Ph.D. Thesis

Department of Computer Science
University of Pisa

Verification of Robustness Property in Chemical Reaction Networks

Lucia Nasti

Supervisor
Paolo Milazzo

Supervisor
Roberta Gori

Anno Accademico 2019/2020

Reviewer
Prof. Maciej Koutny
Newcastle University, School of Computing
Newcastle upon Tyne, United Kingdom

Reviewer
Dr. Christoph Zechner
Max Planck Institute of Molecular Cell Biology and Genetics
Dresden, Germany

Quando cade,
solo allora galleggia
il fiore del loto.

— Yahamoto Yosaburō

There is no learning without having to pose a question.
And a question requires doubt.

— Richard Feynman

To my family.
To my friends.
To my love, Ottavio.

ABSTRACT

Cells are very complex to analyze because they consist of many components that interact with each other, producing multiple sequences of chemical reactions, which regulate their behavior. Besides, several fluctuations can alter the cell functionalities, such as internal errors propagation, and variations in the concentrations of chemical species.

Regarding this, a significant biological property is *robustness*, defined as the observed capacity of the system to maintain its functionalities despite the presence of internal or external perturbations. In nature, different mechanisms ensure this property, such as redundancy, modularity, system control, and structural stability. Redundancy and modularity represent the architectural characteristics of a biological pathway, resulting in multiple copies of structures and compartments, which, having the same functions, avoid the possible presence of errors and failures. Conversely, system control and structural stability are dynamical properties, expressed as the capacity to adapt to environmental changes.

In this context, Computer Science can help research in Biology in many different ways. Through simulations, for instance, it is possible to mimic the internal dynamics of a natural system and to predict its functions. Moreover, model-based analysis techniques can be used to interpret some non-intuitive aspects of a biological system.

In this thesis, we propose a new definition to formally describe a specific notion of robustness, the *initial concentration robustness*. This has the purpose of analyzing how perturbations in the initial concentrations of the involved chemical species (identified as inputs) can alter the system behavior at the steady state. Therefore, we developed a theoretical framework, based on Petri net formalism, and we applied it to different known biological networks available in the literature.

To understand the behavior of a biological system, we should simulate it considering all the possible combinations of the initial values, which implies a huge computational effort. To face this issue, we found a sufficient condition that allows knowing if the concentration of an output species is monotonic

concerning the concentration of an input species (which is the perturbed substance). By monotonic, we mean that increasing (or decreasing) the concentration of the input, the concentration of the output, at each time step, increases (or decreases) consequently. If the sufficient condition is met, we can drastically reduce the number of simulations, testing the model only on the extreme values of the input concentration range.

Finally, we apply our theoretical framework to the case study of Becker-Döring equations, a model which describes the aggregation kinetics of particle clusters. In particular, we study the robustness of this model using the proposed robustness formalism as well as other analytical approaches.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LISTINGS

ACRONYMS

BD Becker-Döring

CRN Chemical Reaction Network

LTL Linear Temporal Logic

PN Petri Net

QFLTL Quantifier-Free Linear Temporal Logic

sd Satisfaction Degree

vd Violation Degree

SRD Standard Reaction Diagram

1 INTRODUCTION

1.1 MOTIVATION

From the discovery of DNA structure, in 1953, there has been a growing interest in the investigation of living cells to understand their morphological and functional organization. Cells are very complex to analyze because they consist of many components that interact with each other, through multiple sequences of chemical reactions which regulate the cell behavior. Besides, several fluctuations can alter their functionalities, such as internal errors propagation, and variations in the concentrations of chemical species.

Regarding this, a significant biological property is *robustness*, defined as the observed capacity of a cell to maintain its functionalities despite the presence of internal or external perturbations. In nature, different mechanisms ensure this property, such as redundancy, modularity, system control, and structural stability [44]. Redundancy and modularity represent the architectural characteristics of a biological pathway, resulting in multiple copies of structures and compartments, which, having the same functions, compensate the possible presence of errors and failures. Conversely, system control and structural stability are dynamical properties, expressed as the capacity to adapt to environmental changes.

We can distinguish three fundamentals and complementary approaches to investigate robustness: *in vivo*, *in vitro* and *in silico*. The first two deal with the branch of methods in which organisms, tissue and cells are studied by direct observation. The last one, instead, deals with all the methods and procedures that abstract the functionalities and dynamical properties of biological networks through the implementation of algorithms and computational models.

In this context, Computer Science can help research in Biology in many different ways. Through simulations, for instance, it is possible to mimic the internal dynamics of a natural system and to predict its functions. Moreover,

model-based analysis techniques can be used to interpret some non-intuitive aspects of a biological system.

Computer scientists developed many formalisms to study systems of interacting components, which can describe new and unexpected issues when applied to a biological context. These formalisms, having specific applications for Systems biology [15], include Petri nets [14, 46, 52], Hybrid systems [3, 39, 48], process calculi such as the π-calculus [62], the Bio-ambients calculus [63], Bio-PEPA [22] and Beta-binders [60]. Other formalisms are based on rewrite rules such as CLS [9], and the k-calculus [23], specifically designed for describing protein interactions. In particular, Hybrid automata represent a dynamical system combining its discrete and continuous aspects, as discussed in [50] where they are used to study an optimal strategy for prostate cancer treatment. Another example of application of Hybrid automata is [54, 55], where we developed a model of the Dopamine System, a neurological component involved in the development of addiction. Petri nets focus on the representation of the network topology and its structural properties, exploring in a general manner, observable phenomena such as steady states or oscillations. Using Petri nets, it is possible to understand in detail the role of single chemical species involved in reactions, as described in [46].

On the other hand, other formalisms were originally defined taking inspiration from the characteristics of cells and their interactions, as P Systems [56, 58], a model for membrane computing, and Reactions Systems [27] which model interactions between biochemical reactions in a rewriting-based formalism. In particular, P Systems are able to represent parallel and distributed systems, considering both the connection between the components of a system and its environment. The work in [58] contains an example of application: in this case, the authors divide the chemical processes into four pathways taking place in three separate regions (environment, cell surface, cytoplasm). This allows describing and reasoning on the chemical reaction network in more manageable way.

From this brief discussion, it emerges that all the cited formalisms have peculiar characteristics which can describe particular features of the explored systems. In this thesis, to verify the robustness in chemical reaction networks we choose the Petri nets formalism for several reasons. In particular, because Petri nets allow us to have a high level of abstraction. This is relevant since robustness is an intrinsic and structural property, as shown in Feinberg's works

[30, 31]. Then, using Petri nets, we can obtain a simple and intuitive representation of the chemical pathways. Indeed, this is not always the case when applying other formal languages, since we would need to specify additional and non-necessary features of the network, making difficult their interpretation.

1.2 CONTRIBUTIONS

The work in this thesis can be divided into three fundamental contributions concerning the verification of robustness property.

The first contribution concerns the formal definition of a specific formulation of robustness, namely *initial concentration robustness*, in which perturbations on the initial concentrations of chemical species can influence the steady state of the system. We propose a new notion of α-robustness, based on continuous Petri nets [33] and interval markings, which extends the notion of *absolute concentration robustness* considered in [70, 71]. More in detail, we identify one or more reactants as the input of the network, and we assign to each of them an interval of initial concentrations. We consider the state when the initial concentration can vary. Then, our definition is satisfied if, at the steady state, the concentration of the designed output is in an interval, whose range is α. Intuitively speaking, α represents an absolute metric to express the extent of the change of the output at the equilibrium. Regarding this, we can observe that:

- the wider are the intervals of the initial interval marking, the more robust is the network. Indeed, it means that the system is able to absorb a higher level of perturbations.
- the smaller is the value of α, the more robust is the system.

As an extension of the concept of α-robustness, we introduce the concept of *relative robustness*, or β-robustness, both to express how robust the system is considering the intensity of perturbations and to compare different systems. We observe both concepts by performing simulations of known biological networks available in the literature.

As already mentioned, the robustness property can occur in different ways within biological systems. Therefore, in literature, there are several works, as

[45, 65], describing it with different approaches. In particular, the framework proposed by Rizk et al. [65] appears very general and able to treat different kinds of robustness. Regarding their proposal, in this thesis we prove that our notion of α-robustness is a specific instance of their definition. Nevertheless, we choose to consider our new definition because we need to abstract the structure of the chemical reaction network efficiently and to identify the input and the output distinctly. In this way, indeed, we can study if there exist structural correlations between the two selected species, which can help us to have more information about the internal dynamics of the system.

In general, the verification of robustness requires a huge number of simulations [70], since we need to test the system behavior for all the possible combinations of initial concentrations of the involved chemical species. Then, to reduce the computational effort and to make our definition effective, we search for sufficient conditions that allows us to analytically infer our robustness property. The sufficient condition relays on the concept of monotonicity of the system.

In this regard, we introduce the second contribution of the thesis. By the extension of the work done by Angeli et al. [4], we derive a graphical approach that can predict if the concentration of the output is monotonic with respect to the initial concentration of the input. If this sufficient condition is verified, in order to assets robustness we can simulate the model only on the extreme values of the input concentration interval, reducing the number of simulations drastically. Then, we verify our sufficient condition of Input-Output monotonicity on two case-studies: the Michaelis-Menten kinetics and the ERK signaling pathway. In particular, these chemical reaction networks underline how to investigate the model dynamics, without making assumptions on its differential equations but examining only its structure.

The third contribution of this thesis concerns the verification of robustness property in Becker-Döring equations, which consist of two main processes: the *aggregation* and the *fragmentation* of clusters of particles. The applications of this kinetic model are multiple and include many fields. Concerning biology, it is employed to study protein aggregation, polymerization, and concentration gradients.

Representing this model as a Petri net, we identify the monomers (C_1) as the input and the output of the network. Indeed, C_1 is the only cluster present at the initial state of the system, and it is the cluster involved in every aggregation

and fragmentation process. By its representation, we find out that we cannot directly apply the methodology we proposed. So, we focus on avoiding the need of many simulations to study the concentration of C_1 at the steady state. We prove analytically that this concentration tends to $\frac{b}{a}$, where b and a are the coefficients of the rates of fragmentation and agglomeration. Then, we apply known results [69], to show how this model can be sensitive to perturbations on the initial concentration.

1.3 STRUCTURE OF THE THESIS

The thesis is organised in 5 chapters, plus introduction and conclusion.

