

Behaviour Preservation of a Biological Regulatory Network when Embedded into a Larger Network

Gilles Bernot *

Laboratoire I3S, UMR 6070 UNSA-CNRS

Algorithmes-Euclide-B

2000, route des Lucioles - B.P. 121

F-06903 Sophia Antipolis CEDEX

bernot@unice.fr

Fariza Tah

IBISC FRE 3190

Université d'Evry-CNRS, Genopole[®]

523 place des Terrasses

F-91000 Evry, France

fariza.tahi@ibisc.univ-evry.fr

Abstract. The main contribution of this work is a mathematical theorem which establishes a necessary and sufficient condition to preserve the behaviour of a genetic regulatory network when it is embedded into a larger network. We adopt the modelling approach of René Thomas, which provides a discrete representation of biological regulatory networks. This framework is entirely formalized using labelled graphs with semantics defined in terms of state graphs with transitions. Our theorem offers the possibility to automatically verify whether a subnetwork has autonomous behaviour. It will allow biologists to better identify relevant sets of genes which should be studied together.

Address for correspondence: bernot@unice.fr, fariza.tahi@ibisc.univ-evry.fr

*We are pleased to acknowledge Genopole[®] for constant support of our researches on regulatory networks, as well as the biologists who worked with us on biological examples, with a special thank to Janine Guespin. We also thank Adrien Richard, Jean-Paul Comet and Randy Thomas for proofreading this article.

1. Introduction

Biological regulatory networks are often mathematically modeled by means of graphs which represent the interactions between biological entities such as macromolecules or genes. Current advances in molecular biology provide us with fairly complete such maps of gene interactions for an increasing number of species. Following the success of the genomic and post-genomic area, this kind of model becomes essential for understanding and predicting cell behaviour in a bottom-up approach.

The regulatory network of a cell, as a whole, is composed of some tens of thousands of genes (e.g. 6000 genes in the yeast genome and between 30.000 and 40.000 genes in mammalian genomes). Unfortunately, rigorous models revealing the precise causal functioning of a regulatory network are currently limited to some tens of genes. The main reason for this limitation is that such models involve a large number of parameters which must be indirectly deduced from the cell behaviour, since they are not directly measurable, leading to many experiments in order to indirectly find the parameter values. Such very precise dynamic models are consequently focused on few genes selected because they cooperate to realize a precise biological function.

Therefore, to predict biological behaviour (or phenotype) from the static knowledge of a regulatory graph, there are two main scientific trends:

- *studying regulatory networks in the large*: probabilistic or statistical approaches can be used to extract general predictions according to some characteristics of the interaction network [11, 16, 21]. Those characteristics can be simple local measurements such as the degree of each node or global measurements such as the graph diameter. Qualitative or logical properties of such large networks are less studied except for some works allowing to focus on some relevant complex properties by simplifying some simple causal cascades [7].
- *studying regulatory networks in detail*: “concentration levels” are attached to each entity belonging to the graph and the goal is to predict their possible evolution [43, 10, 8, 12]. Additional parameters are then required in order to precisely deduce the possible trajectories in the space of possible states. Those parameters define the nature of the considered interactions as well as their relative strengths. This scientific trend is mostly used to analyze in detail the functioning of well identified biological functions inside the cell [40, 23, 28, 22, 6, 44, 18, 32, 33, 20, 24, 1]. Lastly, Petri nets are well studied in computer sciences since more than thirty years and there are also some hopes to take benefits of this corpus: pioneering works in this area are for example proposed in [5, 14].

This article belongs resolutely to the second trend. Nevertheless, we believe that it is possible to increase the number of considered genes *via* a decomposition of the graph into some kind of “modules.” The modules should be studied independently and we should be able to deduce the global behaviour from the knowledge of the individual behaviour of each module. The notion of module remains to be defined in general and this article constitutes a first step towards this general definition.

We adopt here the formalism of René Thomas [38, 37, 41, 39, 36, 3]. It has the advantage to be a logical modelling approach which can benefit from very powerful automated tools in computer science such as *model checking* [9, 4]. This discrete modelling has been proved compatible with differential modelling (ODE) [29, 30] and has won a following. Several theoretical results, which are deeply relevant for biology, have been established over about thirty years (in particular, feedback circuits in the regulatory graphs have been extensively studied [38, 30, 41, 35, 15, 29, 13, 39, 33, 12, 26, 2, 25, 27, 19, 31]).

In continuation of this school, we establish a non trivial theorem about *embeddings* of biological regulatory networks. Given a sub-network embedded into a larger one, we give a necessary and sufficient condition to ensure that its “isolated behaviour” is not affected by the embedding. Therefore, our work constitutes a step towards a modular approach, trying to split regulatory networks into smaller parts which can be treated independently [34].

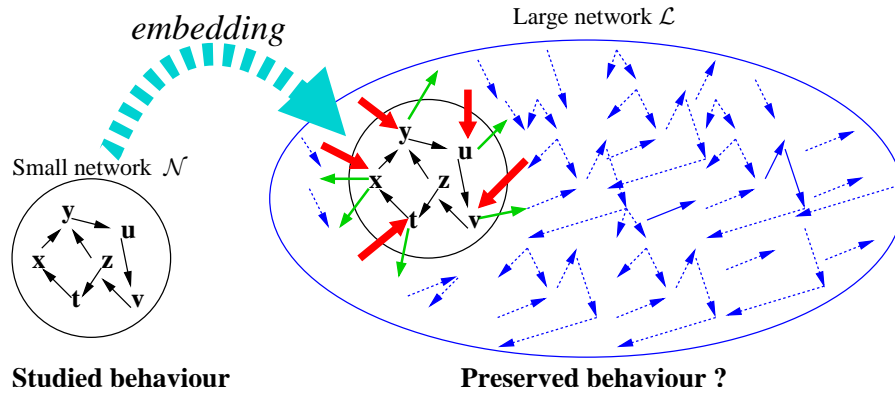


Figure 1. Embedding of a biological regulatory network

More precisely, as illustrated in Figure 1, an embedded network \mathcal{N} communicates with the rest of a larger network \mathcal{L} through two kinds of arrows: *input* arrows which go from a gene outside of \mathcal{N} to a gene of \mathcal{N} (bold arrows) and *output* arrows which go from a gene of \mathcal{N} to the outside (grey arrows). We demonstrate that only the input arrows can influence the behaviour of \mathcal{N} and we give a necessary and sufficient condition to preserve the behaviour of \mathcal{N} . This condition is expressed on the parameters associated to each input arrow. The advantage of our result is that it establishes the equivalence between:

- a global behavioural property of the network
- and a limited collection of algebraic constraints on local parameters, which is easier to verify mechanically.

Our result offers an effective verification of the behaviour preservation without checking all the possible states of the network and without enumerating all possible dynamics. This necessary and sufficient condition only involves the static description of the network, which makes possible an algorithmic verification with reasonable CPU time.

2. Biological Regulatory Network (BRN): Static Description

The logical framework for Biological Regulatory Networks (BRN for short) of René Thomas [37] is based on an abstract view of the interactions between objects of a biological system, mainly genes or macromolecules. Models within this framework mainly describe the nature of the interactions (inhibition or activation) and their strength without explicitly considering the detailed underlying biochemical reactions. They allow us to predict the dynamics of a BRN.

Graph theory is very convenient to support this kind of description, because graphs capture the “static” knowledge about the interactions. In the next section, we will see that graphs are also a useful tool to represent possible dynamics of the biological system, deduced from the static description. In the article, we will introduce the useful concepts on graphs on a “call-by-need” basis, under the title “Graph terminology.”

Graph terminology: A *directed graph* G is a set V of *vertices* together with a set E of *edges*. An edge (u, v) goes from a vertex u of V to another vertex v of V .

A directed graph is said to be *labelled* when each edge carries an information, called the *label* of the edge. Labelled directed graphs can be drawn with a diagram as in Figure 2 where α , β and γ are the labels. It is also possible to put labels on vertices if necessary.

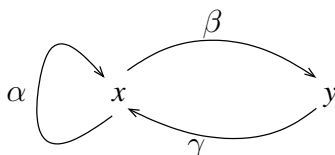


Figure 2. An example of labelled graph

A *predecessor* of a vertex v in a graph G is a vertex u of G such that there is an edge from u to v . The set of predecessors of v is denoted by $G^{-1}(v)$. In Figure 2, x and y are the predecessors of x : $G^{-1}(x) = \{x, y\}$ and $G^{-1}(y) = \{x\}$.

In the terminology of biological regulatory networks it is convenient to call “variables” the considered biological objects. One variable can cover any useful concept such as gene, RNA, protein, as well as a rough abstraction of those three facets of a gene, or a cluster of genes and so on. Such an abstraction is considered adequate if it preserves the ability to associate an abstract “concentration level” or “expression level” to a variable at any time.

If we represent variables as vertices of a directed graph then an edge (u, v) represents an action of u on v : the expression level of v with respect to the expression level of u , after a sufficient delay for u to act on v , is most of the time a sigmoid. let us moreover assume that u also acts on w as in Figure 3a, then three intervals are relevant for the expression level of u (Figure 3b): in the interval called “0” u neither acts on v nor on w , in the interval called “1” it acts on v only and finally in the interval called “2” it acts on both variables.

In Figure 3b, the sigmoid of v being increasing, we say that u is an *activator* of v and, the sigmoid of w being decreasing, we say that u is an *inhibitor* of w . Finally Figure 3c summarizes all these types of information: u begins to activate v at the interval numbered 1 and u begins to inhibit w at the interval numbered 2. This is formally expressed on the graph as: the threshold of (u, v) is “1” with the sign “+” and the threshold of (u, w) is “2” with the sign “−”.

This idea to make a partition of the set of possible expression levels into several consecutive intervals allows us to consider a *finite* set of *discrete* expression levels. Considered expression levels for a variable u will therefore be integer values 0, 1, 2, ... up to a given boundary b_u . This constitutes a valuable discrete partitioning of a continuous global interval.

Predicting the dynamics of the expression level of the variables is of course a major issue. In the framework of René Thomas, a variable v tends towards an expression level which depends only on the

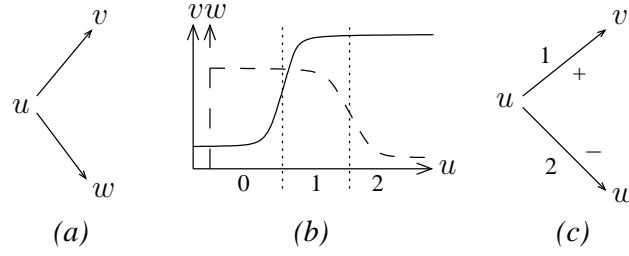


Figure 3. Interactions of variables and expression levels

expression level of its activators and inhibitors. More precisely, it only depends on the set ω of its “resources.” In the following definition, the expression level towards which v tends is denoted by $K_{v,\omega}$.

