Spatial information to restrict the dynamics of genetic regulatory networks

Matthieu Manceny[1,2], *Marc Aiguier*[1], *Pascale Le Gall*[1,2]
Joan Hérisson[2], *Ivan Junier*[2] *and François Képès*[2].

1. **École Centrale Paris, MAS Laboratory, France**
2. **Epigenomics Project, University of Évry, France**

Abstract

In the course of understanding the functioning of cellular processes, modelling frameworks for biological networks are mandatory in order to reason on the models and their properties. One of the main problems with such modelling framework is to determine the dynamics of gene regulatory networks (GRN). Formal techniques, most of them based on model checking, have been applied to select valid dynamics, that is dynamics consistant with biological experiments expressed by temporal properties. The problem is that these formal techniques rapidly become intractable because dynamics associated to the GRN are most of the time very numerous. Recently, it has been observed in *in vivo* experiments and in genomic and transcriptomic studies, that spatial informations are necessary to better understand both the mechanisms and the dynamics of GRN. In this paper we propose to extend the modelling framework of R. Thomas in order to introduce such spatial information between genes. We will show how these further informations allow us to restrict dynamics of GRN. We will illustrate our approach on two classical models of GRN: the mucus production in *Pseudomonas aeruginosa* and the lytic/lysogenic switch in the lambda phage.

Keywords: Gene regulatory networks, Spatial information, Dynamics, Discrete mathematical modelling.

1 Introduction

To understand genetic regulatory networks, modeling frameworks and simulation techniques are often useful since the complexity of the interactions between constituents of the network (mainly genes and proteins) makes intuitive reasoning difficult dJ2002. Nevertheless, simulation techniques are in practice difficult to manage because biological systems are either large, complex or partially known. Indeed, the lack of precise knowledge about the system (are all constituents/interactions taken into account? Which values are given to parameters? Which is the confidence on these parameters?...) is one of the more accurate difficulties to handle computationally all possible hypotheses on the system. Most of the time, parameters of the model have to be inferred from a set of biological experiments. Unfortunately these parameters are rarely measurable and modelling process has to focus on the search of values that lead to a dynamics which is coherent with experiments. Hence, it seems necessary to embed within the model some biological knowledge in order to reduce the complexity of

searching parameters. The notion we are interested in is the spatial relation between genes.

Indeed, recent experiments have shown that both in eukaryotes [6] and in bacteria [2] gene transcription occurs in discrete foci where several RNA polymerases (the transcribing elements) co-localize. This suggests that genes tend also to co-localize in space in order to optimize transcription rates. Such scenario is supported by genomic and transcriptomic analysis [7, 3]. These have revealed that the genes that are regulated by a given transcription factor and the gene that codes for the transcription factor tend to be located periodically along the DNA [7]. In this way, the genes can be easily co-localized in the three-dimensional space according to a solenoidal structure, even in the presence of several kinds of transcription factors [8]. As a result, the effect of a transcription factor is enhanced due to the spatial proximity of the targets. This phenomenon is reminiscent of the local concentration effect that has been formalized by Müller-Hill [9] a decade ago. *Local concentration* simply means that the interaction between molecules that are able to interact with each other is all the more efficient when molecules are close to each other. This straightforward statement is crucial to understand genome organization because genomes seem to have evolved in order to optimize the proximity of reactive groups [9, 8, 10].

In this article, we propose a simple scheme in order to include the notion of distances into genetic regulatory networks (GRN) and to study their effect upon the dynamics of the network. Our approach is based on the discrete modeling of genetic regulatory networks that has been introduced by René Thomas. In this type of models, a discrete concentration level, called expression level, is associated to each gene of the GRN. This abstracts the continuous concentration of the protein that is coded by the gene. Within this scope, the notion of distance will help us to solve dilemma and conflicts during the search for logical parameters that define the dynamics of regulatory networks.

Within the scope of spatial information, two aspects will be discussed in this article: *privileged interaction* and *cluster of genes*.

- *The notion of cluster* expresses the notion of co-regulation, that is a set of spatially closed genes that are expressed at the same time due to the expression of a single regulating gene (*i.e.* the presence of a single transcription factor).
- *the notion of privileged interaction* is an ubiquitous concept in biology. For instance, specific interactions (*e.g.* between a transcription factor and DNA) in contrast to non-specific interactions, local concentration phenomena, short distances between a gene (or a protein) and the genes it regulate are examples of privileged interactions that maintain the good operation of cells.

In the case of spatial information, privileged interactions are mainly based on the notion of gene neighbourhood, which measures the proximity of two given genes modulating the interaction force between the two genes. If all genes are close to each other, or conversely are far away from each other, then the resulting dynamics is identical to the classical approach of Renée Thomas. In contrast, the notion of distance will allow us to reduce, eventually to one, the number of dynamics to consider.

We derive results for Boolean GRNs, where genes have only two expression levels, and multivalued GRNs where genes have a finite number of expression levels. In particular, we study two aspects: i) the static part of the model which describes the interaction occurring in the system, interaction graphs being a classical representation of this aspect , and ii) the dynamics resulting from these interactions. We illustrate our results on three classical models of GRN: the mucus production in Pseudomonas aeruginosa, the lytic/lysogenic switch in the lambda phage and the functioning of the operon lactose.

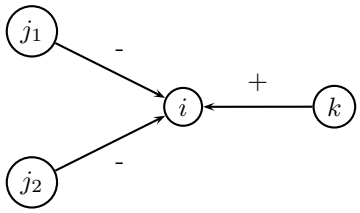


Fig. 1: Example of interaction graph

2 Interaction graph

Interaction graphs are a classical discrete representation of the static part of a genetic regulatory network. It is common to Boolean and multivalued GRN. The GRN is represented by an oriented graph where nodes abstract the proteins or genes which play a role in the system and edges abstract the known interactions inside the considered system. An interaction $(a \to b)$ can be either an activation or an inhibition, which will imply different behaviours considering the dynamics: in an *activation*, the increase of the expression level of a leads to an increase of the expression level of b, the edge is labelled by the sign $+$ and a is an activator of b; in an *inhibition*, the increase of a leads to a decrease of b, the edge is labelled by the sign $-$ and a is an inhibitor of b.

Definition 1 (Interaction graph)
An interaction graph *is a labeled directed graph* $G = (V, E, S)$ *where :*

1. V *is a finite set whose elements are called* variables.
2. $E \subseteq V \times V$ *is the set of* interactions. *For any* $i \in V$, V_i^- *denotes the set of predecessors of* i, *that is elements of* V *which have an action on* i *and* V_i^+ *denotes the set of successors of* i, *that is elements of* V *on which* i *has an action:*

$$V^-(i) = \{j | j \in V, (j,i) \in E\} \qquad V^+(i) = \{j | j \in V, (i,j) \in E\}$$

3. $S : E \to \{+, -\}$ *associates to each interaction its* sign. *An interaction can be either an* activation *(+ sign) or an* inhibition *(− sign).*

Definition 2 (Activators and inhibitors)
Let $G = (V, E, S)$ *be an interaction graph, and let* $i \in V$ *be a gene. We denote by* $A(i)$ *(resp.* $I(i)$*) the set of* activators *(resp.* inhibitors*) of* i*:*

$$A(i) = \{j | j \in V^-(i), S(j,i) = +\} \qquad I(i) = \{j | j \in V^-(i), S(j,i) = -\}$$

Example 1 (Example of interaction graph)
Let us exemplify definition 1 with the toy interaction graph from figure 1 where a gene i *is inhibited by* j_1 *and* j_2 *and activated by* k. *The sign of each interaction is directly expressed by labelling edges.*

3 Dynamics

The *dynamics of a GRN* consists in the evolution of each gene expression level step by step. Two kinds of dynamics are usually considered: either Boolean (genes have only two levels of expression) or multivalued (genes have several levels of expression). In both cases, several dynamics can be associated to an interaction graph, and the main problem is to reduce the number of dynamics we have to consider.

