Biological Regulatory Networks (BRN) modeled as an institution: property preservation along embedding

Pascale LE GALL

with Mbarka MABROUKI and Marc AIGUIER

MAS - École Centrale Paris
Programme Epigénomique - Université d'Evry - Génopole

January, 2008
IFIP Working Group 1.3

Biological regulatory networks (BRN) :

genes or derived products as proteins interact with others such that cell behaviours are regulated

Goal :
Understanding the functioning of biological regulatory networks in order to predict some knowledge about behaviours

Motivation:

- Focusing on an isolated part of the global system, perceived as having a particular biological function
- and studying in which cases properties of biological regulatory networks are preserved.

Qualitative modeling frameworks for BRN

multivalued discrete approach developed by R. Thomas:

- constituent concentrations are abstracted by integers to denote thresholds from which they can act on other constituents
- biological systems are described by an interaction graph defining the static part
- a huge but finite set of state transition graphs defining all the possible dynamics, or models.
- discretization preserves qualitative biological observations/experiments expressed as temporal properties

Decomposing BRN as sub-BRNs expressing biological functions

- As expected, model-checking technics are efficient to study small BRNs, but cannot cope with large BRNs.
- In practice, biologists study small BRNs, of particular importance w.r.t a biological function.
- Interactions between BRNs are studied afterwards, whether properties of sub-BRNs are preserved or not.

The general ideal

Following an institution-like approach, we provide a logic to characterize dynamic of BRNs

- **Syntax**

  Signatures are interaction graphs
  Sentences are CTL-X formulas (Computational Tree Logic without the neXt temporal operator)

- **Semantics**

  Models are a particular sub-class of state transition systems
  Satisfaction relation is the usual one for CTL

and we establish the satisfaction condition with some restricted conditions on the embedding of a BRN within an other one.

Definition

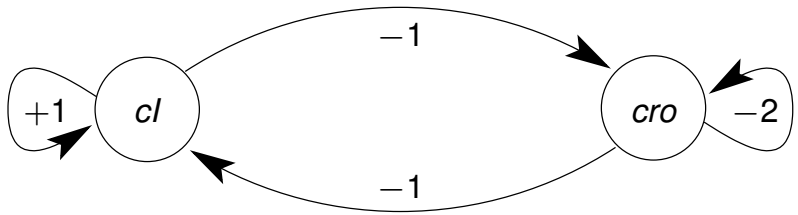
An institution $\mathcal{I} = (\mathit{Sig}, \mathit{Sen}, \mathit{Mod}, \models)$ consists of

- a category Sig of signatures,
- a functor $\mathit{Sen} : \mathit{Sig} \to \mathit{Set}$ giving for each signature Σ a set, element of sentences,
- a contravariant functor $\mathit{Mod} : \mathit{Sig}^{op} \to \mathit{Cat}$ giving for each signature Σ a category of Σ-models
- a $|\mathit{Sig}|$-indexed family of satisfaction relations $\models_\Sigma \subseteq |\mathit{Mod}(\Sigma)| \times \mathit{Sen}(\Sigma)$

such that the satisfaction condition holds:
$\forall \sigma : \Sigma \to \Sigma'$, $\forall \mathcal{M}' \in |\mathit{Mod}(\Sigma')|$, $\forall \varphi \in \mathit{Sen}(\Sigma)$,

$$\mathcal{M}' \models_{\Sigma'} \mathit{Sen}(\sigma)(\varphi) \Leftrightarrow \mathit{Mod}(\sigma)(\mathcal{M}') \models_\Sigma \varphi$$

Biological regulatory graph : Example



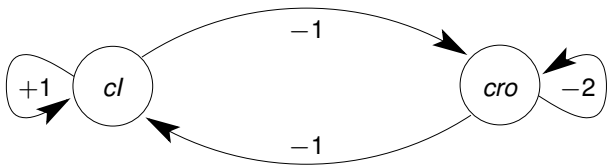
- Vertex = genes
- Edges = interactions (activation or inhibition)
- Labels = $\begin{cases} Sn : Sign\ ((+)\ for\ activation,\ (-)\ for\ inhibition) \\ Th : Threshold\ (interaction\ level) \end{cases}$

