

Coarse-grain Stochastic Abstractions of Biological Pathways

Blaise Genest, IRISA

Joint with:

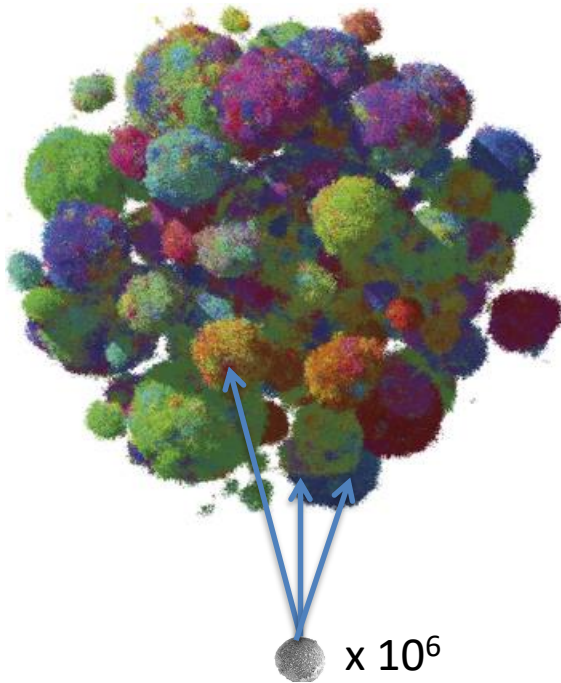
Gregory Batt, Eric Fabre,
Matthieu Pichené, Sucheendra Palaniappan.

[Bioinformatics 2017]

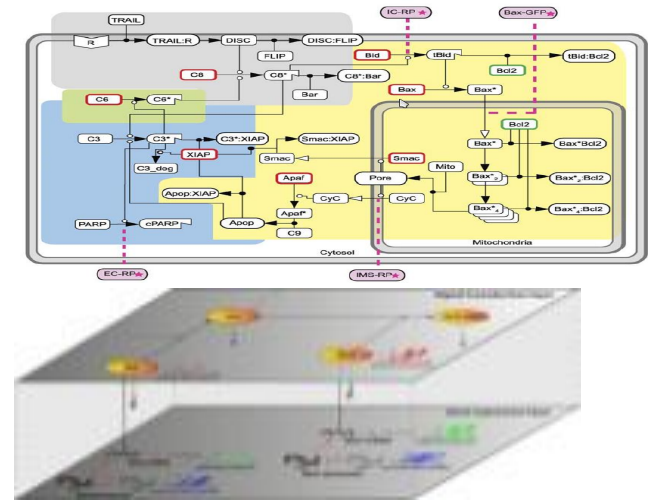


Setting

Population of cells.



for each cell, model of the cell:



Could be ODEs, stochastic model,
Hybrid stochastic deterministic model...

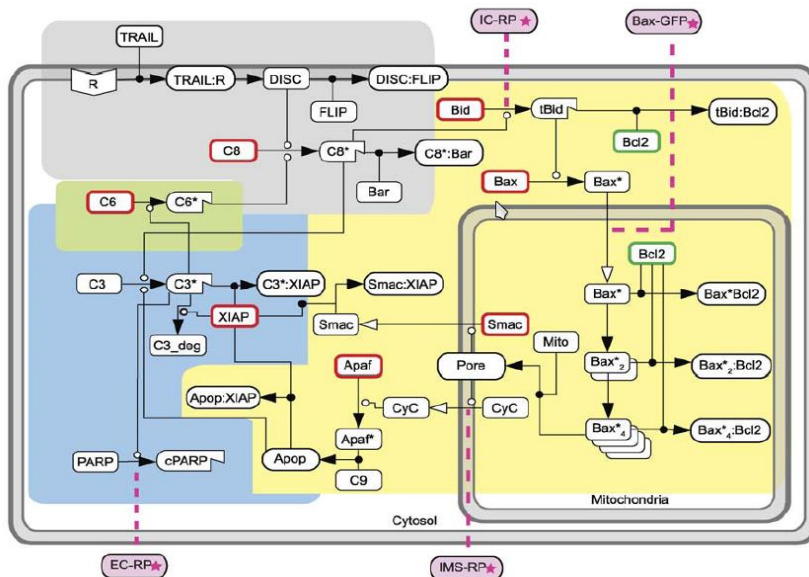
Question: How does the population behave?

Simple algorithm: simulate **10⁶** low level models and see.