In reality, the evolution of a given gene expression level does not depend on all the genes of the GRN, but only on the genes which have an action on the given gene, that is its predecessors. More precisely, not all the predecessors of a given gene have an effect on its expression level, but only the predecessors *with a sufficient expression level*, the interaction is then said to be *effective*. The notion of sufficient expression level is detailed in the following, considering we are in Boolean or multivalued dynamics. The function giving the evolution of a gene considering its effective predecessors is called *logical parameters*.

3.1 Evolution function in Boolean and multivalued dynamics

3.1.1 Boolean dynamic states

In *Boolean dynamics*, genes have only two level of expression corresponding to a low concentration denoted 0, or to an high concentration denoted 1. Hence, an interaction $a \to b$ is effective if and only if the level of expression of a is high, *i.e.* equal to 1. The knowledge of the expression levels of all the genes define a *Boolean dynamic state*.

Definition 3 (Boolean dynamic states)
Let $G = (V, E, S)$ be a GRN and let $i \in V$ be a gene. We denote by $\mathbb{X}_b(G)$ the set of boolean dynamic states of G[1]*: $\mathbb{X}_b(G) = \{0, 1\}^{|V|}$.*

For $x = (x_1, ..., x_{|V|}) \in \mathbb{X}_b(G)$, $x_i \in \{0, 1\}$ is the expression level of gene i.

Example 2 (Boolean dynamic states of the interaction graph from fig. 1)
Interaction graph from fig. 1 is composed of four genes. Thus, there are 16 possible dynamic states which are the elements of $\{0, 1\}^4$. For example, the dynamic state $(x_i = 1, x_{j_1} = 1, x_{j_2} = 0, x_k = 0)$.

3.1.2 Multivalued dynamic states

Thresholds. In *multivalued dynamics*, genes have several possible level of expression, and thus the notion of "sufficient expression level" is more complex because the dynamics depends on a set of parameters called *threshold parameters*. When a gene i acts on several targets, on j and k for example, it is often known that the level of i mandatory for an action on j to be effective is higher than the level necessary for the action of i on k.

Definition 4 (Thresholds parameters)
Given a GRN ($G = V, E, S$), thresholds parameters *are represented by a function $T : E \Rightarrow \mathbb{N}^*$ which associated to each interaction of a GRN its threshold. T is such that such that*

$$\forall (i, j) \in E, T(i, j) \neq 1 \Leftrightarrow \exists k \in E : T(i, k) = T(i, j) - 1$$

In other word, if an interaction outgoing from a variable i is labelled by a threshold α greater than 2, then there exist interactions outgoing from i labelled by $1, \ldots, \alpha - 1$. This well represents the qualitative nature of thresholds in GRN, and an interaction (j, i) will be effective if and only if the expression level of j is above the threshold of (j, i). Obviously, several threshold parameters can be associated to a single interaction graph.

[1] Let us recall that $|V|$ denotes the number of elements in the set V.

Dynamic states. The multivalued dynamics states of a GRN depends on both the interaction graph and the threshold parameters. Indeed, a gene can take as many values as the greatest outgoing threshold.

Definition 5 (Multivalued dynamic states)
Let $G = (V, E, S)$ be a GRN, and let T be a set of threshold parameters associated to G. We denote for all $i \in V$ $b_i = max\{T(i,j) | j \in V^+(i)\}$). The set of possible level of expression for a gene i is $\mathbb{X}_i(G,T) = \{0, 1, ..., b_i\}$.

We denote by $\mathbb{X}_m(G,T)$ the set of multivalued dynamic states of G, associated to T: $\mathbb{X}_m(G,T) = \prod_{i \in V} \mathbb{X}_i(G,T)$.

In the following, many definitions refer either to multivalued or to Boolean dynamics states. If the definition is identical in both cases, we refer to dynamic states (either Boolean or multivalued) using the notation $\mathbb{X}$ (instead of $\mathbb{X}_b(G)$ or $\mathbb{X}_m(G,T)$).

3.1.3 Evolution function

Given a GRN G, and $\mathbb{X}$ a set of *dynamic states* (either Boolean or multivalued), then, each dynamics associated to a GRN can can be modelled by an evolution function $f = (f_1, ..., f_n) : \mathbb{X} \to \mathbb{X}$ (with n the number of genes). Each application f_i gives the evolution of gene i considering the dynamic state $x = (x_1, ..., x_n)$ of the GRN G:

- if $x_i < f_i(x)$ then the level of i is increasing,
- if $x_i > f_i(x)$ then the level of i is decreasing,
- if $x_i = f_i(x)$ then the level of i is stable.

The dynamics defined by a evolution function f is commonly represented by the *transition graph* of f, that is a graph where nodes are elements of $\mathbb{X}$ and there is an edge from x to x' iff $f(x) = x'$.

3.2 Effective interactions and logical parameters

Considering a dynamic state, the evolution of a given gene expression level does not depend on all the genes of the GRN, but only on the predecessors *with a sufficient expression level*, and called *effective predecessors.*

3.2.1 Boolean effective predecessors

In Boolean dynamics, only genes with an expression level equal to 1 may influence other genes.

Definition 6 (Boolean effective activators and inhibitors)
Let $G = (V, E, S)$ be a interaction graph, and let $i \in V$ be a gene. Given a dynamic state $x \in \mathbb{X}_b(G)$, We denote by $A_b^*(i,x)$ (resp. $I_b^*(i,x)$) the set of Boolean effective activators *(resp.* Boolean effective inhibitors*) of i:*

$$A_b^*(i,x) = \{j | j \in V^-(i), S(j,i) = +, x_j = 1\}$$

$$I_b^*(i,x) = \{j | j \in V^-(i), S(j,i) = +, x_j = 1\}$$

We denote by $w_b^*(i,x)$ the set of effective predecessors of i:

$$w_b^*(i,x) = A_b^*(i,x) \cup I_b^*(i,x)$$

For every gene j in $w_b^*(i,x)$, the interaction $(j,i) \in E$ is said to be effective.

3.2.2 Multivalued effective predecessors

The evolution function in multivalued dynamics depends on whether or not the concentrations of genes are under or above their thresholds.

