ALMA MATER STUDIORUM · UNIVERSITY OF BOLOGNA

School of Science
Department of Physics and Astronomy
Master Degree in Physics

Dynamical systems on graphs

Supervisor:
Prof. Armando Bazzani

Submitted by:
Riccardo Scheda

Academic Year 2019/2020

Sommario

La cellula vivente è un sistema complesso governato da molti processi che non sono ancora stati compresi: il processo di differenziazione cellulare è uno di questi. La differenziazione cellulare è il processo in cui le cellule di un tipo specifico si riproducono e danno origine a diversi tipi di cellule. La differenziazione cellulare è regolata dai cosiddetti Gene Regulatory Networks (GRN). Un GRN è una raccolta di regolatori molecolari che interagiscono tra loro e con altre sostanze nella cellula per governare i livelli di espressione genica di mRNA e proteine. Kauffman propose per la prima volta nel 1969 di modellare GRN attraverso le cosiddette Random Boolean Networks (RBN). I RBN sono reti in cui ogni nodo può avere solo due possibili valori: 0 o 1, dove ogni nodo rappresenta un gene in GRN che può essere "on" oppure "off". Queste reti possono modellare GRN perché l'attività di un nodo rappresenta il livello di espressione di un gene nell'intera regolazione.

In questo lavoro di tesi ci avvaliamo di un modello matematico per sviluppare e riprodurre una possibile rete di regolazione genica per i linfociti T del sistema immunitario, che presentano una natura bistabile.

Abstract

Real living cell is a complex system governed by many process which are not yet understood: the process of cell differentiation is one of these. Cell differentiation is the process in which cells of a specific type reproduces themselves and give arise to different type of cells. Cell differentiation is governed by the so called Gene Regulatory Networks (GRNs). A GRN is a collection of molecular regulators that interact with each other and with other substances in the cell to govern the gene expression levels of mRNA and proteins. Kauffman proposed for the first time in 1969 to model GRN through the so called Random Boolean Networks (RBN). RBNs are networks in which each node can have only two possible values: 0 or 1, where each node represent a gene in GRN which can be "on" or "off". These networks can model GRNs because the activity of one node represents the expression level of one gene among the whole regulation.

In this thesis work we make use of a mathematical model to develop and reproduce a possible Gene Regulatory Network for the T-cells of the immune system, which present a bistable nature.

Contents

Introduction **8**

1 Cell Differentiation **10**
1.1 Definition 10
1.2 The role of Gene expression in cellular differentiation 12
1.3 Cell differentiation in Immune system 13

2 Immunosystem and T-Cells **15**
2.1 T-Cells 15
2.2 Life cycle of T lymphocytes 17
2.3 Activation of T lymphocytes 18
2.4 Immunotherapy 19

3 Gene Regulatory Networks **21**
3.1 Definition 21
3.2 Modelling GRNs 23
3.3 Principle of the models 25

4 Random Boolean Networks **26**
4.1 Random Boolean Networks 26
4.2 The model 27
4.3 Topology 27
4.4 Dynamics 28
4.5 Phase transitions 30

4.6 Perturbations 32

5 Kramer Rate Theory **34**

5.1 Fokker-Planck equation 34

5.2 Bistability and Escape Problems 35

5.2.1 Diffusion in a Double-Well Potential in one dimension 35

5.2.2 Behaviour for $D = 0$ 35

5.2.3 Behaviour if D is small 36

5.3 Kramer Rate Theory 38

5.4 First passage time and transition rate escaping from a barrier 40

6 The model **44**

6.1 The underlying philosophy of the model 44

6.2 The mathematical model and related problems 45

7 Analysis **57**

7.1 Implementation 57

7.2 Discrete evolution 58

7.3 Loops and Control nodes 59

7.4 Noise 59

Bibliography **63**

List of Figures

1.1 *Schematic representation of cellular differentiation.* 11

3.1 *Schematic representation of a Gene Regulatory Network.* 22

4.1 Set of all possible networks with $N = 2$ and $K = 1$. 28
4.2 *A small network with* $N = 4$ *and* $K = 1$. 29
4.3 *The state space of the network shown in Figure 4.2, if the functions copy, copy, invert, invert are assigned to the four nodes. The numbers in the squares represent states, and arrows indicate the successor of each state. States on attractors are shaded.* 30
4.4 *Phase diagram for the* $N - K$ *model. The shaded area corresponds to the chaotic phase, whereas the white region corresponds to the chaotic phase. The curve separating both regions is the critical phase.* 32
4.5 *Jump from one attractor to one other in the state space.* 33

5.1 *Example of double-well potential.* 36
5.2 *The distribution (dashed) explains the bistability of the system, since, we have two local maxima of the potential in a and c..* 37
5.3 *Example of double-well potential.* 38

6.1 *Example of random boolean network.* 48
6.2 *Possible behavior for the condition (6.4); the units are arbitrary and scale with the network dimension.* 50
6.3 *Example of a network composed by two competitive subnetworks.* 53

7.1 *Plot of the average activity of the nodes with network of increasing size. In the case of $K = 1$ (i.e. the average number of incoming link for each network is one), the average activity decreases exponentially with the size of the network; in the case of $K = 2$ instead, the average activity of the nodes remains stable with the network size.* 58
7.2 *Plot of the number of the outgoing links depending on the network size. We can see that the average tends to the parameter K.* 59
7.3 *Plot of the number of independet loops in the networks depending on the network size.* 60
7.4 *Plot of the effect of the noise on the average activity on the network. In blue the noise works on the nodes on the network, while in red the noise works on the links. Number of nodes for each network: 10; Number of realizations for each value of noise: 100;* 61

Introduction

The cell is a paradigmatic example of complex system governed by many processes which are not yet understood: the process of cell differentiation is one of these. Cell differentiation is the dynamical process in which stem cells reproduce and give arise to different type of cells. Waddington in 1957 [1] proposed to model cell differentiation with an epigenetic landscape in which lay different type of cells. This epigenetic landscape can be seen as a potential in a physical system in which different type of cells are attracted by the different wells of this potential. From this model, recent studies on omics data propose ways to find epigenetic landscape for different cells[2][3][4], using stochastic processes.

Now, it is well known that cell differentiation is governed by the so called Gene Regulatory Networks (GRNs). A GRN is a collection of molecular regulators that interact with each others and with other substances in the cell to define the gene expression levels of mRNA and proteins.

Disruption of these processes by inappropriate regulatory signals and by mutational rewiring of the network can lead to tumorigenesis[5].

Kauffman proposed for the first time in 1969[6] to model GRNs through the so called Random Boolean Networks (RBN). RBN are networks in which each node can have only two possible values: 0 or 1, where each node represents a gene in GRN which can be "on" or "off"[7]. The evolution of the state of the network is given by some boolean functions, depending on the connectivity of the nodes. So each node will have one boolean function which defines the next state during the descrete evolution. From this networks, one can find some periodical structures called *attractors*, which can be associated to different type of cells by a biological point of view[9][8][10].

Omics data available are subject to noise, and we can't expect to see directly connections between different genes, so we want to build a model able to see some statistical properties among these networks.

In Chapters 1 and 3 we make a biological introduction of Cell Differentiation and of Gene Regulatory Network. In Chapter 4 we present the concept of Random Boolean Networks, a model proposed for the first time by Kauffman to model Gene Regulatory Networks. In Chapter 6 we propose our model based on Random Boolean Networks but with some differences, where networks can have also hinibitory links betwee clusters, showing a bifurcation phenomena among the activity of the network. In Chapter 7 we make a numerical analysis of the thoretical model proposed in 6.

Chapter 1

Cell Differentiation

In this Chapter we explain Cell Differention process and Gene Regulatory Networks.

1.1 Definition

Cell differentiation is the process whereby stem cells become progressively more specialized. The differentiation process occurs both during the development of a multicellular organism and during tissue repair and cell turnover in the adulthood. Gene expression, and therefore its regulatory mechanisms, plays a critical role in cell differentiation. Stem cells are undifferentiated biological cells which can both reproduce themselves, self-renewal ability, and differentiate into specialized cells, potency. The principles underlying cellular differentiation remain among the most enigmatic in biology. We are required to explain the spontaneous generation of a multiplicity of cell types from the single zygote, to deduce a natural tendency of a system to become increasingly heterogeneous, then to stop differentiating.

Among the important characteristics of cell differentiation are:

- initiation of change;
- stabilization of change after cessation of stimulus;

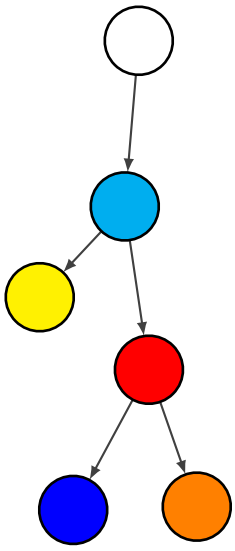

Figure 1.1: *Schematic representation of cellular differentiation.*

- the efficacy of many substances, exogenous and endogenous, as inductive stimuli;
- progressive limitation in the number of developmental pathways open to any small region of the embryo;
- restricted periods during which a cell is competent to respond to an inductive stimulus; the discreteness of cell types, that is, the mutually exclusive constellations of properties by which cells differ;
- a requirement for a minimal and preferably heterogeneous cell mass to initiate differentiation in many instances, and to maintain it in some;
- the occurrence of metaplasia between undifferentiated cell types, or from an undifferentiated type to a specialized type, but the lack of metaplasia (the isolation) between specialized cell types;
- cessation of differentiation.

Cells are thought to differ due to differential expression of, rather than structural loss of, the genes. Differential activity of the genes raises at least two questions which are not always carefully distinguished: the capacity of the genome to behave in more than one mode; and mechanisms which insure the appropriate assignment of these modes to the proper cells.

Within multicellular organisms, tissues are organized in communities of cells that work together to carry out a specific function. The exact role of a tissue in an organism depends on what types of cells it contains. For example, the endothelial tissue that lines the human gastrointestinal tract consists of several cell types. Some of these cells absorb nutrients from the digestive contents, whereas others secrete a lubricating mucus that helps the contents travel smoothly. However, the multiple cell types within a tissue don't just have different functions. They also have different transcriptional programs and may well divide at different rates. Proper regulation of these rates is essential to tissue maintenance and repair. Stem cells typically have the capacity to mature into many different cell types. *Transcription factors* (TF), which are proteins that regulate which genes are transcribed in a cell, appear to be essential to determining the pathway particular stem cells take as they differentiate. For example, both intestinal absorptive cells and goblet cells arise from the same stem cell population, but divergent transcriptional programs cause them to mature into dramatically different cells. Whenever stem cells are called upon to generate a particular type of cell, they undergo an asymmetric cell division. With asymmetric division, each of the two resulting daughter cells has its own unique life course. In this case, one of the daughter cells has a finite capacity for cell division and begins to differentiate, whereas the other daughter cell remains a stem cell with unlimited proliferative ability.

1.2 The role of Gene expression in cellular differentiation

Gene expression is a complex process regulated at several stages in the synthesis of proteins. In addition to the DNA transcription regulation, the expression of a gene may be controlled during RNA processing and transport (in eukaryotes), RNA trans-

lation, and the post-translational modification of proteins. This gives rise to genetic regulatory systems structured by networks of regulatory interactions between DNA, RNA, proteins and other molecules: a complex network termed as a *gene regulatory network* (GRN) (see Chapter 3). Some kind of proteins are the transcription factors that bind to specific DNA sequences in order to regulate the expression of a given gene. The power of transcription factors resides in their ability to activate and/or repress transcription of genes. The activation of a gene is also referred to positive regulation, while the negative regulation identifies the inhibition of the gene. The regulation of gene expression is essential for the cell, because it allows to control the internal and external functions of the cell. Furthermore, in multicellular organisms, gene regulation drives the processes of cellular differentiation and morphogenesis, leading to the creation of different cell types that possess different gene expression profiles, and these last therefore produce different proteins that have different ultrastructures that suit them to their functions. Therefore, with few exceptions, all cells in an organism contain the same genetic material, and hence the same genome. The difference between the cells are emergent and due to regulatory mechanisms which can turn on or off genes. Two cells are different, if they have different subsets of active genes.

1.3 Cell differentiation in Immune system

Cell differentiation concerns all type of cells, and in particular one of the most important systems in the organisms: the Immune system. Immune system is a host defense system comprising many biological structures and processes within a organims that protects against disease. To function properly, an immune system must detect a wide variety of agents, known as pathogens, from viruses to parasitic worms, and distinguish them from the organism's own healthy tissue. In many species, there are two major subsystems of the immune system: the innate immune system and the adaptive immune system.

Cellular differentiation concerns with immune system beacuse it is composed by a wide range of different cells: *B lymphocytes, T lymphocytes, Basophil, Eosinophil,* etc. In the next Chapter we will concentrate on the so called *T cells*, which are important for

the immune system and in particular because they are the focus of our mathematical model.

Chapter 2

Immunosystem and T-Cells

In this Chapter we briefly explain what T-cells are and why they are important for immune system.