Issue: High complexity model

=> **Abstractions**

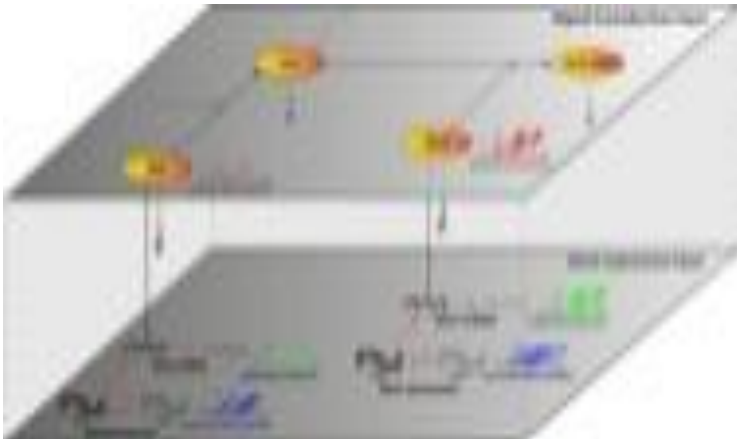
An example: TRAIL-induced apoptosis



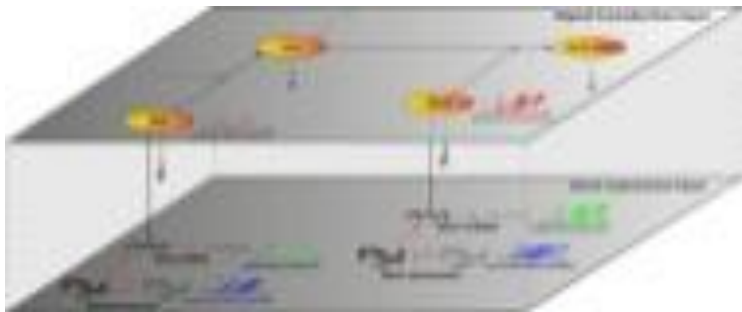
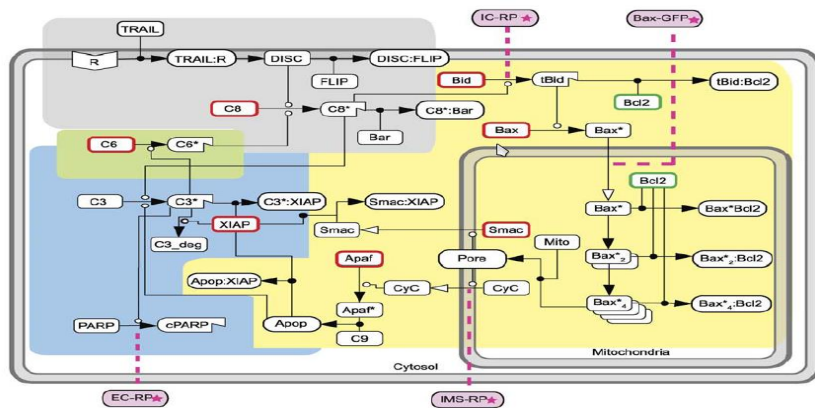
Hybrid stochastic deterministic model
[François Berteaux, Szymon Stoma, Dirk Drasdo, Gregory Batt. Modeling dynamics of cell-to-cell variability in TRAIL-induced apoptosis explains fractional killing and predicts reversible resistance. PLoS Computational Biology 2014]

52 ODE species, 96 reactions
+ 40 stochastic variables
(simple gene turnover model).

Very efficient implementation by
François Berteaux (56sec/1000 simul.)
1 simulation step represents 1 second.

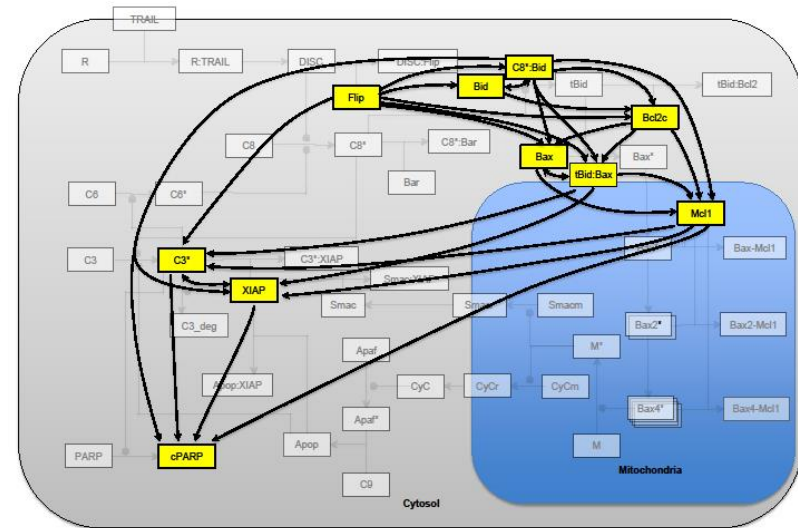
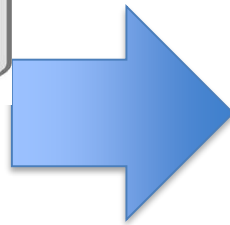


Abstracting the model for TRAIL-induced apoptosis



52 ODE species, 96 reactions
+ 40 stochastic variables

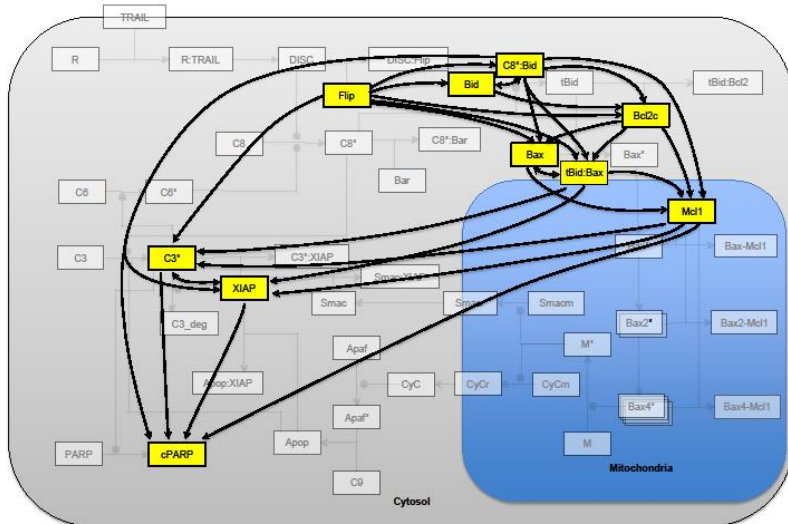
1 simulation step represents 1 second
(fine grain)



Around 10 variables (=species concentration)

1 time step corresponds to 15 min
(coarse grain)

Abstracting the model for TRAIL-induced apoptosis



Problematic:

Choose variables.