- Chapter 2 provides some background concepts useful for understanding the rest of the thesis. The Chapter is divided into three main sections: the first one focuses on the chemical reaction network (CRN) and the chemical kinetics, the other sections focuses on two formalisms, Petri nets and Linear Temporal Logic (LTL). In particular, in Section 2.1 we focus on the representation of chemical reactions, considering the two main methods used to describe them: the *deterministic* and *stochastic approach*. Moreover, the Section gives an overview of the steady state calculation. Section 2.2 recalls some definitions of Petri nets, which we choose as formal language to give the definition of robustness property. Section 2.3.1 presents some background concepts on LTL, which is the formalism used in [64] and in Chapter 3.

- Chapter 3 presents the state of the art in the verification of robustness properties in CRNs. In particular, Section 3.1 focuses on the role of *local* and *global* sensitivity analysis in the study of biological networks. In the second part of this Chapter, we focus on the definition of robustness property, with a particular emphasis on three approaches. Section 3.2 focuses on the general definition of robustness given by Kitano [45], Section 3.3.1 depicts the work done by Rizk et al. [65], which extends the cited previous work. Finally, in Section 3.4, we delineate the work done by Shinar and Feinberg [30, 31, 70].

- Chapter 4 provides a new formal definition of robustness against perturbations to the initial concentrations of species, namely *absolute initial*

concentration robustness, based on Petri nets. In Section 4.2, we demonstrate the validity of our definition by applying it to the models of four different robust biochemical networks: the EnvZ/OmpR osmoregulatory signalling system 4.2.1, the isocitrate dehydrogenase regulatory system 4.2.2, the bacterial chemotaxis 4.2.3 and a model of the enzyme activity at saturation 4.2.4. Finally, in Section 4.3, we prove that our definition of *absolute initial concentration robustness* is an instance of the general Definition 12 given by Rizk et al. and described into detail in Chapter 3.

- Chapter 5 is divided in two parts. The first part focuses on the monotonicity in CRNs, showing the works presented in [4, 5, 24]. The second part contains one of the main contributions of the thesis. In Section 5.3, we give our notion of monotonicity between an input and an output species. Then, in Section 5.2, we present our Input-Output monotonicity theorem, able to reduce the computational cost of simulations. Finally, in Section 5.4 we apply our theorem on two examples: Michaelis-Menten kinetics and ERK signaling pathway. The source code implemented to verify the sufficient conditions on graphs is freely available online [75].

- Chapter 6 presents the Becker-Döring equations. In the first part, we introduce in detail the mathematical properties of this system, showing the analytical solution of the steady state. In the second part, we apply the *Deficiency Zero Theorem* [30], described in Section 3.4 of Chapter 3, to verify the robustness property of this model. The source code implemented to simulate the model is freely available online [74].

- Finally, Chapter 7 presents the conclusion of the thesis and depicts possible future works.

1.4 PUBLISHED MATERIAL

Part of the material presented in this thesis has appeared in some publication, in particular:

- the formalization of absolute initial concentration robustness, presented in Chapter 4, has appeared in [53];

- the notion of Input-Output monotonicity, presented in Section 5.2, has appeared in [35].

Other original contributions of this thesis will be submitted for publications in extended form, in particular:

- the definition of relative robustness, namely β-robustness, presented in Section 4.1.1, and the proof of correspondence between Definition 16 and Definition 12, presented in Section 4.3, will be submitted as part of a journal extension of the work [53].

- Our sufficient condition 1, presented in Section 5.2 and the examples of application, presented in Section 5.4, will be submitted as journal contribution with the title:
  - Roberta Gori, Paolo Milazzo, Lucia Nasti and Federico Poloni. Efficient analysis of Chemical Reaction Networks Dynamics based on Input-Output monotonicity. (2019)

- The work and the results presented in Chapter 6 will be submitted in extend form with the title:
  - Verification and analysis of Robustness in Becker-Döring equations. (2019)

2

BACKGROUND AND NOTIONS

In this chapter we introduce some notions that will be assumed in the rest of the thesis. In the first part, we focus on the representation of chemical reactions, considering the two main methods that we can use to describe them: the *deterministic* and *stochastic approach*. Then, we concentrate on the steady state calculation. In the second part, we show two formalisms, useful to describe the dynamics of chemical reaction networks: Petri nets and Linear temporal logic.

2.1 CHEMICAL REACTIONS

A chemical reaction is a transformation that involves one or more chemical species, in a specific situation of volume and temperature.

The chemical elements that are transformed are called *reactants*; on the other hand, those that are the result of the transformation are called *products*. We can represent a chemical reaction as an equation, showing all the species involved in the process.

A simple example of chemical reaction is the following elementary reaction:

$$a\mathrm{A} + b\mathrm{B} \xrightarrow{k_1} c\mathrm{C} + d\mathrm{D} \tag{1}$$

In this case, A, B, C, D are the species involved in the process: A and B are the reactants, C and D are the products. The parameters a, b, c, d are called *stoichiometric coefficients* and represent the multiplicities of reactants and products participating in the reaction. They are always integers, because elementary reactions involve whole molecules.

The arrow is used to indicate the direction in which a chemical reaction takes place. When we have only one arrow, as in Example (1), it means that the reaction is *irreversible*, which means that it is not possible to have the opposite process.

The symbol k_1, referred to as *kinetic rate*, is a real positive number representing the rate at which the process occurs and has to be associated to a specific arrow. Intuitively speaking, it represents the speed of a reaction, that is how quickly the reaction will take place.

In the case there are two arrows in opposite directions, this means that it is possible to have both the direct and inverse transformations. Considering the following reactions:

$$a\mathrm{A} + b\mathrm{B} \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} c\mathrm{C} + d\mathrm{D} \tag{2}$$

with the same notation, the symbol k_{-1} is the kinetic rate for the inverse reaction, where the reactants and products are exchanged. Reaction (2) is then called *reversible*.

To describe chemical reactions inside a system, we have to define the dynamical principles governing the interaction among reactants and their products. These are formalized within the kinetic theory of reactions.

2.1.1 *Chemical kinetics*

Given a closed pot and a well-stirred mixture of chemical species, which we can call *composition*, the chemical kinetics helps to investigate the temporal evolution of all the species involved. Indeed, the mixture will not remain constant in time, since there will be some of the species produced and others consumed [29, 31]. Therefore, we can study how the *concentration* of elements changes during time. We can define the chemical concentration as the number of particles of a specific species per unit volume of the mixture. Conventionally, the concentration of a chemical species is represented by the symbol denoting the element, written between square brackets, as $[A]$.

The *rate* of a reaction measures the change in concentration of the chemical species per unit time. Mathematically, we can express the rate r as follows:

$$r = \frac{\Delta \mathtt{Concentration}}{\Delta \mathtt{Time}}$$

where $\Delta \mathtt{Concentration}$ represents how much the concentration of the species increases or decreases on the considered range of time $\Delta \mathtt{Time}$. Since the con-

centrations of reactants decrease during the time, a minus precedes their rates, while the concentrations of products increase. To get into details, we can consider a certain *closed* system, in which the reaction (2) can occur. Here, by closed we mean that the system is not allowed to exchange energy and particles with the external environment. It follows that the chemical rates of each species involved in the reaction (2) are:

$$r_A = -\frac{\Delta[A]}{\Delta Time} \qquad r_B = -\frac{\Delta[B]}{\Delta Time} \qquad r_C = +\frac{\Delta[C]}{\Delta Time} \qquad r_D = +\frac{\Delta[D]}{\Delta Time}$$

To study the relation between reaction rates and molecular component concentrations, we can use the empirical *law of mass action*, which provides a complete picture of the component concentrations at all future time points. The law of mass action states that:

Definition 1. *The rate of a reaction is proportional to the product of the concentrations of the individual reactants involved.*

Referring again to (2), applying the law of mass action, we obtain:

$$r_1 = k_1[A]^a[B]^b \qquad\qquad r_2 = k_{-1}[C]^c[D]^d$$

considering in the first rate r_1 the reaction in which A and B are the reactants, and for the rate r_2 the opposite transformations, in which C and D are the reactants.

In general, after a sufficiently long amount of time, particular kinds of systems can reach a chemical equilibrium. This is properly formalized within the so-called *collision theory* [21], that is, given a system, the particles inside it are constantly moving and hit each other with an average uniform distribution, causing chemical changes. In agreement to that, a macroscopic equilibrium is, from a microscopic point of view, a dynamical equilibrium, that is, if both direct and inverse processes are allowed, we have equilibrium when the processes occur at the same rate.

For what concerns Example (2), we have:

$$k_1[A]^a[B]^b = k_{-1}[C]^c[D]^d \quad \Rightarrow \quad \frac{[C]^c[D]^d}{[A]^a[B]^b} = \frac{k_1}{k_{-1}}.$$

2.1.2 *Deterministic and Stochastic Approaches*

From the computational point of view, a chemical system can be essentially described by two mathematical methods [11]:

- the deterministic approach,
- the stochastic approach.

The deterministic approach is the simplest method to solve a chemical model and it involves the application of Ordinary Differential Equations (ODEs).

We can use an ODE to describe how the variation of the concentration of chemical species changes over time. Hence, in a deterministic point of view, we have a differential equation for each species involved in chemical transformations. Each differential equation is written applying the law of mass action: the concentration variation of the chemical species is determined by the rates of the reactions in which the species appears as product minus the rates of the reactions in which it appears as reactant.

Thus, considering the reaction (2), we obtain:

$$\begin{aligned}
\frac{d[A]}{dt} &= \overbrace{-ak_1[A]^a[B]^b}^{\text{direct reaction term}} \overbrace{+ak_{-1}[C]^c[D]^d}^{\text{inverse reaction term}} \\
\frac{d[B]}{dt} &= -bk_1[A]^a[B]^b + bk_{-1}[C]^c[D]^d \\
\frac{d[C]}{dt} &= +ck_1[A]^a[B]^b - ck_{-1}[C]^c[D]^d \\
\frac{d[D]}{dt} &= +dk_1[A]^a[B]^b - dk_{-1}[C]^c[D]^d.
\end{aligned}$$

Since in the system under study, there are four different species A, B, C and D, we have four differential equations. In each one, we isolated the term describing the direct reaction (in which A and B are the reactants and C and D are the products) from the one describing the inverse reaction (in which C and D are the reactants and A and B are the products). Roughly speaking, if the species appears as reactant its reaction rate will be preceded by a minus (as we saw in details in Section 2.1.1), otherwise if it appears as product its reaction rate will be preceded by a plus. Hence, we implicitly considered, for each element, its process of destruction and production.

