CENTRO DE INVESTIGACIÓN Y DE ESTUDIOS AVANZADOS DEL INSTITUTO POLITÉCNICO NACIONAL

UNIDAD ZACATENCO

DEPARTAMENTO DE CONTROL AUTOMÁTICO

"Computational controllability analysis of complex biological Gene Regulatory Networks: The Epithelial-to-Mesenchymal transition in the context of epithelial cancer as study case"

T H E S I S

Presented by

JOSÉ EDUARDO CHAIREZ VELOZ

To obtain the degree of

DOCTOR OF PHILOSOPHY

In the field of

AUTOMATIC CONTROL

Thesis advisors:

Prof. Juan Carlos Martínez García
Prof. María Elena Álvarez-Buylla Roces

Mexico City, Mexico. July, 2021

CENTRO DE INVESTIGACIÓN Y DE ESTUDIOS AVANZADOS DEL INSTITUTO POLITÉCNICO NACIONAL

UNIDAD ZACATENCO

DEPARTAMENTO DE CONTROL AUTOMÁTICO

Análisis de controlabilidad de redes de regulación genética biológicas complejas: La transición Epitelio-Mesénquima en el contexto de cáncer epitelial como caso de estudio

T E S I S

Que presenta

JOSÉ EDUARDO CHAIREZ VELOZ

Para obtener el grado de

DOCTOR EN CIENCIAS

En la especialidad de

CONTROL AUTOMÁTICO

Directores de tesis:

Dr. Juan Carlos Martínez García
Dra. María Elena Álvarez-Buylla Roces

Ciudad de México, México. Julio, 2021

Dedication

To my beloved parents,
Sara and José,
who have always given me uncountable love and unconditional support.

To my three life warrior siblings,
Elisa, Diego and Nahúm,
who have constantly taught me that happiness is only real when shared.

Acknowledgments

I would like to deeply thank my advisors Prof. Juan Carlos Martínez García and Prof. María Elena Álvarez-Buylla Roces for their exceptional mentorship. I am sincerely grateful for all their incredible support, patience and academic and life lessons. Their solid experience in wide different scientific areas prompted me to continue until the end.

I am grateful to my committee members for their constructive comments, remarks and guidance throughout my stay in the Department of Automatic Control at CINVESTAV-IPN.

I express my very warm gratitude to my beloved parents; Sara and José, as well as my siblings; Erick Nahúm, Elisa and Diego, for their love, kindness and extraordinary support during this very long journey. Along with my family, I also warmly thank Irasema, Erika, Acenett, José Manuel, Manuel Alejandro, Karina, Fernando and Gerardo for their hugs and love during my happy times, but specially for those powerful words to never give up in my darker hours.

I thank the DCA community for giving me a pleasant time during all these years. Particularly to my dear friends Víctor and Marco, from whom I have learned a lot, specially in our talks during the lunch time as well as those virtual conversations in the middle of this pandemic.

Last but not least, I would like to thank the Mexican people that by means of CONACYT, through the scholarship with CVU: 482456, have supported this research.

Contents

List of Figures

List of Tables

Resumen

La transición Epitelial a Mesenquimal (EMT) está involucrada en la morfogénesis, la regeneración tisular y la progresión del cáncer. Este proceso dinámico se caracteriza por una serie de transiciones de estado celular, en las que la pertenencia al tejido epitelial pierde sus características epiteliales y adquiere en consecuencia propiedades mesenquimales. Debido a la importancia de comprender las circunstancias que permiten que tal transición pueda darse, se han comenzado a proponer algunos modelos de redes de regulación genética con base experimental para descubrir el núcleo regulador de EMT y, por lo tanto, contribuir a la comprensión de la regulación de EMT y guiar así experimentos mediante la generación de hipótesis comprobables. Sin embargo, los análisis sistemáticos que dilucidan el papel específico que juegan los genes implicados en la transición (adquiridos por su colaboración en la red), no sólo entre los dos estados celulares, sino también entre estados intermedios, son todavía muy escasos. En un intento por contribuir a este respecto proponemos un procedimiento analítico basado en enfoques algebraicos y construido alrededor del así denominado enfoque del Producto Semi Tensor (STP). Ilustramos el procedimiento a través de la exploración de las propiedades de controlabilidad de la red de regulación genética subyacente a la inmortalización de las células epiteliales, expresado en términos Booleanos, que recupera los perfiles de expresión génica específicos que corresponden a los fenotipos epitelial, senescente y mesenquimatoso. Nuestros hallazgos sugieren que existen 17 transiciones inducibles diferentes entre los tres fenotipos principales, mediante una entrada Booleana externa adecuada conectada a un gen particular de la red estudiada. Estas transiciones se prueban aquí tanto en el modelo Booleano como en su correspondiente modelo continuo aproximado. Además, encontramos que la matriz de controlabilidad calculada asociada a la red proporciona información relevante en torno a las perturbaciones que dan lugar a algunos estados intermedios, ya que sus entradas indican si el estado i (cualquier

configuración de activación genética posible) es accesible desde el estado j, bajo un conjunto de entradas Booleanas admisibles. Finalmente, discutimos esta información desde una perspectiva biológica y llegamos a la conclusión de que el análisis de controlabilidad puede darnos efectivamente una idea del papel de los genes en el contexto de la red en su conjunto. Esto es particularmente importante cuando se aborda la dinámica EMT asociada con el inicio y la progresión de las trayectorias del estado celular involucradas en la enfermedad.

Summary

As a complex developmental process, Epithelial-to-Mesenchymal Transition (EMT) is involved in morphogenesis, tissue regeneration and cancer progression. This dynamical process is characterized by a series of cell-state transitions, in which belonging to epithelial tissue loose their epithelial characteristics, and gain mesenchymal properties (e.g., increasing motility). Therefore, some experimentally grounded gene regulatory networks (GRNs) models have started to be proposed to uncover the EMT regulatory core, and thus contribute to the understanding of the EMT regulation and to guide experiments by generating testable hypotheses. Nevertheless, systematic analysis that elucidate the specific role that the involved genes (acquired by their collaboration in the network), play in transitioning not only between the two cell states, but also among intermediate states, are still very scarce. In an attempt to contribute in such a need, we propose an analytical procedure based on algebraic approaches and built around Semi–Tensor Product approach (STP). We illustrate the procedure through the exploration of the structural controllability properties of the low-dimensional Boolean GRN underlying the immortalization of epithelial cells, which recovers the specific gene expression profiles that correspond to the epithelial, senescent, and mesenchymal stem-like phenotypes. Our findings suggest that there exist 17 different inducible transitions between the three main phenotypes, by a suitable external Boolean input connected to a particular gene of the network. These transitions are tested both in the Boolean and in its approximated continuous model. In addition, we found that the computed controllability matrix associated to the aforementioned EMT-GRN, provide relevant information that concerns the perturbations that give rise to some intermediate states, since its entries indicate whether the i-state (any possible gene activation configuration) is reachable from the j-state, under a set of admissible Boolean inputs. Finally, we discuss this data from a biological perspective, and we con-

clude that controllability analysis can give us insights on the role of the genes in the context of the network as a whole. This is particularly important when tackling EMT dynamics associated to the onset and progression of cell-state trajectories involved in disease.

Notation

$\mathbb{R}$	Set of real numbers
$\mathbb{R}^n$	Space of n-dimension vectors with entries in $\mathbb{R}$
$\mathbb{N}$	Set of natural numbers
$:=$	"is defined as"
$\mathcal{M}_{m\times n\times p}$	Set of $m \times n \times p$ matrices with entries in $\mathbb{R}$
$\mathcal{M}_{m\times n}$	Set of $m \times n$ matrices with entries in $\mathbb{R}$
$\mathcal{M}_n$	Set of $n \times n$ matrices with entries in $\mathbb{R}$
$Id(i_1, \ldots, i_k; n_1, \ldots n_k)$	Ordered multi-index
$\text{lcm}(p, q)$	Least common multiple of p and q
$\neg$	Negation (NOT)
$\wedge$	Conjunction (AND)
$\vee$	Disjunction (OR)
$\rightarrow$	Conditional
$\leftrightarrow$	Biconditional
δ_n^k	kth column of I_n
$\mathcal{D}$	Set $\{T, F\}$ or $\{1, 0\}$
Δ	Set $\{\delta_2^1, \delta_2^2\}$
$\otimes$	Kronecker product
$\ltimes$	Left semi-tensor product
$\mathcal{L}_{m\times n}$	Set of $m \times n$ logical matrices
$\delta_k[i_1 \ldots i_s]$	Logical matrix with $\delta_k^{i_j}$ as its jth column
$\text{Col}(A)$	Set of columns of matrix A
$\text{Col}_i(A)$	ith column of matrix A

$\mathrm{Blk}_i(A)$	ith block of matrix A
$\mathrm{tr}(A)$	Trace of A
$\mathcal{P}(k)$	Set of proper factors of k
$W_{[m,n]}$	Swap matrix with index (m,n)
Ω	Limit set
$\mathcal{B}_{m\times n}$	Set of $m\times n$ Boolean matrices
$+_{\mathcal{B}}$	Boolean addition for Boolean matrices
$\sum_{\mathcal{B}}$	Boolean sum for Boolean matrices
$\ltimes_{\mathcal{B}}$	Boolean product for Boolean matrices
$A^{(k)}$	Boolean power of Boolean matrix A
$\delta_{j_1}^{i_1}\to\delta_{j_2}^{i_2}$	State $\delta_{j_1}^{i_1}$ transit to $\delta_{j_2}^{i_2}$

Introduction

This manuscript is concerned by the field of systems biology in the context of the application of methodologies of analysis that are typical of the automatic control theory. This focused on the understanding of biological phenomena associated with the dynamics of gene regulatory networks. In particular, we are concerned with the analysis of the reachability properties of these networks within the framework of the study of developmental dynamics, mainly cell differentiation and morphogenesis. This manuscript summarizes the PhD research effort carried out by the author as a member of a a multidisciplinary team that includes both experimental biologists and researchers immersed in the fields of theoretical biology, dynamical systems and automatic control theory. The present research continues some previous work done in the context of a MSc thesis [14]. The basic objective of the study here reported concerns the design and evaluation of computational tools for the determination of the exogenous inputs that underlie biological development trajectories. We firmly believe that the uncovering of systems based on regulatory constraints potentiates the deep understanding of the basic principles that underlie development. In what follows we detail our methodological approach.

The continuing effort to understand how genes (and the gene regulatory machinery) relate to the phenotypes of living organisms leads to increasing volumes of experimental information resulting from the study of the complex biomolecular systems involved in cell differentiation and morphogenesis. In this context, the use of mathematical modeling has become not only useful but mandatory [6, 7]. Gene Regulatory Networks (GRNs) models have been shown to be particularly useful to this end. Such complex biomolecular networks that underlie cell differentiation are mainly constituted by nonlinear interactions, presenting positive and negative functional feedback loops, which are critical in the establishment of their dynamic properties. In fact, recent studies have shown that the com-

bination of such functional feedback loops are crucial for the delimitation of the network's stable configurations [8]. These stable configurations, referred in the scientific literature on dynamical systems as network's attractors, have been consistently interpreted as the cellular gene expression profiles that emerge during cell differentiation. In addition, there also exists some transitory gene activation configurations (commonly known as *basin of attraction*) which are, in a developmental process context, the undifferentiated states that eventually converge to those differentiated states. In general, the identification of developmental GRNs comes from the integration of well–curated functional data of the genes (as well as some other molecular agents) involved and their interactions, frequently via the bioinformatics methods that currently characterize high throughput genomics and others. The characterization of these gene regulatory networks and their subsequent analysis has mainly focused on recovering observed gene activation configurations for studied cell types and validate them via robustness and mutant analyses, as well as new available experimental data. Such networks have, hence, recovered the qualitative and complex nature of regulatory modules that are involved in different aspects of plant and animal development [6, 15, 25, 31, 32], and in some cases modules that are relevant for the understanding of human development dynamics and for biomedical applications [22, 30, 45, 58].

Out of bioinformatics, there have been a number of attempts to model the aforementioned developmental GRNs through formal approaches, for instance:

- linear models [13, 37, 59];
- systems of Ordinary Differential Equations [53] (ODEs);
- Bayesian networks [47];
- models that include stochastic parameters [54].

Nevertheless, the model of GRN dynamics that has received the most attention because of its successful application to the understanding of these biological regulatory networks, is the discrete Boolean network model, introduced by Stuart Kauffman [39]. In that formal modeling approach, gene expression is binarily quantized (*i.e* the gene activity is described in terms of an on/off switch), and its temporal change in gene activity is considered to occur in discrete–fixed steps. GRNs are then conceptualized as *state machines*. Besides, the state of expression of each gene is functionally related to the state of expression of

some other genes (the so–named *regulators*). For each gene the regulatory interactions that define its activity state is formally described by a specific logical rule. Even if continuous–time models can be seen as more realistic, as far as modeling of gene regulatory networks are concerned, they require to take into consideration many free parameters which are hard to constrain from experimental data, while discrete Boolean models greatly reduce the number of these parameters while still capturing the essential qualitative network dynamics. Recent studies support that the seeming simplistic discrete Boolean formalism, which emphasizes on qualitative details, may answer realistic biological questions through experimental observations [35] and/or experimental data [23, 27, 49]. Therefore, experimentally grounded discrete Boolean GRNs models have become an important modeling tool, widely used nowadays in systems biology research.

Despite the progress in the understanding and the uncovering of such relatively low–dimensional GRN modules, there exists a lack of systemic analyses geared to unravel how such mostly conserved regulatory modules interact with signaling mechanisms or micro–environmental cues, and also to further uncover the modular nature of the larger networks involved. Thus, the overall regulatory functionality of a given GRN results, then, from both the interconnection of its constitutive modules and the responses from exogenous *stimuli*. Conditioned by their topological connectivity, such exogenous inputs alter the regulatory dynamics and consequently, in some cases, a differentiated state is forced to move to another, also referred as attractor–to–attractor transition (A-to-A transition). Specifically, studies have mainly focused on A-to-A transitions, considering that they may constitute changes in cell–fate decisions. Nevertheless, the attractor to Basin or vice versa (A–to–B), and Basin to Basin transition (B–to–B) have not been explored yet. In this context, a control theory perspective is convenient. Roughly speaking, the state transition from an initial state or initial condition (x_0) of the given system to any state or final condition (x_d) by using a set of admissible controls (u) in finite time (t), is known as *reachability from the origin*. This is a systems based key concept developed by automatic control theory. It is also said that x_d is reachable from x_0 in finite time when actuated by u. Hence, this property captures in functional terms the structural constraints that limit the extent of manipulability to which a given complex GRN can accept from its entourage [28]. In terms of reachability, recent published results have established a starting point to understand the role that nodes, edges, and in general topological connectivity, may

be playing in the reachability properties of complex networks [21, 41–43]. However, those results are focused on the properties of the complex networks, left behind the nature of the system that they describe. In attempt to contribute to such a need, in this work we first proposed a systematic analysis which harnesses the reachability properties of dynamical systems, applicable to experimentally grounded Boolean GRN, to then provide a more general exploration of all possible transition through the uncovering of the controllability properties of the Boolean GRNs. Because of the inherent computational complexity that arises when studying large scale complex networks [1, 2], we restrict our study to the analysis of low–dimensional discrete Boolean GRN (at the module level). Our main purpose is to gain some biologically–motivated insights about how developmental, physiological and/or environmental cues could be acting on specific genes (as well as some other non–genetic nodes), and consequently promoting A-to-A, B-to-A or B-to-B transitions . We must point out that, in the context of biological development, core regulatory modules consist of low–dimensional discrete GRN. Thus, the proposed reachability analysis methodology is quite useful for the study of developmental transitions.

The rest of this manuscript is organized as follows:

We dedicate Chapter 1 to the technical background that is required to follow the development of our proposed methodology. We present the mathematical tools (*i.e.* the so-called Semi–Tensor Product), that led us to the discrete Boolean algebraic description of gene regulatory networks. We also profit to review the continuous–time description of this kind of networks. We illustrate how the modeling tools are applied, taking into consideration a set of developmental gene regulatory networks.

Chapter 2 is concerned by the computational reachability analysis of Boolean Control gene regulatory networks, based on the algebraic representation of these nonlinear dynamical systems. The chapter tackles the characterization for the reachable subsets of Boolean Control gene regulatory networks. The methodological approach is illustrated via the analysis of the Floral Organ Specification Gene Regulatory Network (FOS-GRN).

We are concerned in Chapter 3 by the computational controllability analysis of Boolean Control gene regulatory networks. Based on the reachability analysis methodology that we expose in Chapter 2, we develop a computational method intended to shape the control input that ensures specific developmental transitions (when possible).

We strongly believe that are our methodological approach can be very useful when

considering the uncovering of the biomolecular dynamical mechanisms that are behind chronic–degenerative diseases. For this, in Chapter 4 we illustrate our methodology (exposed in Chapter 2 and Chapter 3), via the uncovering of the biological role played by some specific genes that are integrated in the gene regulatory network that underlies the Epithelial–to–Mesenchymal transition in the context of the developmental dynamics of epithelial cancer.

We conclude this report with the general conclusion and some complementary information included in the appendices. In Appendix A we include the logical rules that characterize the discrete Boolean mathematical models that we consider in our research work. We include in Appendix B some information related to the research study turning around the Floral Organ Specification gene regulatory network. Finally, Appendix C is concerned by the research works exposed to the consideration of the scientific community, and that report what is discussed in this manuscript.

Chapter 1

Background

Introduction

THIS first chapter has as its main purpose to briefly introduce the chosen mathematical tools underlying the analysis of the dynamic behavior of gene regulatory networks described in discrete Boolean terms, their reachability properties of as well as their controllability characteristics. The regulatory networks concerned by this approach are related to developmental biological processes, mainly involved in gene regulation. Then, we review the Semi–Tensor Product (STP) of matrices [3], which have been introduced to the control theory community by Daizhang Cheng and collaborators in [17]. This Semi–Tensor Product allows the rewriting of logical rules (representing the modulation of the involved genes dynamics) as algebraic operators. This is intended to allow the application of analytic control theory standard techniques to the concerned gene regulatory networks (described in discrete Boolean terms). Thus, the given discrete Boolean gene regulatory network (in fact a nonlinear dynamical system) can be transformed into an equivalent algebraic form (or discrete–time linear system) to identify its *attractors*, *i.e.* stable configurations corresponding to specific cell phenotypes, as well as its corresponding *basins of attractors*, also referred as transitory states that eventually converge to its corresponding attractor.

We also review in this chapter the continuous–time approximation of discrete Boolean gene regulatory network, intended to uncover dynamical behaviors that cannot be easily represented in Boolean terms (*e.g.* decay rates of the involved gene transcripts). Also, in order to illustrate how to perform the dynamic analysis of gene regulatory networks, we

deal with some examples at the end of this chapter, including:

- the Floral Order Specification Gene Regulatory Network (**FOS-GRN**) model [6] and:
- the Cell-Cycle Gene Regulatory Network (**ACC-GRN**) [49], both concerning *Arabidopsis thaliana*.

These core gene regulatory networks will be analyzed throughout this thesis report. We conclude the chapter summarizing its main contributions.

We can now proceed with the presentation of the algebraic representation of gene regulatory networks via the Semi–Tensor product.

1.1 Algebraic representation of gene regulatory networks: the Semi–Tensor product approach

We dedicate this section to briefly summarize the mathematical tools turning around the a application of the so–called Semi–Tensor Product approach to the algebraic representation of synchronous discrete–time Boolean networks describing gene regulatory dynamics at the cellular level. In what follows $\mathcal{M}_{m\times n}$ stands for the set of matrices of dimensions $m \times n$, where m denotes the number of columns and n denotes the number of rows. Let us begin with the following definition:

Definition 1.1 ([17]) *Let $A \in \mathcal{M}_{m\times n}$, $B \in \mathcal{M}_{p\times q}$, and $\alpha = lcm(n, p)$ be the least common multiple of integers n and p. The left Semi–Tensor Product (STP) of A and B is defined as:*

$$A \ltimes B = \left(A \otimes I_{\frac{\alpha}{n}}\right)\left(B \otimes I_{\frac{\alpha}{p}}\right), \tag{1.1}$$

where $\otimes$ denotes the Kronecker product[1] and I_k is the identity matrix of dimensions $k \times k$, with k being a positive integer.

[1] If A is a $m \times n$ matrix and B is a $p \times q$ matrix, then the Kronecker product $A \otimes B$ is the $mp \times nq$ block matrix given by:

$$\mathbf{A} \otimes \mathbf{B} = \begin{bmatrix} a_{11}\mathbf{B} & \cdots & a_{1n}\mathbf{B} \\ \vdots & \ddots & \vdots \\ a_{m1}\mathbf{B} & \cdots & a_{mn}\mathbf{B} \end{bmatrix}.$$

Remark 1.1 *Note that if $n = p$, then $A \ltimes B = AB$,* i.e. *the conventional matrix product of A and B. So, the STP of matrices is just a generalization of the conventional matrix product. It would be expected that a commutative property does not exist as in the conventional matrix product, however with some auxiliary tools, as the swap matrix defined below, a pseudo–commutative property appears.*

Hereafter, the symbol $\ltimes$ for STP is omitted in most cases if no confusion arises. Let us know recall the swap matrix.

Definition 1.2 ([17]) *A swap matrix $W_{[m,n]}$ is an $mn \times mn$ matrix, defined as follows. Its rows and columns are labeled by double index (i, j), the columns are arranged by ordered multi–index $Id(i, j; m, n)$, and the rows are arranged by the ordered multi–index $Id(j, i; n, m)$. The element at position $[(I, J), (i, j)]$ is then:*

$$w_{(IJ),(ij)} = \begin{cases} 1, & I = i \text{ and } J = j, \\ 0, & \text{otherwise.} \end{cases}$$

Therefore, let $X \in \mathbb{R}^m$, $Y \in \mathbb{R}^n$ be two–column vectors and $W_{[m,n]}$ a swap matrix, then $W_{[m,n]}XY = YX$. Later on, it is shown that this matrix is particularly relevant for the rewriting of the Boolean functions as algebraic operators.

Let us now tackle the definition of structure matrix of Boolean operators, that will be useful when rewriting a Boolean discrete–time regulatory network in algebraic terms (*i.e.* as a state–based represented dynamical system).

1.2 Structure matrix of Boolean operators

Boolean variables, say x_i, take binary values 1 or 0 or, equivalently *true* or *false*. If it is defined the set $\mathcal{D} = \{1, 0\}$, then $x_i \in \mathcal{D}$. For compactness, consider the following notation extensively used throughout this work:

- δ_k^i is the i-th column of identity matrix I_k, and:

$$\Delta_k := \left\{ \delta_k^i \mid i = 1, 2, \ldots, k \right\}$$

is the set of all columns of I_k. In cases where $k = 2$, then $\Delta_2 := \Delta$.

Thus, if both binary values are identified as:

$$true := 1 \sim \delta_2^1$$

and:

$$false := 0 \sim \delta_2^2,$$

a Boolean variable can also be denoted by $x \in \Delta$. Note that, henceforth, $\mathcal{D}$ will be identified as Δ.

We deal now with the interaction between Boolean operators. This requires logical connectives, such as:

- *Negation* ($\neg$) or NOT,
- *Conjunction* ($\wedge$) or AND,
- *Dis-junction* ($\vee$) or OR,
- *Conditional* ($\rightarrow$)
- *Bi-conditional* ($\leftrightarrow$).

In such cases, a connective is also called a Boolean operator. We can now introduce the concept of Boolean function.

Definition 1.3 ([17]) *A Boolean function f of Boolean variables $\{x_1, x_2, \ldots, x_n\}$ is a Boolean expression involving $\{x_1, x_2, \ldots, x_n\}$ and some possible statements (or constants), joined by connectives or Boolean operators. Hence, a Boolean function is a mapping*

$$f : \Delta^n \rightarrow \Delta.$$

It is also called an n-ary operator.

Remark 1.2 *It is clear that a Boolean operator (σ) is also a Boolean function as well as a r-ary operator. For instance, negation is a 1-ary operator, while conjunction, dis–junction,*

conditional and bi–conditional are all 2*-ary operators. Thereby, an* r*–ary Boolean operator is a mapping*

$$\sigma : \Delta^r \to \Delta.$$

Vector forms of Boolean operators

In order to describe the vector form of Boolean operators, the following two definitions are here recalled:

Definition 1.4 ([17]) *A matrix* $L \in \mathcal{M}_{n\times m}$ *is called a logical matrix if* $Col(L) \subset \Delta_n$*, where* $Col(L)$ *stands for the set of columns of matrix* L*. The set of* $n \times m$ *logical matrices is denoted by* $\mathcal{L}_{n\times m}$*. If* $L \in \mathcal{L}_{n\times m}$*, then it has the form:*

$$L \;=\; [\delta_n^{i_1}\ \delta_n^{i_2}\ \cdots\ \delta_n^{i_m}] = \delta_n[i_1\ i_2\ \cdots\ i_m]$$

And:

Definition 1.5 ([17]) *A matrix* $M_\sigma \in \mathcal{L}_{2\times 2^r}$ *is a structure matrix of the* r*-ary Boolean operator* σ *if*

$$\sigma(x_1, x_2, \ldots, x_r) = M_\sigma \ltimes_{i=1}^{r} x_i$$

Based on previous definitions, a summary of the structure matrices for some fundamental Boolean operators such as NOT, AND, OR, conditional and bi-conditional is shown in Table 1.1, since those are the interconnections that commonly appear on biological Boolean networks models. The construction of the Boolean operators is omitted, but details can be reviewed in [17].

Moreover, we also use auxiliary matrices such as the power–reducing matrix $M_r = \delta_4[1\ 4]$ and the dummy matrix $M_u = \delta_2[1\ 2\ 1\ 2]$ (see [17]), involved in obtaining the structure matrix of a Boolean function f. We have then the following:

Boolean operator	Symbol	Structure matrix
NOT	$\neg$	$M_n = \delta_2[2\ 1]$
AND	$\wedge$	$M_c = \delta_2[1\ 2\ 2\ 2]$
OR	$\vee$	$M_d = \delta_2[1\ 1\ 1\ 2]$
Conditional	$\rightarrow$	$M_i = \delta_2[1\ 2\ 1\ 1]$
Biconditional	$\leftrightarrow$	$M_e = \delta_2[1\ 2\ 2\ 1]$

Table 1.1: Fundamental Boolean operators and their structure matrices

Theorem 1.1 ([17]) *Given a Boolean function f with Boolean variables $\{x_1, x_2, \ldots, x_n\}$, there exists a unique $M_f \in \mathcal{L}_{2\times 2^r}$, called the structure matrix of f, such that*

$$f(x_1, x_2, \ldots, x_n) = M_f \ltimes_{i=1}^{n} x_i$$

Remark 1.3 *The algebraic representation of a given Boolean function as a structure matrix leads directly to the transformation of a Boolean gene regulatory network into a discrete–time dynamical system represented in algebraic terms. The structure matrix allows then the construction of the set of Boolean operators that define the networks as a transition map,* i.e. *the structure matrix gives rise to a (dynamical) state–space representation of the network.*

Boolean Networks

A set of discrete–time Boolean functions f_i, for $i = 1, 2, \ldots, n$, can comprise a Boolean network.

> Although Boolean networks can be synchronous or asynchronous, we are only interested here in synchronous Boolean networks, so the asynchronous ones are out of our scope.

A synchronous Boolean Network (BN) is a discrete-time dynamic system of n Boolean variables or *nodes*, $x_i \in \mathcal{D} = \{0, 1\}$. Its *state of expression* is given by a vector of Boolean variables $(x_1(t)\ x_2(t)\ \cdots x_n(t))^T$, and changes, in discrete fixed–steps, according to the

following dynamic equation:

$$x_i(t+1) = f_i(x_1, x_2, \ldots, x_n), \tag{1.2}$$

where:

- $\{x_1, x_2, \ldots, x_n\}$ is the set of the regulators of the node x_i.
- $f_i : \Delta^n \to \Delta$ is the Boolean function related to the node x_i, which is built in agreement to the combinatorial action of its regulators.

  For further notational purposes, in some cases, the set of Boolean functions that conforms a BN can be simply indicated with Σ.

Remark 1.4 *The reason why we are only interested here in synchronous Boolean networks is because we are using them to model* transcriptional *gene regulatory networks. Experimental evidence shows that transcriptional regulation, in the context of developmental processes, can be well qualitative described by discrete–time Boolean synchronous terms. However, we are conscious of the fact that in general actual gene regulatory networks are more complex than deterministic discrete–time synchronous Boolean networks (e.g. they show nonlinear stochastic dynamical behaviors).*

Gene Regulatory Boolean Networks

In experimentally grounded Boolean Gene Regulatory Networks models (B-GRNs), a particular case of BN, the set of f_i, $i = 1, 2, \ldots, n$, expresses the relationship between the n-nodes that share regulatory interactions, involved in the process of interest and derived from experimental molecular data [7]. In fact, such B-GRNs may describe dynamics of complex systems, for example, a cell cycle regulation, cell replication or a developmental module. B-GRNs are integrated by nodes that are related with proteins, hormones, metabolites, genes, etc., which can activate or inactivate each other.

For simplicity, we hereafter refer to nodes as genes indistinctly, but in cases where biological discussion is particular relevant, we retake the biological nature of the node (that can be or not a gene product). In this context, the state of expression of a B-GRN, given by the set of genes $\{x_1, x_2, \ldots, x_n\}$, is also known as the gene profile.

Remark 1.5 *In this chapter, we are mainly concerned about studying the dynamic behavior of B-GRNs only governed by the regulatory interactions among its genes. Such networks are named as* autonomous *B-GRNs or simply GRNs, and both are used indistinctly. However, further on, in cases where an exogenous input is intentionally added to the network, in order to particularly modify its dynamic (also called as* non–autonomous *B-GRN or Boolean Control GRN), we will be more specific about the type of B-GRNs that we are working with.*

Gene profiles

A gene profile of a B-GRN can be given in equivalent different ways:

- scalar form, *i.e.* $(1\ 0\ 1)$;
- vector form, *i.e.* δ_8^3;
- decimal equivalent, also referred as D_e, *i.e.* $k \in \mathbb{Z}_+$.

Later on, those notations become particular relevant to show graphically the dynamic transitions of networks.

We provide now an academic example to illustrate these issues:

Example 1.1 ***Representations of gene profiles*** *Suppose that a B-GRN with nodes* $\{x_1, x_2, x_3\}$ *is given. Its gene profile, denoted by:*

$$x(t) = (x_1(t)\ \ x_2(t)\ \ x_3(t))$$

at time $t = 0$*, is* $x(0) = (x_1(0)\ \ x_2(0)\ \ x_3(0)) = x_0$*. Assume* $x_0 = (1\ 0\ 1)$ *in its scalar form. Thus:*

1. **Scalar form to D_e.**
   It is easy to check that the binary number $(1\ 0\ 1)_2$ *can be converted to its decimal, using* $(1 \times 2^2) + (0 \times 2^1) + (1 \times 2^0) = 5_{10} = D_e$*. Thus, we denote* $(1\ 0\ 1) \sim 5$ *to say that* $(1\ 0\ 1)$ *"is equivalent to"* 5*.*

2. **Scalar to vector form**.
   In vector form $x_0 = \delta_i^j$*, where* $i = 2^n$ *and* n *is the number of nodes. On the other*

hand, j can be obtained using $j = (2^n - D_e)$. It follows that $x_0 = \delta_{2^3}^{2^3-5} = \delta_8^3$. Thus, we denote $(1\ 0\ 1) \sim \delta_8^3$ to say that $(1\ 0\ 1)$ "is equivalent to" δ_8^3.

3. **Vector form to D_e.**
Given $x_0 = \delta_i^j = \delta_8^3$, D_e is calculated by $D_e = (i - j)$. So, $D_e = 8 - 3 = 5$. Thus, we denote $\delta_8^3 \sim 5$ to say that δ_8^3 "is equivalent to" 5.

■

Remark 1.6 *If no confusion arises, the three notations illustrated with Example 1.1 will be used indistinctly.*

It is now time to deal with the algebraic representation of discrete–time gene regulatory Boolean networks.

1.3 Algebraic form of B-GRNs

To convert the B-GRN described by Eq. (1.2) into an algebraic form or conventional discrete–time linear representation, we define the state-space vector as:

$$x(t) = x_1(t)x_2(t)\dots x_n(t) := \ltimes_{i=1}^n x_i(t).$$

Thus, using Theorem 1.1, there exists structure matrices M_{f_i}, $i = 1, 2, \dots, n$ for every Boolean function f_i associated to $x_i(t+1)$, so $x_i(t+1) = M_{f_i}x(t)$. In addition:

$$\begin{aligned} x(t+1) &= \ltimes_{i=1}^n x_i(t+1) \\ &= M_{f_1}x(t)M_{f_2}x(t)\dots M_{f_n}x(t) \end{aligned} \tag{1.3}$$

Hence, $x(t+1)$ can be expressed as:

$$x(t+1) = Lx(t), \tag{1.4}$$

where $x(t) \in \Delta_{2^n}$, and $L \in \mathcal{L}_{2^n \times 2^n}$ is called the *network transition matrix* of the B-GRN.

Remark 1.7 *We must point out that* Eq.(1.4) *is enough to fully describe the dynamics of the network, since it constitutes a linear mapping with respect to each argument.*

Remark 1.8 *The advantage of the system representation given by Eq.* (1.4)*, as an algebraic representation of the corresponding discrete–time Boolean gene regulatory networks, comes from the fact that control theory standard analytic tools can be directly applied to explore the stability–related properties of the network, which can then be used to recover structural reachability information.*

We tackle the issues mentioned in the previous remark in what follows.

1.4 Fixed–point and cyclic attractors of B-GRNs

Suppose that a given B-GRN has been converted into its algebraic form given by Eq. (1.4). Then:

Definition 1.6 (Fixed–point and cyclic attractors) *An initial configuration $x_0 \in \Delta_{2^n}$ is called a fixed–point attractor if $Lx_0 = x_0$. Otherwise, the set $\{x_0, Lx_0, \ldots, L^k x_0\}$ is a cycle with length k if $L^k x_0 = x_0$ and its elements are pairwise distinct. In both cases, attractors are denoted by Λ. Moreover, we define the attractor's set as:*

$$\Omega = \bigcup_{i=1}^{k} \Lambda_i,$$

where $k \in \mathbb{Z}_+$ is the k-th attractor.

Next, we recall the following result:

Theorem 1.2 ([17]) *Consider the Boolean network as described by Eq.* (1.2)*. $\delta_{2^n}^i$ is its fixed–point attractor if and only if, in its algebraic form* (1.4)*, the diagonal element l_{ii} of the network transition matrix L equals* 1*. Consequently, the number of fixed–point attractors of the network, denoted by N_e, is equal to the number of i for which $l_{ii} = 1$. Equivalently,*

$$N_e = trace(L).$$

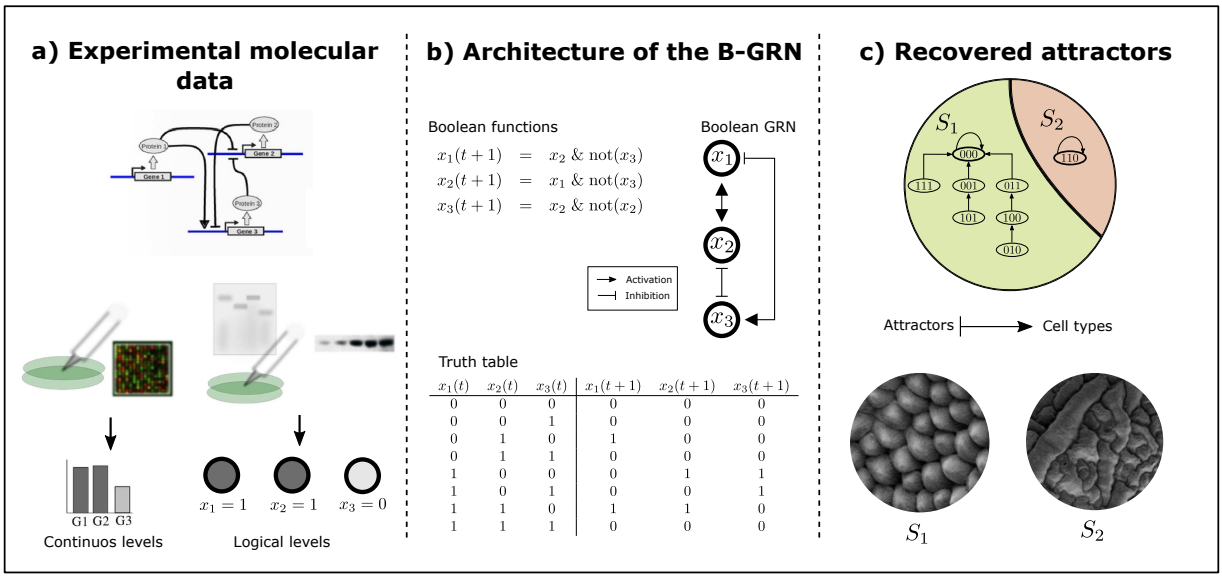


Figure 1.1: From experimental data on gene function and interactions to a Boolean GRN and its attractors. **a)** The starting point is a set of experimental molecular data of the involved genes. Genes expression profiles are determined by either continuous or Boolean levels, for being later integrated in Boolean or differential equations, respectively. In cases such as Boolean models as shown in **b)**, Boolean functions f_i are built in agreement to the combinatorial action of its regulators, and express the relationship between the genes that share regulatory interactions. Once the architecture of the B-GRN is given, diagrams as truth tables can be used to obtain the steady–states or attractors that are commonly related to specific cell types, as it is shown in **c)**.

For ease of statement, we say that $\mathrm{Col}_i(L)$ is a diagonal nonzero column of L if $l_{ii} = 1$.

Remark 1.9 *Despite we are mainly concerned about fixed–point attractors of B-GRNs since these stable gene configurations have been constantly interpreted as the cellular gene expression profiles of differentiated states (see Figure 1.1), we also briefly review how to identify cyclic attractors.*

Later on, an example of a B-GRN with a single cyclic attractor is given and all the further analysis might be also applied on it. So, consider the theorem below.

Theorem 1.3 ([17]) *The number of cycles of length s, denoted by N_s, is inductively*

determined by:

$$\begin{aligned} N_1 &= N_e, \\ N_s &= \frac{trace(L^s) - \sum_{k\in\mathcal{P}(s)} kN_k}{s} \end{aligned}$$

where $k, s \in \mathbb{Z}_+$. $2 \leq s \leq 2^n$ *is called a proper factor of* k *if* $s < k$ *and* $\frac{k}{s} \in \mathbb{Z}_+$.

Remark 1.10 *Besides,* $\mathcal{P}(k)$ *is the set of proper factors of* k*, e.g.* $\mathcal{P}(6) = \{1, 2, 3\}$. *Since* $x(t)$ *can have at most* 2^n *possible values, the length of any cycle is less than or equal to* 2^n*, hence the upper bound of* s *is also given by* 2^n. *In cases where:*

$$trace(L^s) - \sum_{k\in\mathcal{P}(s)} kN_k > 0$$

we say that s *is a nontrivial power. Now, assume that* s *is a nontrivial power. Denote by* l^s_{ii} *the* (i, i)*-th entry of matrix* L^s. *We then define*

$$\Lambda_s = \{i \mid l^s_{ii} = 1\}, \qquad s = 1, 2, \ldots 2^n$$

and

$$D_s = \Lambda_s \bigcap_{i\in\mathcal{P}(s)} \Lambda_i^C,$$

where Λ_i^C *is the complement of* Λ_i.

From the aforementioned argument, the next result is given:

Proposition 1.1 ([17]) *Let* $x_0 = \delta^i_{2^n}$. *Then* $\{x_0, Lx_0, \ldots, L^s x_0\}$ *is a cycle with length* s *if and only if* $i \in D_s$.

In cyclic attractors, a common metric related with its duration is the transient period:

$$r_0, \tag{1.5}$$

i.e. the minimum number of transient steps that lead any gene profile h to the set of attractors Λ_i. In other words, according to [17], in the sequence of matrices:

$$L^0 = I_{2^n}, L, L^2, \ldots, L^r$$

there will be eventually two equal matrices. Let $r_0 < r$ be the smallest i such that L^i appears again in the sequence. Hence:

$$r_0 = \text{argmin}_{0 \leq i < r} \left\{ L^i \in \left\{ L^{i+1}, L^{i+2}, \ldots, L^r \right\} \right\} \tag{1.6}$$

Remark 1.11 *Using Theorem 1.3 and Proposition 1.1, cyclic attractors in B-GRNs can be iteratively constructed. Therefore, for all the* 2^n *initial configurations, the steady–states or attractors,* i.e. Ω*, corresponding to specific gene expression profiles in Boolean GRN models, as well as their transient times and basins, can be easily obtained.*

We can now introduce the following:

Definition 1.7 (Basin of a given attractor) *There exists a set of gene activation configurations* (S_i) *that, in a finite time* t*, converges to* Λ_i*. Such set,* S_i*, is called the basin of attractor* Λ_i*. More precisely, a gene profile* $h \in S_i$ *if and only if the trajectory* $x(t, h)$ *with* $x(0, h) = h$*, satisfies* $x(t, h) \in \Lambda_i$ *for* $t \geq T_t$*.*

We conclude here our brief exposition on the algebraic representation of synchronous discrete Boolean gene regulatory networks, via the Semi–Tensor product. In what follows we shall deal with the continuous–time approximation of these networks.

