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

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 of cells in which cells of a specific type reproduces themselves and give raise 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 gene reproduce a posbbile Gene Regulatory Network for the T-cells of the immune system, which present a bistable nature.

Contents

List of Figures

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 [13] 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[10][11][12], 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[7].

Kauffman proposed for the first time in 1969[1] 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"[4]. 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[5][6][9].

Omics data available are subject to noise, and we cant 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 2 we make a biological introduction of Cell Differentiation and of Gene Regulatory Network. In Chapter 3 we present the concept of Random Boolean Networks, a model proposed for the first time by Kauffman to model Gene Regulatory Networks. In Chapter 5 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 we make a numerical analysis of the thoretical model proposed in 5.

Chapter 1

Cell Differentiation and cancer cells

1.1 Gene Regulatory Networks

All steps of gene expression can be modulated, since passage of the transcrip- tion of DNA to RNA, to the post-translational modification of the protein produced. Hence, gene expression is a complex process regulated at several stages in the synthesis of proteins. In addition to the DNA transcription reg- ulation, the expression of a gene may be controlled during RNA processing and transport (in eukaryotes), RNA translation, and the post-translational modification of proteins. This gives rise to genetic regulatory systems struc- tured by networks of regulatory interactions between DNA, RNA, proteins and other molecules [6]: a complex network termed as a gene regulatory network (GRN). Some, noteworthy, kind of proteins are the transcrip- tion 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 (though they

all possess the genotype, which follows the same genome sequence) 4 . Therefore, with few exceptions, all cells in an organism contain the same genetic material [6], and hence the same genome (the haploid set of chromosomes of a cell). 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.

1.2 Cell Differentiation

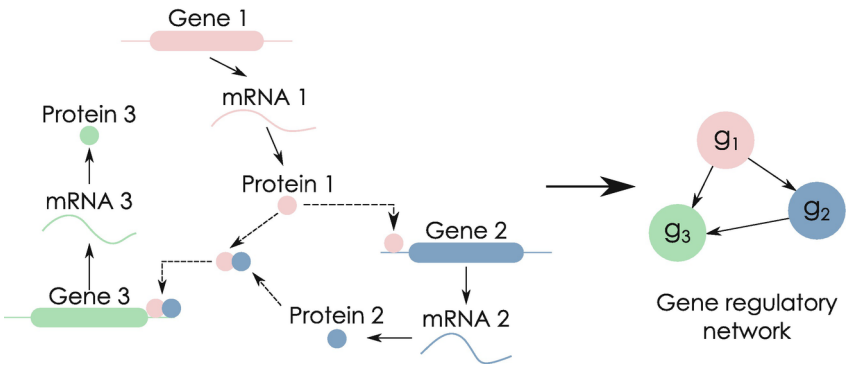
Cell differentiation is the process whereby stem cells become progressively more specialized. The differentiation process occurs both during the devel- opment of a multicellular organism and during tissue repair and cell turnover in the adulthood. Gene expression, and therefore its regulatory mechanisms, plays a critical role in cell differentiation; as described in the previous section. Stem cells are undifferentiated biological cells which can both reproduce themselves, self-renewal ability, and differentiate into specialized cells, po- tency.

Chapter 2

Gene Regulatory Networks

2.1 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. 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 [37]. 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.

2.2 Role within the cell

Gene Regulatory Networks (GRNs) control biological process of all organisms. The complex control systems underlying development have probably been evolving for more than a billion years. They regulare the expression of thousand of genes in any given biological process. They are essentially hardwired genomic regulatory codes, the role of which is to speicify 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.

2.3 Bo

In recent years, single-cell-resolution measurements have revealed unprecedented levels of cell-to-cell heterogeneity within tissues. The discovery of this ever-present heterogeneity is driving a more nuanced view of cell phenotype, wherein cells exist along a continuum of cell-states, rather than conforming to discrete classifications. The comprehensive view of diverse cell states revealed by single cell measurements is also affording new opportunities to discover molecular regulators of cell phenotype and dynamics of lineage commitment (Trapnell et al., 2014; Olsson et al., 2016; Briggs et al., 2018). For example, single cell transcriptomics have revealed the widespread nature of multilineage priming (MLP), a phenomenon wherein individual, multipotent cells exhibit promiscuous coexpression of genes associated with distinct lineages prior to commitment (Nimmo et al., 2015). In principle, mathematical modeling of gene regulatory network 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.