Through the deterministic approach, we can analyze the time evolution of the model as a *continuum*. This method has the advantage of requiring a small amount of time, due to a modest computational effort. Nevertheless, the differential equations have to be solved numerically on the computer and some physical aspects of how reactions occur are approximated.

To overcome this problem, the stochastic approach has been developed to involve the results coming from collision theory. Indeed, reactants collide with a certain probability distribution, which increases as their concentration increases. Therefore, the difference from the previous way of thinking is the introduction of a probability function, that ultimately implies:

- reactions do not occur necessarily in the same infinitesimal time interval;
- it is possible to have a time interval without a certain process occurring.

Explicitly, considering a closed system of fixed volume, in which N chemical species can interact through M possible reactions R_μ, we assume the existence of M constants $c_\mu dt$, representing the average probability to have a reaction at the infinitesimal time interval dt.

In this perspective, we focus the attention on the Stochastic Simulation Approach, a method to study how a system changes randomly, according to some given probabilities. In this regard, one of the most common approaches is the well-known Gillespie's Algorithm [34]. Gillespie's Algorithm is formalized on the basis of the collision theory and the main questions it aims to answer are:

1. *Which is the next reaction?*
2. *When does it occur?*

To answer them, the final purpose is to calculate the *Reaction Probability Density Function* (PDF) $P(\tau, \mu)$. Assuming that in a chemical solution, $S_1, S_2 ..., S_n$ are the only molecules present, a *state* of the under study system is the tuple $(X_1, X_2, ..., X_n)$ representing a solution containing X_i molecules of S_i for each i in $1, ..., n$. The PDF is defined as follows:

Definition 2. *Given a fixed volume* V*, a set of species* $S = \{S_1, S_2, ..., S_n\}$*, and a set of reactions* $R_\mu = \{R_1, R_2, ..., R_M\}$*, the* $P(\tau, \mu)$ *is the probability that, given a state* $(X_1, ..., X_n)$ *at a time* t*, the* next *reaction in the system of volume* V *will occur in the infinitesimal time interval* $[t + \tau, t + \tau + d\tau]$ *and will be reaction* R_μ*.*

Even these two approaches appear as very different, they are deeply *complementary* because they analyze distinct features of a biological system. Indeed, the intrinsic system traits determine the reason why we choose one approach rather than the other.

Generally, a deterministic differential equation models well describe a network characterized by a large number of simultaneous events. In such cases, the network behavior corresponds to the average of all these events. On the other hand, a stochastic approach is well suited to describe a system in which random effects become more pronounced and are not well expressed by the deterministic model.

To remark the differences between the two approaches, we proceed by comparing the two methods by simulations. Consider the chemical reaction network in Example 3, inspired by the Example 29 in [34]:

$$\begin{aligned} X + Y &\xrightarrow{k_1} 2Y + X \\ 2Y &\xrightarrow{k_2} Z \end{aligned} \tag{3}$$

whose differential equations describing its behavior are the following:

$$\begin{cases} \frac{d[X]}{dt} = 0 \\ \frac{d[Y]}{dt} = k_1[X][Y] - k_2[Y]^2 \\ \frac{d[Z]}{dt} = k_2[Y]^2 \end{cases} \tag{4}$$

We apply Gillespie's Algorithm to study the random fluctuation of species Y, given these initial conditions: $Y = 3000$, $X = 10$, $Z = 0$, $k_1 = 0.5$, and $k_2 = 0.005$. In Figure 1, we show the simulation result. Instead, in Figure 2, we simulate the system in Example 3, solving numerically its differential equations 4.

As we can notice, comparing Figure 1 and Figure 2, the simulations agree, confirming that for large concentrations of reactants, individual decay events have negligible effects on the overall behavior. In this context, using a deterministic approach is convenient because we can study the asymptotic behavior of the system, neglecting those aspects that are not prominent for its understanding. In particular, in this perspective, using a deterministic model facilitates its

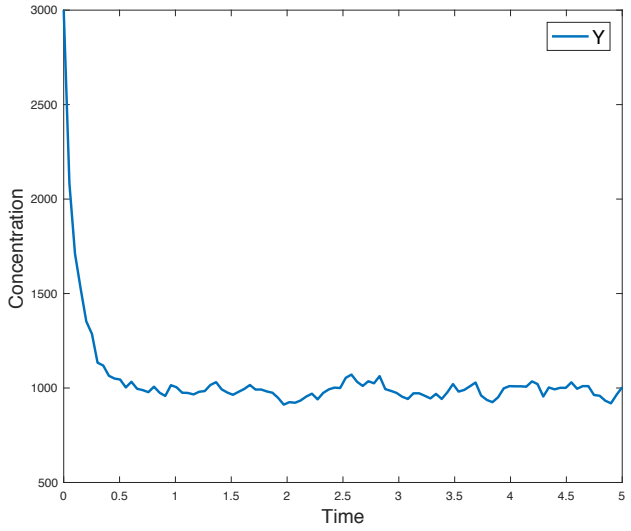

Figure 1: Simulation result of Example 3. In this case, we apply the Gillespie's Algorithm to study the random fluctuation of chemical species Y. The set of parameters are: $Y = 3000$, $X = 10$, $Z = 0$, $k_1 = 0.5$ and $k_2 = 0.005$.

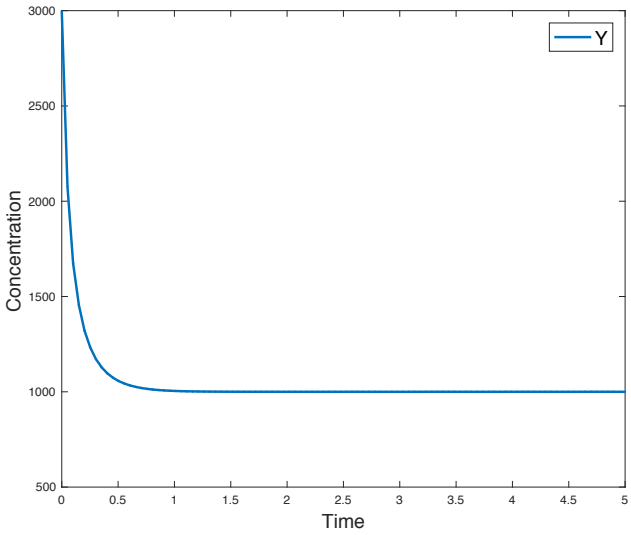

Figure 2: Simulation result of Example 3. In this case, we apply a deterministic approach to study how the concentration variation of species Y changes over time. The set of parameters are: $Y = 3000$, $X = 10$, $Z = 0$, $k_1 = 0.5$ and $k_2 = \frac{0.005}{2}$.

analytical study, as we will see for the steady state calculation in the following Section 2.1.4.

2.1.3 *Chemical Reaction Networks definition*

A Chemical Reaction Network (CRN) is a set of reactions. Formally, it is defined as:

Definition 3 (Chemical reaction network)**.** *Let $\mathcal{R}_i$ be a list of reactions, where the index i takes value in $\mathcal{R} = \{1, 2, ..., n_r\}$, and S_j be the set of the chemical elements involved in the network, where $j \in \{1, 2, ..., n_s\}$, representing the species taking part in the reactions. Then, the chemical reactions are denoted as follows:*

$$\mathcal{R}_i\colon \sum_{j\in\mathcal{S}} \alpha_{ij} S_j \rightleftharpoons \sum_{j\in\mathcal{S}} \beta_{ij} S_j$$

where α_{ij} and β_{ij} are nonnegative integers called the stoichiometry coefficients.

The stoichiometry coefficient matrix is defined as:

$$[\Gamma]_{ij} = \beta_{ij} - \alpha_{ij}$$

for all $i \in \mathcal{R}$ and all $j \in \mathcal{S}$. Based on these definitions, the rate of a reaction is expressed as a function of the concentrations of the species taking part to the chemical transformations. The vector of species concentrations is defined as $S = [S_1, S_2, ..., S_{n_s}]'$, which describes the concentration variation during time, and as a function of it, the vector of reaction rates is defined as $R(S) = [R_1(S), R_2(S), ..., R_n(S)]'$.

Systems are described by differential equations, in a deterministic approach:

$$\frac{dS}{dt} = \Gamma \cdot R(S),\ S \in \mathbb{R}^{n_s}_{\geqslant 0}.$$

where $\frac{dS}{dt}$ represents the concentration variation of species $s \in S$, Γ is the matrix of stoichiometric coefficients and $R(S)$ is the vector of reaction rate, given by concentration of species times constant rate.

2.1.4 *Steady State of Chemical Reaction Network*

A chemical reaction network can reach a steady state condition when the rates of decay are equal to the rates of production of each species involved in the network. In general, almost all biological models have a steady state behavior, even if some systems display sustained oscillations. To find the steady state,

it depends on the complexity of the biological model. Consider the following open system, characterized by two reactions:

$$\xrightarrow{k_1} A \xrightarrow{k_2} \tag{5}$$

The system will reach a steady state when the difference between the reaction rates of the system represented in Example 5 is equal to 0:

rate of change of species $[A]$ = rate of production of A − rate of decay of A

As described in Section 2.1.2, we express the rate of change of A as follows:

$$\frac{d[A]}{dt} = k_1 - k_2[A] \tag{6}$$

The steady state is mathematically expressed by setting the derivative of A with respect to time to zero, then:

$$\begin{aligned} &\frac{d[A]}{dt} = 0 \\ &k_1 - k_2[A] = 0 \\ &k_1 = k_2[A] \end{aligned}$$

with $A = 0$, $k_1 = 1$ and $k_2 = 0.5$, we proceed by substitution and we obtain:

$$\begin{aligned} &k_1 = k_2[A] \\ &[A] = \frac{k_1}{k_2} \\ &[A] = \frac{1}{0.5} = 2 \end{aligned} \tag{7}$$

The result obtained in Formula (7), it is also confirmed by the simulation, represented in Figure 3.