2.1 T-Cells

T-cell, also called *T lymphocyte*, is a type of leukocyte (white blood cell) that is an essential part of the immune system. T-cells are one of two primary types of lymphocytes (B cells being the second type) that determine the specificity of immune response to antigens (foreign substances) in the body [11]. T-cells originate in the bone marrow and mature in the thymus. In the thymus, T-cells multiply and differentiate into different type of cells: *T helper*, *regulatory T-cells* and *cytotoxic T-cells*; further, during their life they can become *memory T-cells*. They are then sent to peripheral tissues or circulate in the blood or lymphatic system. In short words, their main roles are:

- **T helper**: once stimulated by the appropriate antigen, helper T-cells secrete chemical messengers called *cytokines*, which stimulate the differentiation of B cells into plasma cells (antibody-producing cells).
- **Regulatory T-cells**: act to control immune reactions, hence their name.
- **Cytotoxic T-cells**: they are activated by various cytokines, bind to and kill infected cells and cancer cells.

Because the body contains millions of T and B cells, many of which carry unique receptors, it can respond to virtually any antigen.

Despite the structural similarities, the receptors on T-cells function differently from those on B cells. The functional difference underlies the different roles played by B and T-cells in the immune system. B cells secrete antibodies to antigens in blood and other body fluids, but T-cells cannot bind to free-floating antigens. Instead they bind to fragments of foreign proteins that are displayed on the surface of body cells. Thus, once a virus succeeds in infecting a cell, it is removed from the reach of circulating antibodies only to become susceptible to the defense system of the T-cell.

Some T-cells recognize class I MHC molecules on the surface of cells; others bind to class II molecules. Cytotoxic T-cells destroy body cells that pose a threat to the individual—namely, cancer cells and cells containing harmful microorganisms. Helper T-cells do not directly kill other cells but instead help activate other white blood cells (lymphocytes and macrophages), primarily by secreting a variety of cytokines that mediate changes in other cells. The function of regulatory T-cells is poorly understood. To carry out their roles, helper T-cells recognize foreign antigens in association with class II MHC molecules on the surfaces of macrophages or B cells. Cytotoxic T-cells and regulatory T-cells generally recognize targeT-cells bearing antigens associated with class I molecules. Because they recognize the same class of MHC molecule, cytotoxic and regulatory T-cells are often grouped together; however, populations of both types of cells associated with class II molecules have been reported. Cytotoxic T-cells can bind to virtually any cell in the body that has been invaded by a pathogen.

T-cells have another receptor, or coreceptor, on their surface that binds to the MHC molecule and provides additional strength to the bond between the T-cell and the targeT-cell. Helper T-cells display a coreceptor called CD4, which binds to class II MHC molecules, and cytotoxic T-cells have on their surfaces the coreceptor CD8, which recognizes class I MHC molecules. These accessory receptors add strength to the bond between the T-cell and the targeT-cell. The T-cell receptor is associated with a group of molecules called the CD3 complex, or simply CD3, which is also necessary for T-cell activation. These molecules are agents that help transduce, or convert, the extracellular binding of the antigen and receptor into internal cellular signals; thus,

they are called signal transducers. Similar signal transducing molecules are associated with B-cell receptors.

2.2 Life cycle of T lymphocytes

When T-cell precursors leave the bone marrow on their way to mature in the thymus, they do not yet express receptors for antigens and thus are indifferent to stimulation by them. Within the thymus the T-cells multiply many times as they pass through a meshwork of thymus cells. In the course of multiplication they acquire antigen receptors and differentiate into helper or cytotoxic T-cells. As mentioned in the previous section, these cell types, similar in appearance, can be distinguished by their function and by the presence of the special surface proteins, CD4 and CD8. Most T-cells that multiply in the thymus also die there. This seems wasteful until it is remembered that the random generation of different antigen receptors yields a large proportion of receptors that recognize self antigens (i.e. molecules present on the body's own constituents) and that mature lymphocytes with such receptors would attack the body's own tissues.

Most such self-reactive T-cells die before they leave the thymus, so that those T-cells that do emerge are the ones capable of recognizing foreign antigens. These travel via the blood to the lymphoid tissues, where, if suitably stimulated, they can again multiply and take part in immune reactions. The generation of T-cells in the thymus is an ongoing process in young animals. In humans large numbers of T-cells are produced before birth, but production gradually slows down during adulthood and is much diminished in old age, by which time the thymus has become small and partly atrophied. Cell-mediated immunity persists throughout life, however, because some of the T-cells that have emerged from the thymus continue to divide and function for a very long time.

2.3 Activation of T lymphocytes

Helper T-cells do not directly kill infected cells, as cytotoxic T-cells do. Instead they help activate cytotoxic T-cells and macrophages to attack infected cells, or they stimulate B cells to secrete antibodies. Helper T-cells become activated by interacting with antigen-presenting cells, such as macrophages. Antigen-presenting cells ingest a microbe, partially degrade it, and export fragments of the microbe (i.e. antigens) to the cell surface, where they are presented in association with class II MHC molecules. A receptor on the surface of the helper T-cell then binds to the MHC-antigen complex. But this event alone does not activate the helper T-cell. Another signal is required, and it is provided in one of two ways: either through stimulation by a cytokine or through a costimulatory reaction between the signaling protein, B7, found on the surface of the antigen-presenting cell, and the receptor protein, CD28, on the surface of the helper T-cell. If the first signal and one of the second signals are received, the helper T-cell becomes activated to proliferate and to stimulate the appropriate immune cell. If only the first signal is received, the T-cell may be rendered anergic, that is, unable to respond to antigen.

Once the initial steps of activation have occurred, helper T-cells synthesize other proteins, such as signaling proteins and the cell-surface receptors to which the signaling proteins bind. These signaling molecules play a critical role not only in activating the particular helper T-cell but also in determining the ultimate functional role and final differentiation state of that T-cell. For example, the helper T-cell produces and displays IL-2 receptors on its surface and also secretes IL-2 molecules, which bind to these receptors and stimulate the helper T-cell to grow and divide. The overall result of helper-T-cell activation is an increase in the number of helper T-cells that recognize a specific foreign antigen, and several T-cell cytokines are produced. The cytokines have other consequences, one of which is that IL-2 allows cytotoxic or regulatory T-cells that recognize the same antigen to become activated and to multiply. Cytotoxic T-cells, in turn, can attack and kill other cells that express the foreign antigen in association with class I MHC molecules, which (as explained above) are present on almost all cells. So, for example, cytotoxic T-cells can attack target T-cells that express antigens made by viruses or bacteria growing within them. Regulatory T-cells may be

similar to cytotoxic T-cells, but they are detected by their ability to suppress the action of B cells or even of helper T-cells (perhaps by killing them). Regulatory T-cells thus act to damp down the immune response and can sometimes predominate so as to suppress it completely.

2.4 Immunotherapy

Early attempts to harness the immune system to fight cancer involved tumour-associated antigens, proteins that are present on the outer surface of tumour cells. Antigens are recognized as "foreign" by circulating immune cells and thereby trigger an immune response. However, many tumour antigens are altered forms of proteins found naturally on the surface of normal cells; in addition, those antigens are not specific to a certain type of tumour but are seen in a variety of cancers. Despite the lack of tumour specificity, some tumour-associated antigens can serve as targets of attack by components of the immune system. For instance, antibodies can be produced that recognize a specific tumour antigen, and those antibodies can be linked to a variety of compounds (such as chemotherapeutic drugs and radioactive isotopes) that damage cancer cells. In this way the antibody delivers the therapeutic agent directly to the tumour cell. In other cases a chemotherapeutic agent attached to an antibody destroys cancer cells by interacting with receptors on their surfaces that trigger apoptosis.

T-cells themselves may be engineered to recognize, bind to, and kill cancer cells. For example, in an experimental treatment for chronic lymphocytic leukemia, researchers designed a virus to induce the expression on patient T-cells of antibody receptors that identified and attached to antigens on malignant B cells and that activated the T-cells, prompting them to destroy the B cells. T-cells removed from patient blood were incubated with the virus and following infection were infused back into the patient. A portion of the engineered cells persisted as memory T-cells, retaining functionality and suggesting that the cells possessed long-term activity against cancer cells.

A similar T-cell therapy, known as *chimeric antigen receptor T-cells* (CAR-T), in which T-cells isolated from a patient's blood are genetically engineered to specifically iden-

tify and target cancer cells and then are infused back into the patient, has been used in the treatment of certain forms of leukemia, including acute lymphocytic leukemia, as well as B-cell lymphoma. The addition, via genetic engineering, of a unique receptor to the T-cell surface that is capable of recognizing a molecule known as MR1, found on cells from a variety of different cancer types, has opened the possibility of expanding CAR-T to the treatment of solid tumours, in addition to cancers of the blood.

Chapter 3

Gene Regulatory Networks

3.1 Definition

Gene regulation controls the expression of genes and, consequently, all cellular functions. Gene expression is a process that involves transcription of the gene into mRNA, followed by translation to a protein, which may be subject to post-translational modification [12]. The transcription process is controlled by transcription factors (TFs) that can work as activators or inhibitors. TFs are themselves encoded by genes and subject to regulation, which altogether forms complex regulatory networks. Cells efficiently carry out molecular synthesis, energy transduction, and signal processing across a range of environmental conditions by networks of genes, which we define broadly as networks of interacting genes, proteins, and metabolites [13]. Formally speaking, a *gene regulatory network* or *genetic regulatory network* (GRN) is a collection of DNA segments in a cell which interact with each other (indirectly through their RNA and protein expression products) and with other substances in the cell, thereby governing the rates at which genes in the network are transcribed into mRNA. In general, each mRNA molecule goes on to make a specific protein (or set of proteins). In some cases this protein will be structural, and will accumulate at the cell-wall or within the cell to give it particular structural properties.

These networks control biological process of all organisms. The complex control systems underlying development have probably been evolving for more than a billion

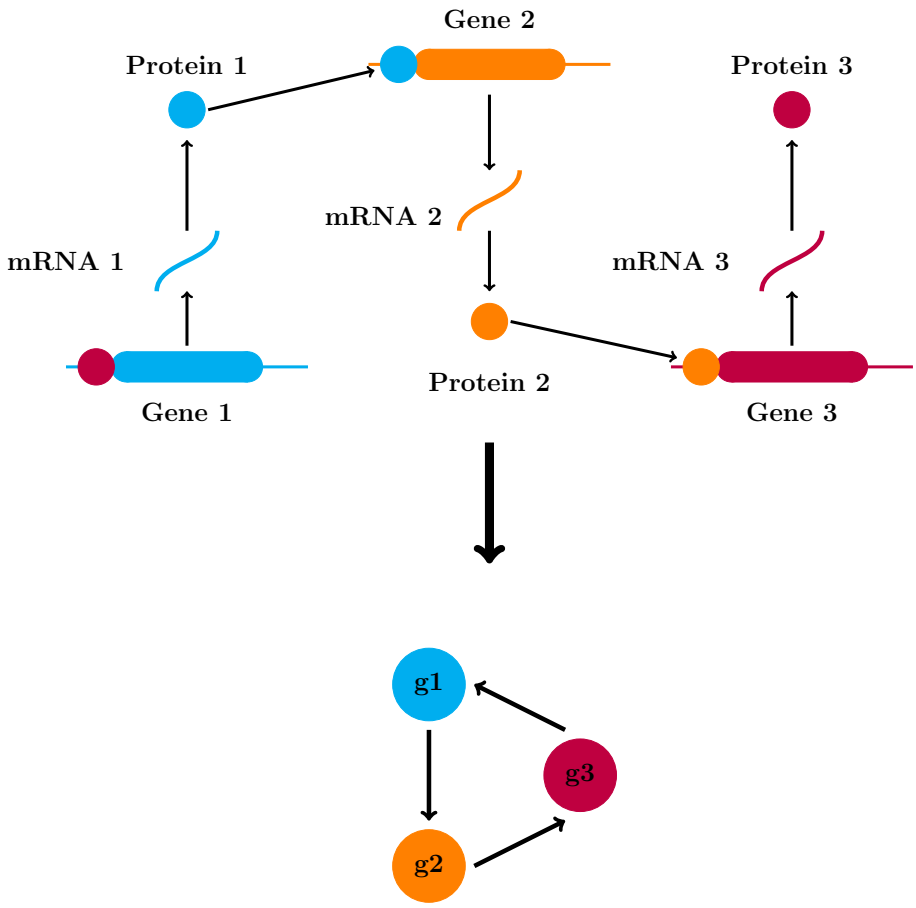

Figure 3.1: *Schematic representation of a Gene Regulatory Network.*

years. They regulate the expression of thousand of genes in any given biological process. They are essentially hardwired genomic regulatory codes, the role of which is to specify the sets of genes that must be expressed in specific spatial and temporal patterns. In physical terms, these control system consist of many thousands of modular DNA sequences. Each module receives and integrates multiple inputs, in the form of regulatory proteins (*activators* and *repressors*) that recognize specific sequences within them. The end result is the precise transcriptional control of the associated genes. Functional linkages between these particular genes, and their associated regulatory modules, define the core networks underlying development. They explain exactly how genomic sequence encodes the regulation of expression of the sets of genes that generate patterns and execute the construction of multiple states of differentiation.