A BRN-signature is a labeled directed graph $G =< V, F, Sn, Th >$ where :

1. V is a finite set of *variables*.
2. $F \subseteq V \times V$ denotes the set of edges.
   For any $i \in V$, G_i^+, resp. G_i^-, denotes the set of successors, resp. predecessors, of i in $< V, F >$.
3. Sn is a mapping from F to $\{+, -\}$.
4. Th is a mapping from F to $\mathbb{N}^*$ such that:

$$\forall i \in V, \forall j \in G_i^+, Th(i,j) = c \wedge c \neq 1 \Rightarrow \exists k \in G_i^+ : Th(i,k) = c-1$$

Dynamic behaviors = traces associating at each time to each gene its concentration level.



We know that $x_{cI} \in \{0, 1\}$ *and* $x_{cro} \in \{0, 1, 2\}$

The state space S_G of $G =< V, F, Sn, Th >$ is the set of mappings $s : V \rightarrow \mathbb{N}$ s.t. $\forall i \in V, s(i) \in \{0, \ldots, b_i\}$.

with $b_i = |\{s \in \mathbb{N}^* \mid \exists j \in G_i^+, Th(i, j) = s\}|$

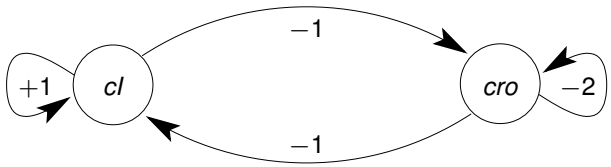
The concentration level of $i \in V$ evolves over time depending on concentration levels of its resources, i.e. i's predecessors having reached a concentration level to affect i's one:

$$R_{G,i}(s) = \begin{cases} \{j \in G_i^- | (Sn(j,i) = + \text{ and } s(j) \geq Th(j,i))\} \\ \cup \\ \{j \in G_i^- | (Sn(j,i) = - \text{ and } s(j) < Th(j,i))\} \end{cases}$$

Resources

cI	cro	$R_{G,cI}$	$R_{G,cro}$
0	0	$\{cro\}$	$\{cI, cro\}$
0	1	$\emptyset$	$\{cI, cro\}$
0	2	$\emptyset$	$\{cI\}$
1	0	$\{cI, cro\}$	$\{cro\}$
1	1	$\{cI\}$	$\{cro\}$
1	2	$\{cI\}$	$\emptyset$

Graph



Hence, a resource is the presence of an activator or the absence of an inhibitor.

- No indication in the signature to decide the concentration level that i can reach.
- This degree of freedom gives rise to a class of possible G-models, so-called dynamics of G.

Let $\kappa = \{(i, w) \mid i \in V \wedge w \subseteq G_i^- \}$ be the set of all subsets of predecessors in G for $i \in V$.

A G-model is a mapping $p : \kappa \to \mathbb{N}$ s. t.:
$\forall i \in V, p((i, \emptyset)) = 0 \wedge (\forall (i, w \neq \emptyset) \in \kappa,\ p((i, w)) \in \{0, \ldots, b_i\})$.

Models: asynchoneous transition system (AST)

Since in the nature, several variables cannot cross a threshold simultaneously, we make evolve one variable i by one unit in the direction its concentration level specified by p.

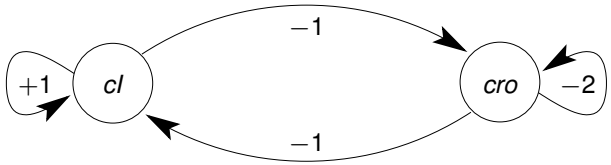
The asynchronous transition system generated from p is a directed graph $GTA((G,p)) =< S_G, T >$ s.t.:

- $\forall s \in S_G, (s,s) \in T \Leftrightarrow \forall i \in V, s(i) = p((i, R_{G,i}(s)))$
- $\forall s \neq s' \in S_G$, $(s,s') \in T$ iff:
  - there exists $i \in V$, s.t.

$$s'(i) = \begin{cases} s(i) + 1 \text{ and } s(i) < p((i, R_{G,i}(s))) \\ s(i) - 1 \text{ and } s(i) > p((i, R_{G,i}(s))) \end{cases}$$

  - and $s'(j) = s(j)$ for every $j \in V \setminus \{i\}$.