Definition 7 (Multivalued effective activators and inhibitors)
Let $G = (V, E, S)$ be a interaction graph, and let T be a set of associated threshold parameters. Let $i \in V$ be a gene. Given a dynamic state x, We denote by $A^*_m(i, x)$ (resp. $I^*_m(i, x)$) the set of multivalued effective activators *(resp.* multivalued effective inhibitors*) of i:*

$$A^*_m(i, x) = \{j | j \in V^-(i), S(j, i) = +, x_j \geq T(j, i)\}$$

$$I^*_m(i, x) = \{j | j \in V^-(i), S(j, i) = -, x_j \geq T(j, i)\}$$

We denote by $w^*_m(i, x)$ the set of multivalued effective predecessors of i:

$$w^*_m(i, x) = A^*_m(i, x) \cup I^*_m(i, x)$$

For every gene j in $w^*_m(i, x)$, the interaction $(j, i) \in E$ is said to be effective.

In the following, many definitions refer either to multivalued or to Boolean effective predecessors. If the definition is identical in both cases, we refer to effective predecessors (either Boolean or multivalued) using the notation w^* (instead of w^*_b or w^*_m).

3.2.3 Logical parameters

Given a dynamic state x, the evolution function of a gene i does not depend on x but on $w^*(i, x)$ (either $w^*_b(i, x)$ or $w^*_m(i, x)$). This allow us to define the *logical parameters* which totally determine the behaviours of the evolution function, and then give a possible dynamics for the GRN.

Definition 8 (Logical parameters)
Let $G = (V, E, S)$ be a interaction graph. Given $f : \mathbb{X} \to \mathbb{X}$ an evolution function for G., we may define for each gene $i \in V$, a set of logical parameters *K_i such that*

- *in Boolean dynamics: $f_i(x) = K_i(w^*_b(i, x))$ and $K_i : 2^{V_i^-} \to \{0, 1\}$;*
- *in multivalued dynamics: $f_i(x) = K_i(w^*_m(i, x))$ and $K_i : 2^{V_i^-} \to \{0, \ldots, b_i\}$.*

Obviously, several different multivalued dynamic states could have the same sets of effective activators or inhibitors. In the following, we usually refer to state by only giving the set of effective activators or inhibitors. Thus, for example, considering a particular gene i, a set $\{j_1, j_2\}$ of effective predecessors of i will refer to any dynamic state x, such that $x_{j_1} \geq T(j_1, i)$ and $x_{j_2} \geq T(j_2, i)$.

Example 3 (Logical parameters for fig. 1)
The gene i has three predecessors which admits two levels of expression. Thus, there is 8 logical parameters K_i to consider: $K_i(\emptyset)$, $K_i(\{j_1\})$, $K_i(\{j_2\})$, $K_i(\{k\})$, $K_i(\{j_1, j_2\})$, $K_i(\{j_1, k\})$, $K_i(\{j_2, k\})$ and $K_i(\{j_1, j_2, k\})$.

To each instantiation of logical parameters corresponds an evolution function defining a possible dynamics for the GRN. We are interested in researching constraints on these logical parameters in order to reduce their possible values, and then to reduce the number of dynamics we have to consider.

Finally, determining the dynamics of a GRN consists in the attribution of values to the different logical parameters. The number of these parameters is huge: given a gene i, there is $2^{|V^-(i)|}$ logical parameters K_i, and each parameter can take at least

two values (in Boolean dynamics). Thus, we have to consider $\prod_{i \in V} 2^{2^{|V^-(i)|}}$ possible logical parameters. For example, just for the interaction graph from fig. 1 we have to consider $2^{2^3} = 256$ possible attributions for the logical parameters of i.

Nevertheless, the attribution of values to logical parameters follows certain rules, linked to the interaction graph, and to the type of interaction (activation or inhibition). These rules are detailed in section 3.3.

3.3 Rules based on interaction graph

The rules presented here are based on the interaction graph. Logical parameters which respect these three rules are said to be *valid*.

3.3.1 Activation/Inhibition rule

This rule is based on the definition of activation or inhibition. If a gene j activates a gene i, we cannot be certain that the increase of j expression level lead to an increase for i, but, it is certain that it cannot lead to a decrease of i expression level.

Definition 9 (Activation/inihibition rule)
Given a GRN $G = (V, E, S)$, and i, j in V two genes such $(j, i) \in E$ (j is a predecessor of i), then:

- $S(j,i) = + \Rightarrow \forall \omega \subseteq V^-(i), K_i(\omega) \leq K_i(\omega \cup \{j\})$
- $S(j,i) = - \Rightarrow \forall \omega \subseteq V^-(i), K_i(\omega) \geq K_i(\omega \cup \{j\})$

3.3.2 Observation rule

This rule expresses how we identify that a predecessor is an activator or an ihibitor. If j is an activator of i, then it exists at least one dynamics state where the high expression level of j leads to a an increase of the expression level of i. In other word, at least one of the previous inequality must be strict.

Definition 10 (Observation rule)
Given a GRN $G = (V, E, S)$, and i, j in V two genes such $(j, i) \in E$ (j is a predecessor of i), then:

- $S(j,i) = + \Rightarrow \exists \omega \subseteq V^-(i), K_i(\omega) < K_i(\omega \cup \{j\})$
- $S(j,i) = - \Rightarrow \exists \omega \subseteq V^-(i), K_i(\omega) > K_i(\omega \cup \{j\})$

3.3.3 Maximum rule

This rule expressed that in a dynamic state where all the activators of a gene are effective and simultaneously none of the inhibitor is effective, then the associated logical parameter is maximum (that is equal to 1 in Boolean dynamics, or b_i in multivalued dynamics). Conversely, if none of the activator is effective, and all inhibitors are, then the logical parameter is minimum, that is equal to 0.

Definition 11 (Maximum rule for Boolean dynamics)
Let $G = (V, E, S)$ be a GRN, and let i in V be a gene. Let x be a Boolean dynamic state. We have:

$$K_i(A(i,x)) = 1 \qquad K_i(I(i,x)) = 0$$

Definition 12 (Maximum rule for multivalued dynamics)
Let $G = (V, E, S)$ be a GRN and T be a set of threshold parameters. Let i in V be a gene. Let x be a multivalued dynamic state. We have:

$$K_i(A(i,x)) = b_i = max\{T(i,j) | (i,j) \in E\} \qquad K_i(I(i,x)) = 0$$

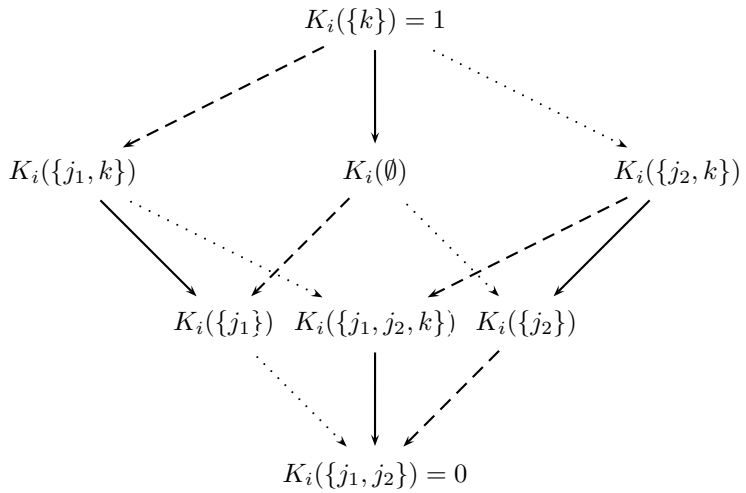


Fig. 2: Relation among logical parameters of the interaction graph from fig. 1.