The regulatory genome is a logic processing system: every regulatory module contained in the genome receives multiple inputs and processes in ways that can be mathematically represented as combinations of logic functions.

Definitive regulatory functions emerge only from the architecture of intergenic linkages, and these functions are not visible at the level of any individual genes. So gene regulatory networks can be determined only by experimental molecular biology in which the functional meaning of given regulatory sequences is directly determined.

GRNs have a complex structure: they are inhomogeneous compositions of different kinds of subnetworks, each performing a specific kind of function. Some subnetworks are used in many processes.

In principle, mathematical modeling of GRN dynamics can provide a theoretical foundation for understanding cell heterogeneity and gene expression dynamics, by quantitatively linking molecular-level regulatory mechanisms with observed cell states. However, due to the molecular complexity of gene regulatory mechanisms, it remains challenging to integrate such models with single-cell data.

3.2 Modelling GRNs

Mathematical models can account for (and at least partially reproduce) observed cellular heterogeneity in two primary ways. First, gene network models are multi-

stable dynamical systems, meaning a given network has the potential to reach multiple stable states of gene expression. These states arise from the dynamic interplay of activation, inhibition, feedback, and nonlinearity [6] [22]. Second, some mathematical models inherently treat cellular noise. This noise, or stochasticity, is modeled in various ways depending on assumptions about the source [14] [15]. Discrete, stochastic models of gene regulation, which track discrete molecular entities, regulatory-protein binding kinetics, and binding states of promoters controlling gene activity, have formed the basis of biophysical theories of gene expression noise due to so-called intrinsic molecular noise [15] [16]. Such stochastic gene regulation mechanisms have also been incorporated into larger regulatory network models using the formalism of stochastic biochemical reaction networks, and have been utilized to explore how molecular fluctuations can cause heterogeneity within phenotype-states and promote stochastic transitions between phenotypes [17] [18].

The quantitative landscape of cellular states is another concept that is increasingly utilized to describe cellular heterogeneity. Broadly, the cellular potential landscape (first conceptualized by Waddington [4][1]) is a function in high-dimensional space (over many molecular observables, typically expression levels of different genes), that quantifies the stability of a given cell state. In analogy to potential energy (gravitational, chemical, electric, etc.), cell states of higher potential are less stable than those of lower potential. The landscape concept inherently accounts for cellular heterogeneity, since it holds that a continuum of states is theoretically accessible to the cell, with low-potential states (in "valleys") more likely to be observed than high-potential states. The landscape is a rigorously defined function derived from the dynamics of the underlying gene network model, according to some choice of mathematical formalism [19][4].

Stochastic modeling of gene network dynamics has been employed in various forms for analysis of single cell measurements. However, few existing analysis methods utilize discrete-molecule, stochastic models, which fully account for intrinsic gene expression noise and its impact on cell-state, to aid in the interpretation of noisy distributions recovered from single cell RNA sequencing data [2][3]. There exists an opportunity to link such biophysical, stochastic models, which reproduce intrinsic noise

and cell heterogeneity in silico, to single cell datasets that characterize cell heterogeneity in vivo. In particular, the landscape of heterogeneous cell states computed from discrete stochastic models can be directly compared to single-cell measurements.

3.3 Principle of the models

GRNs may be interpreted as an idealized dynamical system of model genes with directional links (transcription factors), updating their state in parallel, according to the combinatorial logic of their inputs, Kauffman's Random Boolean Networks [6][23]. There is justified debate as to whether parallel (synchronous) updating, and the on-off characterization of genes, are valid idealizations when applied to real genomic networks, given that transcription is asynchronous and driven at different rates. However, gene activity at the molecular scale consists of discrete events occurring concurrently. Variable protein concentrations can be accounted for by genes being on for some fraction of a given time span. The RBN idealization is arguably a valid starting point for gaining insights into gene network dynamics. In a cell type's gene expression pattern over a span of time (i.e. its space- time pattern), a particular gene may, broadly speaking, be either on, off, or changing. If a large proportion of the genes are changing, chaotic dynamics, the cell will be unstable. On the other hand, dynamics that settles to a pattern where a large proportion of the genes are permanently on or o (frozen) may be too in exible for adaptive behavior. Cells constantly need to adapt their gene expression pattern in response to a variety of hormone and growth/di erentiation factors from nearby cells. The definition of a cell type may be more correctly expressed as a set of closely related gene expression patterns, allowing an essential measure of exibility in behavior.

In the next Chapter we will concentrate on the Random Boolean Networks proposed by Kauffmann in 1969.

Chapter 4

Random Boolean Networks

In this chapter we explain the basic concepts of Random Boolean Network proposed for the first time by Kauffman.

4.1 Random Boolean Networks

Random Boolean networks (RBNs) were introduced in 1969 by S. Kauffman as a simple model of genetic systems [6]. Each gene was represented by a node that has two possible states, "on" (corresponding to a gene that is being transcribed) and "off" (corresponding to a gene that is not being transcribed). There are altogether N nodes, and each node receives input from K randomly chosen nodes, which represent the genes that control the considered gene. Furthermore, each node is assigned an *update boolean function* that prescribes the state of the node in the next time step, given the state of its input nodes. This update function is chosen from the set of all possible update functions according to some probability distribution. Starting from some initial configuration, the states of all nodes of the network are updated in parallel. Since configuration space is finite and since dynamics is deterministic, the system must eventually return to a configuration that it has had before, and from then on it repeats the same sequence of configurations periodically.

4.2 The model

Let's consider a network of N nodes. The state of each node at a time t is given by $\sigma_i(t) \in \{0,1\}$ with $i = 1,\ldots,N$. The N nodes of the network can therefore together assume 2^N different states. The number of incoming links to each node i is denoted by k_i and is drawn randomly independently from the distribution $P(k_i)$. The dynamical state of each $\sigma_i(t)$ is updated synchronously by a Boolean function Λ_i:

$$\Lambda_i : \{0,1\}^{k_i} \to \{0,1\}$$

An update function specifies the state of a node in the next time step, given the state of its K inputs at the present time step. Since each of the K inputs of a node can be on or off, there are $M = 2^K$ possible input states. The update function has to specify the new state of a node for each of these input states. Consequently, there are 2^M different update functions. For example let's consider a network with $K = 1$, so all the functions Λ_i receives the input from one single node. In general each element receives inputs from exactly K nodes, so we have a dynamical system defined from:

$$\sigma_i(t+1) = \Lambda_i(\sigma_{i_1}(t), \sigma_{i_2}(t), ..., \sigma_{i_K}(t)). \tag{4.1}$$

So, the randomness of these network appears at two levels: in the connectivity of the network (which node is linked to which) and the dynamics (which function is attributed to which node).

4.3 Topology

For a given number N of nodes and a given number K of inputs per node, a RBN is constructed by choosing the K inputs of each node at random among all nodes. If we construct a sufficiently large number of networks in this way, we generate an ensemble of networks. In this ensemble, all possible topologies occur, but their statistical weights are usually different. Let us consider the simplest possible example, $N = 2$ and $K = 1$, shown in Figure 4.1. There are 3 possible topologies.

Topologies (a) and (b) have each the statistical weight $1/4$ in the ensemble, since each of the links is connected in the given way with probability $1/2$. Topology (c) has

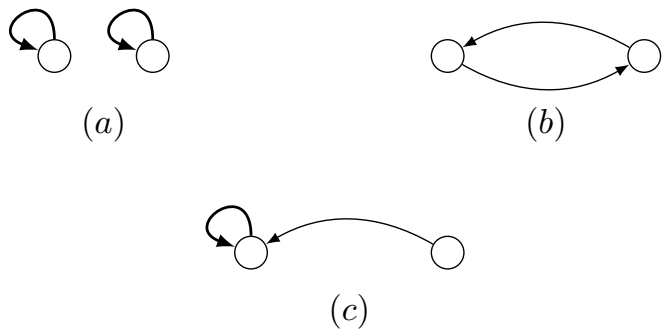

Figure 4.1: Set of all possible networks with $N = 2$ and $K = 1$.

the weight 1/2, since there are two possibilities for realizing this topology: either of the two nodes can be the one with the self-link.

While the number of inputs of each node is fixed by the parameter K, the number of outputs (i.e. of outgoing links) varies between the nodes. The mean number of outputs must be K, since there must be in total the same number of outputs as inputs. A given node becomes the input of each of the N nodes with probability $\frac{K}{N}$. In the thermodynamic limit $N \to \infty$ the probability distribution of the number of outputs is therefore a Poisson distribution:

$$P_{out}(k) = \frac{K^k}{k!} e^{-K}$$

4.4 Dynamics

All nodes are updated at the same time according to the state of their inputs and to their update function. Starting from some initial state, the network performs a trajectory in state space and eventually arrives on an *attractor*, where the same sequence of states is periodically repeated. Since the update rule is deterministic, the same state must always be followed by the same next state. If we represent the network states

by points in the 2^N-dimensional state space, each of these points has exactly one "output", which is the successor state. We thus obtain a graph in state space. The size or length of an attractor is the number of different states on the attractor. The basin of attraction of an attractor is the set of all states that eventually end up on this attractor, including the attractor states themselves. The size of the basin of attraction is the number of states belonging to it. The graph of states in state space consists of unconnected components, each of them being a basin of attraction and containing an attractor, which is a loop in state space. The transient states are those that do not lie on an attractor. They are on trees leading to the attractors.

Let us illustrate these concepts by studying the small $K = 1$ network shown in Figure 4.2, which consists of 4 nodes:

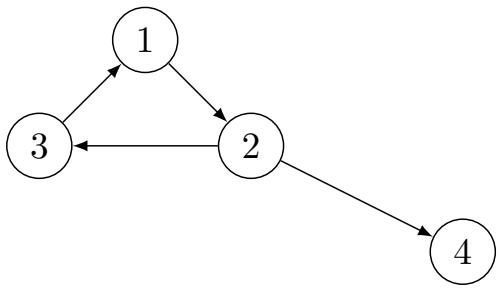

Figure 4.2: *A small network with* $N = 4$ *and* $K = 1$.

If we assign to the nodes 1,2,3,4 the functions invert, invert, copy, copy, an initial state 1111 evolves in the following way:

$$1111 \rightarrow 0011 \rightarrow 0100 \rightarrow 1111$$

This is an attractor of period 3. If we interpret the bit se- quence characterizing the state of the network as a number in binary notation, the sequence of states can also be written as

$$15 \rightarrow 3 \rightarrow 4 \rightarrow 15$$

The entire state space is shown in Figure 4.3:

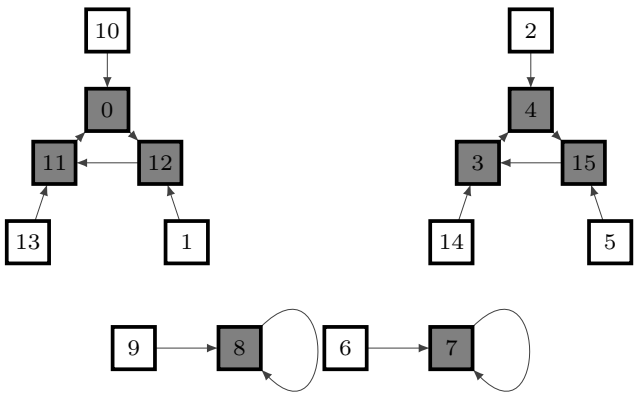

Figure 4.3: *The state space of the network shown in Figure 4.2, if the functions copy, copy, invert, invert are assigned to the four nodes. The numbers in the squares represent states, and arrows indicate the successor of each state. States on attractors are shaded.*

There are 4 attractors, two of which are fixed points (i.e., attractors of length 1). The sizes of the basins of attraction of the 4 attractors are $\Omega_1 = 6, \Omega_2 = 6, \Omega_3 = 2, \Omega_4 = 2$. If the function of node 1 is a constant function, fixing the value of the node at 1, the state of this node fixes the rest of the network, and there is only one attractor, which is a fixed point.

4.5 Phase transitions

In RBNs, as well as in many dynamical systems, three phases can be distinguished: *ordered, chaotic, and critical*. These phases can be identified with different methods, since they have several unique features. these dynamical phases is related to "sensitivity to initial conditions", "damage spreading", and "robustness to perturbations"

which are different ways of measuring the stability of a network. We can "mutate", "damage" or "perturb" a node of a RBN by flipping its state. We can also change a connection between two nodes, or in the lookup table of a node. Since nodes affect other nodes, we can measure how much a random change affects the rest of the network. In other words, we can measure how the damage spreads. This can be done by comparing the evolution of a "normal" network and a "perturbed" network. In the ordered regime, usually the damage does not spread: a "perturbed" network "returns" to the same path of the "normal" network. This is because changes cannot propagate from one green island to another. In the chaotic phase, these small changes tend to propagate through the network, making it highly sensitive to perturbations [21]. An other feature is the convergence versus divergence of the trajectories in state space of the network dynamics. In the ordered phase, similar states tend to converge to the same state. In the chaotic regime, similar states tend to diverge. At the edge of chaos, nearby states tend to lie on trajectories that neither converge nor diverge in state space. Living systems, or computing systems, need certain stability to survive, or to keep information; but also flexibility to explore their space of possibilities. This has lead people to argue that life and computation occur more naturally at the edge of chaos or at the ordered regime close to the edge of chaos [24][21].