Mathematical models of gene regulatory network dynamics can account for (and at least partially reproduce) observed cellular heterogeneity in two primary ways. First, gene network models are multistable 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 (Kauffman, 1969; MacArthur et al., 2009; Huang, 2012). Second, some mathematical models inherently treat cellular noise. This noise, or stochasticity, is modeled in various ways depending on assumptions about the source (Peccoud and Ycart, 1995; Arkin et al., 1998; Kepler and Elston, 2001; Swain et al., 2002). 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 (Peccoud and Ycart, 1995;

Thattai and van Oudenaarden, 2001; Kepler and Elston, 2001; Pedraza and Paulsson, 2008). 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 (Feng and Wang, 2012; Sasai et al., 2013; Zhang and Wolynes, 2014; Tse et al., 2015).

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 (Wang et al., 2011; Huang, 2012; Waddington, 2014) 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 (Wang et al., 2011; Bhattacharya et al., 2011; Huang, 2012; Zhou et al., 2016). For stochastic gene network models that inherently treat noise, the landscape is directly obtained from the computed probability distribution over cell-states (Cao and Liang, 2008; Micheelsen et al., 2010; Feng and Wang, 2012; Tse et al., 2015).

Stochastic modeling of gene network dynamics has been employed in various forms for analysis of single cell measurements. For example, application of noisy dynamical systems theory has shed light on cell-state transitions (Mojtahedi et al., 2016; Jin et al., 2018; Lin et al., 2018). Stochastic simulations of gene network dynamics have been used to develop and/or benchmark tools for network reconstruction (Schaffter et al., 2011; Dibaeinia and Sinha, 2019; Bonnaffoux et al., 2019) Stochastic model-aided analysis of single-cell measurements has been demonstrated to yield insights on gene regulatory mechanisms (Munsky et al., 2018). 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. 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.

2.4 summary Kauffman

We know most of the genes, the coding regions, some of the cis regulatory sites and transcription factors, some of the protein components of cell signaling cascades that are driven by transcription and translation, and in turn feedback to regulate gene activities. Let me refer to this whole system as the genetic regulatory network. One of the outstanding problems of contemporary systems biology is to understand the structure, logic and dynamics of this network within and between cells.

real gene networks are random after 3.8 billion years of selection.

Chapter 3

Random Boolean Networks

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

3.1 Random Boolean Networks

Random Boolean networks (RBNs) were introduced in 1969 by S. Kauffman as a simple model of genetic systems. 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 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: it is on an attractor.

3.2 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 $i = 1, ..., N$. 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:

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

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:

$$\sigma_i(t + 1) = \Lambda_i(\sigma_{i_1}(t), \sigma_{i_2}(t), ..., \sigma_{i_K}(t)). \tag{3.1}$$

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).

3.3 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 statis- tical weights are usually different. Let us consider the simplest possible example, $N = 2$ and $K = 1$, shown in Figure 3.1. 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

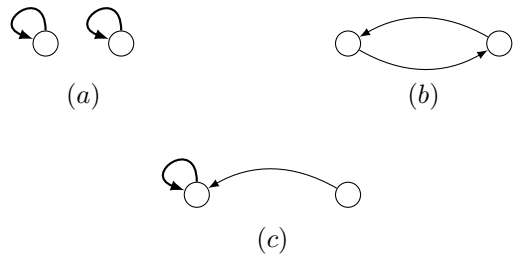


Figure 3.1: Set of all possible networks with $N = 2$ and $K = 1$.

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 outgo- ing 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:

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

3.4 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 3.2, which consists of 4 nodes:

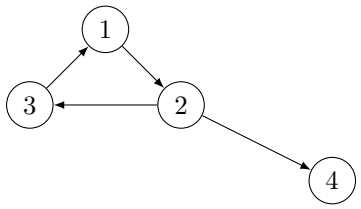


Figure 3.2: A small network with $N = 4$ and $K = 1$.