3.3.4 Example

Example 4 (Valid parameters for fig. 1)
The Activation/Inhibition rule imposes the following inequalities to be respected.

- *k is an activator of i:*

$$K_i(\emptyset) \leq K_i(\{k\}) \qquad K_i(\{j_1\}) \leq K_i(\{j_1, k\})$$
$$K_i(\{j_2\}) \leq K_i(\{j_2, k\}) \qquad K_i(\{j_1, j_2\}) \leq K_i(\{j_1, j_2, k\})$$

- *j_1 is an inhibitor of i:*

$$K_i(\emptyset) \geq K_i(\{j_1\}) \qquad K_i(\{j_2\}) \geq K_i(\{j_1, j_2\})$$
$$K_i(\{k\}) \geq K_i(\{j_1, k\}) \qquad K_i(\{j_2, k\}) \geq K_i(\{j_1, j_2, k\})$$

- *j_2 is an inhibitor of i:*

$$K_i(\emptyset) \geq K_i(\{j_2\}) \qquad K_i(\{j_1\}) \geq K_i(\{j_1, j_2\})$$
$$K_i(\{k\}) \geq K_i(\{j_2, k\}) \qquad K_i(\{j_1, k\}) \geq K_i(\{j_1, j_2, k\})$$

The observation rule imposes that in each previous point, at least one of the inequalities is strict.

Finally, the maximum rule imposes that $K_i(\{k\}) = 1$ and $K_i(\{j_1, j_2\}) = 0$.

This can be resumed in the graph from figure 4, where an arrow from i to j mean $i \geq j$, and this inequality is strict for at least one arrow of each type (plain, dashed or doted arrow).

4 Constraints based on spatial information

4.1 Conflicts and dilemma

Despite the above rules, possible dynamics of a real genetic regulatory network are often too numerous. Indeed, for one interaction graph, several logical parameters are

valid, that is several dynamics exists. In fact we can identify the situations which lead to the existence of different logical parameters, that is dynamics states where the three above rules do not allow us to determine an unique values for logical parameters:

- *Conflicts* occur when a gene is simultaneously activated and inhibited,
- *Dilemma* occur when all the activators (resp. inhibitors) of a gene are not effective.

Definition 13 (Conflicts and dilemma)
Let $G = (V, E, S)$ be an interaction graph, $i \in V$ be a gene and $x \in \mathbb{X}$ be a dynamic state.

- *x is a* situation of conflict *for gene i if, and only if, $A^*(i,x) \neq \emptyset$ and $I^*(i,x) \neq \emptyset$*
- *x is a* situation of dilemma *for gene i if, and only if, ($A^*(i,x) \neq \emptyset$ and $A^*(i,x) \neq A(i)$) or ($I^*(i,x) \neq \emptyset$ and $I^*(i,x) \neq I(i)$)*

In the following, we will focus on the determination of logical parameters. Thus, conflict and dilemma will refer to parameters, that is $K_i(w^*(i,x))$ is a conflict (resp. a dilemma) if and only if x is a situation of conflict (resp. dilemma) for gene i. In other words, if $w^*(i,x) = \omega$:

- $K_i(\omega)$ is a conflict iff $\omega \cap A(i) \neq \emptyset$ and $\omega \cap I(i) \neq \emptyset$,
- $K_i(\omega)$ is a dilemma iff $A(i) \not\subset \omega \not\subset I(i)$ or $I(i) \not\subset \omega \not\subset A(i)$.

In this model, $K_i(\emptyset)$ is neither a conflict nor a dilemma, but corresponds to the basal situation, where a gene i is not activated or inhibited.

Example 5 (Conflicts and dilemma in fig. 1)
Let us consider the 8 possible dynamic states, and the associated logical parameters for gene i for the interaction graph from fig. 1:

- *$K_i(\{j_1\})$ and $K_i(\{j_2\})$ are dilemma;*
- *$K_i(\{j_1, j_2, k\})$ is a conflict;*
- *$K_i(\{j_1, k\})$, $K_i(\{j_2, k\})$ are both conflict and dilemma.*

Note that $K_i(\{k\})$ and $K_i(\{j_1, j_2\})$ are neither conflict nor dilemma: the former correspond to a situation where i is fully activated and is not inhibited, the latter corresponds to the reverse situation.

In the next section we will introduce some spatial information within GRN. This information which enable us to solve some situations of conflicts and dilemma, and then restrict the dynamics of genetic regulatory networks.

4.2 Spatial information

Spatial information is used to help us to solve dilemma and conflicts during the search for logical parameters that define the dynamics of regulatory networks. This notion is captured through two notions:

The spatial information we add in GRN is based on two concepts:

- The notion of cluster expresses the notion of co-regulation, that is a set of spatially closed genes that are expressed at the same time due to the expression of a single regulating gene (i.e. the presence of a single transcription factor).
- The notion of distances between a gene (or a protein) and the genes it regulate is modelled through privileged interactions. Privileged interactions are an ubiquitous concept in biology, more general that the notion of distance. For instance, specific interactions (e.g. between a transcription factor and DNA) in contrast to non-specific interactions, local concentration phenomena are examples of privileged interactions.

4.2.1 Privileged interactions and neighbourhood relation

In the case of spatial information, privileged interactions are mainly based on the notion of gene neighbourhood, which measures the proximity of two given genes modulating the interaction force between the two genes. If all genes are close to each other, or conversely are far away from each other, then the resulting dynamics is identical to the classical approach of René Thomas. In contrast, the notion of distance as privileged interactions will allow us to reduce, eventually to one, the number of dynamics to consider.

Definition 14 (GRN with privileged interactions)
A GRN with privileged interactions *is a tuple $G = (V, E, S, P)$ such that (V, E, S) is a GRN, and P is a subset of E denoting the set of privileged interactions.*

By abuse of notations, given a gene $i \in V$, we denote by $P(i)$ the set of privileged predecessors of i: $P(i) = \{j | (j, i) \in P\}$

Obviously, all definitions used for GRN are extended to GRN with privileged interactions.

GRN are classically represented by a interaction graph. The privileged interactions can be easily representing in the graph, using dashed arrow for the elements of P and plain arrows for the elements of $E - P$ (see section 4.3.1 for example).

4.2.2 Clusters

Clusters represent groups of genes which are simultaneously activated or inhibited by a same gene. Formally, to each gene we associate a partition of its target genes.