Definition (Biological Regulatory Networks): A biological regulatory network (BRN for short) is a triple $\mathcal{N} = (V, E, K)$ where:

- $N = (V, E)$ is a labelled directed graph.
- Each vertex v of V is called a variable of the BRN, and is provided with a strictly positive integer $b_v \in \mathbb{N}^*$, called the boundary of v .
- Each edge (u, v) of E is labelled by a couple (t, ε) where t , called threshold, is an integer between 1 and b_u and ε , called sign, belongs to $\{+, -\}$. When $\varepsilon = +$, u is called an inducer of v . When $\varepsilon = -$, u is called an inhibitor of v .
- $K = \{ K_{v,\omega} \mid v \in V \text{ and } \omega \subset N^{-1}(v) \}$ is a family of integers such that $0 \leq K_{v,\omega} \leq b_v$ for any variable v and for any subset ω of the predecessors of v in N . The family K is called the set of parameters of \mathcal{N} .

In the literature [29, 30], it is often additionally required that if $\omega_1 \subset \omega_2$ then $K_{v,\omega_1} \leq K_{v,\omega_2}$. This restriction means that the more resources a variable owns, the more its expression level tends to grow. In this article, all our results remain valid with or without this restriction.

In the following, when several BRNs are considered, we note $V^{\mathcal{N}}$ the set of variables of a BRN \mathcal{N} , $V^{\mathcal{P}}$ the set of variables of a BRN \mathcal{P} , etc. and similar notations apply for E and K . The exponent is omitted when the considered BRN is obvious from the context.

3. Dynamics of BRNs

It is generally impossible for a human to predict the behaviour of a biological system by looking at its static description alone. One of the main advantages of the René Thomas’ approach is to offer a qualitative representation of the system dynamics which is predictable using a computer. This section gives the corresponding formal definitions. See [4] for a more detailed presentation.

3.1. States and Resources of a Variable

According to our BRN framework, the state of a system is characterized by the expression level of each variable. Thus, a state can be defined as a map which associates with each variable its current expression level.

Definition (States of a BRN): Given a BRN \mathcal{N} , a state of \mathcal{N} is a mapping $\mu : V \rightarrow \mathbb{N}$ such that for any variable $v \in V$, we have $0 \leq \mu(v) \leq b_v$. The integer $\mu(v)$ is then called the expression level of v .

So far, at a given time, each variable v has a given expression level. The evolution of this expression level is driven by the current state of the inducers and inhibitors of v . Figure 4 shows that:

- an inducer of v is an actual resource of v only when it passes its threshold
- an inhibitor plays a symmetric role; it is a resource only when it does not pass its threshold.

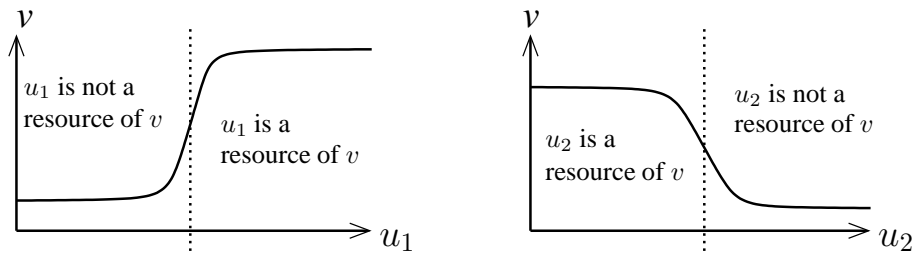


Figure 4. Induction (resp. inhibition) of v by u_1 (resp. u_2)

Definition (Resources of a Variable): Given a state μ and a variable v of a BRN \mathcal{N} , the set of resources of v is the set $\omega_v(\mu)$ containing all the variables u of \mathcal{N} such that:

- u is a predecessor of v in the underlying directed graph of \mathcal{N}
- the edge (u, v) is labelled by (t, ε) and
 - i) if ε is the sign “+” then $\mu(u) \geq t$
 - ii) if ε is the sign “-” then $\mu(u) < t$

The set of variables $\omega_v(\mu)$ is consequently the subset of $N^{-1}(v)$ containing both the inducers of v whose expression level has reached the threshold and the inhibitors of v whose expression level has not reached the threshold.

3.2. State Graph

The dynamics of a BRN is defined by a graph that defines which states can be reached from any given state.

If a variable v has an expression level $\mu(v)$ at a given time, then this expression level is attracted towards $K_{v, \omega_v(\mu)}$. Let us assume for instance that $K_{v, \omega_v(\mu)} = \mu(v) + n$ with $n > 1$. The variations of

concentration levels being continuous *in vivo*, the expression level $\mu(v)$ cannot directly jump to $K_{v, \omega_v(\mu)}$. It takes the neighbor value $\mu(v) + 1$, as defined below:

Definition (κ functions): Given a state μ and a variable v of a BRN \mathcal{N} , $\kappa_v(\mu)$ is defined as follows:

- if $\mu(v) < K_{v, \omega_v(\mu)}$ then $\kappa_v(\mu) = \mu(v) + 1$
- if $\mu(v) = K_{v, \omega_v(\mu)}$ then $\kappa_v(\mu) = \mu(v)$
- if $\mu(v) > K_{v, \omega_v(\mu)}$ then $\kappa_v(\mu) = \mu(v) - 1$

The function κ represents a “first step” evolution of the expression level of v from its current expression level $\mu(v)$ to its “target level”¹ $K_{v, \omega_v(\mu)}$.

Kinetics *in vivo* make highly improbable that two variables go through their threshold exactly at the same time. Therefore, variables vary asynchronously, which leads to a state graph where only one variable evolves at a time.

Definition (“Asynchronous” State Graph): The state graph of a BRN \mathcal{N} is the directed graph S whose vertices are all the possible states of \mathcal{N} and such that there is an edge from a state μ to a state μ' iff there exists a variable v verifying:

- $\mu'(v) = \kappa_v(\mu) \neq \mu(v)$
- for any variable $v' \neq v$ we have $\mu'(v') = \mu(v')$

An edge of the state graph from μ to μ' is usually denoted as $(\mu \rightarrow \mu')$ and is called a transition.

Figure 5 shows the underlying labelled graph of a BRN, the table of target levels according to an arbitrarily fixed set of parameters, and the corresponding state graph. Let us consider for instance the state “1 0” (i.e. $\mu(x) = 1$ and $\mu(y) = 0$). The arrow $x \rightarrow x$ being labelled by $(1, +)$ in Figure 5, x is a resource of itself and the arrow $y \rightarrow x$ being labelled by $(1, +)$, y is not a resource of x . Thus, x is attracted towards $K_{x, x}$. Similarly, the arrow $x \rightarrow y$ being labelled by $(2, +)$, x is not a resource of y and the arrow $y \rightarrow y$ being labelled by $(2, -)$, y is a resource of itself (because it does not pass its threshold). Thus, y is attracted towards $K_{y, y}$. So, the state “1 0” is attracted towards the target state “2 2” (as $K_{x, x} = 2$ and $K_{y, y} = 2$), which tends to increase the expression levels of x and y . Only one variable evolves at a time by one unit, thus two transitions start from the state “1 0” in the state graph: (“1 0” \rightarrow “2 0”) and (“1 0” \rightarrow “1 1”).

Graph terminology: A path of length n from a vertex x_0 to a vertex x_n in a directed graph is a sequence of edges of the form $(x_0, x_1) (x_1, x_2) (x_2, x_3) \cdots (x_{n-1}, x_n)$. If $x_0 = x_n$ then the path is called a *circuit*.

For example, $(x, y) (y, x)$ is a circuit in the graph of Figure 5.

Remark: The condition $\kappa_v(\mu) \neq \mu(v)$ for asynchronous state graphs in the previous definition removes elementary circuits of the state graph where a state points to itself. Indeed, many approaches in the literature would add an elementary circuit on the state “2 2” in the state graph of Figure 5. In the literature, such cases allow the user to identify stable (regular) steady states. According to our definition, these states are identified as the ones with no outgoing transition.

¹The target level $K_{v, \omega_v(\mu)}$ is also called *image* in the literature or sometimes ‘(local) attractor’ or ‘focal point’

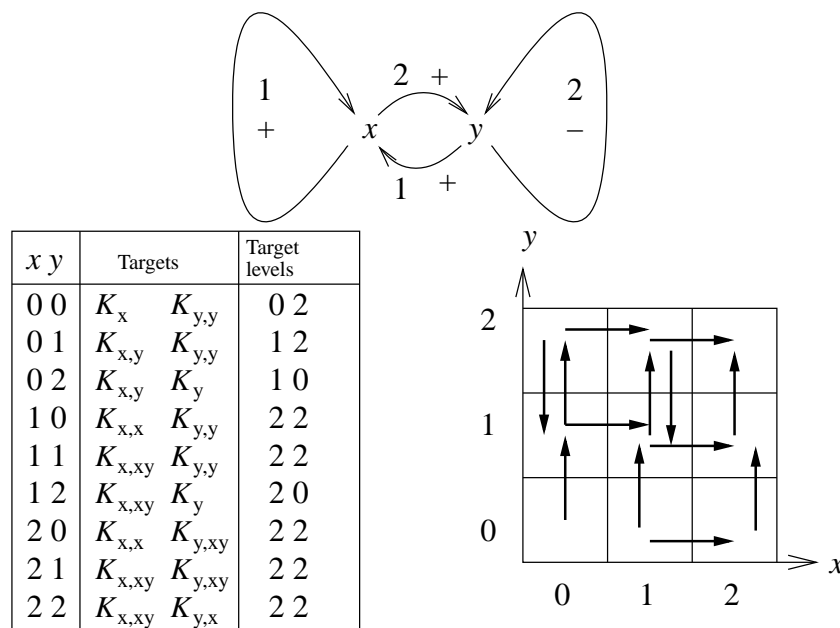


Figure 5. Example of a BRN and its state graph

According to the regulatory network approach, the *behaviour* of a cell is observed through variations of the quantities of diverse macromolecules produced in the cell. This is expressed in René Thomas's approach by variations of the expression levels belonging to the state graphs. Each path in the state graph represents a possible evolution of the biological system. Notice that two BRNs have the same behaviour if and only if they have the same set of paths, i.e. the same state graphs. Therefore, studying the behaviour of BRNs amounts to studying their state graphs.

4. Embeddings of BRNs

Studying the behaviour of the whole regulatory network of a cell would generate a combinatorial explosion which is entirely outside current know how. Biological knowledge begins to provide us with some fairly complete static description of gene interactions in some species (*Yeast*, *E.coli*, ...). However, there is a huge gap between the static knowledge of a regulatory graph and the knowledge of the corresponding regulatory network dynamics.

In practice, we are consequently restricted to the study of smaller regulatory networks, focusing on sets of genes participating in targeted biological functions. These networks are therefore sub-networks of the whole cell regulatory network and hopefully, the studied behaviour of each of them will not be affected by its relationships with the whole network. In this article, our goal is precisely to offer rigorous conditions to reach this hope.