Considering Figure 3, we can identify two particular periods: the *transient* and the *steady state* periods. We talk about *transient period* to identify the time-course that leads from the initial state to the long-time behavior. It displays the immediate feedback to a perturbation [41]. Instead, we talk about *steady state*, to describe when, after a certain transient period, the concentrations of a

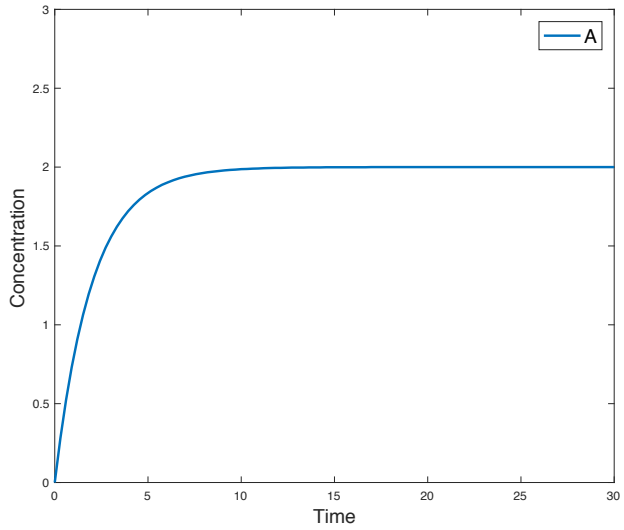

Figure 3: Simulation result of Example 5. The set of parameters are: $A = 0$, $k_1 = 1$ and $k_2 = 0.5$.

chemical reaction network remain apparently unchanged [11]. In this case, at the microscopic level, chemical transformations continue, hence the particles continue to be expressed and decay; however, at the macroscopic level, the changes are not evident in the system behavior, because the particles number no longer changes. In Figure 4, we identify the transient and the steady state periods on the plot. In the rest of the thesis, with the notation $[\cdot]_{ss}$, we will identify the concentration of a species at the steady state.

To study how long it takes before this asymptotic behavior is reached depends on the parameters and the initial conditions. Setting different rates and initial conditions, indeed, the system could reach another steady state or the same steady state but at different time. As example, consider the simulation result in Figure 5. In this case, the set of parameters are: $A = 0$, $k_1 = 10$, $k_2 = 0.5$. By substitution, in Formula (7), we obtain:

$$\begin{aligned} k_1 &= k_2[A] \\ [A] &= \frac{k_1}{k_2} \\ [A] &= \frac{10}{0.5} = 20 \end{aligned} \tag{8}$$

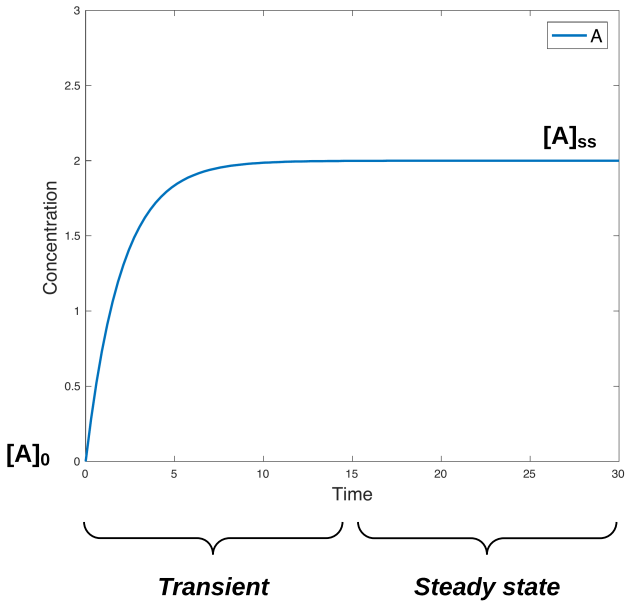

Figure 4: Simulation result of Example 5, in which we show the *transient* phase and the *steady state* phase. As for simulation in Figure 3, the set of parameters are: $A = 0$, $k_1 = 1$ and $k_2 = 0.5$. With the notation $[A]_0$ we identify the initial concentration of the species A, instead with the notation $[A]_{ss}$ we identify the concentration of the species A at the steady state.

Thus, the system reaches another steady state, as shown in Figure 5. The most significant theorems to know *a priori* if a chemical reaction network exhibits one or multiple steady states are the *Zero Deficiency Theorem* and *One Deficiency Theorem* [30], which we will show in details in Sections 3.4.1 and 3.4.2.

2.2 PETRI NETS

In literature there are many mathematical formalisms from Computer Science, that can be used to describe a biological system at different complexity levels.

In this thesis, we consider Petri nets, because this method has the advantage to provide a graphical support, which abstracts further the system description. As demonstrated by many biologists, in fact, graphical representations are often used to provide intuition regarding the network main features [16].

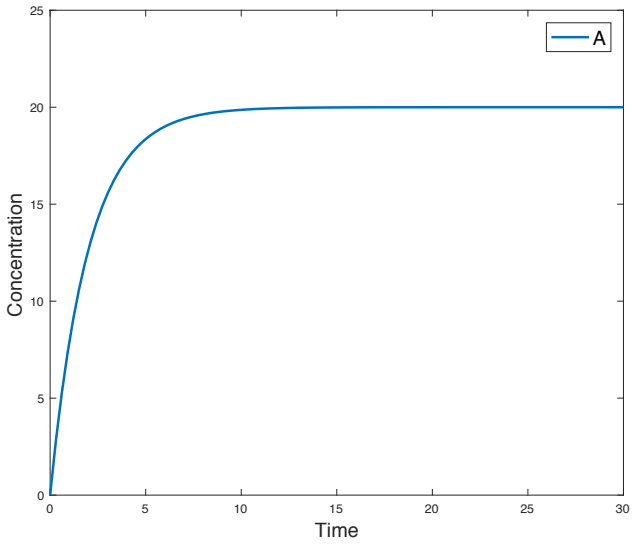

Figure 5: Simulation result of Example 5, in which the set of parameters are: $A = 0$, $k_1 = 10$ and $k_2 = 0.5$.

2.2.1 *Petri nets formalism definition*

The Petri nets are a formal language to model concurrent systems. This method, developed by Carl Adam Petri in 1962, has many applications in different areas, since it is used to model static and dynamic behavioural aspects, it is a valid tool to study communication protocols, decision models, neural networks, concurrent and parallel programs [52].

Formally, we can define a Petri net N as a quintuple:

$$N =< P, T, F, W, m_0 >$$

where:

- P is the set of *places*, conceptually one for each considered kind of system resource;
- T is the set of *transitions* that consume and produce resources;
- $F \subseteq (P \times T) \bigcup (T \times P)$ is the set of arcs connecting places with transitions;
- $W : F \to \mathbb{N}$ is a function that defines the *weight* of each arc, representing the number of resources of the same kind are needed for the transition to be fired;

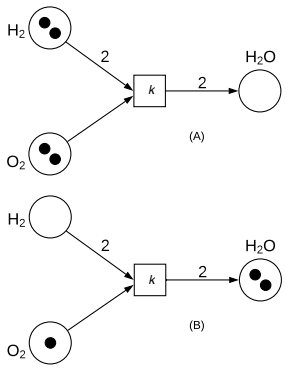

Figure 6: Example of use of Petri net. In this case, it shows how to represent the chemical reaction: $2\,H_2O + O_2 \longrightarrow 2\,H_2O$. (A) and (B) represent two different markings for the same Petri net. The marking in (B) is obtained from the one in (A) as the result of firing transition t.

- m_0 is the *initial marking*, that is the initial distribution of *tokens* (representing resource instances) among places. A marking is defined formally as $m : P \to \mathbb{N}$, with $m \in M$, where M is domain of all markings.

The tokens are movable objects, assigned to places, that are consumed by transitions in the input places and produced in the output places. Graphically, a Petri net is drawn as a graph with nodes representing places and transitions. Circles are used for places and rectangles for transitions. Tokens are drawn as black dots inside places. Graph edges represent arcs and are labeled with their weights. For simplicity, the labels of arcs with weight 1 is omitted.

Figure 6 shows a simple example of Petri net modeling the chemical reaction $2\,H_2O + O_2 \longrightarrow 2\,H_2O$ taken from [52]. In sub-figure (A), each place, H and O, has two tokens: the transition is enabled since it requires two tokens from H_2 and only one from O_2. Sub-figure (B) shows the situation after the transition has been fired: the tokens are moved to the output places. Note that in (B) the transition is no longer enabled.

To faithfully model biochemical networks, we use continuous Petri nets [33], in which the marking of a place is no longer an integer, but a positive real number, called token value, representing the concentration of chemical species. To each transition is associated a chemical rate, which represents a continuous flow. Continuous Petri nets can be defined as follows:

$$N =< P, T, F, C, m_0 >$$

where

- P and T are respectively the sets of *places* and *transitions*;
- $F : (P \times T) \bigcup (T \times P) \to \mathbb{R}_{\geqslant 0}$ represents the set of arcs in terms of a function giving the weight of the arc as result: a weight equal to 0 means that the arc is not present;
- $C : T \to \mathbb{R}_{\geqslant 0}$ is a function, which associates each transition with a *rate*;
- m_0 represents the initial marking. A marking is defined formally as $m : P \to \mathbb{R}_{\geqslant 0}$.

The dynamics of a continuous Petri net can be expressed in terms of ODEs (in agreement with the standard mass action kinetics of chemical reactions). Each place corresponds to a continuous variable whose value corresponds the place's marking. The dynamics of the variable is expressed by a differential equation consisting of a summation of terms corresponding to the transitions connected to the place. The term has a positive sign if the the place is connected to the transition by an outgoing arc. The sign is negative otherwise. Moreover, the term is the product of the weight of the arc with the values of the variables corresponding to all the places providing resources to the transition (i.e., having an outgoing arc connecting them to the transition). Those variables have as exponent the weight of the arc connecting them to the transition.

For example, let us consider a continuous version of the Petri net in Figure 6 by assuming that the rate of the transition is k and that the marking is given in terms of continuous variables rather than of (discrete) numbers of tokens. The ODEs describing the dynamics of the Petri net are as follows:

$$\frac{d[H_2]}{dt} = -2k[H_2][O_2] \qquad \frac{d[O_2]}{dt} = -k[H_2][O_2] \qquad \frac{d[H_2O]}{dt} = +2k[H_2][O_2]$$

An alternative (stochastic) dynamics can be given by using the terms of the ODEs computed for each transition as rates of a Continuous Time Markov Chain (CTMC). Both ODEs and CTMCs offer standard analytical ways to compute the steady state of the system.