Definition 15 (Clusters)
Let $G = (V, E, S)$ be a GRN. A cluster over G is a function $C : V \to 2^V$ such that for all i in V, $C(i) = \{C_i^1, ..., C_i^{p_i}\}$ is a partition over $V^+(i)$, that is:

- $\cup_{k=1}^{p_i} C_i^k = V^+(i)$
- *for all k, k', $k \neq k' \Rightarrow C_i^k \cap C_i^{k'} = \emptyset$*

4.3 Constraints based on privileged interactions

4.3.1 Influence of privileged interactions

The main idea of the privileged interactions is that their "force" of interaction is higher than the force of non privileged interactions. Figure 3 presents the idea of how using the privileged interactions, representing by dashed lines, to solve conflict in case of Boolean dynamics.

Direct influence. The main idea is that if two interactions occurs simultaneously, then the privileged one is preferred. The following constraint is an extension of the maximum rule. The maximum rule imposes that if all the activators of a gene are effective, and none of the inhibitors is, then the expression level of that gene is maximum. The constraint indicates that if none of privileged activators are effective, and some privileged inhibitors of the considered gene are effective, then the expression level cannot be maximum.

Note that in Boolean dynamics, the expression level of a gene is either 0 or maximum. Then saying that a gene expression level is not maximum is equivalent to giving it a value 0.

Definition 16 (Direct influence of privileged interactions in Boolean dynamics)
Let $G = (V, E, S, P)$ be a GRN with privileged interactions. Let $i \in V$ be a gene and $x \in \mathbb{X}_b(G)$ a Boolean dynamic state. Then

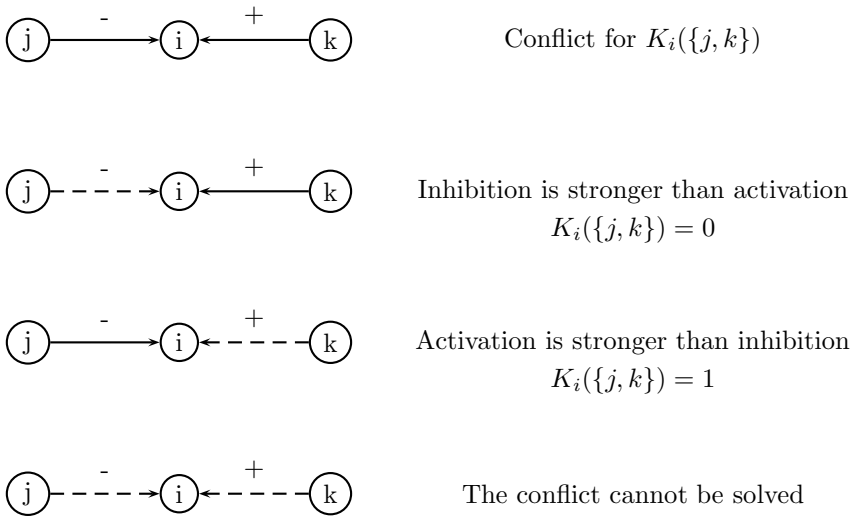


Fig. 3: Idea of how solving conflicts with privileged interactions

- *If* $A_b^*(i,x) \cap P(i) \neq \emptyset$ *and* $I_b^*(i,x) \cap P(i) = \emptyset$ *then,* $K_i(w_b^*(i,x)) = 1$.
- *If* $I_b^*(i,x) \cap P(i) \neq \emptyset$ *and* $A_b^*(i,x) \cap P(i) = \emptyset$ *then,* $K_i(w_b^*(i,x)) = 0$.

Definition 17 (Direct influence of privileged interactions in multivalued dynamics) *Let* $G = (V, E, S)$ *be a GRN and* T *an associated threshold funtion. Let* N *be a neighbourhood relation associated to* G. *Let* $i \in V$ *be a gene and* $x \in \mathbb{X}_m(G)$ *a multivalued dynamic state. Then*

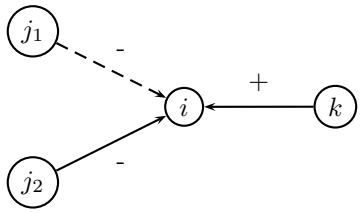
- *If* $A_m^*(i,x) \cap P(i) \neq \emptyset$ *and* $I_m^*(i,x) \cap P(i) = \emptyset$ *then,* $K_i(w_m^*(i,x)) > 0$.
- *If* $I_m^*(i,x) \cap P(i) \neq \emptyset$ *and* $A_m^*(i,x) \cap P(i) = \emptyset$ *then,* $K_i(w_m^*(i,x)) < b_i = max\{T(i,j) | (i,j) \in E\}$.

Relative influence. The level of expression of non privileged predecessors is not important compared to the presence of absence of privileged ones. In other words, the value of a logical parameter for a set of effective genes, whose at least one is a privileged predecessor, remains the same whatever non privileged predecessors becoming effective.

Definition 18 (Relative influence of privileged interactions in Boolean dynamics) *Let* $G = (V, E, S, P)$ *be a GRN with privileged interactions. Let* $i \in V$ *be a gene and* $\omega \subseteq V^-(i)$ *a set a predecessor of* i *such that* $\omega \cap P(i) \neq \emptyset$. *Let* $j \in V^-(i)$ *such that* $j \notin P(i)$ *(*j *is a a non privileged predecessor of* i*). Then*

- *If* $K_i(\omega) = 0$ *then* $K_i(\omega \cup \{j\}) = 0$.
- *If* $K_i(\omega) = 1$ *then* $K_i(\omega \cup \{j\}) = 1$.

Definition 19 (Relative influence of privileged interactions in multivalued dynamics) *Let* $G = (V, E, S, P)$ *be a GRN with privileged interactions, and* T *an associated threshold function. Let* $i \in V$ *be a gene and* $\omega \subseteq V^-(i)$ *a set a predecessor of* i *such that* $\omega \cap P(i) \neq \emptyset$. *Let* $j \in V^-(i)$ *such that* $j \notin P(i)$ *(*j *is a non privileged predecessor of* i*). Then*



j_1	j_2	k	Evolution	
0	0	0	$K_i(\emptyset)$	
0	0	1	$K_i(\{k\})$	$= 1$
0	1	0	$K_i(\{j_2\})$	$=$ Dilemma
0	1	1	$K_i(\{j_2, k\})$	$=$ Dilemma + conflict
1	0	0	$K_i(\{j_1\})$	$= 0$ (j_1 effective, no more dilemma)
1	0	1	$K_i(\{j_1, k\})$	$= 0$ (j_1 effective, no more dilemma nor conflict)
1	1	0	$K_i(\{j_1, j_2\})$	$= 0$
1	1	1	$K_i(\{j_1, j_2, k\})$	$= 0$ (j_1 effective, no more conflict)

Fig. 4: Logical parameters when one inhibitor is the only privileged predecessor.

- *If* $K_i(\omega) < b_i$ *then* $K_i(\omega \cup \{j\}) < b_i$.
- *If* $K_i(\omega) > 0$ *then* $K_i(\omega \cup \{j\}) > 0$.

4.3.2 Influence of privileged interactions in the interaction graph from fig. 1

We study here the different possible privileged interactions for the interaction graph from fig. 1.