4.1. Partial BRN

According to this aim, an obvious preliminary condition is to consider *all* interactions *inside* each considered sub-network. The definition of partial BRN below reflects this preliminary condition.

Graph terminology: Given a directed graph G , a *subgraph* of G is a graph G' whose set of vertices is included in the set of vertices of G and whose set of edges is included in the set of edges of G . If G' contains all the edges of G which connect vertices of G' (i.e. for all vertices u and v of G' , the edge (u, v) belongs to G' iff it belongs to G), then G' is called a *partial graph* of G .

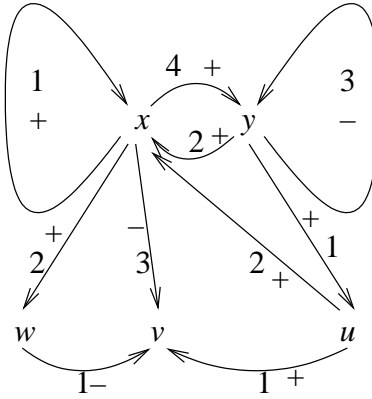
Definition (Partial BRN): A BRN \mathcal{P} is a partial BRN of a BRN \mathcal{L} iff:

- the underlying directed graph of \mathcal{P} is a partial graph of the underlying directed graph of \mathcal{L}
- any edge of \mathcal{P} has the same label in \mathcal{L}
- for any variable v of \mathcal{P} and for any subset ω of the predecessors of v in \mathcal{P} , we have $K_{v,\omega}^{\mathcal{P}} = K_{v,\omega}^{\mathcal{L}}$

Notice that in this definition, all the parameters of \mathcal{L} of the form $K_{v,\omega}^{\mathcal{L}}$ such that v or any variable of ω does not belong to \mathcal{P} are ignored. Mainly, the object of the remainder of this article is to study the effect of these ignored “external variables” on the behaviour of \mathcal{P} .

4.2. Level Folding

Let us consider the example of Figure 6. The graph of Figure 5 is not directly a partial graph of Figure 6



Parameters:

$$K_x=0, K_{x,x}=4, K_{x,y}=2, K_{x,u}=0, K_{x,xy}=4, K_{x,xu}=4, K_{x,yu}=4, K_{x,xyu}=3,$$

$$K_y=0, K_{y,x}=3, K_{y,y}=3, K_{y,xy}=3,$$

$$K_u=0, K_{u,y}=2,$$

$$K_v=0, K_{v,x}=0, K_{v,u}=0, K_{v,w}=0, K_{v,xu}=1, K_{v,xw}=1, K_{v,uw}=1, K_{v,xuw}=1,$$

$$K_w=0, K_{w,x}=1$$

Figure 6. A BRN which embeds the BRN of Figure 5

because the thresholds of edges (x, y) , (y, x) and (y, y) differ. Nevertheless, assuming that we remove

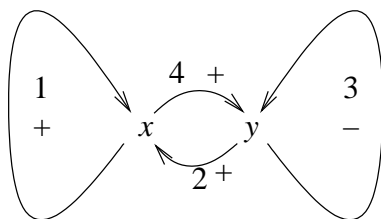
all the edges starting from x or y to u , v or w , the threshold of the edge (x, y) becomes the second threshold among the ones of edges starting from x . Similarly, the threshold of (y, x) becomes the first threshold starting from y and the threshold of (y, y) becomes the second one. Consequently, when u , v and w are removed from Figure 6, it is sensible to consider that we retrieve Figure 5. In other words, the graph of Figure 5 is not directly a partial graph of Figure 6 but is in fact *embedded* in the graph of Figure 6.

This example reveals the usefulness of a *folding function* in order to keep the thresholds consecutive in the graph.

Definition (Level Folding): Let \mathcal{N} be any BRN and let $u \in V$ be a variable of \mathcal{N} . The set of the threshold values of the outgoing edges of u can be sorted in strictly increasing order. Let t_0, t_1, \dots, t_d be the strictly increasing sequence such that $t_0 = 0$, $t_d = b_u$ (boundary of u) and the intermediate t_i are the sorted threshold values.

Given an expression level $l \in [0, b_u]$ of u , there exists a unique i such that $t_i \leq l < t_{i+1}$ (or $i = d$ if $l = b_u$). We define the folding function $\rho_u : [0, b_u] \rightarrow [0, d]$ by $\rho_u(l) = i$.

For example, if the underlying directed graph of \mathcal{N} is the one of Figure 7 then $\rho_x(0) = 0$, $\rho_x(1) = \rho_x(2) = \rho_x(3) = 1$, $\rho_x(4) = 2$, $\rho_y(0) = \rho_y(1) = 0$, $\rho_y(2) = 1$ and $\rho_y(3) = 2$.



Parameters:

$$K_x=0, K_{x,x}=4, K_{x,y}=2, K_{x,xy}=4, \\ K_y=0, K_{y,x}=3, K_{y,y}=3, K_{y,xy}=3$$

Figure 7. Example of a non canonical BRN

4.3. Embeddings of BRNs

Given a BRN, it is always possible to use the folding function in order to minimize its thresholds and parameters. The resulting BRN takes a form which can be considered as canonical.

Definition (Canonical Form): For any BRN \mathcal{P} , the BRN \mathcal{N} obtained by replacing:

- the boundary b_v of each variable v by $\rho_v(b_v)$
- each threshold s of any edge (u, v) by $\rho_u(s)$
- each parameter $K_{v,\omega}$ by $\rho_v(K_{v,\omega})$

is called the canonical form of \mathcal{P} . The function which associates to a BRN its canonical form is obviously idempotent. A BRN which is equal to its canonical form is said to be canonical.

According to this definition, Figure 5 defines a canonical BRN which is the canonical form of the BRN of Figure 7.

Definition (Embedding of BRNs): Let \mathcal{N} and \mathcal{L} be two BRNs. We say that \mathcal{N} is embedded in \mathcal{L} iff there exists a partial BRN \mathcal{P} of \mathcal{L} whose canonical form is equal to the canonical form of \mathcal{N} .

Remark: This implies that the variables of \mathcal{N} belong to \mathcal{L} ($V^{\mathcal{N}} \subset V^{\mathcal{L}}$). Moreover, \mathcal{N} being given, \mathcal{P} is unique if it exists (since $V^{\mathcal{P}} = V^{\mathcal{N}}$).

For example, the BRN defined in Figure 5 is embedded in the BRN of Figure 6 because the BRN of Figure 7 is a partial BRN of the one of Figure 6 whose canonical form is equal to the BRN of Figure 5 (which is already canonical).

4.4. Preservation of Behaviour

When a BRN \mathcal{N} is embedded in a BRN \mathcal{L} , the preservation of the behaviour of \mathcal{N} means intuitively that each variable of \mathcal{L} which does not belong to \mathcal{N} cannot modify the behaviour of \mathcal{N} . In other words, whatever the expression level of those variables is, it is unable to modify “noticeably” the expression level of the variables of \mathcal{N} . A transition in \mathcal{L} is “noticeable” only if it induces a state modification in the *folded version* of the set of states of \mathcal{N} .

Definition (State Folding): Let \mathcal{P} be any BRN and let $\mu : V^{\mathcal{P}} \rightarrow \mathbb{N}$ be a state of \mathcal{P} . The folded state $\rho(\mu)$ is the state $\mu' : V^{\mathcal{P}} \rightarrow \mathbb{N}$ defined by $\mu'(u) = \rho_u(\mu(u))$ for any variable u of \mathcal{P} .

In order to consider all possible expression levels of a variable without modifying the state of the other variables, we technically need the notion of state assignment.

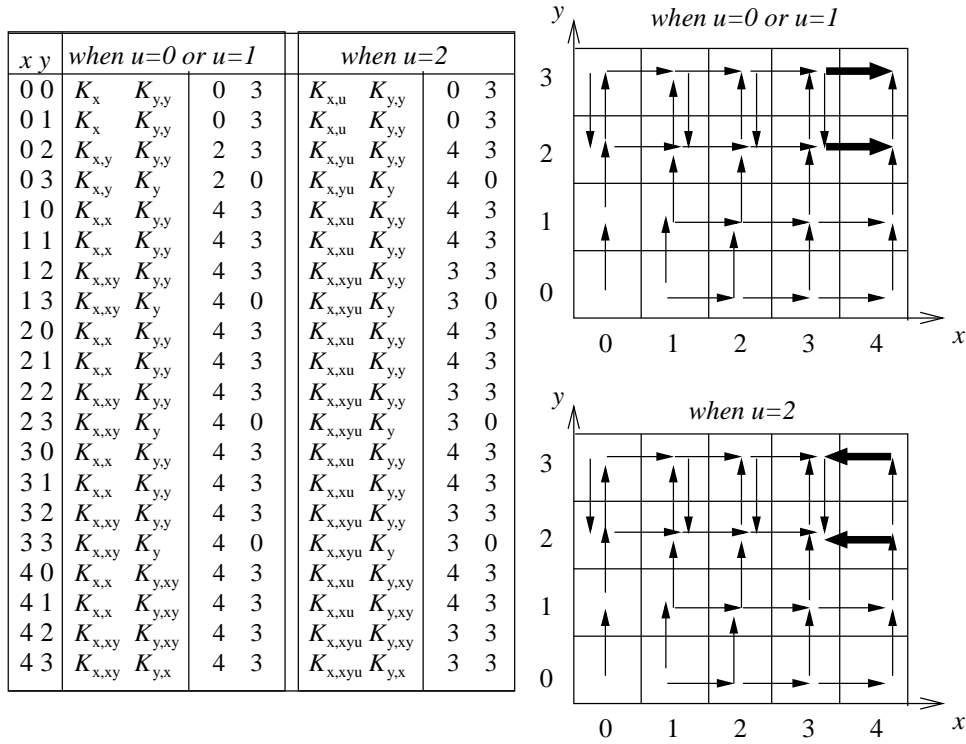
Notation (State Assignment): Let μ be a state of a BRN \mathcal{L} , let u be a variable of \mathcal{L} and let l be an integer such that $0 \leq l \leq b_u$. We note $\mu^{[u \leftarrow l]}$ the state μ' such that $\mu'(u) = l$ and $\mu'(u') = \mu(u')$ for all variables $u' \neq u$. We say that $[u \leftarrow l]$ is a state assignment.

Finally, we can define the preservation of behaviour as follows:

Definition (Preservation of Behaviour): An embedding of a BRN \mathcal{N} into a BRN \mathcal{L} preserves the behaviour of \mathcal{N} iff for any transition $(\mu \rightarrow \mu')$ of $S^{\mathcal{L}}$ such that $\rho(\mu|_{\mathcal{N}}) \neq \rho(\mu'|_{\mathcal{N}})$ and for any state assignment $[u \leftarrow l]$ such that u does not belong to \mathcal{N} , the transition $(\mu^{[u \leftarrow l]} \rightarrow \mu'^{[u \leftarrow l]})$ also belongs to $S^{\mathcal{L}}$.