Hereinafter, we refer to continuous Petri nets simply as Petri nets and we assume their dynamics to be expressed in terms of ODEs.

2.2.1.1 *Example of use of Petri nets*

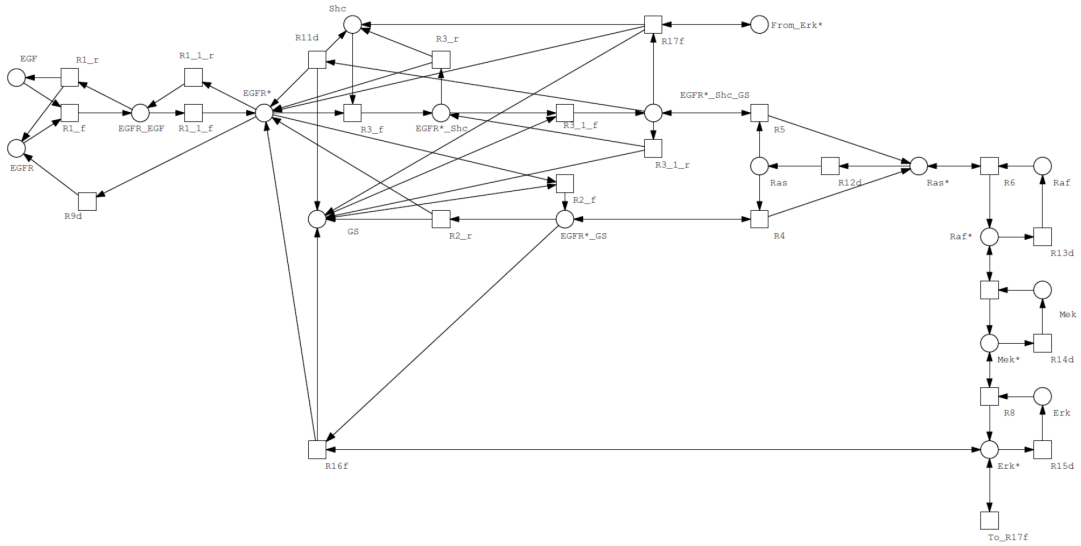

Figure 7: Petri net model of EGFR Ras- MAPK signaling.

Petri nets are an essential tool in many research fields, including Systems biology. Indeed, through this approach, we can analyze many biochemical systems, such as signaling pathways, metabolic networks, and gene regulatory networks [46].

The advantages of using Petri nets are the following:

- *Intuitive visualization*, which allows different levels of abstraction, useful when the model is very complex and difficult to analyze;
- *qualitative analysis*, which is crucial to describe the network topology and its structural properties;
- *quantitative analysis* to explore, in a general manner, possible observable behaviors (such as steady state or oscillations);
- *model validation* and *verification* through mathematical formalization, which offers an unambiguous representation of the model.

In general, the timeless discrete representation is the natural choice to describe the architectural properties of a biological system; instead, the timed continuous one is suggested for the quantitative analysis. As demonstrated in

[33, 38], both approaches provide different but complementary viewpoints on the same model. In particular, in the absence of exact parameters, Petri nets are indispensable to understand the dynamic qualities of a system.

An example of the use of Petri nets is in [14], where the authors present a Petri net model of the Epidermal Growth Factor Receptor (EGFR) signaling to the RasRaf-Mek-Erk (Ras-MAPK) pathway. The EGFR-Ras-MAPK pathway is strongly implicated in the development and progression of cancer, and, therefore, it is studied to develop suitable drugs that intervene in individual proteins, involved in this pathway.

By developing a Petri net model, as in Figure 7, it emerged that some proteins limited the functionality of the anti-cancer drugs. In particular, these proteins (places in the model) were siphons, which in Petri nets formalism means that they obstructed any signals. Intuitively, a siphon is a place that when it becomes empty, it remains empty.

2.3 LINEAR TEMPORAL LOGIC

As defined in Section 2.2.1, through the use of Petri nets, we can examine the behavior of a biological network as a continuous process. The trace obtained by a sequence of markings describes exactly the time evolution of the system.

In [64], as we will see in details in Section 3.3.1, the authors use a different approach to describe the dynamical properties of a biological network. Its dynamics is represented by a numerical trace, which is a discrete sequence of tuples describing system's evolution with time. Thus, the numerical trace corresponds to a discrete representation of a continuous process. Then, the dynamical system features are specified by the use of the *Linear Temporal Logic* (LTL), developed to express the behavior of discrete dynamical systems.

2.3.1 *Formal definition of LTL*

Linear Temporal Logic is a logical formalism. It provides a mathematical notion to express systems behaviors [6], based on a linear-time perspective. The temporal logic is necessary to specify the relative order of events, expressed by elementary modalities, which combined can express complex dynamical properties. Typical properties are oscillations (when a behavior recurs infinitely),

reachability (when the system can reach a given state), invariance (when a property is always true), inevitability (when a system has to reach a given state), response (an event causes a specific behavior) [28].

A basic LTL formula ϕ consists of atomic propositions $a \in AP$, Boolean connectors ($\wedge$, $\vee$, $\neg$, $\implies$), and two basic modal operators:

- $\mathbb{X}\phi$ or $\circ$ ("next") means that a given formula ϕ is true in the next state;
- $\phi\mathbb{U}\phi$ or $\cup$ ("until") means that given two formulas ϕ_1 and ϕ_2, the formula ϕ_1 is true, until the formula ϕ_2 becomes true.

The "until" operator $\cup$ allows to derive other two temporal modalities, defined as follows:

- $\mathbb{F}\phi$ or $\diamond$ "eventually (in the future)" means that a given formula ϕ is true now or sometime in the future, defined as $\diamond\phi \stackrel{def}{=} true \cup \phi$;
- $\mathbb{G}\phi$ or $\square$ "globally" means that a given formula ϕ is true now and forever, defined as $\square\phi \stackrel{def}{=} \neg \diamond \neg\phi$.

The LTL formulae are formed according to the following grammar:

$$\phi :: true \,|a|\, \phi_1 \wedge \phi_2 \,|\, \neg\phi \,|\circ\phi|\, \phi_1 \cup \phi_2$$

The atomic proposition a, with $a \in AP$ is a state label, representing an assertion about the value of a system variable that has to be evaluated, such as the concentration of a chemical species.

A LTL formula ϕ represents a property of a trace, which is a infinite path. Given a path and a formula ϕ, we can formulate precisely when ϕ holds on the path. For example, the trace T in Figure 8 satisfies the formula $\phi_1 = \mathbb{F}(x \vee \neg y)$ because it is true in the first state of the trace T. Instead, the trace does not satisfy the formula $\phi_2 = \mathbb{G}(y)$ because y is not true in all states.

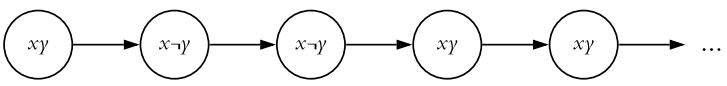

Figure 8: Example of Linear Temporal Logic.

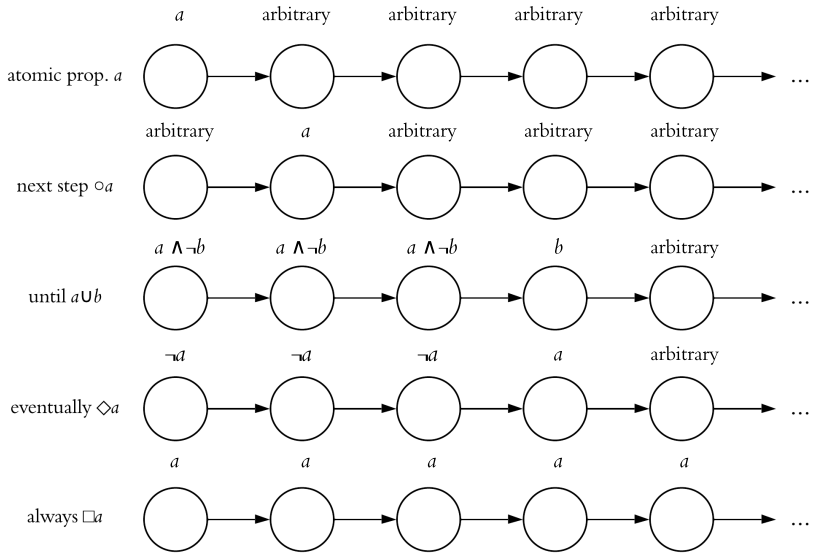

Figure 9: Intuitive sketch of LTL semantics.

2.3.2 *Semantics of LTL*

To precisely formulate when a path satisfies an LTL formula, we define the semantics of LTL formula ϕ by providing a satisfaction relation $\models$ such that $\sigma \models \phi$ if and only if a property ϕ is satisfied by a trace σ. [6]:

Definition 4. *Let ϕ be an LTL formula over AP and $\sigma \in (2^{AP})^\omega$ be a trace. The satisfaction relation $\models \subseteq (2^{AP})^\omega \times \mathrm{LTL}$ is the smallest relation with the following properties:*

$$
\begin{aligned}
&\sigma \models \mathrm{true} \\
&\sigma \models a \quad \mathrm{iff} \quad a \in A_0 \quad (\mathrm{i.e.}, A_0 \models a) \\
&\sigma \models \phi_1 \wedge \phi_2 \quad \mathrm{iff} \quad \sigma \models \phi_1 \quad \mathrm{and} \quad \sigma \models \phi_2 \\
&\sigma \models \neg\phi \quad \mathrm{iff} \quad \sigma \not\models \phi \\
&\sigma \models \circ\phi \quad \mathrm{iff} \quad \sigma[1...] = A_1A_2A_3... \models \phi \\
&\sigma \models \phi_1 \cup \phi_2 \quad \mathrm{iff} \quad \exists j \geqslant 0.\sigma[j...] \models \phi_2 \quad \mathrm{and} \quad \sigma[i...] \models \phi_1, \quad \forall 0 \leqslant i < j.
\end{aligned}
$$

Here, for $\sigma = A_0A_1A_2... \in (2^{AP})^\omega$, $\sigma[j...] = A_jA_{j+1}A_{j+2}...$ is the suffix of σ starting in the $(j+1)$st symbol A_j.

