Probabilistic Gene Network

Kristine Joy E. Carpio[1,3], Gilles Bernot[2],
Jean-Paul Comet[2] and Francine Diener[1]

[1] Laboratoire J.A. Dieudonné, Université de Nice - Sophia Antipolis, France
[2] CNRS UMR 7271, Laboratoire I3S, Université de Nice - Sophia Antipolis, France

Abstract

In this article we present a modelling framework that links the well known modelling framework of gene network introduced by R. Thomas and Markov chains. In a first development we introduce a Markov chain having as state space the set of all possible states of the R. Thomas models: we generate the transition probabilities by examining all the possible parameterizations of the interaction graph. The second development focuses on a stochastic framework where several parameterizations of a same qualitative gene interaction graph are considered and transition probabilities allow one to jump from a state to another one which can potentially be in another parameterized model. The idea is to consider only parameterized qualitative models of R. Thomas which abstract biological knowledge, and to use transition probabilities to allow to jump from one to another, if information coming from biological experiments reinforces the belief in a particular model.

1 Introduction

Regulatory networks are models based on graphs which are used to obtain a simpler view of gene regulation [6, p. 101]. Gene regulation is defined as the process of turning genes on and off which is made possible by a network of interactions that includes chemically modifying genes and using regulatory proteins. Gene regulation guarantees that appropriate genes are expressed at proper times specially during early development where cells begin to take on specific functions; it also helps an organism respond to its environment [8].

The different frameworks for modelling gene networks can be classified into three main groups. The systems of differential equations have been largely used in order to represent a lot of systems with a lot of details (transcription, traduction, transports ...). The second group consists of stochastic frameworks like Markov chains. The Markov modelling framework supposes that, given the past and the present, the future only depends on the present [9, p. 163].

[3]On research leave from De La Salle University, Philippines. Funded by Emma in the framework of the EU Erasmus Mundus Action 2.

This framework is well adapted to biological systems but supposes a strong effort in the enumeration of all entities (and interactions between them) that play a role in the system. The third group of approaches consists of qualitative frameworks in which details have been abstracted and only main causalities have been taken into account. Two paradigmatic frameworks can be classified in this group: the Boolean networks first introduced by Kauffman [5] and the multi-valuated modelling framework first introduced by R. Thomas [13].

In this paper we present a modelling framework that links the well known modelling framework of gene network introduced by R. Thomas and Markov chains. In a first development we introduce a Markov chain having as state space the set of all possible states of the R. Thomas models: we generate the transition probabilities by examining all the possible parameterizations of the interaction graph. Thus the Markov chain represents the possible behaviours obtained by superposition of all parameterized models. We then extend this stochastic framework to a Markov chain in which we distinguish the states of each parameterized model and where the probabilities are computed on a smaller set of parameterizations. The idea is to consider only parameterized qualitative models of R. Thomas which represent well the biological knowledge and to use transition probabilities to allow the system to jump from a particular dynamics to another one.

The earliest qualitative model for a gene regulatory network was introduced by Kauffman [5]. In Kauffman's model, a gene is modelled as a binary variable (0 or 1) which takes only one of the possible Boolean functions of its inputs. When the gene is on it takes the value of 1, otherwise it takes 0. The outputs of a gene at time $t+1$ depends only on the activity at time t. In this group of qualitative modelling frameworks, we can also cite the framework of R. Thomas in which each gene can have several levels of expression [13]. Thomas' model allows the gene to be represented as a multilevel logical variable $(0, 1, 2, \ldots)$; the number of possible values depends on the number of distinct actions it does on the network. In this case, the actions refer to a gene acting as an activator or repressor of some of the genes in the network. For each distinct action, a threshold value is assigned to specify from which expression level the influence takes place. So a variable with n distinct actions has n thresholds and this variable becomes an $(n+1)$-level variable. Allowing multilevel logical variables guarantees that no two distinct actions can happen simultaneously.

In order to illustrate our modelling approach, we focus on the gene regulatory network of the pathogen *Pseudomonas aeruginosa*, more specifically on the subsystem which is responsible for mucus production in the lungs of individuals with cystic fibrosis. Although the global gene regulatory network of

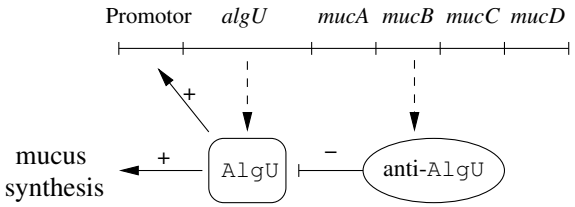

Figure 1: Portion of the gene regulatory network of the pathogen *Pseudomonas aeruginosa*, responsible for the mucus production; an arrow indicates activation or stimulation while a T-symbol represents repression [1, 2].

this pathogen consists of 690 genes and 1020 regulatory interactions between their products [3], the subsystem controlling the mucus production consists of some genes and proteins, see Figure 1. Because the mucus production worsens the respiratory problem of the patients which is often the cause of death [2, p. 75], elucidating the behaviour of this subsystem may be of great help to address this outcome.