Very early in the studies of RBNs, people realized in simulations that the networks with $K \leq 2$ were in the ordered regime, and networks with $K \geq 3$, were in the chaotic regime. In Figure 4.4 we can appreciate characteristic dynamics of RBNs in different phases. We can identify phase transitions in RBNs in different ways. The main idea is to measure the effect of perturbations, the sensitivity to initial conditions, or damage spreading. This is analogous to Lyapunov exponents in continuous dynamics. The phase transitions can be statistically or analytically obtained. Derrida and Pomeau were the first to determine analytically that the critical phase (edge of chaos) was found when $K = 2$ [25][32][27]. following the model of Kauffman, RBN which most represent biological GRNs are those wich has $K = 2$ [24], because in the frozen phase (where $K = 1$) networks are too simple to represent real regulatory networks; while in the caothic phase (where $K = 3$) the time scales of the networks cycles grow exponentially, which is not biologically pheasible. In Chapter 7 we will see the differences

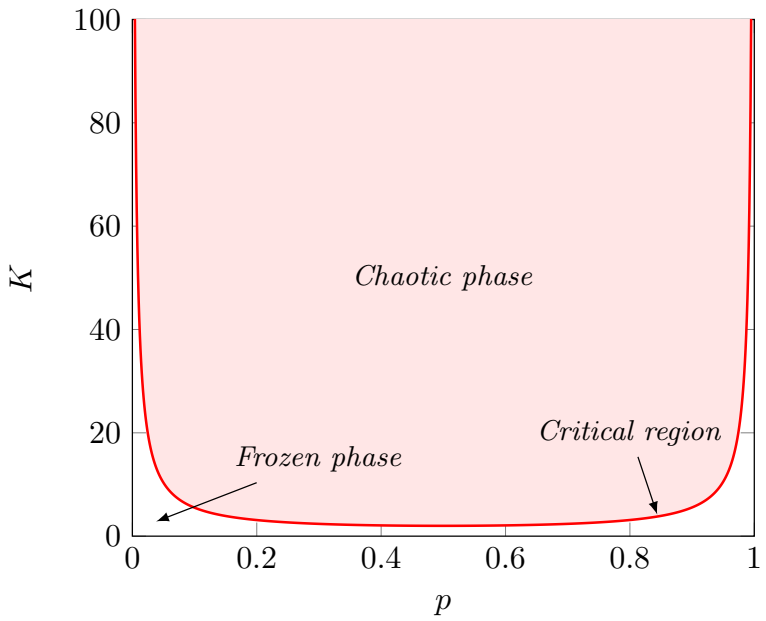

Figure 4.4: *Phase diagram for the $N - K$ model. The shaded area corresponds to the chaotic phase, whereas the white region corresponds to the chaotic phase. The curve separating both regions is the critical phase.*

in $K = 1$ networks and $K = 2$ networks.

4.6 Perturbations

We consider attractors in the state space as gene regulatory networks of different cells, where different attractors represent cells of different type, and where for example cancer cells lay in one specific attractor [8][9]. Now, if we suppose that different cell types lay in different attractors, we suppose that the jump from an attractor to one other is given by a perturbation in the binary sequence of the genes. So for example we take the previos network, and consider that we are in the state 12 of the first attractor:

$$12 \to 11 \to 0 \to 12$$

And suddenly we change the state of the third node from 0 to 1:

$$1100 \rightarrow 1110$$

We change the system to have the state 14 and so we jump into the second attractor:

$$14 \rightarrow 3 \rightarrow 4 \rightarrow 15 \rightarrow 3$$

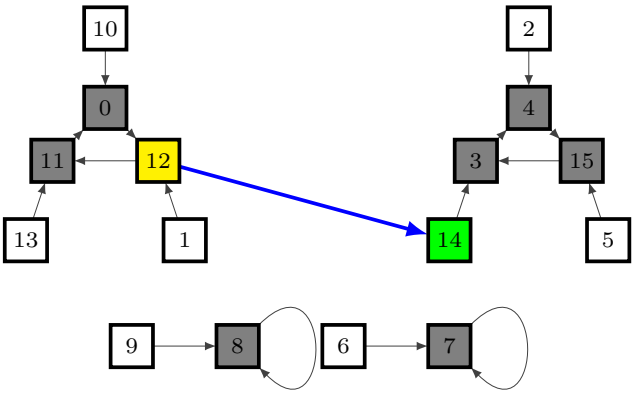

Figure 4.5: *Jump from one attractor to one other in the state space.*

So the branching pathways of differentiation between attractors in a RBN in the ensembles create a directed graph showing which attractors can be perturbed to reach which attractors.

In fact, if we consider all the possible stochastic perturbations in the binary sequence of the genes, we can get the all the possible transitions between the attractors, and what we obtain is an other network, where the nodes are the attractors and the frequencies of transitions can be used to build a random walk on this network[9].

Chapter 5

Kramer Rate Theory

In this Chapter we make a briefly introduction in stochastic process and in Kramer Rate Theory.

5.1 Fokker-Planck equation

Fokker-Planck equation is a second order parabolic partial differential equation, and it's born during the investigation of the Brownian motion in a radiation field by Fokker and Planck [28]. This equation describe the time evolution of a probability distribution $\rho(x,t)$ in a stochastic process. In one dimension, the Fokker-Planck equation (FPE) takes the simple form:

$$\frac{\partial \rho(x,t)}{\partial t} = -\frac{\partial}{\partial x}\Big(A(x,t)\rho(x,t)\Big) + \frac{1}{2}\frac{\partial^2}{\partial x^2}\Big(B(x,t)\rho(x,t)\Big) \tag{5.1}$$

with the initial condition:

$$\rho(x,t_0|x_0,t_0) = \delta(x-x_0) \tag{5.2}$$

The first term in the equation 5.1 is the drift equation, while the last term is the diffusion equation. Every stochastic process described by a conditional probability satisfying the FPE is equavalent to the Ito stochastic differential equation (SDE)

$$dx(t) = A(x(t),t)dt + \sqrt{B(x(t),t)}dw_t \tag{5.3}$$

and the two descriptions are complementary to each other.

5.2 Bistability and Escape Problems

In Nature exist some systems which have at least two possible stable states. This kind of systems are known as *bistable systems* [28]. For such systems there are some problems of interest:

- How stable are the various states relative to each other?
- How long does it take for a system to switch spontaneously from one state to onother?
- How is the path in the relevant state space?
- How does a system relax from an unstable state?

These questions can all be answered reasily for one-dimensional diffusion processes.

5.2.1 Diffusion in a Double-Well Potential in one dimension

We consider the system in which we have the probability density $\rho(x,t)$ of a prarticle, which obeys the Fokker-Planck equation:

$$\frac{\partial \rho}{\partial t} = \frac{\partial}{\partial x} U'(x)\rho + D\frac{\partial^2}{\partial x^2}\rho \tag{5.4}$$

The shape of $U(x)$ is shown in Figure 5.3.

There are two minima at a and c and in between, a local maximum b. The stationary distribution is

$$\rho_s(x) = \mathcal{N} e^{-\frac{U(x)}{D}} \tag{5.5}$$

where $\mathcal{N}$ is the normalization constant.This distribution explains the bistability of the system, since, we have two local maxima of the potential in a and c. (see Figure 5.2).

5.2.2 Behaviour for $D = 0$

In this case, $x(t)$ obeys the differential equation

$$\frac{dx}{dt} = -U'(x) \tag{5.6}$$

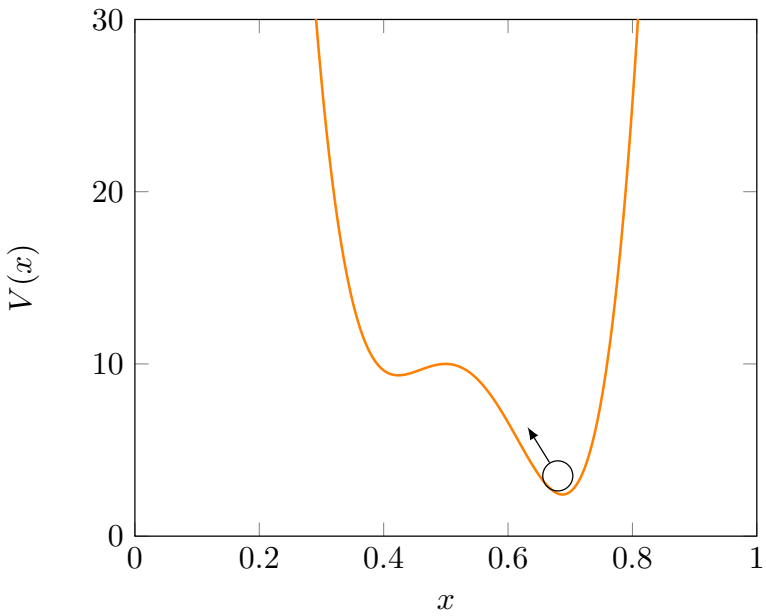

Figure 5.1: *Example of double-well potential.*

with initial condition $x(0) = x_0$. Now, since

$$\frac{dU(x)}{dt} = U'(x)\frac{dx}{dt} = -(u'(x))^2 < 0$$

$x(t)$ always moves in such a way as to minimise $U(x)$, and stops only when $U'(x)$ is zero. Thus, depending on whether x_0 is grater than or less than b, the particle ends up at c or a, respectively. Once the particle is at a or c, it stays there. If it starts exactly at b, it also stay there, thoug the slightest perturbation drives it to a or c. Thus, b in an *unstable* stationart point and a and c are stable.

5.2.3 Behaviour if D is small

With the addition of noise, the situation changes. The stationary state can be approximated asymptotically as folows. Assuming $U(x)$ is everywhere sufficiently smooth, we can write

$$\begin{aligned} U(x) &\simeq U(a) + \frac{1}{2}U''(a)(x-a)^2 \\ &\simeq U(c) + \frac{1}{2}U''(c)(x-c)^2 \end{aligned} \tag{5.7}$$

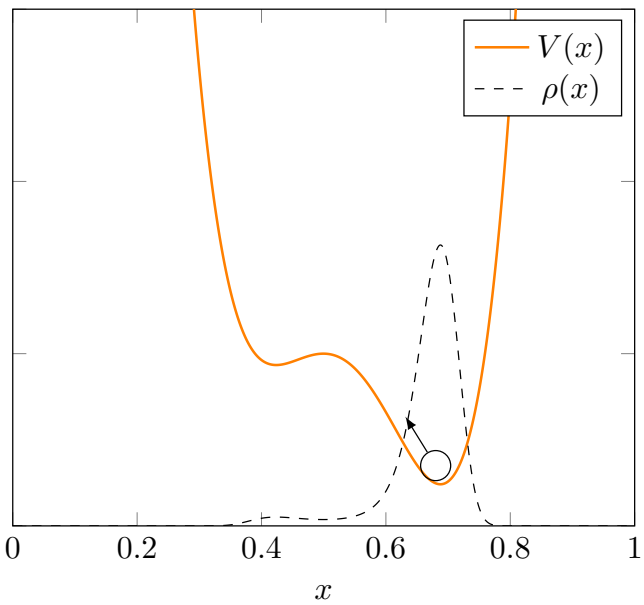

Figure 5.2: *The distribution (dashed) explains the bistability of the system, since, we have two local maxima of the potential in a and c..*

for $|x-a|$ and $|x-c|$ small, respectively. If D is very small, then we may approximate

$$\begin{aligned}\rho_s(x) &\simeq \mathcal{N}\exp\left(-\frac{U(a)}{D}-\frac{1}{2}U''(a)\frac{(x-a)^2}{D}\right)\\ &\simeq \mathcal{N}\exp\left(-\frac{U(c)}{D}-\frac{1}{2}U''(c)\frac{(x-c)^2}{D}\right)\end{aligned} \tag{5.8}$$

And zero elsewhere. So that:

$$\mathcal{N}^{-1} \simeq e^{-\frac{U(a)}{D}}\sqrt{\frac{2\pi D}{U''(a)}}+e^{-\frac{U(c)}{D}}\sqrt{\frac{2\pi D}{U''(c)}} \tag{5.9}$$

Now suppose as in drawn in Figure 5.2

$$U(a) > U(c)$$

Then for small enough D, the second term is larger than the first and $\mathcal{N}^{-1}$ can be approximated by it alone. So, substituing into 5.8 we find

$$\rho_s(x) = \sqrt{\frac{U''(c)}{2\pi D}}\exp\left(-\frac{1}{2}U''(c)\frac{(x-c)^2}{D}\right) \tag{5.10}$$

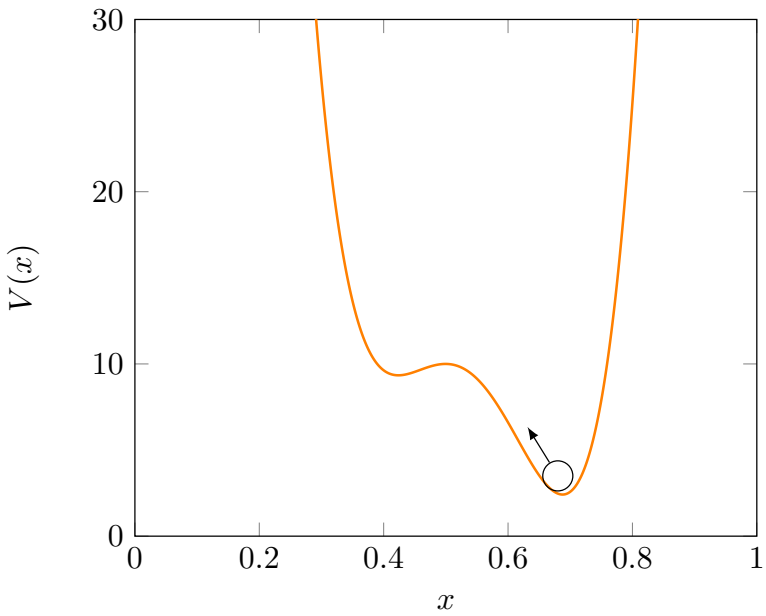

Figure 5.3: *Example of double-well potential.*

for

$$|x - c| \sim \sqrt{D}$$

and zero otherwise. This means that in the limit of very small D, the deterministic stationary state at which $U(x)$ has an absolute minimum is the more stable state in the sense that in the stochastic stationary state, $\rho_s(x)$ is very small everywhere except in its immediate vicinity. This means, differently from the previous results, that due to the noise the equilibrium points a and c are no longer equally stable.