Variables Representation.

Describe evolution of variables.
(correlations between variables)

Around 10 variables (=species)

1 time steps corresponds to 15 min.

Main tool: **Information theory.**

- Entropy,
- (Conditional) Mutual information
- ...

Abstracting the model for TRAIL-induced apoptosis

Around 10 variables (= species)

=> should allow to understand if apoptosis will be triggered

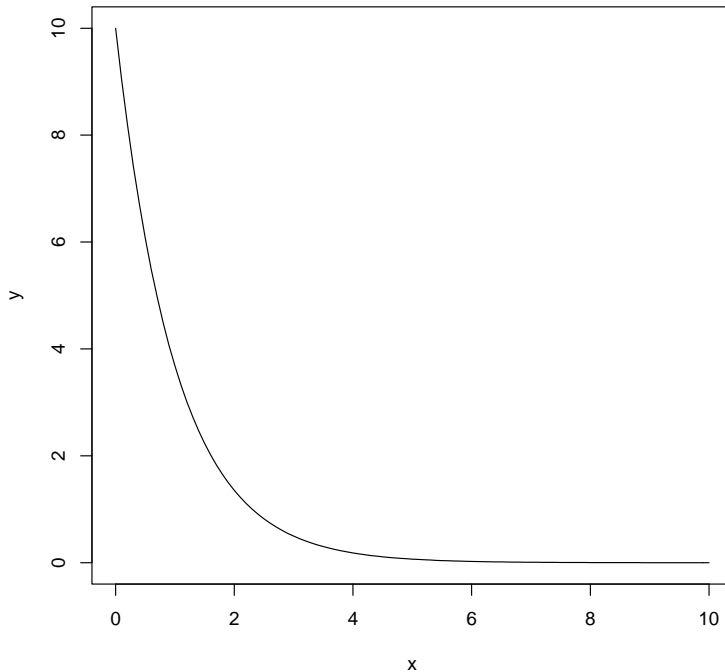
1 time steps corresponds to 15 min

=> that's the time step we are interested in.

Variables Representation

=> few discretized values: around 5
(ex: « high/low » concentrations)

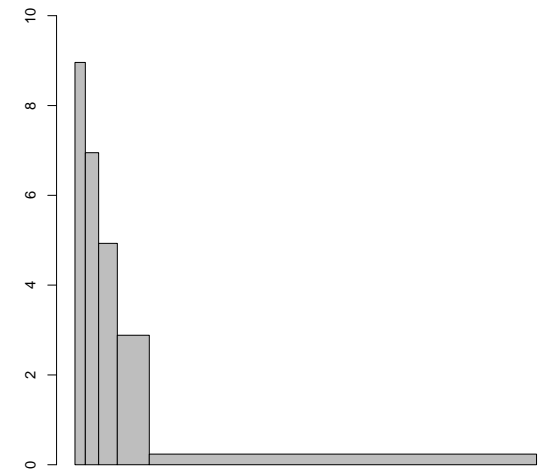
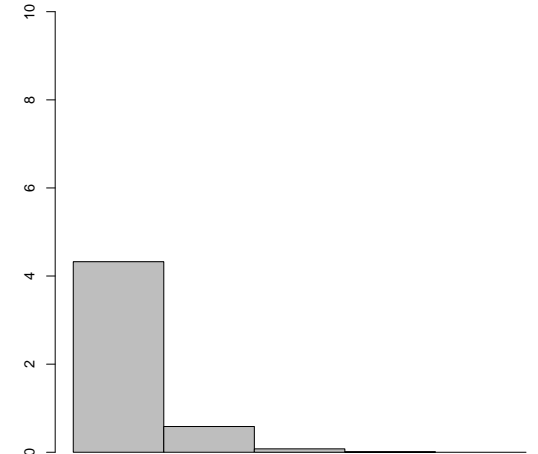
Discretization Methods



uniform



Max entropy



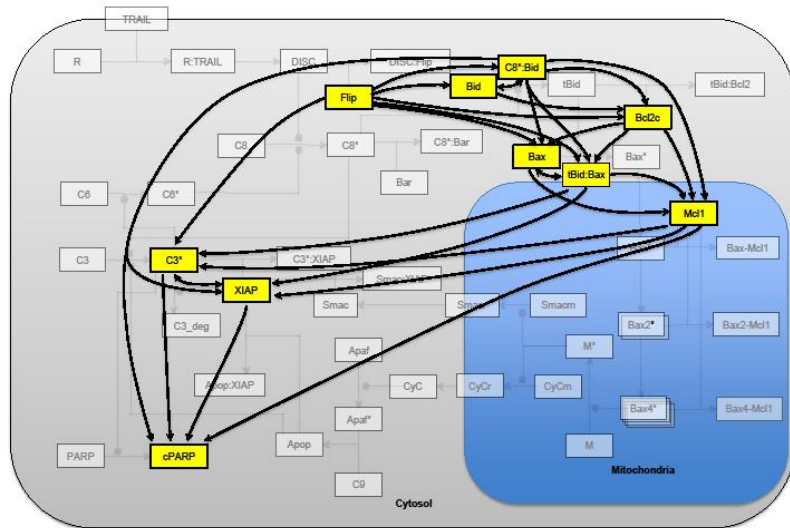
Ours: default = **uniform**. (e.g. native species)
check **entropy** of discretized variable.
If too low (<0.4), use **max entropy**
(e.g. for polymerized complexes)

...
(ex : Lloyd-
Max reducing
distorsion)

5 discretized values

Choosing Variables

should allow to understand if apoptosis will be triggered



Correlate the initial value of variables
With death/alive signal of the cell
at the end of the pathway (8 hours).