The paper is organised as follows. Section 2 is devoted to sketch the qualitative modeling framework of R. Thomas. Section 3 explains how to build a Markov chain from the set of all possible parameterizations of an interaction graph. We can then push this idea forward and propose, when biological knowledge allows to reduce the set of possible parameterizations, a unique stochastic model where it becomes possible to jump from one qualitative model to another, see Section 4. Finally Section 5 is devoted to conclusion and discussion.

2 Reminding of R. Thomas' Modelling Framework

The biological regulatory network controlling the mucus production in *Pseudomonas aeruginosa* can be abstracted by the simple directed graphs of Figure 2 in which positive and negative signs indicate activation and repression, respectively, following the direction of the edge they label. These interaction graphs would suffice if we are only interested in applying Kauffman's model but if we want to apply Thomas' model there must be a threshold indicator for each distinct action of the gene as seen in Figure 2.

Such interaction graphs, as those in Figure 2, are called *biological regulatory graphs* [1, Definition 1] and are represented by graphs $G = (V, E)$, where V is the set of genes in the network and E represents the set of interactions between the genes in V. Each vertex $v \in V$ has a boundary b_v that is less than or equal to its out-degree (unless its out-degree is zero in which case we take the boundary to be one) while each edge is labelled by an ordered pair

containing the threshold t and action ε (activation "$+$" or repression "$-$").

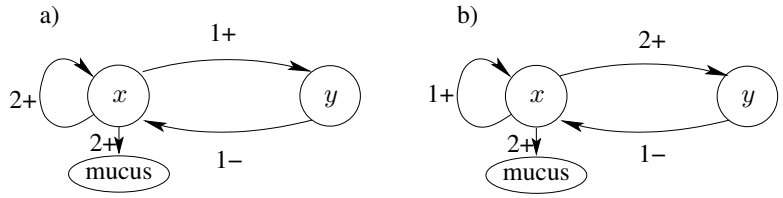

Figure 2: Two simple interaction graphs representing the system controlling the production of mucus in *Pseudomonas aeruginosa* (see Figure 1). The variable x denotes the gene algU and the protein AlgU while y denotes the gene mucB and the anti-AlgU. Both biological regulatory graphs differ by the labelling of edges outgoing from node x: the thresholds are not the same.

In Figure 2, we have $V = \{x, y\}$ and $E = \{(x \to x), (x \to y), (y \to x)\}$. The interaction $x \to$ mucus is not taken into consideration since mucus has no action backward toward x and y: it is a by-product of x which is produced when x is at its highest level (the regulation ($x \to$ mucus) is labelled by the threshold 2 since x has two distinct actions on y and on itself). The variable y has a unique action (repression of x) so the only possible threshold of the regulation ($y \to x$) is 1. Lastly note that Figure 2 does consider two possible biological regulatory graphs because the ordering between thresholds labelling edges outgoing from node x is not well known: we have to consider the two possible orderings.

In order to build the the dynamics of a biological regulatory graph, we first introduce the states of the network. A *state* of a regulatory network is a tuple denoted by $(n_{v_1}, \ldots, n_{v_p})$, where p denotes the number of genes and for each $n_{v_i} \in \mathbb{N}$ (natural numbers / non negative integers) $n_{v_i} \leq b_v$ [1, Definition 3]. We have now to define the *resources* of a vertex v_i with respect to a state $(n_{v_1}, \ldots, n_{v_p})$. Given a regulatory network, a state $(n_{v_1}, \ldots, n_{v_p})$ and an edge $(v_i \to v_j)$ with label (t, ε), the vertex v_i is a resource of v_j if and only if $n_{v_i} \geq t$ and $\varepsilon = +$ or $n_{v_i} < t$ and $\varepsilon = -$ [1, Definition 4]. The intuition is that the absence of an inhibitor plays the same role as the presence of an activator. Finally, a *biological regulatory network* refers to the biological regulatory graph $G = (V, E)$ together with a set of parameters $\mathscr{K} = \{k_{v,\omega}\}$, where $v \in V, \omega \subset G^{-1}(v) = \{u \mid (u \to v)$ is an edge in $G\}$ and $k_{v,\omega} \leq b_v$ [1, Definition 2]. The parameter $k_{v,\omega}$ gives the value towards which v is attracted when the set of resources of v is ω.