1.5 Continuous approximations of B-GRNs

Discrete approaches do not contemplate aspects such as the differences in gene expressions decay rates, saturation rates and other quantitative parameters of biological gene regulatory networks. Continuous–time descriptions are the right models for such purposes. Fortunately, useful continuous–time models can be obtained from discrete–time models. That is the case for discrete–time Boolean gene regulatory networks. In fact, previously approaches have been used to such an end [7, 61].

The approximation used throughout this work consists in transforming each Boolean

function f_i (which depends on the regulators) into a continuous function $\bar{f}_i$ as follows:

$$\left\{\begin{array}{rcl} f_i & \rightarrow & \bar{f}_i \\ x_i(t) \ \& \ x_j(t) & \rightarrow & x_i(t) \cdot x_j(t) \\ x_i(t) \mid x_j(t) & \rightarrow & x_i(t) + x_j(t) - x_i(t) \cdot x_j(t) \\ !x_i(t) & \rightarrow & 1 - x_i(t) \end{array}\right\} \tag{1.7}$$

After that, we adopted a system of Ordinary Differential Equations (ODEs) of the form:

$$\dot{x}_i = \Theta\left[\bar{f}_i(x_1, x_2, \ldots, x_n)\right] - x_i \tag{1.8}$$

The considered input functions display a saturation behavior characterized by the following logistic function:

$$\Theta\left[\bar{f}_i(x_1, x_2, \ldots, x_n)\right] = \frac{1}{1 + \exp\left[-b\left[\bar{f}_i(x_1, x_2, \ldots, x_n)\right] - \epsilon\right]} \tag{1.9}$$

where ϵ is a threshold level (usually $\epsilon = 0.5$), and b the input saturation rate.

Remark 1.12 *We notice that for $b >> 1$ in Eq. (1.9), the input function displays a dichotomic behavior. It must be pointed out that the numerical solution of the given ODEs system can be conducted using a standard numerical algebra computer–based solver.* Matlab$^{\text{TM}}$ *provides such a solver.*

In what follows we summarize in algorithmic terms the procedure to tackle the analysis of the dynamical behavior of B-GRN, via the mathematical tools that we have just exposed.

1.6 Dynamic analysis of B-GRNs using STP and its continuous approximation

Until this point, we are now ready to present the first algorithm that mainly incorporates:

(i) The transformation of B-GRN into a discrete–time linear system using the STP approach, see Theorem 1.1.

Algorithm 1 Dynamic analysis of B-GRNs using STP and its continuous approximation

```
1: procedure DynamicAnalysis(Σ)
2:     n ← number of genes
3:     for i = 1 : n do
4:         f_i ← Boolean function of the i-th gene
5:         Rewrite f_i as algebraic operators using Thm. 1.1
6:         M_{f_i} ← structure matrix of f_i
7:     end for
8:     Compute the network transition matrix L using Eq. (1.3)
9:     Find the fixed–point or cyclic attractors using Thm. 1.2 or Prop. 1.1, respectively.
10:    Λ_p ← fixed–point attractors
11:    for j = 1 : p do
12:        S_j ← basin of attraction of the p-th attractor
13:        r_{0j} ← transient time of the j-th attractor
14:    end for
15:    Convert the Boolean functions into a ODEs system, using Eqs. (1.7), (1.8) and
   (1.9).
16: end procedure
```

(ii) The computation of fixed–point and/or cyclic attractors via: Theorem 1.2, Theorem 1.3 and Proposition 1.1.

(iii) The continuous approximation of the network by using Eqs. (1.7), (1.8) and (1.9), for gene relative expression comparative purposes.

In short, our summarizing Algorithm 1 leads with both Boolean and continuous dynamic analysis of a given B-GRN.

In what follows we shall illustrate how to apply our algorithm with some illustrative examples, including some biologically meaningful ones.

1.7 Illustrative examples

In the previous sections we exposed our methodology of analysis of the behavior of gene regulatory networks via the Semi–Tensor Product approach. In this section, we show in a detailed way how to perform our proposed dynamic analysis of B-GRN, *i.e.* we shall apply Algorithm 1. For this we shall consider some illustrative examples that will be later

useful for the purposes of this thesis manuscript. Before the detailed application of the methodology we shall introduce in what follows the set of discrete Boolean equations that comprises each of the considered B-GRNs.

Example 1.2: The TwoFixed BN**.** This example just corresponds to a designed small BN consisted of three interacting nodes. This simple idealized regulatory network has been constructed only for an illustrative purpose, which intentionally recovers two fixed–point attractors.

Example 1.3: The FOS-GRN**.** This example deals with one of the most studied GRN developmental module that has been used to uncover important aspects of early flower development, the Floral Organ Specification GNR in *Arabidopsis thaliana* (FOS-GRN). The FOS-GRN corresponds to a core regulatory module and is described as an autonomous dynamical system in Boolean terms (see the set of Boolean equations on Table A.1 and its attractors on Figure A.1) and characterized from experimental evidence. This network example recovers ten fixed–point attractors related to different cell types during flower development [6], so it becomes a first attempt to show how further analyses are relevant in biologically inspired GRNs.

Example 1.4: The ACC-GRN**.** This third example corresponds to the experimentally grounded GRN involved in the cyclic behavior of *Arabidopsis thaliana*, see [49], where the cell cycle is thoroughly revisited. The ACC-GRN, where its regulatory interactions are also described through Boolean functions (see Table A.3), recovers a single 11-length cyclic attractor, presented in Figure A.2, which is related to the G_1, S, G_2 and M phases of the cell cycle (CC).

> Notice that this example is here introduced to further illustrate how the CC can be modified and, in coordination with cell differentiation, understand the emergence of morphogenetic patterns in multicellular organisms.

We can proceed now with our illustrative examples.

Example 1.2 The TwoFixed BN

Consider the Boolean network:

$$\Sigma_{\text{TwoFixed}} : \begin{cases} x_1(t+1) &= f_1(x_2, x_3) = x_2 \wedge \neg x_3, \\ x_2(t+1) &= f_2(x_1, x_3) = x_1 \wedge \neg x_3, \\ x_3(t+1) &= f_3(x_1, x_2) = x_1 \wedge \neg x_2. \end{cases} \tag{1.10}$$

The topology of the BN (also referred as network graph) is established by the set of Boolean functions $\Sigma_{TwoFixed}$, as shown in Figure 1.2 a).

1. *We first convert every Boolean equation f_i as a structure matrix M_{f_i}:*

$$\begin{aligned} x_1(t+1) &= M_{f_1}x(t) = \delta_2[2\ 1\ 2\ 2\ 2\ 1\ 2\ 2]x(t), \\ x_2(t+1) &= M_{f_2}x(t) = \delta_2[2\ 1\ 2\ 1\ 2\ 2\ 2\ 2]x(t), \\ x_3(t+1) &= M_{f_3}x(t) = \delta_2[2\ 2\ 1\ 1\ 2\ 2\ 2\ 2]x(t). \end{aligned}$$

2. *Let us multiply the set of equations above:*

$$\begin{aligned} x(t+1) &= M_{f_1}x(t)M_{f_2}x(t)M_{f_3}x(t) \\ &= \delta_8\left[8\ 2\ 7\ 5\ 8\ 4\ 8\ 8\right]x(t) \\ &= Lx(t). \end{aligned}$$

   As we can see, in both preceding steps the STP is applied on the BN to convert it into a discrete–time linear system of the form $x(t+1) = Lx(t)$, and thus obtain its dynamic graph which comprises (see Figure 1.2 b)):

   - *the transient states;*
   - *its m-attractors of the set Ω, as well as:*
   - *its basins.*

3. *It follows that $trace(L) = 2$, so the system (1.10) has two fixed–point attractors. To find them, the condition $Col_i(L) = \delta_{2^n}^i$ must be satisfied. It holds for $i = 2$ and $i = 8$, which means that both $\delta_8^2 \sim (1,1,0) \sim 6$ and $\delta_8^8 \sim (0,0,0) \sim 0$ are the attractors*

or, equivalently, $\Omega = \{\delta_8^2, \delta_8^8\}$. *Moreover, it is easy to calculate that the smallest repeating power is* $r_0 = 3$, *also known as the transient period.*

Then, we compute the basins S_m *for each attractor* Λ_m *for* $m = 1, 2$, *where* m *is the number of attractors, as shown in Table 1.2.*

m	Λ_m	S_m
1	$\delta_8^8 \sim (0,0,0) \sim 0$	$\{\delta_8^1 \sim (1,1,1) \sim 7, \delta_8^3 \sim (1,0,1) \sim 5, \delta_8^4 \sim (1,0,0) \sim 4,$
		$\delta_8^5 \sim (0,1,1) \sim 3, \delta_8^6 \sim (0,1,0) \sim 2, \delta_8^7 \sim (0,0,1) \sim 1,$
		$\delta_8^8 \sim (0,0,0) \sim 0\}$
2	$\delta_8^2 \sim (1,1,0) \sim 6$	$\{\delta_8^2 \sim (1,1,0) \sim 6\ \}$

Table 1.2: Summary of the attractors and basins of the TwoFixed BN

Note that, for this particularly case, the TwoFixed BN is intentionally built to recover two fixed–point attractors $\Omega = \{\delta_8^2, \delta_8^8\}$, *with the main objective to illustrate how Algorithm 1 can be applied on the concerned BNs. Moreover, Boolean transitions can be simulated by observing the changes of its* D_e *at the s-step, in a graph named the Boolean dynamic transition plot. For instance, it is shown that, considering* $x_0 = \delta_8^6 \sim 2 \in S_1$, *this initial condition eventually converges to its attractor* Λ_1 *in* 3 *steps, after passing through the transient states as follows*

$$x_0 = \delta_8^6 \to \delta_8^4 \to \delta_8^3 \to \delta_8^8 = \Lambda_1, \tag{1.11}$$

where "$\to$*" means "transit to". In fact, Eq. (1.11) represents an autonomous state trajectory or simply state trajectory, and it is depicted in Figure 1.2 c).*

4. *Finally, by using Eqs. (1.7), (1.8) and (1.9), we transform the system (1.10) into its continuous approximation as follows:*

i	f_i	$\bar{f}_i$
1	$x_2 \wedge \neg x_3$	$x_2 \cdot (1 - x_3)$
2	$x_1 \wedge \neg x_3$	$x_1 \cdot (1 - x_3)$
3	$x_1 \wedge \neg x_2$	$x_1 \cdot (1 - x_2)$

Thus, the system of ODEs for (1.10) is given by:

$$\begin{aligned}\dot{x}_1 &= \Theta\left[\bar{f}_1(x_2, x_3)\right] - x_1,\\ \dot{x}_2 &= \Theta\left[\bar{f}_2(x_1, x_3)\right] - x_2,\\ \dot{x}_3 &= \Theta\left[\bar{f}_3(x_1, x_2)\right] - x_3,\end{aligned}$$

where:

$$\Theta\left[\bar{f}_i(x_1, x_2, \ldots, x_n)\right] = \frac{1}{1 + exp\left[-15\left[\hat{f}_i(x_1, x_2, \ldots, x_n)\right] - 0.5\right]}.$$

The solution of the ODEs system is displayed in the continuous dynamic transition plot, where, in contrast with the Boolean plots, here gene relative expressions can be individually compared. Notice that in Figure 1.2 d), contrary to Boolean dynamic transition plot, transient states cannot be clearly observed.

■

In short, Figure 1.2 illustrates the results that can be obtained when applying Algorithm 1 to a small and simple BN as the TwoFixed BN. In the next example a biological inspired GRN is explored to show the importance of such dynamic analysis and its results.

We hope that this simple academic example has fulfilled our objective of facilitating the understanding of the analysis tools that we are proposing. We can now consider a more significant example from a systems biology point of view.

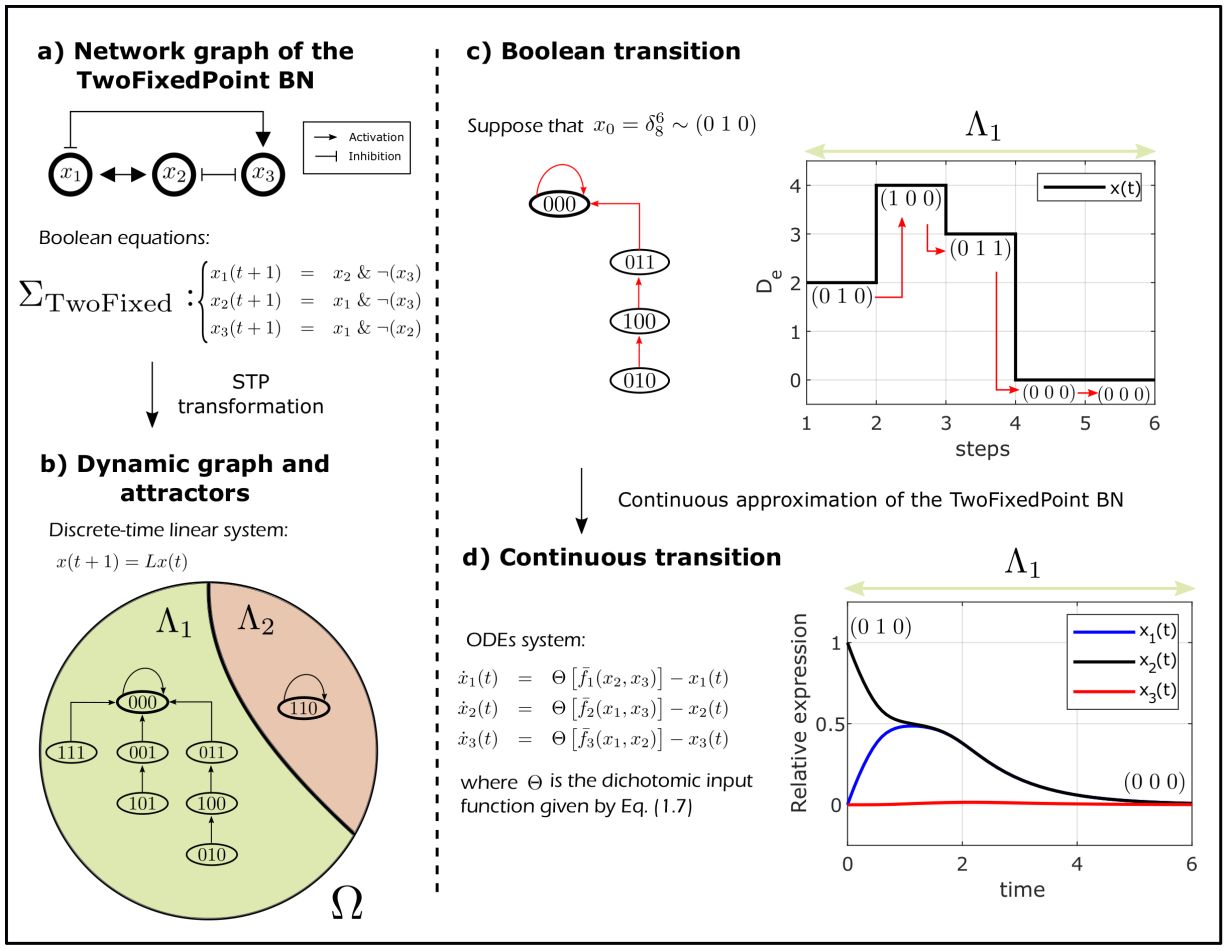


Figure 1.2: Dynamic analysis of the TwoFixed BN. **a)** Network graph of the TwoFixed BN based on Σ_{TwoFixed}. **b)** Once the BN is converted into its discrete–time linear representation, it is possible to obtain its dynamic graph which comprises attractors, basins and transient states. **c)** Then, $x_0 = \delta_8^6$ is selected to visualize its state trajectory to Λ_1, through the Boolean dynamic transition plot. Finally, the dynamic analysis concludes with **d)** the continuous approximation of the BN by using Eqs. (1.7), (1.8) and (1.9). Results are displayed in the continuous dynamic transition plot.

Example 1.3 The FOS-GRN

Let us now illustrate the proposed analysis tools through an example that has played a very important role in the consolidation of formal treatment in the study of morphogenesis. The example we include here concerns the determination of cell fate in flower conformation in Arabidopsis thaliana.

Description of the FOS-GRN

The discrete Boolean FOS-GRN model in its latest version [6], successfully modeled from experimental data for the 13 genes ($n = 13$) and their interactions, proposes a robust functional module that recovers the gene activations configurations characterizing floral organ primordia [5]. The set of Boolean equations $\Sigma_{FOS\text{-}GRN}$ that conform this B-GRN are fully presented in Table A.1. However, in compact notation we say that:

$$\Sigma_{FOS\text{-}GRN}: \quad \Big\{ x_i(t+1) = f_i(x_1, x_2, \dots x_n), \qquad i = 1, 2, \dots, n. \tag{1.12}$$

It has been shown in previous studies that, given an initial condition and using exploratory techniques [7], this autonomous dynamical system, where the state trajectories only depend on its initial conditions, can only converge to one of ten fixed–point attractors as shown in Figure 1.3 a), as a result of its network graph (see Figure 1.3 b)).

Four attractors are associated to the inflorescence meristem (IM), where four subzones (I_1, $i = 1, 2, 3, 4$) are distinguished. The other six attractors correspond to the FM that arises from the flank of the IM. FM is divided into four concentric regions that will eventually give rise to the flower organ primordia: sepals (SE), petals (PE1 and PE2), stamens (ST1 and ST2) and carpels (CAR). We must point out that in the cases such as petals and stamens, gene UFO can be expressed or non-expressed, so two different D_e can be assigned according to its gene profile. As shown in Figure 1.3 c), for illustrative purposes on further sections, we represent these attractors with colored concentric circles. Finally, a summary of the attractors and its D_e is presented in Figure 1.3 d).

Dynamic analysis

1. *The network graph of the biological GRN shown in Figure 1.4 a), involved in the floral organ specification in Arabidopsis thaliana, is built according to Eq. (1.12).*

Given the autonomous FOS-GRN, the Boolean equations $\Sigma_{FOS\text{-}GRN}$ related to each node of the network are rewritten as algebraic operators, following Theorem 1.1. However, the structure matrices M_{f_i}, $i = 1, 2, \ldots, n$ cannot be displayed here since they are a big set of data. Instead, they are computationally conducted and storaged.

2. *The rewriting of the step above allows the transformation of the dynamical system into its algebraic form, described by:*

$$x(t+1) = Lx(t),$$

where $x(t) = \ltimes_{i=1}^{n} x_i(t)$ is the STP of the molecular components of the FOS-GRN, meaning:

$$\begin{aligned} x(t) \;=\; & AG(t)AP1(t)AP2(t)AP3(t)EMF1(t)FT(t)FUL(t) \\ & LFY(t)PI(t)SEP(t)TFL1(t)UFO(t)WUS(t) \end{aligned}$$

and L is a matrix of dimensions $2^{13} \times 2^{13}$, also called the transition matrix of the network. Once L is computed, we are able to obtain the dynamic graph of the system (1.12), its attractors and basins as it is depicted in Figure 1.4 b). Notice that, in contrast with the small TwoFixed BN, FOS-GRN has 2^{13} different gene activation configurations or gene profiles. Therefore, it is not possible to have a clear plot of the dynamic graph of this network. Nevertheless, we can easily track any state trajectory caused by an arbitrary initial condition x_0 by computing.

3. *Since $trace(L) = 10$, the Boolean FOS-GRN has ten fixed–point attractors. To find Ω, the condition $Col_i(L) = \delta_{2^n}^i$ must be satisfied. Hence, the attractor's set is:*

$$\Omega = \{\delta_{8192}^{7932}, \delta_{8192}^{7930}, \delta_{8192}^{7929}, \delta_{8192}^{7931}, \delta_{8192}^{4952}, \delta_{8192}^{4424}, \delta_{8192}^{4422}, \delta_{8192}^{2312}, \delta_{8192}^{2310}, \delta_{8192}^{2824}\}$$

with $r_0 = 9$.

Then, the basins S_m for each Λ_m, $m = 1, 2, \ldots, 10$ are computed. In contrast with Table 1.2, the big set of elements of each basin cannot be displayed here. Instead, they are storaged and only their sizes are shown in Table 1.3.

The recovered fixed point or cyclic attractors, that correspond to some specific gene

activation configurations, must be equivalent to those biological observables. Ergo, to contrast results between discrete–time linear and Boolean models, the ten attractors of $\Sigma_{FOS\text{-}GRN}$ were also compared with those gene profiles previously reported in [6], conducted in R software (see Figure A.1) and as expected, they were the same.

m	Λ_m	Name tag	Basin size
1	$\delta_{8192}^{7932} \sim (0,0,0,0,1,0,0,0,0,0,1,0,0) \sim 260$	I_1	136
2	$\delta_{8192}^{7930} \sim (0,0,0,0,1,0,0,0,0,0,1,1,0) \sim 262$	I_2	136
3	$\delta_{8192}^{7929} \sim (0,0,0,0,1,0,0,0,0,0,1,1,1) \sim 263$	I_3	72
4	$\delta_{8192}^{7931} \sim (0,0,0,0,1,0,0,0,0,0,1,0,1) \sim 261$	I_4	72
5	$\delta_{8192}^{4952} \sim (0,1,1,0,0,1,0,1,0,1,0,0,0) \sim 3240$	SE	812
6	$\delta_{8192}^{4424} \sim (0,1,1,1,0,1,0,1,1,1,0,0,0) \sim 3768$	PE1	12
7	$\delta_{8192}^{4422} \sim (0,1,1,1,0,1,0,1,1,1,0,1,0) \sim 3770$	PE2	824
8	$\delta_{8192}^{2312} \sim (1,0,1,1,0,1,1,1,1,1,0,0,0) \sim 5880$	ST1	94
9	$\delta_{8192}^{2310} \sim (1,0,1,1,0,1,1,1,1,1,0,1,0) \sim 5882$	ST2	3064
10	$\delta_{8192}^{2824} \sim (1,0,1,0,0,1,1,1,1,1,0,0,0) \sim 5368$	CAR	2970

Table 1.3: Summary of the attractors and basin sizes of the FOS-GRN

Next, the following three initial conditions:

$$x_{01} = \delta_{8192}^{5319}, \qquad x_{02} = \delta_{8192}^{4356}, \qquad x_{03} = \delta_{8192}^{4103}$$

that eventually converge to ST1-attractor, are tested. The observed autonomous state trajectories were:

$$\begin{aligned} x_{01} &= \delta_{8192}^{5319} \to \delta_{8192}^{4551} \to \delta_{8192}^{327} \to \delta_{8192}^{2376} \to \delta_{8192}^{2132} = ST1, \\ x_{02} &= \delta_{8192}^{4356} \to \delta_{8192}^{4551} \to \delta_{8192}^{327} \to \delta_{8192}^{2376} \to \delta_{8192}^{2132} = ST1, \\ x_{03} &= \delta_{8192}^{4103} \to \delta_{8192}^{5448} \to \delta_{8192}^{328} \to \delta_{8192}^{237} \to \delta_{8192}^{2132} = ST1, \end{aligned}$$

for each initial configuration. For a better visualization in Figure 1.4 c), the corresponding D_e is plotted for each x_0.

4. *Lastly, we use Eqs. (1.7), (1.8) and (1.9) to transform the system (1.12) into its continuous–time approximation. Thus, the system of ODEs is:*

$$\dot{x}_i = \Theta\left[\bar{f}_i(x_1, x_2, \ldots, x_n)\right] - x_i; \qquad i = 1, 2, \ldots, 13,$$

where:

$$\Theta\left[\bar{f}_i(x_1, x_2, \ldots, x_n)\right] = \frac{1}{1 + exp\left[-15\left[\hat{f}_i(x_1, x_2, \ldots, x_n)\right] - 0.5\right]}$$

The numerical solutions of the ODEs system caused by each x_0 are showed in Figure 1.4 d). At the top of these continuous transition plots, we indicate the corresponding gene profile of each initial condition, whereas, at the bottom, the gene activation configuration of the ST1 attractor. Besides, the relative expression of those genes that notoriously change over the time are displayed in thicker lines (i.e. AG, AP1, EMF1, FT, FUL and WUS on the x_{01} to ST1 continuous transition plot).

■

As shown with the previous example, the Semi–Tensor Product representation allows, in a quiet easy way, the systemic analysis of the considered network. From a systems biology perspective the advantages are obvious. This is particularly important when the characterization of developmental trajectories is concerned.

In the following last example, we briefly introduce a B-GRN that only converges to a single cyclic attractor, in order to fully cover all types of attractors. We must point out that there exist cases where both fixed–point and cyclic attractors can appear on the same B-GRN (*i.e.* the mammalian cell cycle reported in [25]), but they are not explored in this thesis report. However, all the results presented in this work can be also applied on them without any problem.

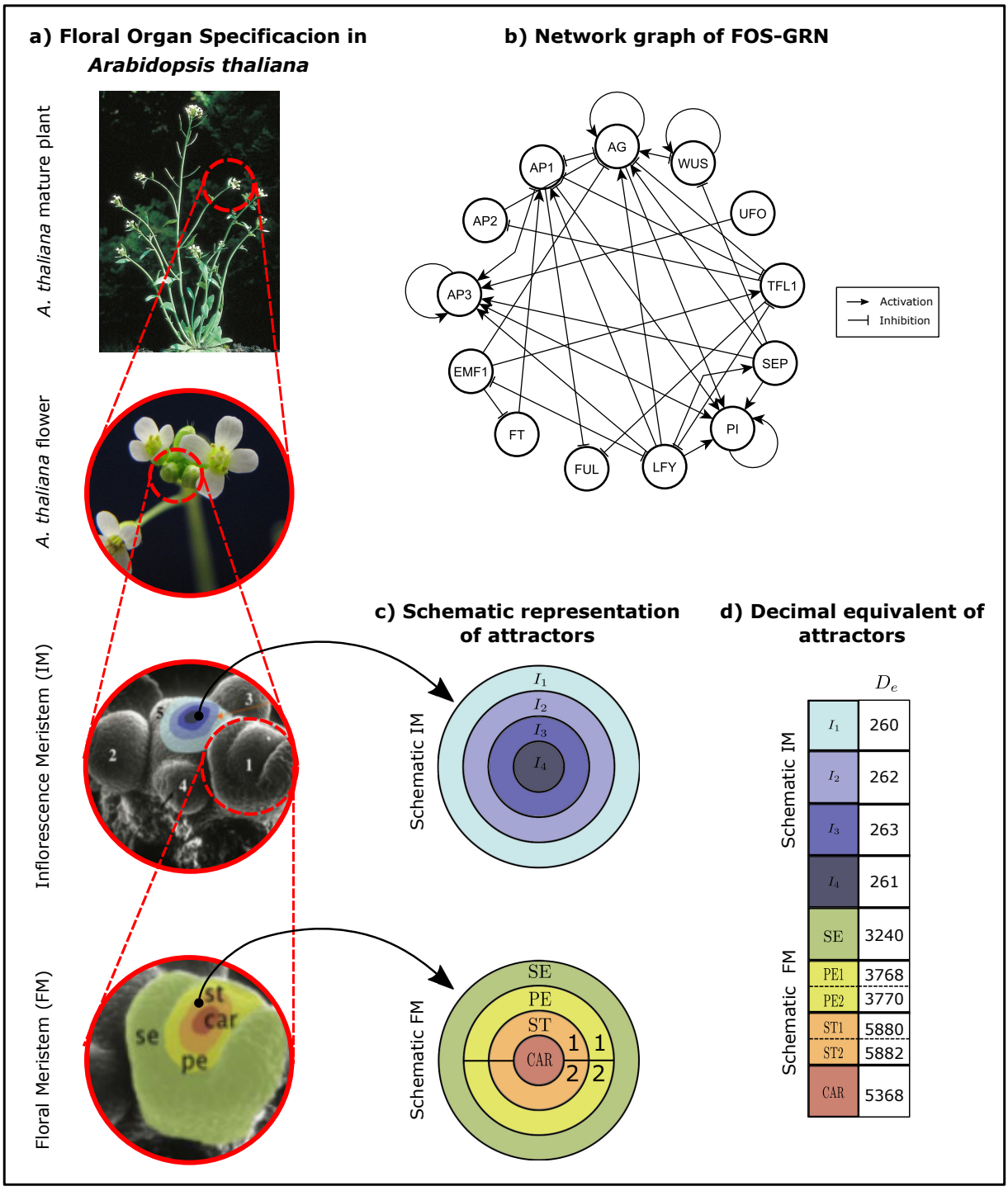


Figure 1.3: The ten fixed–point attractors of the FOS-GRN.

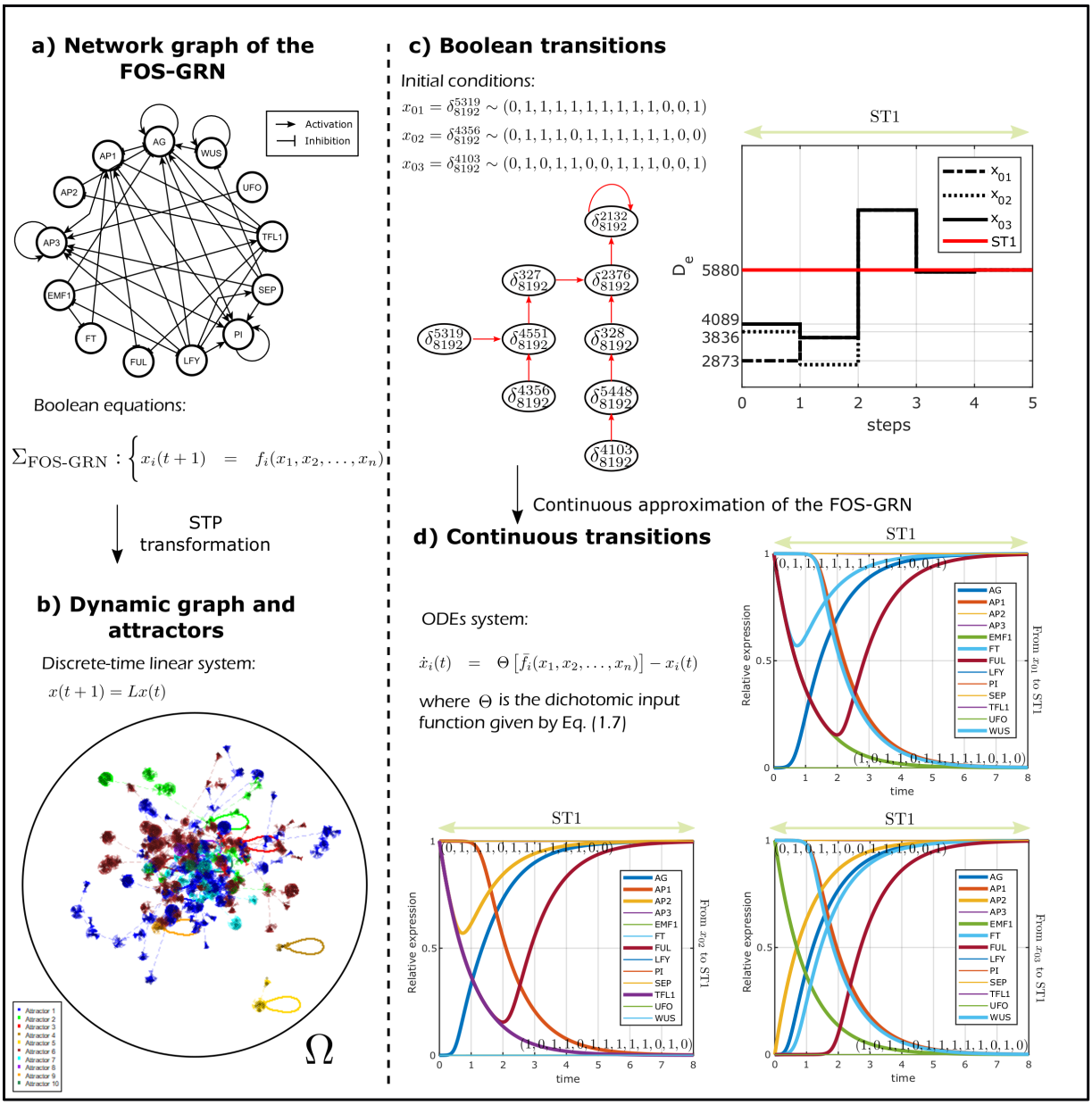


Figure 1.4: Dynamic analysis of the FOS-GRN. **a)** The network graph of the FOS-GRN is composed by 13 genes as well as several non-linear interactions among them. **b)** After a STP transformation, $\Sigma_{\text{FOS-GRN}}$ is rewritten as a discrete-time linear system of the form $x(t+1) = Lx(t)$. Hence, the dynamic graph that includes both attractors and basins can be plotted. Finally, three diferent $x_0 \in S_8$ were tested to observe its **c)** Boolean transition as well as **d)** continuous transition plots.

We present now our final illustrative example.

Example 1.4 The ACC-GRN

We present here a third illustrative example, whose purpose is to show, in the context of morphogenesis, how the proposed system analysis tool allows the study of cyclic behaviors. Let us then proceed.

Description of the ACC-GRN

The spatio-temporal regulation of cell cycle (CC) in multicellular organisms is essential during morphogenetic process since it relies on the coordinated progression of CC and its variations, i.e. CC arrest, reactivation and endoreduplication, which in coordination with cell differentiation, allow the emergence of morphogenetic patterns. Its experimentally grounded regulatory interactions that interlink these processes can be also integrated into a modular structured GRN and mathematically described through discrete–time Boolean models, in similar way that the aforementioned FOS-GRN.

Some low–dimensional Boolean GRNs have been proposed to recover the cyclic gene activation configurations observed in different CC stages, and they have been validated via robustness and mutant analysis. In this direction, we here review a B-GRN that recovers the cyclic behavior of Arabidopsis thaliana CC, referred hereafter as ACC-GRN.

The ACC-GRN model was first proposed by Ortiz–Gutierrez and collaborators in [49], and it is conformed by 14 genes ($n = 14$) and mostly nonlinear interactions as well as some feedback loops, as shown in Figure 1.5 a). In a compact notation, we describe the ACC-GRN as:

$$\Sigma_{ACC\text{-}GRN}: \quad \Big\{ x_i(t+1) = f_i(x_1, x_2, \dots x_n), \qquad i = 1, 2, \dots, 15. \tag{1.13}$$

However, the full set of Boolean equations can be reviewed in Table A.3. This biological inspired B-GRN converges into a single cyclic attractor of 11-length that comprises the G_1, S, G_2 and M phases (see Figure 1.5 b)).

Notice in Figure 1.5 b) that, in the cases such as G_1, G_2 and M phases, they are conformed by 3 different gene profiles; two of them are associated to the phase in itself, while the remaining is related to a checkpoint of transition to the subsequent phase (i.e.

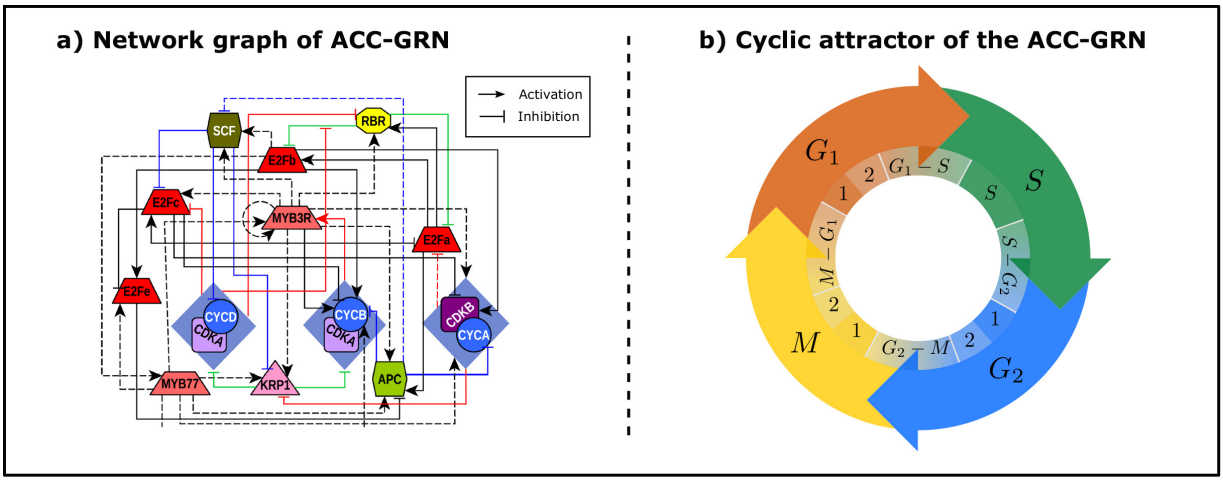


Figure 1.5: The single cyclic attractor of the ACC-GRN

$G_2 - M$ checkpoint). On the other hand, for the S phase there is only a single gene activation configuration and a checkpoint of transition to the G_2 phase.

Dynamic analysis

1. *The corresponding network graph of the ACC-GRN shown in Figure 1.6 a), involved in the CC behavior in Arabidopsis thaliana, is built in agreement to the set of Boolean equations $\Sigma_{ACC\text{-}GRN}$. We then rewrite them into algebraic operators to obtain the structure matrices M_{f_i}, $i = 1, 2, \ldots, n$, which cannot be displayed here since they are a big set of data. Instead, they are computationally conducted and storaged.*

2. *The STP transformation allows the representation of the ACC-GRN in its algebraic form, as follows:*

$$x(t+1) = Lx(t),$$

where $x(t) = \ltimes_{i=1}^{n} x_i(t)$ is the STP of the molecular components of the ACC-GRN, meaning:

$$\begin{aligned} x(t) \;=\; & CYCD31(t)SCF(t)RBR(t)E2Fa(t)E2Fb(t)E2Fc(t)E2Fe(t) \\ & MYB77(t)MYB3R14(t)CYCB11(t)CDKB11(t)CYCA23(t) \\ & KRP1(t)APCC(t) \end{aligned}$$

	Λ_1	Name tag
1	$\delta_{16384}^{8192} \sim (1,0,0,0,0,0,0,0,0,0,0,0,0,0) \sim 8192$	$G_1 - S$
2	$\delta_{16384}^{7040} \sim (1,0,0,1,0,0,1,0,0,0,0,0,0,0) \sim 9344$	S
3	$\delta_{16384}^{6272} \sim (1,0,0,1,1,1,1,0,0,0,0,0,0,0) \sim 10112$	$S - G_2$
4	$\delta_{16384}^{2112} \sim (1,1,0,1,1,1,1,1,0,0,0,0,0,0) \sim 14272$	$G_2 - 1$
5	$\delta_{16384}^{10498} \sim (0,1,0,1,1,0,1,1,1,1,1,1,1,0) \sim 5886$	$G_2 - 2$
6	$\delta_{16384}^{9220} \sim (0,1,1,0,1,1,1,1,1,1,1,1,0,0) \sim 7164$	$G_2 - M$
7	$\delta_{16384}^{9796} \sim (0,1,1,0,0,1,1,0,1,1,1,1,0,0) \sim 6588$	$M - 1$
8	$\delta_{16384}^{9924} \sim (0,1,1,0,0,1,0,0,1,1,1,1,0,0) \sim 6460$	$M - 2$
9	$\delta_{16384}^{9923} \sim (0,1,1,0,0,1,0,0,1,1,1,1,0,1) \sim 6461$	$M - G_1$
10	$\delta_{16384}^{14039} \sim (0,0,1,0,0,1,0,0,1,0,1,0,0,1) \sim 2345$	$G_1 - 1$
11	$\delta_{16384}^{10498} \sim (1,0,1,0,0,1,0,0,0,0,1,0,1,1) \sim 10507$	$G_1 - 2$

Table 1.4: Summary of the single cyclic attractor of the ACC-GRN

and L is a matrix of dimensions $2^{14} \times 2^{14}$. Once L is computed, we are able to obtain the dynamic graph of the system (1.13), its single cyclic attractor and basin as it is depicted in Figure 1.6 b).

3. *Using Theorem 1.2, we conclude that the system does not have fixed–point attractors. Nevertheless, to explore cyclic attractors on the ACC-GRN, we use Theorem 1.3 and Prop. 1.1. Hence:*

$$\begin{aligned} \delta_{16384}^{8192} \quad &\to \quad \delta_{16384}^{7040} \to \delta_{16384}^{6272} \to \delta_{16384}^{2112} \to \delta_{16384}^{10498} \to \delta_{16384}^{9220} \to \\ \delta_{16384}^{9796} \quad &\to \quad \delta_{16384}^{9924} \to \delta_{16384}^{9923} \to \delta_{16384}^{14039} \to \delta_{16384}^{10498} \end{aligned}$$

is the 11-length cyclic attractor, as expected according to [49]. A summary of the gene profiles and tags of each element of the cycle is presented in Table 1.4.

Now, let us consider $x_0 = \delta_{16384}^{8192} \sim (1,0,0,0,0,0,0,0,0,0,0,0,0,0)$, which is the $G_1 - S$ checkpoint. As expected, in Figure 1.6 c) we illustrate the cyclic dynamics through the Boolean transition plot, where, after 11 steps such cycle is repeated over and over.

4. *Lastly, we use Eqs. (1.7), (1.8) and (1.9) to transform the system 1.13 into its*

continuous approximation. Thus, the system of ODEs is:

$$\dot{x}_i = \Theta\left[\bar{f}_i(x_1, x_2, \ldots, x_n)\right] - x_i \qquad i = 1, 2, \ldots, 15,$$

where

$$\Theta\left[\bar{f}_i(x_1, x_2, \ldots, x_n)\right] = \frac{1}{1 + exp\left[-15\left[\hat{f}_i(x_1, x_2, \ldots, x_n)\right] - 0.5\right]}.$$

The numerical solution of the ODEs system originated by x_0 is showed in Figure 1.6 d). On these graphs, we first plotted the CDK-Cyclin activity and the KRP1 inhibitor (the biological relevance of these genes will be later discussed) in order to confirm that the oscillations are maintained. Additionally, according to a very preliminary analysis, which will be extensively discussed on a subsequent section, we found the potentiality of the genes MYB3R14 and APCC to disrupt the cycle behavior of the ACC-GRN, so we also included them on a continuous transition plot for further discussions.