- Figure 4 present the case where one inhibitor is the only privileged predecessor. In that case, as soon as the concentration of the privileged gene is under the threshold, conflict and dilemma appears between other genes. When the concentration is above the threshold, conflict and dilemma are solved.
- Figure 5 present the case where all inhibitors are the privileged predecessors. There is no conflict and the dynamics is unique.
- Figure 6 present the case where all activators are the privileged predecessors. There is no conflict, but some dilemma remains.

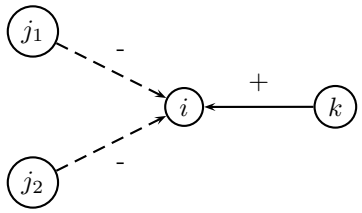
4.3.3 Toward an unique Boolean dynamics

Necessary condition. A necessary condition for a GRN to have an unique dynamics is "the set of privileged predecessors is either included in the set of activators or inhibitors".

Theorem 1 (Necessary condition to a unique dynamic)
A GRN with privileged interactions (V, E, S, P) *has a unique dynamics only if the following condition is true: for all* $i \in V$*, we have*

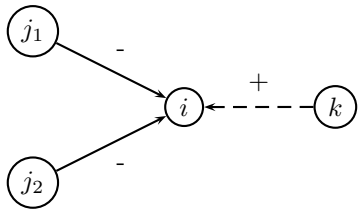
$$P(i) \subseteq A(i) \text{ or } P(i) \subseteq I(i)$$

Proof 1 *Without this condition, there exist one activator and one inhibitor which are privileged predecessors of a given gene, and when the concentration of theses two gene is above their respective threshold, a conflict occurs.*



j_1	j_2	k	Evolution	
0	0	0	$K_i(\emptyset)$	
0	0	1	$K_i(\{k\})$	$= 1$
0	1	0	$K_i(\{j_2\})$	$= 0$ (j_2 effective, no more dilemma)
0	1	1	$K_i(\{j_2, k\})$	$= 0$ (j_2 effective, no more dilemma nor conflict)
1	0	0	$K_i(\{j_1\})$	$= 0$ (j_1 effective, no more dilemma)
1	0	1	$K_i(\{j_1, k\})$	$= 0$ (j_1 effective, no more dilemma nor conflict)
1	1	0	$K_i(\{j_1, j_2\})$	$= 0$
1	1	1	$K_i(\{j_1, j_2, k\})$	$= 0$ (j_1 and j_2 effective, no more conflict)

Fig. 5: Logical parameters when all inhibitors are privileged predecessors.



j_1	j_2	k	Evolution	
0	0	0	$K_i(\emptyset)$	
0	0	1	$K_i(\{k\})$	$= 1$
0	1	0	$K_i(\{j_2\})$	$=$ Dilemma
0	1	1	$K_i(\{j_2, k\})$	$= 1$ (k effective, no more dilemma nor conflict)
1	0	0	$K_i(\{j_1\})$	$=$ Dilemma
1	0	1	$K_i(\{j_1, k\})$	$= 1$ (k effective, no more dilemma nor conflict)
1	1	0	$K_i(\{j_1, j_2\})$	$= 0$
1	1	1	$K_i(\{j_1, j_2, k\})$	$= 1$ (k effective, no more conflict)

Fig. 6: Logical parameters when all activators are privileged predecessors.

This condition is not a sufficient one, because when the concentration of privileged predecessors is under their threshold, conflict and dilemma may occur between other genes.

Necessary and sufficient condition for no conflict. A necessary and sufficient condition to have no conflict is given by the equality in previous condition. In other words, there is no conflict if the set of privileged predecessors is either equal to activators or inhibitors.

Theorem 2 (Necessary and sufficient condition to a no conflict situation) *A GRN with privileged interactions (V, E, S, P) has no situation of conflict if the following condition is true: for all $i \in V$, we have*

$$P(i) = A(i) \text{ or } P(i) = I(i)$$

Proof 2 *In this case, it is clear that we cannot have any conflict since the privileged predecessors have the same type of interaction. Thus, the condition is sufficient.*

If the condition is not respected, two cases have to be considered:

- $P(i) \subsetneq A(i)$ *or* $P(i) \subsetneq I(i)$
- $P(i) \cap A(i) \neq \emptyset$ *and* $P(i) \cap I(i) \neq \emptyset$

In both cases, it is easy to find conflict. Thus, the condition is necessary

Nevertheless, if the concentration of all privileged predecessors is under thresholds, then a situation of dilemma may occur.

Necessary and sufficient condition for no dilemma. Dilemma occur when two genes having the same action (either activation or inhibition) are not effective simultaneously. Thus, a necessary and sufficient condition to have no dilemma is that either there is only one gene for a given action, or each predecessor having this action is a privileged predecessor of the target.

Theorem 3 (Necessary and sufficient condition to a no dilemma situation) *A GRN with privileged interactions (V, E, S, P) has no situation of dilemma iff the two following conditions are true: for all $i \in V$, we have*

- $A(i) \subseteq P(i)$ *or* $|A(i)| = 1$
- $I(i) \subseteq P(i)$ *or* $|I(i)| = 1$

Proof 3 *The proof is let to the reader.*

Necessary and sufficient condition for a unique Boolean dynamics. The dynamics is unique iff there is no dilemma nor conflict, which directly lead to the following necessary and sufficient condition.

Theorem 4 (Necessary and sufficient condition to a no dilemma situation) *A GRN with privileged interactions (V, E, S, P) has a unique dynamics iff for all $i \in V$ one the following condition is true:*

- $A(i) = P(i)$ *and* $|I(i)| = 1$
- $|A(i)| = 1$ *and* $I(i) = P(i)$

Proof 4 *The theorem is a direct consequence of theorems 2 and 3.*

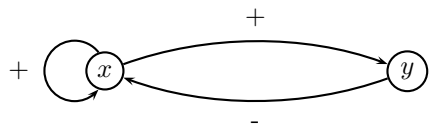


Fig. 7: Interaction graph for the mucus production system in *P. aeruginosa*

4.4 Constraints based on clusters

Clusters information is used in multivalued dynamics. It influences the threshold function to consider. In other words, given a cluster we only consider the threshold function verifing that genes of a same cluster are simultaneously activated or inhibited, that is the threshold between the source and the target of the considered genes are the same.

Definition 20 (Cluster and threshold)
Let $G = (V, E, S)$ be a GRN, and C an associated cluster. Then the compatible threshold functions are such that: for all i in V, for all k, k' in $\{j | j \in V, (i, j) \in E\}$

$$\exists p \in \mathbb{N}, k \in C_i^p, k' \in C_i^p \Rightarrow T(i, k) = T(i, k')$$

5 Mucus production in *Pseudomonas aeruginosa*

Pseudomonas aeruginosa are bacteria that secrete mucus (alginate) in lungs affected by cystic fibrosis, but not in common environment. As this mucus increases respiratory defficiency, this phenomenon is a major cause of mortality. Details of the regulatory network associated with the mucus production by *Pseudomas aeruginosa* are described by Govan and Deretic [4] but a simplified genetic regulatory network has been proposed by Guespin and Kaufman [5], see Fig.7.