An easy way to represent the dynamics of a regulatory network is to associate with each state, the state towards which the system is attracted, when considering that each variable v changes at the same time to its current attrac-

tion value $k_{v,\omega}$ (ω being the current set of ressources of v). This defines the so-called *synchronous state graph* $\mathscr{S} = (S, T)$: The set of vertices S contains all possible states, and the edges of T are of the form $(n_{v_1}, \ldots, n_{v_p}) \to (k_{(v_1,\omega_1)}, \ldots, k_{(v_p,\omega_p)})$ such that for every i, ω_i is the set of resources of v_i at the state $(n_{v_1}, \ldots, n_{v_p})$ [1, Definition 5]. Unfortunately, the parameters $k_{v,\omega}$ are not measurable in vivo [1, p. 342] so we are left with several possibilities which results to obtaining several synchronous state graphs.

The synchronous state graph is not well adapted to represent evolution of the biological system because it is improbable that two (or more) genes reach their thresholds exactly at the same time and because a gene cannot directly jump two or more consecutive thresholds. To correct these drawbacks, one has to *desynchronize* each transition. Each transition $(n_{v_1}, \ldots, n_{v_p}) \to (n'_{v_1}, \ldots, n'_{v_p})$ is replaced by the set of its *desynchronizations* which are of the form $(n_{v_1}, \ldots, n_{v_i-1}, n_{v_i}, n_{v_i+1}, \ldots, n_{v_p}) \to (n_{v_1}, \ldots, n_{v_i-1}, n_{v_i} + \delta, n_{v_i+1}, \ldots, n_{v_p})$ for i such that $n_{v_i} \neq n'_{v_i}$ and $\delta = 1$ when $n_{v_i} < n'_{v_i}$, otherwise $\delta = -1$ [1, Definition 6]. The desynchronization step allows some states to transition to more than one other state. Thus, the dynamics of the regulatory graph is represented by the *asynchronous state graph* $\mathscr{S}' = (S, T')$ where the set S of vertices is the set of states and the set T' of transitions contains all desynchronized transitions of the synchronous state graph [1, Definition 7]. Note that two different synchronous state graphs may lead to the same asynchronous state graph since the desynchronization step can reduce two distinct synchronous transitions to the same set of desynchronized transitions.

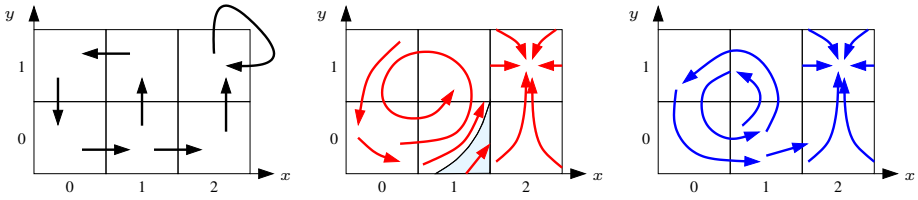

Figure 3: A possible qualitative dynamics in the modelling framework of R. Thomas (left). This asynchronous state graph can correspond to an inward spiral (centre) or to an outward spiral (right).

The global modelling approach consists of identifying all variables of the system, as well as their interactions and then the identification of parameters. Unfortunately, sometimes, it is not clear which parameters to choose. Consider the possible qualitative dynamics shown in Figure 3 (left) where we see that the state $(2, 1)$ is a stable state and that the system presents a counterclockwise oscillation between states which have a level of x less or equal to 1. It is clear that both the inward spiral (Figure 3 centre) and the outward spiral (right) are

represented by the same qualitative model. But it could be more convenient to represent the inward spiral by the model where transition $(1,0) \to (2,0)$ does not appear: the small part of the domain $(1,0)$ from which the stable domain $(2,1)$ is reachable, is integrated in the domain $(2,0)$, and when a grand tour is done in the inward spiral, it become impossible to reach the stable domain $(2,1)$.

3 Markov Chains in Gene Regulatory Networks

When Kauffman proposed the Boolean model, the output of the genes at time $t+1$ were only dependent on the activity of its inputs at time t [5, p. 441] which resembles the Markov property where given the past and the present, the future only depends on the present [9, p. 163]. In Kauffman's model a gene at time t transitions only to exactly one state at $t+1$ while in a Markov chain, the transition probabilities allows the system to go from a state at time t to more than one other state at $t+1$. In this section, we discuss briefly several ways of setting up Markov chains for gene regulatory networks based on available literature (see the works of Skornyakov *et al.* [12], Kim *et al.* [7] and Shmulevich *et al.* [11]) and then we give a basic Markov chain that represents the asynchronous dynamics of the interaction graphs of Figure 2.