We must point out that the rest of the genes were not intentionally displayed on Figure 1.6 d), because this information cannot be clearly observed on the plots. However, they are computationally available for future purposes.

■

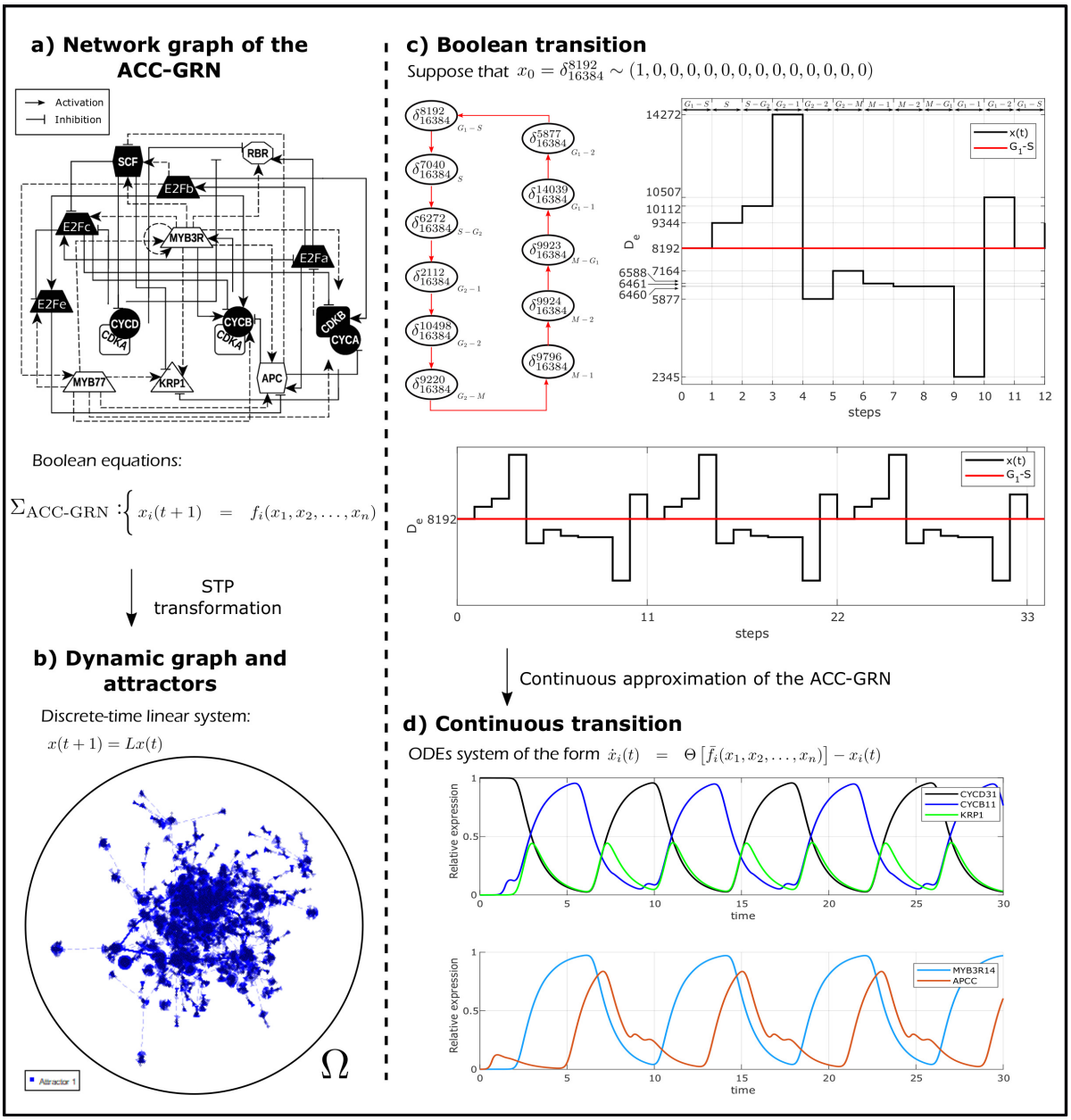


Figure 1.6: Dynamic analysis of the ACC-GRN. **a)** The network graph is built according to $\Sigma_{\text{ACC-GRN}}$. **b)** After the STP transformation, the ACC-GRN is rewritten as $x(t+1) = Lx(t)$, and thus recover its single cyclic attractor and basin. **c)** A Boolean transition plot is shown to observe the discrete cyclic dynamic behavior, given $x_0 = G_1 - S$, which is repeated every 11 steps. **d)** Finally, two continuous transition plots are displayed to evaluate if oscillations are maintained in the continuous version of ACC-GRN.

We concluded our illustrative examples. We proceed now to discuss the significance of our proposed methodology.

1.8 Discussion

Discrete–time linear representations of autonomous B-GRNs, *i.e.* discrete–time Boolean gene regulatory networks without inputs, become the starting point of our thesis work. As shown throughout this chapter, we are able to fully describe the dynamic behavior of these networks via Eq. (1.4), by the use of STP transformation [17]. Notice that the network transition matrix L, involved in such linear mapping, is unique as well as quite useful in recovering attractors (either fixed–point or cyclic), basins and transient states. Moreover, with some continuous representations of these B-GRNs, based on Eqs. (1.7), (1.8) and (1.9), we provided a complete dynamic analysis of the concerned GRN in its both discrete and continuous versions, which is integrated in our first proposed Algorithm 1. This systematic study comprises both qualitative (*i.e.* stable states restricted by its topology) as well as quantitative (*i.e.* relative gene decay variations) information of networks.

To illustrate how the aforementioned dynamic analysis can be performed on B-GRNs, Examples 1.2 (TwoFixed BN), 1.3 (FOS-GRN) and 1.4 (ACC-GRN) were introduced for the first time in the present manuscript. These examples, indeed, will be reviewed and discussed several times in subsequent chapters from a control theory perspective. In contrast with the TwoFixed BN, which has been intentionally built to recover only two fixed–point attractors with a simple topology of three genes, the FOS-GRN and the ACC-GRN are both complex biological GRN, that have been raised from experimental data and validated via mutant analysis [6, 49]. Although there have been some previous dynamic analyses applied on these two last networks, they have been mostly based on extensive iterative techniques (*i.e.* stochastic explorations of random initial conditions). In this context, the linear representations of these networks become very useful since in future chapters we will discuss how some deterministic control inputs can modify the dynamic of the B-GRNs by promoting specific state trajectories. What we have just presented in this chapter is our first step in understanding the generic principles underlying the emergence of developmental trajectories. Since there is no single representation of genetic regulatory networks, the choice of representation depends on the objective pursued.

Since development is essentially a self–organizing process that occurs when the underlying genetic regulatory networks interact with the environment in which they are embedded, a formal representation aimed at elucidating precisely the dynamic consequences of such interaction is what It required. The representation based on the Semi–Tensor Product is well aimed at satisfying the goal we are pursuing. However, we are fully aware of its intrinsic limitations. The most important limitation of the approach discussed resides in the computational cost incurred in transforming the logical rules into the algebraic operations that give rise to the network transition matrix. This is because the set of logical rules that defines the network are presented in the transformation as a concatenation. Due to this, the proposed methodology is aimed at dealing with networks of limited dimension, such as those that occur in the regulation and transcription of biological development processes.

1.9 Conclusion

In this chapter, we reviewed the STP approach, a mathematical tool that allows the rewriting of a deterministic B-GRN expressed in logical terms into its algebraic form, which corresponds to a discrete–time linear state–space representation. We constrained our development to the synchronous case. As shown in our exposition, the linear mapping is sufficient to fully describe the dynamic behavior of the transformed network. Moreover, we also included in the proposed methodology a continuous–time approximation of the B-GRN that allows to take into consideration, for instance, differences in both genetic expressions decay–rates and saturation rates. We summarized the methodology via an algorithm that provides both qualitative as well as quantitative information of autonomous B-GRNs dynamics. As illustrated, the proposed methodology provides a computer–based tool that successfully tackles the understanding of the dynamical behavior of gene regulatory networks. Specifically, we show with the examples that involve floral morphogenesis and the cell cycle, the usefulness of the methodology to systematically recover the structural indicators associated with stability. As pointed out in the Discussion section, because of the involved computational cost, we are well aware of the limitations of the methodology. Taking into account this limitation, we strongly believe that our proposal can be very useful when considering its application to the reduced–size gene regulatory networks

involved in developmental processes, the so–called core gene regulatory modules.

Chapter 2

Computational reachability analysis of Boolean Control GRNs

Introduction

DESPITE the progress in understanding and uncovering of some low-dimensional Boolean GRN modules as the examples in the previous chapter, we still lack methodologies of analysis geared to unravel how such mostly conserved regulatory modules interact with signaling mechanisms or micro–environmental cues, and also to further uncover the modular nature of the larger networks involved. We are concerned in the present chapter by this subject. We must point out that the overall regulatory functionality of a given Boolean GRN comes from:

1. The dynamical processes carried out by the interconnection of its constitutive modules.

2. The response of the network resulting from exogenous *stimuli*.

As far as the interactivity of the regulatory network with its surroundings is concerned, this is mediated by the exogenous *stimuli*. Moreover, the consequences of the *stimuli* on the behavior of the network are conditioned by the involved network topology. Which is to say, there exists an intimate relationship between the functionality of the network and how exogenous *stimuli* act on *specific* network nodes. Thus, conditioned by their topological connectivity, the exogenous inputs alter the Boolean dynamics and consequently, in some

cases, the state–space is completely restructured, *i.e.* a state trajectory is produced when the network is interacting with its entourage. Specifically, studies have mainly focused on attractor transitions, since they may constitute changes in cell–fate decisions.

Remark 2.1 *As we discussed previously, we are very conscious of the role played by stochastic fluctuations on the functionality of gene regulatory processes. However, our methodological approach is based on a deterministic point of view. This because, in the context of biological development, the structural robustness of core transcriptional regulatory modules allows its description via deterministic discrete–time Boolean networks and still get some important insights about the underlying biological principles. The structure of developmental networks allows then a deterministic representation. This does not mean that stochastic processes are not present in the dynamical behavior of this kind of networks. We can say that our methodological choice subsumes stochastic fluctuations in the (idealized) deterministic exogenous* stimuli. *This is customary in deterministic robust control* [67]*, we follow then such an approach.*

From the perspective of control theory, roughly speaking, the possibility of transfer the zero state or initial condition (x_0) of the given system to any state or final condition (x_d) by using a set of admissible controls (u) in finite–time (t), is known as *reachability from the origin*. It is also said that x_d is reachable from x_0 in finite–time when actuated by u. Hence, this property captures in functional terms the structural constraints that limit the extent of manipulability to which a given complex GRN can accept from its entourage [28]. In terms of structural controllability, recent published results have established a starting point to understand the role that nodes, edges, and in general topological connectivity, may be playing in the controllability of complex networks [20, 41, 43]. However, those results are focused on the properties of the complex networks, left behind the nature of the system that they describe. In an attempt to contribute to such a need, in this chapter we propose a systematic analysis which harnesses the reachability properties of dynamical systems, applicable to experimentally grounded Boolean GRN. Because of the inherent computational complexity that arises when studying large scale complex networks [1, 2], we restrict our study to the analysis of low–dimensional discrete Boolean GRN (at module level). Therefore, we constraint our research effort to the reachability properties of core gene regulatory networks in the context of biological development.

> Our main purpose is to gain some biologically-motivated insights about how developmental, physiological and/or environmental cues could be acting on specific genes, and consequently promoting attractor transitions.

The reachability analysis is also oriented to the uncovering of the circumstances underlying transient dynamics between considered given attractors, which define developmental trajectories. In fact, to observe this, we retake the TwoFixedBN, FOS-GRN and the ACC-GRN as examples at the end of the present chapter.

We recall in what follows the algebraic representation of the Boolean control gene regulatory networks. The Semi–Tensor Product approach previously presented is then extended to take into consideration exogenous *stimuli*. We are inspired in what follows by the work exposed in [17].

2.1 Algebraic form of Boolean Control GRNs (BC-GRNs)

In order to understand how a BN could be interacting with its entourage, and consequently modifying its dynamics, a set of Boolean inputs:

$$u_j(t), \quad j = 1, \ldots, m,$$

can be connected in some specific genes. In our approach both AND and OR Boolean operators are tested as connectors between the BN and u. We must point out that some other Boolean operators could be taken into account, but we restricted our analyses to these ones, and thus mimic a spontaneous switch-off and switch-on on the node, respectively. Hence, the BN is converted into a *non-autonomous* BN or, more precisely, a Boolean Control Network (BCN), described by:

$$x_i(t+1) = \bar{f}_i\left(x_1, x_2, \ldots, x_n, u_1, u_2, \ldots, u_m\right) \tag{2.1}$$

where $\{x_1, x_2, \ldots, x_n, u_1, u_2, \ldots, u_m\}$ are the regulators of the node x_i, which now also depends on the set of $u_j \in \Delta_{2^m}$. $\bar{f}_i : \mathcal{D}^{n+m} \to \mathcal{D}$ is the Boolean function related to the i-th node. Unlike the f_i in Eq. (1.2), $\bar{f}_i$ integrates the set of u_j as components that arbitrarily

overrule the original regulatory Boolean function of the concerned gene.

A similar procedure to the one presented in the Chapter 1, holds to convert the BCN described by Eq. (2.1) into its algebraic form (see [17]). In this case, both STP of:

$$x(t) = \ltimes_{i=1}^{n} x_i$$

and:

$$u(t) = \ltimes_{j=1}^{m} u_j$$

are considered, in order to together yield:

$$x(t+1) = \bar{L}u(t)x(t), \tag{2.2}$$

where $\tilde{L} = LW_{[2^n,2^m]}$, so $\bar{L} \in \mathcal{L}_{2^n \times 2^{n+m}}$. Moreover, $\tilde{L}$ is the network transition matrix of the BCN in its algebraic form. In fact, Eq. (2.2) is also known as the *discrete–time bilinear representation* of the BCN.

The algebraic form given by Eq. (2.2) can be also applied on B-GRNs. If it is the case, we name these networks as non-autonomous B-GRNs or Boolean Control GRNs (BC-GRNs)

Remark 2.2 *$\bar{L}$ can be split into 2^m blocks, referred to as $Blk_i(\bar{L})$, $i = 1, 2, \ldots, 2^m$, where the first block of columns corresponds to $u(t+1) = \delta_{2^m}^1$, the second block corresponds to $u(t+1) = \delta_{2^m}^2$, and so on. Namely, the entries of $\bar{L}$, l_{ij}, dictates whether or not there exists a set of Boolean control inputs such that the i-state is* reachable *from j-state in one step, based on whether $l_{ij} > 0$ or not.*

The concept of reachability, based on the information recovered from $\bar{L}$, is first presented in Remark 2.2. However, let us discuss it from a control theory point of view. Roughly speaking, the possibility to transfer the zero state or initial condition (x_0) to any state or final condition (x_d) under a suitable control input (u) in finite time (t) is called as controllability from the origin or, more often, reachability, which is a structural property of the dynamic system [16]. In fact, this statement agrees with our goal to expose the potential genes of the BC-GRN, which are involved in quite specific state transitions, since reachability can be extended to Boolean systems, considering a finite number of steps (s) as a finite time. So, exhaustive characterization of the reachable subsets (also known as

reachability analysis) associated to particular gene profiles, *i.e.* the fixed–point attractors, as $x(0) = x_0$, is sufficient to elucidate the role that potential genes play in complex B-GRNs.

Remark 2.3 *When speaking about the control input,* i.e. *u, we do not establish its specific molecular nature. At this level the control input is just an abstract sort of switch. However, for some examples we later tackle this issue when discussing how the entourage could be interacting with the concerned specific regulatory modules according to available biological data.*

This conceptual approach applied in the context of BC-GRNs, together with its discrete–time bilinear representations, is our first attempt to contribute in the setting up of systematic analyses that allow to uncover structural reachability properties in such nonlinear abstract dynamical systems.

On the other hand, notice that in Eq. (2.2):

$$\begin{array}{llll} if & t=0, & then & x(1) = \tilde{L}x(0)u(0), \\ if & t=1, & then & x(2) = \tilde{L}x(1)u(1) = \tilde{L}\tilde{L}x(0)u(0)u(1) = \tilde{L}^2x(0)u(0)u(1), \\ & \vdots & & \vdots \\ if & t=s-1 & then & x(s) = \tilde{L}^s x(0)u(0)u(1)\dots u(s-1). \end{array}$$

Thus, the next theorem, proposed by Cheng and collaborators in [17], can be introduced:

Theorem 2.1 ([17]) *x_d is reachable from x_0 at the s-th step by controls of Boolean sequences of length s if and only if*

$$x_d \in Col\left\{\tilde{L}^s x_0\right\}. \tag{2.3}$$

Remark 2.4 *Eq. (2.3) means that x_d is equal to a column of $\tilde{L}^s x_0$. In cases where such equation holds for the k-th column, the controls are defined by*

$$u(0)u(1)\dots u(s-1) = \delta^k_{2^{ms}} \tag{2.4}$$

and thus all the u_i, $i = 0, 1, \ldots s-1$ can be determined.

Theorem 2.1 and Eq. (2.4) set the stage to explore the reachability properties of a B-GRN. However:

> We restrict our exploration to the reachable subsets related to the attractor's set, since we are mainly concerned about transitions among pairs of attractors.

Remark 2.5 *Nevertheless, we must point out that the reachable subset of any initial configuration x_0 can be explored even though it is a transitory state. In fact, this scenario would be discussed in further sections, particularly when we address cyclic dynamic behaviors.*

In what follows we shall tackle the caracterization of the reachable subsets associated to a given discrete–time Boolean gene regulatory network.

2.2 Characterization of the reachable subsets of BC-GRNs

To find the potential genes that promote *non–autonomous* state trajectories among pairs of given attractors in a BC-GRN, we propose in this section a systematic procedure (see Algorithm 2). The proposed methodology harnesses the structural reachability properties of these networks, based on the result given by Eqs. (2.3) and (2.4).

Algorithm 2, once the attractors are identified, tests if adding a control input to the i-node (through either AND or OR connector) is sufficient to promote a transition from one attractor to another.

> If so, we say that the final attractor x_d is reachable from the initial attractor x_0 by a control input u via the i-node.

Such available state trajectories promoted by the action of control inputs on the nodes can be tested by discrete–time simulation, via Boolean dynamic transition plots.

Later in this chapter we shall illustrate how to characterize the reachable subsets with the same examples that have been used in the previous chapter. Before that we shall consider in the continuous–time approximation of controlled discrete–time Boolean gene regulatory networks.

Algorithm 2 Characterization of the reachable subsets associated to the fixed–point attractors in a Boolean GRN

```
n ← number of nodes/genes
m ← number of control inputs
p ← number of attractors
s ← steps
for i = 1 : s do
    for j = 1 : n do
        Add control input in j-node
        Compute L̃
        for k = 1 : p do
            x_0 ← k-attractor
            Compute L̃^i x_0
            for l = 1 : p do
                x_d ← l-attractor
                if x_d ∈ Col{L̃^i x_0} then
                    Add trajectory from x_0 to x_d via j-node
                    r ← rth column of L̃^i x_0
                    u = u(0)u(1)...u(i−1) = δ^r_{2^{mi}}
                    Save control sequence u
                end if
            end for
        end for
    end for
end for
```

2.3 Continuous approximations of BC-GRNs

In order to recover not only qualitative but also quantitative information, *i.e.* the decay/saturation rates of the genes, several approaches have been used to describe Boolean GRNs models as continuous systems [7, 56, 60, 61]. In this same direction, we are interested in testing whether or not the Boolean inputs obtained through the computational reachability analysis can drive the system from x_0 to x_d in a continuous version of the system. To this end, we adopt here a system similar to ODEs as the one taken into consideration in Chapter 1:

$$\dot{x}_i = \Theta\left[\hat{f}_i\left(x_1, x_2, \ldots, x_n, u_1, u_2, \ldots, u_m\right)\right] - x_i, \tag{2.5}$$

where:

$$\hat{f}_i : \mathcal{D}^{n+m} \to C\,[0,1]$$

is the continuous function related to the i-th gene, which now not only depends on its regulators but also on its control inputs, that results from the transformation of the Boolean function $\bar{f}_i$ given by Eq. (2.1), as follows:

$$\left\{ \begin{array}{rcl} \bar{f}_i & \to & \hat{f}_i, \\ x_i(t) \;\&\; x_j(t) & \to & x_i(t) \cdot x_j(t), \\ x_i(t) \mid x_j(t) & \to & x_i(t) + x_j(t) - x_i(t) \cdot x_j(t), \\ !x_i(t) & \to & 1 - x_i(t). \end{array} \right\} \tag{2.6}$$

Since the set of u_j, $j = 1, 2, \ldots, m$, are the Boolean control sequences obtained via the reachability analysis, they are consequently time–dependent parameters. We approximate its values through piece–wise linear functions, which are then interpolated to obtain the value of the time–dependent terms as $\hat{f}_i\,(x_1, x_2, \ldots, x_n, u_1, u_2, \ldots, u_m)$.

Besides, we consider that the input-response function associated to each gene is the following logistic function:

$$\Theta\left[\hat{f}_i(x_1, x_2, \ldots, x_n, u_1, u_2, \ldots, u_m)\right] = \frac{1}{1 + \exp\left[-15\left[\hat{f}_i(x_1, x_2, \ldots, x_n, u_1, u_2, \ldots, u_m)\right] - 0.5\right]}. \tag{2.7}$$

This function also displays a dichotomic behavior for an $\epsilon = 0.5$, and $b >> 1$, as it has been previously stated on Remark 1.12. Finally, the numerical solution of the ODEs system is conducted using a standard solver provided by Matlab$^{\text{TM}}$.

In what follows we shall proceed to illustrate how to apply our proposed methodology, *i.e.* we illustrate how to use Algorithm 2. For this we shall consider some examples. In fact, we modify the autonomous systems that we have previously taken into consideration adding now the control inputs that model the interaction of the given regulatory network with the entourage.

2.4 Illustrative examples

When a control input is adequately added to an autonomous BN, what results from the study of the resulting non–autonomous systems provides quite interesting information. If interpreted in the right way, this information can show us how the network could be interacting with its entourage. Therefore, we can get some biologically–motivated insight into the functional mechanisms underlying patterns of connectivity in modularly structured GRNs. In this context, this section is intended to illustrate how to perform the reachability analysis on BCNs via the methodology summarized by Algorithm 2. In fact, we recall those three dynamical systems that have been previously introduced on Section 1.7, in order to explore now their structural reachability properties directly associated to specific control inputs.

First, Example 2.1 shows, step by step, the reachability analysis procedure, as well as the graphical information that can be obtained (*i.e.* Boolean/continuous transition plots and bifurcation diagrams). Next, Example 2.2 allow us to explore the reachability properties of the biological FOS-GRN, and thus elucidate the role that genes play in the context of the network as a whole. Finally, in Example 2.3 we explore the reachability properties of the dynamical systems that display a single cyclic attractor. *Let us then proceed.*

Example 2.1 The Control TwoFixed BN

Consider the following Boolean control system:

$$\Sigma_{BC\text{-}TwoFixed}: \quad \Big\{ x_i(t+1) = \bar{f}_i(x_1, x_2, x_3, u), \tag{2.8}$$

originated by the addition of a control input u at the i-th gene on the TwoFixedBN. If we separately test OR and AND operator on each gene of Eq. (2.8), the BCNs that arise are listed below:

Case 1: *u at x_1 with an OR operator:*

$$\Sigma_1 : \begin{cases} x_1(t+1) &= (x_2 \wedge \neg x_3) \vee u, \\ x_2(t+1) &= x_1 \wedge \neg x_3, \\ x_3(t+1) &= x_1 \wedge \neg x_2. \end{cases}$$

Case 2: *u at x_2 with an OR operator*

$$\Sigma_2 : \begin{cases} x_1(t+1) &= x_2 \wedge \neg x_3, \\ x_2(t+1) &= (x_1 \wedge \neg x_3) \vee u, \\ x_3(t+1) &= x_1 \wedge \neg x_2. \end{cases}$$

Case 3: *u at x_3 with an OR operator*

$$\Sigma_3 : \begin{cases} x_1(t+1) &= x_2 \wedge \neg x_3, \\ x_2(t+1) &= x_1 \wedge \neg x_3, \\ x_3(t+1) &= (x_1 \wedge \neg x_2) \vee u. \end{cases}$$

Case 4: *u at x_1 with an AND operator*

$$\Sigma_4 : \begin{cases} x_1(t+1) &= (x_2 \wedge \neg x_3) \wedge u, \\ x_2(t+1) &= x_1 \wedge \neg x_3, \\ x_3(t+1) &= x_1 \wedge \neg x_2. \end{cases}$$

Case 5: *u at x_2 with an AND operator*

$$\Sigma_5 : \begin{cases} x_1(t+1) &= x_2 \wedge \neg x_3, \\ x_2(t+1) &= (x_1 \wedge \neg x_3) \wedge u, \\ x_3(t+1) &= x_1 \wedge \neg x_2. \end{cases}$$

Case 6: *u at x_3 with an AND operator*

$$\Sigma_6 : \begin{cases} x_1(t+1) &= x_2 \wedge \neg x_3, \\ x_2(t+1) &= x_1 \wedge \neg x_3, \\ x_3(t+1) &= (x_1 \wedge \neg x_2) \wedge u. \end{cases}$$

Let

$$x(t) = x_1(t)x_2(t)x_3(t).$$

We can then express the BCNs Σ_i, $i = 1, 2, \ldots 6$, *into their algebraic forms, according to Eq. (2.2) as:*

$$x(t+1) = \tilde{L}_i x(t) u(t)$$

where $\tilde{L}_i \in \mathcal{L}_{8\times 16}$ *is the network transition matrix for the* i*-th BCN. Thus, for each case we have:*

$$\begin{aligned}
\tilde{L}_1 &= \delta_8[4\ 8\ 2\ 2\ 3\ 7\ 1\ 5\ 4\ 8\ 4\ 4\ 4\ 8\ 4\ 8],\\
\tilde{L}_2 &= \delta_8[6\ 8\ 2\ 2\ 5\ 7\ 5\ 5\ 6\ 8\ 2\ 4\ 6\ 8\ 6\ 8],\\
\tilde{L}_3 &= \delta_8[7\ 8\ 1\ 2\ 7\ 7\ 5\ 5\ 7\ 8\ 3\ 4\ 7\ 8\ 7\ 8],\\
\tilde{L}_4 &= \delta_8[8\ 8\ 2\ 6\ 7\ 7\ 5\ 5\ 8\ 8\ 4\ 8\ 8\ 8\ 8\ 8],\\
\tilde{L}_5 &= \delta_8[8\ 8\ 2\ 4\ 7\ 7\ 5\ 7\ 8\ 8\ 4\ 4\ 8\ 8\ 8\ 8],\\
\tilde{L}_6 &= \delta_8[8\ 8\ 2\ 2\ 7\ 8\ 5\ 6\ 8\ 8\ 4\ 4\ 8\ 8\ 8\ 8].
\end{aligned}$$

Then, we want to know if a certain destination state x_d *can be reached at the* s*-th step from* x_0. *Since we are mainly concerned about the reachable subsets related to the fixed–point attractors:*

$$\{\Lambda_1 = \delta_8^8, \Lambda_2 = \delta_8^2\},$$

these ones will be our initial as well as our final conditions. Since we restrict our exploration to the transitions from Λ_1 *to* Λ_2, *and vice versa, we name it as a characterization of the associated reachable subsets.*

Note that in Theorem 2.1, the reachable states are conditioned by the number of steps s, *so we fixed it using* r_0 *as an upper limit, namely* $1 \leq s \leq r_0$. *As shown in Chapter 1 in Section 1.7, the transient period for this network is* $r_0 = 3$. *Hence, we can calculate* $\tilde{L}^s x_0 \in \mathcal{L}_{2^n \times 2^{sm}}$ *for* $1 \leq s \leq 3$.

In order to show how the characterization of the reachable subsets occurs, let us consider the BCN on Example 2.1 given by Σ_4, *its network transition matrix* $\tilde{L}_4$ *and* $x_0 = \Lambda_2 = \delta_8^2$ *as an illustrative example. Then, according to Eq. (2.3):*

$$\begin{aligned}
s = 1, \quad \tilde{L}_4^1 x_0 &= \delta_8[2\ \ 6],\\
s = 2, \quad \tilde{L}_4^2 x_0 &= \delta_8[2\ \ 6\ \ 4\ \ 8],\\
s = 3, \quad \tilde{L}_4^3 x_0 &= \delta_8[2\ \ 6\ \ 4\ \ 8\ \ 5\ \ 5\ \ 8\ \ 8].
\end{aligned}$$

It is clear that at $s = 2$ *the state* $\delta_8^8 \sim (0,0,0)$ *can be reached, which corresponds to* Λ_1. *For this case, note that in the 4th column we have* δ_8^8, *which means that the control* δ_4^4 *can drive* $x_0 = \delta_8^2 \sim (1,1,0)$ *to* $x_d = \delta_8^8 \sim (0,0,0)$. *Thus, according to Eq. (2.4):*

$$u(0)u(1) = \delta_4^4.$$

Converting $4-4$ *to binary form yields* 00, *which means that the corresponding control input is*

$$u(0) = 0, \qquad u(1) = 0.$$

In short, we say that an attractor transition, from Λ_2 *to* Λ_1, *can be promoted in two steps by connecting a control input* u *at* x_1 *with an AND operator, under a Boolean sequence* $u(0) = 0$, $u(1) = 0$. *However, we cannot rule out those additional sequences that appear when* $s = 3$, *particularly if we are exploring biological GRNs since it may indicate relevant changes on gene concentration observed in real experiments. In fact, we will discuss this situation in further sections. Unless something different is specified, we use the smallest* s *such that the* x_d *is reached.*

As shown above, this process is repeated until all the six cases are explored. We summarize the results in a table, where the crossover of row and column contains the perturbed gene to transit from x_0 *to* x_d. *In bold letters, the control input switches-on the gene while the reminder switches-off. The (*) represents the same attractor as initial to final condition, so no control input is needed.*

x_0/x_d	Λ_1	Λ_2
Λ_1	*	$\mathbf{x_2}$
Λ_2	x_1, x_2, $\mathbf{x_3}$	*

Table 2.1: Available trajectories among all possible pairs of attractors in the Control TwoFixed BN.

Once we have identified the reachable subsets related to the two fixed–point attractors (see Table 2.1), we are able to produce composed transitions among states. Let us consider the trajectory:

$$\Lambda_2 \xrightarrow[3]{x_2} \Lambda_1 \xrightarrow[2]{\mathbf{x_2}} \Lambda_2,$$

which means in a compact notation, that starting from Λ_2*-attractor and under a suitable*

switching-off sequence of length 3 on x_2, Λ_1 will be eventually reached. Then, being on Λ_1 and a switching-on of length 2 x_2, Λ_2 will be finally reached, as shown in Fig. 2.1.a).

The Boolean sequences for the control inputs are

$$\begin{aligned} \Lambda_2 \xrightarrow[3]{x_2} \Lambda_1 : &\qquad u_1(0) = 0,\ u_1(1) = 1,\ u_1(2) = 1, \\ \Lambda_1 \xrightarrow[2]{\mathbf{x_2}} \Lambda_2 : &\qquad u_2(0) = 1,\ u_2(1) = 1, \end{aligned}$$

or simply:

$$u_1(t) = (0, 1, 1)$$

and:

$$u_2(t) = (1, 1).$$

However, since we are interested in plotting a composed trajectory, some additional points about control inputs must be taken into consideration:

- *If u is connected with an AND operator and for some reason we cannot affect the original Boolean function, we fix $u = 1$. Otherwise, if u is connected with an OR, we fix $u = 0$.*
- *To visualize the initial condition of the BCN in a plot, we fix the u_i as stated above.*
- *If a composed trajectory is plotted, the Boolean sequences of the control inputs must be increased according to its effect by each attractor transition.*

Taking into account what was stated before, the BCN that produces the desired trajectory can be described by:

$$\begin{aligned} x_1(t+1) &= x_2 \ \wedge\ \neg x_3, \\ x_2(t+1) &= ((x_1 \ \wedge\ \neg x_3) \wedge u_1) \vee u_2, \\ x_3(t+1) &= x_1 \ \wedge\ \neg x_2, \end{aligned} \tag{2.9}$$

where the suitable control sequences u_1 and u_2 are:

$$\begin{aligned} u_1(t) &= (1,0,1,1,1,1,1,1,1,1), \\ u_2(t) &= (0,0,0,0,0,0,0,1,1,0). \end{aligned} \tag{2.10}$$

Analyzing the u_1 by sections (which also holds for u_2):

$$u_1(t) = (\overbrace{1}^{a}, \overbrace{0,1,1}^{b}, \overbrace{1,1,1}^{c}, \overbrace{1,1}^{d}, \overbrace{1}^{e}),$$

where: (see Fig. 2.1.b)

- *a-section is the x_0 of the BCN;*
- *b-section is the action of control u_1;*
- *c-section simulates a stationary state;*
- *d-section is the action of control u_2;*
- *e-section is the x_d of the BCN.*

Finally, given Eqs. (2.9) and (2.10), it is possible to recover quantitative information about the Control TwoFixed BN through its continuous approximation, via Eqs. (2.5), (2.6) and (2.7). We separated the trajectory into two attractor transitions to find the minimum time of manipulation Δ_{t_j}, and also the enough extra time for all the genes to reach the new state (see Fig. 2.1.c). In this way a plot is generated, called as a bifurcation diagram, with the parameter value of Δ_{t_j} in the x-axis, whereas in the y-axis the total sum of the single gene expression values for the n genes. If a change in the sum occurs, this is the minimum time of manipulation to produce the transition. The time values found after the numerical solution of the ODEs system are presented as follows

Trajectory	*Minimum Δ_{t_j}*	*Extra time*
$\Lambda_2 \xrightarrow{x_2} \Lambda_1$	*0.92*	*4.08*
$\Lambda_1 \xrightarrow{\mathbf{x_2}} \Lambda_2$	*1.16*	*3.39*

In Fig. 2.1 d), we show the relative concentration after considering the minimum Δ_{t_j} and the extra–time. All the information is graphically summarized in Fig. 2.1.

■

We must point out that the dynamic analysis carried out in the previous example is repeated for the rest of examples but condensed. We can proceed now to our second illustrative example.

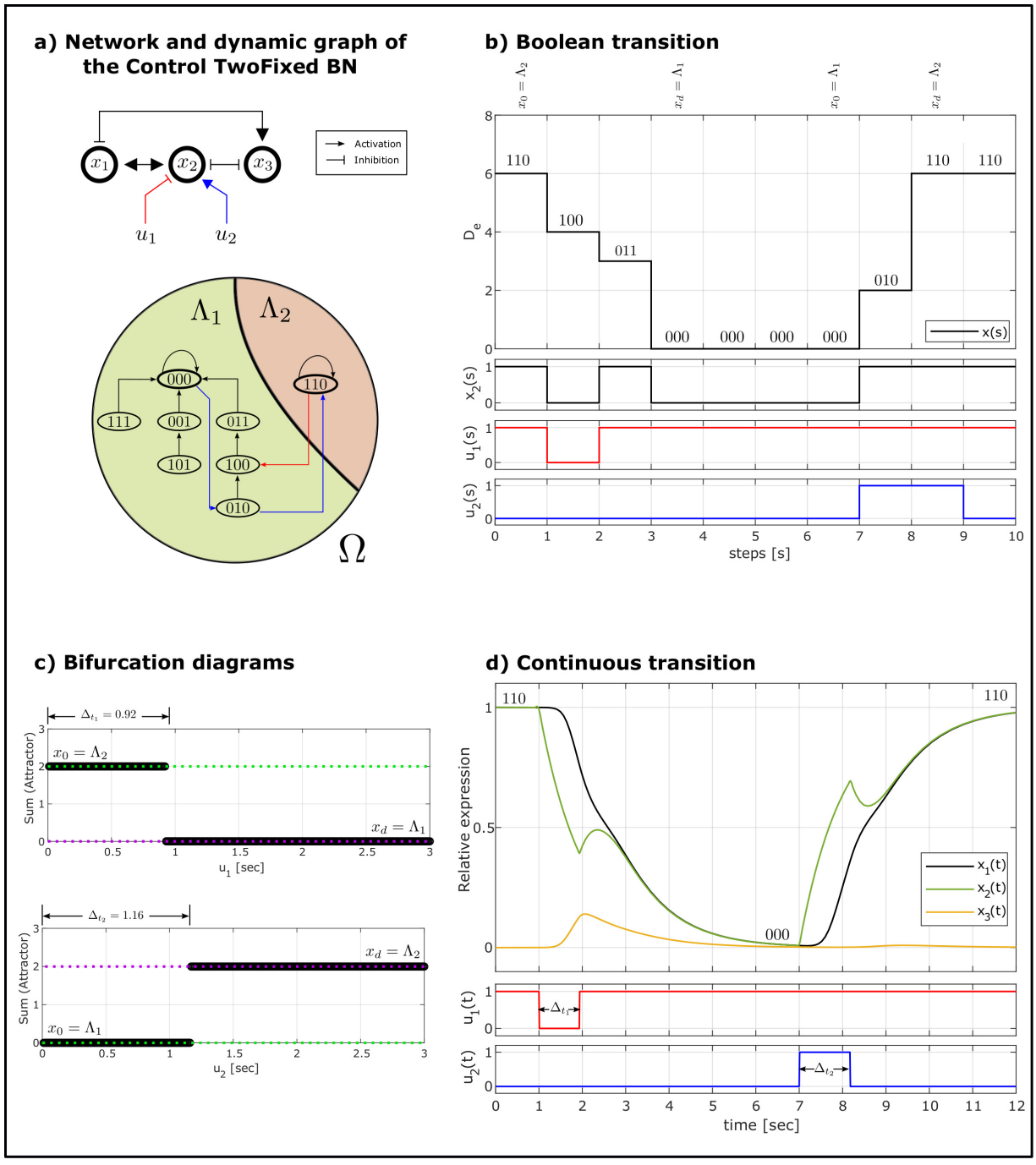


Figure 2.1: Dynamic analysis of the Control TwoFixed BN.

Example 2.2 The BC-FOS-GRN

In order to uncover the genes that promote attractor transitions in this example, we transformed the FOS-GRN into BC-FOS-GRN, by adding a control input on the i-th gene. We tested separately two possible Boolean operators (taking them as connectors) between original Boolean functions f_i and u to create the new Boolean function $\bar{f}_i$;

- *the OR operator (switch-on), and*

- *the AND operator (switch-off).*

Thus, we adopted a new system of Boolean equations:

$$\Sigma_{BC\text{-}FOS\text{-}GRN}: \quad \Big\{ x_i(t+1) = \bar{f}_i(x_1, x_2, \dots x_n, u), \qquad i = 1, 2, \dots, 13 \tag{2.11}$$

which now also depends on u. After that, we applied the STP transformation but now for the new system BC-FOS-GRN, and the system became a discrete-time bilinear system of the form

$$x(t+1) = \tilde{L}x(t)u(t),$$

where $\tilde{L}$ is a matrix of dimensions $2^{14} \times 2^{13}$, also called as the transition matrix of the controlled network. Then, to characterize the reachable subsets of the FOS-GRN given the set of attractors and the computed $\tilde{L}$, we performed the Algorithm 2 for $s = 1, 2, \dots, r_0$.

As we mentioned above, we called potential genes (result of reachability analysis) those ones that, changing their level of expression by the interaction with the control input u, promote a trajectory from one attractor to another. Particularly for this example, a total of 79 available trajectories were found on FOS-GRN -at least under the two connectors tested here- and they were summarized in Table 2.2. Interestingly, we first noticed that some trajectories between attractors are not allowed. For example, a considerable gap appeared on those transitions from floral primordia to inflorescence-like meristems, indicating that, it is difficult to return to a previous stage once the developmental program of flowering initiates. Hence, mutants that produce flower with inflorescence-like characteristics, for example ag-1 mutants grown in short day conditions [48], are probably not caused by a return to an inflorescence state but they may have a novel potentiality of the IM state now to choose between two cell-fate decisions: floral or inflorescence-like meristems, as had been demonstrated for other MADS mutants by the epigenetic landscape analysis of the

x_0/x_d	I_1	I_2	I_3	I_4	SE	PE1	PE2	ST1	ST2	CAR
I_1	*	**UFO**		**WUS**	**AP1**,**FT** EMF1,TFL1					**LFY**,**AP1** **FT** EMF1,TFL1
I_2	UFO	*	**WUS**				**AP1**,**FT** EMF1,TFL1		**LFY**,**AP1** **FT** EMF1,TFL1	
I_3		WUS	*	UFO					**LFY**,**AP1** **FT** EMF1,TFL1	
I_4	WUS		**UFO**	*						**LFY**,**AP1** **FT** EMF1,TFL1
SE					*	**AP3**	**UFO**			**AG**,**WUS** **TFL1** AP1,AP2
PE1					AP3,PI SEP,LFY	*	**UFO**	**AG**,**WUS** **TFL1** AP2		**TFL1** AP1
PE2						UFO	*		**AG**,**WUS** **TFL1** AP1,AP2	
ST1					**EMF1** AG,LFY			*	**UFO**	**EMF1** AP3,PI,SEP AG,LFY
ST2							**EMF1** AG,LFY	UFO	*	
CAR					**EMF1** AG,LFY			**AP3**	**UFO**	*

Table 2.2: Available trajectories among all possible pairs of attractors induced by potential genes. The crossover of row and column contains the gene to transit from x_0 to x_d. In bold letters, the control input switches-on the gene while the reminder switches-off. The (*) represents the same attractor as initial to final condition, so no control input is needed.