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:

$$1111 \to 0011 \to 0100 \to 1111$$

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

$$15 \to 3 \to 4 \to 15$$

The entire state space is shown in Figure 3.3:

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 6,6,2,2. 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. Its basin of

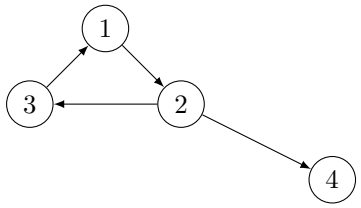


Figure 3.3: The state space of the network shown in Figure 3.2, if the functions copy, copy, invert, invert are assigned to the four nodes. The numbers in the squares represent states, and ar- rows indicate the successor of each state. States on attractors are shaded.

attraction is of size 16. If the functions of the other nodes remain unchanged, the state space then looks as shown in Figure 3.4

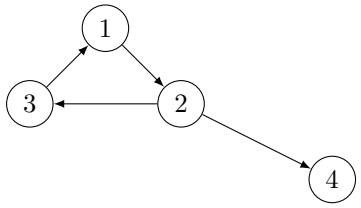


Figure 3.4: The state space of the network shown in Figure 3.2, if the functions 1, copy, invert, invert are assigned to the four nodes.

Before we continue, we have to make the definition of attractor more precise: as the name says, an attractor attracts states to itself. A periodic sequence of states (which we also call cycle) is an attractor if there are states outside the attractor that lead to it. However, some netaworks contain cycles that cannot be reached from any state that is not part of it. For instance, if we removed node 4 from the network shown in Figure 2.2, the state space would only contain the cycles shown in Figure 2 C, and not the 8 states leading to the cycles. In the follow- ing, we will use the word cycle whenever we cannot be confident that the cycle is an attractor.

3.5 Applications

Let us now make use of the definitions and concepts introduced in this section in order to derive some results concerning cycles in state space. First, we prove that in an ensemble of networks with update rule 1 (biased functions) or rule 2 (weighted classes), there is on an average exactly one fixed point per network. A fixed point is a cycle of length 1. The proof is slightly different for rule 1 and rule 2. Let us first choose rule 2. We make use of the property that for every update function the inverted function has the same probability. The inverted function has all 1s in the output replaced with 0s, and vice versa. Let us choose a network state, and let us determine for which fraction of networks in the ensemble this state is a fixed point. We choose a network at random, prepare it in the chosen state, and perform one update step. The probability that node 1 remains in the same state after the update, is 1/2, because a network with the inverted function at node 1 occurs equally often. The same holds for all other nodes, so that the chosen state is a fixed point of a given network with probability 2^N . This means that each of the 2 N states is a fixed point in the proportion 2^N of all networks, and therefore the mean number of fixed points per network is 1. We will see later that fixed points may be highly clustered: a small proportion of all networks may have many fixed points, while the majority of networks have no fixed point.

Next, we consider rule 1. We make now use of the property that for every update function a function with any permutation of the input states has the same probability. This means that networks in which state A leads to state B after one update, and networks in which another state C leads to state B after one update, occur equally often in the ensemble. Let us choose a network state with n 1s and Nn 0s. The average number of states in a network leading to this state after one update is $2^N p^n (1p^{Nn}$. Now, every state leads equally often to this state, and therefore this state is a fixed point in the proportion $p^n(1p)^{Nn}$ of all networks. Summation over all states gives the mean number of fixed points per network, which is 1.

Finally, we derive a general expression for the mean number of cycles of length L in networks with $K = 2$ inputs per node. The generalization to other values of K is straightforward. Let $\langle C_L \rangle_N$ denote the mean number of cycles in state space of length L, averaged over the ensemble of networks of size N . On a cycle of length L, the state of

each node goes through a sequence of 1s and 0s of period L. Let us number the 2^L possible sequences of period L of the state of a node by the index j, ranging from 0 to $m = 2^L 1$. Let n_j denote the number of nodes that have the sequence j on a cycle of length$L$, and $(P_L)^j_{l,k}$ the probability that a node that has the input sequences l and k generates the output sequence j. This probability depends on the probability distribution of update functions.