Remark: The condition $\rho(\mu|_{\mathcal{N}}) \neq \rho(\mu'|_{\mathcal{N}})$ allows us to ignore the states of \mathcal{N} which point to themselves as in Remark of Section 3.2 above.

Let us consider for example the embedding of the BRN of Figure 5 into the BRN of Figure 6. Figure 6 contains five variables which make it difficult to draw its whole state graph. Fortunately, the variables v and w are resources neither of x nor of y . Consequently, only the different values of the variable u have to be considered to detect the transitions of the form $(\rho(\mu|_{\{x,y\}}) \rightarrow \rho(\mu'|_{\{x,y\}}))$ such that $\rho(\mu|_{\{x,y\}}) \neq \rho(\mu'|_{\{x,y\}})$. Figure 8 gives the state graph of the BRN of Figure 6 restricted to $\{x, y\}$ according to these remarks. Consider, for example, the transition from $(x = 3, y = 3)$ to $(x = 4, y = 3)$ when $u = 0$ or $u = 1$. If we assign $[u \leftarrow 2]$, this transition disappears. Consequently, the embedding of variables x and y (i.e. Figure 5) into the BRN of Figure 6 does *not* preserve behaviour, as $\rho_x(3) = 1 \neq \rho_x(4) = 2$. This is indeed the case for the four bold transitions of Figure 8.

Figure 8. The state graph of the BRN of Figure 6 restricted to $\{x, y\}$

5. Main Results

In this section, we prove the main theorem of this article. It establishes that the preservation of behaviour defined above is equivalent to a considerably smaller set of conditions, limited to the parameters of some “frontier” variables of \mathcal{N} . State folding plays an important role and we first prove (Section 5.1) that it makes sense to fold an entire state graph; then we prove (Section 5.2) that canonical forms “commute” with state graph foldings. Lastly, Sections 5.3 and 5.4 give and prove the main theorem.

5.1. State Graph Folding

State folding defined in the previous section preserves the resources of each variable, as shown below.

Proposition: *Let v be a variable of a BRN \mathcal{P} , let μ be a state of \mathcal{P} and let \mathcal{C} be the canonical form of \mathcal{P} . The set of resources of v in \mathcal{P} w.r.t. μ is equal to the set of resources of v in \mathcal{C} w.r.t. $\rho(\mu)$.*

Formally: $\omega_v^{\mathcal{P}}(\mu) = \omega_v^{\mathcal{C}}(\rho(\mu))$.

Proof:

Let u be any variable of $P^{-1}(v)$ and let t be the threshold of $(u \rightarrow v)$. As t is a threshold, ρ_v is defined in such a way that for any expression level l of u we have: $\rho_v(l) \geq \rho_v(t) \Leftrightarrow l \geq t$. Applied to $l = \mu(u)$, since $\rho_u(\mu(u)) = (\rho(\mu))(u)$, it comes $(\rho(\mu))(u) \geq \rho(t) \Leftrightarrow \mu(u) \geq t$. By contraposition, it comes

$(\rho(\mu))(u) < \rho(t) \Leftrightarrow \mu(u) < t$. Thus, by definition of \mathcal{C} , u is a resource of v in \mathcal{P} w.r.t. μ iff it is a resource of v in \mathcal{C} w.r.t. $\rho(\mu)$. \square

This proposition makes it reasonable to introduce the following definition of folded state graphs.

Definition (State Graph Folding): Let \mathcal{N} be a BRN embedded in a BRN \mathcal{L} . The folded state graph of the embedding, by notation abuse $\rho^{\mathcal{N}}(S^{\mathcal{L}})$, is defined by:

- the nodes of $\rho^{\mathcal{N}}(S^{\mathcal{L}})$ are all the possible states of the canonical form of \mathcal{N}
- the edges of $\rho^{\mathcal{N}}(S^{\mathcal{L}})$ are the transitions of the form $(\rho(\mu|_{\mathcal{N}}) \rightarrow \rho(\mu'|_{\mathcal{N}}))$ such that $(\mu \rightarrow \mu')$ is a transition of $S^{\mathcal{L}}$ and such that $\rho(\mu|_{\mathcal{N}}) \neq \rho(\mu'|_{\mathcal{N}})$, where “ $\mu|_{\mathcal{N}}$ ” stands for the restriction of μ to $V^{\mathcal{N}}$ (remind that $V^{\mathcal{N}} \subset V^{\mathcal{L}}$).

Let us again consider the embedding of the BRN of Figure 5 into the BRN of Figure 6. Figure 9 gives the folded state graph of the embedding according to previous definition. In the state graph at the left of Figure 9, the singular values (threshold) drawn with bold lines are the values where $\rho(x)$ or $\rho(y)$ changes. The folded graph at the right of Figure 9 is obtained by keeping only the transitions which cross the bold lines. All the other transitions are ignored.

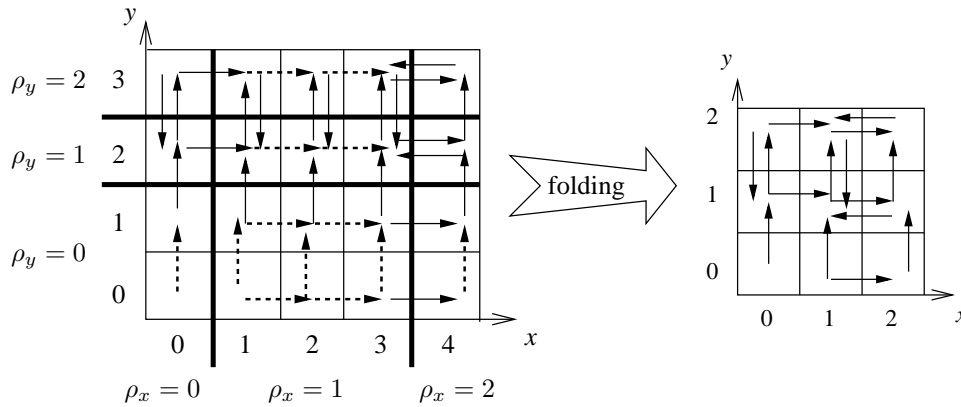


Figure 9. Construction of the folded state graph resulting from the state graphs obtained in Figure 8

5.2. Preservation of state graph

Let \mathcal{N} be a BRN embedded in a BRN \mathcal{L} . Let \mathcal{P} be the partial graph of \mathcal{L} corresponding to the embedding of \mathcal{N} . Since the behaviour of \mathcal{N} is reflected by its state graph, the preservation of its behaviour should imply that the state graph $S^{\mathcal{N}}$ of \mathcal{N} is equal to the state graph $S^{\mathcal{P}}$ of \mathcal{P} . However, two questions must be addressed before proving such a theorem:

- The expression levels of \mathcal{P} are not comparable with the ones of \mathcal{N} due to the folding stuff seen before. Consequently, we have to compare the *folded* versions of $S^{\mathcal{N}}$ and $S^{\mathcal{P}}$.

- When considering \mathcal{N} or \mathcal{P} individually, there are two ways to get a folded view of their behaviour: either one could consider the folded state graphs of \mathcal{N} and \mathcal{P} or one could consider the state graphs of their canonical forms. So, should we compare the folded state graphs of \mathcal{N} and \mathcal{P} or should we compare the state graphs of their canonical forms?

The following theorem solves both of these questions, because it has a corollary which establishes that the folded state graph of a BRN is equal to the state graph of its canonical form (which solves the second question according to the previous definition), and it proves that if an embedding preserves the behaviour then the folded state graphs are equal.

Theorem (Preservation of State Graph): *If the embedding of the BRN \mathcal{N} into the BRN \mathcal{L} preserves the behaviour of \mathcal{N} then the folded state graph $\rho^{\mathcal{N}}(S^{\mathcal{L}})$ is equal to the state graph $S^{\mathcal{C}}$ of the canonical form \mathcal{C} of \mathcal{N} .*

Proof:

Let \mathcal{P} be the partial BRN of \mathcal{L} whose canonical form is \mathcal{C} .

$S^{\mathcal{C}} \subset \rho^{\mathcal{N}}(S^{\mathcal{L}})$: Let $(\nu \rightarrow \nu')$ be any transition of $S^{\mathcal{C}}$. Let us prove that there exists a transition $(\mu \rightarrow \mu')$ of $S^{\mathcal{L}}$ such that $\rho(\mu|_{\mathcal{N}}) = \nu$ and $\rho(\mu'|_{\mathcal{N}}) = \nu'$. This means that $\rho(\mu|_{\mathcal{C}}) = \nu$ and $\rho(\mu'|_{\mathcal{C}}) = \nu'$ because \mathcal{C} and \mathcal{N} have the same set of variables.

Let v be the variable of \mathcal{C} such that $\nu'(v) = \kappa_v^{\mathcal{C}}(\nu) = \nu(v) + \delta$ (with $\delta = \pm 1$). Since ρ_v is surjective and monotonic, there exists a level of expression l_v of v in \mathcal{L} such that $\rho_v(l_v) = \nu(v)$ and $\rho_v(l_v + \delta) = \nu'(v)$. For the same reason, there exists a state η of \mathcal{P} such that $\eta(v) = l_v$ and $\rho(\eta) = \nu$. Now let us consider the variables of $L^{-1}(v)$ which do not belong to \mathcal{P} : it is always possible to assign them expression levels such that they are not resources of v in \mathcal{L} . Let us also choose arbitrary expression levels for the other variables of \mathcal{L} which do not belong to \mathcal{P} : this defines a state μ of \mathcal{L} . By construction, $\omega_v^{\mathcal{L}}(\mu) = \omega_v^{\mathcal{P}}(\eta) = \omega_v^{\mathcal{C}}(\nu)$ and $\rho(\mu|_{\mathcal{C}}) = \nu$ and $\rho(\mu^{[v \leftarrow l_v + \delta]}|_{\mathcal{C}}) = \nu'$. Consequently, it is sufficient to prove that $(\mu \rightarrow \mu^{[v \leftarrow l_v + \delta]})$ is a transition of $S^{\mathcal{L}}$.