It has been observed that mucoid *P. aeruginosa* can continue to produce mucus isolated from infected lungs. It is commonly thought that the mucoid state of *P. aeruginosa* is due to a mutation which cancels the inhibition of gene x. An alternative hypothesis has been made: this mucoid state can occur by reason of an epigenetic modification, *i.e.* without mutation [5]. The models compatible with this hypothesis are constructed in [1]. We show here that distances are useful for reducing such a set of models.

5.1 Boolean dynamics

The logical parameters to consider are the following ones:

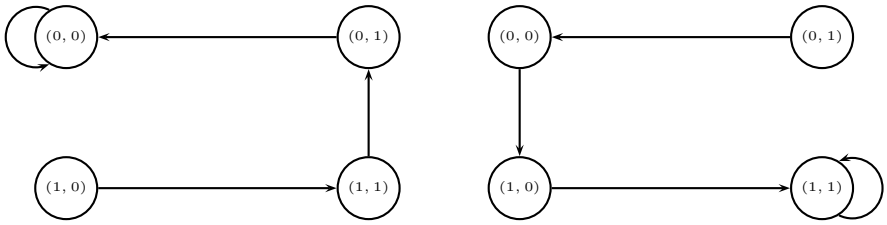
- $K_y(\emptyset)$ and $K_y(\{x\})$ for the gne y,
- $K_x(\emptyset)$, $K_x(\{x\})$, $K_x(\{y\})$ and $K_x(\{x, y\})$for gene x.

These parameters can take values 0 and 1, and thus, without any further consideration this leads to $2^2 \times 2^4 = 64$ possible dynamics. Obviously, this number is decreased considering the rules previously presented:

- For y: $K_y(\emptyset) = 0$ and $K_y(\{x\}) = 1$ due to the observation rule.
- For x:
  - the maximum rule leads to $K_x(\{x\}) = 1$ and $K_x(\{y\}) = 0$,
  - then the observation rule leads to two possible dynamics: either ($K_x(\emptyset) = 1$ and $K_x(\{x, y\}) = 1$) or ($K_x(\emptyset) = 0$ and $K_x(\{x, y\}) = 0$).

x	y	K_x	K_y
0	0	0	0
0	1	0	0
1	0	1	1
1	1	0	1

x	y	K_x	K_y
0	0	1	0
0	1	0	0
1	0	1	1
1	1	1	1



Inhibition stronger than activation Activation stronger than inhibition

Fig. 8: Valid dynamics for the interaction graph of mucus production in *P. aeruginosa*

The two possible dynamics are due to the conflict between x and y, and then the knowledge of privileged interactions among the activation of x by itself or the inhibition of x by y would lead to the determination of a unique dynamic.

- If both the interactions are privileged ones (or conversely are not privileged ones) then the two dynamics remains valid.
- If the inhibition is privileged and not the activation, then $K_x(\emptyset) = 0$ and $K_x(\{x, y\}) = 0$.
- If the activation is privileged and not the inhibition, then $K_x(\emptyset) = 1$ and $K_x(\{x, y\}) = 1$.

The two dynamics are represented on figure 8.

5.2 Multivalued dynamics

Logical parameters. In that case, gene x may have three level of expression, and thus the logical parameters to consider are the following ones:

- $K_y(\emptyset)$ and $K_y(\{x\})$ which can take value in $\{0, 1\}$
- $K_x(\emptyset)$, $K_x(\{x\})$, $K_x(\{y\})$ and $K_x(\{x, y\})$ which can take value in $\{0, 1, 2\}$

Without any considerations, these parameters lead to $2^2 \times 3^4 = 324$ different models for the dynamics.

The value of some logical parameters can be inferred with the rules previously described:

- For y: $K_y(\emptyset) = 0$ and $K_y(\{x\}) = 1$ due to the observation rule.
- For x:
  - the maximum rule lead to $K_x(\{x\}) = 2$ and $K_x(\{y\}) = 0$
  - the observation rule implies that we cannot have ($K_x(\{x, y\}) = 0$ and $K_x(\emptyset) = 2$) or ($K_x(\{x, y\}) = 2$ and $K_x(\emptyset) = 0$).

We finally have 7 different models for logical parameters.

Thresholds. In multivalued dynamics, we also have to consider the different possible threshold functions. The case of gene y is easily solved, because y has only one successor, and thus any valid threshold function T is such that $T(y, x) = 1$. For gene x, three threshold function are valid:

- $T(x, y) = 1$ and $T(x, x) = 1$, which is similar to Boolean dynamics;
- $T(x, y) = 1$ and $T(x, x) = 2$, the activation of y occurs before the activation of x;
- $T(x, y) = 2$ and $T(x, x) = 1$, the activation of y occurs after the activation of x.

Figures 9 and 10 give the seven possible dynamics considering the two last solutions.

With privileged interactions. Many dynamics exist due to the conflict between the activation of x by itself, and its inhibition by y. Privileged interactions between these two interactions may reduce the number of dynamics to consider. Obviously, if both of interactions are privileged, conversely if none of them are, then the 7 dynamics remains valid for each threshold function.

- If the activation of x by itself is the only privileged interaction, then $K_x(\{x, y\}) > 0$.
- If the inhibition of x by y is the only privileged interaction, then $K_x(\{x, y\}) < 2$.

In both cases, the knowledge of privileged interactions allows us to refute 4 models over the 14.

References

[1] G. Bernot, J.-P. Comet, A. Richard, and J. Guespin. Application of formal methods to biological regulatory networks: Extending Thomas' asynchronous logical approach with temporal logic. *Journal of Theoretical Biology*, 229(3):339–347, 2004.

[2] J. E. Cabrera and D. J. Jin. The distribution of rna polymerase in escherichia coli is dynamic and sensitive to environmental cues. *Mol Microbiol*, 50(5):1493–1505, 2003.

[3] A.-S. Carpentier, B. Torresani, A. Grossmann, and A. Henaut. Decoding the nucleoid organisation of bacillus subtilis and escherichia coli through gene expression data. *BMC Genomics*, 6(1):84, 2005.

[4] J. Govan and V. Deretic. Microbial pathogenesis in cystic fibrosis: mucoid pseudomonas aeruginosa and burkholderia cepacia. *Microbiol rev.*, 60(3):539–74, 1996.

[5] J. Guespin-Michel and M. Kaufman. Positive feedback circuits and adaptive regulations in bacteria. *Acta Biotheor.*, 49(4):207–218, 2001.

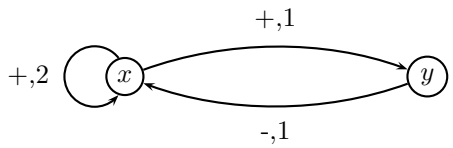
[6] D. A. Jackson, A. B. Hassan, R. J. Errington, and P. R. Cook. Visualization of focal sites of transcription whitin human nuclei. *J. Cell Biol.*, 164:515–526, 2004.

[7] F. Képès. Periodic transcriptional organization of the e.coli genome. *J Mol Biol*, 340(5):957–964, 2004.