Then

$$\langle C_L \rangle_N = \frac{1}{L} \sum_{n_j} \frac{N!}{n_0! \dots n_m!} \prod_j \left(\sum_{l,k} \frac{n_l n_k}{N^2} (P_L)^j_{l,k} \right)^{n_j} \tag{3.2}$$

The factor 1/L occurs because any of the L states on the cycle could be the starting point. The sum is P over all possibilities to choose the values n j such that j n j = N . The factor after the sum is the number of different ways in which the nodes can be divided into groups of the sizes n 0 , n 1 , n 2 , . . . , n m . The product is the probability that each node with a sequence j is connected to nodes with the sequences l and k and has an update function that yields the output sequence j for the input sequences l and k. This formula was first given in the beautiful paper by Samuelsson and Troein [10].

Chapter 4

Cancer attractors

In questo capitolo viene fatta una piccola introduzione al modello classico preda-predatore di Lotka-Volterra.

4.1 The model

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). In the simplest case, we introduce a stochastic dynamics using the probability $p_i(t)$ that the node i is in the state $\sigma_i = 1$ (then $1 - p_i(t)$ is the probability to get $\sigma_i = 0$) and we define a linear equation for the probability evolution

$$\dot{p}_i(t) = \sum_j \mathcal{P}_{ij} p_j(t) - \gamma_i p_i(t) \tag{4.1}$$

where $\mathcal{P}_{ij}$ are transition probability rates and γ_i^{-1} defines the mean lifetime of the excited state. The meaning of the rates $\mathcal{P}_{ij}$ is the rate at which the excited state of the node j increases (or decreases if $\mathcal{P}_{ij} < 0$) the probability of a transition to the excited state of the node i. Since $0 \leq p_i \leq 1$ for all i, this space should be invariant for the dynamics. This condition depends on the spectral properties of the matrix

$$\mathcal{P}_{ij} - \gamma_j \delta_{ij} \tag{4.2}$$

associated to the system. Let consider the case $\mathcal{L}_{ij} \geq 0$ (i.e. we have no inhibitory link), the first quadrant is clearly invariant and if we define

$$\sum_i \mathcal{P}_{ij} = \hat{\gamma}_j > 0$$

the matrix

$$\mathcal{L}_{ij} = \mathcal{P}_{ij} - \hat{\gamma}_j \delta_{ij}$$

is a Laplacian matrix and the system (4.1) can be written in the form

$$\dot{p}_i(t) = \sum_j \mathcal{L}_{ij} p_j(t) - \Delta\gamma_i p_i(t) \qquad \Delta\gamma_i = \gamma_i - \hat{\gamma}_i$$

and by assumption we have $\gamma_i > \hat{\gamma}_i$. The eigenvalues of the matrix $\mathcal{L}_{ij}$ have all negative real part except the null eigenvalue. It follows that all the eigenvalue of the matrix (4.2) has negative real part and the dynamics is a contraction towards the origin: a stable solution (i.e. without any external stimulus the system relaxes to the $\sigma_i = 0$ state). A non trivial stationary can be achieved only if an external stimulus is inserted

$$\dot{p}_i(t) = \sum_j \mathcal{P}_{ij} p_j(t) - \gamma_i p_i(t) + \epsilon f_i(t) \tag{4.3}$$

The stationary solution has to satisfy $p_i \in [0, 1]$ so that $f_i(t) \geq 0$ otherwise we can have negative probability when $p_i \simeq 0$. The case of a Laplacian matrix

$$\hat{\gamma}_i = \gamma_i$$

we get another possible stationary solution for $\mathcal{L}_{ij} p_j^* = 0$ in the first quadrant and the subspace $\sum p_i = 0$ is invariant and the dynamics is a contraction in this subspace (in general). Then the system a stable stationary solution even in absence of an external stimulus.
The presence of inhibitory links complicates the model and one has to prove that

- 1) there exists a physical space: an invariant cone in the first quadrant where the dynamics is a contraction towards the origin;
- 2) the external stimulus maintains the solution in the physical space.