For the derived operator $\diamond$ and $\Box$ the expected result is:

$$\sigma \models \diamond\phi \quad \text{iff} \quad \exists j \geqslant 0.\sigma[j...] \models \phi$$
$$\sigma \models \Box\phi \quad \text{iff} \quad \forall j \geqslant 0.\sigma[j...] \models \phi.$$

In Figure 9, we add an intuitively sketch of the semantics of temporal modalities.

3

ROBUSTNESS PROPERTIES IN CHEMICAL REACTION NETWORKS

Many factors affect the internal and external behavior of a biological system, which is, for this reason, hard to study. To solve the uncertainty connected to its understanding, there exist several approaches, such as numerical simulations and experimental methods. In this context, Systems biology emerges as a powerful tool to investigate network dynamics merging computational methods and real data.

The development of predictive models requires information about the set of initial conditions, which characterizes biological systems and which are unfortunately challenging (or even impossible) to measure. Moreover, some parameters are affected by fluctuations that alter ordinary system behavior. Concerning this, one of the most efficient tools is *sensitivity analysis*, to study how distinct inputs result in a qualitative change of the system behavior. Therefore, applying sensitivity analysis we can verify if a model is *robust*.

Robustness is an observed biological characteristic, for which a system preserves its functions despite the presence of perturbations. Since its implication in particular diseases (such as diabetes and cancer), there are several works concerning this property.

In the first part of this Chapter, we will examine the role of local and global sensitivity analysis, explaining their differences and applications [42, 81].

In the second part, we focus on the robustness property, showing three significant works, representing a step forward in the study of this topic. The work in [45] delineates the first attempt towards a mathematical formalization of robustness property, albeit in a very general perspective, and it represents the starting point of the work in [64]. This latter one, indeed, suggests a specific formal definition for robustness, using a general computational framework based on temporal logic. The last one [70], done by Guy Shinar and Martin Feinberg, focuses on a particular acceptation of robustness definition, namely *absolute initial concentration robustness*, in which only the perturbations on the initial concentration of chemical species affect the CRN. In addition, the work

[70] lays the foundations to prove how the structure of a CRN characterizes its behavior.

3.1 SENSITIVITY ANALYSIS APPROACHES TO STUDY SYSTEMS BIOLOGY MODELS

Sensitivity analysis is a set of methods that allows understanding different aspects of systems [81]. It is an invaluable tool for:

- investigating the perturbation effects;
- determining which is the most (or the least) contributing parameter among all others, or estimate the parameters value;
- predicting and understanding the behavior of a biological model;
- guiding experimental analysis;
- establishing the initial conditions of a Systems biology models;

among the others. The noisy nature of biological data makes it difficult (usually impossible) to know the exact parameters. For this reason, mathematical models are the best method to determine such values [26]. In this context, both biological and mathematical models should be examined to reduce the uncertainty about parameters values, to valuate if the extracted information is realistic. Moreover, parameter values can be also influenced by the form used to shape the problem, hence also the mathematical model has to be valuated, as described in [26].

3.1.1 *Local and global sensitivity analysis*

Another interesting point of view on the study of biological models is to investigate their dynamics, examining the connections among the chemical species. Sensitivity analysis moves in this direction, dividing into two fundamental approaches: a *local* and *global* perspective.

Local sensitivity analysis is an investigation concerning the effects of small perturbations in a network. Given a model, described by a set of differential

equations, we can vary its initial conditions to study how this affects its dynamics. We identify with the term *input* all the perturbed parameters, such as the initial concentration of substances in a CRN. With the term *output*, we refer to all variables, significant to analyze the model response to disturbances, such as the steady state concentration of a chemical species.

Mathematically, local sensitivity analysis is performed by computing the first-order partial derivatives of the output with respect to the input, around a local point in the parameter space [81], and it is usually estimated with a perturbation in a single variable at a time. In particular, in a CRN we could study the impact of a small perturbation on the steady state values of a CRN, applying the *absolute local sensitivity analysis*, defined as follows:

Definition 5 (Absolute local sensitivity)**.** *Given a system modeled by a set of differential equations, let* P *and* S *be two parameters of the model, respectively the* input *and the* output *of the model. The absolute local sensitivity of* S *at steady state* (S_{ss}) *w.r.t to a variable* P *is defined as:*

$$\frac{dS_{ss}}{dP}.$$

This partial derivative represents the slope of the tangent to the continuation curve, and it can be used to predict how a small perturbation on P can affect the system at the steady state [41]. We can notice that the Definition 5 does not specify the entity of S and P, but it refers to them generically as "parameters": indeed, we can vary any model parameters (such as initial concentrations, kinetic rates, etc.).

To make an example of this approach, we recall the CRN 5 in Chapter 2. The steady state formula of the species A (the system output) is:

$$[A]_{ss} = \frac{k_1}{k_2}$$

Identifying k_2 as the model input, we can study how it influences $[A]_{ss}$:

$$\frac{d[A]_{ss}}{dk_2} = \frac{d}{dk_2}\left(\frac{k_1}{k_2}\right) = -\frac{1}{k_2^2}.$$

Choosing a nominal value for the input, such as its original value in the simulation represented in Figure 3, we can calculate the absolute local sensitivity coefficient:

$$-\frac{1}{k_2^2} = -\frac{1}{(0.5)^2} = -4,$$

which describes the impact on $[A]_{ss}$. Adding a perturbation in k_2, we can calculate directly the amount of changes in the steady state of A, using the following linearization formula, as described in [41]:

$$S_{ss} \cdot (P_0 + \Delta P) \approx S_{ss} \cdot (P_0) + \Delta P \cdot \left(\frac{dS_{ss}}{dP}\right)$$

Increasing k_2 of 0.1, we obtain:

$$S_{ss} \cdot (P_0 + \Delta P) \approx -0.6$$

and this result is confirmed in simulation reported in Figure 10.

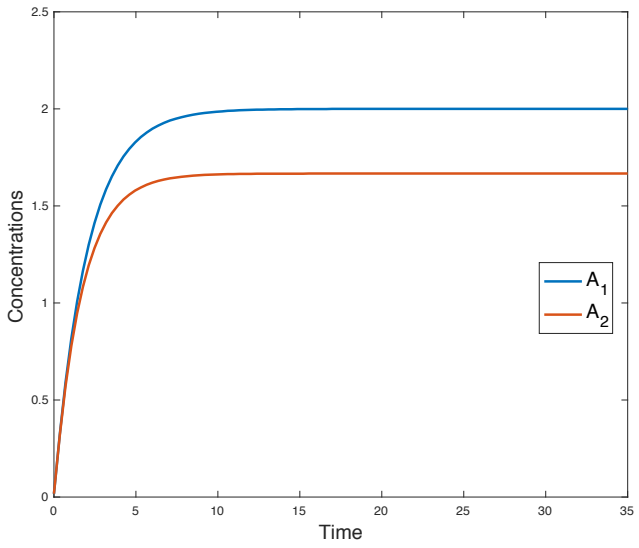

Figure 10: Simulation result of Example 5, presented in Chapter 2. We compare two simulations to show how the steady state concentration of A is influenced by k_2. The set of parameters are: $A_1 = 0$, $k_1 = 1$, $k_2 = 0.5$ and $A_2 = 0$, $k_1 = 1$, and the perturbed $k_2 = 0.5 + 0.1$.

The same approach can be used to describe the relative perturbations effect.

Definition 6 (Relative local sensitivity)**.** *Given a system modeled by a set of differential equations, let* P *and* S *be two parameters of the model, respectively the* input *and the* output *of the model. The relative local sensitivity is the relation between the size of a relative perturbation in* P *to a relative change in* S_{ss}*. It is defined as:*

$$\frac{\frac{dS_{ss}}{S_{ss}}}{\frac{dP}{P}} = \frac{P}{S_{ss}} \frac{dS_{ss}}{dP}.$$

To calculate the derivatives can be more or less arduous: it depends on the models complexity and the number of parameters, which we have to consider for the analysis [81]. Indeed, to explore multiparameter responses, using local sensitivity analysis, is possible but it is costly from a computational point of view. Moreover, this approach does not consider the model non-linearity and the interaction among different parameters, hence we have to proceed carefully in the data interpretation.

The global sensitivity analysis, on the other hand, is the set of approaches, studying the output of a model with respect to *large* variations in the input parameters [81], and typically involves sampling the space of parameters values and determining the corresponding system behavior [41]. Thus, while the local sensitivity analysis is estimating the derivatives at a particular point, in the global sensitivity the evaluation of the perturbations concerns the entire system behavior, and it is particularly indicated for determining which inputs are more significant.

As we described above, the local sensitivity analysis gives information about a small region in parameter space only. Thus, to explore the full parameter space, we can adopt one of the possible global sensitivity approaches. Briefly, we introduce some of them:

- *Average local sensitivities*: it is a local/global method, which combines both approaches. It computes the local sensitivity analysis, not only on a single set of values but on a large range of values, in a global fashion. Then, the average sensitivity coefficient is calculated, which differs from the coefficient found by a local approach. In this way, the sensitivity on the whole parameter space is tested. The main issue of the average local sensitivities is the expensive computational cost;

- *Variance-based sensitivity analysis*: it is an approach that can be applied in different ways, through many techniques [20]. In general, there are three basic principles [59]:

1. the inputs factors are stochastic variables so that the model induces a distribution in the output space;
2. the variance of the output distribution represents the model uncertainty;
3. the sensitivity measure is given by the contribution of an input factor on the output variance.

One of the possible techniques is the *First-order indices* [59, 66], which measures the direct contribution to the output variance from individual input factors. The effect on the output variance is expressed by the formula:

$$\frac{var_X[E(Y|X)]}{var(Y)}$$

where X and Y are respectively the input and output variables, the $[E(Y|X)]$ is the expectation of Y given a fixed value of X, while the var_X is taken over all the possible values of X. Then, First-order indices measure the single contribution of an input on the output variance.