5.3 Kramer Rate Theory

We consider the Smoluchowski equation:

$$dx = -V'(x)dt + \sqrt{2T}dw_t \tag{5.11}$$

where $V(x)$ is a double well potential with x_a and x_c local minima separated by a saddle point x_b: To compute the transition probability from the left well to the right

well we consider the stationary solution of the Fokker-Planck equation:

$$\frac{\partial \rho}{\partial t} = \frac{\partial}{\partial x} U'(x)\rho + T\frac{\partial^2}{\partial x^2}\rho \tag{5.12}$$

with the boundary condition that we have a source in the point $x_- < x_a$ and an absorbing boundary at the point $x_+ > x_c$. We assume that the temperature T is much les of the potential barrier $V_b - V_a$ and we look for a solution that reduces to the form:

$$e^{-\frac{V(x)}{T}}$$

in the vicinity of the point x_a that vanishes at the absorbing barrier and gives a constant current J between the two wells. Let

$$\rho(x) = C(x)e^{-\frac{V(x)}{T}} \tag{5.13}$$

and the particle density current $-J = V'(x)\rho + \frac{\partial \rho}{\partial x}T$ reads

$$V'(x)C(x)e^{-\frac{V}{T}} + C'(x)Te^{-\frac{V}{T}} - V'(x)C(x)e^{-\frac{V}{T}} = -J$$

so that:

$$C'(x) = -\frac{J}{T}e^{\frac{V}{T}}$$

and we integrate with the condition $C(x_+) = 0$:

$$C(x) = \frac{J}{T}\int_x^{x_+} e^{\frac{V(y)}{T}}dy$$

The distribution reads:

$$ho(x) = \frac{J}{T}e^{-\frac{V(x)}{T}}\int_x^{x_+} e^{\frac{V(y)}{T}}dy.$$

When$x \simeq x_a$ the integral

$$\int_x^{x_+} e^{\frac{V(y)}{T}}dy$$

is stationary since $V^(x_a) = 0$ and $ho(x)$ has the behavior of $e^{-\frac{V(x)}{T}}$. Then we have to compute the number of particles n_a in the left well

$$n_a = \frac{J}{T}\int_{-\infty}^{x_b} e^{-\frac{V(x)}{T}}\int_x^{x_+} e^{\frac{V(y)}{T}}dydx$$

We approximate the integrals using the saddle point method:

$$e^{\frac{V(y)}{T}} \simeq e^{\frac{V_b}{T} - \frac{\omega_b^2}{2T}(y-x_b)^2}$$

$$e^{\frac{V(x)}{T}} \simeq e^{\frac{V_a}{T} - \frac{\omega_a^2}{2T}(x-x_a)^2}$$

so we compute:

$$\frac{n_a}{J} = \frac{1}{T} e^{\frac{V_b - V_a}{T}} \int_{-\infty}^{x_b} dx e^{-\frac{\omega_a^2}{2T}(x-x_a)^2} \int_{x}^{x_+} e^{-\frac{\omega_b^2}{2T}(y-x_b)^2} dy$$

Then we can extend both the integrals between $-\infty$ and $+$
$infty$:

$$\int_{-\infty}^{+\infty} dx e^{-\frac{\omega_a^2}{2}(x-x_a)^2} = \frac{\sqrt{2\pi T}}{\omega_a} \frac{n_a}{J} = \frac{2\pi}{\omega_a \omega_b} e^{\frac{V_b - V_a}{T}}$$

And we find the transition probability rate:

$$k_{a \to c} = \frac{J}{n_a} \simeq \frac{\omega_a \omega_b}{2\pi} e^{\frac{V_b - V_a}{T}}$$

5.4 First passage time and transition rate escaping from a barrier

We first consider the generic diffusion problem

$$\frac{\partial \rho}{\partial t}(x,t) = \mathbb{L}_{FP} \rho(x,t)$$

where $\mathbb{L}_{FP}$ is a generic Fokker-Planck operator and we assume that $x \in \Omega$ and $\partial\Omega$ is an absorbing boundary condition: $\rho(x,t) = 0$ for $x \in \partial\Omega \ \forall t$. By definition the probability to be in Ω at time t for an initial condition $\delta(x-a)$ is:

$$P_\Omega = \int_\Omega \rho(x,t|a) dx$$

Let t_a the first passage time at $\partial\Omega$ for a realiation starting from a at $t = 0$ and let $p(t|a)$ be the probability distribution, we have by definition

$$\int_t^\infty p(s|a) ds = P_\Omega(t|a)$$

since $\partial\Omega$ is an absorbing boundary condition. Then

$$p(t|a) = \int_\Omega \frac{\partial}{\partial t}\rho(x,t|a)dx = \int_\Omega \mathbb{L}_{FP}\rho(x,t|a)dx$$

To get an equation for the average first passage time $\tau(a)$

$$\tau(a) = \int_0^\infty tp(t|a)dt = \int_\Omega \int_0^\infty t\frac{\partial}{\partial t}\rho(x,t|a)dtdx = -\int_0^\infty \int_\Omega \rho(x,t|a)dxdt$$

we use the adjoint operator $\mathbb{L}_{FP}^\dagger$ and taking into account the boundary conditions

$$\mathbb{L}_{FP}^\dagger \tau(a) = -\int_0^\infty \int_\Omega \mathbb{L}_{FP}^\dagger \rho(x,t|a)dxdt$$

Remark: $\mathbb{L}_{FP}^\dagger$ acts on the initial state a. From the Kolmogorov relation for Markov processes we get

$$\frac{\partial}{\partial s}\int_\Omega dt\rho(x,t|x_0,s)\rho(x_0,s|a,0)dx_0 = 0$$

so that

$$\begin{aligned}\int_\Omega dx_0\rho(x_0,s|a,0)\frac{\partial}{\partial s}\rho(x,t|x_0,s) + \rho(x,t|x_0,s)\frac{\partial}{\partial s}\rho(x_0,s|a,0) = \\ \int_\Omega dx_0\rho(x_0,s|a,0)\frac{\partial}{\partial s}\rho(x,t|x_0,s) + \rho(x,t|x_0,s)\mathbb{L}_{FP}\rho(x_0,s|a,0)\end{aligned} \tag{5.14}$$

Then:

$$\int_\Omega dx_0\left[\frac{\partial}{\partial s}\rho(x,t|x_0,s) + \mathbb{L}_{FP}^\dagger\rho(x,t|x_0,s)\right]\rho(x_0,s|a,0)$$

and since we consider a stationary process

$$\frac{\partial}{\partial s}\rho(x,t|x_0,s) = \frac{\partial}{\partial s}\rho(x,t-s|x_0,0) = -\frac{\partial}{\partial t}\rho(x,t-s|x_0,0)$$

Letting $s \to 0$, one gets the adjoint equation

$$\frac{\partial}{\partial t}\rho(x,t|a,0) = -\mathbb{L}_{FP}^\dagger\rho(x,t|a,0)$$

Accoring to the previous result, the average first passage time reads

$$\mathbb{L}_{FP}^\dagger\tau(a) = -\int_0^\infty \int_\Omega \mathbb{L}_{FP}^\dagger\rho(x,t|a)dxdt = \int_\Omega \int_0^\infty \frac{\partial}{\partial t}\rho(x,t|a,0)dtdx = -\int_\Omega \delta(x-a)dx = -1$$

since $\lim_{t\to\infty} \rho(x,t|a) = 0$ due to the presence of the absorbing boundary condition. In an explicit form we have the equation

$$\mathbb{L}^{\dagger}_{FP}\tau(a) = -1 \tag{5.15}$$

We remark that the computation of the average first passage time requires the knowloedge of the evolution at all times, however this result can be related to the escape probability rate. Let us apply the first passage time to the computation of the Kramer probability escape from a potential well. We start from the Fokker-Planck equation

$$\frac{\partial}{\partial x}\frac{\partial V}{\partial x}\rho + D\frac{\partial^2 \rho}{\partial x^2} = \frac{\partial \rho}{\partial t}$$

where the potential $V(x)$ has a minimum at $x = x_0$ and a saddle point at $x = x_1$. The average firste time $\tau(x)$ id written:

$$-\frac{\partial V}{\partial x}\frac{\partial \tau}{\partial x} + D\frac{\partial^2 \tau}{\partial x^2} = -1$$

and we get

$$\frac{\partial}{\partial x}\exp\left(-\frac{V(x)}{D}\right)\frac{\partial \tau}{\partial x} = -\frac{1}{D}\exp\left(-\frac{v(x)}{D}\right)$$

The previous equation can be integrated

$$\frac{d\tau}{dx} = -\frac{1}{D}\exp\left(\frac{V(x)}{D}\right)\int_{-\infty}^{x} dz \exp\left(-\frac{V(z)}{D}\right)$$

where we assume $\lim_{x\to-\infty} V(x) = \infty$. The final result is

$$\tau(x) = \frac{1}{D}\int_{x}^{x_a} \exp\left(\frac{V(y)}{D}\right)\int_{-\infty}^{y} \exp\left(-\frac{V(z)}{D}\right) dzdy \tag{5.16}$$

Then the Kramer's estimate follows by approximating the integral at the critical points of the potential (x_a is not se near the local maximum x_1) so that the result is independent from x. The transition rate is definded:

$$k_c = \frac{1}{\tau}$$

The fact that k_c does not depend on x is due to the fast relazation scale time inside the potential well with respect to the escape time scale from the potential well and to

the quasi-stationary distribution that concentrates the particles near the local minimal point where the effect of the potential can be approximated by a parabolic potential: then we estimate $\tau \simeq \tau(x)$ where $x \simeq x_0$. In a analogous way one considers the stationary solution in presence of a local source at $x_s \leq x_0$ and the absorbing barrier at $x_1 lex_a$. The stationary solution obeys to the differential equation

$$\frac{\partial}{\partial x}\frac{\partial V}{\partial x}\rho(x) + D\frac{\partial^2 \rho}{\partial x^2} = -J\delta(x - x_s)$$

thath can be integrated

$$\rho(x) = \frac{J}{d}\exp\left(\frac{-V(x)}{D}\int_x^{x_a}\exp\left(\frac{V(y)}{D}\right)\Theta(y - x_s)dy\right.$$

The parameter J defines the current at the absorbgin barrier and the transition rate can be computed as the ratio between J and the number of particles in the potential well, $k_x = J/N$ where:

$$N = \frac{J}{D}\int_{-\infty}^{x_1}\exp\left(\frac{-V(x)}{D}\right)\int_x^{x_a}\exp\left(\frac{V(y)}{D}\right)\Theta(y - x_s)dydx \tag{5.17}$$

and the Kramer's estimate follows by approximating the integral 5.17. Assuming that $V(x)$ has a local minimum A and a local maximum B $x_s < x_a < x_b$.

Chapter 6

The model

Models of boolean networks proposed by Kauffmann are limited and can't represent perfectly biological networks because of some considerations: simple RBNs present cahotic behaviours and attractors are not sufficient to explain cellular differentiation. In this Chapter we present the theoretical model of GRNs based on RBNs.