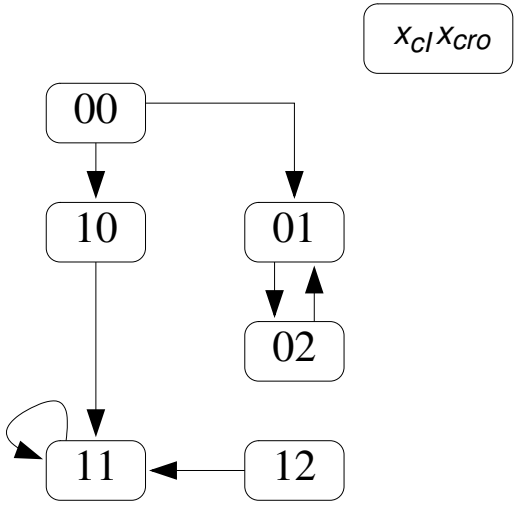
Graph



Resources and a model

cI	cro	$R_{G,cI}$	p_{cI}	$R_{G,cro}$	p_{cro}
0	0	$\{cro\}$	1	$\{cI, cro\}$	2
0	1	$\emptyset$	0	$\{cI, cro\}$	2
0	2	$\emptyset$	0	$\{cI\}$	1
1	0	$\{cI, cro\}$	1	$\{cro\}$	1
1	1	$\{cI\}$	1	$\{cro\}$	1
1	2	$\{cI\}$	1	$\emptyset$	0

Asynchroneous model



Sentences $Sen(G)$

Sentences are CTL-X formulas whose atomic formulas are comparisons between a concentration level of a variable with some threshold values.

- Atomic formulas are of the form $(i \sim s)$ where $i \in V$, $s \in \{0, \ldots, b_i\}$ and $\sim \in \{=, <, >\}$.
- Formulas are of the form:

  $$ATOM \mid For \Rightarrow For \mid For \wedge For \mid For \vee For \mid \neg For$$
  $$AG\ For \mid EG\ For \mid AF\ For \mid EF\ For \mid A[For\ U\ For] \mid E[For\ U\ For]$$

  A (for All path), E (there exists one path), F(there exists one state in the path), G (for all states in the path), U (until).

$i \geq s$ (resp. $i \leq s$) will denote $i = s \vee i > s$ (resp. $i = s \vee i < s$).

Satisfaction relation

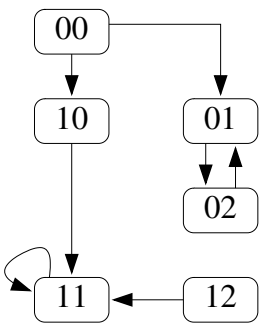
Satisfaction relation between models and sentences for BRN is derived form the usual one between transition systems and CTL-X formulas.

For a model p with $GTA((G,p))$ as associated AST:

$$p \models_G \varphi \Leftrightarrow GTA((G,p)) \models \varphi$$

with for $s \in S_G$,

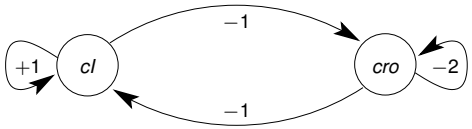
$$L(s) = \{i > l, i < l', i = l'' \mid \begin{array}{l} i \in V, l, l', l'' \in \{0, 1, \ldots, b_i\}, \\ s(i) > l, s(i) < l', s(i) = l'' \end{array} \}$$

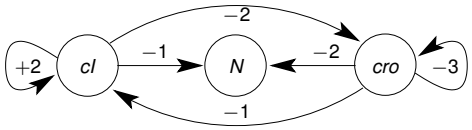


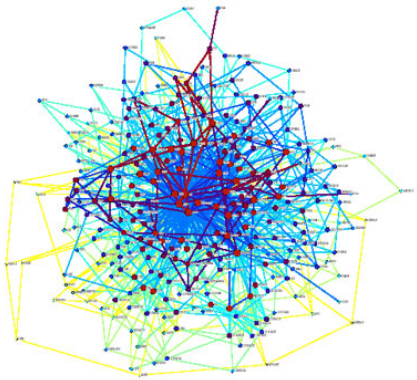
$$\models \qquad x_{cI} = 1 \wedge x_{cro} = 0 \Rightarrow EG(x_{cI} = 1)$$

Embedding : motivation and objectives

Biologists study small BRNs viewed as a biologic function.