Another solution could be to introduce boundary conditions so that $p_i \geq 0$ in any case (the system is non linear in such a case).
The eigenvalues of the matrix (4.2) define the different relaxation time scale the process and determine its rectivity to the change of the external stimulus: in a typical problem one consider a slowly varying external stimulus so that the system could be considered i a quasi stationary state

$$\sum_j \mathcal{L}_{ij} p_j - \Delta\gamma_i p_i = -\epsilon f_i(t) \qquad \frac{df_i}{dt} \ll 1$$

the derivative is small with respect to the eigenvalues of th matrix (adiabatic approximation). On the other hand we have the effect of a correlated noise (we need to introduce a correlation in order have a continuous function $f_i(t)$). The problem is to study the relation between the solution and the spectral properties of the matrix $\mathcal{L}_{ij}$: we simplify the equation by assuming $\Delta\gamma_i = \Delta\gamma$ so that if λ is an eigenvalue of $\mathcal{L}_{ij}$ then $\lambda - \Delta\gamma$ is an eigenvalue of the matrix (4.2) and we assume that the dynamics is perturbed by

$$\dot{p}_i(t) = \sum_j \left(\mathcal{L}_{ij} + \Delta\mathcal{L}_{ij}\right) p_j(t) - \Delta\gamma p_i(t) + \epsilon f_i(t) \tag{4.4}$$

where the perturbation $\Delta\mathcal{L}_{ij}$ is a Laplacian matrix ($\sum_i \Delta\mathcal{L}_{ij} = 0$ and we assume $< \mathcal{L} >= 0$) that can represent an error in the measure of the transition rates $\mathcal{L}_{ij}$ or possible evolution of network due to in time. In the first case we have an ensemble of transition matrices and we have to study the eigenvalue distribution due to perturbation and the possible presence of bifurcation phenomena. In the second case we have a stochastic differential equation (since $\Delta\mathcal{L}_{ij}(t)$ can be represented as a realization of a stochastic process). The possible approach are Perturbation Theory, Random Matrix Theory and Statistical Physics Methods for random matrices. The external signal form the environment (the environmental node) can be considered in the adiabatic approximation (to be justified form a biological point of view).
The underlying stochastic process on the graph is defined by assigning the state $\xi_i(t) \in [0, 1]$ at each node i according to a probability distribution $\pi_i(t)$ that evolves as

$$\dot{\pi}_i(t) = \sum_j \mathcal{L}_{ij}\xi_j(t) - \Delta\gamma\xi_i(t) + \epsilon f_i(t)$$

By discretizing the dynamics for a time step Δt we have the evolution

$$\pi_i(t+\Delta t) = \pi(t) + \sum_j \mathcal{L}_{ij}\xi_j(t)\Delta t - \Delta\gamma\Delta t\xi_i(t) + \epsilon f_i(t)\Delta t$$

and $\xi(t+\Delta t)$ realized according to the distribution $\pi_i(t+\Delta t)$ (stochastic cellular automata). The average dynamics is computed by

$$\begin{aligned} \dot{<}\pi_i(t)> &= \sum_j \mathcal{L}_{ij} < \xi_j(t) > -\Delta\gamma < \xi_i(t) > +\epsilon f_i(t) \\ &= \sum_j \mathcal{L}_{ij} p_j(t) - \Delta\gamma p_i(t) + \epsilon f_i(t) = \dot{p}_i(t) \end{aligned}$$

and we recover the average equation (4.1). But the stochastic dynamics gives information on the applicability of the average approximation and the variability at the critical states (at bifurcation of the spectrum of $\mathcal{L}$). The stochastic dynamics can be studied for stochastic connection matrices $\mathcal{L} + \Delta\mathcal{L}$.

Chapter 5

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.

5.1 The underlying philosophy of the model

The dynamical model is a schematic representation of the activity of genetic network. 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.

5.2 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

$$\sigma_i(t+1) = \Phi_i(\sigma(t)) = \Theta\left(\sum_j A_{ij}\sigma_j(t)\right) \tag{5.1}$$

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

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

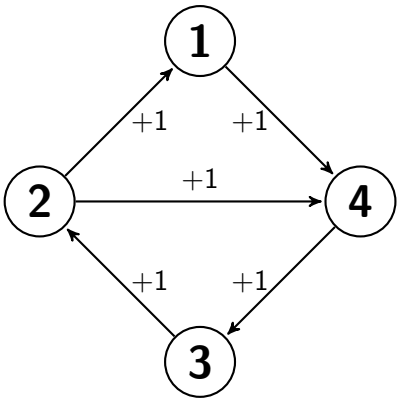
so that $A_{ij} = 0$ if $i \in S_A$ and $j \in S_{\bar{A}}$: i.e. there is not a cause-effect connection between $S_{\bar{A}}$ and S_A and the state $\sigma_i = 1$ for $i \in S_{\bar{A}}$ is an equilibrium state. Therefore we have as many equilibrium states as many partitions S_A and $S_{\bar{A}}$ there 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