XAL2 regulatory network module [52].

On the other hand, we can interpret those available trajectories either as:

(1) *developmental*

(2) *or homeotic one.*

First, we focus on the trajectories that occur in normal development, for instance, inside the sub-zones of IM and from IM to floral primordia. Both UFO and WUS are necessary for the transitions of the sub–zones (I_i) of the IM, as a consequence of changes in the positional information caused by the displacement of the cells since the shoot apical meristem is constantly growing. Indeed, we found the developmental trajectory

$$I_4 \xrightarrow[1]{\boldsymbol{UFO}} I_3 \xrightarrow[1]{WUS} I_2 \xrightarrow[1]{UFO} I_1,$$

which means, in a compact notation, that starting from the I_4 attractor and switching-on

UFO (in bold letters), I_3 will be eventually reached, then being on I_3 and switching-off WUS, I_2 will be now reached, and finally, starting from I_2 attractor and switching-off UFO, I_1 will be the final attractor. Although the necessary steps to reach each x_d from x_0 are not specified, the shorter trajectory is hereafter considered for all the cases.

To illustrate the induced trajectories previously discussed, we generated Figure 2.2 with three mainly sections; on the top-left we present the control inputs that are added to the original network as well as the desired IM-to-IM transitions (from I_4 to I_1); on the bottom-left, the initial attractor state and its progressive change under the effect of the controllers (Boolean transition); while on the right-side the same before but for the continuous approximation of the network. Particularly, Figure 2.2.a) shows the graph of the desired transitions inside IM (from the youngest to the oldest sub-zone), red line indicates a switch-off, whereas blue line a switch-on. In Figure 2.2.b), the attractors (tagged with its D_e) are visited as expected, and such Boolean transitions only require one step to be induced by the controllers u_1, u_2 and u_3, respectively. On the other hand, to observe quantitative changes (i.e. the relative gene expression of the UFO and WUS on the induced trajectories), and evaluate if such attractor transitions are maintained on the continuous approximation of the model, we first plotted on Figure 2.2.c) the bifurcation diagrams for each separated transition. Notice that, interestingly, we reasoned that could be a direct equivalence of one step at least for these two genes UFO and WUS. Eventually, such IM's will give rise to FM's, and they will sub–differentiate in the flower organ primordia. It is has been reported in [10, 26, 44] that some of the characterized regulators which determine these meristematic cells are LFY, AP1 and AP2 while, on the other hand, the activity of the FM is counteracted by the gene TFL1 [50]. Thus, an expected attractor for floral meristems will include on its gene profile LFY=AP1=AP2=1 and TFL1=0. Although the FOS-GRN do not recover an attractor with the configuration of the floral meristem, we reasoned that the induced trajectories (by potential genes) from the inflorescence meristem to the floral organs may be passing by a transitory state that corresponds to the FM.

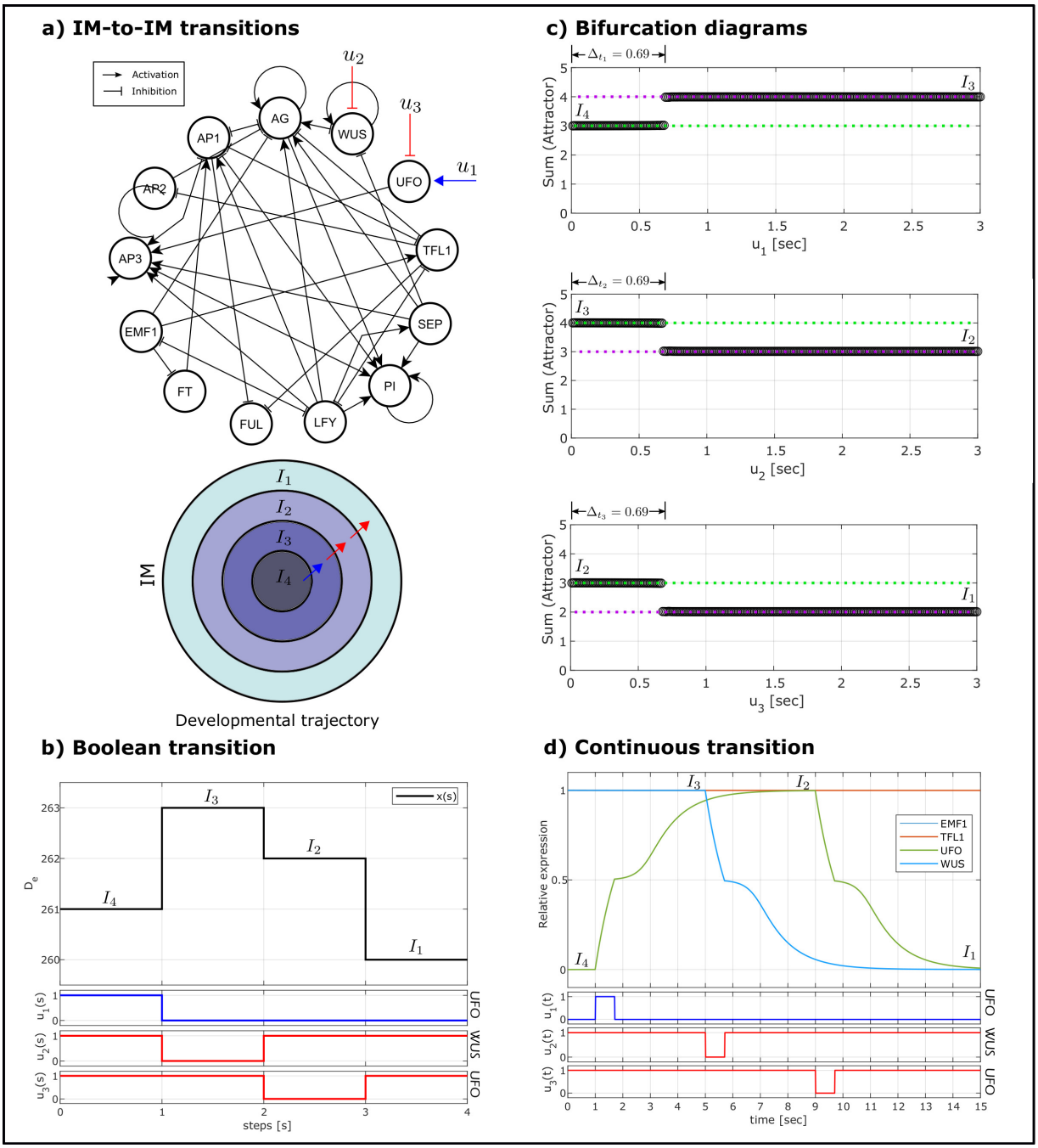


Figure 2.2: Developmental induced trajectories inside the IM. **a)** A trajectory from the youngest (I_4) to the oldest (I_1) sub–zones of the IM is tested on **b)** the Boolean model of the FOS-GRN, as well as on **d)** its approximated continuous–time model.

To test this, we selected trajectories (labeled with x_i, i=1,2,3,4) from:

$$x_1: \quad I_1 \xrightarrow[6]{\boldsymbol{FT}} SE$$
$$x_2: \quad I_2 \xrightarrow[7]{\boldsymbol{FT}} PE2$$
$$x_3: \quad I_3 \xrightarrow[8]{\boldsymbol{FT}} ST2$$
$$x_4: \quad I_4 \xrightarrow[8]{\boldsymbol{FT}} CAR$$

which are displayed on Figure 2.3 a). On the other hand, Figure 2.3.b) shows how each of the four cases of induced trajectories has a transitory state where AP1, AP2, LFY and TFL1 have the expected configuration for FM-like attractor (for further information about the transitory states, see Figure B.1 on Appendix B). Besides, according to Table 2.2, the potential genes AP1, LFY, EMF1 and TFL1 also work. While FT, AP1 and LFY must be switched-on, EMF1 and TFL1 must be switched-off. This is actually what happens during development, for example, long day grown plants, where FT is active, produce less secondary inflorescences than short day grown plants where FT is not transcribed [18]. But is even more interesting that floral homeotic genes such as AP3, PI and AG, despite being necessary, are not sufficient to form floral organs, in agreement with the previous hypothesis on the ABC model of floral patterning [23]. Actually, the control of AP1 in the 35S::AP1-GR ap1 cal plants that grow inflorescence meristems until the change to floral meristems when treated with dexamethasone, showed than only one treatment of dexamethasone is sufficient to form the four types of floral organs [62]. These experiments were done after the publication of the FOS-GRN [24], therefore, the reachability analysis can give us predictions on how biological modules could be controlled.

Secondly, we focus on the trajectories that can be interpreted as homeosis, which are associated to transitions between floral organ primordia (FOP). Unlike IM–to–FOP developmental trajectories induced by LFY, AP1, FT, EMF1 and TFL1, FOP-to FOP homeotic trajectories can be promoted by AP2, AP3, PI, AG and SEP, which are recognized by their mutant homeotic flower phenotypes. Even WUS and UFO are able to cause such transitions, showing that alterations in floral patterning can be caused by changing spatial cues in the meristem. Now, we take as an example of a control element a node not presented in the network, the BELLRINGER gene [12]. In the blr-4 missense mutant AG is dere-

pressed, and some flowers develop carpels in place of sepals in the first whorl. This whorl 1 and carpels are often fused and form a gynoecium that encloses the rest of the flower [9]. In fact, such switch-on of AG to transit from sepal to carpel primordium is found with the reachability analysis.

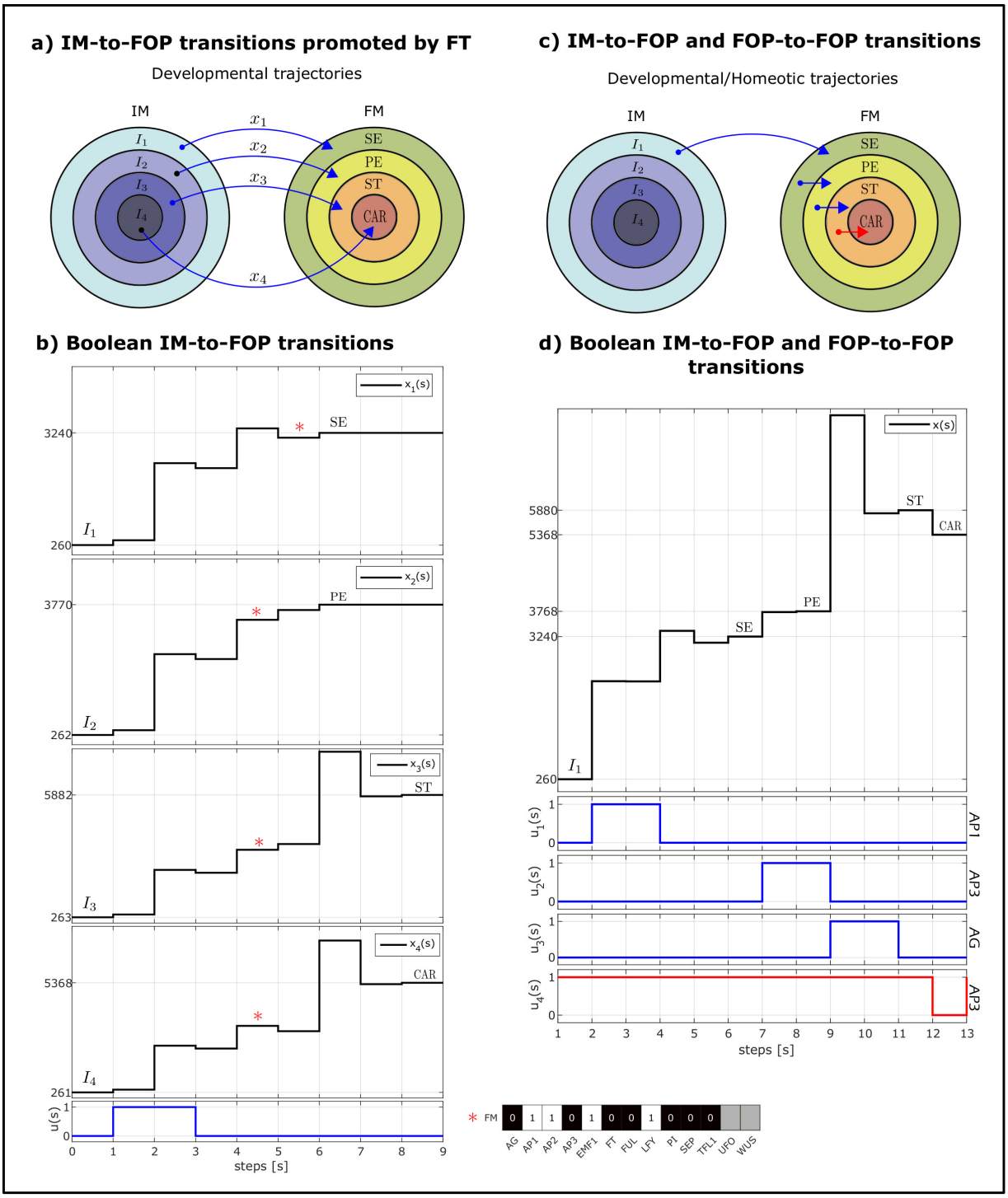


Figure 2.3: Developmental/Homeotic trajectories on the IM and FM

The role of potential genes is maintained on the continuous FOS-GRN model

Until this point, we were only interested in qualitative changes on the dynamics of the FOS-GRN under the effect of Boolean sequences connected to the potential genes. However, quantitative information as the minimum time of manipulation of these potential genes and their impact on the relative decay/saturation rates of the rest of the genes, have not been explored yet. To illustrate this, we first selected a composed developmental/homeotic trajectory

$$I_1 \xrightarrow[5]{\boldsymbol{AP1}} SE \xrightarrow[2]{\boldsymbol{AP3}} PE1 \xrightarrow[3]{\boldsymbol{AG}} ST1 \xrightarrow[1]{AP3} CAR,$$

which mimics the sequence of attractor transition previously recovered by different levels of noise in the updating rules of the network [5]. First, we tested the control inputs obtained through reachability analysis on the BC-FOS-GRN (see Figure 2.3.c), only to verify that the expected composed transition occurs. Then, using Eqs. (2.5), (2.6) and (2.7), we obtained the continuous version of the BC-FOS-GRN. In order to find the minimum time of manipulation of the potential genes by the control inputs u_j in the BC-FOS-GRN, we separated the composed developmental/homeotic trajectory into four individual transitions, and consequently four different ODEs system were also approximated as stated before. For each attractor transition, we take the initial attractor as x_0 of the ODEs system, and we then increased the duration of the j-th control input (Δ_{t_j}), interpolating its value to obtain the value of time-dependent terms at a specified time. Moreover, we considered an extra time for each interval Δ_{t_j} in such a way that the potential gene and the rest of the genes have enough time to relax to its new state after the manipulation by the control input. The resulting time values are

Trajectory	*Minimum* Δ_{t_j}	*Extra time*
$I_1 \xrightarrow{\boldsymbol{AP1}} SE$	*1.10*	*5.5*
$SE \xrightarrow{\boldsymbol{AP3}} PE1$	*1.61*	*5.5*
$PE1 \xrightarrow{\boldsymbol{AG}} ST1$	*0.61*	*5.5*
$ST1 \xrightarrow{AP3} CAR$	*0.71*	*5.5*

note that the continuous functions that changed by the addition of the u_j are

$$\begin{aligned}\hat{f}_{AP1}(\cdot) &= \hat{f}_{AP1}(\cdot) + u_1 - \hat{f}_{AP1}(\cdot) \cdot u_1 \\ \hat{f}_{AP3}(\cdot) &= \left(\hat{f}_{AP3}(\cdot) + u_2 - \hat{f}_{AP3}(\cdot) \cdot u_2\right) \cdot u_4 \\ \hat{f}_{AG}(\cdot) &= \hat{f}_{AG}(\cdot) + u_3 - \hat{f}_{AG}(\cdot) \cdot u_3\end{aligned}$$

.

Figure 2.4.a) shows the graphs obtained for each attractor transition, and as we can see, despite AP3 promotes trajectories from SE to PE1 and from ST1 to CAR, the necessary Δ_{t_j} is different. We finally integrated all the individual transitions in Figure 2.4.b), in the x-axis we show the relative concentration of the 13 genes involved in the FOS-GRN, and in the y-axis the time intervals for the four control inputs. This plot suggests that there exists a minimum trigger threshold which, once it has been crossed, provokes an imminent attractor transition.

■

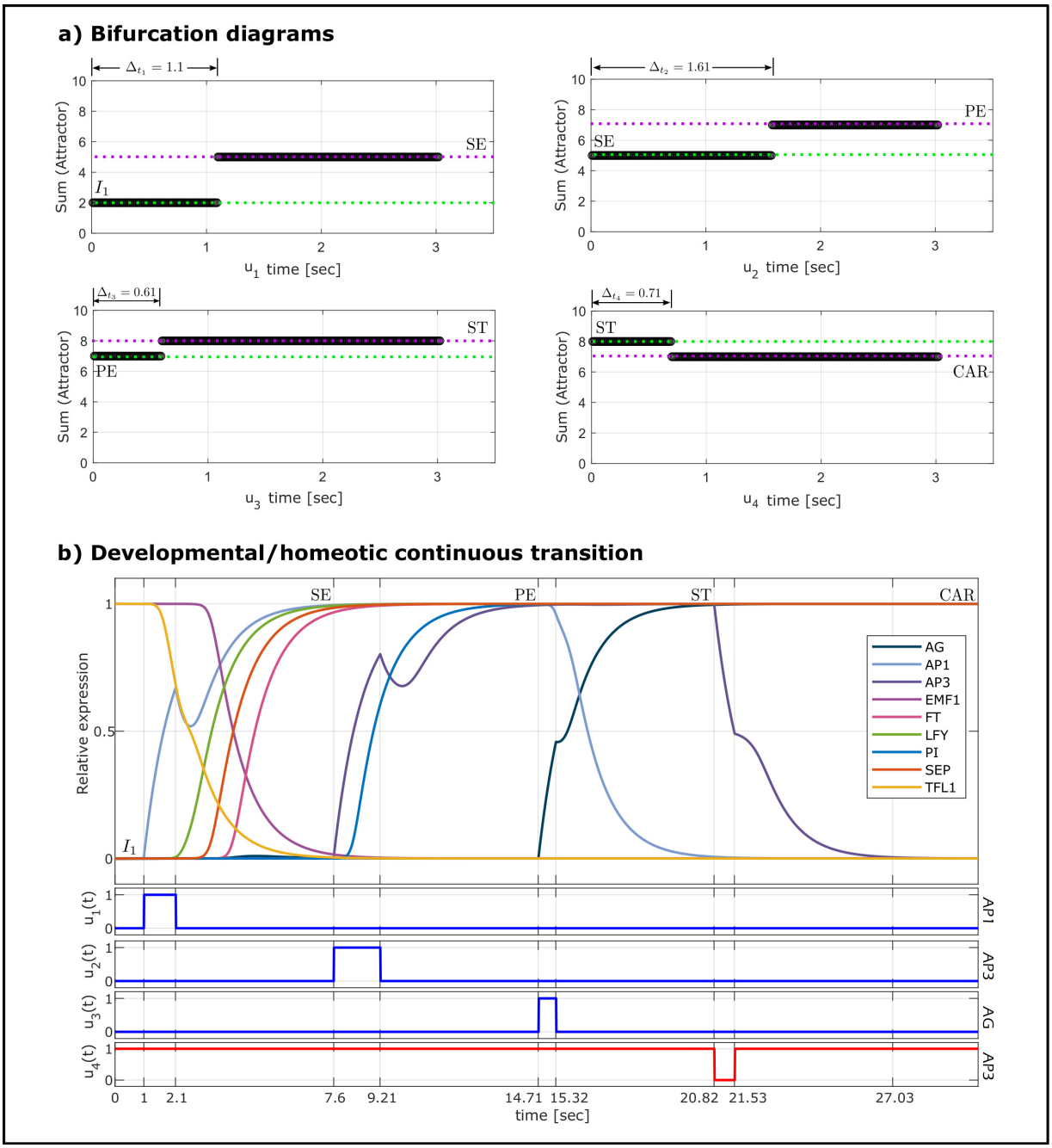


Figure 2.4: A developmental/homeotic trajectory in the FOS-GRN that mimics the sequence of gene profiles observed in real flowers. The trajectory I_1 – **AP1** →SE– **AP3** →PE1– **AG** →ST1–AP3 →CAR is tested.

In what follows we shall consider the reachability analysis of controlled gene regulatory networks that display structural cyclic behavior.

Example 2.3 The BC-ACC-GRN

In this example, we show that the reachability analysis can be also applied on networks that have a single cyclic attractor. In this context, we can distinguish four different main purposes of this analysis: (1) to shorten the cycle, (2) to lengthen the cycle, (3) to maintain a specific state of the cycle or (4) to avoid states of the cycle. Particularly for the ACC-GRN, we are interested in exploring the last two purposes since they have biological relevance in CC-arrest and endocycle in Arabidopsis thaliana, respectively.

Consider the following Boolean control system

$$\Sigma_{BC\text{-}ACC\text{-}GRN}: \quad \Big\{ x_i(t+1) = \bar{f}_i(x_1, x_2, \dots x_n, u), \qquad i = 1, 2, \dots, 15 \tag{2.12}$$

which is the result of the addition of u on the i-th gene of the ACC-GRN (hereafter referred as BC-ACC-GRN). After applying the STP transformation, the system is given by its algebraic form:

$$x(t+1) = \tilde{L}x(t)u(t),$$

where $\tilde{L} \in \mathcal{L}_{2^{16}\times 2^{15}}$, also called as the transition matrix of the BC-ACC-GRN. In contrast with the results of the reachability analysis of networks with fixed–point attractors, here the potential gene continuously changes since every step of time a new state is reached by its own cyclic dynamics. Nevertheless, we discriminated those trajectories that arise as a consequence of this phenomena, and only take into consideration those ones that are particularly relevant for our main objectives: endocycle and CC-arrest.

In this direction, we tested two different trajectories for endocycle:

$$G_2 - 1 \xrightarrow[4]{E2Fe} G_1 - 1$$

and:

$$G_2 - M \xrightarrow[3]{\boldsymbol{APCC}} G_1 - 1$$

The corresponding results are shown in Figure 2.5. Note that the length of the switching-off sequence of E2Fe is greater than the one of the switching-on sequence of APCC by one step.

On the other hand, we also found that a low-level of E2Fa or a high-level of MYB3R14, originated by an external control input u, is sufficient to drive the system from G_1-stage to CC arrest (an extended G_1-stage). Such CC-arrest holds as long as the duration of the action u. To illustrate this, let us consider the following induced trajectories:

$$G_1 - S \xrightarrow[hold]{E2Fa} S$$

and:

$$G_1 - 1 \xrightarrow[hold]{\boldsymbol{MYB3R14}} G_1 - 2$$

The graphical results of both induced disruptions (CC arrest and endocycle) are presented in Figure 2.5. Finally, the continuous approximation of the BC-ACC-GRN, via Eqs. (2.5), (2.6) and (2.7), when genes E2Fa and MYB3R14 are perturbed is shown in Figure 2.6.

■

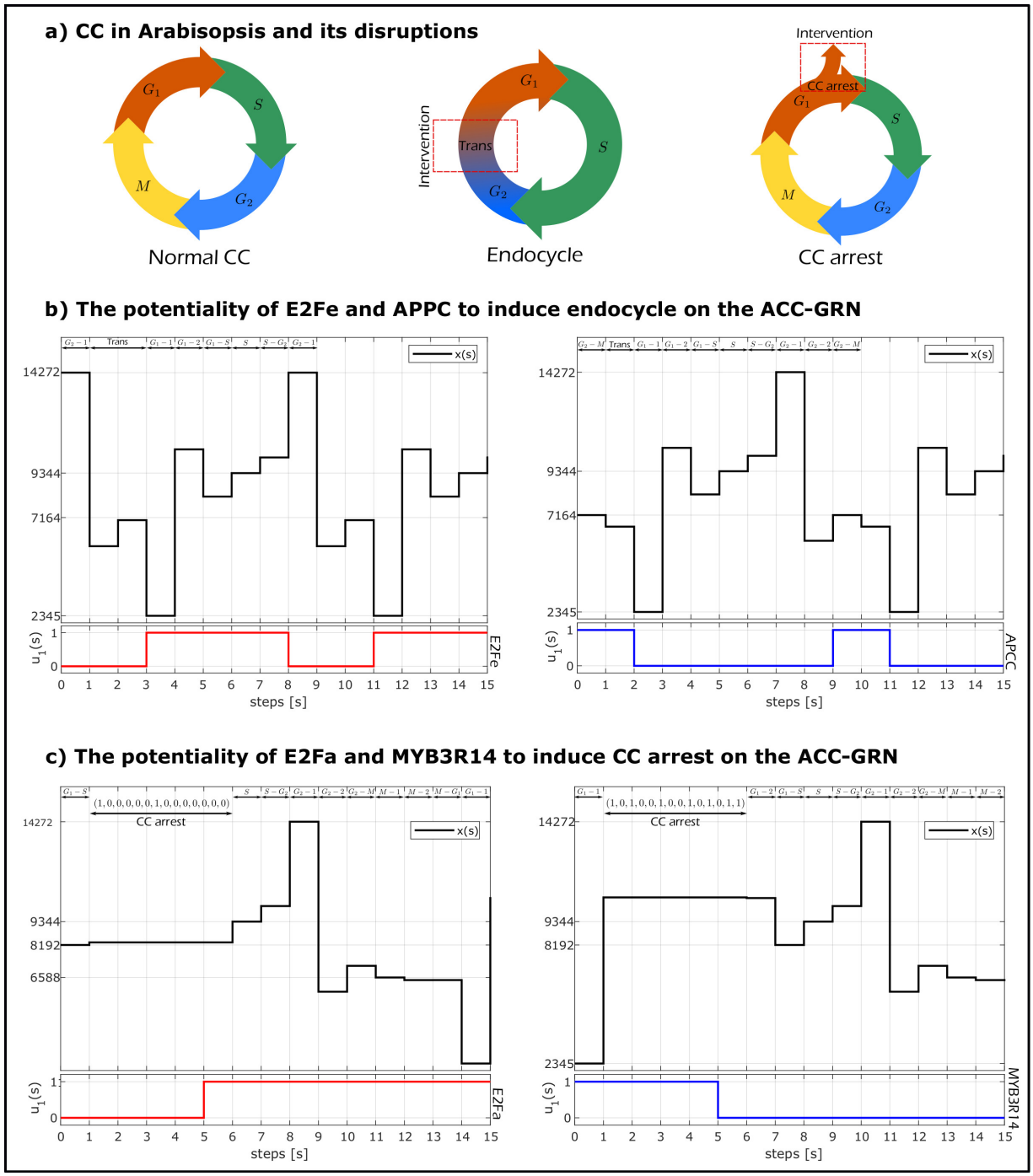


Figure 2.5: Induced disruptions on the BCC-ACC-GRN

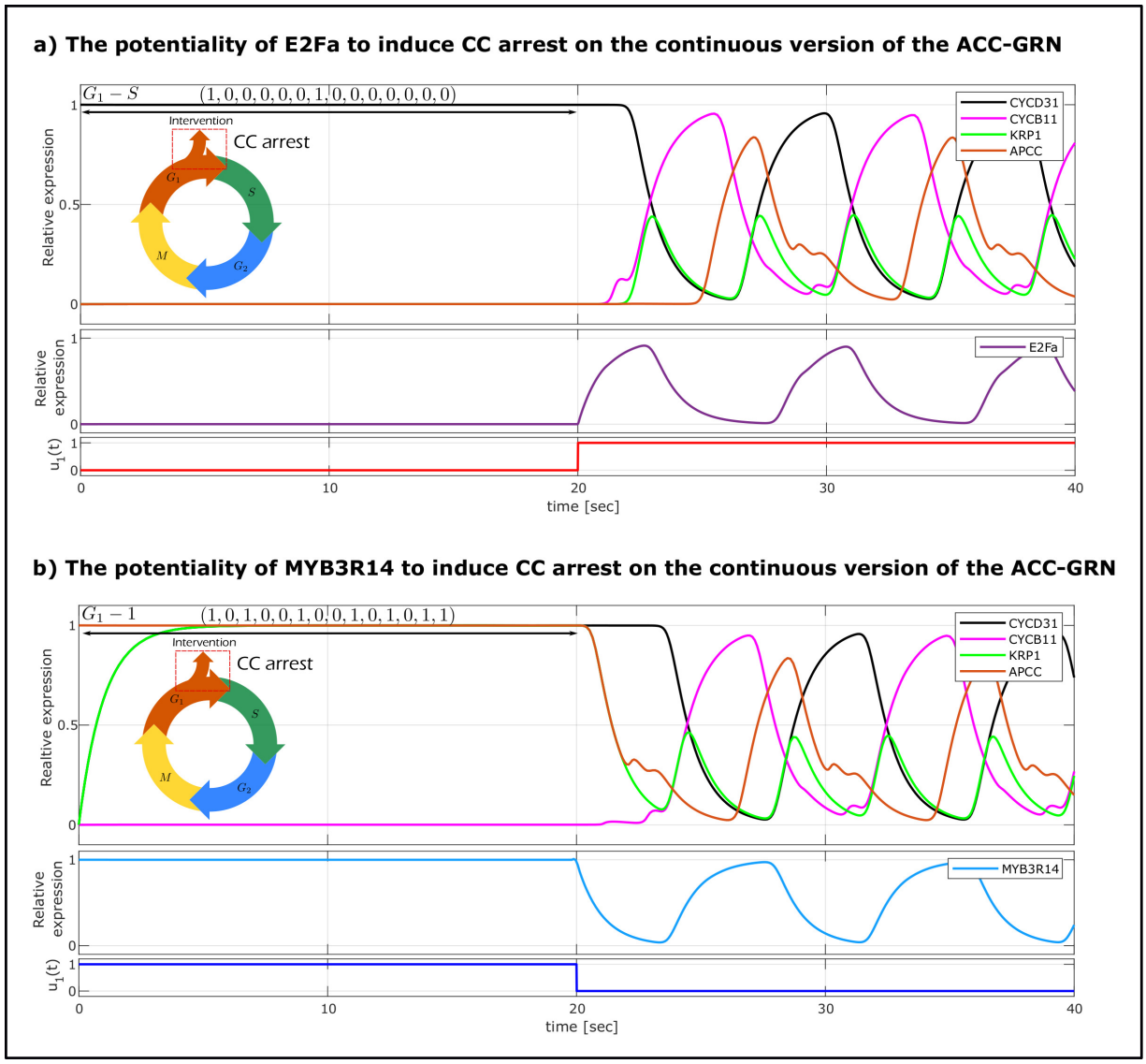


Figure 2.6: Induced disruptions promoted by potential genes on the continuous version of the BC-ACC-GRN.

As can be seen with the considered examples, the reachability analysis is an useful tool to uncover some of the phenomena that underlie developmental dynamics. To be more specific, reachability analysis allows the determination of the key role that play specific genes some well–known observed transitions (as illustrated with the examples that deal with the FOS-GRN model and the cell cycle). We can now discuss some aspects of the biological significance of what we have just shown.

2.5 Discussion

A common issue in control theory is to find a control input which drives a dynamical system from an initial state to a desired final state in finite time. In fact, there exists several ways to find such input depending on the nature of the given dynamical system. For example, in some multi–stable systems, suitable variations on a single parameter are sufficient to change from one steady–state (or attractor) to another. In this same direction, such transitions in experimental grounded Gene Regulatory Networks (GRNs), promoted by an external *stimuli* in a particular gene, originated from responses of changes in physiological, environmental and/or developmental cues, could implicate a change in cell–fate decision during development. Some authors have proposed different methodologies to capture the effect of changes on the level of expression of the genes by external elements in GRNs that underlie realistic cases of cell differentiation, *e.g.*:

- bifurcation analysis of a single parameter [19, 35],
- stochastic combinatorial Boolean interventions [57],
- uncertainty in the updating Boolean rules [5],

but fewer have used mathematical approaches based on underlying structural properties given by their topology and their dynamics.

Here, we propose an analytical procedure which harnesses the structural reachability property of a low–dimensional synchronous Boolean GRN. As concrete study cases, we use the regulatory module underlying floral organ determination in *Arabidopsis thaliana* during early stages of flower development (FOS-GRN), since it constitutes a model–of–choice for the study of differentiation and morphogenesis in multicellular organisms [5, 19]; as well as the gene regulatory module proposed by Ortiz-Gutierrez and collaborators in [49] which is involved in the cell cycle in *Arabidopsis thaliana*, to show the effect of external *stimuli* in promoting disruptions on the cyclic behavior. Our proposal is built around the Semi–Tensor Product approach [17] (STP), which led us to transform the BC-FOS-GRN and the BC-ACC-GRN via the inclusion of an external Boolean input to a specific gene. This results in each case in obtaining a discrete–time bilinear systems. Now, the new network transition matrix $\tilde{L}$ encompasses structural reachability properties of the Boolean Control GRNs. We say that a state is *reachable* from an initial state, if a qualitative change

on the state–space configuration is observed when an exogenous Boolean input is connected to a specific gene, modifying its state to describe knockout (On) or over–expressed (Off) states, and changing the corresponding updating rule for the gene during simulation. In spite of the advances in reachability analysis of GRNs [33], the exploration of the reachable subsets corresponding to attractors in experimental grounded GRN, that explains the role of the genes on such attractor transitions, has been left behind. So, we focus exclusively on attractor transitions and thus we limit the scope of our conclusions.

In a more general context, reachability analysis could also provide some useful tools to uncover the basic generic principles that underlie chronic–diseases in humans. This because the onset and the progression of chronic–degenerative diseases involves the disruption of some gene regulatory networks. Nowadays, aspects related to chronic–degenerative diseases, implicate cell undifferentation and cell reprogramming which result from abnormalities in the cell regulation. Therefore, the elimination of health problems in this context is also subjected, necessarily, to techniques as the design of strategies of cell reprogramming in order to return cells to a suitable state or to revoke the *unhealthy* trajectory. Thus, our proposed methodology can be useful to explore some potential state–space system based therapeutic strategies. At this point in the discussion we leave this possibility as an open question, but we will return to this issue at the end of the present manuscript when we address the epithelial—to–mesenchymal transition in the context of epithelial cancer.

2.6 Conclusion

In this chapter we presented a computational analysis that harnesses the structural reachability properties of B-GRNs, in order to promote transitions among pairs of attractors, which we named as reachability analysis. The proposed framework consisted of carrying on a systematic exploring of the potentiality of single genes to promote qualitative changes on the state–space configuration. The considered approach is focused on the connection of a control input u designed to arbitrarily overrules the original regulatory Boolean function of the concerned gene. Thus, we mimicked spontaneous switching actions on such genes through Boolean OR and AND operators, that gain experimental relevance since they have the same effects as induced knock-ins and knock-outs in real experiments, respectively. We showed how biological insights are derived by applying this methodological framework to

experimentally grounded B-GRNs that may have multiple fixed–point attractors (*i.e.* the FOS-GRN) or a single cyclic attractor (*i.e.* the ACC-GRN). Future studies should explore if the results of this analysis can be also applied on combinations of both aforementioned types of attractors.

Additionally, we showed that the role that genes play in attractor transitions are also maintained on the continuous version of GRNs, recovering not only qualitative but also quantitative information about their potentiality. In both the considered examples, which is to say FOS-GRN and ACC-GRN, this is particularly relevant since our findings suggests that there exists a strong relationship between the impact of specific genes in the system dynamics of the network, and their observed biological function during developmental processes.

Chapter 3

Computational controllability analysis of Boolean Control GRNs

Introduction

In the previous chapter of this manuscript it has been shown that reachability analysis applied on BC-GRNs provides biological insights into the specific role that also specific genes play in the context of network as a whole. This has been illustrated with the core gene regulatory network that underlies flower order specification in *Arabidopsis thaliana* and also with the network that regulates cell cycle. We must point out that until this point, the systematic exploration has been only focused on attractor transitions (either fixed–point or cyclic attractors). Nevertheless, a natural question arises:

> *Is it possible to expand this result to all 2^n distinct gene activation configurations, significant when tackling dynamic transitions among both stationary as well as transitory states?.*

In this direction, such question is intended to be answered throughout this chapter via the controllability matrix C for BCNs, proposed by Cheng and collaborators in [66], since its entries c_{ij} denote if the i-state is reachable from the j-state in finite time, under a set of admissible control inputs u. Considering this, our contribution here is mainly aimed at exploring the potentiality of the n-genes of BC-GRNs through the *computational controllability analysis*, resulting in a 3D array of k-controllability matrices, where $k = 1, 2, \ldots, n$, or in a compact notation C_k.

The proposed extension C_k contains all the dynamic information that characterizes the available induced trajectories promoted by potential genes, among not only pairs of attractors but also with transient states. In fact, such trajectories are classified into three main categories:

- firstly, the attractor to attractor transitions (A-to-A), that are equal to those ones obtained through the reachability analysis;
- secondly, the basin to attractor transitions (A-to-B), or vice versa (B-to-A);
- and thirdly, the basin to basin transitions (B-to-B).

We must point out that the last two cases, indeed, gain biological relevance since such dynamics are associated to the onset and progression of cell-state trajectories involved in disease (*i.e.* the Ephitelial-to-Mesenquimal Transition, further discussed in detail). Moreover, in this chapter we will discuss and compare the impact of Boolean sequences connected to specific genes, that results from our controllability analysis, with the effect of relative gene decay variations proposed by Davila-Velderrain *et al.* [19], concluding that both are complementary approaches. For instance, a control input connected with an AND operator that holds over time, is similar/equivalent to an induced gene decay rate. On the contrary:

> In cases where a suitable Boolean sequence, also referred as *train of information*, is necessary to promote an specific state trajectory, such transition cannot be induced by using a single gene decay variation.

Finally, taking as a case–of–study the Control TwoFixed BN, we deal with some illustrative examples in the last section in order to illustrate the applicability of our proposed computational controllability analysis on complex biological GRNs by exploring A-to-A, A-to-B, B-to-A, B-to-B and combinations of these transitions. We conclude the chapter discussing the potential relevance of this contribution from a biological perspective.

3.1 A brief summary on Boolean algebra

In this section, we briefly review the Boolean algebra introduced by Daizhang Cheng and collaborators in [17], which allows the computation of the controllability matrix of BCN. Basically, it is summarized in three main Boolean operations:

1. If $x, y \in \mathcal{D}$, it is defined Boolean addition and the Boolean product as follows

$$\begin{aligned} x +_{\mathcal{B}} y &= x \vee y \\ x \times_{\mathcal{B}} y &= x \wedge y \end{aligned}$$

   $\{\mathcal{D}, +_{\mathcal{B}}, \times_{\mathcal{B}}\}$ forms a Boolean algebra.

2. Consider $X = (x_{ij}) \in \mathcal{B}_{m\times n}$ and $Y = (y_{ij}) \in \mathcal{B}_{m\times n}$. It is defined the matrix Boolean addition as

$$X +_{\mathcal{B}} Y := \left(x_{ij} +_{\mathcal{B}} y_{ij}\right)$$

3. Consider $X \in \mathcal{B}_{m\times n}$ and $Y \in \mathcal{B}_{n\times p}$. It is defined the matrix Boolean product as

$$X \times_{\mathcal{B}} Y := Z \in \mathcal{B}_{m\times p},$$

   where

$$z_{ij} = \sum_{k=1}^{n}{}_{\mathcal{B}}\, x_{ik} \times_{\mathcal{B}} y_{kj}$$

In cases where $X \in \mathcal{B}_{n\times n}$, then

$$X^{(2)} := X \times_{\mathcal{B}} X$$

3.2 The controllability matrix $\mathcal{C}$ for BC-GRNs

From a Control Theory perspective, a dynamic system is said to be *controllable* at the time t_0, if it is possible to drive the initial state $x(t_0) = x_0$ to any other state $x(t_f) = x_d$, under a set of admissible controls $u(t)$, in a finite time interval $(t_f - t_0)$ [38]. This mathematical concept is clearly consistent with the question that was left open in the last chapter:

> *Is it possible to explore the reachable sets of all 2^n gene activation configurations of the BC-GRN?.*

In order to tackle this question let us consider a BCN with n nodes and m input nodes.