This amounts to the demonstration that $\kappa_v^{\mathcal{L}}(\mu) = l_v + \delta$. When $\delta = 1$, it means that $K_{v, \omega_v^{\mathcal{L}}(\mu)}^{\mathcal{L}} > l_v$ and since ρ_v is an increasing function, $\rho_v(K_{v, \omega_v^{\mathcal{L}}(\mu)}^{\mathcal{L}}) > \rho_v(l_v)$ is a sufficient condition. We have $\rho_v(K_{v, \omega_v^{\mathcal{L}}(\mu)}^{\mathcal{L}}) = \rho_v(K_{v, \omega_v^{\mathcal{P}}(\eta)}^{\mathcal{P}}) = K_{v, \omega_v^{\mathcal{C}}(\nu)}^{\mathcal{C}}$ as \mathcal{C} is the canonical form of \mathcal{P} and we have $\rho_v(l_v) = \nu(v)$. So, the sufficient condition becomes $K_{v, \omega_v^{\mathcal{C}}(\nu)}^{\mathcal{C}} > \nu(v)$ which precisely means that $(\nu \rightarrow \nu')$ is a transition of $S^{\mathcal{C}}$. A similar reasoning applies if $\delta = -1$ and this ends this part of the proof.

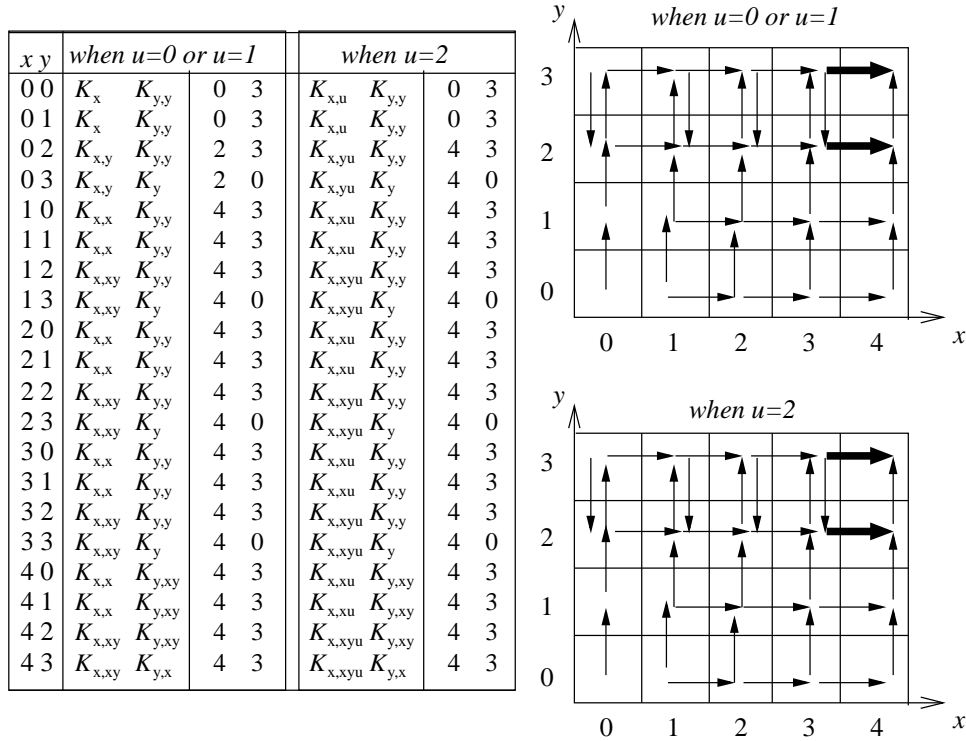
$\rho(S^{\mathcal{L}}) \subset S^{\mathcal{C}}$: Let $(\mu \rightarrow \mu')$ be any transition of $S^{\mathcal{L}}$ such that $\rho(\mu|_{\mathcal{N}}) \neq \rho(\mu'|_{\mathcal{N}})$ (i.e. $\rho(\mu|_{\mathcal{C}}) \neq \rho(\mu'|_{\mathcal{C}})$).

We will first prove that there exists a transition $(\eta \rightarrow \eta')$ of $S^{\mathcal{P}}$ such that $\rho(\mu|_{\mathcal{C}}) = \rho(\eta)$ and $\rho(\mu'|_{\mathcal{C}}) = \rho(\eta')$, and we will secondly prove that $(\rho(\eta) \rightarrow \rho(\eta'))$ belongs to $S^{\mathcal{C}}$.

Let v be the unique variable such that $\mu'(v) = \kappa_v^{\mathcal{L}}(\mu) \neq \mu(v)$. It necessarily belongs to \mathcal{C} as $\rho(\mu|_{\mathcal{C}}) \neq \rho(\mu'|_{\mathcal{C}})$. Let us make an induction on the number n of variables in $\omega_v^{\mathcal{L}}(\mu)$ which do not belong to \mathcal{C} . If $n = 0$ then $\omega_v^{\mathcal{L}}(\mu) = \omega_v^{\mathcal{P}}(\mu|_{\mathcal{C}})$ and we can choose $\eta = \mu|_{\mathcal{C}}$ and $\eta' = \mu'|_{\mathcal{C}}$ because $K_{v, \omega_v^{\mathcal{L}}(\mu)}^{\mathcal{L}} = K_{v, \omega_v^{\mathcal{P}}(\mu)}^{\mathcal{P}}$, thus $\kappa_v^{\mathcal{L}}(\mu) = \kappa_v^{\mathcal{P}}(\eta)$. If $n > 0$, then let us consider u , one of these variables, and let us choose l_u , an expression level of u such that u is not a resource of v . Since the embedding preserves the behaviour, $(\mu^{[u \leftarrow l_u]} \rightarrow \mu'^{[u \leftarrow l_u]})$ also belongs to $S^{\mathcal{L}}$, it has $n - 1$ variables in $\omega_v^{\mathcal{L}}(\mu)$ which do not belong to \mathcal{C} , and $\rho(\mu|_{\mathcal{C}}) = \rho(\mu^{[u \leftarrow l_u]})$ and $\rho(\mu'|_{\mathcal{C}}) = \rho(\mu'^{[u \leftarrow l_u]})$. This ends the first facet of the proof.

Let $(\eta \rightarrow \eta')$ be any transition of $S^{\mathcal{P}}$ such that $\rho(\eta) \neq \rho(\eta')$. Let v be the variable such that $\eta'(v) =$

$\kappa_v^{\mathcal{P}}(\eta) = \eta(v) + \delta$ with $\delta = \pm 1$. Let us assume that $\delta = 1$. Consequently, $K_{v, \omega_v^{\mathcal{P}}(\eta)}^{\mathcal{P}} > \eta(v)$ and since $\rho_v(\eta'(v)) = \rho_v(\eta(v) + 1) \neq \rho_v(\eta(v))$ and since ρ_v is an increasing function, we get $\rho_v(K_{v, \omega_v^{\mathcal{P}}(\eta)}^{\mathcal{P}}) > \rho_v(\eta(v))$. We know that $\rho_v(K_{v, \omega_v^{\mathcal{P}}(\eta)}^{\mathcal{P}}) = K_{v, \omega_v^{\mathcal{C}}(\rho(\eta))}^{\mathcal{C}}$ because \mathcal{C} is the canonical form of \mathcal{P} . Thus, by construction of ρ_v , we have $\kappa_v^{\mathcal{C}}(\rho_v(\eta(v))) = \rho_v(\eta(v)) + 1 = \rho_v(\eta(v) + 1) = \rho_v(\eta'(v))$. Consequently, $(\rho(\eta) \rightarrow \rho(\eta'))$ is a transition of $S^{\mathcal{C}}$. When $\delta = -1$, we follow a similar proof (replacing “ $>$ ” by “ $<$ ”), which ends the proof of the theorem. \square



State graph of Figure 8 with $K_{x,xyu} = 4$ instead of 3.

Figure 10. Modified state graph of the BRN of Figure 6 restricted to $\{x, y\}$

To illustrate the previous theorem, let us come back to the example of Figure 6 and let us modify only one parameter with $K_{x,xyu} = 4$ instead of 3. This defines another embedding of the BRN of Figure 5. Figure 10 gives the corresponding state graph (still restricted to variables x and y). This new embedding preserves the behaviour of the BRN of Figure 5 because the state graph restricted to x and y does not change, whatever the values of u , v and w are. The folded state graph is then constructed in Figure 11 and we see that, as established by our theorem, it is equal to the state graph of Figure 5.

Corollary: *The canonical form \mathcal{C} of a BRN \mathcal{N} preserves the behaviour of \mathcal{N} and consequently, the state graph of \mathcal{C} is equal to the folded state graph of \mathcal{N} .*

Proof:

There is no variable u of \mathcal{N} which does not belong to \mathcal{C} . Consequently, the embedding of \mathcal{C} into \mathcal{N}

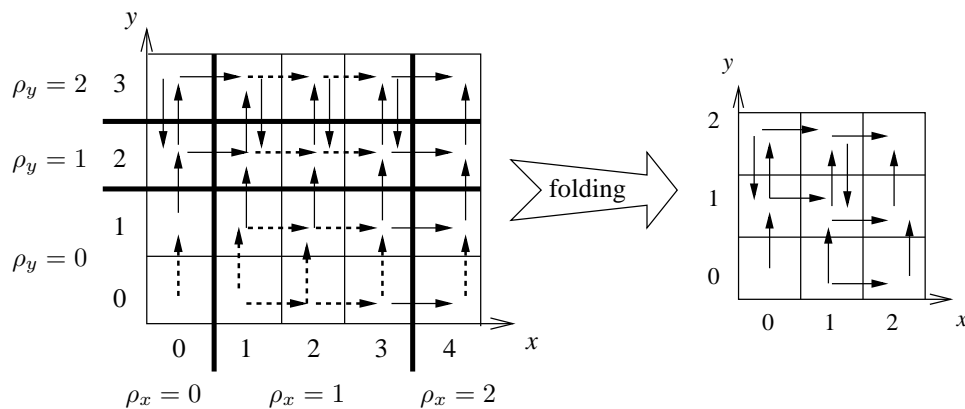


Figure 11. Folding of the modified state graph obtained in Figure 10

preserves the behaviour of \mathcal{C} and the previous theorem establishes that the folded state graph $\rho^{\mathcal{N}}(S^{\mathcal{N}})$ is equal to the state graph $S^{\mathcal{C}}$ of \mathcal{C} . \square

5.3. Local Dynamics

Preservation of behaviour as defined in section 4.4 is based on an exhaustive knowledge of the state graph of the global BRN. We have shown that this definition reflects our intuition of behaviour preservation. However, this definition is difficult to directly check in practice, because it involves a large number of transitions in the complete state graph (e.g. as in Figure 10 for only two variables). In order to more easily check whether an embedding preserves the behaviour, we need to establish local criteria which can be checked on a limited number of edges of the underlying BRN without constructing the whole state graph. The definition of local dynamics preservation below does not refer to state graphs. It only relies on edges and parameters of the global BRN. It requires two intermediate notions: ω -compatibility and the *sigum* function.

Definition (ω -compatibility): Let \mathcal{L} be a BRN, let v be a variable of \mathcal{L} and let ω a subset of $L^{-1}(v)$. An expression level l_v of v is said to be compatible with ω iff the following condition is satisfied:

if $v \in L^{-1}(v)$, let t be the threshold of the edge (v, v) , then $v \in \omega$ iff $l_v \geq t$.