In these three articles [7, 11, 12], a state can be thought of as a snapshot of the activity level of all the genes with respect to a given time. In [11], these states were referred to as maps. The Markov chain was applied to Kauffman's Boolean model of a gene regulatory network but it requires that the cooperation between interactions are well specified. In [7], the Markov chain allowed each gene to take three states, namely -1 (under-expressed), 0 (equivalently-expressed) and 1 (over-expressed) and it makes use of conditional probabilities to compute the transition probabilities. The Probabilistic Boolean Network (PBN) [11] addresses the deterministic nature of Kauffman's Boolean model. Both works [7, 11] can be easily extended to Thomas' model. However, the updates on all the genes in a PBN are done synchronously to simplify computation while preserving the generic properties of global network dynamics [11].

In Thomas' modelling framework, there are several possible values for the parameters k_{v_i,ω_i}, where v_i denotes the ith gene while ω_i denotes the ith gene's resources. The variability of these parameters results to potentially enormous (exponential) number of synchronous state graphs, but this number can be largely trimmed by considering the following constraints:

$$k_{v,\emptyset} = 0 \text{ and } \omega \subseteq \omega' \Rightarrow k_{v,\omega} \leq k_{v,\omega'}. \tag{1}$$

When a gene v has no resources, its expression level is not supposed to increase. Hence, a value of zero is assigned to $k_{v,\emptyset}$. When a gene loses some of

its resources its expression level may drop while increasing its resources may increase its expression level. Because of the constraints in (1), the number of synchronous state graphs for each regulatory graph in Figure 2 is reduced to 28 which is now a reasonable number of graphs to work with. These synchronous state graphs can be obtained by playing with the different values of the parameters of Table 1 with the previous constraints of inclusion of resources in mind.

a)

State		Next State	
x	y	x	y
0	0	$k_{x,\{y\}}$	0
0	1	0	0
1	0	$k_{x,\{y\}}$	$k_{y,\{x\}}$
1	1	0	$k_{y,\{x\}}$
2	0	$k_{x,\{x,y\}}$	$k_{y,\{x\}}$
2	1	$k_{x,\{x\}}$	$k_{y,\{x\}}$

b)

State		Next State	
x	y	x	y
0	0	$k_{x,\{y\}}$	0
0	1	0	0
1	0	$k_{x,\{x,y\}}$	0
1	1	$k_{x,\{x\}}$	0
2	0	$k_{x,\{x,y\}}$	$k_{y,\{x\}}$
2	1	$k_{x,\{x\}}$	$k_{y,\{x\}}$

Table 1: Tables giving the parameters of interaction graphs according to the current state: Tables a) and b) correspond to Figures 2(a) and 2(b), respectively.

We now recall the foundations of Markov chains in discrete time on a countable state space. Let p_{ij}^n denote the probability that state j can be reached from state i in n steps. If $n = 1$ we have the entries p_{ij} of the transition matrix $\mathbf{P}$ while the n-step transition probabilities p_{ij}^n $(n > 1)$ are contained in the matrix $\mathbf{P}^n$. If for any couple of states (i, j) we can find an $n \in \mathbb{N}^+$ (positive integers) such that $p_{ij}^n > 0$ and $p_{ji}^n > 0$, then we say that the states *communicate* with each other. This indicates that all the states belong to a unique class (Markov chain is irreducible). A state i has period d if $p_{ii}^n = 0$ whenever n is not divisible by d and d is the greatest integer with this property [9, p. 169]. In an irreducible aperiodic Markov chain the states are either all transient or null recurrent (finite number of visits) or positive recurrent (infinite number of visits) with a unique stationary distribution $\{\pi_j, j = 1, 2, \ldots\}$, where $\pi_j = \lim_{n\to\infty} p_{ij}^n > 0$ [9, Theorem 4.3.3]. Note that an irreducible Markov chain with a finite state space cannot have transient states because the chain will eventually stop once it has visited all the states in a finite number of time which should not be the case [9, p. 170]. Thus, in such a chain, all the states must be positive recurrent.

Let μ_{jj} denote the expected number of transitions needed to return to state j starting from j. When state j is positive recurrent $\mu_{jj} < \infty$ [9, p. 173] and when the Markov chain is aperiodic and irreducible, we have $\lim_{n\to\infty} p_{ij}^n = 1/\mu_{jj}$ [9, Theorem 4.3.1]. It follows that in such a Markov chain, we have $\pi_j = 1/\mu_{jj}$. An aperiodic irreducible positive recurrent Markov chain is called *ergodic* [9, p. 177]. In an ergodic Markov chain we have a limiting matrix

$\mathbf{\Pi} = \lim_{n\to\infty} \mathbf{P}^n$ with all rows having the same vector $\boldsymbol{\pi} = (\pi_1, \pi_2, \ldots)$ of positive probabilities with sum equal to 1; the probability π_i denotes the long-run proportion of time that the Markov chain stays in state i