Such a network is described in dynamical terms as follows:

$$\Sigma : \begin{cases} x_1(t+1) &= f_1(x_1, x_2, \ldots, x_n, u_1, u_2, \ldots, u_m), \\ x_2(t+1) &= f_2(x_1, x_2, \ldots, x_n, u_1, u_2, \ldots, u_m), \\ \vdots & \vdots \quad \vdots \\ x_n(t+1) &= f_n(x_1, x_2, \ldots, x_n, u_1, u_2, \ldots, u_m), \end{cases}$$

where $x_i \in \mathcal{D}$, and $f_i : \mathcal{D}^{n+m}$, $i = 1, 2, \ldots n$ are Boolean functions. Setting $x(t) = \ltimes_{i=1}^{n} x_i$, $u(t) = \ltimes_{j=1}^{m} u_j$, then its algebraic form is:

$$x(t+1) = Lu(t)x(t), \tag{3.1}$$

where $L \in \mathcal{L}_{2^n \times 2^{n+m}}$. Then

$$\mathcal{C} := \sum_{s=1}^{2^{m+n}}{}_{\mathcal{B}} \sum_{i=1}^{2^m}{}_{\mathcal{B}} \mathrm{Blk}_i \left(L^{(s)}\right) = \sum_{s=1}^{2^{m+n}}{}_{\mathcal{B}} M^{(s)} \in \mathcal{B}_{2^n \times 2^n} \tag{3.2}$$

is the *controllability matrix* proposed by Cheng and collaborators in [66], where its entries c_{ij} indicate whether the i-state is reachable from the j-state, under a set of admissible Boolean inputs u at the s-th time step, if and only if $c_{ij} > 0$. Besides, $\sum_{\mathcal{B}}$ and $L^{(s)}$ denote an iterative Boolean addition ($+_{\mathcal{B}}$) and Boolean product ($\times_{\mathcal{B}}$), respectively (see Appendix 3.1 for a brief review of Boolean algebra). Following [66], the three fundamental properties of matrix $\mathcal{C}$ are summarized as follows:

1. $x_d = \delta_{2^n}^i$ is reachable from $x_0 = \delta_{2^n}^j$ if and only if $c_{ij} > 0$.
2. The system is controllable at $x_0 = \delta_{2^n}^j$ if and only if $\mathrm{Col}_j(\mathcal{C}) > 0$.
3. The system is controllable if and only if $\mathcal{C} > 0$.

Remark 3.1 *We must point out that in cases where the inequalities:*

- $c_{ij} > 0$ *(for matrix elements),*
- $Col_j(\mathcal{C}) > 0$ *(for column of a matrix) or*
- $\mathcal{C} > 0$ *(for full matrices)*

are used, it means that their entries are positive, otherwise not, since they are the result of binary operations.

In what follows, we provide an academic example to illustrate the information that can be obtained from the computed controllability matrix of a BCN.

From now until the end of this chapter, we will be showing each result by taking the Control TwoFixed BN as an illustrative example.

Example 3.1 ***The Control TwoFixed BN: Computation of $\mathcal{C}$***

Recall Example 2.1. We shall consider the particular case when the dynamic system is perturbed on the gene x_3:

$$\Sigma_4 : \begin{cases} x_1(t+1) &= (x_2 \ \wedge \ \neg x_3) \wedge u, \\ x_2(t+1) &= x_1 \ \wedge \ \neg x_3, \\ x_3(t+1) &= x_1 \ \wedge \ \neg x_2. \end{cases}$$

Setting $x(t) = \ltimes_{i=1}^{3} x_i$, $u(t) = u$, we have:

$$x(t+1) = Lu(t)x(t), \tag{3.3}$$

where:

$$L = \delta_8[8 \ 2 \ 7 \ 5 \ 8 \ 4 \ 8 \ 8 \ 8 \ 6 \ 7 \ 5 \ 8 \ 8 \ 8 \ 8].$$

Suppose that we are interested in answering the following questions:

1. *Is δ_8^8 reachable from $x_0 = \delta_8^2$?*

First, according to Eq. (3.2) and considering $s = 1$ *and* $s = 2$*, we have*

$$M^{(1)} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{bmatrix} \quad \textit{and} \quad M^{(2)} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}.$$

We can extract then the elements:

$$\left(M^{(1)}\right)_{82} = 0, \qquad \textit{and} \qquad \left(M^{(2)}\right)_{82} > 0,$$

which means that $x(2) = \delta_8^8 \sim (0,0,0)$ *is reachable from* $x(0) = \delta_8^2 \sim (1,1,0)$ *at the second step. This result is clearly the same that we obtained when considering the reachability analysis in Example 2.1.*

2. Is the system controllable, or controllable at any point?

After an straightforward computation, we can check the controllability matrix:

$$\mathcal{C} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}$$

According to the properties of the controllability matrix given above, we concluded that

the system is not controllable. However, $x_d = \delta_8^8$ *is reachable from any* x_0.

■

Remark 3.2 *Note that, this is the controllability matrix only for one particular case of Example 2.1. We propose in what follows a systematic analysis in the next section, referred in this work as controllability analysis, to explore the potentiality of individual genes in Boolean GRNs models when perturbed, involved on transitioning among pairs of states of the network. Consequently, a set of* $\mathcal{C}$ *that rises from the exploration of gene by gene is taken into consideration.*

3.3 An extension of $\mathcal{C}$

Consider a BCN described as:

$$\Sigma : \quad \Big\{ x_i(t+1) = f_i(x_1, x_2, \dots x_n, u), \qquad i = 1, 2, \dots, n, \tag{3.4}$$

originated by the addition of a control input u at the $i-th$ gene. The proposed controllability analysis, which harnesses the structural controllability properties of the network, consists of five main steps:

- *Step 1.* Initialize $k = 1$.
- *Step 2.* Add a control input u on the k-th node of the BN. If $k = n$, go to Step 5; otherwise, go to the next step.
- *Step 3.* Convert the BCN into its algebraic form given by Eq. (3.1), and then compute the k-th controllability matrix $\mathcal{C}_k$ by using Eq. (3.2).
- *Step 4.* Set $k = k + 1$ and go back to Step 2.
- *Step 5.* Save $\mathcal{C}_k$, $k = 1, 2, \dots, n$. Stop.

After a straightforward computation, the set of C_k, $k = 1, 2, \dots, n$ contains all the *available state trajectories* among pairs of states, when the k-th gene is perturbed, if and only if $(c_{ij})_k > 0$. Later on, we incorporate this procedure into the last algorithm of this thesis work.

Let us now illustrate our approach via an illustrative example.

Example 3.2 ***The Control TwoFixed BN: Computation of*** $\mathcal{C}_k$

Let consider again our previous Example 2.1. For this we build six controlled networks Σ_i, $i = 1, 2, \ldots 6$. *Each one of these networks comes as a consequence of the integration of the exogenous input to each of the involved genes in the given autonomous network through both the Boolean OR operator and the AND operator. For the resulting controlled networks we compute the corresponding controllability matrix as follows:*

- *First case:* u *connected to the system via the OR operator:*

$$\Sigma_1 : \left\{ \begin{array}{rcl} x_1(t+1) & = & (x_2 \wedge \neg x_3) \vee u \\ x_2(t+1) & = & x_1 \wedge \neg x_3 \\ x_3(t+1) & = & x_1 \wedge \neg x_2 \end{array} \right\} \Longrightarrow \mathcal{C}_1 = \begin{bmatrix} 1 & 0 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}$$

$$\Sigma_2 : \left\{ \begin{array}{rcl} x_1(t+1) & = & x_2 \wedge \neg x_3 \\ x_2(t+1) & = & (x_1 \wedge \neg x_3) \vee u \\ x_3(t+1) & = & x_1 \wedge \neg x_2 \end{array} \right\} \Longrightarrow \mathcal{C}_2 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}$$

$$\Sigma_3 : \left\{ \begin{array}{rcl} x_1(t+1) & = & x_2 \wedge \neg x_3 \\ x_2(t+1) & = & x_1 \wedge \neg x_3 \\ x_3(t+1) & = & (x_1 \wedge \neg x_2) \vee u \end{array} \right\} \Longrightarrow \mathcal{C}_3 = \left[\begin{array}{cccccccc} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{array} \right]$$

- *Second case: u connected to the system via the AND operator:*

$$\Sigma_4 : \left\{ \begin{array}{rcl} x_1(t+1) & = & (x_2 \wedge \neg x_3) \wedge u \\ x_2(t+1) & = & x_1 \wedge \neg x_3 \\ x_3(t+1) & = & x_1 \wedge \neg x_2 \end{array} \right\} \Longrightarrow \mathcal{C}_4 = \left[\begin{array}{cccccccc} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{array} \right]$$

$$\Sigma_5 : \left\{ \begin{array}{rcl} x_1(t+1) & = & x_2 \wedge \neg x_3 \\ x_2(t+1) & = & (x_1 \wedge \neg x_3) \wedge u \\ x_3(t+1) & = & x_1 \wedge \neg x_2 \end{array} \right\} \Longrightarrow \mathcal{C}_5 = \left[\begin{array}{cccccccc} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{array} \right]$$

$$\Sigma_6 : \left\{ \begin{array}{rcl} x_1(t+1) & = & x_2 \wedge \neg x_3 \\ x_2(t+1) & = & x_1 \wedge \neg x_3 \\ x_3(t+1) & = & (x_1 \wedge \neg x_2) \wedge u \end{array} \right\} \Longrightarrow \mathcal{C}_6 = \left[\begin{array}{cccccccc} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 1 & 1 & 1 & 1 \end{array} \right]$$

Remark 3.3 *By counting the $c_{ij} > 0$ of Σ_i, $i = 1, 2, \ldots 6$, it is easy to check that there are 31, 27, 22, 16, 18, and 15 available trajectories, respectively.*

For a better visualization C_k can be indeed represented by a 3D-array as shown in Figure 3.1, where Figure 3.1.a) and Figure 3.1.b) show the set of controllability matrices obtained for each connector. In the front of such array, we present the controllability matrix underlying u connected with an OR or AND operator at x_1, then behind the u connected at x_2, and so on. Besides, it is also maintained the blue color for OR operator and red color for AND operator as before. Note that in pink circles we highlighted the element of $\mathcal{C}_k$ that corresponds to the same attractor as initial and as final condition. Namely, $c_{22} > 0$ means that the desired trajectory is $\delta_8^2 \longrightarrow \delta_8^2$ so, regardless the connector, any control sequence is valid.

Moreover, we integrated the available trajectories when the k-th gene is perturbed, either by an OR and AND operator, into a single controllability matrix (see Figure 3.1.c)).

In Figure 3.1. a) $c_{ij} > 0$ in green represents that we can use both AND and OR operator as a connector to promote such trajectory. On the other hand, for a better visualization, we rearranged in blocks the controllability matrix according to its attractors and basins as shown in Figure 3.1. d), as presented in Table 1.2. In other words, for the attractor $\Lambda_1 = \delta_8^8$ its corresponding basin is given by:

$$S_1 = \left\{ \delta_8^1, \delta_8^3, \delta_8^4, \delta_8^5, \delta_8^6, \delta_8^7, \delta_8^8 \right\},$$

and for the attractor $\Lambda_2 = \delta_8^2$ *its basin is:*

$$S_2 = \left\{\delta_8^2\right\}.$$

Thus, the indexes were then sorted by basins (in increasing order); the first section of the matrix is composed by the elements:

$$\{i, j\} = \{1, 3, 4, 5, 6, 7, 8\},$$

while the second section only for:

$$\{i, j\} = \{2\}.$$

■

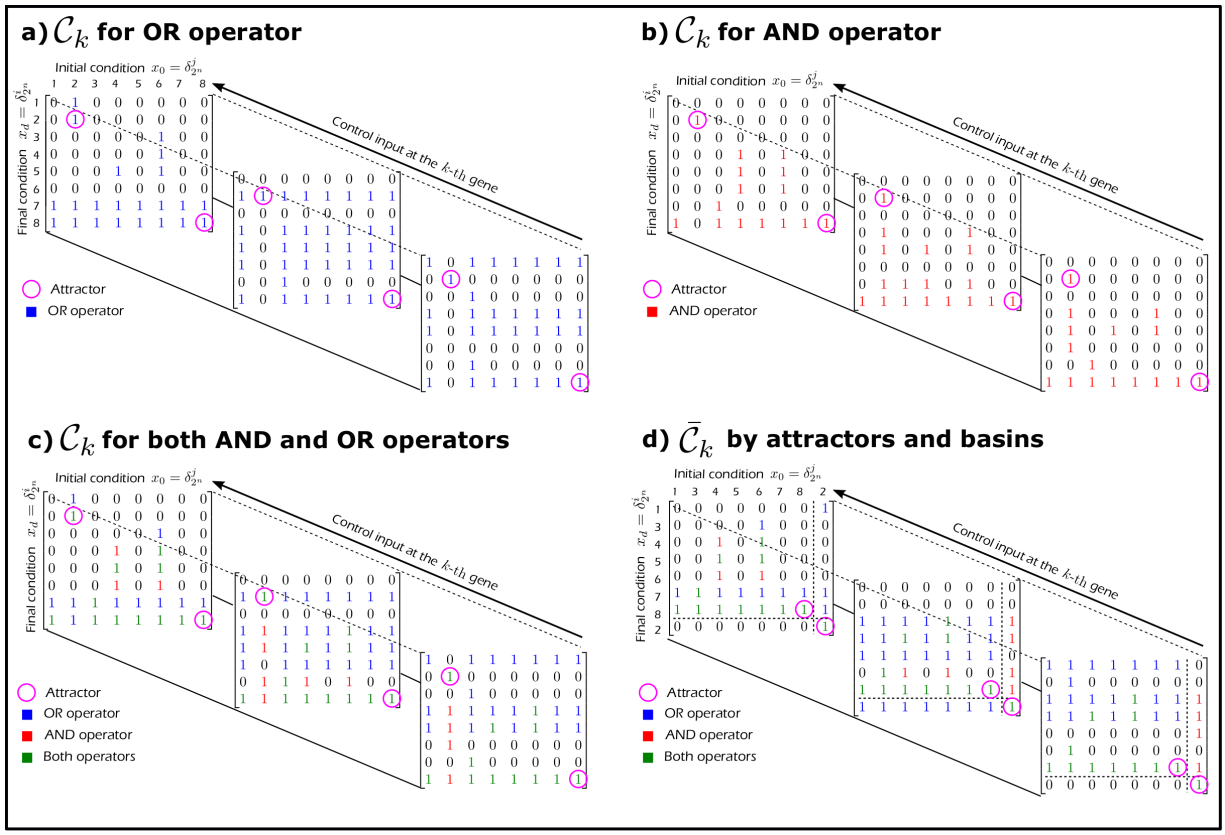


Figure 3.1: 3-D array representation of the controllability matrix $\mathcal{C}_k$ for the Control TwoFixed BN. **a)** $\mathcal{C}_k$ when u is connected to the autonomous system via the OR operator at the k-th gene (blue color). Remember that $c_{ij} > 0$ means that it is an available trajectory. Note that, pink circles highlight attractors. In a similar way, **b)** shows $\mathcal{C}_k$ when u is connected to the autonomous system via the AND operator at the k-th gene (red color). In **c)** both trajectories were integrated, and green color represents that any operator as a connector can promote such trajectory. Finally, in **d)** we show the controllability matrix sorted by its attractors and basins.

3.4 Available state transitions

The available trajectories, also referred as available state transitions, were classified into 3 different categories:

1. Attractor-to-Attractor transitions (A-to-A)

These steady–state transitions are particularly relevant for developmental processes represented via B-GRNs models, since its attractors correspond to cell types and thus push them towards or away from certain behaviours. The information recovered from this type of transitions is related to the necessary alterations (potential genes) in the dynamic system in order to switch between its multiple attractors.

Let $\Lambda_1 = \delta_{2^n}^i \in \Omega$ and $\Lambda_2 = \delta_{2^n}^j \in \Omega$. We say that an *A-to-A transition* is available by connecting a control input u at the k-th gene, if $(c_{ij})_k > 0$. In compact notation:

$$\delta_{2^n}^i \xrightarrow{k-\text{th gene}} \delta_{2^n}^j \qquad \text{if} \qquad (c_{ij})_k > 0 \tag{3.5}$$

2. Basin-to-Attractor transitions (B-to-A or A-to-B)

In cases where the system cannot be driven from one attractor to another in a specific number of steps, we consider this type of dynamic transition.

Let $x_0 = \delta_{2^n}^i \in S_1$ and $\Lambda_1 = \delta_{2^n}^j \in \Omega$. We say that a *B-to-A transition* is available, by connecting a control input u at the k-th gene, if $(c_{ij})_k > 0$. In fact, the same compact notation of Eq. (3.5) holds here.

3. Basin-to-Basin transitions transitions (B-to-B)

Finally, our last case is the Basin-to-Basin transition. We explore this transition on the Chapter 4, in the context of ephitelial cancer when considering the EMT-GRN.

Let $x_0 = \delta_{2^n}^i \in S_1$ and $x_d = \delta_{2^n}^j \in S_2$. We say that a *B-to-B transition*, represented by $x_0 \longrightarrow x_d$, is available if and only if $(c_{ij})_k > 0$. The same compact notation of Eq. (3.5) holds for this transition.

Remark 3.4 *Note that, our reachability analysis only addresses the A-A transitions, so we can say that controllability analysis give us more solid information about the role of the genes in transitory states.*

In the next section we propose a technique for displaying the information provided by the controllability matrix with respect to the available trajectories when the size of the matrix is considerable.

3.4.1 Available State Transitions Distribution (ASTD)

The last example provides information about the use of the controllability matrix $\bar{C}_k$, organized by attractors and basins. Its $2^3 = 8$ different gene profiles allow us to visualize all the $(c_{ij})_k > 0$, and consequently all the different available trajectories. However, in cases where the size of the network increases, the size of $\bar{C}_k$ also does. In consequence, the available trajectories cannot be fully displayed via represented with the techniques previously considered. So we instead propose a visualization tool that we named *Available State Transitions Distribution* matrix, or simple ASTD. This tool is in fact a heat-map (square matrix) of $2^m \times 2^m$, where m is the number of attractors. The sections into which this map is divided contain all the states of the attraction basins, one section for each basin, including in each section the gene profile of the associated attractor. Moreover, the available trajectories are counted and displayed in percentage in each section of the heat-map. To ease the exposition we illustrate in this section the proposed visualization technique with the data associated to our current example.

Example 3.3 ***ASTD matrix of the TwoFixed BN***

Recall Example 3.2, where the set $\bar{\mathcal{C}}_k$ has been already computed. Specifically, let us consider the case where $k = 4$ when a control input u is connected to the x_1 gene through an AND operator (spontaneous switch-off). The controllability matrix, sorted by its attractors and basins, is shown in Fig. 3.3. a). Note that, for a better visualization, the matrix has been colored in black (not available state transition) and white (available state transition), as well as separated in 2^m blocks with a red dot line, where $m = 1, 2$ index the number of attractors (and then of basins of attraction).

There are 16 available induced state transitions, which is to say:

- *11 of the available induced state transitions are in the upper–left block which corresponds to those ones that occur inside the basin of S_1;*
- *4 of the available induced state transitions are in the upper-right block related to transitions that are available from a state of S_2 to a state of S_1;*

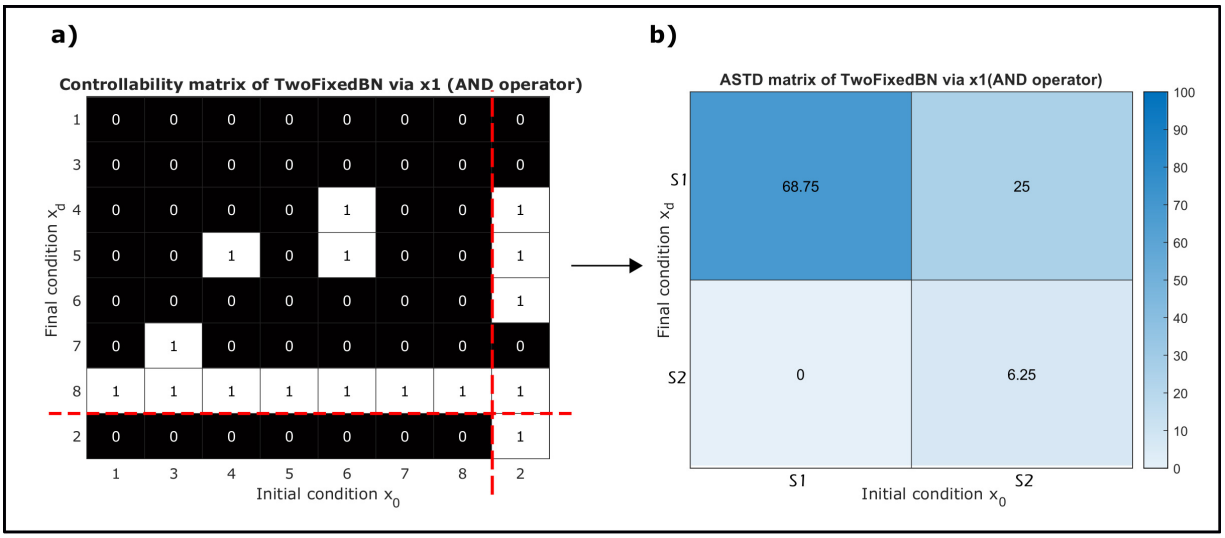


Figure 3.2: The construction of the ASTD matrix. a) The computed controllability matrix $\bar{C}_k$ is sectioned according to the number of attractors and its basins. As we can see there are 16 available state transitions, distributed in 11, 4, 0 and 1 in the upper-left, upper right, bottom-left and bottom-right block, respectively. Then, b) they are counted to be displayed, in percentages, by sections in a heatmap, named as the ASTD matrix. Note that, if we could desire to transit from S_1 to S_2, it is not possible at least under the spontaneous switch-off of x1 (0%). In contrast, there exists several possible induced transitions inside the $S1$ block (68.75%).

- *1 of the available induced state transitions is in the lower-right block.*

Interestingly, in the bottom-left block there are not available induced state transitions from basin S_1 to basin S_2.

To add some qualitative information on the difficulty that involves an available transition, we transform the previous information in a percentage indicator, as shown in Fig. 3.2, b). Visualized in this way we can easily identify, for instance, for which of the attractors it is easy to transit towards it.

This systematic analysis is repeated for all the n genes of the BC-GRN.

■

3.5 Discussion

In previous chapters, it has been illustrated that the topological connectivity patterns restrict the information flux as well as impose structural constraints on the overall behavior of the network [29, 46], mainly in what concerns the attractor transitions.

In this direction, we here compute the controllability matrix of BC-GRNs, which allows a full visualization of the available trajectories among pairs of gene activation configurations. This is particularly relevant since we are concerned about finding control strategies that drive the modular dynamic system from one configuration to another, commonly related to understand how a particular cell type (in cases where the gene profile corresponds to a fixed-point attractor) can change, under the effect of micro-environments, alterations originated from other B-GRN modules, etc.

Although reachability analysis (see the previous chapter) can be quite helpful in this situation, there are some questions that still cannot be answered with that method. For instance, if an attractor-to-attractor transition is not available under a suitable control input u in a finite number of steps, we could investigate if the system can be driven *near of* the attractor, and then let the system autonomously converge. In fact, such concerning induced state trajectories can be found via the controllability analysis, by obtaining first the controllability matrix C_k for each k-th gene, then sorting the information by basins $\bar{C}_k$ for both AND and OR operators.

Matrix $\bar{C}_k$, however, is a complex set of data that can be more successfully interpreted if we sectioned it in its attractors and corresponding basins, assigning then a percentage for each block. Thus, we introduce a visualization tool that we named as the ASTD matrix, which basically is a heat-map of the available induced state transitions among pairs of gene activation configurations. Hence, the different scenarios of state trajectories (A-A, A-B or B-B transitions), presented in Section 3.4, can be easily analyzed.

Finally, we suppose that the information coded by the ASTD matrix can be discussed from a biological perspective, since a high percentage of a certain block of the matrix could be directly related to the robustness of the network to avoid induced transitions, at least by a suitable manipulation of the concerned gene. A complete exploration of all the genes would provide a preliminary insight about the relevance directly associated with its functionality in the network as a whole, namely the cell fates outcomes. In the following chapter, a B-GRN module related to the Epithelial-to-Mesenchymal Transition proposed

by Mendez-Lopez and collaborators [45], is studied by all the algorithms that have been presented in previous sections, to later discuss the result from a biologically substantiated perspective.

3.6 Conclusion

In this chapter, we addressed the controllability analysis of BC-GRNs, using mathematical tools derived from the Semi–Tensor Product approach. The systematic exploration proposed in the chapter allowed us to elucidate the available induced trajectories among pairs of the attractors of the network as well as among states of their corresponding basins. Thus, the number of state trajectories is expanded, as we had suspected in the beginning of this chapter, complementing the results obtained from the reachability analysis. To visualize all the transitions, we built a 3D array controllability matrix, named as C_k, which is built by exploring individual alterations on the genes, and sorted by basins. Results are summarized and shown in the Available State Transition Distribution matrix (ASTD), a heat-map that includes the number of total available state trajectories in percentage form. Hence, in the next chapter we will be able to evaluate the qualitative changes that occur in a biological inspired GRN, when a suitable control input is added into the network.

Taken together, we are now ready to enhance the understanding of the dynamic effects of ordered single alterations of genes on B-GRNs, which is quite helpful to uncover *drug targets* on complex diseases such as cancer, just to provide a quite appealing potential application.

Chapter 4

Uncovering the role that genes play in GRNs: The Ephitelial-to-Mesenchymal Transition as a Case–of–Study

Introduction

In the previous chapters we exposed the reachability and controllability systemic analysis that we conceived as tools intended to uncover the structural role played by individual network nodes in the transient properties of networks described in discrete–time Boolean terms. The two proposed methodologies conform our solid Semi–Tensor Product based algorithms and we use them to depict the dynamical characteristics of different state transitions scenarios. In the context of systems biology, since attractor to attractor transitions have a fundamental role in morphogenetic dynamics (this because a particular given attractor corresponds to a specific cell phenotype and then a potential developmental stage), we are particularly interested in that kind of state transitions. As discussed before, some attractor to attractor transitions cannot be reached as a consequence of an exogenous stimuli acting on a single node. However, new state trajectories appear, including attractor to basin and basin to basin, when all the possible state transitions are explored. The main objective of this chapter is to analyze what results when the

proposed analysis methods are applied to a core gene regulatory network that has high biological significance. For this reason we decided to apply our methodology to a network that comes from the field of biology of medical systems. Henceforth, we take in this chapter as a case–of–study the gene regulatory network that explain in causal terms the Epithelial-to-Mesenchymal Transition in the context of epithelial cancer (see [45]). This network determines cell–fate in the metastasis of cancerous tumors through the regulation of a transition between two cell types: epithelial cells, which line the external and internal surfaces of many organs, and mesenchymal stem cells, which are multipotent connective tissue cells that can differentiate into other type of cells (such as muscles or bones, for instance). It is worth noting that Epithelial-to-Mesenchymal Transition is part of normal developmental processes such as embryogenesis and tissue healing. Nevertheless, this transition is of one of main mechanisms of tumor metastasis. For this reason, it is in the context of the study of cancer diseases that the understanding of the specific role of the nodes of the associated regulatory network is being addressed in an increasingly intensive way. This in order to conceive therapeutic strategies aimed at stopping the emergence and / or progression of the disease. We are not naive enough to claim that understanding the role of a specific gene in reducing the probability of transition leads directly to the development of an applicable therapeutic method. However, we are convinced that the best route to conceive therapeutic strategies must be accompanied by a deep understanding of the dynamics that govern the regulatory networks involved. It is in this tenor that we inscribe our effort.

To fully cover all the previous results of this manuscript, now applying all the proposed algorithms to the discrete–time Boolean model of the core gene regulatory network that underlies the Epithelial-to-Mesenchymal transition, we follow the next procedure:

1. Description of the Epithelial-to-Mesenchymal Transition GRN.
2. Dynamic analysis of the Epithelial-to-Mesenchymal Transition GRN.
3. Reachability analysis of the Epithelial-to-Mesenchymal Transition GRN.
4. Controllability analysis of the Epithelial-to-Mesenchymal Transition GRN.
5. Discussion of the results with biological observables.

Remark 4.1 *At the end of this chapter we exemplify a very interesting transition according to the obtained results; it is a backward non-common trajectory, from Mesenchymal-to-Epithelial state. Interestingly, the manipulation of the genes p16 and NFkB are the main targets to promote such transition. This is carried out in the context of the conceptualization of potential therapeutic strategies intended to modulate the onset and progression of epithelial cancer.*

4.1 Description of the Ephitelial-to-Mesenchymal Transition GRN (EMT-GRN)

As a complex developmental process, Epithelial-to-Mesenchymal Transition (EMT) is involved in:

- morphogenesis,
- tissue regeneration and,
- cancer progression.

This dynamical process is characterized by a series of cell–state transitions, in which belonging to epithelial tissue, they loose their epithelial characteristics, and gain mesenchymal properties (*e.g.* increasing motility). Therefore, some experimentally grounded gene regulatory networks (GRNs) models have started to be proposed to uncover the EMT regulatory core, and thus contribute to the understanding of the EMT regulation and to guide experiments by generating testable hypotheses.

Throughout this work, we particularly used as a case–of–study the core gene regulatory network underlying Epithelial-to-Mesenchymal Transition (EMT-GRN) in the context of epithelial cancer, proposed by Méndez-Lopez and collaborators [45], and expressed in discrete–time Boolean terms. This complex Boolean network, grounded on experimental data, integrates 9 different interacting molecular components; Snai2, ESE2, p16, E2F, Cyclin, TELasa, NFkB, Rb and p53 (see the corresponding set of Boolean functions on Supplementary Table A.5). Although this net is of moderate size, it was necessary to carry out a pre-conditioning to facilitate the application of our techniques. This, by considering

adjusting the logical rules, does not change the dynamic properties of the network. We address this issue in what follows.

4.1.1 Simplified version of the EMT-GRN (S-EMT-GRN)

Despite its seeming low–dimensionality, almost all the Boolean functions involve many logical interactions between many binary variables, which increase the computational cost for further analyses. To solve this, we minimized such interactions through logic inference by using the software Wolfram Mathematica for mathematical analysis. This map–method, used for the synthesis of combinatorial logic circuits, helped us to simplify the Boolean functions of the EMT-GRN to the minimum, and they are presented in Supplementary Table A.6. We must point out that this new simplified version of the EMT-GRN, hereafter referred to as S-EMT-GRN, recovers exactly the same attractors as well as the same truth tables for each Boolean variable as the original EMT-GRN, which is to say the dynamical properties of the network remain invariant. All the computational analyses were first performed in the S-EMT-GRN in order to reduce the computational complexity, and the results were then tested on the original EMT-GRN (*i.e.* the Boolean inputs, which drive the system from one configuration to another).

In compact notation, the set of Boolean equations $\Sigma_{\text{S-EMT-GRN}}$ that conform this B-GRN underlying the immortalization of epithelial cells, can be represented by:

$$\Sigma_{\text{S-EMT-GRN}}: \quad \{x_i(t+1) = f_i(x_1, x_2, \ldots, x_n), \quad i = 1, 2, \ldots, 9\}\,. \tag{4.1}$$

It has been shown in previous studies (see [45]), that this autonomous dynamic system recovers the specific gene expression profiles (fixed–point attractors) that correspond to the :

- epithelial (*Epi*),

- senescent (*Sen*),

- mesenchymal (*Mes*)

stem–like phenotypes. This means that the set of attractors of the network corresponds to the known gene profiles of the cells that constitute epithelial, senescent, and mesenchymal,

tissues. In fact, in what follows we corroborated this result by using the algebraic form of BN.

4.2 Dynamic analysis of the S-EMT-GRN

Given the autonomous low–dimensional Boolean S-EMT-GRN, the Boolean functions related to each node of the network, $\Sigma_{\text{S-EMT-GRN}}$, are rewritten as algebraic operators, following for this Theorem 1.1. Thus, the structure matrices M_{f_i}, $i = 1, 2, \ldots, n$, can be computed.

This rewriting enables us to transform the dynamical system into its algebraic form described by:

$$x(t+1) = Lx(t),$$

where $x(t) = \ltimes_{i=1}^{n} x_i(t)$ is the STP of the molecular components of the S-EMT-GRN, meaning:

$$x(t) = \text{Snai2}(t)\text{ESE2}(t)\text{p16}(t)\text{E2F}(t)\text{Cyclin}(t)\text{TELasa}(t)\text{NFkB}(t)\text{Rb}(t)\text{p53}(t)$$

and $L \in \mathcal{L}_{2^9 \times 2^9}$ is the network transition map between the updated and the current state of expression.

Given L, we proceed to identify the attractors of the network corresponding to those gene activation configurations experimentally observed. This for all the 2^9 possible initial conditions $x_0 \in \Delta_{2^9}$. Since $trace(L) = 3$, this network recovered three fixed–point attractors in accordance to Méndez *et al.* in [45], each corresponding, respectively, to the epithelial, senescent and mesenchymal stem–like cellular phenotypes.

The attractor's set, conformed by those gene profiles that satisfy the condition:

$$Col_i(L) = \delta_{2^n}^i$$

is:

$$\Omega = \{\delta_{512}^{332}, \delta_{512}^{313}, \delta_{512}^{228}\},$$

with $r_0 = 4$ (see Eq. (1.5)). Besides, the basins S_m for each Δ_m, $m = 1, 2, 3$ are computed. However, because of the size of S_m, the elements are not displayed here. Instead, a

m	Λ_m	Name tag	Basin size
1	$\delta_{512}^{332} \sim (0,1,0,1,1,0,1,0,0) \sim 180$	*Epi*	92
2	$\delta_{512}^{313} \sim (0,1,1,0,0,0,1,1,1) \sim 199$	*Sen*	132
3	$\delta_{512}^{228} \sim (1,0,0,0,1,1,1,0,0) \sim 284$	*Mes*	288

Table 4.1: Summary of the attractors and basin sizes of the S-EMT-GRN

summary of this information is shown in Table 4.1.

Remark 4.2 *The recovered fixed–point attractors via the algebraic representation are the same as the ones reported in [45], which were there obtained using standard simulation also written in the R language (see Fig. A.3).*

The analysis of the data provided by Table 4.1 leads us to the following:

Remark 4.3 *As can be seen in Table 4.1, the basin of attraction corresponding to the mesenchymal cellular phenotype comprises 56.25 percent of the 512 activation profiles of the network, and the basin of attraction corresponding to the epithelial cellular phenotype attractor comprises 17.96 percent. This means, in colloquial terms that it is easier for the network to remain in the mesenchymal configuration state than to remain in the epithelial cellular phenotype.*

We now explore some autonomous state trajectories using the Boolean as well as the continuous models of the S-EMT-GRN. Let us test three different x_0 that eventually converge to each attractor of the network. We have then:

$$x_{01} = \delta_{512}^{512}, \qquad x_{02} = \delta_{512}^{352}, \qquad x_{03} = \delta_{512}^{448}$$

The observed autonomous state trajectories are:

$$x_{01} = \delta_{512}^{512} \to \delta_{512}^{71} \to \delta_{512}^{378} \to \delta_{512}^{364} \to \delta_{512}^{332} = Epi$$

$$x_{02} = \delta_{512}^{352} \to \delta_{512}^{331} \to \delta_{512}^{298} \to \delta_{512}^{313} = Sen$$

$$x_{03} = \delta_{512}^{448} \to \delta_{512}^{17} \to \delta_{512}^{186} \to \delta_{512}^{242} \to \delta_{512}^{228} = Mes$$

for each initial configuration. For a better visualization in Fig. 4.1 c), the corresponding D_e is plotted for each x_0.

We use Eqs. (1.7)–(1.9) to transform the system (4.1) into its continuous approximation. Thus, the system of ODEs that results from the transformation is given by:

$$\dot{x}_i = \Theta\left[\bar{f}_i(x_1, x_2, \ldots, x_n)\right] - x_i; \qquad i = 1, 2, \ldots, 13,$$

where:

$$\Theta\left[\bar{f}_i(x_1, x_2, \ldots, x_n)\right] = \frac{1}{1 + \exp\left[-15\left[\hat{f}_i(x_1, x_2, \ldots, x_n)\right] - 0.5\right]}.$$

The numerical solutions of the ODEs system caused by each x_0 are displayed in Fig. 4.1 d). At the top of these continuous transition plots, we indicate the corresponding gene profile of each initial condition, whereas, at the bottom, the gene activation configuration of the *Epi*, *Sen* and Mes attractors. Besides, the relative expression of the genes that change over the time are displayed in thicker lines.

The previous analysis corresponds to the application of Algorithm 1 (see Chapter 1), to the transformed core gene regulatory network that underlies the Epithelial-to-Mesenchymal Transition. We can now proceed to the application of our second algorithm, which is to say Algorithm 2 exposed in Chapter 2.

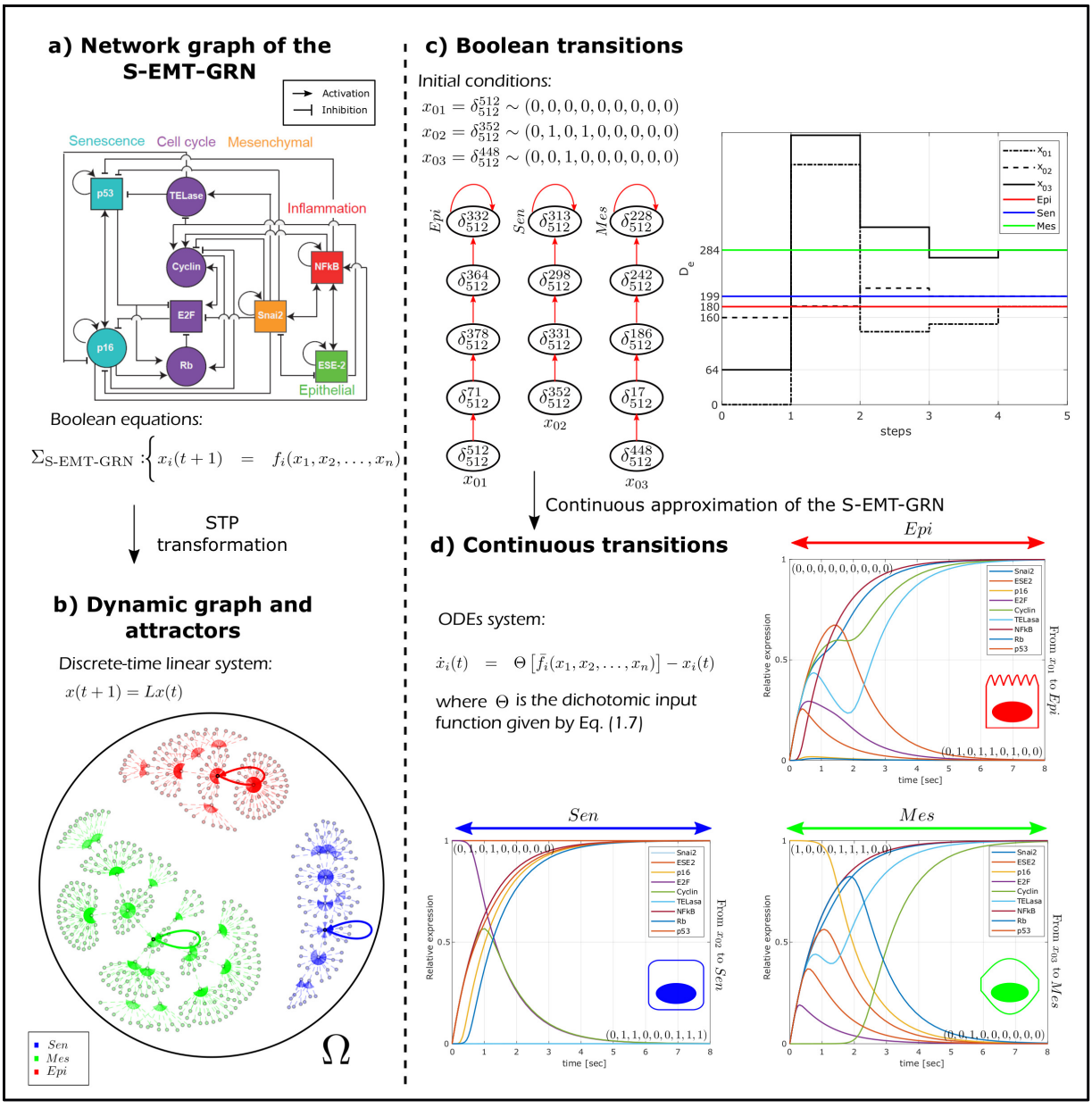


Figure 4.1: Dynamic analysis of the S-EMT-GRN. a) The network graph of the S-EMT-GRN is composed by 9 genes as well as several non-linear interactions among them. **b)** After an STP transformation, $\Sigma_{\text{S-EMT-GRN}}$ is rewritten as a discrete–time linear system of the form $x(t+1) = Lx(t)$. Hence, the dynamic graph that includes both attractors and basins can be plotted. Finally, three different x_0 were tested to observe its **c)** Boolean transition as well as **d)** continuous transition plots.

x_0/x_d	Epi	Sen	Mes
Epi	*	NFkB, **p16**, **p53**	ESE2, **Snai2**
Sen	p16	*	ESE2, **Snai2**
Mes	Snai2		*

Table 4.2: Available trajectories among all possible pairs of attractors induced by potential genes. The crossover of row and column contains the gene to transit from x_0 to x_d. In bold letters, the control input switches-on the gene while the reminder switches-off. The (*) represents the same attractor as initial to final condition, so no control input is needed.

4.3 Reachability analysis of the S-EMT-GRN

In order to uncover the genes that promote attractor transitions for the discrete–time Boolean gene regulatory networks that we are taking into consideration in the current example, we proceed to transform the S-EMT-GRN into a BC-S-EMT-GRN, by adding a control input on the i-th gene. We tested separately two possible Boolean operator among the original Boolean functions f_i and the exogenous input u to create the new Boolean function $\bar{f}_i$: the OR operator (switch-on), and the AND operator (switch-off). Thus, we adopted the new system of Boolean equations given by:

$$\Sigma_{\text{BC-S-EMT-GRN}}: \quad \Big\{ x_i(t+1) = \bar{f}_i(x_1, x_2, \dots x_n, u), \qquad i = 1, 2, \dots, 9, \tag{4.2}$$

which now also depends on u. We apply then the STP transformation but now to BC-S-EMT-GRN, and consequently the system became a discrete–time bilinear system of the form:

$$x(t+1) = \tilde{L}x(t)u(t),$$

where $\tilde{L}$ is a matrix of dimensions $2^{10} \times 2^9$, also called *the transition matrix of the controlled network*. Then, to characterize the reachable subsets of BC-S-EMT-GRN, given the set of attractors and the computed $\tilde{L}$, we performed Algorithm 2 for $s = 1, 2, \dots, r_0$.