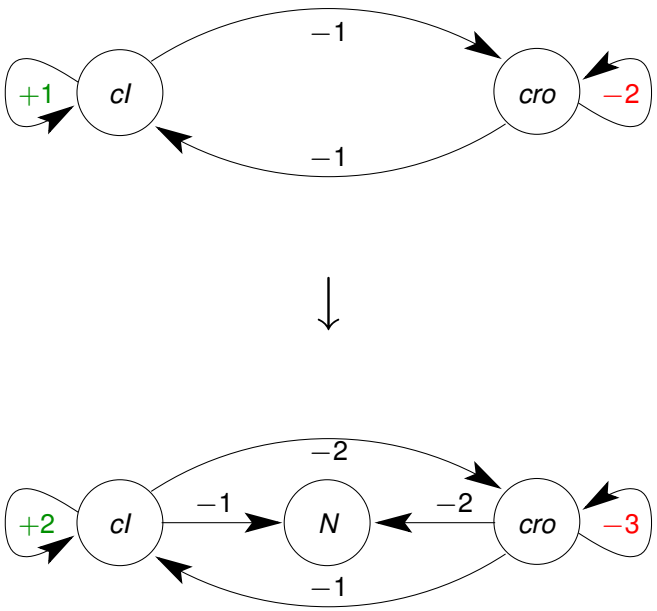




- Embedding = inclusion of a graph G in a graph G'
- Embedding preserves genes, interactions, and order between thresholds in relation to each gene.

Effect : Shifting the thresholds

Example



Signature embedding

$G =< V, F, Sn, Th >$ and $G' =< V', F', Sn', Th' >$ signatures.

An embedding $G \rightarrow G'$ is an injective mapping $\sigma : V \rightarrow V'$ s.t.:

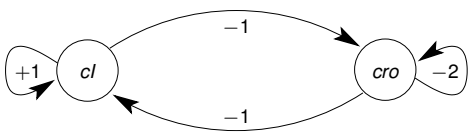
1. $\forall i, j \in V, (i, j) \in F \Leftrightarrow (\sigma(i), \sigma(j)) \in F'$
2. $\forall i, j \in V, (i, j) \in F, Sn(i, j) = Sn'(\sigma(i), \sigma(j))$
3. $\forall i \in V, \forall j, k \in G_i^+,$
   $Th(i, j) = Th(i, k) \Leftrightarrow Th'(\sigma(i), \sigma(j)) = Th'(\sigma(i), \sigma(k))$
4. $\forall i \in V, \forall j, k \in G_i^+,$
   $Th(i, j) < Th(i, k) \Leftrightarrow Th'(\sigma(i), \sigma(j)) < Th'(\sigma(i), \sigma(k))$
5. $\forall j \in V, \ \forall k' \in V',$
   $(k', \sigma(j)) \in F' \Rightarrow \exists i \in V, (i, j) \in F \wedge \sigma(i) = k'$

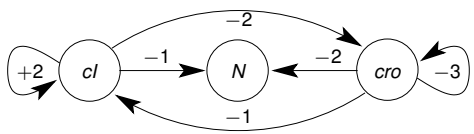
Translation of formulas along signature embedding $\sigma : g \to G$

Idea : Translating a threshold into an interval of values.
Notation: $\sigma_{g_1}(0) = 0$ and
for $l \neq 0$, $\sigma_{g_1}(l) = Th_G(g_1, g_2)$ for g_2 s.t. $Th_g(g_1, g_2) = l$,
Example $\sigma_{cl}(1) = 2$

- For all $(i = l)$, $Sen(\sigma)(x_g = l) = x_g \geq \sigma_g(l) \wedge x_g < \sigma_g(l+1)$
- For all $(i > l)$, $Sen(\sigma)(x_g > l) = x_g \geq \sigma_g(l+1)$
- For all $(i < l)$, $Sen(\sigma)(x_g < l) = x_g < \sigma_g(l+1)$
- Other symbols are handled as usual

Example





$$Sen(\sigma)(AG(x_{cro} = 1)) = AG(1 \leq x_{cro} < 3)$$

Reduced model along a signature embedding $\sigma : G \to G'$

The reduction of a model along a signature embedding is defined up to some restrictions on thresholds.