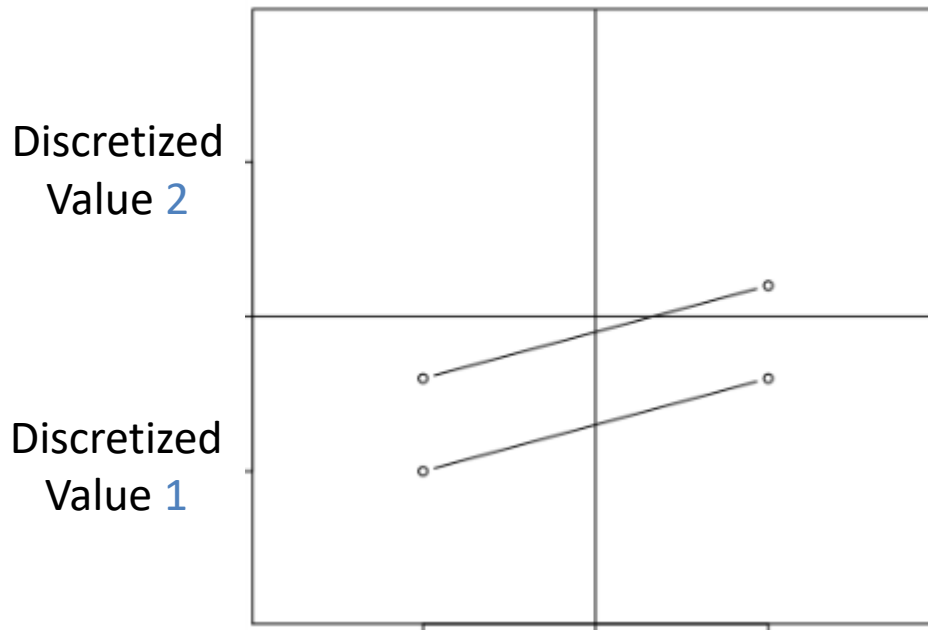
using mutual information for that.

⇒ Select 6 **native** variables in that way.

⇒ Add the death signal (cPARP > some high level)

Evolution: Stochastic Model

Represent **evolution** of discretized variables.



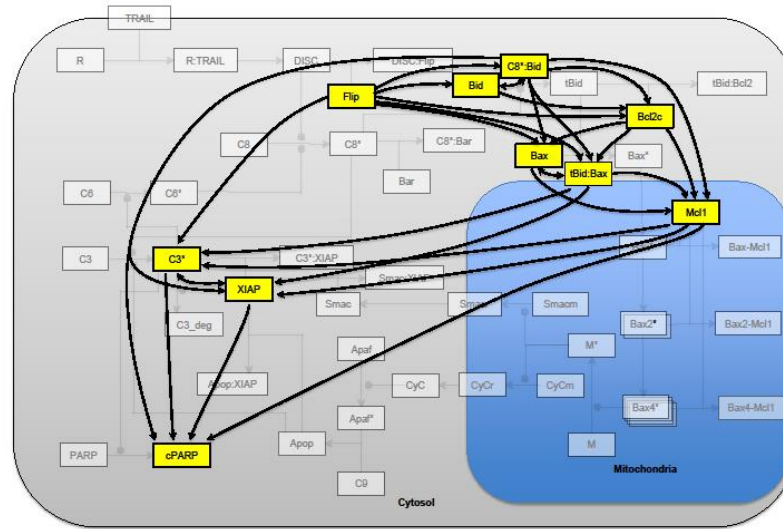
Even deterministic models (e.g. ODEs) become non-deterministic because of discretization:

Cope with that using **stochasticity** (e.g. Markov chains).

$\text{Prob}(\text{value 1} \Rightarrow \text{value 1})=0.8$

$\text{Prob}(\text{value 1} \Rightarrow \text{value 2})=0.2$

Markov Chains



For performance reason, we use compact representation of Markov chains, in the form of **Dynamic Bayesian Networks (DBNs)**.

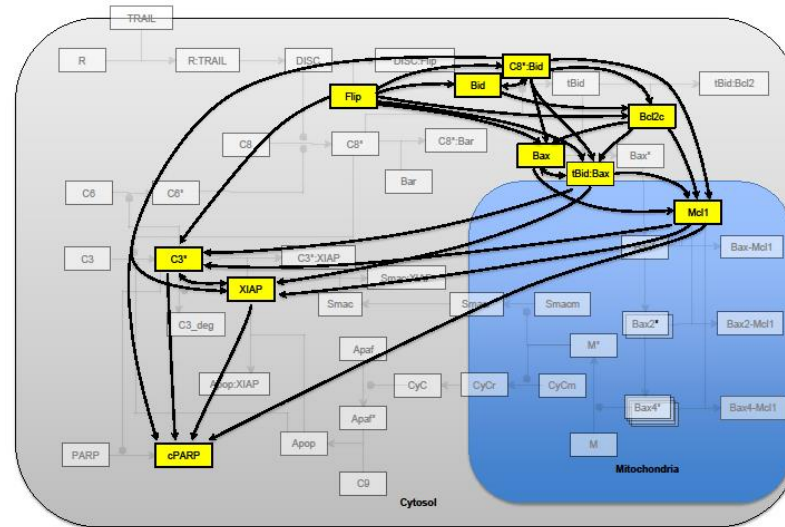
Set of variables, and:

$$\text{Proba}(\text{Cvar}_1^{t+1}=\text{low} \dots \text{Cvar}_{10}^{t+1}=\text{low} \mid \text{Cvar}_1^t=\text{low}, \dots \text{Cvar}_{10}^t=\text{high}\dots)=0.01$$

....

$$\Rightarrow 5^{20} = 95 \times 10^{12} \text{ values. Bit too much.}$$

DBNs



Evolution of each variable is dependant only on past (no cross dependencies)

Each variable depends only on few variables at the previous time points called « parents »

Set of variables, and:

$$\text{Proba}(\text{Cvar}_1^{t+1}=\text{high} \mid \text{Cvar}_1^t=\text{low}, \text{Cvar}_2^t=\text{high}) = 0.1.$$

.....