As we mentioned above, we called *potential genes* (resulting from the reachability analysis) those ones that, changing their level of expression by the interaction with the control input u, promote a *trajectory* from one attractor to another. Particularly for this example, a total of 9 available trajectories were found on BC-S-EMT-GRN -at least under the two connectors tested here- and they are summarized in Table 4.2.

According to the results in the previous Table 4.2, we can now explore an induced transition from *Sen* to *Epi* attractor and backwards. The genes p16 (switching-off) and NFkB (switching-on) were selected for this objective, respectively. So, the desired trajectory is

$$Sen \xrightarrow[5]{\text{p16}} Epi \xrightarrow[3]{\text{NFkB}} Sen \tag{4.3}$$

Then, using Eqs. (2.5), (2.6) and (2.7), we obtained the continuous version of the BC-S-EMT-GRN. In order to find the minimum time of manipulation of the potential genes by the control inputs u_j in the BC-S-EMT-GRN, we separated the composed trajectory into two individual transitions, and consequently two different ODEs system were also approximated as stated before. For each attractor transition, we take the initial attractor as an x_0 of the ODEs system, and we then increased the duration of the j-th control input (Δ_{t_j}), interpolating its value to obtain the value of time–dependent terms at specified time. Moreover, we considered an extra time for each interval Δ_{t_j} in such a way that the potential gene and the rest of the genes have enough time to relax to its new state after the manipulation by the control input. The resulting time values are

Trajectory	Minimum Δ_{t_j}	Extra time
$Sen \xrightarrow{\text{p16}} Epi$	2.19	5.5
$Epi \xrightarrow{\text{NFkB}} Sen$	0.77	5.5

note that the continuous functions that changed by the addition of the u_j are

$$\begin{aligned} \hat{f}_{\text{p16}}(\cdot) &= \hat{f}_{\text{p16}}(\cdot) \cdot u_1 \\ \hat{f}_{\text{NFkB}}(\cdot) &= \hat{f}_{\text{NFkB}}(\cdot) \cdot u_2 \end{aligned}$$

In order to illustrate the induced trajectories that we have previously discussed, we generated Figure 4.2. The included image has four main sections:

Section one: On the top-left we present the control inputs that are added to the S-EMT-GRN, *i.e.* Figure 4.2.a), to promote the *Sen*-to-*Epi* transition, and then *Epi*-to-*Sen* transition via p16 and NFkB, respectively.

Section two: On the bottom-left, the initial attractor state and its Boolean progressive change under the effect of the controllers. In Fig. 4.2.b), the attractors (tagged with

its D_e, *i.e.* its decimal equivalent representation) are visited as expected, and the number of steps of such Boolean A-to-A transitions correspond with the computed through the reachability analysis.

Section tree: On the other hand, to observe quantitative changes (*i.e.* the relative gene expression of the p16 and NFkB on the induced trajectories), and evaluate if such attractor transitions are maintained on the continuous approximation of the model, we first plotted on Fig. 4.2.c) the bifurcation diagrams for each separated transition.

Section four: Finally, on Fig. 4.2. d) we tested the results on the continuous model of the S-EMT-GRN.

Remark 4.4 *Notice that, interestingly, on the continuous model the time of manipulation of p16 is bigger than NFkB (see Figure 4.2.c)), so we reasoned that could be a direct equivalence of steps and time.*

In the next section, let us now expand the number of available trajectories according to the results presented in Chapter 3, computing the ASTD matrix for each gene.

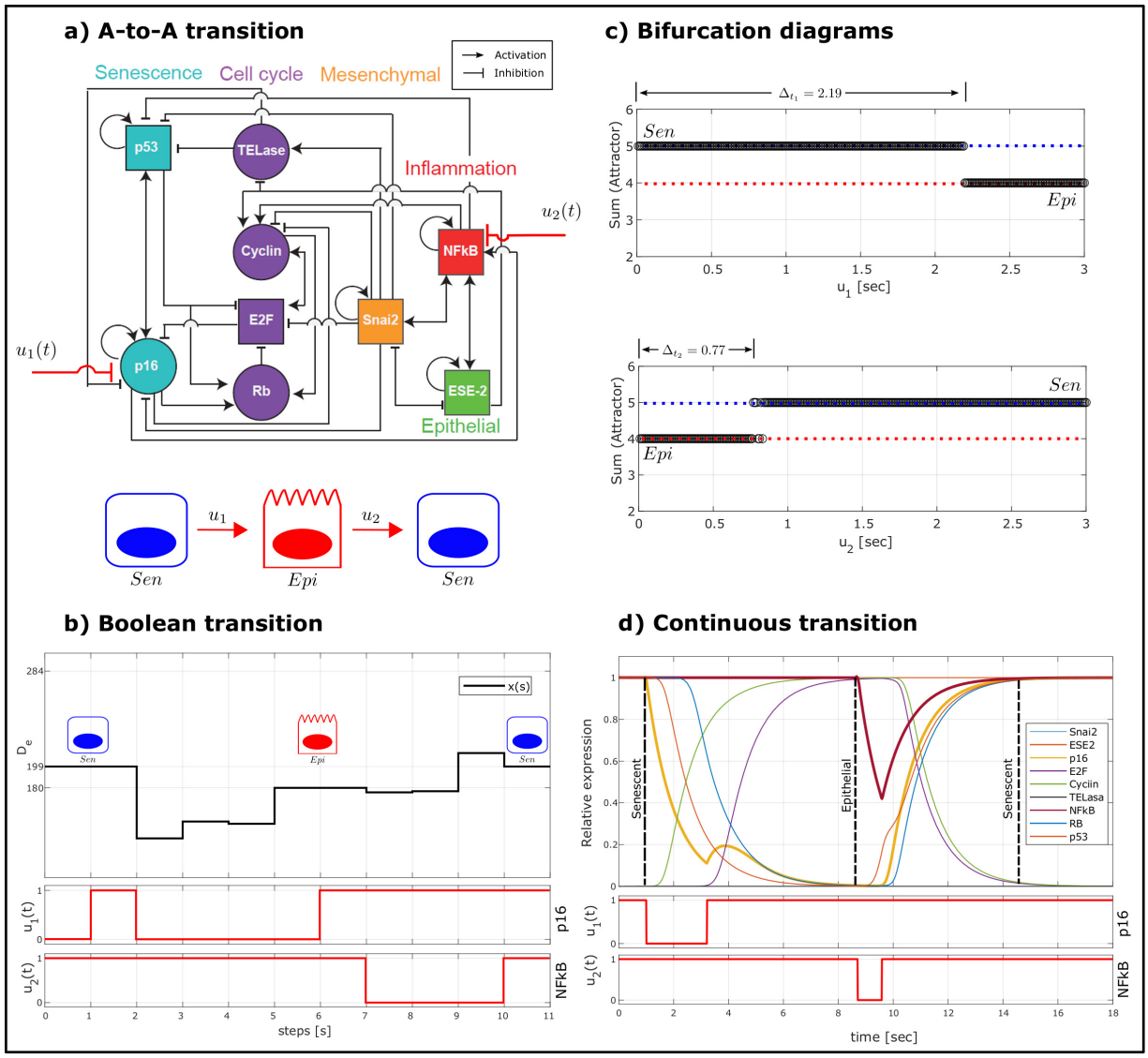


Figure 4.2: Senescent-to-Epithelial-to-Senescent transition. **a)** A trajectory from *Sen* to *Epi* to *Sen* attractor. It is tested on **b)** the Boolean model of the S-EMT-GRN, as well as on **d)** its approximated continuous model.

4.4 Controllability analysis of the S-EMT-GRN

Previously we induced A-to-A state transitions, since we conclude that it was possible to do it via the exploiting of the information provided by the reachablity analysis. Now, as an illustration of the possibilities provided by the controllability analysis, we shall explore the

possibility to steer the Epithelial-to-Mesenchymal core gene regulatory network, described in terms of a discrete–time Boolean network, in order to get a reversible transition, *i.e.* an induced Mesenchymal-to-Epithelial Transition. In this direction, we recall the result in Chapter 3 to compute the controllability matrix C_k, which corresponds to the sort information in blocks that include the attractor and its basin states.

According to the algorithm presented above, it is now necessary to add a control entry for each gene of the EMT network. Then, each of the controllability matrices are calculated and arranged by blocks for each of the given attractors, which are taken as initial and final states defining the beginning and the end of differentiation trajectories, respectively. The controllability matrix has $2^n = 512$ entries, that correspond to all the gene profiles that the network can configure. Since the network has 9 nodes and also the control inputs that are tested in this analysis are connected to the network via AND and OR logical operators, there are a total number of 18 different matrices that conform the matrix set identified by C_k. In this matrix are all the induced trajectories (A-to-A, B-to-A, ...) by the different possible manipulations of the genes of the network, through spontaneous on and off switches. It is important to emphasize that in this analysis those trajectories that had previously been found through the analysis of reachability are recovered.

In Figure 4.3, the results of the controllability analysis for the network are shown. A control input connected to the corresponding Boolean function through an AND (switches off) is taken into account. The representation of the results, as mentioned in the previous chapter, is provided through a heat map. This representation counts the total number of available trajectories and display such quantities as percentages. For example, in the upper left part of the graph, corresponding to the manipulation of node $Snai2$, there are no possible trajectories from Epithelial and Senescent steady–states to the Mesenchymal phenotype. We represent this fact just including 0% in the corresponding section. For the same case the highest percentage of possible induced trajectories are within the same mesenchymal attractor. This means that any initial condition belonging to the basin, which will eventually reach the attractor's configuration, can be manipulated to lengthen or shorten this immanent trajectory. Similarly, the exploration and results for the rest of the genes are shown in this figure.

On the other hand, when the control inputs intended to manipulate the nodes are connected via an OR connector we obtain the corresponding results through the ASTD

matrix as shown in Figure 4.4. Thus, both figures integrate the 18 graphs of the matrix C_k.

Once performed the controllability analysis, which also includes the transient configurations of the network, the total number of induced trajectories is expanded as had been assumed at the beginning of the current chapter. For this, in a condensed way, in Figure 4.4, the tables of the analysis of reachability and controllability are included.

Remark 4.5 *Note that, in the second analysis, 8 new induced trajectories appear. For instance, one of particular relevance is the trajectory that leads from the mesenchymal phenotype to the senescent phenotype. This induced trajectory did not appear when carrying out the reachability analysis. This is because now the transient configurations are also taken into account as initial conditions, so transitions are no longer restricted only to the gene profiles of the attractors taken as initial conditions.*

To visualize the results coming out from the controllability analysis and the associated complexity of the manipulation of these transient configurations or "intermediate states" that eventually, if there were no control action, would lead to the attractor to which they belong, in the next section we shall explore in a detailed manner transitions from the mesenchymal phenotypic state to the epithelial phenotypic state.

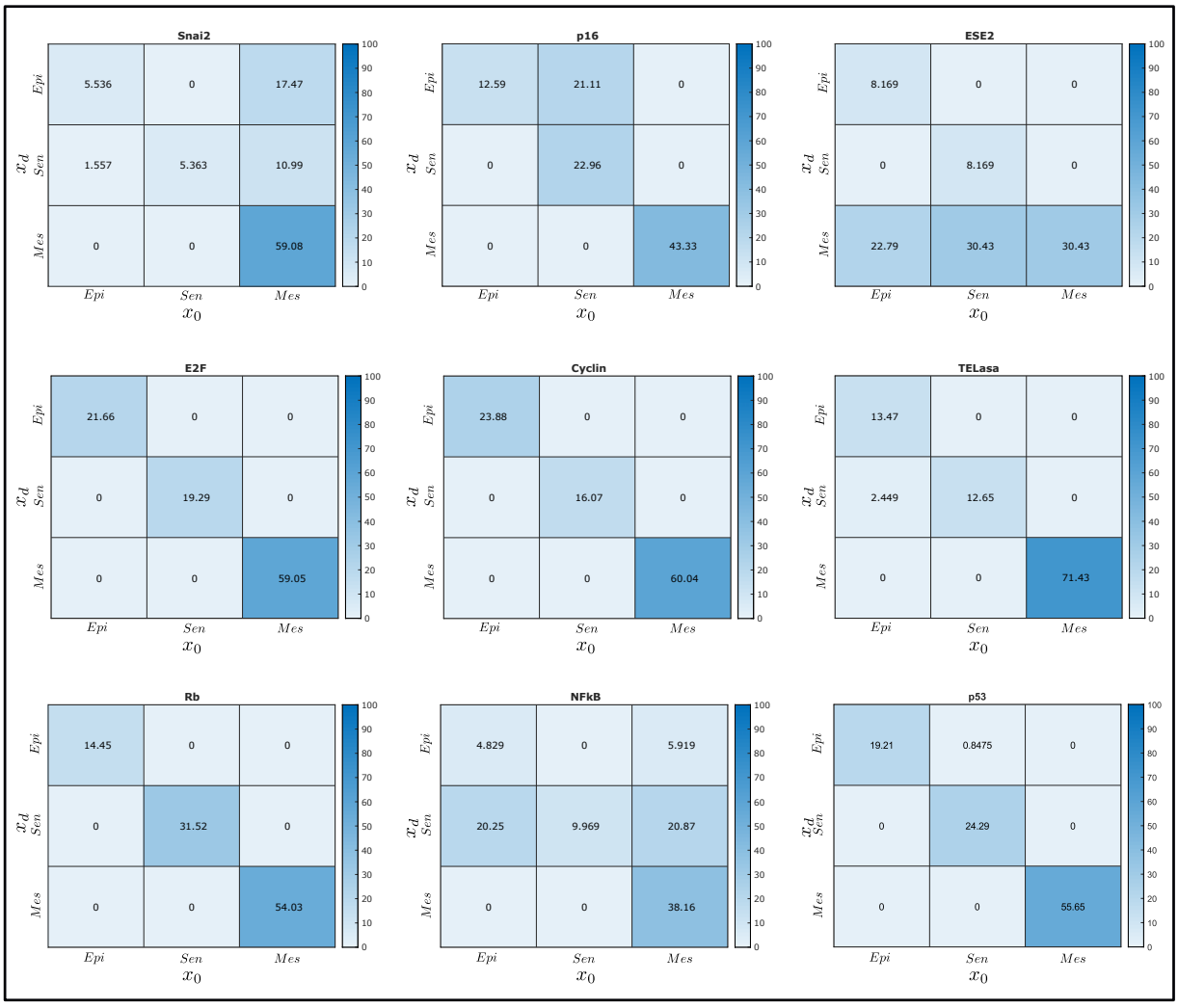


Figure 4.3: ASTD matrices obtained from spontaneous switching-offs of a single gene. For each gene of the S-EMT-GRN, a control input u is connected via an AND operator in the concerned gene, overruling the original Boolean function and consequently a spontaneous inhibition of such gene is induced. Then, we compute the controllability matrix $\bar{C}_k$, which has been separated in 2^m blocks, where $m = 1, 2, 3$ are the number of attractors (and then of basins of attraction). The total number of available trajectories, including A-to-A, B-to-A and B-to-B transitions, are counted to be displayed in percentages in the previous blocks, in a heat-map form.

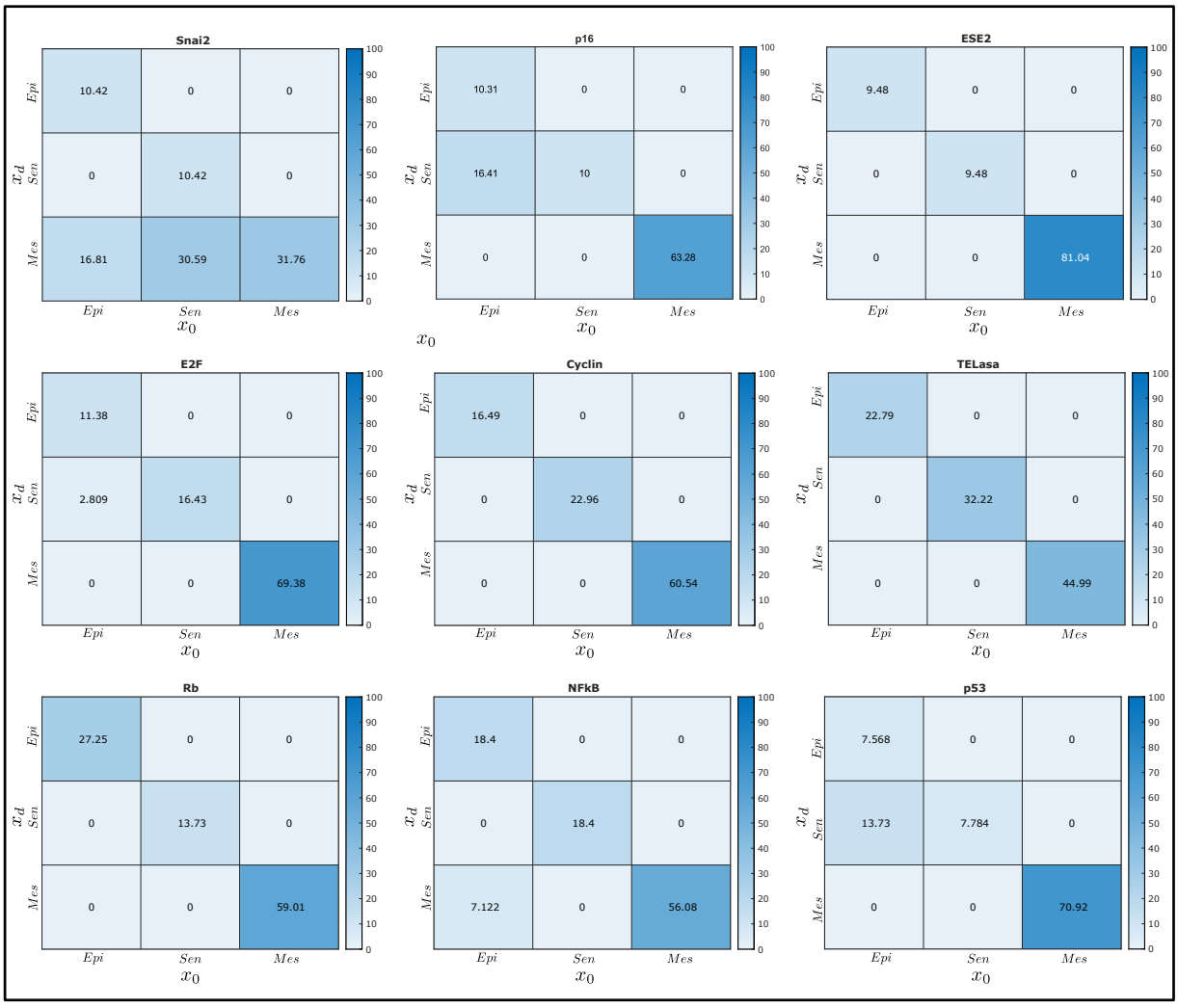


Figure 4.4: ASTD matrices obtained from spontaneous switching-ons of a single gene. For each gene of the S-EMT-GRN, a control input u is connected via an OR operator in the concerned gene, overruling the original Boolean function and consequently a spontaneous activation of such gene is induced. Then, we compute the controllability matrix $\bar{C}_k$, which has been separated in 2^m blocks, where $m = 1, 2, 3$ are the number of attractors (and then of basins of attraction). The total number of available trajectories, including A-to-A, B-to-A and B-to-B transitions, are counted to be displayed in percentages in the previous blocks, in a heat-map form.

a) Available trajectories via reachability analysis

x_0/x_d	*Epi*	*Sen*	*Mes*
Epi	*	NFkB, **p16**, **p53**	ESE2, **Snai2**
Sen	p16	*	ESE2, **Snai2**
Mes	Snai2		*

b) Available trajectories via controllability analysis

x_0/x_d	*Epi*	*Sen*	*Mes*
Epi	*	NFkB, Snai2, TELasa, **p16**, **p53**, **E2F**	ESE2, **Snai2**, **NFkB**
Sen	p16, p53	*	ESE2, **Snai2**
Mes	Snai2, NFkB	Snai2, NFkB	*

Figure 4.5: Available trajectories obtained in both analyses. **a)** After applying the reachability analysis, where the x_0 and x_d are only taken as the specific gene profiles corresponding to attractors, we found 9 possible transitions among pairs of attractors. Note that, under this analysis, it is not possible to transit from the *Mes* to the *Sen* attractor. Nevertheless, if we expand the number of possible x_0 and x_d to those gene profiles also related to the elements of basins, **b)** the number of total available trajectories increases to 17.

4.4.1 Interventions of intermediate states on EMT-GRN

In this section, we named *intermediate states* those gene profiles that correspond to the transient states of the network (*i.e.* the gene profiles of the network that are not attractors), for instance, a state that is a member of the basin of *Mes* attractor:

$$x_0 = \delta_{512}^{385} \rightarrow \delta_{512}^{50} \rightarrow \delta_{512}^{250} \rightarrow \delta_{512}^{228} = Mes \tag{4.4}$$

In other words, they are those initial conditions x_0 that, if no control input u were connected, the autonomous system would eventually converge to the attractor that characterizes the basin where x_0 is chosen.

Although the more common gene activation configurations found in gene expression analyses are the ones corresponding to attractors, these intermediate states are also quite relevant. Why? Because in recent multiple mutations single–cell experiments (see for instance [36, 63]), it is possible to induce a particular desired gene profile, regardless of whether they are or not a configuration that defines a given attractor. In this direction, we

provide here a theoretical framework which can be very useful to help/guide these kind of experiments, by using the controllability analysis where all 2^n gene profiles are explored, including attractors and the corresponding basins of attraction. In fact, our approach is oriented to *intervene* intermediate states and thus revoke imminent transitions, via suitable control inputs in specific genes. Later on, we will indeed intervene the intermediate state x_0 of the Eq. (4.4).

In the case of the BC-S-EMT-GRN, a particularly relevant induced state trajectory is the one associated to the Mesenchymal-to-Epithelial Transition (MET). This transition has an important therapeutic meaning. This because in the context of epithelial cancer MET corresponds to an induced transition from a cellular unhealthy state to a cellular healthy (for a specific tissue). Hence, the desired trajectory is:

$$Mes \longrightarrow Sen \longrightarrow Epi$$

This trajectory can be divided into two sections:

- from Mes to Sen,

- and from Sen to Epi.

We apply the proposed controllability analysis to find the suitable control sequence that allows to drive the dynamic system from Mes to Sen attractor since, according to the table on the bottom of Fig. 4.5, this state transition is possible only if we explore and induce a B-to-A transition. In fact, there are two potential genes, Snai2 and NFkB, that could promote such an induced transition. On the other hand, we apply the control sequence obtained via reachability analysis for the transition from Sen to Epi attractor due to there is an available A-to-A transition by manipulating the network node that corresponds to p16. Fortunately, MET transition is quite useful to elucidate that both analyses are complementary and can be applied simultaneously, and thus provide a full relevant information on the functionality of the genes of the network (when considering differentiation dynamics).

Once the two complementary analyses are applied, we built the full desired trajectory as follows:

$$Mes \xrightarrow[8]{\text{NFkB}} Sen \xrightarrow[5]{\text{p16}} Epi$$

The result of the controllability analysis indicates that if we fix as initial condition:

$$x_0 = \delta_{512}^{385} \sim (0,0,1,1,1,1,1,1,1) \sim 127,$$

which is a state that belongs to the basin of Mes attractor, we are able to intervene this intermediate state under the control sequence:

$$u_1(t) = (0,0,1,1,1,1,1,1,1)$$

that acts on the activation state of gene NFkB through an AND operator (spontaneous switching-off). Thus after 8 steps the network attains the Sen gene profile. On the left of Fig. 4.6, we show the transition from a state of the basin of attraction of Mes to the Sen attractor (in dotted line). Note that, the black line indicates that if the transitory state is not intervened, the system eventually converge to the Mes attractor as expected.

For the second section, once the dynamic system has reached the gene profile corresponding to the Sen attractor, the reachability analysis showed that the control sequence given as:

$$u_2(t) = (0,0,1,1,1)$$

that modulates the activation state of gene p16 through an AND operator (spontaneous switching-off), is able now to promote a transition to the Epi attractor. Since both control input will be connected simultaneously, it is necessary to complete each sequence to promote the full transition as follows:

$$\begin{aligned} u_1(t) &= (0,0,1,1,1,1,1,1,1,1,1,1,1,1), \text{ on NFkB} \\ u_2(t) &= (1,1,1,1,1,1,1,1,0,0,1,1,1), \text{ on p16} \end{aligned}$$

Finally, the complete induced state reprogramming trajectory, which is quite relevant in the context of epithelial cancer when finding potential genes that can be therapeutic targets, is depicted on the right of the Fig. 4.6.

Remark 4.6 *It is important to mention that the previous control action is not the unique strategy to promote MET transition. However, we specifically pointed out these genes because they are strongly related with inflammatory processes [45], which are the key features*

on understanding the origin of unhealthy states on tissues in the context of chronic—degenerative diseases.

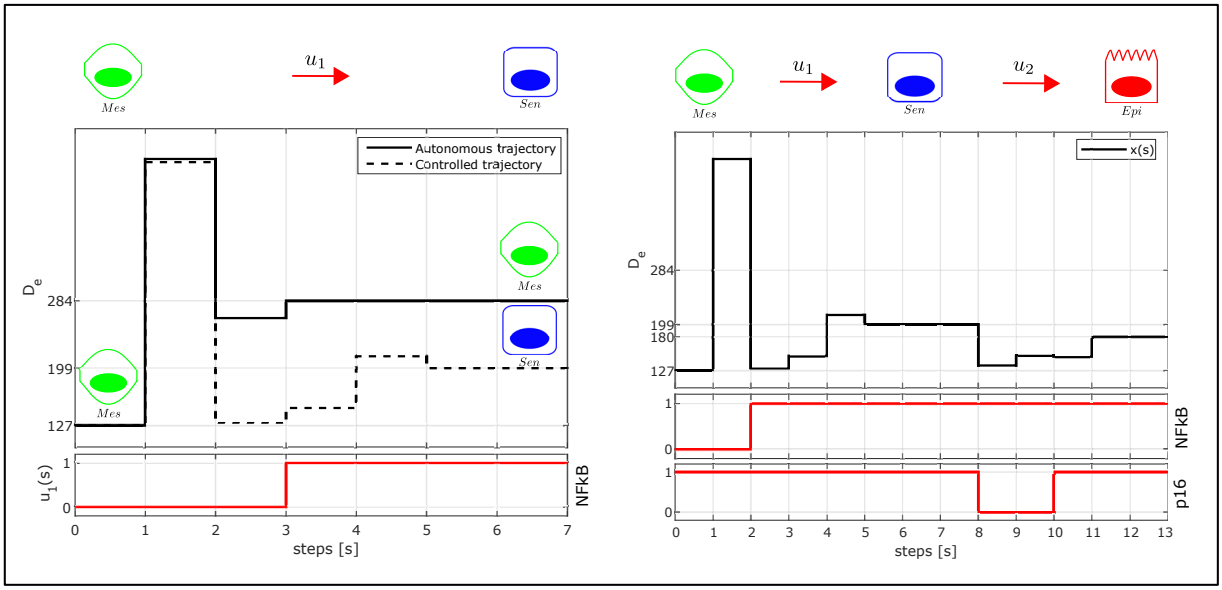


Figure 4.6: Mesenchymal to Epithelial Transition.

Figure 4.7: Reversing the Epithelial-to-Mesenchymal Transition via exogenous control. For the core gene regulatory network that underlies the epithelial to mesenchymal transition, modeled by the discrete–time Boolean network EMT-GRN, the image on the left represents an induced transition from the mesenchymal attractor to the senescent attractor. The induction starts from an element located in the attraction basin of the mesenchymal attractor (that is, the depicted transition belongs to the B-to-A transition class). By manipulating the activation state of the NFκb node, first sustaining an inhibitory action and then releasing the control action, the network is stirred from the given initial condition to the senescent attractor. On the other hand, the image on the right concerns the transition from the senescent attractor to the epithelial attractor. Once attained the final condition via the control action depicted in the image on the left, a complementary control action is now applied. This steers the network's configuration to the epithelial state. This is attained by applying a train of inhibition actions of the activation state of the node corresponding to p16. This thus shows the reversible transition process induced by control from the mesenchymal state to the epithelial state (this manipulating the activation state of node NFκb and p16 via a well–designed control action).

4.5 Discussion

In this chapter we complement the reachability analysis methodology with a proposed controllability analysis. The complementary nature of both approaches to induce phenotypic transitions has been illustrated in this chapter via the induction of a Mesenchymal-to-Epithelial Transition via a composed control action that modulates the behavior of Epithelial-to-Mesenchymal gene regulatory network. Since this core gene regulatory network characterizes the onset and progression of epithelial cancer, the design of the control action can be interpreted in conceptual terms as a therapeutic strategy intended to reverse the unhealthy cellular phenotypic state. Henceforth, the combined method can be useful to explore the regulatory circumstances that potentiates reversed transitions. Since chronic inflammation potentiates epithelial cancer, it is quite natural from a medical systems biology point of view to chose NFκb and p16 as therapeutic targets. However, the considered core gene regulatory network is not isolated. It is embedded into a complex regulatory machinery that includes different modules, *e.g.* the module that regulates the inflammatory response and the module that regulates the cell cycle (see [34] and the references therein). This complexity is increased when the nodes that instantiate the Epithelial-to-Mesenchymal core gene regulatory network to specific tissues are included. This finding leads us to suggest that the proposed methodology can be used to explore the connection patterns between modules that make reverse transitions possible (*e.g.* through the modification of the membership of the basins of attraction of the attractors involved). This is the control law could be used to determine the structural and functional characteristics of the interconnection patterns that must exist between core gene regulatory modules to enhance cellular reprogramming reported in the scientific literature. This may also be of great importance in the context of the interpretation, from the perspective of medical systems biology of the processes underlying tissue regeneration (see for instance [64]).

Returning to the therapeutic aspect, an actual control strategy requires taking into account transition times, as well as the unavoidable presence of stochasticity (among other important aspects). This necessarily leads to the transformation of our proposed methodology to one that uses continuous–time descriptions of the dynamics of regulation. In the specific case of the transition from the epithelial state to the mesenchymal state, under the action of the inflammatory response, the continuous–time model presented in [4] could serve as a starting point.

Although the presented methodology focuses on elucidating the role that certain nodes play in the phenotypic transition dynamics, by acting as access routes to the stimuli that induce them, it would be necessary to complement our study by also exploring interconnection patterns between core gene regulatory modules. Interconnection patterns that are translated into the setting of specific properties of the attraction basins, which condition transition probabilities due to stochastic noise and not to deterministic stimuli. In this regard, we think that the formalization of stochastic landscapes should be enriched to include aspects of reachability. Recent work is allowing the experimental characterization of the dynamics of biological development in specific tissues. This through the use of analysis methods based on single cells (see for instance [55]), which has the potential to clarify the action of natural modulation patterns that can be understood in the light of their comparison with the conceptual control strategies that induce transitions.

An aspect that also seems significant to us from what we present in this chapter concerns the possibility of introducing a metric that quantifies the significance that each specific gene in a regulatory network possesses in terms of its role in the induction of transition dynamics between phenotypic states. As illustrated with the example of the Mesenchymal-to-Epithelial Transition, a specific combination of actions to modulate the activity of the NFkB and p16 genes allows the transition shown, from which the existence of associations between specific genes and differentiation (or reprogramming, if it is the case) processes that are also specific, which is manifested in the ASTD matrix. This matrix can be used to conceptualize the suggested metric.

4.6 Conclusion

In this chapter we have illustrated how to exploit both the reachability analysis and the controllability analysis as a means of inducing phenotypic transitions by designed controlling actions consisting of specific deterministic exogenous stimuli. Our discussion has been limited to the generic model of the network that underlies the Epithelial-to-Mesenchymal Transition described in discrete–time Boolean terms in the context of epithelial cancer, *i.e.* system S-EMT-GRN as described in Table A.6. By applying the proposed methodology, the nodes of the network whose activation state can be manipulated to induce transitions are identified. Subsequently, control actions are designed and when applied, transitions are

induced, both between attractors and between intermediate states and attraction basins. The procedure is illustrated by constructing a control law that allows Mesenchymal-to-Epithelial Transition to take place. In this case, the activity states of the NFkB and p16 genes are manipulated to achieve the desired goal. In this regard, we have illustrated how ASTD matrices guide the design process of the considered control actions.

In the context of medical systems biology, we have discussed the potential usefulness of the results presented in this chapter, particularly with regard to the detection of therapeutic targets. These are nodes of the network whose manipulation can aid in the transition between unhealthy phenotypic states and healthy phenotypic states. It is clear to us that this possibility is inscribed in a conceptual perspective that must adapt to clinical reality, to the fact that the network considered here is not isolated from its environment, as well as to the inevitable presence of stochasticity in regulatory dynamics.

Conclusion and Perspectives

THROUGHOUT this manuscript we have addressed the study of the structural properties linked to the reachability of core gene regulatory networks described through discrete–time Boolean models. This in the context of the study of biological development from the perspective of automatic control theory. We have restricted our study to the deterministic case, without this meaning that we are not aware of the stochastic nature of the transcriptional regulation of the biomolecular dynamics of biological development. This perspective has been translated into the use of methods of algebraic representation of the networks through the so–called Semi–Tensor Product, which allows gene regulatory networks under study to be represented in algebraic terms as discrete–time dynamic systems. Once such a representation has been achieved, the exploration of the circumstances that allow or not the transitions between steady–states of the dynamic system can then be approached. In terms of dynamic systems, the steady–states of a developmental gene regulatory network correspond to its set of attractors. Thus, a specific developmental trajectory can then be described formally in terms of paths through the attractor landscape. Addressing the study of the structural properties of gene regulatory networks from the perspective of the theory of automatic control has two fundamental objectives. The first focuses on discovering the characteristics of the structural constraints that make the transitions that underlie development possible. The second focuses on characterizing the specific role of the nodes that make up the regulatory network. The algebraic treatment allow us to address these two aspects in a systematic way through the development of efficient numerical algorithms, without this meaning that there are no significant limitations in the scope of what can be done with them. More specifically, the use of the Semi–Tensor Product, which allows translating the logical rules that shape Boolean networks to a matrix representation, involves a high computational cost as the number of nodes and

interactions grows. This limits its application to networks of moderate size. Fortunately for us, biological development involves networks of moderate size, the so-called core gene regulatory networks. As long as it is a question of studying these networks in themselves, the problem of computational cost does not represent a significant challenge. This is no longer the case when addressing the study of what results in dynamic terms when networks are interconnected. Although developmental networks do not work in isolation, partial modularity of transcriptional regulation makes possible to use algebraic analysis methods. In addition, the study of reachability properties gives rise to the identification of nodes that could be acting as bridges that allow interconnection between networks. Identifying which nodes must be modulated to drive a specific transition between attractors in a given regulatory network allows us to explore the specific origin of exogenous stimuli, as due to the action on the network of other regulatory networks. If this is important in the context of the study of developmental dynamics, it is even more so when considering problems inherent to the biology of medical systems. For this reason, in this research work we have illustrated our systematic methodology for exploring the properties of gene regulatory networks, first considering an academic example, then a famous example that has been taken from the field of study of floral morphogenesis, as well as an example concerning the presence of a cyclic attractor, to finish with the example that concerns the Epithelial-to-Mesenchymal Transition in the context of epithelial cancer. It is through this last example that we have explored the possibility of reversing the transition from the mesenchymal cell state to the epithelial cell state, by manipulating the state of activity of certain specific nodes of the underlying gene regulatory network. More specifically, it was demonstrated in Chapter 4 that a piecewise control action involving the modulation of the activity of the nodes corresponding to NFκB and p16 can lead the network from the dominant mesenchymal state to the epithelial state. In terms of the treatment of cancerous disease, the proposed control law can be implemented in different ways, with those that follow a preventive perspective being the preferred ones. This use of the methodology that we present in this manuscript is part of the effort to provide the community dedicated to biomedical research with tools that allow them to discover the specific role of certain sets of genes in the onset and progression of degenerative diseases. Our work is conceived as a guide to address the system based understanding of the dynamic processes that underlie the progression of chronic degenerative diseases, from the perspective of automatic control

theory, as a complement to the bioinformatics approaches that so far dominate the field of medical science that is immersed in the so–called genomic medicine.

To end, we are aware of the existence of other methodologies for controlling the dynamics of gene regulatory networks, some of which not only consider the manipulation of node activity, but also the modification of the network edges, inherent stochasticity, as well as closed-loop control schemes (see for instance [11, 40, 65] and the references therein). However, the algebraic and deterministic approach that we follow here has as its main advantage the fact that the Boolean description of regulatory interactions between nodes is relatively easy to obtain from experimental data. Furthermore, as has been illustrated by following this approach, it is possible to systematize in computational terms the exploration of the manipulation opportunities offered by the network, which can guide the design of therapeutic strategies in the context of medical systems biology. We are aware of the existence of other methodologies for controlling the dynamics of genetic regulation networks, such as those proposed in [51], which use the so-called Process Hitting method, which focuses on the construction of the discrete–time model of gene regulatory networks from the evidence of transitions in the available data. One perspective of this work is to adapt the proposed methodology to such an approach. In addition, the proposed methodological approach is further explored in terms of continuous–time descriptions, to take into account aspects such as the influence of differences between time–scales on the dynamics that govern the reachability properties of networks.

Appendix A

Boolean functions of the B-GRN modules

Introduction

In this appendix we include the sets of logical rules that define the discrete–time Boolean core gene regulatory networks for the examples covered in this manuscript (excluding the academic one, because for that example the corresponding logical rules were presented when the example was introduced), which is to say:

- The Floral Organ Specification core gene regulatory network (FOS-GRN).
- The *Arabidopsis thaliana* Cell Cycle core gene regulatory network (ACC-GRN).
- The Epithelial-to-Mesenchymal core gene regulatory network (EMT-GRN).

En each case we also include the corresponding set of attractors. As far as FOS-GRN, we also include the regulatory functions that correspond to the continuous–time approximation. For the gene regulatory network that underlies the Epithelial-to-Mesenchymmal transition in the context of epithelial cancer we also include the Boolean regulatory rules that define the simplified version, *i.e.* system S-EMT-GRN.

$x_i(t+1) =$	$f_i(x_1(t), x_2(t), \ldots x_n(t))$
$AG(t+1) =$	$(!EMF1 \,\&\, !AP2 \,\&\, !TFL1\,) \mid (\,!EMF1 \,\&\, !AP1 \;\&\; LFY\,) \mid (\,!EMF1 \,\&\, !AP2 \;\&\; LFY)$ $\mid (\,!EMF1 \,\&\, !TFL1 \;\&\; LFY \;\&\; (\,AG \;\&\; SEP\,)) \mid (\,!EMF1 \;\&\; (\,LFY \;\&\; WUS))$
$AP1(t+1) =$	$(\,!AG \,\&\, !TFL1\,) \mid (\,FT \;\&\; LFY \,\&\, !AG\,) \mid (\,FT \,\&\, !AG \,\&\, !PI\,) \mid (\,LFY \,\&\, !AG \,\&\, !PI)$ $\mid (\,FT \,\&\, !AG \,\&\, !AP3\,) \mid (\,LFY \,\&\, !AG \,\&\, !AP3\,)$
$AP2(t+1) =$	$!TFL1$
$AP3(t+1) =$	$(\,LFY \;\&\; UFO\,) \mid (\,PI \;\&\; SEP \;\&\; AP3 \;\&\; (\,AG \mid AP1\,))$
$EMF1(t+1) =$	$!LFY$
$FT(t+1) =$	$!EMF1$
$FUL(t+1) =$	$!AP1 \,\&\, !TFL1$
$LFY(t+1) =$	$!EMF1 \mid !TFL1$
$PI(t+1) =$	$(\,LFY \;\&\; (\,AG \mid AP3\,)) \mid (\,PI \;\&\; SEP \;\&\; AP3 \;\&\; (\,AG \mid AP1\,))$
$SEP(t+1) =$	LFY
$TFL1(t+1) =$	$!AP1 \;\&\; (\,EMF1 \,\&\, !LFY\,)$
$UFO(t+1) =$	UFO
$WUS(t+1) =$	$WUS \;\&\; (\,!AG \mid !SEP\,)$

Table A.1: Boolean functions of the FOS-GRN

A.1 System FOS-GRN

A.1.1 Boolean functions

The Floral Organ Specification Gene Regulatory Network on *Arabidopsis thaliana* (FOS-GRN), characterized from experimental evidence and proposed by Álvarez-Buylla and collaborators [6], constitutes a model–of–choice for the study of morphogenesis regulation in multi–cellular organisms. This developmental gene regulatory network is up to now described as an autonomous deterministic discrete–time dynamical system in Boolean terms as a network conformed by 13 genes. These, the nodes of the network, change its activation state over the time according to the Boolean function related to the gene x_i, for $i = 1, 2, \ldots, 13$. The set of Boolean functions is shown in Table A.1.

A.1.2 Set of attractors

For FOS-GRN we have ten fixed-point attractors; four of them specifically related to both the structural configuration and the dynamical behavior of the inflorescence meristem $(I_i, i = 1, 2, 3, 4)$, whilst the reminder members of the attractors set are related to the phenotypic traits of the primordial cells shaping sepals, petals, stamens, and carpels. The configurations of each attractor are presented in Figure A.1.