Given a signature embedding $\sigma : G \to G'$ and a G'-model p', the reduced G-model from p' denoted $p'_{|\sigma}$ is defined as follows: $\forall (i, w) \in \kappa$,

$$p'_{|\sigma}((i,w)) = \begin{cases} Th(i,j) & \text{if } \exists j \in V, \\ & Th'(\sigma(i),\sigma(j)) = max_{(i,k)\in F}\{Th'(\sigma(i),\sigma(k)) \mid \\ & \qquad Th'(\sigma(i),\sigma(k)) \leq p'((\sigma(i),\sigma(w)))\} \\ 0 & otherwise \end{cases}$$

CTL-X formulas are preserved through embedding of biological regulatory networks.

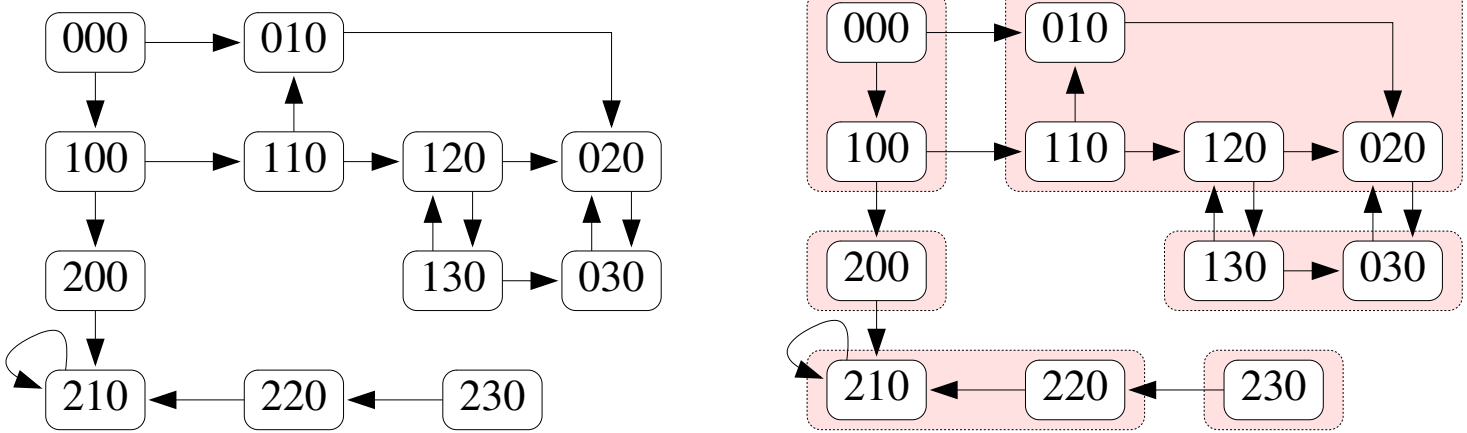
Theorem

For $\sigma : G \to G'$ embedding, p' G'-model and $\varphi \in Sen(G)$,

$$p' \models \sigma(\varphi) \quad \Longleftrightarrow \quad p'_{|_\sigma} \models \varphi$$

Let us consider $\sigma : G \to G'$,
p' a G'-model,
$(S_{G'}, T') = GTA((G', p'))$ its associated ATS
$\varphi \in Sen(G)$.

Let us define a partition of the state space of $GTA((G', p'))$, taking into account shifting of thresholds through the embedding σ.



