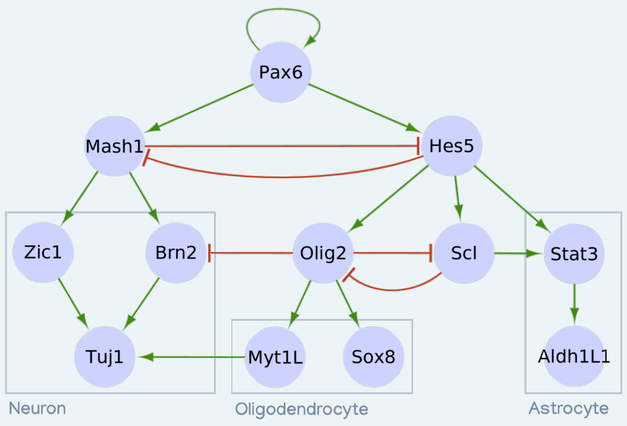
Logical approach to identify Boolean networks modelling cell differentiation

Stéphanie Chevalier[(1)], Andrei Zinovyev[(3)], Christine Froidevaux[(1)], Loïc Paulevé[(2)]

(1) LRI UMR 8623, Université Paris-Sud/Paris-Saclay, CNRS, France
(2) LaBRI UMR 5800, Univ. Bordeaux, Bordeaux INP, CNRS, France
(3) Institut Curie UMR 900, Inserm, MINES ParisTech, France

Exhaustive, non-redundant and scalable Inference of Boolean Networks

1) using (at most) the prior knowledge influences



Prior Knowledge Networks *(PKN)*

→ activation
⊣ inhibition

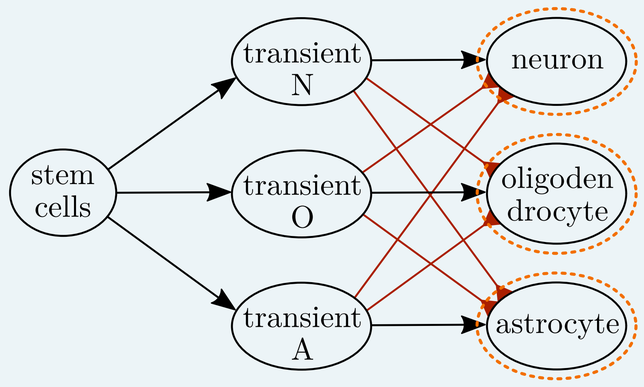
example:
PKN on the Central Nervous System development[1] *(CNS dev.)*

2) verifying the dynamical properties of differentiation

Differentiation data

- partial observations *(a set of binarized gene expression)*
- positive reachability
- negative reachability
- stability features *(trap spaces/attractors/fixpoints)*

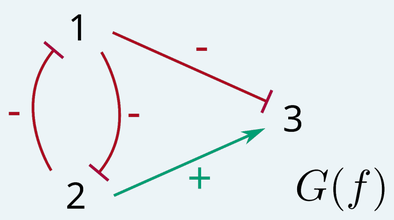
example:
dynamical model on CNS dev.



Context

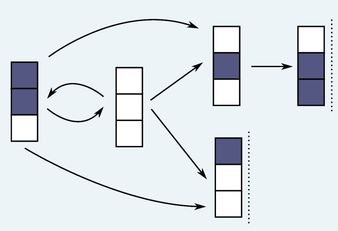
Boolean networks (BN) are intelligible and usable for modelling complex dynamics without extensive knowledge

influence graph ← *compatible with* — BN — *given a semantics* → dynamics



$G(f)$

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



Semantics: BoNesis applies the **most permissive semantics**[2] which guarantees the capture of all possible behaviours with a quantitative model

Our methodology: BoNesis

Logic program in Answer-Set Programming:

Domain of the BNs: PKN

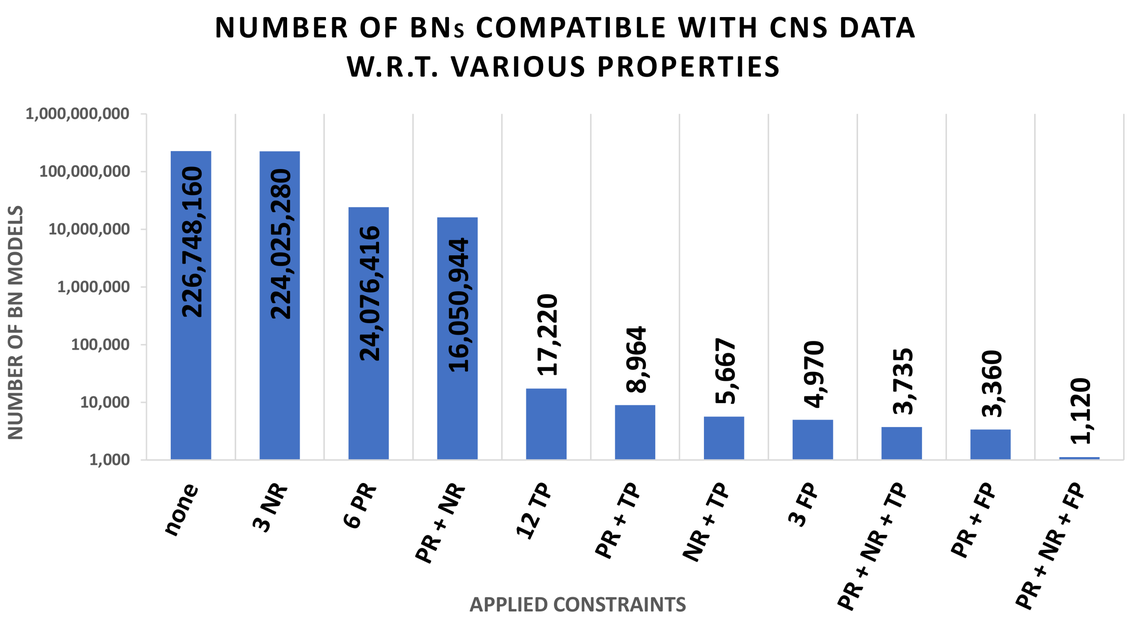
scalable up to
~100 nodes with *NR*
>1000 wo *NR*

Constraints on the rules:

- **partial observations** *(obs.)*
  n observed genes among m≥n genes in the network
- **perturbations** *overexpression/KO*
- **existence and absence of trajectories:**
  (PR) **positive reachability**
  (NR) **negative reachability**
- **stable behaviours:**
  (TP) **trap space**
  from an obs., a set of genes are fixed
  (FA) **general attractor**
  an obs. belongs to an attractor
  (FP) **fixpoint**
  an obs. belongs to a fixpoint

solver *(clingo)* → **The complete set of compatible BNs**

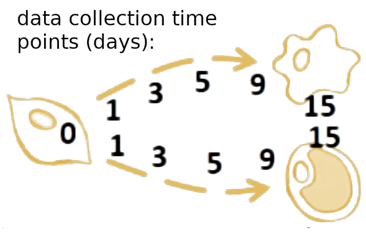
Impact of the constraints on the BN inference



1) Application with RNA-Seq data to study mesenchymal differentiation

Differentiation data

binarization *(tool RefBool[4])* of RNA-Seq bulk data collected at multiple time points



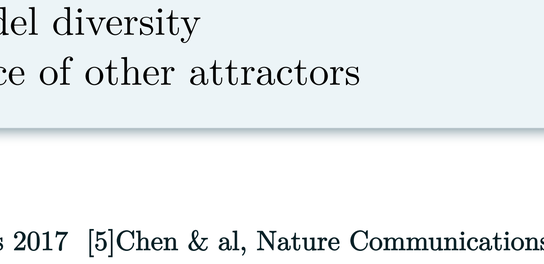
PKN

the network of transcription factors extracted from MetaCore (~1000 nodes)

2) Application with scRNA-Seq data to study hematopoiesis

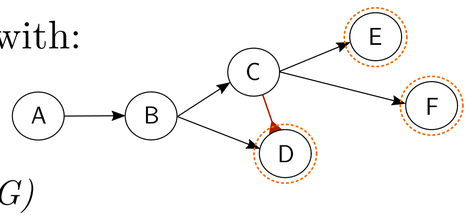
Differentiation data

from a pseudo-time trajectory,
(built from single-cell data on differentiating cells[3] and the tool STREAM[5])



construction of a dynamical model with:

- observations at extremities of branches, containing the binarized expression of 19 transition genes *(TG)*
- 5 PR + 1 NR + 3 TP or 3 FP



PKN

- without NR: the whole SIGNOR database (5656 nodes)
- with NR: the strongly connected component with the most TG (186 nodes - 19 TG)

- **Results** *in 10-30 min with ~16-32 Go*
  partial enumeration of compatible BNs
  (with optimization on their size for application 1)

- **Future work**
  analyse model diversity
  constraints on absence of other attractors

Contact
schevalier@lri.fr

References
[1]Qiu & al, Nature Methods 2017 [2]Chatain & al, CoRR 2018 [3]Nestorowa & al, Blood 2016 [4]Jung & al, Bioinformatics 2017 [5]Chen & al, Nature Communications 2019