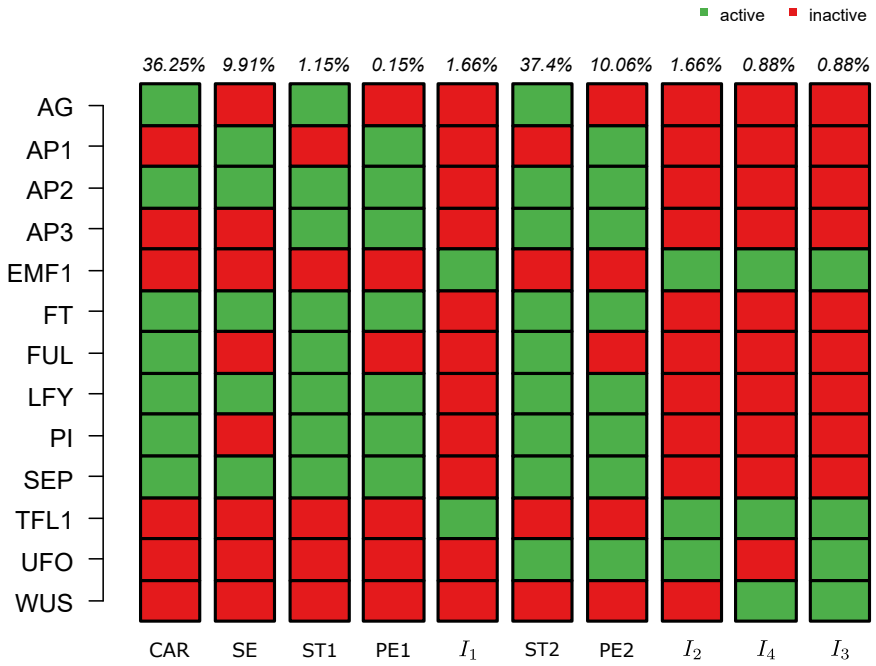


Figure A.1: Gene profiles of the the ten fixed-point attractors of FOS-GRN.

A.1.3 Set of continuous functions

The ODEs system is described by:

$$\dot{x}_i = \frac{1}{1 + \exp\left[-b\left[\hat{f}_i(x_1, x_2, \ldots, x_n)\right] - \epsilon\right]} - x_i$$

where the corresponding continuous–time $\hat{f}_i$ functions are shown in Table A.2. The parameters satisfy $\epsilon = 0.5$ and $b >> 1$.

A.2 System ACC-GRN

A.2.1 Boolean functions

As far as the discrete–time Boolean logical functions that constitute describe the *Arabidopsis thaliana* cell cycle core gene regulatory network, they are shown in Table A.3.

$\hat{f}_{AG}(\cdot) =$	$LFY \cdot (1 - EMF1) \cdot (1 - AP1 \cdot AP2 \cdot (1 - WUS) \cdot (1 - AG \cdot SEP\,(1 - TFL1))) +$ $(1 - EMF1) \cdot (1 - TFL1) \cdot (1 - AP2) -$ $LFY \cdot (1 - EMF1) \cdot (1 - AP1 \cdot AP2 \cdot (1 - WUS) \cdot (1 - AG \cdot SEP\,(1 - TFL1))) \cdot$ $(1 - EMF1) \cdot (1 - TFL1) \cdot (1 - AP2)$
$\hat{f}_{AP1}(\cdot) =$	$(1 - AG) \cdot (1 - TFL1 \cdot (1 - LFY \cdot FT))$
$\hat{f}_{AP2}(\cdot) =$	$(1 - TFL1)$
$\hat{f}_{AP3}(\cdot) =$	$(LFY \cdot UFO) + (PI \cdot SEP \cdot AP3 \cdot (AG + AP1 - AG \cdot AP1)) -$ $(LFY \cdot UFO) \cdot (PI \cdot SEP \cdot AP3 \cdot (AG + AP1 - AG \cdot AP1))$
$\hat{f}_{EMF1}(\cdot) =$	$(1 - LFY)$
$\hat{f}_{FT}(\cdot) =$	$(1 - EMF1)$
$\hat{f}_{FUL}(\cdot) =$	$(1 - AP1) \cdot (1 - TFL1)$
$\hat{f}_{LFY}(\cdot) =$	$(1 - EMF1 \cdot TFL1)$
$\hat{f}_{PI}(\cdot) =$	$(LFY \cdot (AG + AP3 - AG \cdot AP3)) + (PI \cdot SEP \cdot AP3 \cdot (AG + AP1 - AG \cdot AP1)) -$ $(LFY \cdot (AG + AP3 - AG \cdot AP3)) \cdot (PI \cdot SEP \cdot AP3 \cdot (AG + AP1 - AG \cdot AP1))$
$\hat{f}_{SEP}(\cdot) =$	LFY
$\hat{f}_{TFL1}(\cdot) =$	$(1 - AP1) \cdot (1 - LFY) \cdot EMF1$
$\hat{f}_{UFO}(\cdot) =$	UFO
$\hat{f}_{WUS}(\cdot) =$	$WUS \cdot (1 - AG \cdot SEP)$

Table A.2: Continuous functions of FOS-GRN

A.2.2 Cyclic attractor

The gene profiles of the corresponding cyclic attractor are shown in Figure A.2.

A.3 System EMT-GRN

A.3.1 Boolean functions

The set of Boolean functions presented in Table A.3, correspondc to those molecular components that are considered on the core GRN underlying the immortalization of epithelial cells, proposed by Méndez-Lopez and collaborators in [45].

A.3.2 Set of attractors

The fixed–point attractors of the discrete–time core gene regulatory network underlying Epithelial-to-mesenchymal transition are shown in Figure A.3.

$x_i(t+1) =$	$f_i(x_1(t), x_2(t), \ldots x_n(t))$
$CYCD31(t+1) =$	$!SCF$
$SCF(t+1) =$	$!APCC\ \&\ (\ (\ E2Fb\ \&\ (\ !RBR\ \mid\ (\ !KRP1\ \&\ CYCD3;1\ )\ )\ )\ \mid MYB3R1/4\ )$
$RBR(t+1) =$	$(\ KRP1\ \mid !CYCD3;1\ )\ \&\ (\ (\ E2Fa\ \&\ !RBR\ )\ \mid MYB3R1/4\ )$
$E2Fa(t+1) =$	$(\ E2Fa\ \mid !E2Fc\ )\ \&\ !\,(\ CDKB1;1\ \&\ CYCA2;3\ )$
$E2Fb(t+1) =$	$E2Fa\ \&\ !RBR$
$E2Fc(t+1) =$	$!\,(\ SCF\ \&\ !KPR1\ \&\ CYCD3;1\ )\ \&\ (\ (\ E2Fa\ \&!RBR\ )\ \mid MYB3R1/4\ )$
$E2Fe(t+1) =$	$(\ !E2Fc\ \mid\ (\ E2Fb\ \&\ (\ !RBR\ \mid\ (\ !KRP1\ \&\ CYCD3;1\ )\ )\ )\ )\ \mid MYB77$
$MYB77(t+1) =$	$E2Fb\ \&\ (\ !RBR\ \mid\ (\ !KRP1\ \&\ CYCD3;1\ )\ )$
$MYB3R1/4(t+1) =$	$MYB77\ \mid\ (\ MYB3R1/4\ \&\ CYCB1;1\ \&\ !KRP1\ )$
$CYCB1;1(t+1) =$	$!APC/C\ \&$ $(\ MYB3R1/4\ \mid MYB77\ \mid\ (\ !RBR\ \mid\ (\ !KRP1\ \&\ CYCD3;1\ )\ \&\ E2Fb\ \&\ !E2Fc\ )\ )$
$CDKB1;1(t+1) =$	$(\ (\ !RBR\ \mid\ (\ !KRP1\ \mid CYCD3;1\ )\ )\ \&\ E2Fb\ \&\ !E2Fc\ )\ \mid MYB3R1/4\ \mid MYB77$
$CYCA2;3(t+1) =$	$!APC/C\ \&\ (\ MYB3R1/4\ \mid MYB77\ )$
$KRP1(t+1) =$	$(\ MYB77\ \mid MYB3R1/4\ )\ \&\ !\,(\ CDKB1;1\ \&\ CYCA2;3\ \&\ SFC\ )$
$APC/C(t+1) =$	$!E2Fe\ \&\ (\ MYB3R1/4\ \mid MYB3R1/4\ \mid\ (\ E2Fa\ \&\ !RBR\ )\ )$

Table A.3: Boolean functions of ACC-GRN

A.3.3 System S-EMT-GRN

In order to reduce the computational complexity for further analyses, in terms of the computational cost that arises due to the operations between matrices, we simplified the aforementioned set of Boolean functions of the EMT-GRN. The result of such simplification is shown in Table A.5, and we refer to this simplified version of the EMT-GRN as S-EMT-GRN.

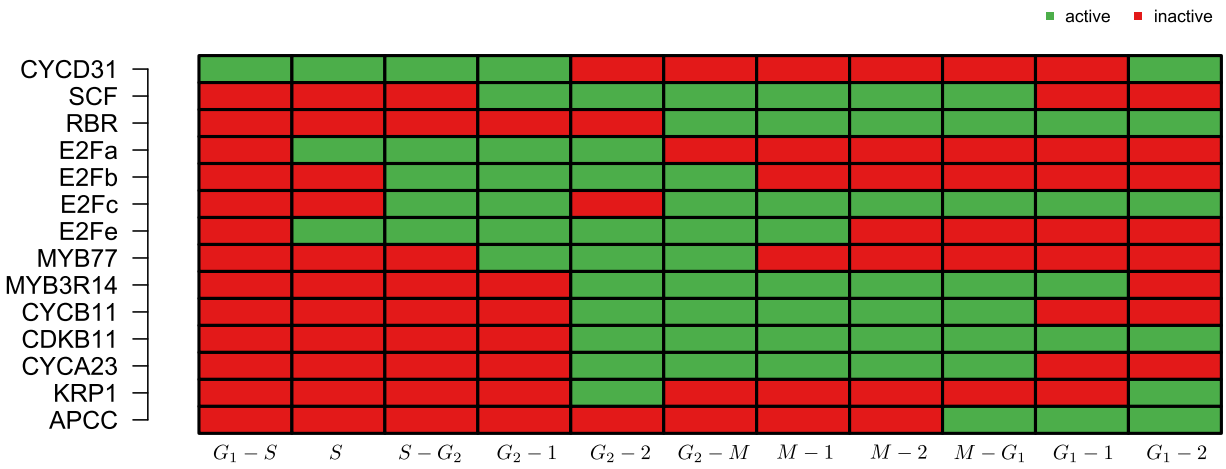


Figure A.2: Gene profiles of the single cyclic 11-length attractor of the ACC-GRN.

$\hat{f}_{CYCD31}(\cdot) =$	$1 - SCF$
$\hat{f}_{SCF}(\cdot) =$	$\min\left(1 - APCC, \max(\min\left(\max\left(1 - RBR, \min(1 - KRP1, CYCD31)\right), E2Fb\right), MYB3R14)\right)$
$\hat{f}_{RBR}(\cdot) =$	$\min(\max(KRP1, 1 - CYCD31), \max(\min(E2Fa, 1 - RBR), MYB3R14))$
$\hat{f}_{E2Fa}(\cdot) =$	$\min(\max(E2Fa, 1 - E2Fc), 1 - \min(CDKB11, CYCA23))$
$\hat{f}_{E2Fb}(\cdot) =$	$\min(E2Fa, 1 - RBR)$
$\hat{f}_{E2Fc}(\cdot) =$	$\min(1 - \min(SCF, \min(1 - KRP1, CYCD31)), \max(\min(E2Fa, 1 - RBR), MYB3R14))$
$\hat{f}_{E2Fe}(\cdot) =$	$\max(1 - E2Fc, \max(\min(E2Fb, \max(1 - RBR, \min(1 - KRP1, CYCD31))), MYB77))$
$\hat{f}_{MYB77}(\cdot) =$	$\min(E2Fb, \max(1 - RBR, \min(1 - KRP1, CYCD31)))$
$\hat{f}_{MYB3R14}(\cdot) =$	$\max(MYB77, \min(MYB3R14, \min(CYCB11, 1 - KRP1)))$
$\hat{f}_{CYCB11}(\cdot) =$	$\min(1 - APCC, \min(\max(MYB3R14, \max(MYB77, \min(\max(1 - RBR,$ $\min(1 - KRP1, CYCD31)), \min(E2Fb, 1 - E2Fc))))))$
$\hat{f}_{CDKB11}(\cdot) =$	$\max(\min(\max(1 - RBR, \min(1 - KRP1, CYCD31)), \min(E2Fb, 1 - E2Fc)),$ $\max(MYB3R14, MYB77))$
$\hat{f}_{CYCA23}(\cdot) =$	$\min(1 - APCC, \max(MYB3R14, MYB77))$
$\hat{f}_{KRP1}(\cdot) =$	$\min(\max(MYB77, MYB3R14), 1 - \min(CDKB11, \min(CYCA23, SCF)))$
$\hat{f}_{APCC}(\cdot) =$	$\min(1 - E2Fe, \max(\min(E2Fa, 1 - RBR), \max(MYB3R14, MYB77)))$

Table A.4: Continuous functions of ACC-GRN

$x_i(t+1) =$	$f_i(x_1(t), x_2(t), \dots x_n(t))$
$Snai2(t+1) =$	($!ESE2$ & $!NFkB$ & $!Snai2$) | ($!ESE2$ & $!NFkB$ & $Snai2$) | ($!ESE2$ & $NFkB$ & $!Snai2$) | ($!ESE2$ & $NFkB$ & $Snai2$) | ($ESE2$ & $NFkB$ & $Snai2$)
$ESE2(t+1) =$	($!NFkB$ & $!Snai2$ & $!ESE2$) | ($!NFkB$ & $!Snai2$ & $ESE2$) | ($!NFkB$ & $Snai2$ & $ESE2$) | ($NFkB$ & $!Snai2$ & $!ESE2$) | ($NFkB$ & $!Snai2$ & $ESE2$)
$p16(t+1) =$	($!p16$ & $!E2F$ & $p53$ & $!TELasa$ & $!Snai2$) | ($!p16$ & $!E2F$ & $p53$ & $!TELasa$ & $Snai2$) | ($!p16$ & $!E2F$ & $p53$ & $TELasa$ & $!Snai2$) | ($!p16$ & $E2F$ & $p53$ & $!TELasa$ & $!Snai2$) | ($!p16$ & $E2F$ & $p53$ & $!TELasa$ & $Snai2$) | ($!p16$ & $E2F$ & $p53$ & $TELasa$ & $!Snai2$) | ($p16$ & $!E2F$ & $!p53$ & $!TELasa$ & $!Snai2$) | ($p16$ & $!E2F$ & $p53$ & $!TELasa$ & $!Snai2$) | ($p16$ & $!E2F$ & $p53$ & $TELasa$ & $!Snai2$) | ($p16$ & $E2F$ & $!p53$ & $!TELasa$ & $!Snai2$) | ($p16$ & $E2F$ & $!p53$ & $!TELasa$ & $Snai2$) | ($p16$ & $E2F$ & $!p53$ & $TELasa$ & $!Snai2$) | ($p16$ & $E2F$ & $!p53$ & $!TELasa$ & $Snai2$) | ($p16$ & $E2F$ & $p53$ & $!TELasa$ & $!Snai2$) | ($p16$ & $E2F$ & $p53$ & $!TELasa$ & $Snai2$) | ($p16$ & $E2F$ & $p53$ & $TELasa$ & $!Snai2$) | ($p16$ & $E2F$ & $p53$ & $TELasa$ & $Snai2$) | ($p16$ & $!E2F$ & $!p53$ & $TELasa$ & $!Snai2$)
$E2F(t+1) =$	($!Rb$ & $!p53$ & $!Snai2$ & $!Cyclin$) | ($!Rb$ & $!p53$ & $!Snai2$ & $Cyclin$)
$Cyclin(t+1) =$	($!ESE2$ & $!E2F$ & $!p16$ & $!NFkB$ & $!Snai2$) | ($!ESE2$ & $!E2F$ & $!p16$ & $NFkB$ & $!Snai2$) | ($!ESE2$ & $!E2F$ & $!p16$ & $NFkB$ & $Snai2$) | ($!ESE2$ & $E2F$ & $!p16$ & $!NFkB$ & $!Snai2$) | ($!ESE2$ & $E2F$ & $!p16$ & $NFkB$ & $!Snai2$) | ($!ESE2$ & $E2F$ & $!p16$ & $NFkB$ & $Snai2$) | ($ESE2$ & $!E2F$ & $!p16$ & $!NFkB$ & $!Snai2$) | ($ESE2$ & $!E2F$ & $!p16$ & $NFkB$ & $!Snai2$) | ($ESE2$ & $E2F$ & $!p16$ & $!NFkB$ & $!Snai2$) | ($ESE2$ & $E2F$ & $!p16$ & $NFkB$ & $!Snai2$)
$TELasa(t+1) =$	($!Snai2$ & $!ESE2$) | ($Snai2$ & $!ESE2$)
$NFkB(t+1) =$	! ($!ESE2$ & $!p16$ & $!Snai2$ & $!NFkB$)
$Rb(t+1) =$	($!Cyclin$ & $!p16$ & $p53$) | ($!Cyclin$ & $p16$ & $!p53$) | ($!Cyclin$ & $p16$ & $p53$) | ($Cyclin$ & $!p16$ & $p53$) | ($Cyclin$ & $p16$ & $!p53$) | ($Cyclin$ & $p16$ & $p53$)
$p53(t+1) =$	($!p53$ & $!NFkB$ & $!TELasa$ & $!p16$ & $!Snai2$) | ($!p53$ & $!NFkB$ & $!TELasa$ & $p16$ & $!Snai2$) | ($!p53$ & $NFkB$ & $!TELasa$ & $p16$ & $!Snai2$) | ($p53$ & $!NFkB$ & $!TELasa$ & $!p16$ & $!Snai2$) | ($p53$ & $!NFkB$ & $!TELasa$ & $p16$ & $!Snai2$) | ($p53$ & $NFkB$ & $!TELasa$ & $p16$ & $!Snai2$) |

Table A.5: Boolean functions of EMT-GRN.

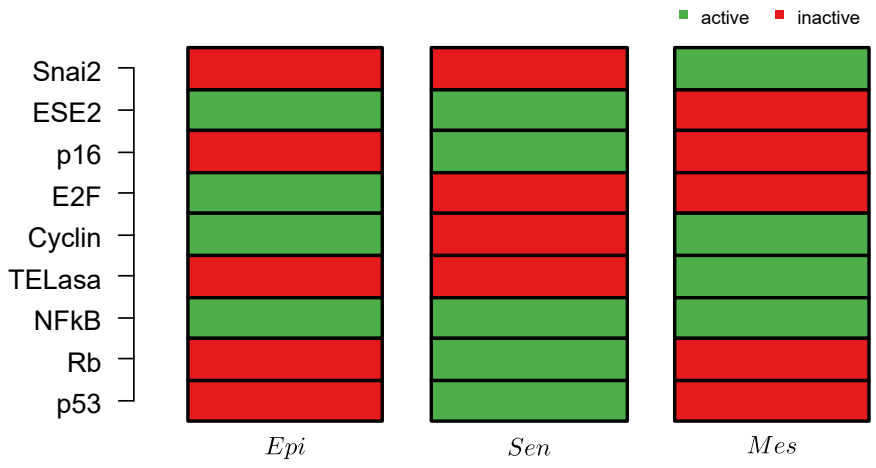


Figure A.3: Gene profiles of three fixed-point attractors of the EMT-GRN.

$x_i(t+1) =$	$f_i(x_1(t), x_2(t), \dots x_n(t))$
$Snai2(t+1) =$	$(\ Snai2\ \&\ NFkB\ )\ \vert\ !ESE2$
$ESE2(t+1) =$	$(\ ESE2\ \&\ !NFkB\ )\ \vert\ !Snai2$
$p16(t+1) =$	$(\ E2F\ \&\ p16\ )\ \vert\ (\ p16\ \&\ !Snai2\ )\ \vert\ (\ p53\ \&\ !Snai2\ )\ \vert\ (\ p53\ \&\ !TELasa\ )$
$E2F(t+1) =$	$!p53\ \&\ !Snai2\ \&\ !Rb$
$Cyclin(t+1) =$	$(\ !ESE2\ \&\ !p16\ \&\ NFkB\ )\ \vert\ (\ !Snai2\ \&\ !p16\ )$
$TELasa(t+1) =$	$!ESE2$
$NFkB(t+1) =$	$Snai2\ \vert\ ESE2\ \vert\ p16\ \vert\ NFkB$
$Rb(t+1) =$	$p16\ \vert\ p53$
$p53(t+1) =$	$(\ !Snai2\ \&\ p16\ \&\ !TELasa\ )\ \vert\ (\ !Snai2\ \&\ !TELasa\ \&\ !NFkB\ )$

Table A.6: Boolean functions of S-EMT-GRN.

$\hat{f}_{Snai2}(\cdot) =$	$Snai2 \cdot NFkB + (1 - ESE2) - Snai2 \cdot NFkB \cdot (1 - ESE2)$
$\hat{f}_{ESE2}(\cdot) =$	$(1 - Snai2) + ESE2 \cdot (1 - NFkB) - (1 - Snai2) \cdot ESE2 \cdot (1 - NFkB)$
$\hat{f}_{p16}(\cdot) =$	$E2F \cdot p16 + p16 \cdot (1 - Snai2) - E2F \cdot p16 \cdot p16 \cdot (1 - Snai2) +$ $p53 \cdot (1 - Snai2) + p53 \cdot (1 - TELasa) - p53 \cdot (1 - Snai2) \cdot p53 \cdot (1 - TELasa) -$ $E2F \cdot p16 + p16 \cdot (1 - Snai2) - E2F \cdot p16 \cdot p16 \cdot (1 - Snai2) \cdot$ $p53 \cdot (1 - Snai2) + p53 \cdot (1 - TELasa) - p53 \cdot (1 - Snai2) \cdot p53 \cdot (1 - TELasa)$
$\hat{f}_{E2F}(\cdot) =$	$(1 - p53) \cdot (1 - Snai2) \cdot (1 - Rb)$
$\hat{f}_{Cyclin}(\cdot) =$	$(1 - Snai2) \cdot (1 - p16) + (1 - ESE2) \cdot (1 - p16) \cdot NFkB -$ $(1 - Snai2) \cdot (1 - p16) \cdot (1 - ESE2) \cdot (1 - p16) \cdot NFkB$
$\hat{f}_{TELasa}(\cdot) =$	$1 - ESE2$
$\hat{f}_{NFkB}(\cdot) =$	$(Snai2 + ESE2 - Snai2 \cdot ESE2) + (p16 + NFkB - p16 \cdot NFkB) -$ $(Snai2 + ESE2 - Snai2 \cdot ESE2) \cdot (p16 + NFkB - p16 \cdot NFkB)$
$\hat{f}_{Rb}(\cdot) =$	$p16 + p53 - p16 \cdot p53$
$\hat{f}_{p53}(\cdot) =$	$((1 - Snai2) \cdot p16 \cdot (1 - TELasa) + (1 - Snai2) \cdot (1 - TELasa) \cdot (1 - NFkB)) -$ $((1 - Snai2) \cdot p16 \cdot (1 - TELasa) \cdot (1 - Snai2) \cdot (1 - TELasa) \cdot (1 - NFkB))$

Table A.7: Continuous functions of S-EMT-GRN

Appendix B

Reachability analysis support information

Introduction

We deal in this appendix with the support material corresponding to the reachability analysis proposed in this manuscript.

B.1 FM-like transitory state appears in the IM-to-FOP transitions.

We found that induced trajectories from any of the sub-zones of the IM to floral organs, visit a transitory state with the gene profile of an FM-like attractor, even though the FOS-GRN does not recover it. To illustrate this, we selected four developmental trajectories; $I_1 - \mathbf{FT} \rightarrow$SE, $I_2 - \mathbf{FT} \rightarrow$PE2, $I_3 - \mathbf{FT} \rightarrow$ST2, $I_4 - \mathbf{FT} \rightarrow$CAR. Since these trajectories are induced by u, the Boolean function of the gene FT changes as follows:

$$FT(t+1) = f_{FT}(\cdot)\,|\,u$$

The results for each IM-to-FOP transition are shown in the following Table B.1. As we noticed, at $s = 4$ a specific gene activation configuration is reached,which is very similar to an expected FM-like state.

State	AG	AP1	AP2	AP3	EMF1	FT	FUL	LFY	PI	SEP	TFL1	UFO	WUS	s
I_1	0	0	0	0	1	0	0	0	0	0	1	0	0	0
	0	0	0	0	1	1	0	0	0	0	1	0	0	1
	0	1	0	0	1	1	0	0	0	0	1	0	0	2
	0	1	0	0	1	0	0	0	0	0	0	0	0	3
FM-like	0	1	1	0	1	0	0	1	0	0	0	0	0	4
	0	1	1	0	0	0	0	1	0	1	0	0	0	5
SE	0	1	1	0	0	1	0	1	0	1	0	0	0	6
I_2	0	0	0	0	1	0	0	0	0	0	1	1	0	0
	0	0	0	0	1	1	0	0	0	0	1	1	0	1
	0	1	0	0	1	1	0	0	0	0	1	1	0	2
	0	1	0	0	1	0	0	0	0	0	0	1	0	3
FM-like	0	1	1	0	1	0	0	1	0	0	0	1	0	4
	0	1	1	1	0	0	0	1	0	1	0	1	0	5
PE	0	1	1	1	0	1	0	1	1	1	0	1	0	6
I_3	0	0	0	0	1	0	0	0	0	0	1	1	1	0
	0	0	0	0	1	1	0	0	0	0	1	1	1	1
	0	1	0	0	1	1	0	0	0	0	1	1	1	2
	0	1	0	0	1	0	0	0	0	0	0	1	1	3
FM-like	0	1	1	0	1	0	0	1	0	0	0	1	1	4
	0	1	1	1	0	0	0	1	0	1	0	1	1	5
	1	1	1	1	0	1	0	1	1	1	0	1	1	6
	1	0	1	1	0	1	0	1	1	1	0	1	0	7
ST	1	0	1	1	0	1	1	1	1	1	0	1	0	8
I_4	0	0	0	0	1	0	0	0	0	0	1	0	1	0
	0	0	0	0	1	1	0	0	0	0	1	0	1	1
	0	1	0	0	1	1	0	0	0	0	1	0	1	2
	0	1	0	0	1	0	0	0	0	0	0	0	1	3
FM-like	0	1	1	0	1	0	0	1	0	0	0	0	1	4
	0	1	1	0	0	0	0	1	0	1	0	0	1	5
	1	1	1	0	0	1	0	1	0	1	0	0	1	6
	1	0	1	0	0	1	0	1	1	1	0	0	0	7
CAR	1	0	1	0	0	1	1	1	1	1	0	0	0	8

Table B.1: FM-like transitory state appears on the IM-to-FOP transitions

Appendix C

Research publications

Introduction

The work reported here has been exposed to the System Biology international research community. In this appendix we include three reports that were presented at the *19th International Conference on Systems Biology (ICSB 2018), October 18 - November 1, 2018, Lyon, France*:

1: Controllability analysis of the core gene regulatory network underlying epithelial-to-mesenchymal transition in the context of epithelial cancer.

2: Identification of gene morphogenetic activity functionalities through reachability analysis: the *Arabidopsis thaliana* flower morphogenesis case.

3: Exploring cell cycle gene regulatory network dynamics through the algebraic analysis of its structural reachability properties.

C.1 Controllability analysis of the core gene regulatory network underlying epithelial-to-mesenchymal transition in the context of epithelial cancer

Eduardo CHAIREZ-VELOZ[1], Jose Luis CALDU-PRIMO[2], Jose DAVILA-VELDERRAIN[3], Alberto SORIA-LÓPEZ[1], Elena R ALVAREZ-BUYLLA[4], Juan Carlos MARTINEZ-GARCIA[1].

[1]Automatic Control Department, Cinvestav-IPN.

[2]PhD Program on Biomedical Science, UNAM, MEXICO CITY, Mexico.

[3]CSIL, Massachusetts Institute of Technology, CAMBRIDGE, United States.

[4]Instituto de Ecología, & Centro de Ciencias de la Complejidad, UNAM, MEXICO CITY, Mexico

19th International Conference on Systems Biology (ICSB 2018), October 18 - November 1, 2018, Lyon, France.

Oral Communications

Controllability analysis of the core gene regulatory network underlying epithelial-to-mesenchymal transition in the context of epithelial cancer

Eduardo CHAIREZ-VELOZ[1], José Luis CALDU-PRIMO[2], José DAVILA-VELDERRAIN[3], Alberto SORIA-LÓPEZ[1], Elena R ALVAREZ-BUYLLA[4], Juan Carlos MARTINEZ-GARCIA[* 1]

[1]Automatic Control Department, Cinvestav-IPN, [2]PhD Program on Biomedical Science, UNAM, MEXICO CITY, Mexico, [3]CSIL, Massachusetts Institute of Technology, CAMBRIDGE, United States, [4]Instituto de Ecología & Centro de Ciencias de la Complejidad, UNAM, MEXICO CITY, Mexico

Secondary topic : Modelling Networks and Circuits

Your abstract : As a complex developmental process, Epithelial-to-mesenchymal transition (EMT) is involved in morphogenesis, tissue regeneration and cancer progression. This dynamical process is characterized by a series of cell-state transitions, in which belonging to epithelial tissue loose their epithelial characteristics, and gain mesenchymal properties (e.g., increasing motility). Therefore, some experimentally grounded gene regulatory networks (GRNs) models have started to be proposed to uncover the EMT regulatory core, and thus contribute to the understanding of the EMT regulation and to guide experiments by generating testable hypotheses. Nevertheless, systematic analysis that elucidate the specific role that the involved genes (acquired by their collaboration in the network), play in transitioning not only between the two cell states, but also among intermediate states, are still very scarce. In attempt to contribute in such a need, we propose an analytical procedure based on algebraic approaches and built around Semi-Tensor Product approach (STP). We illustrate the procedure through the exploration of the structural controllability properties of the low-dimensional Boolean GRN underlying the immortalization of epithelial cells, which recovers the specific gene expression profiles that correspond to the epithelial, senescent, and mesenchymal stem-like phenotypes. Our findings suggest that there exist 9 different inducible transitions between the three main phenotypes, by a suitable external Boolean input connected to a particular gene of the network. These transitions are tested both in the Boolean and in its approximated continuous model. In addition, we found that the computed controllability matrix associated to the aforementioned EMT-GRN, provide relevant information that concerns the perturbations that give rise to some intermediate states, since its entries indicate whether the i-state (any possible gene activation configuration) is reachable from the j-state, under a set of admissible Boolean inputs. Finally, we discuss this data from a biological perspective, and we concluded that controllability analysis can give us insights on the role of the genes in the context of the network as a whole. This is particularly important when tackling EMT dynamics associated to the onset and progression of cell-state trajectories involved in disease.

Disclosure of Interest: None Declared

Controllability analysis of the core gene regulatory network underlying Epithelial-to-Mesenchymal transition in the context of epithelial cancer

Chairez-Veloz J.[1], Caldú-Primo J.L.[2,3], Dávila-Velderrain J.[3,4], Soria-López A.[1], Álvarez-Buylla E.R.[3,5], Martínez-García J.C.[1]

[1] Automatic Control, CINVESTAV-IPN, México [2] Biomedical Sciences, UNAM, México [3] Centro de Ciencias de la Complejidad, UNAM, México, [4] CSIL, MIT, United States, [5] Instituto de Ecología, UNAM, México

Speaker: José Eduardo Chairez Veloz
jechairez@ctrl.cinvestav.mx

The 19th International Conference on Systems Biology, Lyon, France.

Outline

Motivation (1/2)

About Gene Regulatory Networks (GRNs) models:

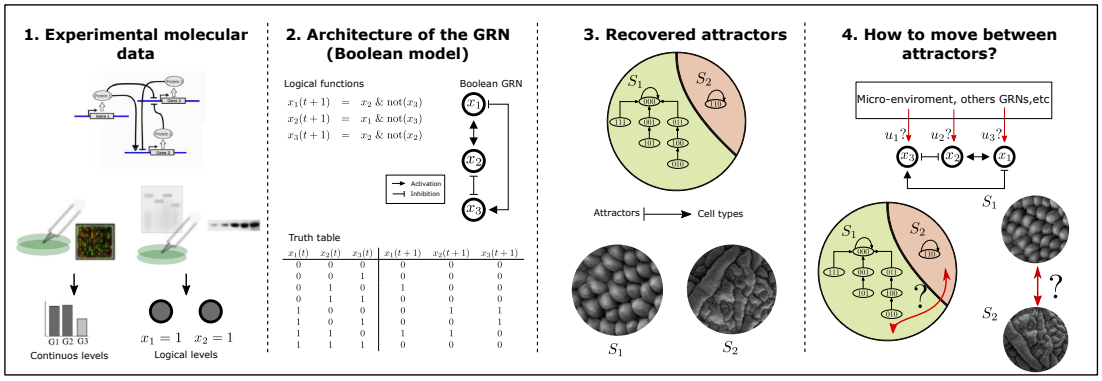


Figure 1: From experimental data on gene function and interactions, to a Boolean dynamic GRN and its attractors.

Motivation (2/2)

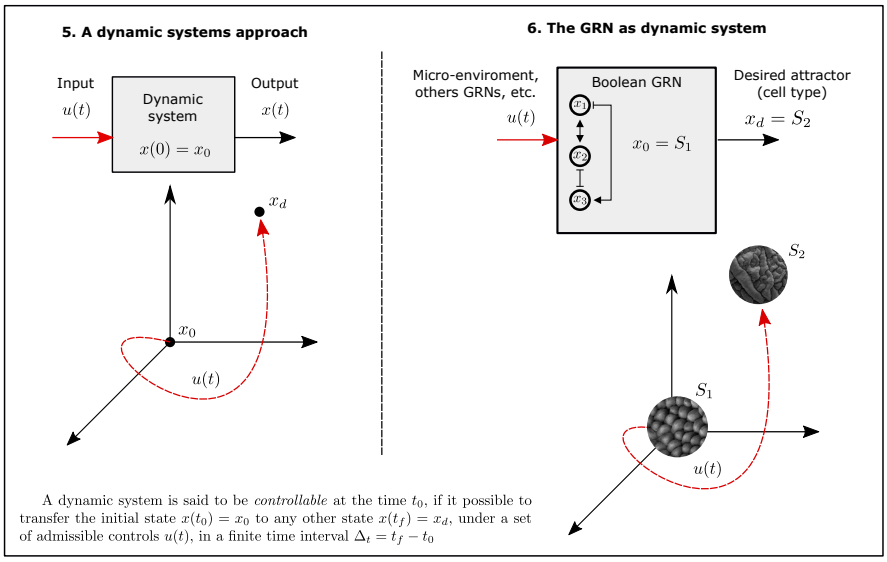


Figure 2: Transition among pairs of attractors from a Control Theory perspective.

Objective

Main objective

Explore the structural controllability properties of the core GRN underlying Ephitelial-to-Mesenchymal transition (EMT-GRN) in the context of epithelial cancer, in order to provide information about the role of the genes in the context of the network as a whole, indicating how they could be interacting with the micro-enviroment or/and others GRN modules.

Mathematical description of GRNs

A synchronous Boolean Network (BN) is a discrete-time dynamic system of n Boolean variables. Its *state of expression* is given by a vector of Boolean variables $\begin{bmatrix} x_1(t) & x_2(t) & \dots & x_n(t) \end{bmatrix}^T$, and changes, in discrete fixed-steps as follows:

$$x_i(t+1) = f_i(x_1, x_2, \dots, x_n) \tag{1}$$

$f_i : \mathcal{D}^n \to \mathcal{D}$ is the Boolean function related to the node x_i, which is built in agreement to the combinatorial action of its regulators. In experimentally grounded Boolean GRN models, a particular case of BN, the set of f_i, $i = 1, 2, \dots, n$, expresses the relationship between the genes that share regulatory interactions, involved in the process of interest and derived from experimental data

Algebraic form of a Boolean GRN via STP [Cheng, 2012]

The Semi-Tensor Product approach (STP) was used to rewrite the set of Boolean functions into algebraic operators. Following, the BN described by the Eq. (1) is thus converted into a conventional discrete-time linear system as follows:

$$x(t+1) = L \ltimes x(t) \tag{2}$$

where $x(t) = \ltimes_{i=1}^{n} x_i(t)$ is the STP of the Boolean variables, and L is called the *network transition matrix* of the BN. The matrix $L \in \mathcal{L}_{2^n \times 2^n}$ is enough to describe the full dynamic of the BN, since Eq. (2) constitutes a linear mapping with respect to each arguments.

Algebraic form of a Boolean Control GRN via STP

In order to understand how the BN could be interacting with its entourage, and consequently modifying its dynamics, a set of Boolean inputs $u_j(t)$ are connected in some specific nodes. Thus,

$$x_i(t+1) = \bar{f}_i\left(x_1, x_2, \ldots, x_n, u_1, u_2, \ldots, u_m\right) \tag{3}$$

In the same way as mentioned before, we used STP to convert (3) into a bilinear system. So,

$$x(t+1) = \bar{L}u(t)x(t) \tag{4}$$

The computation of the controllability matrix ($\mathcal{C}$)

Once the network is given in its algebraic form, described by (4), the controllability matrix, according to [Cheng, 2011], can be computed by:

$$\mathcal{C} = \sum_{s=1}^{2^{m+n}}{}_{\mathcal{B}} \sum_{i=1}^{2^{m}}{}_{\mathcal{B}} \mathrm{Blk}_i \left(\bar{L}^{(s)} \right) \in \mathcal{B}_{2^n \times 2^n} \tag{5}$$

where $\sum_{\mathcal{B}}$ and $L^{(s)}$ denote an iterative Boolean addition ($+_{\mathcal{B}}$) and Boolean product ($\times_{\mathcal{B}}$), respectively. Its entries indicate whether the i-state (any possible gene activation configuration) is reachable from the j-state, under a set of admissible Boolean inputs

Properties of matrix ($\mathcal{C}$) [Cheng, 2011]

1. $x_d = \delta_{2^n}^i$ is reachable from $x_0 = \delta_{2^n}^j$ if and only if $c_{ij} > 0$.
2. The system is controllable at $x_0 = \delta_{2^n}^j$ if and only if $\mathrm{Col}_j(\mathcal{C}) > 0$.
3. The system is controllable if and only if $\mathcal{C} > 0$.

Main result: the controllability analysis (1/3)

The proposed controllability analysis, which harnesses the structural controllability properties of the network, consists in five main steps:

- *Step 1.* Initialize $k = 1$. Then, given the Boolean GRN in its algebraic form, recover the experimentally observed attractors.
- *Step 2.* Add a control input u on the k-th node of the BN. If $k = n$, go to Step 5; otherwise, go to the next step.
- *Step 3.* Convert the BCN into its algebraic form, and then compute the k-th controllability matrix $\mathcal{C}_k$.
- *Step 4.* Set $k = k + 1$ and go back to Step 2.
- *Step 5.* Save $\mathcal{C}_k$, $k = 1, 2, \ldots, n$. Stop.

Main result: the controllability analysis: (2/3)

After a straightforward computation, the set of C_k, $k = 1, 2, \ldots, n$ contains all the *available state trajectories* among pairs of states, when the k-th gene is perturbed, if and only if $(c_{ij})_k > 0$. Furthermore, those trajectories can be restricted to a subset associated to the steady-states transitions, also called as *attractor transitions*, particular relevant for developmental processes in Boolean GRNs models.