Let us define the mapping $B : S_G \rightarrow 2^{S_{G'}}$ as follows:
$\forall s \in S_G$, $B(s) \subseteq S_{G'}$ verifying: $s' \in B(s)$ if for every i in V:

- if $s(i) = 0$, then $s'(\sigma(i)) \geq 0$ and $s'(\sigma(i)) < min_{(i,k)\in F}\{Th'(\sigma(i), \sigma(k)) \mid Th'(\sigma(i), \sigma(k)) > 0\}$
- otherwise, let j be any variable in G_i^+ s.t. $s(i) = Th(i, j)$ then $s'(\sigma(i)) \geq Th'(\sigma(i), \sigma(j))$ and $s'(\sigma(i)) < min_{(i,k)\in F}\{Th'(\sigma(i), \sigma(k)) \mid Th'(\sigma(i), \sigma(k)) > Th'(\sigma(i), \sigma(j))\}$

Proposition 1

The mapping B makes a partition of $S_{G'}$, i.e.

1. $\forall s, s' \in S,\ B(s) \cap B(s') = \emptyset$, and
2. $\bigcup_{s\in S_G} B_s = S_{G'}$.

A binary relation R is called a *divergence blind stuttering* (dbs) relation iff it is symmetric and

$$r\ R\ s \Longleftrightarrow \begin{cases} L(r) = L(s) \\ (r, r') \in T \Rightarrow & \exists s_0, s_1, \ldots, s_n \text{ finite path }, n \geq 0, (s_0 = s) \\ & \wedge (\forall i < n, r\ R\ s_i) \wedge r'\ R\ s_n \end{cases}$$

The largest dbs relation is an equivalence relation noted $\simeq_{dbs}$.

Proposition 2

Note $P = \{B(s) | s \in S_G\}$. Then, we have:
P is a dbs equivalence.

Preliminary

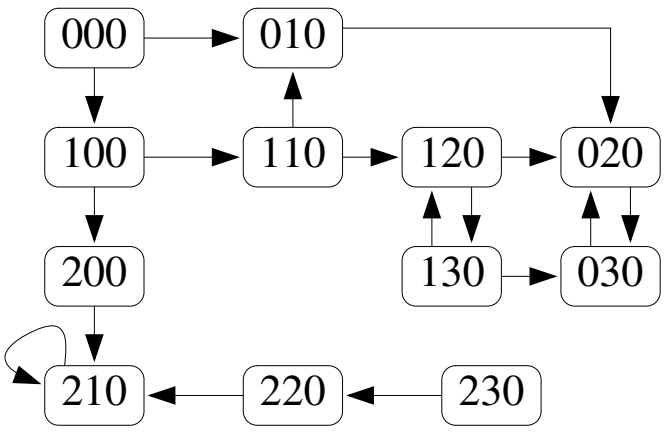
The quotient of a transition system (S, T) by $\simeq_{dbs}$ is denoted $(S, T)_{/\simeq_{dbs}}$.
The equivalence relation $\simeq_{dbs}$ preserves CTL-X formulas, i.e. (S, T) and $(S, T)_{/\simeq_{dbs}}$ satisfy the same formulas.

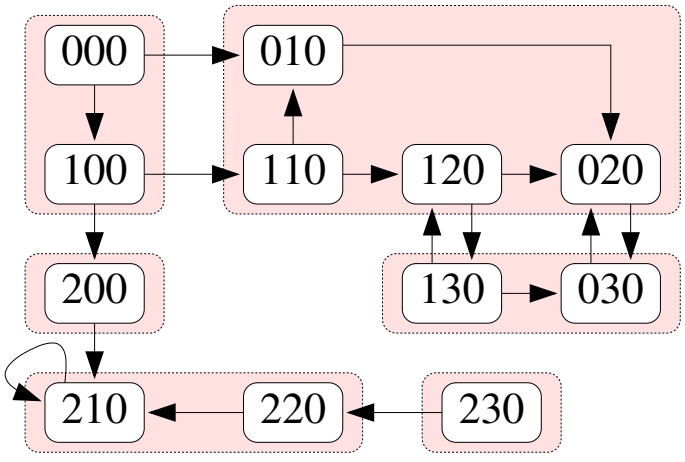
Proposition 3

$(S_{G'}, T')_{/\simeq_{dbs}}$ and $GTA(G, p'_{|\sigma})$ are isomorphic.