=> $5^3 = 125$ values for each variable => 1250 values for 10 variables.

Parent relations

Represent evolution of variables.
(correlations between variables)

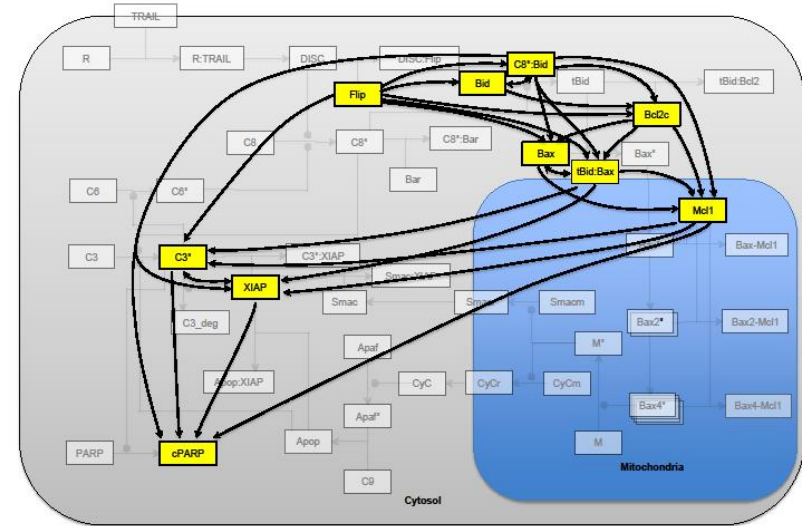
Use conditional mutual information
to select **parents**:

Select Y maximizing $MI(X^{t+1}, Y^t)$:
Quantifies how much Y^t can predict X^{t+1}

Select Z maximizing $CMI(X^{t+1}, Z^t | Y^t)$
Quantifies how much Z^t can predict X^{t+1} on top of what Y^t tells

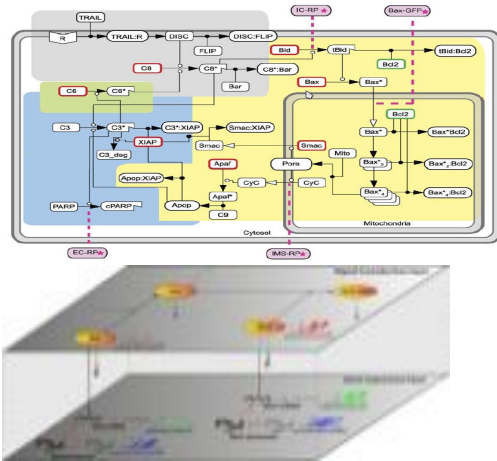
...

We found that **4 parents** was a good compromise between accuracy and speed

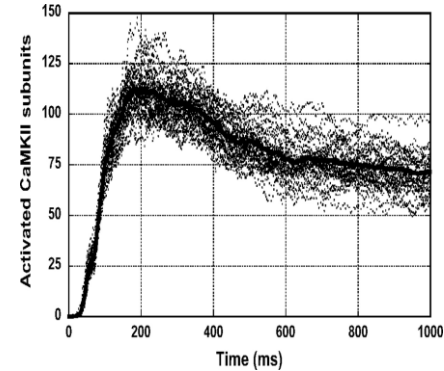


Filling the Conditional Proba. Tables

HSD model



50 000 simulations



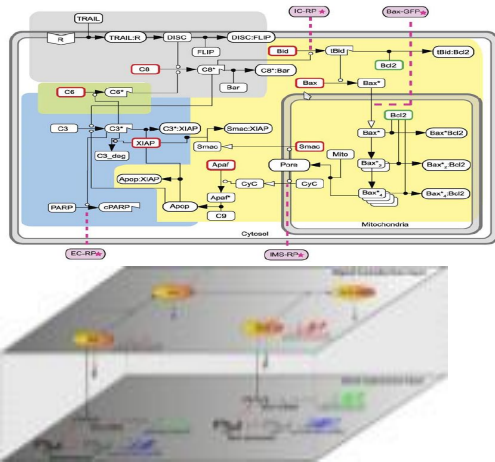
...

$$\begin{aligned} \text{Proba}(X^{t+1}=\text{high} \mid Y^t=\text{low}, Z^t=\text{high}) &= \frac{\text{nbr_simu}(X^{t+1}=\text{high} \mid Y^t=\text{low}, Z^t=\text{high})}{\text{nbr_simu}(Y^t=\text{low}, Z^t=\text{high})} \\ &= \frac{100}{1000} = 0.1 \end{aligned}$$

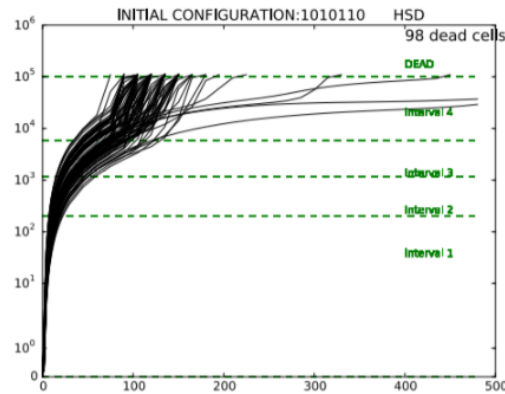
...