Main result: the controllability analysis (3/3)

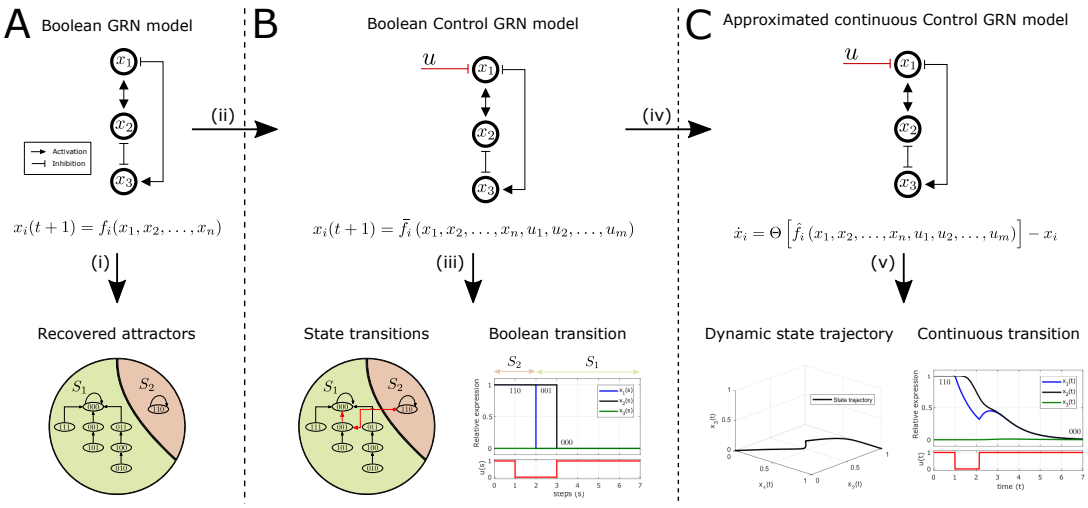


Figure 3: Overview of the controllability analysis

The Boolean EMT-GRN model [Méndez-López, 2017]

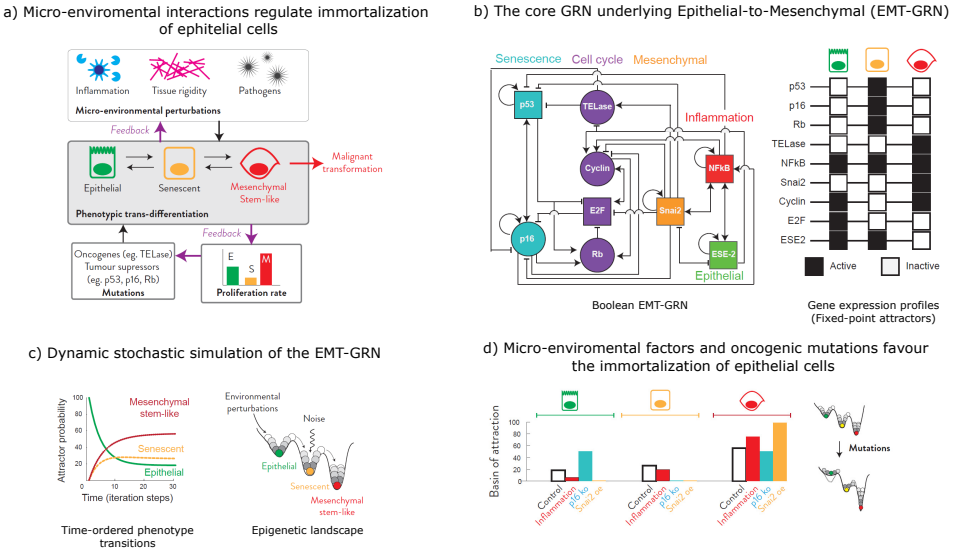


Figure 4: GRN underlying the immortalization of ephitelial cells

Attractor transitions on the EMT-GRN

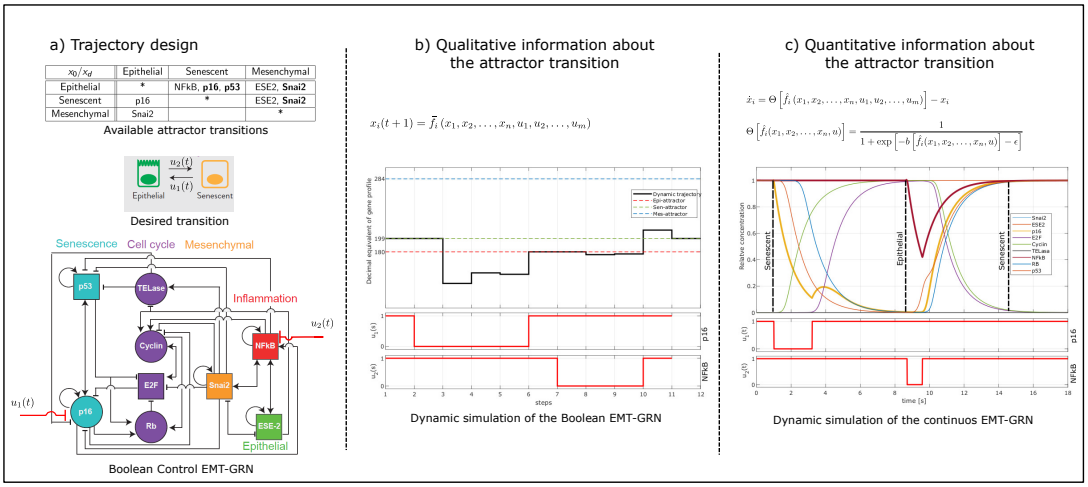


Figure 5: Sen-to-Epi transition via p16, and Epi-to-Sen transition via NFkB

Conclusions and perspectives [1/2]

- A systematic analysis that harnesses the structural controllability properties of Boolean GRN was proposed. Particularly, we applied this approach on the EMT-GRN.
- Such controllability analysis provided information about how some nodes of the network must be altered in order to produce an attractor transition.
- We validated that such intervention of the node produced the same attractor transition on the continuous GRN model, recovering not only qualitative but also quantitative information, *i.e.* the relative concentration of the genes during the transition.

Conclusions and perspectives [2/2]

- Particularly, we observed on the continuous EMT-GRN that switching-off NFkB during an interval of time, is sufficient to promote an Epi-to-Sen transition, even though both cases NFkB is high. This result suggests that some genes of the network only require a momentary manipulation to produce a qualitative change. Such information cannot be obtained by mutant analyses.
- All this information must be discussed from a biological perspective as future work, since this is particularly relevant when tackling EMT dynamics associated to the onset and progression of cell-state trajectories involved in disease.

References


Méndez-López et al. (2017)
Gene regulatory network underlying the immortalization of epithelial cells
BMC Systems Biology 11(24).


Cheng, D. & Qi, H. (2012)
Controllability and observability of Boolean Control Networks.
IET Control. Theory & Appl. (6) 24772484


Cheng, D. & Qi, H. (2011)
Analysis and Control of Boolean Networks.
First Edition. Springer

C.2 Identification of gene morphogenetic activity functionalities through reachability analysis: the Arabidopsis thaliana flower morphogenesis case

Chairez-Veloz J.E.[1], Chávez-Hernández C.E. [2], Soria-López A.[4], Álvarez-Buylla E.R.[2,3], Martínez-García J.C.[4].

[1]PhD Program on Automatic Control, Cinvestav-IPN.

[2]Centro de Ciencias de la Complejidad, UNAM, MEXICO CITY, Mexico.

[3]Instituto de Ecología, UNAM, MEXICO CITY, Mexico

[4]Automatic Control Department, Cinvestav-IPN.

19th International Conference on Systems Biology (ICSB 2018), October 18 - November 1, 2018, Lyon, France.

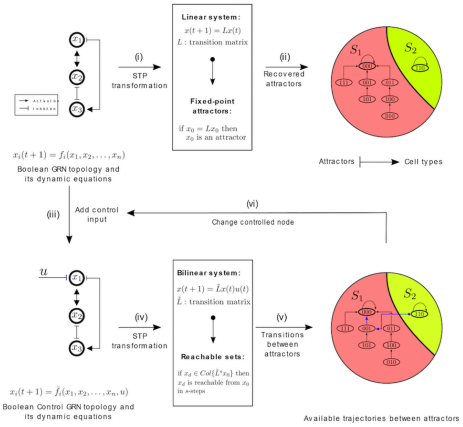
Identification of gene morphogenetic activity functionalities through reachability analysis: the *Arabidopsis thaliana* flower morphogenesis case

Chairez-Veloz J.E.[1], Chávez-Hernández C.E.[2], Soria-López A.[4], Álvarez-Buylla E.R.[2,3], Martínez-García J.C.[4]

[1]*PhD Program on Automatic Control, Cinvestav-IPN,* [2]*Centro de Ciencias de la Complejidad, UNAM,* [3]*Instituto de Ecología, UNAM,* [4]*Automatic Control, Cinvestav-IPN*

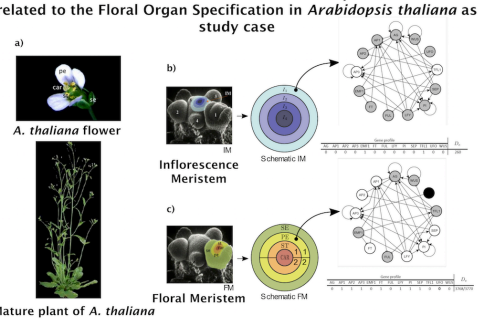
Abstract: We are concerned here by transient dynamics in morphogenetic developmental patterns. More specifically, we explore the transitions between cell types understood as dynamical attractors of the underlying developmental Gene Regulatory Networks (GRNs). Taking as a case of study the Floral Organ Specification in *Arabidopsis thaliana* (FOS-GRN), we illustrate a methodology intended to uncover the structural properties of reachability in abstract dynamical systems. The proposed methodology, based on an algebraic description of the low-dimensional FOS-GRN (first described in Boolean discrete terms), addresses the uncovering of the functional role played by specific genes. In particular, it is shown that certain genes, for example, the Wushel (WUS) and Unusual Floral Organs (UFO) genes acquire, in the context of their collaboration in the regulatory network, sufficient properties for the determination of well-identified morphogenetic transitions (*i.e.* a suitable manipulation of WUS on petal primordium causes a transition to sepal primordium), showing that alterations in floral patterning can be caused by changing spatial cues in the meristem. Even if we do not establish the molecular nature of the manipulation, we concluded that reachability analysis can give us some useful insights into the specific role that individual genes have in the context of the network as a whole.

Proposed methodology for the reachability analysis of Boolean GRNs



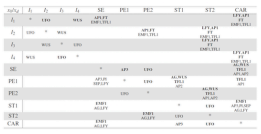
The starting point is an experimentally grounded low-dimensional Boolean GRN. *(i)* Its set of logical rules are rewriting into algebraic operators via Semi-Tensor product (STP) [1]. *(ii)* This linear representation enable us to find the attractors (fixed-point or cyclic), which are commonly associated to distinct cell types in some GRN, according to their gene profile. *(iii)* A Boolean control input u is connected in the i-th gene, overruling its corresponding logical rule. *(iv)* We then rewrite the set of new logical rules by using STP. *(v)* This bilinear representation allow us to identify as available trajectories, when u is able to promote a transition from an initial attractor to another. Finally, *(vi)* the process is repeated until all the genes are explored.

The Boolean GRN that recovers the ten fixed-point attractors related to the Floral Organ Specification in *Arabidopsis thaliana* as study case

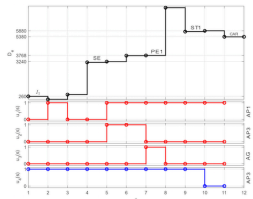


(a) This low-dimensional Boolean GRN recovers ten fixed-points attractors, in its last version in [2]. **(b)** Four of them are associated with both the structural and the dynamical behavior of the inflorescence meristem (IM), found in the apex of mature plants. **(c)** The other six corresponds to sepal (SE), petal (PE1 and PE2), stamen (ST1 and ST2), and carpel (CAR) primordial cells, that emerge from the flanks of the flower meristem (FM) in helicoidal patterns

Available homeotic and developmental trajectories found in the FOS-GRN through the reachability analysis



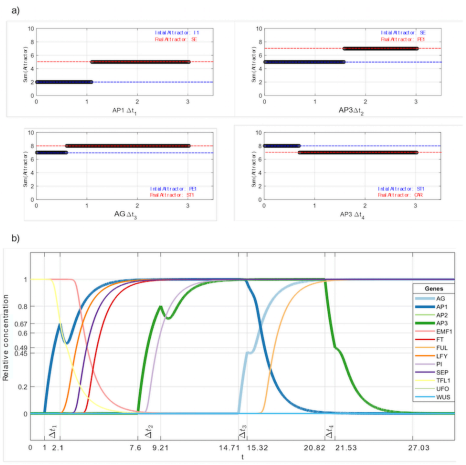
a) Available trajectories among all possible pairs of attractors



b) Inducing the sequence of gene activation observed in real flowers

In **(a)**, the crossover contains the potential gene to transit from x_0 to x_d. In bold letters, the control inputs switch-on the gene while the reminder switch-off. Moreover, the (*) means that the same attractor is the initial and the final condition. **(b)** shows the developmental and homeotic trajectory that mimics the chronological sequence observed during flower morphogenesis, in this case induced by external control inputs.

The role of the potential genes is maintained on the continuous FOS-GRN model



In **(a)**, the bifurcation diagrams for the minimum time of manipulation of each gene by u, sufficient to produce qualitative changes in the sum of the active genes. Based on these times,, each control input is built and their values are interpolated in the solution of the ODE's system. In **(b)**, we show the relative concentrations of each gene under the action of u.

Contact: Corresponding author: Juan Carlos Martínez-García (juancarlos.martinezg@gmail.com)

References: [1] Cheng, D. and Qi, H, (2012) *Controllability and Observability of Boolean Networks,* IET Control Theory & Applications, 6 (16).
[2] Álvarez-Buylla , E.R. *et. al.*, (2010) *A From ABC genes to regulatory networks, epigenetic landscapes and flower morphogenesis.* Semin. Cell Dev. Biol. 21, 108-117

Acknowledgements: UNAM Papiit-208517, CONACYT 240180 and 280629

C.3 Exploring cell cycle gene regulatory network dynamics through the algebraic analysis of its structural reachability properties

Chairez-Veloz J.E.[1], Ortiz-Gutiérrez E.[2], Arciniega-Gonzalez J.A.[2], Álvarez-Buylla E.R.[2,3], Martínez-García J.C.[4].

[1]PhD Program on Automatic Control, Cinvestav-IPN.

[2]Centro de Ciencias de la Complejidad, UNAM, MEXICO CITY, Mexico.

[3]Instituto de Ecología, UNAM, MEXICO CITY, Mexico

[4]Automatic Control Department, Cinvestav-IPN.

19th International Conference on Systems Biology (ICSB 2018), October 18 - November 1, 2018, Lyon, France.

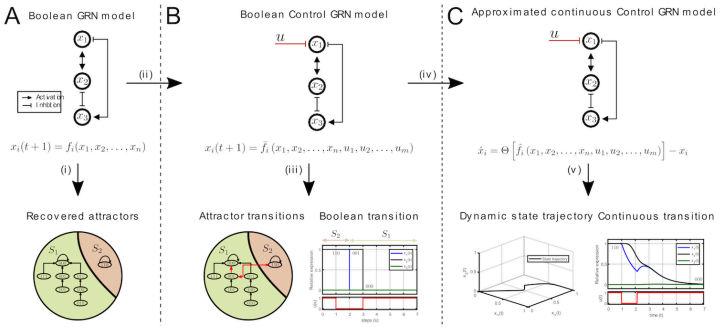
Exploring cell cycle gene regulatory network dynamics through the algebraic analysis of its structural reachability properties

Chairez-Veloz J.E.[1], Ortiz-Gutiérrez E.[2], Arciniega-González J.A.[2], Álvarez-Buylla E.R.[2,3], Martínez-García J.C.[4]

[1]*PhD Program on Automatic Control, Cinvestav-IPN,* [2]*Centro de Ciencias de la Complejidad, UNAM,* [3]*Instituto de Ecología, UNAM,* [4]*Automatic Control, Cinvestav-IPN*

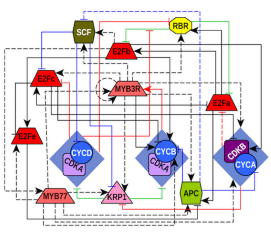
Abstract: The spatio-temporal regulation of cell cycle (CC) in multicellular organisms is essential during morphogenetic process since it relies on the coordinated progression of CC and its variations, *i.e.* CC arrest, reactivation and endoreduplication, which in coordination with cell differentiation, allow the emergence of morphogenetic patterns. Its experimentally grounded regulatory interactions that interlink these processes can be integrated into a modular structured Gene Regulatory Network (GRN) and mathematically described through discrete Boolean models. Some low-dimensional Boolean GRNs have been proposed to recover the cyclic gene activation configurations observed in different CC stages, and they have been validated via robustness and mutant analysis. In spite of the recent interest and advances in CC-GRNs models, systematical analyses that elucidate the role of individual genes when perturbated by combinatorial switched-on and off actions, implicated in disrupting the cyclic behavior, are still very scarce. Following this idea, we put forward an analytical procedure which harnesses the structural reachability properties of Boolean GRNs in its algebraic form, obtained via Semi-Tensor Product approach (STP). As a concrete study case, we used the Boolean GRN model that recovers the cyclic behavior of *Arabidopsis thaliana* CC. Our findings suggests that their molecular components acquire a novel potentiality when are temporally switched-on and off throughout the CC progression, not only in the Boolean GRN, but also in the approximated continuous model. For example, we found that suitable time-manipulations by external control inputs in genes such as APC/C and MYB3R1/4 induce endoreduplication and provoke CC arrest, respectively. We concluded that reachability analysis can give us insights about how developmental, physiological and/or environmental cues could be acting on individual genes to coordinated CC network dynamics with spatio-temporal CC regulation.

Proposed methodology for the reachability analysis of Boolean GRNs.

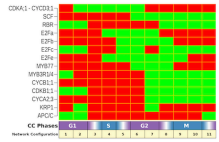


The starting point is **(A)** an experimentally grounded low-dimensional Boolean GRN model in which, *(i)* its Boolean functions are rewriting into algebraic operators via Semi-Tensor product (STP), also called as the algebraic form of the network [1]. The attractors, either fixed-points or cyclics, are thus identified in the linear system, after an exhaustive exploration for every initial condition. *(ii)* A Boolean control input u is connected in the i-th gene, in order to overrule its corresponding Boolean function, and thus promote a spontaneous switch-off/on on the gene. **(B)** The new Boolean Control GRN model is then transformed into its algebraic form via STP. This bilinear representation enables us to *(iii)* identify the reachable subsets, related to all the previous computed attractors, which we name them as available trajectories. In **(C)**, we also recover quantitative information about the relative concentration of the gene activity by *(v)* simulating the action of the previous Boolean control input in the approximated continuous model.

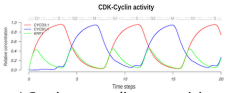
The Boolean CC-GRN that recovers the cyclic behavior of *Arabidopsis thaliana* CC as study case.



a) GRN of the *A. thaliana* CC



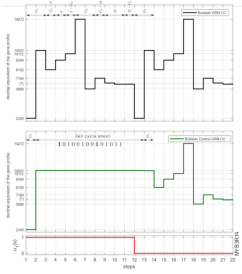
b) Boolean cyclic gene activity



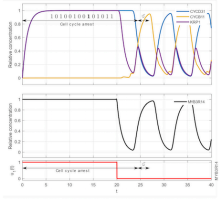
c) Continuous cyclic gene activity

This Boolean CC-GRN was proposed by Ortiz-Gutierrez *et. al.* in [2]. The model recovers the cyclic behavior of *Arabidopsis thaliana* cell cycle (referred hereafter as ACC-GRN). **(a)** It is conformed by 14 nodes and mostly nonlinear interactions, which (**b**) converges into a single cyclic attractor that comprises the G1, S, G2 and M phases. In **(c)**, the activity of the CDK-Cyclin obtained from the continuous version of ATCC-GRN. Such activity is a limit factor to pass the G1/S and the G2/M checkpoints.

MYB3R1/4 promotes the CC arrest in *Arabidopsis thaliana*, according to the reachability analysis of the ACC-GRN.



a) MYB3R1/4 promotes CC arrest in the Boolean model



b) MYB3R1/4 promotes CC arrest in the continuous model

(a) After the reachability analysis of the Boolean ACC-GRN, we found that a high level of MYB3R1/4, originated by an external control input u, is sufficient to drive the system from a G1-phase to CC arrest. This CC arrest holds during the action of u. Additionally, **(b)** the CDK-Cyclin activity is also disrupted in the continuous model. This result is particular relevant since it has been recently experimentally found that MYB3R1/4 has also a repressive function during the CC, even though it was originally described as an activator of the CC [3].

Contact: Corresponding author: Juan Carlos Martínez-García (juancarlos.martinezg@gmail.com)

References: [1] Cheng, D. and Qi, H (2012) *Controllability and Observability of Boolean Networks*, IET Control Theory & Applications, 6 (16).
[2] Ortiz-Gutiérrez E. *et. al.* (2015) *A dynamic Gene Regulatory Network Model that Recovers the Cyclic Behavior of Arabidopsis thaliana* Cell Cycle. PLoS Comput Biol 11(9):e1004486.
[3] Chen P, *et. al.* (2017) *Arabidopsis R1R2R3-Myb proteins are essential for inhibiting cell division in response to DNA damage.* Nat Comm 8 (635).

Acknowledgements: UNAM Papiit-208517, CONACYT 240180 and 280629

Bibliography

[1] T. Akutsu, M. Hayashida, W Ching, and M Ng. Control of Boolean networks: Hardness results and algorithms fortree structured networks. *Theor. Biol.*, 244:670–679, 2007. doi: 10.1016/j.jtbi.2006.09.023.

[2] T. Akutsu, S. Kosub, A. A. Melkman, and T. Tamura. Finding a periodic attractor of a Boolean network. *Transactions on Comput. Biol.*, 9:1410–1421, 2012. doi: 10.1109/TCBB.2012.87.

[3] Tatsuya Akutsu, Morihiro Hayashida, Wai Ki Ching, and Michael K. Ng. Control of Boolean networks: Hardness results and algorithms for tree structured networks. *Journal of Theoretical Biology*, 244(4):670–679, 2007. ISSN 00225193. doi: 10.1016/j.jtbi.2006.09.023.

[4] M. Ali Al-Radhawi and Eduardo D. Sontag. Analysis of a reduced model of epithelial-mesenchymal fate determination in cancer metastasis as a singularly-perturbed monotone system, 2020.

[5] Elena R. Álvarez-Buylla, lvaro Chaos, Maximino Aldana, Mariana Benítez, Yuriria Cortes-Poza, Carlos Espinosa-Soto, Diego A. Harta Sánchez, R. Beau Lotto, David Malkin, Gerardo J. Escalera Santos, and Pablo Padilla-Longoria. Floral morphogenesis: Stochastic explorations of a gene network epigenetic landscape. *PLoS ONE*, 3 (11), 2008. ISSN 19326203. doi: 10.1371/journal.pone.0003626.

[6] Elena R. Alvarez-Buylla, Eugenio Azpeitia, Rafael Barrio, Mariana Benítez, and Pablo Padilla-Longoria. From ABC genes to regulatory networks, epigenetic landscapes and flower morphogenesis: Making biological sense of theoretical approaches.

Seminars in Cell and Developmental Biology, 21(1):108–117, 2010. ISSN 10849521. doi: 10.1016/j.semcdb.2009.11.010.

[7] Eugenio Azpeitia, José Davila-velderrain, Carlos Villarreal, and Elena R. Álvarez-Buylla. Flower Development. *Flower Development*, 1110:441–469, 2014. doi: 10.1007/978-1-4614-9408-9. URL `http://link.springer.com/10.1007/978-1-4614-9408-9`.

[8] Eugenio Azpeitia, Daniel Muñoz, Stalin andGonzález-Tokman, Mariana Esther Martínez-Sánchez, Nathan Weinstein, Naldi Aurélien, Elena R. Álvarez-Buylla, David A Rosenblueth, and Luis Mendoza. The combination of the functionalities of feedback circuits is determinant for the attractors' number and size in pathway-like Boolean networks. *Scientific Reports*, 42023(7), 2017. doi: 10.1038/srep42023.

[9] X. Bao, R. Franks, J. Levin, and Z. Liu. Repression of AGAMOUS by BELLRINGER in floral and inflorescence meristems. *The Plant Cell*, 16:1478–1489, 2004. doi: 10.1105/tpc.0211.

[10] J.L. Bowman, J. Alvarez, D. Weigel, E. M. Meyerowitz, and D.R. Smyth. Control od flower development in Arabidopsis thaliana by APETALA1 and interacting genes. *Development*, 3:721–743, 1993. ISSN 1477-9129. doi: 10.1016/j.anbehav.2009.06.0.

[11] Simone Bruno, M. Ali Al-Radhawi, Eduardo D. Sontag, and Domitilla Del Vecchio. Stochastic analysis of genetic feedback controllers to reprogram a pluripotency gene regulatory network. In *2019 American Control Conference (ACC)*, pages 5089–5096, 2019. doi: 10.23919/ACC.2019.8814355.

[12] M. E. Byrne. Phyllotatic pattern and stem cell fate are determined by the Arabidopsis homebox gene BELLRINGER. *Development*, 130:3941–3950, 2003. doi: 10.1242/dev.0062.

[13] Richard Casey, Hidde De Jong, and Jean Luc Gouzé. Piecewise-linear models of genetic regulatory networks: Equilibria and their stability. *Journal of Mathematical Biology*, 52(1):27–56, 2006. ISSN 03036812. doi: 10.1007/s00285-005-0338-2.

[14] José Eduardo Chairez. Análisis de propiedades de controlabilidad en redes de regulación genética del desarrollo de organismos vivientes (directores de tesis: Juan Carlos

Martínez García, Osbaldo Resendis Antonio (CCG-UNAM)). Master's thesis, Centro de Investigación y de Estudios Avanzados del IPN, Departamento de Control Automático, México, 2014.

[15] Tzu Min Chan, William Longabaugh, Hamid Bolouri, Hua Ling Chen, Wen Fang Tseng, Chung Hao Chao, Te Hsuan Jang, Yu I. Lin, Shao Chin Hung, Horng Dar Wang, and Chiou Hwa Yuh. Developmental gene regulatory networks in the zebrafish embryo. *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms*, 1789(4):279–298, 2009. ISSN 18749399. doi: 10.1016/j.bbagrm.2008.09.005. URL `http://dx.doi.org/10.1016/j.bbagrm.2008.09.005`.

[16] Chi-Tsong Cheng. *Linear System Theory and Design*. Oxford University Press, Inc., New York, third edit edition. ISBN 0195117778.

[17] Daizhan. Cheng, Qi Hongsheng., and Li Zhiqiang. *Analysis and Control of Boolean Networks*. Springer-Verlag London, first edit edition, 2011. ISBN 9783540707820. doi: 10.1007/978-1-4471-0207-6.

[18] L. Corbesier, C. Vincent, S. Jang, F. Fornara, Q. Fan, I. Searle, A. Giakountis, S. Farrona, L. Gissot, C. Turnbull, and G. Coupland. FT protein movement contributes to long-distance signaling in floral induction of Arabidopsis. *Science*, 316:1030–1033, 2007. doi: 10.1126/science.11417.

[19] José Davila-Velderrain, Carlos Villarreal, and Elena R. Alvarez-Buylla. Reshaping the epigenetic landscape during early flower development: Induction of attractor transitions by relative differences in gene decay rates. *BMC Systems Biology*, 9(1): 1–14, 2015. ISSN 17520509. doi: 10.1186/s12918-015-0166-y.

[20] D. Del Vecchio, H. Abdallah, Y. Qian, and J. Collins. A blueprint for a synthetic genetic feedback controller to reprogram cell fate. *Cell system*, 4:109–120, 2017. doi: 10.1016/j.cels.2016.12.001.

[21] Domitilla Del Vecchio, Hussein Abdallah, Yili Qian, and James J. Collins. A Blueprint for a Synthetic Genetic Feedback Controller to Reprogram Cell Fate. *Cell Systems*, 4(1):109–120, 2017. ISSN 24054720. doi: 10.1016/j.cels.2016.12.001. URL `http://dx.doi.org/10.1016/j.cels.2016.12.001`.

[22] Frank Emmert-Streib, Matthias Dehmer, and Benjamin Haibe-Kains. Gene regulatory networks and their applications understanding biological and medical problems in terms of networks. *Frontiers in Cell and Developmental Biology*, 38, 2014. doi: 10.3389/fcell.2014.00038.

[23] C Espinosa-Soto, P Padilla-Longoria, and Elena R. Alvarez-Buylla. A gene regulatory network model for cell-fate determination during Arabidopsis thalianal flower development that is robust and recovers experimental gene expression profiles. *Plant Cell*, 16(11):2923–2939, 2004. ISSN 1040-4651. doi: 10.1105/tpc.104.021725.2.

[24] C Espinosa-Soto, P Padilla-Longoria, and Elena R. Alvarez-Buylla. A gene regulatory network model for cell-fate determination during Arabidopsis thaliana flower development that is robust and recovers experimental gene expression profiles. *Plant Cell*, 16(11):2923–2939, 2004. ISSN 1040-4651. doi: 10.1105/tpc.104.021725.2.

[25] Adrien Fauré, Aurélien Naldi, Claudine Chaouiya, and Denis Thieffry. Dynamical analysis of a generic Boolean model for the control of the mammalian cell cycle. *Bioinformatics*, 22(14):124–131, 2006. ISSN 13674803. doi: 10.1093/bioinformatics/btl210.

[26] C. Ferrándiz, Q. Gu, R. Martienssen, and M. Yanofsky. Redundant regulation of meristem identity and plant architecture by FRUITFULL, APETALA1 and CAULIFLOWER. *Development*, 127:725–734, 2000. doi: 10.1046/j.1365-313x.1999.00442.

[27] Mónica L. García-Gómez, Eugenio Azpeitia, and Elena R. Álvarez-Buylla. A dynamic genetic-hormonal regulatory network model explains multiple cellular behaviors of the root apical meristem of Arabidopsis thaliana. 13(4):1–36, 2017. ISSN 15537358. doi: 10.1371/journal.pcbi.1005488.

[28] Alexander Gates and Luis Rocha. Control of complex networks requires both structure and dynamics. *Scientific Reports*, 6:1–11, 2016. doi: 0.1038/srep24456.

[29] Alexander J. Gates and Luis M. Rocha. Control of complex networks requires both structure and dynamics. *Scientific Reports*, 6(April):1–11, 2016. ISSN 20452322. doi: 10.1038/srep24456. URL `http://dx.doi.org/10.1038/srep24456`.

[30] Sourish Ghosh and Anirban Basu. Network medicine in drug design: Implications for neuroinflammation. *Drug Discovery Today*, 17(11-12):600–607, 2012. ISSN 13596446. doi: 10.1016/j.drudis.2012.01.018. URL `http://dx.doi.org/10.1016/j.drudis.2012.01.018`.

[31] Clare E. Giacomantonio and Geoffrey J. Goodhill. A boolean model of the gene regulatory network underlying mammalian cortical area development. *PLoS Computational Biology*, 6(9), 2010. ISSN 1553734X. doi: 10.1371/journal.pcbi.1000936.

[32] Aitor González, Claudine Chaouiya, and Denis Thieffry. Dynamical analysis of the regulatory network defining the dorsal-ventral boundary of the drosophila wing imaginal disc. *Genetics*, 174(3):1625–1634, 2006. ISSN 00166731. doi: 10.1534/genetics.106.061218.

[33] Gregor Gössler. Component-based modeling and reachability analysis of genetic networks. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 8 (3):672–682, 2011. ISSN 15455963. doi: 10.1109/TCBB.2010.81.

[34] Florian R. Greten and Sergei I. Grivennikov. Inflammation and cancer: Triggers, mechanisms, and consequences. *Immunity*, 51(1):27–41, 2019. ISSN 1074-7613. doi: https://doi.org/10.1016/j.immuni.2019.06.025. URL `https://www.sciencedirect.com/science/article/pii/S107476131930295X`.

[35] Sui Huang, Yan Ping Guo, Gillian May, and Tariq Enver. Bifurcation dynamics in lineage-commitment in bipotent progenitor cells. *Developmental Biology*, 305(2): 695–713, 2007. ISSN 00121606. doi: 10.1016/j.ydbio.2007.02.036.

[36] Soyeong Jun, Hyeonseob Lim, Honggu Chun, Ji Hyun Lee, and Duhee bang. Single-cell analysis of a mutant library generated using CRISPR-guided deaminase in human melanoma cells. *Commun Biol*, 3(154), 2020. doi: 10.1038/s42003-020-0888-2.

[37] Mitra Kabir, Nasimul Noman, and Hitoshi Iba. Reverse engineering gene regulatory network from microarray data using linear time-variant model. *BMC Bioinformatics*, 11(11):1–15, 2010. ISSN 14712105. doi: 10.1186/1471-2105-11-S1-S56.

[38] R. Kalman. Mathematical description of linear dynamical systems. *J.S.I.A.M Control*, 1(638):152–192, 1963.

[39] S. A. Kauffman. Metabolic stability and epigenesis in randomly constructed genetic nets. *Journal of Theoretical Biology*, 22(3):437–467, 1969. ISSN 10958541. doi: 10.1016/0022-5193(69)90015-0.

[40] Koichi Kobayashi. Design of fixed points in boolean networks using feedback vertex sets and model reduction. *Complexity*, 2019, 2019. doi: 10.1155/2019/9261793. URL `https://doi.org/10.1155/2019/9261793`.

[41] H Li and Y Wang. Minimum-Time State Feedback Stabilization of Constrained {Boolean} Control Networks. *Asian Journal of Control*, 18(5):1–10, 2015. ISSN 15618625. doi: 10.1002/asjc.1234.

[42] Yang Yu Liu, Jean Jacques Slotine, and Albert Lszl Barabási. Controllability of complex networks. *Nature*, 473(7346):167–173, 2011. ISSN 00280836. doi: 10.1038/nature10011.

[43] C. Luo, R. Shao, and Y. Zheng. Controllability of Boolean networks via input controls under Harveys update scheme. *Chaos*, 2016, 2016. doi: 10.1063/1.4941728.

[44] O. Mantegazza, V. Gregis, M. Chiara, C. Selva, G. Leo, D. Horner, and M. Kater. Gene coexpression patterns during early development of the native Arabidopsis reproductive meristem: novel candidate developmental regulators and patterns of functional redundancy. *The Plant Journal*, 79:861–877, 2014. doi: 10.1111/tpj.12585.

[45] Luis Fernando Méndez-López, José Davila-Velderrain, Elisa Domínguez-Hüttinger, Christian Enríquez-Olguín, Juan Carlos Martínez-García, and Elena R. Alvarez-Buylla. Gene regulatory network underlying the immortalization of epithelial cells. *BMC Systems Biology*, 11(1):1–15, 2017. ISSN 17520509. doi: 10.1186/s12918-017-0393-5.

[46] T. Nepusz and T. Vicsek. Controlling edge dynamics in complex networks. *Nature Physics*, 8:568–573, 2012. doi: 10.1038/NPHYS2327.

[47] Hasna Njah and Salma Jamoussi. Weighted ensemble learning of Bayesian network for gene regulatory networks. *Neurocomputing*, 150(PB):404–416, 2015. ISSN 18728286. doi: 10.1016/j.neucom.2014.05.078. URL `http://dx.doi.org/10.1016/j.neucom.2014.05.078`.

[48] J. Okamuro, B. Den Boer, C. Lotys-Prass, W. Szeto, and D. Jofuku. Flowers into shoots: Photo and hormonal control of a meristem identity switch in Arabidopsis. *Proc. Natl, Acad. Sci. USA*, 93:13831–13836, 1996. doi: 10.1073/pnas.93.24.13831.

[49] Elizabeth Ortiz-Gutiérrez, Karla García-Cruz, Eugenio Azpeitia, Aaron Castillo, María de la Paz Sánchez, and Elena R. Álvarez-Buylla. A Dynamic Gene Regulatory Network Model That Recovers the Cyclic Behavior of Arabidopsis thaliana Cell Cycle. *PLoS Computational Biology*, 11(9):1–28, 2015. ISSN 15537358. doi: 10.1371/journal.pcbi.1004486.

[50] F. Parcy, K. Bomblies, and D. Weigel. Interaction of LEAFY, AGAMOUS and TERMINAL FLOWER1 in maintaining floral meristem identity in Arabidopsis. *Development*, 129:2519–2527, 2002.

[51] Loïc Paulevé, Courtney Chancellor, Maxime Folschette, Morgan Magnin, and Olivier Roux. Analyzing Large Network Dynamics with Process Hitting. In Luis Fariñas del Cerro and Katsumi Inoue, editors, *Logical Modeling of Biological Systems*, pages 125 – 166. Wiley, July 2014. URL `https://hal.archives-ouvertes.fr/hal-01060490`.

[52] R.V. Perez-Ruiz, B. Garcá-Ponce, N. Marsh-Martínez, Y. Ugartechea-Chirino, M. Villajuana-Boqueni, S. de Folter, E. Azpeitia, J. Dávila-Velderrain, D. Cruz-Sánchez, A. Garay-Godoy, M. P. Sánchez, J.M. Estévez-Palmas, and E. R. Álvarez-Buylla. XAANTAL2 (AGL14) is an important componentof the complex gene regulatory network that underlies Arabidopsis shoot apical meristem transitions. *Molecular Plant*, 8:796–813, 2015. doi: 10.1016/j.molp.2015.01.017.

[53] A. Polynikis, S. J. Hogan, and M. di Bernardo. Comparing different ODE modelling approaches for gene regulatory networks. *Journal of Theoretical Biology*, 261(4):511–530, 2009. ISSN 00225193. doi: 10.1016/j.jtbi.2009.07.040. URL `http://dx.doi.org/10.1016/j.jtbi.2009.07.040`.

[54] Hana El Samad, Mustafa Khammash, Linda Petzold, and Dan Gillespie. Stochastic Modeling of Gene Regulatory Networks. *International Journal of Robust and Nonlinear Control*, (244725):1–6, 2005. ISSN 1049-8923. doi: 10.1002/rnc.1018.

[55] Geoffrey Schiebinger, Jian Shu, Marcin Tabaka, Brian Cleary, Vidya Subramanian, Aryeh Solomon, Joshua Gould, Siyan Liu, Stacie Lin, Peter Berube, Lia Lee, Jenny Chen, Justin Brumbaugh, Philippe Rigollet, Konrad Hochedlinger, Rudolf Jaenisch, Aviv Regev, and Eric S. Lander. Optimal-transport analysis of single-cell gene expression identifies developmental trajectories in reprogramming. *Cell*, 176(4):928–943.e22, 2019. ISSN 0092-8674. doi: https://doi.org/10.1016/j.cell.2019.01.006. URL `https://www.sciencedirect.com/science/article/pii/S009286741930039X`.

[56] Thomas Schlitt and Alvis Brazma. Current approaches to gene regulatory network modelling. *BMC Systems Biology*, 8:S9, 2007. ISSN 17520509. doi: 10.1186/1471-2105-8-S6-S9.

[57] Steven Nathaniel Steinway, Jorge Gomez Tejeda Zañudo, Paul J Michel, David J Feith, Thomas P Loughran, and Reka Albert. Combinatorial interventions inhibit TGFβ-driven epithelial-to-mesenchymal transition and support hybrid cellular phenotypes. *npj Systems Biology and Applications*, 1(1):15014, 2015. ISSN 2056-7189. doi: 10.1038/npjsba.2015.14. URL `http://www.nature.com/articles/npjsba201514`.

[58] Xiaodian Sun, Fang Hu, Shuang Wu, Xing Qiu, Patrice Linel, and Hulin Wu. Controllability and stability analysis of large transcriptomic dynamic systems for host response to influenza infection in human. *Infectious Disease Modelling*, 1(1):52–70, 2016. ISSN 24680427. doi: 10.1016/j.idm.2016.07.002. URL `http://linkinghub.elsevier.com/retrieve/pii/S2468042716300082`.

[59] E. P. Van Someren, L. F. A. Wessels, and M. J. T. Reinders. Linear modeling of genetic networks from experimental data. *Proc. Int. Conf Intell. Syst. Mol. Biol.*, 8 (2):355–366, 2000.

[60] Nedumparambathmarath Vijesh, Swarup Chakrabarti, and Janardanan Sreekumar. Modeling of gene regulatory networks: A review. *Journal of Biomedical Science and Engineering*, 22(6):223–231, 2013. doi: 10.4236/jbise.2013.62A027.

[61] Carlos Villarreal, Pablo Padilla-Longoria, and Elena R. Alvarez-Buylla. General theory of genotype to phenotype mapping: Derivation of epigenetic landscapes from n-node complex gene regulatory networks. *Physical Review Letters*, 109(11):1–5, 2012. ISSN 00319007. doi: 10.1103/PhysRevLett.109.118102.

[62] F. Wellmer, M. Alves-Ferreira, A. Dubois, J. Riechmann, and E. Meyerowitz. Genome-wide analysis of gene expression during early Arabidopsis flower development. *PLoS Genetics*, 2:1012–1024, 2006. doi: 10.1371/journal.pgen.00201.

[63] Chen weng, Jiajia Xi, Haiyan Li, Gu Anniya Cui, jian, Sisi Lai, Konstantin Leskov, Luxin Ke, Fulai jin, and Yan Li. Single-cell lineage analysis reveals extensive multimodal transcriptional control during directed beta-cell differentiation. *Nat Metab*, 2 (12):1443–1458, 2020. doi: 10.1038/s42255-020-00314-2.

[64] Shiro Yui, Luca Azzolin, Martti Maimets, Marianne Terndrup Pedersen, Robert P. Fordham, Stine L. Hansen, Hjalte L. Larsen, Jordi Guiu, Mariana R.P. Alves, Carsten F. Rundsten, Jens V. Johansen, Yuan Li, Chris D. Madsen, Tetsuya Nakamura, Mamoru Watanabe, Ole H. Nielsen, Pawel J. Schweiger, Stefano Piccolo, and Kim B. Jensen. Yap/taz-dependent reprogramming of colonic epithelium links ecm remodeling to tissue regeneration. *Cell Stem Cell*, 22(1):35–49.e7, 2018. ISSN 1934-5909. doi: https://doi.org/10.1016/j.stem.2017.11.001. URL `https://www.sciencedirect.com/science/article/pii/S1934590917304538`.

[65] Jorge Gómez Tejeda Zañudo, M Tyler Guinn, Kevin Farquhar, Mariola Szenk, Steven N Steinway, Gábor Balázsi, and Réka Albert. Towards control of cellular decision-making networks in the epithelial-to-mesenchymal transition. *Physical Biology*, 16(3):031002, mar 2019. doi: 10.1088/1478-3975/aaffa1. URL `https://doi.org/10.1088/1478-3975/aaffa1`.

[66] Y. Zhao, H. Qi, and Cheng D. Input-State incidence matrix of Boolean control networks and its applications. *Systems Control Letters*, 59:767–774, 2010. doi: 10.1016/j.syconle.2010.09.002.

[67] Kemin Zhou, John Comstock Doyle, Keith Glover, et al. *Robust and optimal control*, volume 40. Prentice hall New Jersey, 1996.

Index