6.1 The underlying philosophy of the model

The dynamical model is a schematic representation of the activity of gene regulatory networks introduced in Chapter 3. We have to discuss the assumptions the define the model from a biological point of view: the main criticism to a model is that its assumptions cannot be justified by the biological mechanisms. Our goal is to model the genetic activity related to a differentiation process of a cell: i.e. this activity is a stable long term activity whose stability is probably controlled by biochemical mechanisms (i.e. methylation processes), but for cancer cells the control dynamics is not so efficient allowing the evolution of different cell populations. Then we assume that this evolution is possible due to the competition of different genetic activities through dynamical mechanisms that can be triggered by the external environmental signals. In particular we assume:

- the long term genetic activity is determined by the presence of small genetic networks that have a stable active dynamical state;

- there exists an eternal control mechanism: the subnetworks have control nodes that prevent the arise of the active state in the subnetwork if there are set to the inactive state;
- once the active state has been established in a subnetwork it remains stable in time without any stimulus, except if an inhibitory stimulus change the state of control nodes;
- the stability and the controllability properties of a subnetwork depends from the existence of loops in the subnetwork: a loop may be related to the activation of metabolic cycles in the cell that define the cell behavior;
- each node of a subnetwork may represent the state of a gene that is connected and regulates the activation of other genes;
- in the cell differentiation mechanism is defined by the competition of different subnetworks that interact in a inhibitory way;
- the mutation mechanism change the connectivity of the network: we may distinguish between permanent changes and dynamical change (i.e. a connection may exist or non exist during time).

The complexity of the model is not a fundamental issue since we want to point out universal behaviors: first of all the existence of bistability or bifurcation phenomena for simple model and the definition of control parameters.

6.2 The mathematical model and related problems

Here we studied the model dynamics in different situations using mathematical methods. The main idea is understand the dynamics of the models to point out the universal properties that are robust and could explain the experimental data. The biological meaning of control parameters is a fundamental task to apply the model to predict the results of new experiments.

We consider a physical system that can be described by an weighted interaction network among nodes that can assume different dynamical states (in the case of a gene network the states $\sigma \in \{0, 1\}$ and we have models similar to spin models). The interaction structure is defined by signed adjacency matrix $A_{ij} \in \{-1, 0, 1\}$ where the sign refers to a cooperative or antagonist interaction between the connected nodes. In the simplest case, we introduce a stochastic dynamics using the probability $p_i(\sigma, t)$ that the node i is in the state σ (we assume $\sigma > 0$) at time t: in a deterministic approach $p_i(\sigma, t) = \delta(\sigma - \sigma(t))$ to denote that the node assume the state $\sigma = 1$ with probability one. The evolution of a deterministic model can be described by the equation

$$\sigma_i(t+1) = \Phi_i(\sigma(t)) = \Theta\left(\sum_j A_{ij}\sigma_j(t)\right) \tag{6.1}$$

where $\Theta(x) \in [0, 1]$ is a threshold sigmoidal function (we assume $A_{ii} = 0$ to avoid self loops).

Remark: the dynamics is a information diffusion on the network. If we consider the linear system

$$\zeta_i(t+1) = \sum_j A_{ij}\zeta_j(t)$$

where ζ_i are non negative integers we have an equivalent dynamics since $\sigma_i = \Theta(\zeta_i)$ and it is possible to study the linear system to derive some properties of the initial system. For example the relaxation time to the solution $\sigma_i = 1 \; \forall \; i$ is for a given initial condition $\sigma_j^0 = \delta_{jk}$ is $t = n$ such that the matrix A^n has positive entries along the whole k-th column. This mens that for each node i there is a walk of length n from the initial node k to i.

We also assume a cause-effect relation so that A_{ij} is a directed graph. The deterministic model is a Hopfield network (each node has at least an input and an output link; the environment nodes has only output links) and one could study the equilibrium states and their stability. An equilibrium condition as follows is characterized as follows: for each i let

$$Q_i(t) = \sum_j A_{ij}\sigma_j(t)$$

then $Q_i > 0$ if $\sigma_i > 0$ and vice versa. Then $A_{ij} \geq 0$(i.e. A_{ij} is a connectivity matrix for a directed network) implies that the non trivial equilibrium is $\sigma_i = 1$: if $\sigma_k = 0$ for

some $k \in K$ then we have

$$\sum_{j \notin K} A_{kj}\sigma_j = 0$$

so $A_{kj} = 0$ for all $j \notin K$ and the network is disconnected. Then we have the trivial solution $\sigma_i = 0$. For each equilibrium solution σ^* we have a stability basin

$$S_{\sigma^*} = \left\{ \sigma \mid \lim_{t \to \infty} \sigma(t) = \sigma^* \right\}$$

If S_{σ^*} defined neighborhood of σ^* the solution is stable or if $S_{\sigma^*} = \{\sigma^*\}$ the solution completely unstable. The stability of the origin depends on the existence of a Ljapounov function: let introduce the network activity

$$\Sigma(t) = \sum_i \sigma_i(t) = \sum_i \Theta(Q_i(t-1)) \geq \Sigma(t-1)$$

since if each node has at least one input link, $A_{ij} = 1$ implies $\sigma_j(t-1) \Rightarrow \sigma_i(t) = 1$ and the activity cannot decrease. The solution $\sigma_i = 0$ is completely unstable. If there would exists an equilibrium solution with $\sigma_k = 0$ for some k then we define S_A the set of nodes s.t.

$$i \in S_A \quad \Rightarrow \quad \sigma_i = 0$$

(obviously $\sigma_k \in S_A$). Let $S_{\bar{A}}$ the complement of S_A, the network dynamics implies

$$0 = \sum_j A_{ij}\sigma_j = \sum_{j \notin S_A} A_{ij}\sigma_j = 0 \qquad \text{if} \quad i \in S_A$$

so that $A_{ij} = 0$ if $i \in S_A$ and $j \in S_{\bar{A}}$: i.e. there is not a cause-effect connection between $S_{\bar{A}}$ and S_A and the state $\sigma_i = 1$ for $i \in S_{\bar{A}}$ is an equilibrium state. Therefore we have as many equilibrium states as many partitions S_A and $S_{\bar{A}}$ there exist such that S_A triggers the activity of $S_{\bar{A}}$ but not vice versa. For any initial condition $\sigma_i(0) = \delta_{ik}$ the possible evolution are a periodic orbit or an equilibrium state: one can detect all the equilibrium conditions by σ^* by the condition

$$\sigma_i^* = 1 \quad \text{if} \quad \sigma_i(t) = 1 \quad \text{for some} \quad t \geq 0$$

The equilibrium states are a semigroup: let σ^a and σ^b two equilibrium states the

$$\sigma^a \cup \sigma^b = \sigma^c$$

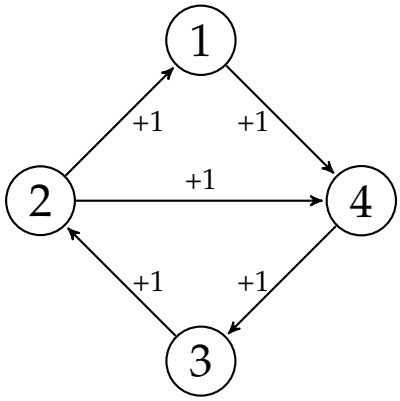

Figure 6.1: *Example of random boolean network.*

is still an equilibrium. An example: if there exit a one directional loop γ in the network and there is no output link from γ to the remaining nodes of the network then $\sigma_i = 1$ for any $i \in \gamma$ is an equilibrium. If the loop is simple (each node has a one input link and one output link) the equilibrium is neutral since any change $\sigma_i = 1 \rightarrow \sigma_i = 0$ creates a periodic orbit (the total activity is constant). But if we we add a link to the loop then we get a stable solution since a single node can trigger the activity of two nodes and the equilibrium is an attractive stationary state (see figure). If a node is accidentally set to zero this anomaly propagates in the loop, until it reaches the node 4 where it is annihilated by the activity of the node (2). The average lifetime of a single perturbation is the average path length to propagate to the node (4) from the initial node (therefore it depends from the loop length or in case of presence of many loops, the average path length is computed considering independent loops).

The boolean network models the propagation of information. By studying the stability problem of the solution $\sigma_i = 1$ it is convenient to introduce the dual dynamics:

$$\sigma_i^c(t+1) = \Theta\left(\prod_{j\sim i} A_{ij}\sigma_j^c(t)\right) = \prod_{j\sim i} A_{ij}\sigma_j^c(t)$$

where $\sigma_i^c = 1 - \sigma_i$ is the dual state of the node and the product is restricted to the nodes connected to i ($A_{ij} \neq 0$): i.e. the node (4) takes the state $\sigma^c = 1$ only if both

the nodes (1) and (2) in that state at previous time. This dynamics is valid for any configuration of the network and the state $\sigma^c = 1$ moves on the network until it reaches an absorbing state for which

$$\prod_{j\sim i} \sigma_j^c(t) = 0 \quad \forall\, i$$

For a given stable equilibrium σ^γ state associated to a loop γ any environmental perturbation that set to zero a activity of a node will destroy the equilibrium after a time equal to the number of the loop nodes minus one. For example in the figure there are two loops ((1) $\to$ (2) $\to$ (3) $\to$ (4)) and ((2) $\to$ (4) $\to$ (3)) if we set to zero the node (4) after three iterations all the nodes will be in the zero state. The two loops are nit independent since one loops contains the other). On the contrary if we set to zero the node (1) one loop remains active. This remark allows to introduce the concept of control node: a node is a control node if its state is able to force the state of the whole network. The effect of a thermal bath could be introduced by assuming that the state of a node is defined as random variable that takes value $\sigma_i(t) = 1$ with probability $p_i(t)$ where

$$p_i(t+1) = \Theta_T\left(\sum_j A_{ij}\sigma_j(t)\right) \tag{6.2}$$

and $\Theta_T(x)$ is a logistic function

$$\Theta_T(x) = \frac{1}{2}\left(1 + \mathrm{tgh}(x/T - \epsilon)\right)$$

where ϵ measures to tendency of the network to be in the idle state when no stimulus is present. The logistic function is a generic sigmoidal function we do not expect that the specific form of $\Theta_T(x)$ is critical for the results.

Remark: since the values of x are quantized to integer in any case, if $\epsilon > T^{-1}$ the idle state is statistically attractive so ϵ could define a critical temperature for the network activation. We recover the deterministic dynamics for $T \to 0$. As a stochastic process we have a Markov process (since the realization of the variable $\sigma_i(t+1)$ depends only on the present state $\sigma_j(t)$ of the network. The dynamics (6.2) is a Markov field: the realization of the variable σ_i depends only from the present state of the network (and not from past states) and only from the states of the connected nodes

$A_{ij} \neq 0$. The last condition (Markov field) means that the realizations of $\sigma_i(t)$ and $\sigma_j(t-1)$ are independent if the nodes are not connected. The transition probabilities depend from the state of the network and one derives the average dynamics

$$< \sigma_i > (t+1) = p_i(t+1) = \left\langle \Theta_T \left(\sum_j A_{ij}\sigma_j(t) \right) \right\rangle \simeq \Theta_T \left(\sum_j A_{ij} p_j(t) \right) \tag{6.3}$$

Then we have two possibilities: if the total average network activity tends to increase

$$\bar{\Sigma}(t+1) = \sum_i p_i(t+1) = \sum_i \Theta_T \left(\sum_j A_{ij} p_j(t) \right) > \bar{\Sigma}(t) \tag{6.4}$$

the equilibrium solution $\sigma_i = 1$ is attractive, on the contrary we have a an average tendency to decrease the network activity. The situation is illustrated in the fig.

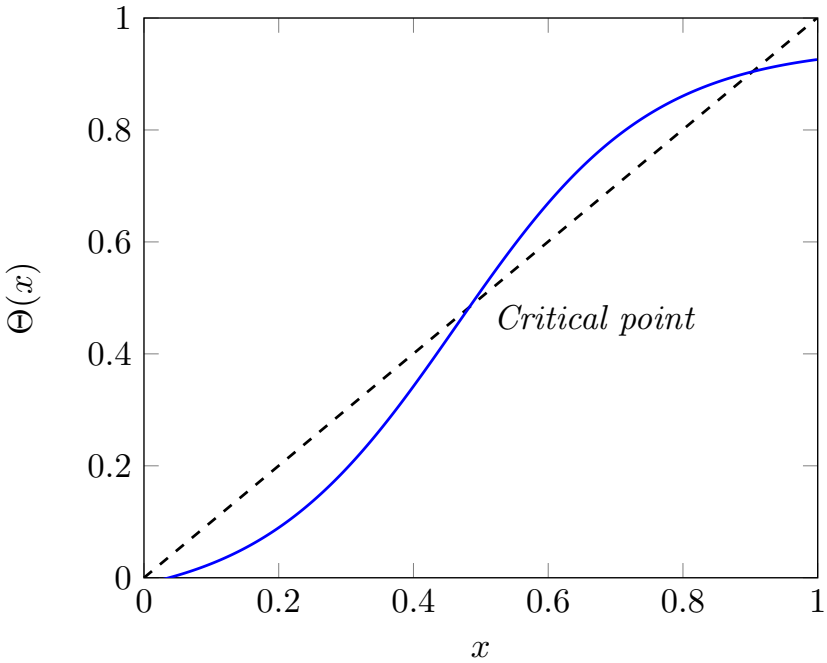

Figure 6.2: *Possible behavior for the condition (6.4); the units are arbitrary and scale with the network dimension.*