In other words, v belongs to ω iff l_v actually makes v a resource of itself.

Terminology: The *sigum* function is the function *sign* which associate -1 to any strictly negative number, 0 to 0 and 1 otherwise.

Definition (Preservation of the Local Dynamics): Let \mathcal{N} be a BRN embedded into a BRN \mathcal{L} and let v be a variable of \mathcal{N} . An edge (u, v) of \mathcal{L} preserves the local dynamics of v iff for any subset ω of $L^{-1}(v)$ which does not contain u , and for any expression level l_v of v compatible with ω , we have:

$$\text{sign}(\rho_v(K_{v, \omega \cup \{u\}}^{\mathcal{L}}) - \rho_v(l_v)) = \text{sign}(\rho_v(K_{v, \omega}^{\mathcal{L}}) - \rho_v(l_v))$$

Intuitively, the previous formula can be explained as follows:

- If at a given time the variable v has a certain expression level l_v and v has a certain set of resources ω , then we know that the expression level of v will increase (resp. decrease) iff $K_{v,\omega}^{\mathcal{L}}$ is greater than (resp. less than) l_v .
- The variable u thus has no impact on the behaviour of v iff $K_{v,\omega \cup \{u\}}^{\mathcal{L}}$ is also greater than (resp. less than) l_v . It can be expressed as: $(K_{v,\omega \cup \{u\}}^{\mathcal{L}} - l_v)$ has the same sign than $(K_{v,\omega}^{\mathcal{L}} - l_v)$.
- Lastly, one should not forget the folding function ρ_v , because we only look at the behaviour of the embedded BRN.

The preservation of local dynamics as defined in this section seems intuitively far less restrictive than the preservation of behaviour of Section 4.4, especially if it is checked only on a small subset of the edges of \mathcal{L} . The goal of Section 5.4 is to prove that the preservation of local dynamics, when checked on a well chosen subset of edges of \mathcal{L} , is in fact equivalent to the preservation of behaviour. From an algorithmic point of view, this result is decisive to make practicable the verification of behaviour preservation.

5.4. The Main Theorem

When \mathcal{N} is embedded into \mathcal{L} the influence of \mathcal{L} on the behaviour of \mathcal{N} necessarily goes through the edges of \mathcal{L} entering into \mathcal{N} . The set of these edges is called the “input frontier.”

Definition (Input Frontier): Let \mathcal{N} be a BRN embedded into a BRN \mathcal{L} . The input frontier of \mathcal{N} in \mathcal{L} is the set of edges (u, v) of \mathcal{L} such that $u \notin \mathcal{N}$ and $v \in \mathcal{N}$.

For example, the input frontier of the embedding of Figure 6 is reduced to the edge (u, x) . Notice that the outgoing edges (going from x or y to u , v or w) are not in the input frontier. Indeed, outgoing edges have no influence on \mathcal{N} .

The following theorem shows that it is sufficient to check the local dynamics on the input frontier.

Theorem (Main Result): Let \mathcal{N} be a BRN embedded into a BRN \mathcal{L} . A necessary and sufficient condition to preserve the behaviour of \mathcal{N} in \mathcal{L} is that all edges of the input frontier preserve the local dynamics.

Proof:

Necessary condition: Let us assume that $(u \rightarrow v)$ in the input frontier does not preserve the local dynamics of v for a given set of resources ω in \mathcal{L} . The values $(\rho_v(K_{v,\omega \cup \{u\}}^{\mathcal{L}}) - \rho_v(l_v))$ and $(\rho_v(K_{v,\omega}^{\mathcal{L}}) - \rho_v(l_v))$ do not share the same sign (more precisely, they are not both positive, or both negative, or both null). Since ρ_v is monotonic, $(K_{v,\omega \cup \{u\}}^{\mathcal{L}} - l_v)$ and $(K_{v,\omega}^{\mathcal{L}} - l_v)$ do not share the same sign. Let μ be a state such that $\omega_v^{\mathcal{L}}(\mu) = \omega$ and let l_u be an expression level of u such that u becomes a resource of v (μ and l_u exist). We necessarily have $\kappa_v^{\mathcal{L}}(\mu) \neq \kappa_v^{\mathcal{L}}(\mu^{[u \leftarrow l_u]})$. Let us assume for example that $\kappa_v^{\mathcal{L}}(\mu) \neq 0$ (a symmetric reasoning applies if $\kappa_v^{\mathcal{L}}(\mu^{[u \leftarrow l_u]}) \neq 0$). Then the transition $(\mu \rightarrow \mu')$, where $\mu' = \mu^{[v \leftarrow \kappa_v^{\mathcal{L}}(\mu)]}$, belongs to $S^{\mathcal{L}}$ while $(\mu^{[u \leftarrow l_u]} \rightarrow \mu'^{[u \leftarrow l_u]})$ does not. Thus, the embedding does not preserve the behaviour of \mathcal{N} in \mathcal{L} .

Sufficient condition: Let $(\mu \rightarrow \mu') \in S^{\mathcal{L}}$ such that $\rho(\mu|_{\mathcal{N}}) \neq \rho(\mu'|_{\mathcal{N}})$ and let $[u \leftarrow l]$ be a state assignment such that $u \notin \mathcal{N}$. Let v be the variable such that $\mu(v) \neq \mu'(v)$: it belongs to \mathcal{N} because $\rho(\mu|_{\mathcal{N}}) \neq \rho(\mu'|_{\mathcal{N}})$. It is sufficient to prove that $\kappa_v^{\mathcal{L}}(\mu) = \kappa_v^{\mathcal{L}}(\mu^{[u \leftarrow l]})$, because it implies that $(\mu^{[u \leftarrow l]} \rightarrow \mu'^{[u \leftarrow l]})$ also belongs to $S^{\mathcal{L}}$.

If $\omega_v^{\mathcal{L}}(\mu) = \omega_v^{\mathcal{L}}(\mu^{[u \leftarrow l]})$ then obviously $\kappa_v^{\mathcal{L}}(\mu) = \kappa_v^{\mathcal{L}}(\mu^{[u \leftarrow l]})$.

If $\omega_v^{\mathcal{L}}(\mu) \neq \omega_v^{\mathcal{L}}(\mu^{[u \leftarrow l]})$ then u belongs to $L^{-1}(v)$ and $(u \rightarrow v)$ belongs consequently to the input frontier. We know that $\rho_v(\kappa_v^{\mathcal{L}}(\mu)) - \rho_v(\mu(v)) \neq 0$ because $\rho(\mu|_N) \neq \rho(\mu'|_N)$. Thus, by definition of κ , and ρ_v being an increasing function, $\rho_v(K_{v, \omega_v^{\mathcal{L}}(\mu)}^{\mathcal{L}}) - \rho_v(\mu(v)) \neq 0$. The preservation of the local dynamics implies that it has the same sign as $\rho_v(K_{v, \omega_v^{\mathcal{L}}(\mu^{[u \leftarrow l]})}^{\mathcal{L}}) - \rho_v(\mu(v))$. Finally, ρ_v being an increasing function, $K_{v, \omega_v^{\mathcal{L}}(\mu)}^{\mathcal{L}} - \mu(v)$ and $K_{v, \omega_v^{\mathcal{L}}(\mu^{[u \leftarrow l]})}^{\mathcal{L}} - \mu(v)$ have the same sign and are different from 0. This implies that $\kappa_v^{\mathcal{L}}(\mu) = \kappa_v^{\mathcal{L}}(\mu^{[u \leftarrow l]})$, which ends the proof. \square

Example: Let us consider the embedding of Figure 6. The edge $(u \rightarrow x)$ constitutes the input frontier.

- Let us assume for example that $l_x = 3$. We have $K_{x, xy=4}$ and $K_{x, xyu=3}$ as well as $\rho_x(1) = \rho_x(2) = \rho_x(3) = 1$ and $\rho_x(4) = 2$. Thus, $\rho_x(K_{x, xy}) - \rho_x(l_x) = 1$ and $\rho_x(K_{x, xyu}) - \rho_x(l_x) = 0$. So, local dynamics is not preserved.
- Looking at Figure 8, one can see why the behaviour is not preserved: the bold arrows of the state graph differ if $u = 2$ (i.e. u is a resource of x) or not.
- As a consequence, when comparing the state graph of Figure 5 with the state graph of Figure 9, one can see that the bold arrows added by the case $u = 2$ in Figure 8 have been added in Figure 9. Thus, the global behaviour is not preserved.

Notice that because $K_{x, y=2}$ and $K_{x, yu=4}$, if l_x is equal to 1, 2 or 3 then $\rho_x(K_{x, y}) - \rho_x(l_x) = 0$ and $\rho_x(K_{x, yu}) - \rho_x(l_x) = 1$. Nevertheless, the local dynamics is preserved because when $l_x \geq 1$, x is a resource of itself. Consequently, only the case where $l_x = 0$ has to be considered and then both expressions have the same sign.

Let us consider now the same embedding except that $K_{x, xyu=4}$ instead of 3.

- Then the local dynamics is preserved because for any possible value of l_x and compatible ω , $\rho_x(K_{x, \omega \cup \{u\}}) - \rho_x(l_x)$ has the same sign than $\rho_x(K_{x, \omega}) - \rho_x(l_x)$ (and remind that (u, x) is the only edge of the input frontier).
- According to the state graph side, we see on Figure 10 that even when $u = 2$, the reverse bold arrows of Figure 8 do not occur.
- As a consequence, the state graph of Figure 11 is identical to the one of Figure 5.

6. Sufficient conditions to find behaviour preserving subnetworks

Let us assume that we want to verify whether the embedding of a subnetwork \mathcal{N} of \mathcal{L} preserves its behaviour. The main advantage of the previous theorem is to avoid building the state graphs and more importantly, to avoid checking the impact of all possible state assignments of the variables which do not belong to \mathcal{N} . It is sufficient to refer to a small number of static parameters of the BRN \mathcal{L} (in the previous example only one edge has to be considered, instead of a state graph of dimension 5).

Since embeddings which preserve behaviour identify sub-regulatory networks whose dynamics cannot be influenced by the rest of the global BRN, those subnetworks are biologically interesting. Indeed,

they “pilot” the whole behaviour of the global BRN since they can influence it *via* their “output frontier” whilst they are not influenced *via* their input frontier.

According to this point of view, one of the advantages of the modelling activity could be to automatically find those “pilot” subnetworks. Notice that this question is algorithmically far more difficult than verifying *a posteriori* that an embedding preserves the behaviour of a *given* subnetwork. Checking all the edges for all possible input frontiers induces a combinatorial explosion.

Nevertheless, our main theorem provides us with a simple sufficient condition. This condition can easily be verified by classical algorithms on graphs.

Corollary: *If \mathcal{P} is a partial BRN of \mathcal{L} whose input frontier is empty then the embedding of \mathcal{P} into \mathcal{L} preserves the behaviour of \mathcal{P} .*

Proof:

Obvious from the main theorem. □

In practice, this sufficient condition already identifies a certain number of interesting pilot subnetworks.