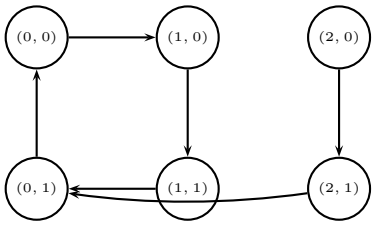
[8] F. Képès and C. Vaillant. Transcription-based solenoidal model of chromosomes. *Complexus*, 1(4):171–180, 2003.

[9] B. Muller-Hill. The function of auxiliary operators. *Mol Microbiol*, 29(1):13–18, 1998 Jul.

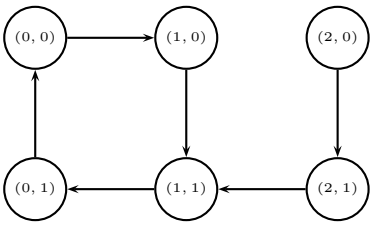
[10] T. Sexton, H. Schober, P. Fraser, and S. Gasser. Gene regulation through nuclear organization. *Nat Struct Mol Biol*, 14(11):1049–1055, 2007 Nov 5.



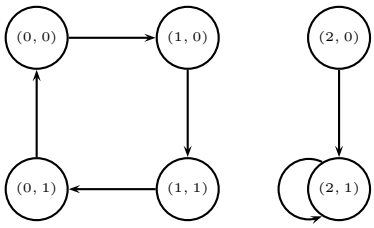
x	y	K_x	K_y
0	0	$K_x(\emptyset)$	0
0	1	0	0
1	0	$K_x(\emptyset)$	1
1	1	0	1
2	0	2	1
2	1	$K_x(\{x,y\})$	1



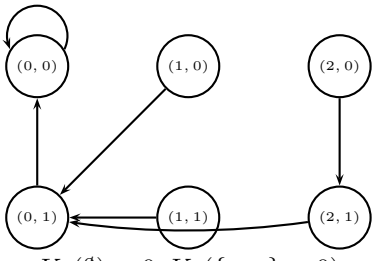
$K_x(\emptyset) = 1, K_x(\{x,y\} = 0)$



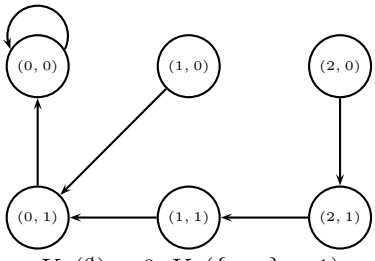
$K_x(\emptyset) = 1, K_x(\{x,y\} = 1)$



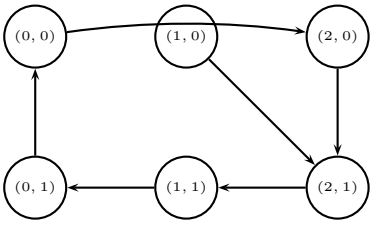
$K_x(\emptyset) = 1, K_x(\{x,y\} = 2)$



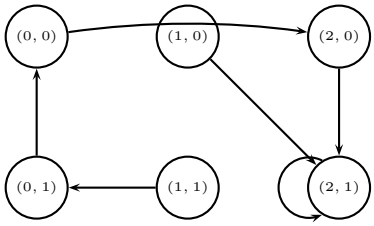
$K_x(\emptyset) = 0, K_x(\{x,y\} = 0)$



$K_x(\emptyset) = 0, K_x(\{x,y\} = 1)$

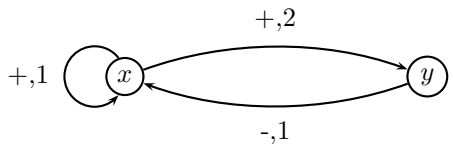


$K_x(\emptyset) = 2, K_x(\{x,y\} = 1)$

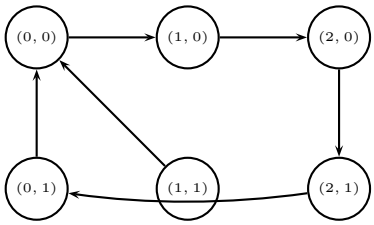


$K_x(\emptyset) = 2, K_x(\{x,y\} = 2)$

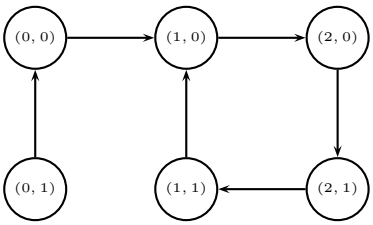
Fig. 9: Valid dynamics for the interaction graph of mucus production in *P. aeruginosa* when the activation of y occurs before the activation of x



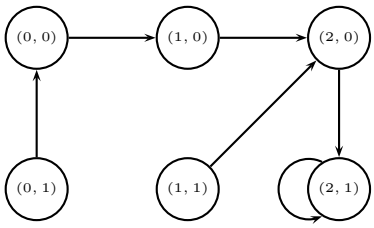
x	y	K_x	K_y
0	0	$K_x(\emptyset)$	0
0	1	0	0
1	0	2	0
1	1	$K_x(\{x,y\})$	0
2	0	2	1
2	1	$K_x(\{x,y\})$	1



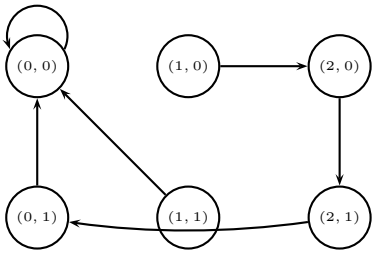
$K_x(\emptyset) = 1, K_x(\{x,y\} = 0)$



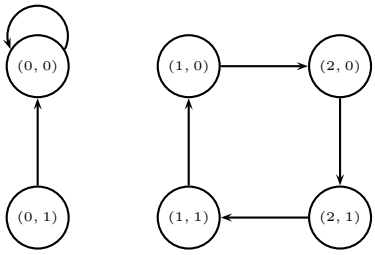
$K_x(\emptyset) = 1, K_x(\{x,y\} = 1)$



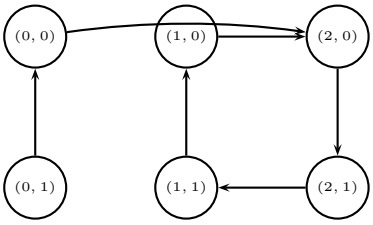
$K_x(\emptyset) = 1, K_x(\{x,y\} = 2)$



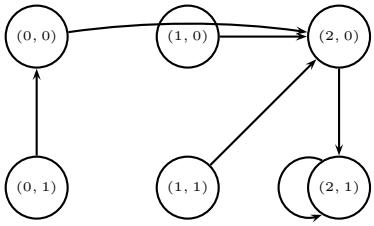
$K_x(\emptyset) = 0, K_x(\{x,y\} = 0)$



$K_x(\emptyset) = 0, K_x(\{x,y\} = 1)$



$K_x(\emptyset) = 2, K_x(\{x,y\} = 1)$



$K_x(\emptyset) = 2, K_x(\{x,y\} = 2)$

Fig. 10: Valid dynamics for the interaction graph of mucus production in *P. aeruginosa* when the activation of y occurs after the activation of x