(A)	(0,0)	(0,1)	(1,0)	(1,1)	(2,0)	(2,1)
(0,0)	12	0	16	0	0	0
(0,1)	28	0	0	0	0	0
(1,0)	9	0	5	9.5	4.5	0
(1,1)	0	21	7	0	0	0
(2,0)	0	0	7.5	0	9	11.5
(2,1)	0	0	0	16.5	8.5	3

(B)	(0,0)	(0,1)	(1,0)	(1,1)	(2,0)	(2,1)
(0,0)	12	0	16	0	0	0
(0,1)	28	0	0	0	0	0
(1,0)	2	0	8	0	18	0
(1,1)	0	6	19	0	0	3
(2,0)	0	0	7.5	0	9	11.5
(2,1)	0	0	0	16.5	8.5	3

Table 2: Sum of the probabilities assigned to each possible transition over all asynchronous state graphs. (A) and (B) are built from Tables 1(a) and 1(b) respectively. These numbers take into account the multiplicity of asynchronous state graphs.

To model the asynchronous dynamics by a Markov chain, we examine all the possible asynchronous state graphs. Since different synchronous state graphs may lead to an identical asynchronous state graph, the number of distinct asynchronous state graphs can be less than the number of distinct synchronous state graphs. In that case, we also have to take into account the multiplicity (number of occurrences) of each distinct asynchronous state graph. In each asynchronous state graph, we assign appropriate probabilities to transitions (the transitions outgoing from a same state receive the same probability if no knowledge contradicts this hypothesis). Once this is done for each asynchronous state graph, we multiply the probabilities assigned to each possible transition by the multiplicity of the asynchronous state graph and take the sum of all such terms over all the possible asynchronous state graphs.

We now set-up the transition probability matrices for Tables 1(a) and 1(b) which are obtained by simply dividing the entries of Table 2 by the total number of synchronous state graphs which is 28 as already mentioned. We have:

$$\mathbf{P}_a = \begin{bmatrix} \frac{3}{7} & 0 & \frac{4}{7} & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ \frac{9}{28} & 0 & \frac{5}{28} & \frac{19}{56} & \frac{9}{56} & 0 \\ 0 & \frac{3}{4} & \frac{1}{4} & 0 & 0 & 0 \\ 0 & 0 & \frac{15}{56} & 0 & \frac{9}{28} & \frac{23}{56} \\ 0 & 0 & 0 & \frac{33}{56} & \frac{17}{56} & \frac{3}{28} \end{bmatrix} \text{ and } \mathbf{P}_b = \begin{bmatrix} \frac{3}{7} & 0 & \frac{4}{7} & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{14} & 0 & \frac{2}{7} & 0 & \frac{9}{14} & 0 \\ 0 & \frac{3}{14} & \frac{19}{28} & 0 & 0 & \frac{3}{28} \\ 0 & 0 & \frac{15}{56} & 0 & \frac{9}{28} & \frac{23}{56} \\ 0 & 0 & 0 & \frac{33}{56} & \frac{17}{56} & \frac{3}{28} \end{bmatrix}.$$

The order of the entries in $\mathbf{P}_a$ and $\mathbf{P}_b$ follow the order given in Table 2. These Markov chains are ergodic, which guarantees the existence of a unique stationary probability π_i which gives the long-run proportion of time in i.

- Evaluating $\lim_{n\to\infty} \mathbf{P}_a^n$, we have $\pi_{(0,0)} = 0.339$, $\pi_{(0,1)} = 0.096$, $\pi_{(1,0)} = 0.304$, $\pi_{(1,1)} = 0.128$, $\pi_{(2,0)} = 0.091$, and $\pi_{(2,1)} = 0.042$. In the long-run, the most visited states are (0,0), (1,0), and (1,1). Given the long-run probabilities, we can also compute the average number of transitions required to return to each state. Recall that μ_{ii} denotes the expected number of transitions needed to return to state i starting at state i. For this chain, the mean return times for the states $(0,0)$, $(0,1)$, $(1,0)$, $(1,1)$, $(2,0)$ and $(2,1)$ are given by $\mu_{11} = 2.949$, $\mu_{22} = 10.424$, $\mu_{33} = 3.285$, $\mu_{44} = 7.818$, $\mu_{55} = 11.014$, and $\mu_{66} = 23.944$, respectively.

