction** (e.g. STREAM) **to obtain differentiation branches and bifurcation points**



Methodology to model from scRNA-seq

scRNA-seq differentiation data: gene measurements across cells at different stage of differentiation

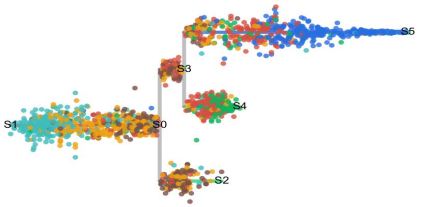
1) From data, we use **trajectory reconstruction** (e.g. STREAM) **to obtain differentiation branches and bifurcation points**



2) For each extremity of branches, **we select a pool of cells** from which **we binarize activity of genes** (possibly unknown for some of them)

Methodology to model from scRNA-seq

Satisfiability problem



3) **We translate the branches into Boolean dynamical properties**:

   a) positive reachability:
   there should be **a trajectory from the beginning to the end of each branch**

   b) negative reachability:
   there should be **no trajectory across branches**

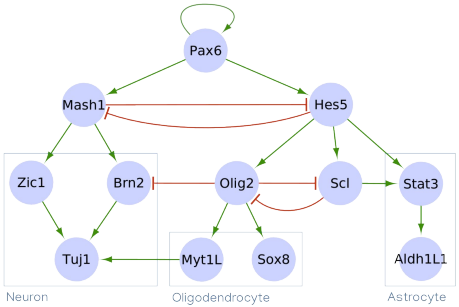
   c) stable properties:
   leafs of the graph are interpreted as **trap spaces** or **attractors** (for now fixed points)

   d) universality in the properties of the reachable fixed points:
   - we can ensure that, from a beginning of a branch, **no other fixed points than the observed one are reachable**
   - we can account for observations in **different mutants**

Methodology to model from scRNA-seq

Satisfiability problem

4) The possible Boolean functions are generated from a **prior knowledge network** (**PKN**)



Can be extract from interaction databases
(e.g. could be a full export of SIGNOR)

Methodology to model from scRNA-seq

Satisfiability problem

5) We use **logic programming** with **Answer-Set Programming** to encode the inference problem:

   we obtain a big equation, where variables relate to the logical functions in the Boolean network

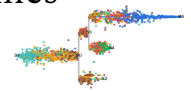
Each solution ➪ BN showing the complete bifurcation process matching with scRNA-seq data

Solver: clingo
Can scale to **BNs with thousands of components** (genes) **depending on the properties** ➤ *see ICTAI 2019 paper*

Main lines of the logic program:

- the description of a BN
- the domain of its functions
  = *PKN*

- the way to compute its dynamic
  = *semantics*
- the properties of its dynamics
  = *observations*



The solver enumerates the solutions
(solutions = BNs compatible with data = models)

Ensemble of models

Samping among possible models and run simulations

Each solution = distinct Boolean network

Exhaustive enumeration
➪ **often too many models**

we can **sample diverse solutions**
➪ build **ensembles of Boolean models**, that all share the dynamical properties related to the scRNA-seq data

we can then **simulate these ensembles**
➪ for instance assess the efficiency of a mutation

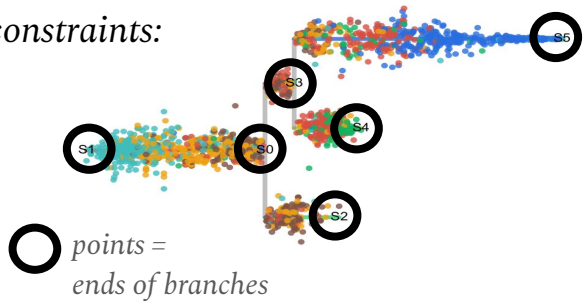
➤ *see CMSB 2020 paper*

Application in progress

Blood cell differentiation

Prior knowledge network: proteins / proteins-family / complexes / phenotypes interactions
from SIGNOR database ➪ 4837 nodes, 20717 edges

Dynamical constraints:



at each point: around 2400 genes with a binarized value

5 positive reachability (trajectory between successive points)

1 negative reachability (no trajectory between branches)

3 trap spaces (phenotype genes in final points stay fixed)

<u>reduction of the graph:</u> the biggest graph without constant nodes that can satisfied the dynamics ➤ 398 nodes

<u>model enumeration on the reduced graph:</u> soon, work in progress

Conclusion

Contribution

Infer Boolean models that reproduce the bifurcations observed in scRNA-seq differentiation data.

Scale to +1000s of genes (work in progress on real data)

Build ensembles of models and reason on them

Work in progress:

Sensitive to binarization and a priori regulatory graph:
need to account for uncertainties

Thank you for your attention !

Do you have questions?

stephanie.chevalier@universite-paris-saclay.fr
loic.pauleve@labri.fr
andrei.zinovyev@curie.fr

Our tool "BoNesis": github.com/bioasp/bonesis

Synthesis of Boolean Networks from Biological Dynamical Constraints using Answer-Set Programming

Stéphanie Chevalier, Christine Froidevaux, Andrei Zinovyev, Loïc Paulevé

Synthesis and Simulation of Ensembles of Boolean Networks for Cell Fate Decision

Stéphanie Chevalier, Vincent Noël, Laurence Calzone, Andrei Zinovyev, Loïc Paulevé

Reconciling qualitative, abstract, and scalable modeling of biological networks

Loïc Paulevé, Juraj Kolcak, Thomas Chatain, Stefan Haar