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

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 $\sigma_i = 1 \to \sigma_i = 0$ 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

$$p_i(t+1) = \Theta_T\left(\sum_j A_{ij}\sigma_j(t)\right) \tag{5.2}$$

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 (5.2) 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

$$< \sigma_i > (t+1) = p_i(t+1) = \left\langle \Theta_T \left(\sum_j A_{ij}\sigma_j(t) \right) \right\rangle \simeq \Theta_T \left(\sum_j A_{ij} p_j(t) \right) \tag{5.3}$$

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

$$\bar{\Sigma}(t+1) = \sum_i p_i(t+1) = \sum_i \Theta_T \left(\sum_j A_{ij} p_j(t) \right) > \bar{\Sigma}(t) \tag{5.4}$$

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 5.1: Possible behavior for the condition (5.4); 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 (5.4) 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 (cfr. Kramer transition rate Theory). 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 (5.1) reads

$$\sigma_i(t+1) = \Theta\left(\sum_j A_{ij}(t)\sigma_j(t)\right) \tag{5.5}$$

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.

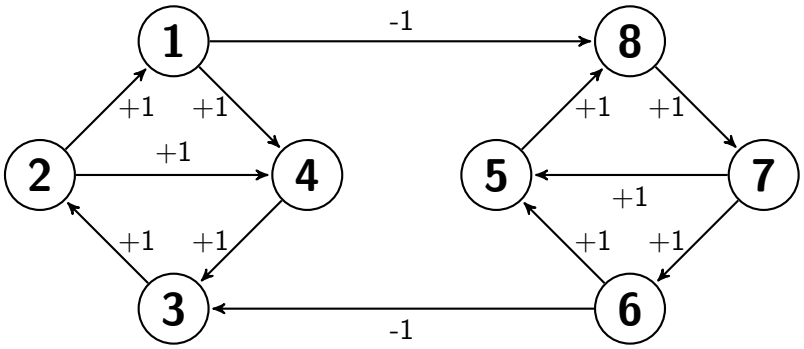
One could say that the network is active if a certain condition is satisfied (for example the average total activity should overcome a given threshold) so that fluctuation may introduce the existence of non active states. Problems for a single network: starting from a loops with a fixed dimension adds 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

$$\sigma_i(t+1) = \Theta\left(\sum_j A_{ij}(t)\sigma_j(t)\right) \tag{5.6}$$

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 the connectivity matrix (to be studied) then it is possible to study the spectral properties of random connectivity matrices (see Catanzaro thesis) and develop a control theory for the network.

Let us consider the existence of competitive networks (see figure) 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

$$\nu_k(t+\Delta t) - \nu_k(t) = \phi(\nu_k(t)) - \gamma\left(H_{kj}\nu_j(t)\right) \tag{5.7}$$

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

$$p(\sigma', t+1) = \sum_\sigma \pi(\sigma', \sigma) p(\sigma, t) \tag{5.8}$$

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 6

Analysis

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

6.1 Implementation

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

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

The first thing to evalute, was to choose the mean number of incoming links for each RBN. In Figure 6.1 we can see that the mean discrete evolution of 100 different realizations of RBNs, with increasing size. In the case of $K = 1$, i.e. the mean number of incoming links for each network is one, the mean activity decreases exponentially with the size of the network; in the case of $K = 2$ instead, the mean activity of the nodes remains stable with networks of increasing size.