In the general case, it is indeed not necessary to compute all the algebraic expressions involved in the definition of the preservation of local dynamics (where one algebraic expression per possible expression level l_v of the variable v is required). Provided that the variable v has no direct influence on itself, one can take benefit of the following result:

Proposition: *Let \mathcal{N} be a BRN embedded into a BRN \mathcal{L} and let (u, v) be an edge of the input frontier such that v has no direct influence on itself (i.e. $(v, v) \notin E^{\mathcal{N}}$). Then (u, v) preserves the local dynamics of v if and only if: for any subset ω of $L^{-1}(v)$ which does not contain u , $\rho_v(K_{v, \omega \cup \{u\}}^{\mathcal{L}}) = \rho_v(K_{v, \omega}^{\mathcal{L}})$.*

Proof:

If $\rho_v(K_{v, \omega \cup \{u\}}^{\mathcal{L}}) = \rho_v(K_{v, \omega}^{\mathcal{L}})$ then the preservation of the local dynamics is obvious. Conversely, if for any expression level l_v compatible with ω we have $\text{sign}(\rho_v(K_{v, \omega \cup \{u\}}^{\mathcal{L}}) - \rho_v(l_v)) = \text{sign}(\rho_v(K_{v, \omega}^{\mathcal{L}}) - \rho_v(l_v))$, then, provided that $v \notin L^{-1}(v)$, it is always possible to choose $l_v = K_{v, \omega}^{\mathcal{L}}$, which proves that $\rho_v(K_{v, \omega \cup \{u\}}^{\mathcal{L}}) = \rho_v(K_{v, \omega}^{\mathcal{L}})$. □

This proposition avoids generating all possible expression levels of the variable v . It is sufficient to test the equality of the folded values of some parameters of the BRN. When the variable v has a direct influence on itself, it is less simple but a similar result holds:

Proposition: *Let \mathcal{N} be a BRN embedded into a BRN \mathcal{L} and let (u, v) be an edge of the input frontier such that v has a direct influence on itself (i.e. $(v, v) \in E^{\mathcal{N}}$) with a threshold s and a sign ε . Then (u, v) preserves the local dynamics of v if and only if for any subset ω of $L^{-1}(v)$ which does not contain u nor v , the four following conditions are satisfied:*

1. $\left\{ \begin{array}{l} \max(K_{v, \omega \cup \{u, v\}}^{\mathcal{L}}, K_{v, \omega \cup \{v\}}^{\mathcal{L}}) \geq s \\ \text{and } \varepsilon = \text{“+”} \end{array} \right\} \implies \rho_v(K_{v, \omega \cup \{u, v\}}^{\mathcal{L}}) = \rho_v(K_{v, \omega \cup \{v\}}^{\mathcal{L}})$
2. $\left\{ \begin{array}{l} \min(K_{v, \omega \cup \{u, v\}}^{\mathcal{L}}, K_{v, \omega \cup \{v\}}^{\mathcal{L}}) < s \\ \text{and } \varepsilon = \text{“-”} \end{array} \right\} \implies \rho_v(K_{v, \omega \cup \{u, v\}}^{\mathcal{L}}) = \rho_v(K_{v, \omega \cup \{v\}}^{\mathcal{L}})$

$$3. \left\{ \begin{array}{l} \min(K_{v, \omega \cup \{u\}}^{\mathcal{L}}, K_{v, \omega}^{\mathcal{L}}) < s \\ \text{and } \varepsilon = "+" \end{array} \right\} \implies \rho_v(K_{v, \omega \cup \{u\}}^{\mathcal{L}}) = \rho_v(K_{v, \omega}^{\mathcal{L}})$$

$$4. \left\{ \begin{array}{l} \max(K_{v, \omega \cup \{u\}}^{\mathcal{L}}, K_{v, \omega}^{\mathcal{L}}) \geq s \\ \text{and } \varepsilon = "-" \end{array} \right\} \implies \rho_v(K_{v, \omega \cup \{u\}}^{\mathcal{L}}) = \rho_v(K_{v, \omega}^{\mathcal{L}})$$

Proof:

Preservation of local dynamics implies condition 1: If $\max(K_{v, \omega \cup \{u, v\}}^{\mathcal{L}}, K_{v, \omega \cup \{v\}}^{\mathcal{L}}) \geq s$ and $\varepsilon = "+"$, then choosing $l_v = \max(K_{v, \omega \cup \{u, v\}}^{\mathcal{L}}, K_{v, \omega \cup \{v\}}^{\mathcal{L}})$ is compatible with the set of resources $\omega \cup \{v\}$. Consequently, from our main theorem, $\text{sign}(\rho_v(K_{v, \omega \cup \{u, v\}}^{\mathcal{L}}) - \rho_v(l_v)) = \text{sign}(\rho_v(K_{v, \omega \cup \{v\}}^{\mathcal{L}}) - \rho_v(l_v))$ and l_v being equal to one of the two parameters $K_{v, \omega \cup \{u, v\}}^{\mathcal{L}}$ or $K_{v, \omega \cup \{v\}}^{\mathcal{L}}$, both signs are null. Consequently $\rho_v(K_{v, \omega \cup \{u, v\}}^{\mathcal{L}}) = \rho_v(K_{v, \omega \cup \{v\}}^{\mathcal{L}}) = \rho_v(l_v)$ which proves the implication.

Preservation of local dynamics implies condition 2: The same reasoning holds when choosing $l_v = \min(K_{v, \omega \cup \{u, v\}}^{\mathcal{L}}, K_{v, \omega \cup \{v\}}^{\mathcal{L}})$.

Preservation of local dynamics implies condition 3: If $\min(K_{v, \omega \cup \{u\}}^{\mathcal{L}}, K_{v, \omega}^{\mathcal{L}}) < s$ and $\varepsilon = "+"$, then choosing $l_v = \min(K_{v, \omega \cup \{u\}}^{\mathcal{L}}, K_{v, \omega}^{\mathcal{L}})$ is compatible with the set of resources ω (which does not contain v). Consequently, our theorem proves the implication in the same way as for the condition 1.

Preservation of local dynamics implies condition 4: The same reasoning holds when choosing $l_v = \max(K_{v, \omega \cup \{u\}}^{\mathcal{L}}, K_{v, \omega}^{\mathcal{L}})$.

Conditions 1 to 4 imply the preservation of local dynamics: From our main theorem it is sufficient to prove that for any subset ω' of $L^{-1}(v)$ which does not contain u , and for any expression level l_v of v compatible with ω' , we have $\text{sign}(\rho_v(K_{v, \omega' \cup \{u\}}^{\mathcal{L}}) - \rho_v(l_v)) = \text{sign}(\rho_v(K_{v, \omega'}^{\mathcal{L}}) - \rho_v(l_v))$. Let us denote ε and s respectively the sign and the threshold of the direct influence of v on itself.

If $\varepsilon = "+"$ then the four following cases can be distinguished.

Case $l_v \geq s$ and $\max(K_{v, \omega' \cup \{u\}}^{\mathcal{L}}, K_{v, \omega'}^{\mathcal{L}}) \geq s$: Since $l_v \geq s$, ω' is of the form $\omega' = \omega \cup \{v\}$. Consequently, condition 1 implies that $\rho_v(K_{v, \omega' \cup \{u\}}^{\mathcal{L}}) = \rho_v(K_{v, \omega'}^{\mathcal{L}})$, a fortiori $\text{sign}(\rho_v(K_{v, \omega' \cup \{u\}}^{\mathcal{L}}) - \rho_v(l_v))$ is equal to $\text{sign}(\rho_v(K_{v, \omega'}^{\mathcal{L}}) - \rho_v(l_v))$.

Case $l_v \geq s$ and $\max(K_{v, \omega' \cup \{u\}}^{\mathcal{L}}, K_{v, \omega'}^{\mathcal{L}}) < s$: $\rho_v(K_{v, \omega'}^{\mathcal{L}}) < \rho_v(s)$ because $K_{v, \omega'}^{\mathcal{L}} < s$ and s being a threshold of \mathcal{N} , the folding function preserves this strict inequality. For the same reason $\rho_v(K_{v, \omega' \cup \{u\}}^{\mathcal{L}}) < \rho_v(s)$. Consequently, $\text{sign}(\rho_v(K_{v, \omega' \cup \{u\}}^{\mathcal{L}}) - \rho_v(l_v)) = \text{sign}(\rho_v(K_{v, \omega'}^{\mathcal{L}}) - \rho_v(l_v)) = -1$, as $\rho_v(s) \leq \rho_v(l_v)$.

Case $l_v < s$ and $\min(K_{v, \omega' \cup \{u\}}^{\mathcal{L}}, K_{v, \omega'}^{\mathcal{L}}) \geq s$: We have the inequalities $\rho_v(K_{v, \omega'}^{\mathcal{L}}) \geq \rho_v(s)$ and $\rho_v(K_{v, \omega' \cup \{u\}}^{\mathcal{L}}) \geq \rho_v(s)$ and $\rho_v(s) > \rho_v(l_v)$ for the same reasons as in the previous case. Consequently, $\text{sign}(\rho_v(K_{v, \omega' \cup \{u\}}^{\mathcal{L}}) - \rho_v(l_v)) = \text{sign}(\rho_v(K_{v, \omega'}^{\mathcal{L}}) - \rho_v(l_v)) = 1$.

Case $l_v < s$ and $\min(K_{v, \omega' \cup \{u\}}^{\mathcal{L}}, K_{v, \omega'}^{\mathcal{L}}) < s$: Since $l_v < s$, ω' does not contain v . Consequently, condition 3 implies that $\rho_v(K_{v, \omega' \cup \{u\}}^{\mathcal{L}}) = \rho_v(K_{v, \omega'}^{\mathcal{L}})$, a fortiori $\text{sign}(\rho_v(K_{v, \omega' \cup \{u\}}^{\mathcal{L}}) - \rho_v(l_v)) = \text{sign}(\rho_v(K_{v, \omega'}^{\mathcal{L}}) - \rho_v(l_v))$.

If $\varepsilon = "-"$ then the same above cases are solved with the same reasoning using conditions 2 and 4. \square

Example:

According to the embedding of the BRN of Figure 5 into the BRN of Figure 6, the next to last proposition can be applied to the edge (u, x) . We have to verify whether for any subset ω of $L^{-1}(x)$ which does not contain u , we have $\rho_x(K_{x, \omega \cup \{u\}}^{\mathcal{L}}) = \rho_x(K_{x, \omega}^{\mathcal{L}})$.