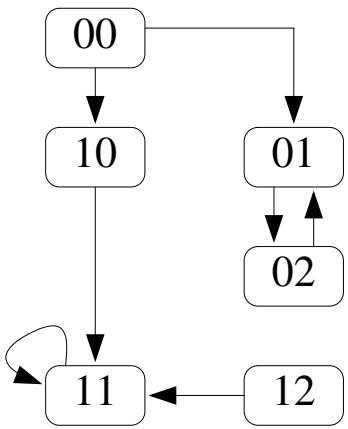
–

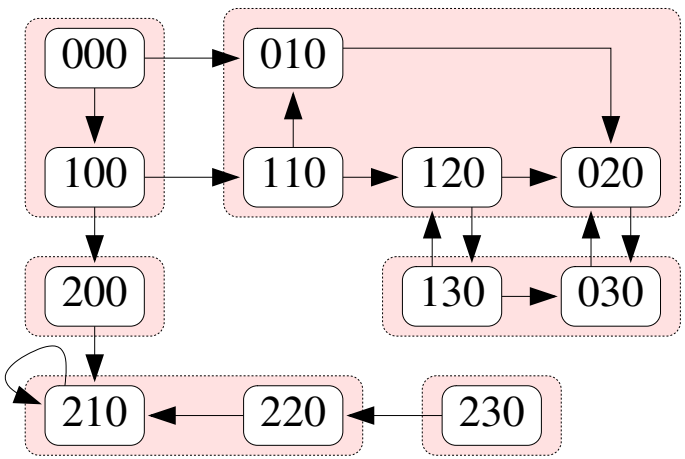
Sketch of the proof:illustration



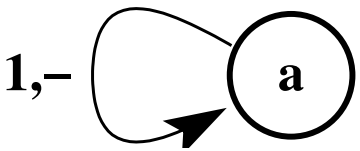


Sketch of the proof:illustration

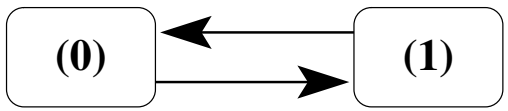


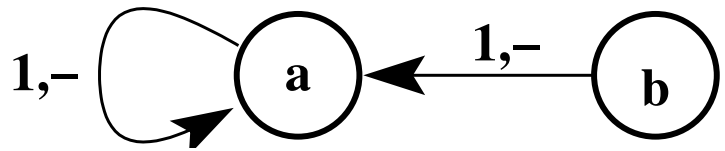


Counter-example: necessity of restrictive conditions on signature embeddings

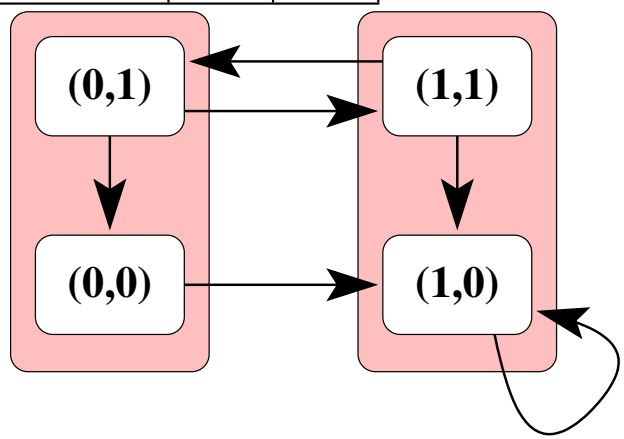


resource ω	p_a
$\emptyset$	0
$\{a\}$	1

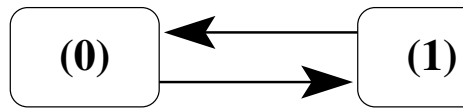




resource ω'	p'_a	p'_b
$\emptyset$	0	0
$\{a\}$	1	
$\{b\}$	1	
$\{a, b\}$	1	



resource ω'	$p'_{\mid_\sigma a}$
$\emptyset$	0
$\{a\}$	1



$AG(AF(a = 0))$ is satisfied by $p'_{\mid_\sigma}$ but not by p'.

Conclusion

- Result :
  - Preservation of properties along simple embedding of biological regulatory networks:
    no new entering edge when embedding a network within a larger one
  - discrete models for BRN modelled as an institution
- Current work : Investigation of some other loose conditions about property preservation for BRN
- Future work : BRN as an application domain to study complex systems.
  A system is considered as complex according to the fact that properties of sub-systems are not preserverd at the level of the global system.