Instead, if we want to analyze if an input factor is itself influenced by the other system parameters, we can use *Total Sensitivity indices* (TS), also known as *Sobol indices*. This method studies the interaction of an input with respect to the output and its correlation with all the other parameters. In this way, we can discern if the input effect is amplified by other factors. As described in [66], assuming that we have three input factors in the model (A, B, C) and that we want to measure the total effect of factor A, we can decompose its effect on the output variance, considering all the interactions between A and the other inputs, as in the following example:

$$TS_A = TS_A + TS_{AB} + TS_{ABC}$$

First-order indices and Total-order indices are useful because they can be applied easily on different models (such as non-linear, monotonic or non-monotonic systems). Moreover, the major disadvantage of variance-based methods is that the number of indices grow geometrically with

the number of inputs, hence the problem resolution could become intractable. Another limitation of this approach is linked to the assumption that the variance is the most significant indicator. Actually, for multimodal or highly-skewed models, to study the output variance could generate an inaccurate data interpretation.

- *Density-based methods*: This set of approaches differs from the variance-based sensitivity analysis because the output variance is not the only considered metric. In this context, the PAWN method [80] is an emerging technique: to describe the output uncertainty the entire model output distribution is considered and not only the variance, as in the Sobol' method, which implicitly assumes that this metric is a sufficient indicator.

In general, local and global sensitivity analysis allows establishing if the model output depends on its inputs factors, based on different approaches. In the following Example 3.1.1.1, we study the output of the a model using a local and a global method. To do this example, we take inspiration from the case study shown in paper [26], where the authors analyse how to study local and global sensitivity can originate different results.

3.1.1.1 *Example of difference between local and global sensitivity analysis*

To compare the local and global sensitivity analysis, we perform both approaches on the following model. Given a chemical reaction network:

$$\begin{aligned} A + B &\xrightarrow{k_1} C \\ C &\xrightarrow{k_2} D \\ D + E &\xrightarrow{k_3} F \\ F &\xrightarrow{k_4} C \end{aligned} \tag{9}$$

we want to study how the output C varies at the steady state, changing the concentrations of other chemical species involved in the network. Randomly, we set the initial concentrations values for each species: $A = 10$, $B = 32$, $C = 0$, $D = 4$, $E = 20$ and $F = 12$. Then, we proceed applying the local sensitivity analysis and the average local sensitivities.

Local sensitivity analysis. We fix the coefficient rates of reactions: $k_1 = 0.5$, $k_2 = 1.6$, $k_3 = 0.3$, $k_4 = 0.02$ and we vary by 2% the initial concentration of one input at time. Then, we simulate the model and we calculate the concentration of the output at the steady state, after the perturbation. In Figure 11 we plot, for each species, the difference between the concentration of C at steady state, without any perturbation, with respect to the concentration of C after the perturbation.

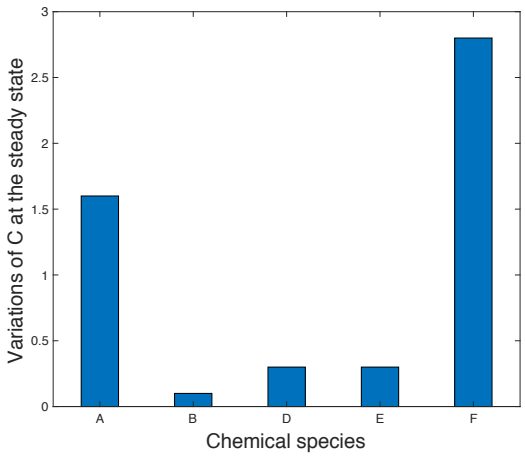

Figure 11: Graphical result of local sensitivity analysis on the chemical reaction network 9. The parameters perturbed are shown on the horizontal axis, on the vertical axis it is represented the variation of the concentration of the output C at the steady state.

Average local sensitivities. As in the local sensitivity analysis, we leave unchanged the coefficient rates of the network: $k_1 = 0.5$, $k_2 = 1.6$, $k_3 = 0.3$, $k_4 = 0.02$. We choose one species at time, varying its initial concentration in a large range: $A = [10, 120]$, $B = [32, 120]$, $D = [4, 120]$, $E = [20, 120]$ and $F = [12, 120]$. We perform the simulations considering all the integer values in the range of possible initial concentrations. Then, for each species, we calculate the average of difference between the concentration of C at the steady state with respect to the concentration of C, after the perturbation. In this way, we apply a sort of local sensitivity analysis in a global fashion. In Figure 12, we represent the average of difference between the concentrations of C at the steady state with perturbation and without perturbation.

Comparing the two Figures 11 and 12, we can notice that we obtain significantly different results. In particular, the chemical species F, which in the local

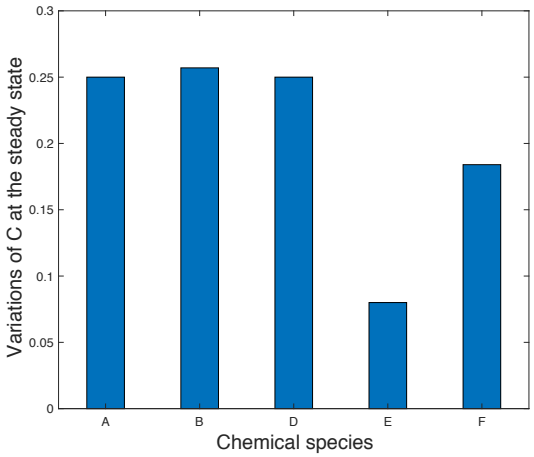

Figure 12: Graphical result of average local sensitivities on the chemical reaction network 9. The parameters perturbed are shown on the horizontal axis, on the vertical axis it is represented the average variation of the concentration of the output C at the steady state.

sensitivity analysis seems the most influential species, in the global sensitivity analysis, has a resized role. This difference shows how local sensitivity analysis could not reflect exactly the relation among parameters. Instead, to use a global approach it guarantees a deeper investigation because, intuitively, we can test a wide range of possible values and not only a single set.

In general, choosing which method to use to analyze a system depends on the system characteristics and on the properties to be verified. For example, the global sensitivity analysis is crucial to analyse the robustness of a system. As we will explain in details in Section 3.2, robustness is a biological property for which a system works despite any kinds of disturbances, which can alter its normal functioning.

In general, it requires many simulations and high computational efforts because many factors have to be considered, such as the numbers of variables, the correlations between input and output, the computational cost of the running model, the uncertainty of initial values. For this reason, other solutions have to be studied, as the possibility to conduct a preliminary analysis on the system, as we will propose in details in Chapter 5.

3.2 ROBUSTNESS MECHANISMS IN NATURE

Robustness is considered a fundamental feature of complex evolving systems. It is the property which allows the system to preserve its functions despite external and internal perturbations. In nature, there exist many examples of biological systems showing robustness, such as the segmental polarity of *Drosophila melagonaster*, the Lac Operon network, the chemotaxis of *Escherichia Coli*. The main mechanisms ensuring the robustness are system control, redundancy, modularity and structural stability [44].

The *system control* is the mechanism attaining an adaptive dynamic response of the network. There are two main instances of it: positive and negative feedback.

The first one is the amplification of the input signals in a pathway, to prevent the presence of noise, which can cause an error cascade. The positive feedback works when a biological system consists of many states, such as a cell cycle, and, intensifying the signal, it guarantees the transition from one state to another.

Instead, the negative feedback works stabilizing the system, which consequently adapts itself to the new environment. The principal representative of this phenomenon is the chemotaxis of *E. Coli* because it shows an evident robust adaptation to environmental changes. The chemotaxis is the process in which bacteria, such as *E. Coli*, move due to spatial and temporal gradients of specific substances, called *attractants* or *repellents* if they attract or repulse the bacteria, respectively. Roughly speaking, while moving bacteria perceive the difference between the previous and the current concentration, sensing the gradients by making temporal comparisons. In the case of attractants, like food for example, bacteria move along straight lines in the direction of the increasing gradient (runs); on the other hand, if they perceive repellents, they are found to change randomly direction (tumbles). With the help of *in vivo* experiments and of simulations, it was shown that another crucial property of chemotaxis is the adaptation: a change in the concentration of a chemical stimulant induces a rapid change in the bacteria's tumbling frequency, which gradually adapts back precisely to its pre-stimulus value, as described in [2].

Redundancy is the mechanism according to which a system has multiple structures with the same functions: the aim, in this case, is to avoid errors or failures. If one or more components do not work efficiently, other elements

replace them. A redundancy example occurs in tissues: cells, with equivalent functionalities, replace each other to avoid error propagation. Or in signalling pathways, such as in the MAP kinase cascade: if some pathways are disabled because of errors or mutations, there are collateral pathways preserving the cellular activity.

Modularity is a mechanism, ensuring that errors and damages remain isolated, without spreading in the other compartments of the system. Thus, the consequences of failure do not affect the whole structure. A clear example is the cell, that can be seen as an entire structure or as a set of functional components (as organella and membranes): if there is an error in one of its parts, it does not involve the entire cell.

The *structural stability* is the quality according to which a system can adapt to changes even in the presence of different *initial* parameter variations. Some examples of it are demonstrated by some gene regulatory circuits, that are stable for a broad range of initial pulses and genetic polymorphisms [43].

Following the above mentioned examples, it emerges that the robustness is both an internal quality and an architectural characteristic of the system, that enables complex systems to evolve after a specific environmental perturbation.

3.3 MATHEMATICAL FORMULATION OF BIOLOGICAL ROBUSTNESS

Delineating the concept of robustness and giving a mathematical formalization of it are fundamental problems in Systems biology. This because robustness needs a precise definition since it is often confused with other notions concerning different biological properties.

The first attempt to give a mathematical representation of robustness is in [45], where there is, first of all, a clear distinction between stability and robustness. The first one is the set of mechanisms that ensures that the system has a steady state; the second one is a mechanism ensuring the system adaptation in the environment.

Formally, robustness is defined as follows:

Definition 7 (Kitano's Robustness)**.** *The robustness* (R) *of a system* (s) *with respect to one of its functions* (a) *can be computed as*

$$R^s_{a,P} = \int_P \psi(p) D^s_a(p) dp$$

where:

- P *is the perturbation space;*
- $\psi(p)$, *with* $p \in P$, *is the perturbation probability;*
- $D^s_a(p)$ *is the evaluation function, which gives the level of maintenance of functionality* a *under perturbation* p.