- For $\mathbf{P}_b$, we obtain $\lim_{n\to\infty} \mathbf{P}_b^n$ which gives the stationary distributions $\pi_{(0,0)} = 0.073$, $\pi_{(0,1)} = 0.022$, $\pi_{(1,0)} = 0.285$, $\pi_{(1,1)} = 0.101$, $\pi_{(2,0)} = 0.347$, and $\pi_{(2,1)} = 0.172$. The most visited states are $(2,0)$, $(1,0)$, and $(2,1)$. For this chain, the mean return times are given by $\mu_{11} = 13.592$, $\mu_{22} = 46.125$, $\mu_{33} = 3.508$, $\mu_{44} = 9.884$, $\mu_{55} = 2.883$, and $\mu_{66} = 5.824$.

Both chains show that 70% of the time in the long-run $y = 0$ which means that x is not inhibited; we then expect that in the long-run $x \neq 0$ most of the time. On one hand, this is true for $\mathbf{P}_b$ since the most visited states in the long-run are the states (2, 0) and (1, 0). On the other hand, in the case of $\mathbf{P}_a$, these both states are visited only close to 40% of the time. Moreover a drawback of setting up a Markov chain this way is the inability to show the steady-states or circuits of the asynchronous state graphs.

4 Probabilistic Gene Network (PGN)

In the previous model, the Markov chain comes from a superposition of all the parameterized qualitative models. In order to distinguish the different asynchronous state graphs, we introduce a Markov chain which memorizes the asynchronous state graph a particular state is in. Because all asynchronous state graphs can differ drastically, we limit this Markov chain to a set of asynchronous state graphs that behave closely (see below).

In PGNs, we take into consideration attractors. The *attractors* of a network are the smallest sets of states from which one cannot escape [10, Section 2.5]. This can be a stable state which is a state without successors or a group of states that demonstrates sustained oscillations without exits. These latter attractors are said to be cyclic and, naturally, it is not possible to reach a stable state starting from a cyclic attractor. Note that every asynchronous state graph has at least an attractor. The stable states of an asynchronous state graph results to having absorbing states in the Markov chain built on it. The presence of absorbing states may result to obtaining an *absorbing Markov chain*; this

happens when it is possible to eventually reach an absorbing state from every state [4, p. 416]. To show a Markov chain is absorbing, we need to find an $n \in \mathbb{N}^+$ such that all the entries of $\mathbf{P}^n$ are non zero, $\mathbf{P}$ being the transition probability matrix. An absorbing Markov chain give the expected times of absorption and the probability of absorption from every transient state.

Let $\mathscr{N} = \{N_1, \ldots\}$ denote a subset of Thomas' networks (asynchronous state graphs). This set can correspond to all asynchronous state graphs that are coherent with some biological knowledge, in that sense, the set is supposed to be largely smaller than the total number of asynchronous state graphs. This set can result e.g. from a filtering step which selects only asynchronous state graphs which are coherent with behavioural properties expressed in a formal language [1]. We introduce an ordering relation between Thomas' networks: for distinct networks N and N', we have $N > N'$ if and only if N' is a subgraph of N. This ordering relation leads to consider the set of models equipped with this relation as a lattice with possibly several minimal elements. Denote by $S = \{s_1, s_2, \ldots\}$ a set of states (this set is common to all asynchronous state graphs).

The intuition is the following. A biological system can be represented by a set of different dynamics (asynchronous state graphs). In a particular environment, the biological system can behave exactly as one of these dynamics but according to some changes of the environment, the behaviour of the biological system can adopt the dynamics of another asynchronous state graph. It becomes natural to allow the Markov chain to jump from one asynchronous state graph to another. But it is unlikely that the biological system jumps from a state of a certain network toward another state in another network with a very different dynamics from the initial network. This is the reason why the jumps are possible only under some conditions on the ordering relation between Thomas' networks.

A probabilistic gene network (PGN for short) on $\mathscr{N}$ satisfies the following conditions:

i. For each $N \in \mathscr{N}$ and each transition $s \to s'$ on the asynchronous state graph N a probability of $P_N(s \to s')$ is attached in such a way that for every s, $\sum_{s' \in S} P_N(s \to s') = 1$.

ii. For each pair of networks $N, N' \in \mathscr{N}$ such that $N > N'$ and there exists no other network N'' such that $N > N'' > N'$, a probability is attached to $(N \to N')$ in such a way that the sum of all such probability for a given N is less than 1.