How good is the Abstraction?

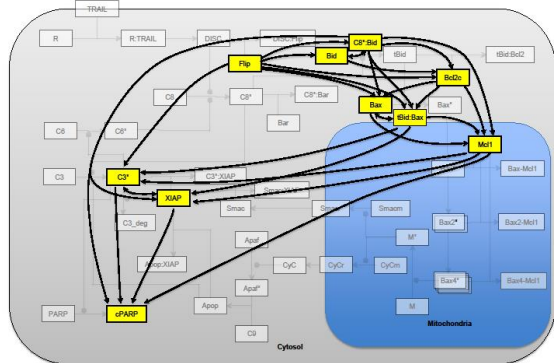
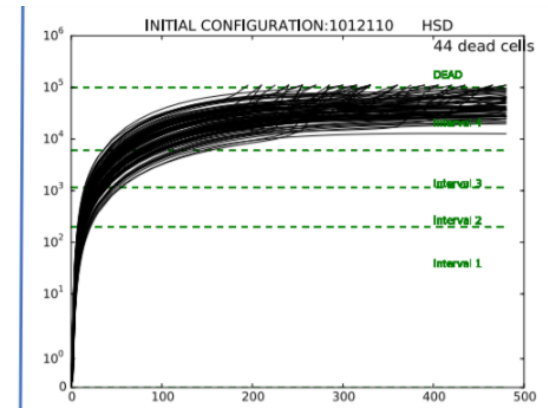
HSD model



100 runs: 98% dead

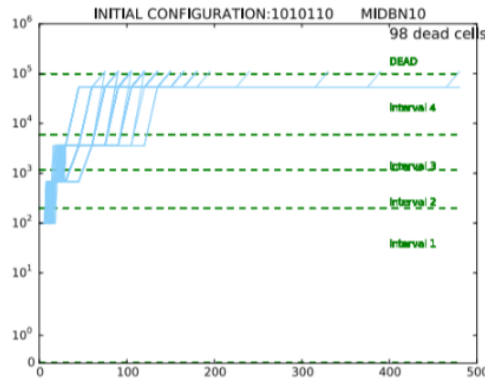


100 runs: 44% dead

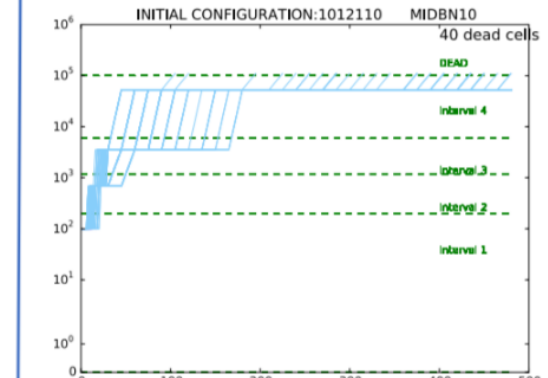


DBN abstraction

100 runs: 98% dead



100 runs: 40% dead

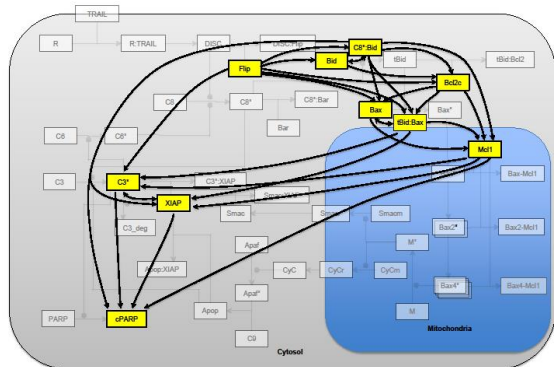
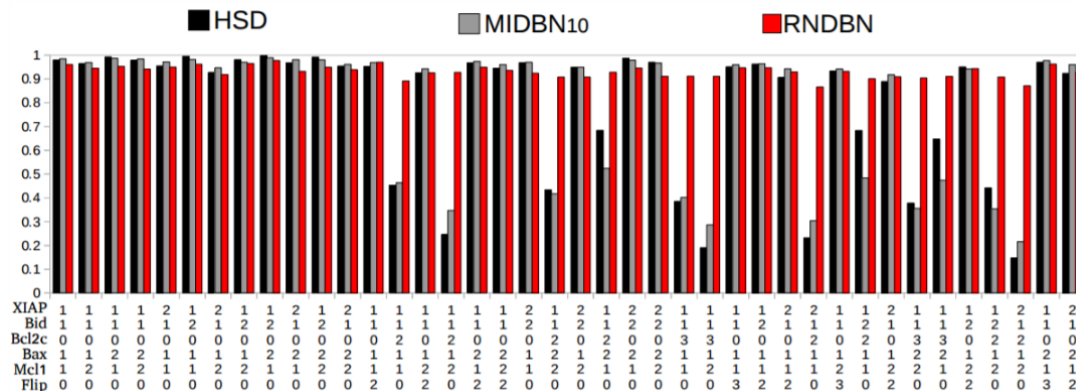
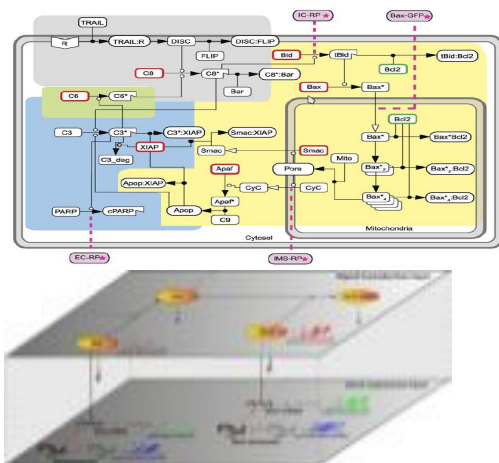


Less antiapop. molecules

More antiapop. molecules

How good is the Abstraction?