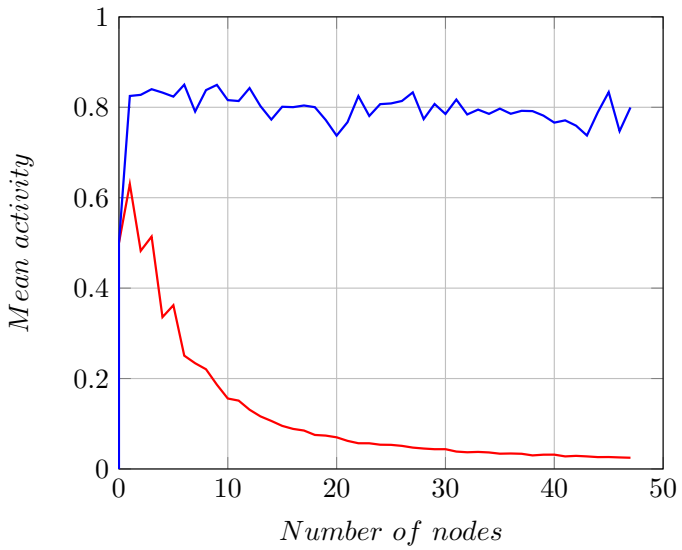


Figure 6.1: *Plot of the mean activity of the nodes with network of increasing size. In the case of $K = 1$ (i.e. the mean number of incoming link for each network is one), the mean activity decreases exponentially with the size of the network; in the case of $K = 2$ instead, the mean activity of the nodes remains stable with the network size.*

6.2 Discrete evolution

As shown in Chapter 5, 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 means 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 mean activity of the network, which is the mean number of nodes with the value

$$\sigma_i(t) = 1$$

6.3 Control nodes

6.4 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 infer 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 propbability p for the node or for the link to be turned off. In Figure 6.2 we can see the behaviour of the mean activity of the network depending on the amount of noise added.

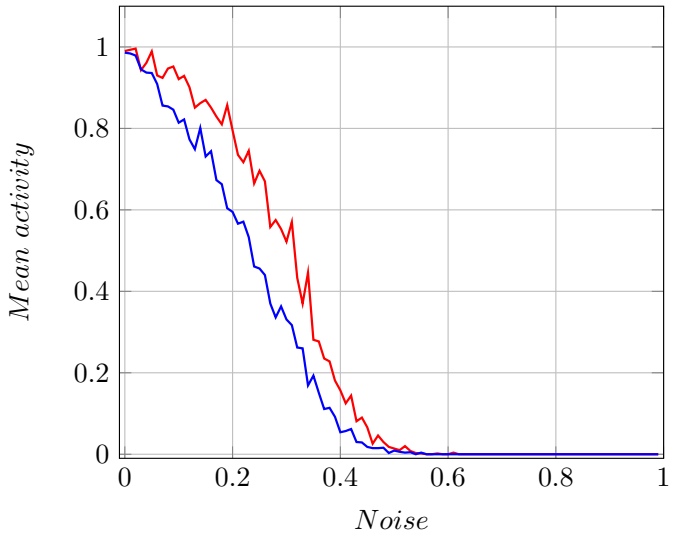


Figure 6.2: *Plot of the effect of the noise on the mean activity on the network. In blue the noise works on the nodes on the network, while in red the noise works on the links.*

Conclusions

Bibliography

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

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

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

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

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

[6] 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: 185193, (2010)

[7] 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)

[8] S. Kauffman, *A proposal for using the ensemble approach to understand genetic regulatory networks*,Journal of Theoretical Biology 230 (2004) 581590 ,(2004)

[9] 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)

[10] 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)

[11] 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)

[12] 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)

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

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

[15] 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)

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

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

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

[19] R. V. Sol, B. Loque, *Phase transitions and antichaos in generalized Kauffman networs*,Physics Letters,(1994)

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

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

[22] A. Rka and A-L. Barabsi, *Statistical mechanics of complex networks*, Reviewes of modern physics, Volume 74,(2002)

[23] N. Masuda , M. A. Porter, R. Lambiotte ,*Random walks and diffusion on networks*,Physics Reports 716717 158,(2017)

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

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

[26] 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)

[27] 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)

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

[29] 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)

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

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

[32] Christopher H. Joyner and Uzy Smilansky, *Dysons Brownian-motion model for random matrix theory - revisited*,arXiv:1503.06417,(2015)

[33] 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)

[34] Xin Kang, Chunhe Li,*Landscape inferred from gene expression data governs pluripotency in embryonic stem cells*,omputational and Structural Biotechnology Journal 18 (2020) 366374,(2020)

[35] Genaro J. Martnez , 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)

[36] Jifan Shi ID , Tiejun Li , Luonan Chen, Kazuyuki Aihara,*Quantifying pluripotency landscape of cell differentiation from scRNA-seq data by continuous birth-death process*, (2019)

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