Once a probabilistic gene network has been established, we define its cor-

responding probabilistic state graph. Let (N, s) denote the set of nodes of the probabilistic state graph, where $N \in \mathscr{N}$ and $s \in S$. The set of edges (transitions) is defined by $(N, s) \to (N', s')$ if and only if $s \to s'$ (with $s' \neq s$) is a transition of N and $P(N \to N') \neq 0$. The probability $P((N, s) \to (N', s'))$ is defined by

$$\begin{aligned} & P((N, s) \to (N', s')) \\ & = \frac{W(N')\mathbf{1}_{[(s \to s') \in N]}(N) P_{N'}(s \to s')}{\sum_{N^* \in \mathscr{N}} \sum_{s^* \in S} W(N^*)\mathbf{1}_{[(s \to s^*) \in N^*]}(N^*) P_{N^*}(s \to s^*)}, \end{aligned} \tag{2}$$

where $\mathbf{1}_A(\cdot)$ is the indicator function and $W(N)$ denotes the weight of the network N which pertains to its multiplicity in the set of all possible asynchronous state graphs of $\mathscr{N}$. When the transition does not involve a change of network, we replace N' by N to compute for the corresponding probability.

The probabilistic state graph is used to set up a Markov chain that would hopefully give a better representation of the dynamics of a gene regulatory network. More precisely the probabilistic state graph allows one to walk inside the set of considered asynchronous state graphs. Thus, a probabilistic gene network gives information not only on the state of each genes but also the network containing the state.

We present an illustration of a Probabilistic Gene Network by considering the networks of Figure 4. For $N = N_1$ (resp. N_2), the value of $P_N(s \to s')$ is computed making the assumption that in the network N the probability of transitioning from s to any of its successor is equally likely which is 1 divided by the number of transitions outgoing from s in N. Moreover we suppose that $\mathscr{N}$ contains twice N_1 and once N_2, which can be interpreted as: staying in N_1 is more likely than jumping from N_1 to N_2. The transition probability matrix is:

$$\begin{array}{c} \\ N_1,(0,0) \\ N_1,(0,1) \\ N_1,(1,0) \\ N_1,(1,1) \\ N_1,(2,0) \\ N_1,(2,1) \\ N_2,(0,0) \\ N_2,(0,1) \\ N_2,(1,0) \\ N_2,(1,1) \\ N_2,(2,0) \\ N_2,(2,1) \end{array} \begin{array}{c} \begin{array}{cccccccccccc} \end{array} \\ \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 2/3 & 0 & 0 & 0 & 0 & 0 & 1/3 & 0 & 0 & 0 & 0 & 0 \\ 2/3 & 0 & 0 & 0 & 0 & 0 & 1/3 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1/3 & 1/3 & 0 & 0 & 0 & 0 & 1/6 & 1/6 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1/3 & 1/3 & 0 & 0 & 0 & 0 & 0 & 1/3 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1/2 & 1/2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix} \end{array}.$$

We can note that it is not possible to escape from $(N_1, (0,0))$ and $(N_1, (2,0))$. This is due to the fact that to jump from an asynchronous state graph to another one, it is mandatory to be in a state which is not stable in the initial asynchronous state graph. Thus stable states in non minimal asynchronous state graphs (in the lattice) become absorbing states of the chain. In this illustration, the resulting Markov chain is an absorbing Markov chain so we can generate the expected time of absorption from each transient state and its corresponding probability of absorption in the stable states as shown in Tables 3 and 4, respectively.

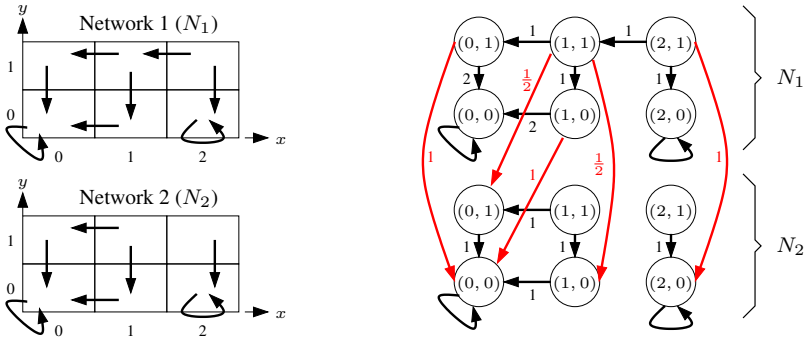

Figure 4: Two asynchronous state graphs (left) associated with interaction graph of Figure 2(a) on which we illustrate the construction of the probabilistic gene network (right). The numbers labelling the transitions (right) correspond to the numerator of Equation 2.

When involved networks have only cyclic attractors, the resulting Markov chain gets divided into several classes whose number would depend on the number of networks in the subset and on the structure of the lattice. Because the chain only allows a transition from a larger network to a smaller one, all the states in a network form a class of transient states except the states in the smallest networks (according to the lattice structure of the set of networks) which constitute different classes of recurrent states.