Remark: the mean field approximation apply when the $\Theta_T(x)$ can be approximated by a linear function locally: i.e. the fluctuations are small enough to approximate the

function by a linear function in the whole fluctuation range. This is certainly not true when we have fat tail fluctuations.
Except for a small initial region, the condition (6.4) can be satisfied up to a critical value of the network activity Σ (if the temperature is not too big), so that the average activity tends to increase. But if the activity is below the critical value then the network activity tend to decrease and the stability of the solution $\sigma_i = 1$ is lost. A connected network tends to be more stable since the quantities $\sum_j A_{ij}\sigma_j$ increase. This picture is clearly an approximation since we neglect the fluctuation effects: if the fluctuations are big (this depends also on the connectivity matrix) we may have a fast transition between the two possible regime and a correction of the critical value. The critical vale is a consequence of the sigmoidal behavior of the $\Theta_T(x)$ function and its depends on the temperature and on the ϵ values. In presence of fluctuations and of two dynamical regimes (active and non active) we expect that the network activity may switch from one regime to another with a characteristic time scale (Kramer transition rate Theory, see Chapter 5). The transition may be triggered by large fluctuations that are both consequence of rare events (in such a case the probability should be exponentially small with respect the activity) but also depend on the network structure (the presence of hub nodes that can change the activity of many nodes amplifies the effect of small fluctuations (i.e. the change of the hub node state) and may introduce fat tail statistic in the fluctuation distribution). A second stochastic effect is related to the fluctuations of the connectivity due to environmental causes: the matrix $A_ij(t)$ is a stochastic process (so that its entries change their value according to a probability distribution). The simplest model can be formulated as follows: we assume that the nonzero entries $A_{ij}(t)$ assume value 1 with a given probability p (independent from the network state) each time step Δt (i.e. we are not simulating a parametric white noise, but a correlated random noise with a define correlation time scale Δt). Δt is the shortest evolution time scale for the system (we need a physical interpretation) and we set $\Delta t = 1$. The effect of a parametric noise is substantially different from the environmental noise and the evolution equation (6.1) reads

$$\sigma_i(t+1) = \Theta\left(\sum_j A_{ij}(t)\sigma_j(t)\right) \tag{6.5}$$

In such a case the average dynamics is not useful and the problem can be studied by the representative dynamics

$$\zeta_i(t+1) = \sum_j A_{ij}(t)\zeta_j(t) \qquad \Rightarrow \qquad \zeta(t) = \prod_{k=1}^{t} A(k)\zeta(0)$$

where the solution is the product of random matrices (there are results on the spectral properties). From a biological point of view means that the interaction of genes depends also by external factors.
One could say that the network is active if a certain condition is satisfied (for example the average total activity should overcome a given threshold) so that fluctuation may introduce the existence of non active states. Problems for a single network: starting from a loops with a fixed dimension adds randomly links to stabilize the equilibrium solutions (existence of sub-loops) and study the robustness of the solution and the recovery times in relation with the connectivity matrix; adding the temperature, study the existence of critical value for the appearance of the equilibrium solution; the thermodynamic limit. The effect of the environmental noise has to be justified from a biological point of view by relating it to the individual variability of the cell phenotypes in an homogeneous population.
In the stochastic models one should also consider the problem that the connectivity matrix is not fixed (for example we have a ensemble of admissible matrices or the existence of the links is a random event). In such a case we have a stochastic dynamics

$$\sigma_i(t+1) = \Theta\left(\sum_j A_{ij}(t)\sigma_j(t)\right) \tag{6.6}$$

where $A_{ij}(t)$ is a random process with value $\in \{0,1\}$ maintaining some average properties of the connectivity; this is an alternative to the environmental noise (parametric noise). This model simulates the fact that the activation of a link may depends on random events (i.e. not only from the existence of the link) so that a genetic network is indeed a stochastic network. In principle any realization of the connectivity matrix A_{ij} has an equilibrium $\sigma_i = 1$ but the robustness of equilibrium can be influenced by the fluctuations. Problem: if the robustness of the equilibrium with respect to the external perturbations (i.e. an external signal on a node) depends on the spectral properties of

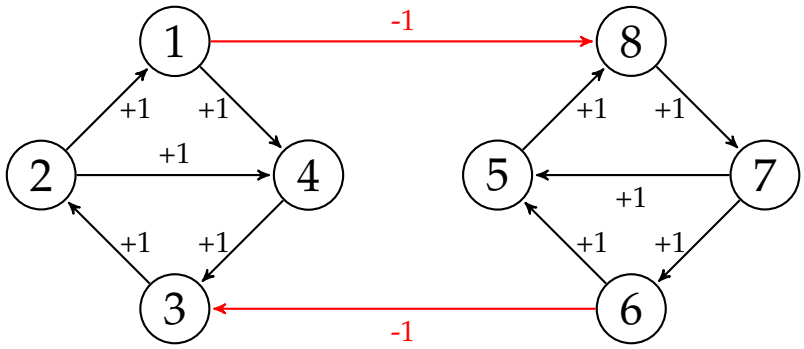

Figure 6.3: *Example of a network composed by two competitive subnetworks.*

the connectivity matrix then it is possible to study the spectral properties of random connectivity matrices and develop a control theory for the network.
Let us consider the existence of competitive networks (see Figure 6.3) that are linked by inhibitory links: if the first network is in an exited state the second network should be completely switched off for a stable equilibrium.

If we start with all the node states set to one we create a frustrated situation, otherwise the network choose one of the two possible stable states. In such a case the presence of an environmental noise could induce the transition to one state to another (to be studied). An external forcing breaks the symmetry.
We expect a transition phase as a function of the temperature: increasing the temperature the node states tends to be independent, but under a threshold the system should choose a stationary state.
The system can be generalized to consider the interactions of different cooperative networks (possibly with different internal structure) that are connected by inhibitory links (in the case of connection with excitatory links we join the subnetwork in a single one). We can introduce a metadynamics where $\nu_k(t)$ is the state of the k subnetwork and we have a relation

$$\nu_k(t+\Delta t) - \nu_k(t) = \phi(\nu_k(t)) - \gamma\left(H_{kj}\nu_j(t)\right) \tag{6.7}$$

where $H_{hk} \geq 0$ is an inhibitory connectivity matrix. $\phi(\nu_k(t))$ describes the tendency of the sub-network to increase its activity and γ the average decreasing of the activity

due to the presence of other sub-networks. This is an effective equation: ν_k should describe the network activity (i.e. it could be the time-average activity of the nodes assuming that the network could be considered in a stationary state). Indeed the evolution time scale Δt could be assumed $\Delta t \gg 1$ so that the subnetwork states are relaxed to a stationary states. The structure of attraction basins of the stable states could be related to a potential in the state space if

$$\nu_k(t+1) - \nu_k(t) = -\frac{\partial}{\partial \nu_k}\left[\frac{\gamma}{2}\sum_{ij}\nu_i H_{ij}\nu_j + \sum_j V(\nu_j)\right]$$

where

$$\phi(\nu) = -\frac{\partial V}{\partial \nu}$$

Since $\phi(\nu) \geq 0$ $V(\nu)$ is increasing. Then we introduce the energy

$$E = \frac{\gamma}{2}\sum_{ij}\nu_i H_{ij}\nu_j + \sum_j V(\nu_j)$$

and the equilibrium are the critical points of the energy. Moreover

$$\begin{aligned} E(t+1) - E(t) &\simeq (\nu_k(t+1) - \nu_k(t))\frac{\partial}{\partial \nu_k}\left[\frac{\gamma}{2}\sum_{ij}\nu_i(t) H_{ij}\nu_j(t) + \sum_j V(\nu_j(t))\right] \\ &= -\frac{1}{2}\frac{\partial}{\partial \nu_k}\left[\frac{\gamma}{2}\sum_{ij}\nu_i(t) H_{ij}\nu_j(t) + \sum_j V(\nu_j(t))\right]^2 \end{aligned}$$

Therefore the energy is a Ljapounov function and the system equilibria are defined by the critical points of the Energy function corresponding to local minima and maxima. Remark: the existence of the Energy implies that H_{ij} is symmetric negative defined

$$\frac{\partial^2 E}{\partial \nu_j \partial \nu_i} = \frac{\partial^2 E}{\partial \nu_i \partial \nu_j}$$

The stochastic effect has to be introduce but it is possible a thermodynamics approach and a thermodynamics equilibrium exists according to the Maxwell-Boltzmann distribution and the detailed balance condition. This means that the whole network does not satisfy this condition, but the metadynamic network realized a reversible

Markov process. The existence of a thermodynamic equilibrium allows to use Maximal Entropy Principle and the Maxwell Boltzmann distribution when we introduce a thermal bath.
We are interested in networks with many different equilibria each one related to exited state of subnetworks (or a combination of subnetworks), in the effect of a thermal noise and in the effect of external forcing. The external we introduce in the network boundary nodes whose state is defined by a given external signal $\sigma_b(t)$ (possible a stochastic process) then the network dynamics reads

$$\sigma_i(t+1) = \Theta\left(\sum_j A_{ij}\sigma_j(t) + \sum_b A_{ib}\sigma_b(t)\right)$$

where A_{ib} is the link between the environmental node b and the node i. It is possible to introduce a probabilistic description of the evolution of the probability that the network is in the state σ' at time $t+1$ according to

$$p(\sigma', t+1) = \sum_\sigma \pi(\sigma', \sigma) p(\sigma, t) \tag{6.8}$$

where

$$\pi(\sigma'|\sigma) = E\left(\delta_{\sigma', \Phi_{\sigma_b}(\sigma)}\right)$$

and the expectation value is computed on the realization of the input noise. $\pi(\sigma'|\sigma)$ is the transition rate per unit time (the continuous limit could be considered). Let σ_{eq} stable equilibrium state the effect of external random perturbations could be to move the network state in a neighborhood of the equilibrium solution or it could induce a transition to other equilibrium basin attractions so that the dynamics starts to perform an intermittence behavior. In such a case the relevant quantities are the residence times in the different basins that can be associated to metastable states.
We introduce the stochasticity in the system assuming that the adjacency matrix is not known: i.e. A_{ij} is a extracted from an ensemble of random matrices. As the result of an experimental one could assume that each entry A_{ij} is a dichotomous random variable with probability p_{ij} to get the value ± 1 (i.e. the link is active). The value $p_{ij} = 0$ is admitted so that the corresponding link it always inactive. The problems are:

1. Classifying the equilibrium states in relation to their robustness with respect the changes in the adjacency matrix;

2. Understanding the representativity of the average dynamics: i.e. substituting the adjacency matrix with an average matrix one highlights the dynamical properties that are correctly described by the average system

3. Pointing out the existence of bifurcation phenomena so that it is possible to divide the ensemble in different communities with similar dynamical behaviors.

Chapter 7

Analysis

In this Chapter we explain the starting implementation and analysis of the model following biological considerations.

7.1 Implementation

The theoretical model has benn analyzed making simulations using Python. The first thing was to create a class *Random Network* to have a random boolean graph as an object, with its nodes and links, represented by a boolean adjacency matrix.

Every random network is a directed graph and is built to avoid self-loops, this averages to create a random, boolean, and non-symmetric adjacency matrix with null trace.

The first thing to evaluate was to choose the average number of incoming links for each RBN. In Figure 7.1 we can see that the average discrete evolution of 100 different realizations of RBNs, with increasing size. In the case of $K = 1$, i.e. the average number of incoming links for each network is one, the average activity decreases exponentially with the size of the network; in the case of $K = 2$ instead, the average activity of the nodes remains stable with networks of increasing size. For this reason, the network considered for this model will take the parameter K constant: $K = 2$. Moreover, we can analyze the average number of outgoing links depending on the network size: In Figure 7.2, we can see that the average number of outgoing links

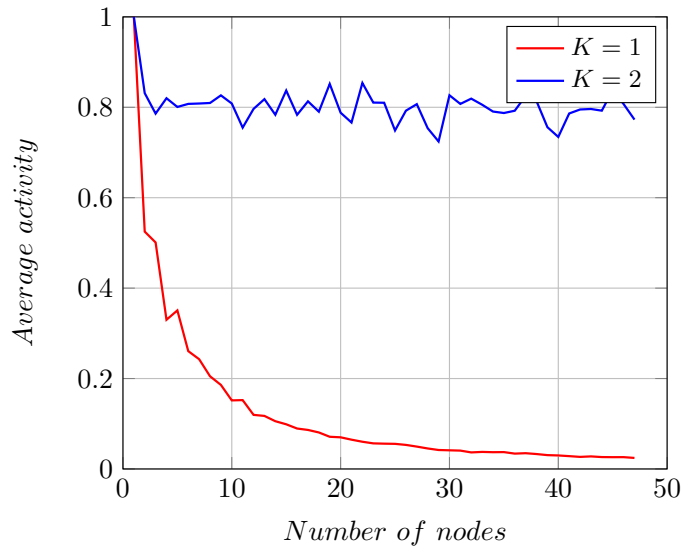

Figure 7.1: *Plot of the average activity of the nodes with network of increasing size. In the case of $K = 1$ (i.e. the average number of incoming link for each network is one), the average activity decreases exponentially with the size of the network; in the case of $K = 2$ instead, the average activity of the nodes remains stable with the network size.*

tends to the parameter K.