ω can be equal to: the empty set, the singletons x or y , or the pair x, y . This would mean:

1. $\rho_x(K_{x, \{u\}}^{\mathcal{L}}) = \rho_x(K_x^{\mathcal{L}})$
2. $\rho_x(K_{x, \{x, u\}}^{\mathcal{L}}) = \rho_x(K_{x, \{x\}}^{\mathcal{L}})$
3. $\rho_x(K_{x, \{y, u\}}^{\mathcal{L}}) = \rho_x(K_{x, \{y\}}^{\mathcal{L}})$
4. $\rho_x(K_{x, \{x, y, u\}}^{\mathcal{L}}) = \rho_x(K_{x, \{x, y\}}^{\mathcal{L}})$

This embedding does not preserve the local dynamics because the false equation is not satisfied: $K_{x, \{x, y, u\}}^{\mathcal{L}} = 3$ and $K_{x, \{x, y\}}^{\mathcal{L}} = 4$, and $\rho_x(4) = 2$ and $\rho_x(3) = 1$.

On the contrary, if we modify the BRN of Figure 6 with $K_{x, \{x, y, u\}}^{\mathcal{L}} = 4$ then the four previous equations are satisfied and consequently the embedding preserves the behaviour.

7. Conclusion and perspectives

The theory of René Thomas is a discrete formalism for modelling biological regulatory networks. This formalism has the advantage to handle the precise dynamics of gene regulation while benefiting from automated logical tools. Our motivation is to develop a modular approach which allows the study of large networks by splitting them into small sub-networks. Our theorem characterizes the sub-networks which “pilot” the rest of the network since they cannot be influenced by the global network. We have given several corollaries which facilitate the algorithmic search for these pilot sub-networks.

The main theorem established in this paper solves a strong view of modularity, since it applies to sub-networks whose behaviour is entirely preserved by embedding. In the future, we would like to develop theorems where the behaviour of the sub-network can be slightly modified by the global network but in a rigorously controlled way.

Several extensions of this work can be considered in the near future. For example, we are interested in sub-networks whose behaviour is preserved only for a subset of possible initial states. We are also interested in sub-networks whose behaviour preserves only some given properties, such as the preservation of the steady states (with possibly different trajectories).

So far, almost all the theoretical work related to René Thomas’ framework has been focused on global properties of the network dynamics. The theoretical result given here opens the way to new studies for biological regulatory networks related to modularity issues.

References

- [1] Altinok, A., Lévi, F., Goldbeter, A.: A cell cycle automaton model for probing circadian patterns of anticancer drug delivery, *Adv Drug Deliv Rev.*, **138**, 2007, 277–288.

- [2] Aracena, J., Demongeot, J., Goles, E.: Fixed points and maximal independent sets in AND-OR networks, *Discrete Applied Mathematics*, **138**, 2004, 277–288.
- [3] Bassano, V., Bernot, G.: Marked Regulatory Graphs: A formal framework to simulate Biological Regulatory Networks with simple automata, *Proc. of the 14th IEEE International Workshop on Rapid System Prototyping, USA*, 2003.
- [4] Bernot, G., Comet, J., Richard, A., Guespin, J.: A Fruitful Application of Formal Methods to Biological Regulatory Networks: Extending Thomas' Asynchronous Logical Approach with Temporal Logic, *Journal of Theoretical Biology*, **229**(3), 2004, 339–347.
- [5] Chaouiya, C.: Petri net modelling of biological networks, *Briefings in Bioinformatics*, **8**, 2007, 210–219.
- [6] Chen, C., Csikasz-Nagy, A., Gyorffy, B., Val, J., Novak, B., Tyson, J.: Kinetic analysis of a molecular model of the budding yeast cell cycle, *Molecular Biology of the Cell*, **11**, 2000.
- [7] Correale, L., Leone, M., Pagnani, A., Weigt, M., Zecchina, R.: Core percolation and onset of complexity in boolean networks, *Phys Rev Lett.*, **96**(1), 2006.
- [8] De Jong, H.: Modeling and Simulation of Genetic Regulatory Systems: A Literature Review, *Journal of Computational Biology*, **9**(1), 2002, 67–103.
- [9] De Jong, H., Geiselmann, J., Hernandez, C., Page, M.: Genetic network analyzer: qualitative simulation of genetic regulatory networks, *Bioinformatics*, 2003, 336–44.
- [10] De Jong, H., Page, M., Hernandez, C., Geiselmann, J.: Qualitative simulation of genetic regulatory networks: Method and application, *In Proc. of the 17th Inter Joint Conf on Artificial Intelligence, USA*, 2001.
- [11] De Jong, H., Tombor, B., Albert, R., Oltvai, Z. N., Barabási, A.-L.: The large-scale organization of metabolic networks, *Nature*, **407**, 2000, 651–654.
- [12] Demongeot, J., Aracena, J., Thuderoz, F., Baum, T.-P., Cohen, O.: Genetic regulation networks: circuits, regulons and attractors, *C.R. Biologies*, **326**, 2003, 171–188.
- [13] Demongeot, J., Kaufman, M., Thomas, R.: Positive feedback circuits and memory, *C R Acad Sci III*, 2000, 69–79.
- [14] Gilbert, D., Heiner, M., Lehrack, S.: A Unifying Framework for Modelling and Analysing Biochemical Pathways Using Petri Nets, *CMSB 2007 (Computational Methods in Systems Biology)*, Springer-Verlag in LNCS/LNBI, 4695, 2007.
- [15] Gouzé, J.: Positive and negative circuits in dynamical systems, *Journal of Biological Systems*, **6**, 1998, 11–15.
- [16] Guelzim, N., Bottani, S. and Bourguin, P., Képès, F.: Topological and causal structure of the yeast transcriptional regulatory network, *Nature Genetics*, **31**, 2002, 60–63.
- [17] Guespin, J., Bernot, G., Comet, J.-P., Mérieau, A., Richard, A., Hulén, C., Polack, B.: Epigenesis and dynamic similarity in two regulatory networks in *Pseudomonas aeruginosa*, *Acta Biotheoretica*, **52**(4), 2004, 379–390.
- [18] Guespin-Michel, J., Kaufman, M.: Positive feedback circuits and adaptive regulations in bacteria, *Acta Biotheoretica*, **49**(4), 2001, 207–218.
- [19] Kaufman, M., Soulé, C., Thomas, R.: A new necessary condition on interaction graphs for multistationarity, *J Theor Biol.*, **4**(248), 2007, 675–85.
- [20] Louis, M., Holm, L., Sanchez, L., Kaufman, M.: A Theoretical Model for the Regulation of Sex-lethal, a Gene That Controls Sex Determination and Dosage Compensation in *Drosophila melanogaster*, *Genetics*, **165**, 2003, 1355–1384.

- [21] Maslov, S., Sneppen, K.: Specificity and stability in topology of protein networks, *Science*, **296**, 2002, 910–913.
- [22] Mendoza, L., Thieffry, D., Alvarez-Buylla, E.: Genetic control of flower morphogenesis in *Arabidopsis thaliana*: a logical analysis, *Bioinformatics*, 1999, 593–606.
- [23] Mutaille, L., Thieffry, D., Leo, O., Kaufman, M.: Toxicity and Neuroendocrine Regulation of the Immune Response: A model Analysis, *Journal of Theoretical Biology*, 1996, 285–305.
- [24] Pérès, S., Comet, J.-P.: Contribution of Computational Tree Logic to biological regulatory network: example from *Pseudomonas Aeruginosa*, *In Internat. workshop on Comput. Methods in System Biology*, 2003.
- [25] Plathe, E., Mestl, T., Omholt, S.: Feedback loops, stability and multistationarity in dynamical systems, *Journal of Biological Systems*, **3**, 1995, 569–577.
- [26] Remy, E., Mossé, B., Chaouiya, C., Thieffry, D.: A description of dynamical graphs associated to elementary regulatory circuits, *Bioinformatics*, **19**(2), 2003, 172–178.
- [27] Remy, E., Ruet, P., Thieffry, D.: Graphic requirements for multistability and attractive cycles in a boolean dynamical framework, *Prépublication IML*, 2005.
- [28] Sanchez, L., Van Helden, J., Thieffry, D.: Establishment of the dorso-ventral pattern during embryonic development of *Drosophila melanogaster*: a logical analysis, *Journal of Theoretical Biology*, **189**(4), 1997, 377–389.
- [29] Snoussi, E.: Qualitative dynamics of piecewise-linear differential equation: a discrete mapping approach, *J. Math. Biology*, 1989, 189–207.
- [30] Snoussi, E., Thomas, R.: Logical identification of all steady states: the concept of feedback loop characteristic states, *B. of mathematical biology*, 1993, 973–991.
- [31] Soulé, C.: Graphical requirements for multistationarity, *ComplexUs*, **1**, 2003, 123–133.
- [32] Sánchez, L., Thieffry, D.: A logical analysis of the gap gene system, *Journal of theoretical Biology*, **211**, 2001, 115–141.
- [33] Sánchez, L., Thieffry, D.: Segmenting the fly embryo: a logical analysis of the pair-rule cross-regulatory module, *Journal of theoretical Biology*, **224**, 2003, 517–537.
- [34] Thieffry, D., Romere, D.: The modularity of biological regulatory networks, *Biosystems*, 1998, 49–59.
- [35] Thieffry, D., Snoussi, E., Richelle, J., Thomas, R.: Positive loops and differentiation, *J. Biol. Systems*, 1995.
- [36] Thieffry, D., Thomas, R.: Qualitative analysis of gene networks, *In Pac Symp Biocomput*, 1998.
- [37] Thomas, R.: Regulatory networks seen as asynchronous automata: A logical description, *J. theoretical Biology*, 1991, 1–23.
- [38] Thomas, R., d’Ari, R.: *Biological Feedback*, *CRC Press*, 1990.
- [39] Thomas, R., Kaufman, M.: Multistationarity, the basis of cell differentiation and memory. II. Logical analysis of regulatory networks in terms of feedback circuits, *Chaos*, **11**, 2001, 180–195.
- [40] Thomas, R., Thieffry, D.: Dynamical behavior of biological regulatory networks-II. Immunity control in bacteriophage lambda, *B. of mathematical biology*, 1995, 277–297.
- [41] Thomas, R., Thieffry, D., Kaufman, M.: Dynamical of biological regulatory networks-I. Biological role of feedback loops and practical use of the concept of the loop-characteristic state, *B. of mathematical biology*, 1995, 247–275.

- [42] Troncale, S., Tahi, F., Campard, D., Vannier, J., Guespin, J.: Modeling and simulation with Hybrid Functional Petri Nets of the role of Interleukin-6 in human early haematopoiesis, *In Proc. of Pacific Symposium on Biocomputing (PSB)*, 2006.
- [43] Tyson, J., Chen, C., Novak, B.: Network Dynamics and Cell Physiology, *Nature Reviews Molecular Cell Biology*, **2**, 2001, 908–916.
- [44] Tyson, J., Novák, B.: Regulation of the Eukaryotic Cell Cycle: Molecular Antagonism, Hysteresis, and Irreversible Transitions, *J. Theoret. Biol.*, **249**, 2001, 249–263.