Transient States	Expected Time of Absorption
$(N_1,(0,1))$	1
$(N_1,(1,0))$	1
$(N_1,(1,1))$	2
$(N_1,(2,1))$	5/3 = 1.666

Transient States	Expected Time of Absorption
$(N_2,(0,1))$	1
$(N_2,(1,0))$	1
$(N_2,(1,1))$	2
$(N_2,(2,1))$	1

Table 3: Expected time of absorption (number of transitions) from any of transient state to any of the absorbing states for the networks in Figure 4.

Transient States	Probability of Absorption			
	$(N_1,(0,0))$	$(N_1,(2,0))$	$(N_2,(0,0))$	$(N_2,(2,0))$
$(N_1,(0,1))$	2/3	0	1/3	0
$(N_1,(1,0))$	2/3	0	1/3	0
$(N_1,(1,1))$	4/9	0	5/9	0
$(N_1,(2,1))$	4/27	1/3	5/27	1/3
$(N_2,(0,1))$	0	0	1	0
$(N_2,(1,0))$	0	0	1	0
$(N_2,(1,1))$	0	0	1	0
$(N_2,(2,1))$	0	0	0	1

Table 4: The transient states in the smaller network can only be absorbed in its own steady states because it is not allowed to leave the network. The transient states in the bigger network can be absorbed by any of the absorbing states of the networks under consideration, see Figure 4.

5 Discussion and conclusion

In this article we mixed two different frameworks of gene networks allowing to take advantage of the formal framework of R. Thomas modelling theory and to use transition probabilities of Markov chains to change the parameterized model. According to Figure 3, the modelling framework of R. Thomas leads to a single unique model, both the inward spiral and the outward spiral. In a natural way, it could be more efficient to represent the inward spiral by the model where transition $(1,0) \to (2,0)$ does not appear. In such a case, it could be interesting to consider in a unique framework both discrete state graphs and to allow the trajectory to jump from one to another, if information coming from biological experiments reinforces the belief in a particular model. For example, longer are the observed traces around the qualitative cycle, bigger the belief in the model representing the inward spiral.

In such a way, it becomes natural to consider each asynchronous state graph as the dynamics of the biological system in a particular context. When the environment changes the context, the qualitative dynamics can also change. Probabilistic Gene Networks presented in this article, constitute a first framework allowing to jump from a qualitative dynamics to another one.

References

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

[2] Bernot, G., J-P. Comet, A. Richard, M. Chaves, J-L. Gouzé, J-L. & F. Dayan, (2013) Modeling and analysis of gene regulatory networks. In *Modeling in Computational Biology and Biomedicine: A Multidisciplinary Endeavor*, eds F. Cazals, P. Kornprobst, pp. 47-80. Springer-Verlag Berlin Heidelberg.

[3] Galán-Vásquez, E., B. Luna, & A. Martinez-Antonio, (2011) The regulatory network of *Pseudomonas aeruginosa*. *Microbial Informatics and Experimentation* **1(1)**, 3. doi:10.1186/2042-5783-1-3

[4] Grinstead, C.M. & J.L. Snell, (1997) Introduction to Probability, 2nd edn (pp. 415-422). *American Mathematical Society*.

[5] Kauffman, S.A., (1969). Metabolic stability and epigenesis in randomly constructed genetic nets. *Journal of Theoretical Biology* **22**: 437-467.

[6] Klipp, E., R. Herwig, A. Kowad, C. Wierling & H. Lehrach, (2005) Systems Biology in Practice: Concepts, Implementation and Application. *Wiley-VCH*, Weinheim.

[7] Kim, S., H. Li, H., E.R. Dougherty, N. Cao, Y. Chen, M. Bittner, & E.B. Suh, (2002) Can Markov chain models mimic biological regulations? *Journal of Biological Systems* **10(4)**: 337-357.

[8] National Institutes of Health. *Talking Glossary of Genetic Terms*. National Human Genome Research Institute.

[9] Ross, S.M., (1996) Stochastic Processes, 2nd edn. *John Wiley & Sons, Inc.*, USA.

[10] Richard, A., J-P. Comet & G. Bernot, (2008) R. Thomas' logical model. Available from: http://www.i3s.unice.fr/ bernot/Teaching/2008-LilleSchool-RichardCometBernot.pdf

[11] Shmulevich, I., E.R. Dougherty, S. Kim & W. Zhang, (2002) Probabilistic Boolean networks: a rule-based uncertainty model for gene regulatory networks. *Bioinformatics* **18(2)**, 261-274.

[12] Skornyakov, V.P., M.V. Skoryakova, A.V. Shurygina & P.V. Skornyakov, (2014) *Finite-state discrete-time Markov chain models of gene regulatory networks*. [Preprint] Available from: http://biorxiv.org/

[13] Thomas, R., (1991). Regulatory networks seen as asynchronous automata: a logical description. *Journal of Theoretical Biology* **153**: 1-23.