7.2 Discrete evolution

As shown in Chapter 6, the discrete time evolution of the network is given by the equation:

$$\sigma_i(t+1) = \Theta\left(\sum_j A_{ij}\sigma_j(t)\right)$$

where A is the connectivity matrix of the network. So this averages that each node which has at least one incoming link with a node which is active, in the next step this node will be active. At each time step we can measure the average activity of the network, which is the average number of nodes with the value:

$$\sigma_i(t) = 1$$

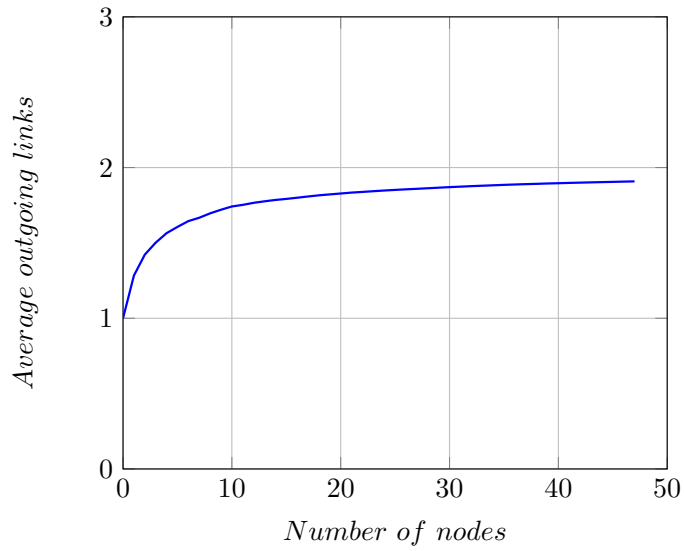

Figure 7.2: *Plot of the number of the outgoing links depending on the network size. We can see that the average tends to the parameter K.*

7.3 Loops and Control nodes

In simple random networks one can always find the number of independent loops. Independent loops determine the complexity of the networks[46], and are important for the construction of the model. In fact from the independent loops one can find the *control nodes* of the network, which are the nodes that theirstate are able to force the state of the whole network, and are the nodes with maximum connectivity. Control nodes determine the whole activity and stability of the network. In RBNs with parameter $K = 2$, we can see in Figure 7.3 that the number of independent loops is linear with the network size.

7.4 Noise

The second thing to evaluate is the effect of the noise on the evolution of the network and the difference between noise and parametric noise, where parametric noise refers to the noise which infers in the links and not on the nodes. To add noise to the system, during the discrete evolution of the network, at each time step there is a

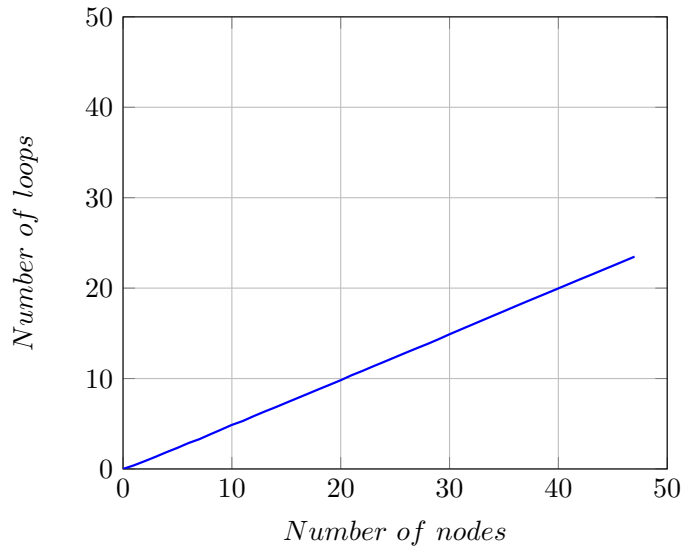

Figure 7.3: *Plot of the number of independet loops in the networks depending on the network size.*

probability p for the node or for the link to be turned off. In Figure 7.4 we can see the behavior of the average activity of the network depending on the amount of noise added. The plot is similar to a sigmoidal functions in both of the cases. In the case of noise added to the links the average activity results to be bigger than the case of noise added to the nodes. This is reasonable in the sense that since the number of links is less than the number of nodes so the effect of the noise among the links is less.

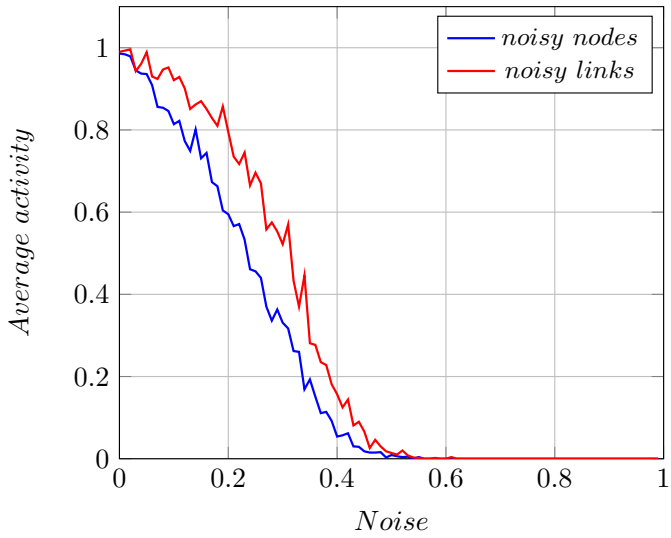

Figure 7.4: *Plot of the effect of the noise on the average activity on the network. In blue the noise works on the nodes on the network, while in red the noise works on the links. Number of nodes for each network: 10; Number of realizations for each value of noise: 100;*

Conclusions

Bibliography

[1] Waddington CH, *The strategy of the genes: a discussion of some aspects of theoretical biology*. London: Allen and Unwin, (1957)

[2] Cameron P. Gallivan, Honglei Ren and Elizabeth L. Read,*Analysis of Single-Cell Gene Pair Coexpression Landscapes by Stochastic Kinetic Modeling Reveals Gene-Pair Interactions in Development*,doi: 10.3389/fgene.2019.01387 ,(2019)

[3] Jifan Shi, Tiejun Li , Luonan Chen, Kazuyuki Aihara,*Quantifying pluripotency landscape of cell differentiation from scRNA-seq data by continuous birth-death process*,https://doi.org/10.1371/journal. pcbi.1007488 ,(2019)

[4] Jin Wang, Kun Zhang, Li Xu, and Erkang Wang ,*Quantifying the Waddington landscape and biological paths for development and differentiation*, https://doi.org/10.1073/pnas.1017017108 ,(2011)

[5] S. Huang, I. Ernberg, S. Kauffman,*Cancer attractors: A systems view of tumors from a gene network dynamics and developmental perspective*, doi:10.1016/j.semcdb.2009.07.003, (2009)

[6] S. A. Kauffman, *Metabolic Stability and Epigenesis in Randomly Constructed Genetic Nets*, J. Theoret. Biol. (1969)

[7] Drossel B.,*Random Boolean Networks*,arXiv:0706.3351 ,(2008)

[8] R.Serra, M. Villani, A. Barbieri, S.A. Kauffman, A. Colacci,*On the dynamics of random Boolean networks subject to noise: Attractors, ergodic sets and cell types.*,J Theor Biol 265: 185–193, (2010)

[9] M. Villani, A. Barbieri, R. Serra,*A Dynamical Model of Genetic Networks for Cell Differentiation*, doi:10.1371/journal.pone.0017703.g001,(2011)

[10] M. Ali Al-Radhawi , Nithin S. Kumar, Eduardo D. Sontag , Domitilla Del Vecchio, *Stochastic multistationarity in a model of the hematopoietic stem cell differentiation network*,doi:10.1109/cdc.2018.8619300, (2018)

[11] Janeway, *Immunobiology*, 9th Edition

[12] E. Davidson, M. Levine, *Gene Regulatory Networks*, doi:10.1073/pnas.0502024102, (2005)

[13] Chen L et al., *Biomolecular networks: methods and applications in systems biology*, Wiley, Hoboken ,(2009)

[14] Peccoud, J. and Ycart, B., *Markovian Modelling of Gene Products Synthesis.*, https://doi.org/10.1006/tpbi.1995.1027, (1995)

[15] T.B. Kepler and T.C. Elston, *Stochasticity in transcriptional regulation: origins, consequences, and mathematical representations*, 10.1016/S0006-3495(01)75949-8 , (2001)

[16] J.M. Pedraza, J. Paulsson, *Effects of molecular memory and bursting on fluctuations in gene expression*, 10.1126/science.1144331, (2008).

[17] Y. Sasai *Cytosystems dynamics in self-organization of tissue architecture*, https://doi.org/10.1038/nature11859, (2013)

[18] B. Zhang and P. G. Wolynes, *Stem cell differentiation as a many-body problem*, https://doi.org/10.1073/pnas.1408561111, (2014)

[19] Zhou et al., *Fast Pyrolysis of Glucose-Based Carbohydrates withAdded NaCl Part 2: Validation and Evaluation of theMechanistic Model*,DOI 10.1002/aic.15107, (2016)

[20] A. Wuensche, *Genomic regulation modeled as a network with basins of attraction*, Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing, (1998)

[21] S. A. Kauffman, *Investigations*, Oxford University Press, (2000)

[22] MacArthur S. et al., *Developmental roles of 21 Drosophila transcription factors are determined by quantitative differences in binding to an overlapping set of thousands of genomic regions*,doi:10.1186/gb-2009-10-7-r80 (2009)

[23] S. A. Kauffman: J. Theor. Biol., 44, Physica D, 10, 145 (1984)

[24] S. Kauffman, *A proposal for using the ensemble approach to understand genetic regulatory networks*,Journal of Theoretical Biology 230 (2004) 581–590 ,(2004)

[25] B. Derrida, *Random Networks of Automata: A Simple Annealed Approximat ion.*, (1985)

[26] B. Derrida and H.Flyvbjerg, *The random map model: a disordere model with deterministic dynamics*, J.Physique, (1987)

[27] B. Derrida, *Spin glasses, random boolean networks and simple models of evolution*

[28] C.W. Gardiner, *Handbook of Stochastic Methods*, Springer, (1985)

[29] Sui Huang , Ingemar Ernberg, and Stuart Kauffman,*Cancer attractors: A systems view of tumors from a gene network dynamics and developmental perspective*, DOI:10.1016/j.semcdb.2009.07.003, (2009)

[30] M. Rybarsch and S. Bornholdt,*On the dangers of Boolean networks: Activity dependent criticality and threshold networks not faithful to biology*, arXiv:1012.3287v1. (2010)

[31] J. Park and M. E. J. Newman,*The statistical mechanics of networks*, DOI: 10.1103/PhysRevE.70.066117 (2004)

[32] C. Gershenso,*Introduction to Random Boolean Networks*, arXiv:nlin/0408006,(2004)

[33] R. V. Solè, B. Loque, *Phase transitions and antichaos in generalized Kauffman networks*,Physics Letters,(1994)

[34] J. T. Lizier, S. Pritam, M. Prokopenko, *Information dynamics in small-world Boolean networks*, (2011)

[35] A. Rèka and A-L. Barabàsi, *Statistical mechanics of complex networks*, Reviewes of modern physics, Volume 74,(2002)

[36] N. Masuda , M. A. Porter, R. Lambiotte ,*Random walks and diffusion on networks*,Physics Reports 716–717 1–58,(2017)

[37] T. Biyikoglu, J. Leydold, P. F. Stadler,*Laplacian Eigenvectors of Graphs*, Springer

[38] Fan R. K. Chung,*Spectral Graph Theory*, CBMS

[39] M. Ali Al-Radhawi, Nithin S. Kumar, Eduardo D. Sontag, Domitilla Del Vecchio ,*Stochastic multistationarity in a model of the hematopoietic stem cell differentiation network*,DOI:10.1109/cdc.2018.8619300.,(2018)

[40] Rushina Shah, Domitilla Del Vecchio,*Reprogramming cooperative monotone dynamical systems*,DOI:10.1109/cdc.2018.8618649, (2018)

[41] Atefeh Taherian Fard and Mark A. Ragan, *Modeling the Attractor Landscape of Disease Progression: a Network-Based Approach*,DOI: 10.3389/fgene.2017.00048, (2017)

[42] Sui Huang, Yan-Ping Guo, Gillian May, Tariq Enver,*Bifurcation dynamics in lineage-commitment in bipotent progenitor cells*,Developmental Biology 305, (2007)

[43] Xin-She Yang and Young Z. L. Yang,*Cellular Automata Networks*,arXiv:1003.4958 , (2010)

[44] Xin Kang, Chunhe Li,*Landscape inferred from gene expression data governs pluripotency in embryonic stem cells*,Computational and Structural Biotechnology Journal 18 (2020) 366–374,(2020)

[45] Genaro J. Martinez , Andrew Adamatzky , Bo Chen , Fangyue Chen , Juan C.S.T. Mora ,*Simple networks on complex cellular automata: From de Bruijn diagrams to jump-graphs*,(2017)

[46] Schnakenberg J., *Network theory of microscopic and macroscopic behavior of master equation systems.*, Reviews of Modern Physics,Vol.48, (1976)