HSD model



DBN abstraction

Model	cell death (HSD: 69.9%)	discerning power (HSD: 100%)	Time / 1000 simulations (HSD: 56s)
<i>MIDBN</i> ₇	70.43%	96.14%	2.13s (26.3X)
<i>MIDBN</i> ₈	69.57%	96.31%	2.64s (21.21X)
<i>MIDBN</i> ₉	69.33%	96.37%	2.98s (18.8X)
<i>MIDBN</i> ₁₀	69.03%	96.84%	3.30s (17X)
<i>MIDBN</i> ₅₈	66.85%	94.12%	73.05s
<i>RNDBN</i>	92.29%	85.53%	299s

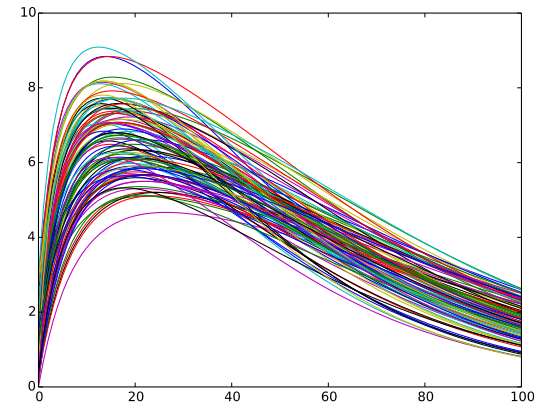
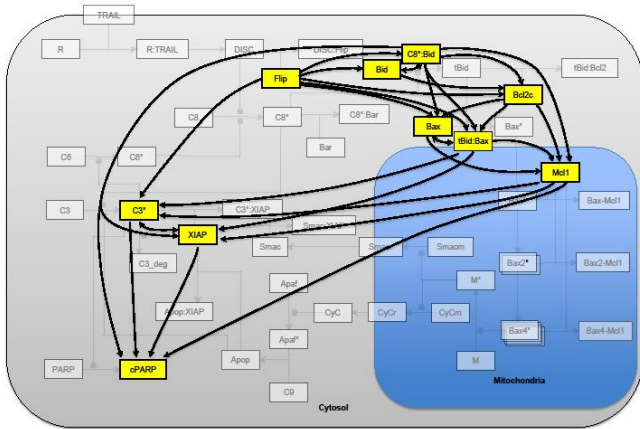
Time efficient:

1 simu CMC **20x faster** than 1 simu HSD

1 simulation of CMC = many simulations of HSD

Simulation vs Inference

To obtain the probability distribution produced by the DBN

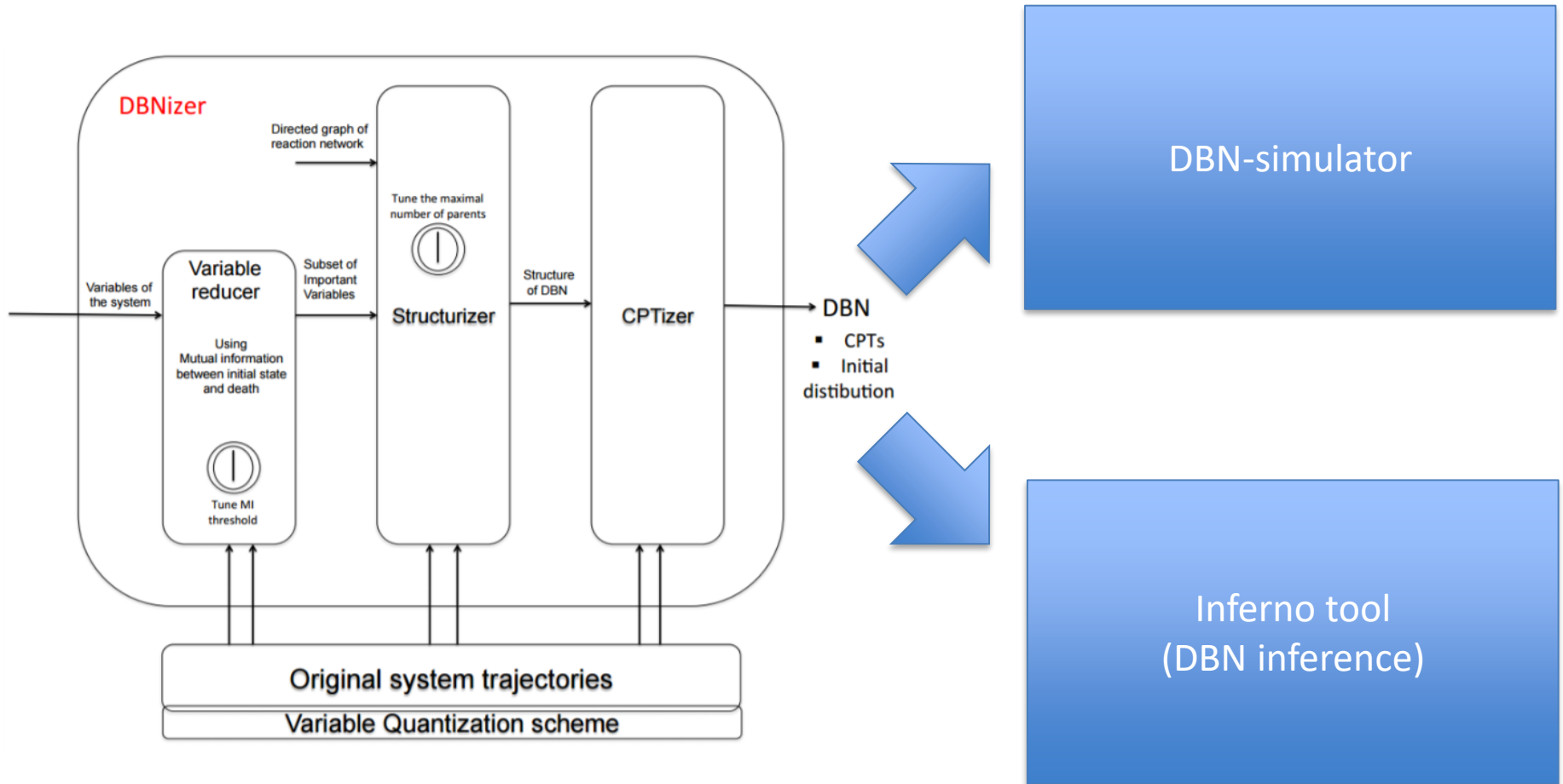


Lots of **simulations**
[HSB'16]

$$P^t(\mathbf{X} = \mathbf{x}) = \sum_{\mathbf{u} \in V^X} P^{t-1}(\mathbf{X} = \mathbf{u}) \prod_{i=1}^n CPT_{t,i}(\mathbf{x}_i | \mathbf{u}_i)$$

Inference (1 computation). ~10sec.
[submitted]

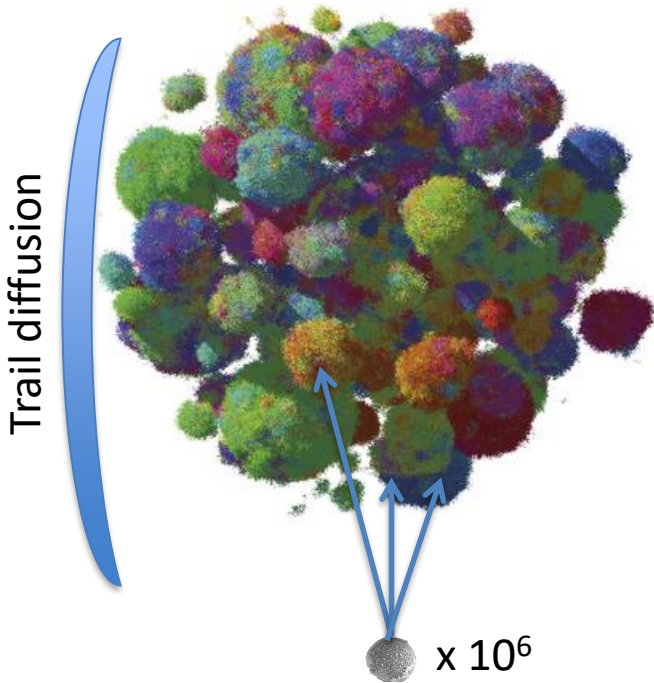
DBNizer: fully automatic abstraction tool



available [freely](https://sucree.bitbucket.io/DBNizer/) at
<https://sucree.bitbucket.io/DBNizer/>

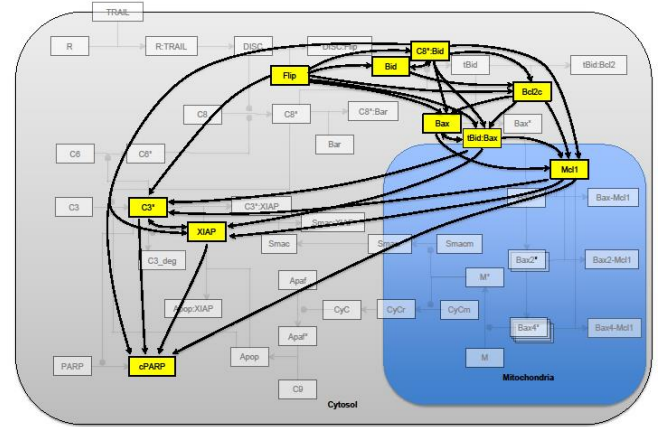
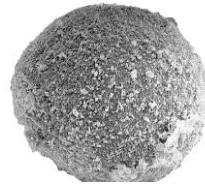
The big picture

Tumor

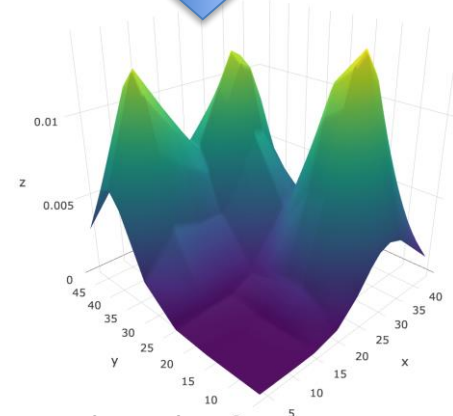


Cell death,
Natural Tumor growth...

for each cell, abstraction DBN:



inference



Describe
The
State of the
Entire
Population
With probabilistic distribution