Definition 8 (Evaluation function)**.** *Given the set* P *of all possible perturbations, let* $A \subset P$ *be the set of perturbations for which the systems fails, and* $p \in P$. *The evaluation function is defined as:*

$$D^s_a(p) = \begin{cases} 0, p \in A \subset P \\ \frac{f_a(p)}{f_a(0)}, p \in P \backslash A \end{cases}$$

The function $D(p)$ is zero when a system does not maintain its functions under a perturbation, otherwise the function returns the relative viability of a function under perturbation $(f_a(p))$ against non-perturbed condition $(f_a(0))$.

Using the approach defined in Definitions 7 and 8, we can compare two systems to know which is more robust between them: a system S_1 can be considered more robust with respect to a system S_2, considering the same function a and the same perturbations set Y if:

$$R^{S_1}_{a,Y} > R^{S_2}_{a,Y}$$

The work presented so far represents the starting point for the work [64], in which what is criticized is the lack of information about how to specify the evaluation function, described in Definition 7, which needs to be defined for each specific problem and re-implemented every time for the computation of robustness. Instead, in [64], the authors propose a general definition of robustness that applies to any biological function expressible in *Linear Temporal Logic* (LTL), described in details in Section 2.3.1.

3.3.1 *Temporal logic semantics of numerical traces*

Before to present, in details, the robustness notion proposed in [64], it is necessary to introduce some background notions. First of all, in [64], numerical

simulations are used to obtain the behavior of a biological system, which is described by a *numerical timed trace*.

A *numerical timed trace*, expressing the evolution of a system with time, is a finite sequence of tuples $T = (s_0, s_1, ..., s_n)$ with $s_i = (t_i, \mathbf{x_i}, \dot{\mathbf{x}}_\mathbf{i})$ where t_i with $i \in [0, n]$ is a sequence of increasing time points, $\mathbf{x_i}$ is the vector of state variable values and $\dot{\mathbf{x}}_\mathbf{i}$ is the derivative of state variable at time t_i.

By a numerical trace, we can depict different biological phenomena, such as the time evolution of a concentration of a chemical species in a network, as represented in Figure 13. In this example, $T = ((0, 2, 0), (1, 6, 4.12), ..., (9, 10, 0))$ is the the associated trace, in which each state variable defines a specific concentration level of B over time.

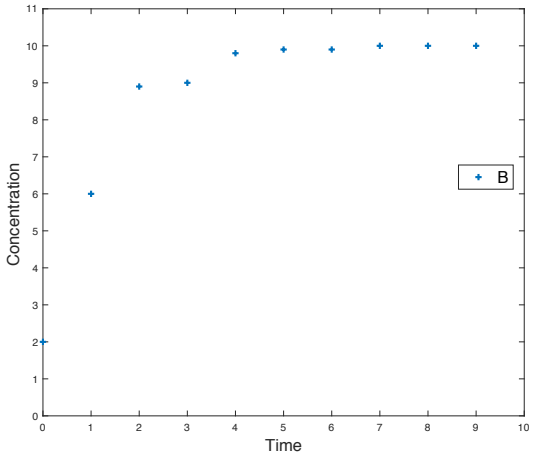

Figure 13: Numerical trace representing the time evolution of the concentration of chemical species B.

As mentioned above, the authors of the work in [64] use LTL to express dynamical properties of biological systems. LTL operators, as seen in details in Section 2.3.1, describe if and when a property ϕ holds on a trace T. Considering again Figure 13, the formula $\phi = \mathbb{F}([B] > 7)$ expresses that at some point the concentration of species B is greater than 7.

Since it is interesting to define how much a numerical trace satisfies a formula ϕ, the authors specify the quantifier-free LTL (QFLTL), which replaces the numerical constants in the atomic propositions of a LTL formula, with free real-valued variables $\mathbf{y}$. Then, in this way, having a formula ϕ and a vector of real-valued variables $\mathbf{y}$, it is possible to know for which values $\mathbf{y}$ the QFLTL

formula $\phi(\mathbf{y})$ holds on T. At this point, the *satisfaction domain* is defined as follows:

Definition 9 (Satisfaction Domain)**.** *Given a QFLTL ϕ formula, for any trace T, the satisfaction domain of $\phi(\mathbf{y})$ is the set of variables $\mathbf{y}$ for which $\phi(\mathbf{y})$ holds. It is defined as:*

$D_{T,\phi(\mathbf{y})} = \{\mathbf{y} \in \mathbb{R}^q | T \models \phi(\mathbf{y})\}$ *where q is the number of constants appearing in ϕ.*

Through this approach, the LTL formula becomes an instance of a more general QFLTL formula obtained by variable abstraction, as described in [64]. Considering again Example 13, and the formula $\phi = \mathbb{F}([B] > 2 \wedge \mathbb{F}[B] < 10)$, it is possible to associate the formula $\phi(\mathbf{y}) = \phi(y_1, y_2) = \mathbb{F}([B] > y_1 \wedge \mathbb{F}([B] < y_2))$. Moreover, concerning trace T in Example 13, the domain is $D_{T,\phi(y_1,y_2)} = \{y_1 \leqslant 10 \wedge y_2 \geqslant 2\}$, since 2 and 10 are respectively the minimum and the maximum values of the trace.

Given a trace $T = (s_0, s_1, ..., s_n)$ and a LTL formula ϕ, the authors define the notion of *violation degree* to quantify how much ϕ must be changed to hold on T. This concept is defined as the *Euclidean distance* between a formula ϕ and the domain of the trace T. It is formally defined as follows:

Definition 10 (Violation Degree)**.** *The violation degree $vd(T, \phi)$ of a formula ϕ w.r.t a trace T is the distance between the actual specification and validity domain $D_{T,\phi((y))}$ of the QFLTL formula $\phi(\mathbf{y})$ obtained by variable abstraction:*

$$vd(T, \phi) = dist(\phi, D_{T,\phi(\mathbf{y})})$$

Considering Example 11 and the formula $\phi_1 = \mathbb{F}([B] > 2 \wedge \mathbb{F}([B] < 10))$, by violation degree, we can compute how much ϕ is distant from the domain: in this case, we obtain $vd(T, \phi_1) = 0$ because ϕ is satisfied by T. Instead, if we consider the formula $\phi_2 = \mathbb{F}([B] > 12 \wedge \mathbb{F}([B] < 3))$, the violation degree is $vd(T, \phi_2) = 2$ meaning that ϕ_2 has to be changed to hold on T.

To define, instead, how much the given LTL formula holds on a given numerical trace, the notion of *satisfaction degree* is introduced as follows:

Definition 11 (Satisfaction Degree)**.** *The satisfaction degree $sd(T, \phi)$ of a formula ϕ w.r.t a trace T is defined as:*

$$sd(T, \phi) = \frac{1}{1+vd(T,\phi)} \in [0, 1],$$

where $vd(T, \phi)$ *represents the violation degree.*

The value obtained by computation of the satisfaction degree ranges between 0 and 1. The satisfaction degree is equal to 1 when the trace T satisfies a formula ϕ, otherwise it tends to 0. The concept of satisfaction degree in [64] characterizes the definition of robustness. As previously explained, for the robustness definition in this work, they get inspiration from the one given in [45], replacing the evaluation function with the concept of satisfaction degree.

Definition 12 (Rizk et al. Robustness)**.** *The robustness of a system is defined as:*

$$R^s_{\phi,P} = \int_{p \in P} prob(p) sd(T_p, \phi) dp$$

where ϕ *is the specification of the functionality in LTL;* T_p *is the numerical trace, representing the system behavior under perturbation* p*;* P *is the set of perturbations.*

The continuous probability distribution characterizes the perturbations, affecting the entire system: each perturbation has its weight, representing how much it can influence the biological behavior under study.

To compare how different systems react to the same kind of perturbations, they give a formal relative robustness definition, this notion was already introduced in other works [10, 45].

Definition 13 (Relative Robustness)**.** *The relative robustness of a system w.r.t a nominal behavior is defined as the system's robustness divided by its satisfaction degree of the reference behaviour (its nominal performance):*

$$R^{s,p*}_{\phi,P} = \frac{R^s_{\phi,P}}{sd(T_{p*}, \phi)},$$

where T_{p*} *denotes the unperturbed, nominal behavior of the system.*

3.4 STRUCTURAL SOURCES OF ROBUSTNESS IN CRN

Biological properties are challenging to study because they require the observation of the system behavior, considering all the possible initial states.

For example, regarding the robustness evaluation of a signaling pathway, all the combinations of initial concentrations of chemical species have to be examined, and this requires many simulations with different hypotheses. For this reason, in literature, numerous works introduce methods and approaches,

which verify this property, avoiding the execution of the entire system. In [30], the *Deficiency One Theorem* and *Deficiency Zero Theorem* are presented, and both of them give crucial information about the steady state of the system, using only linear algebra and without any simulation.

The Deficiency One Theorem is the basis of one of the most relevant contributions [70]. This work defines a sufficient condition that, if met, guarantees the robustness of the system under study, looking only at its chemical reaction network structure. Chemical reactions can exhibit cryptic behaviors, due to the fact that non-linear differential equations govern their dynamics. Nevertheless, researches have achieved relevant progress using simple models, which have the advantage to reduce the complexities of large systems. Thus, the approach in [70] gets general information (such as oscillations, stable or unstable steady states), looking only at the architecture of the CRN, and avoiding the resolution of differential equations.

The Deficiency Zero Theorem is fundamental to study the sensitivity of a CRN. As explained in [69, 71], if the Deficiency Zero Theorem is met, the CRN is not robust but it can be more or less sensitive to the variation of initial concentrations of chemical species.

3.4.1 *The Decificiency One Theorem and the robustness verification*

As mentioned in the Introduction of Section 3.4, the work done by Shinar and Feinberg analyzes the *absolute concentration robustness* of a CRN, only looking to its structure. Informally, this property is verified when a chemical species, involved in the CRN, maintains the same concentration at the steady-state, regardless of the initial conditions of the system. Before going into detail, we introduce some auxiliary definitions, laying the foundation for the structural analysis of a chemical reaction.

To investigate the architectural properties of a CRN means to imagine it as a graph, with a specific notation defined *Standard Reaction Diagram* (SRD). Considering the following CRN:

$$\begin{aligned} A + B &\xrightarrow{\alpha} 2B \\ B &\xrightarrow{\beta} A, \end{aligned} \tag{10}$$

we build the graph represented in Figure 14.

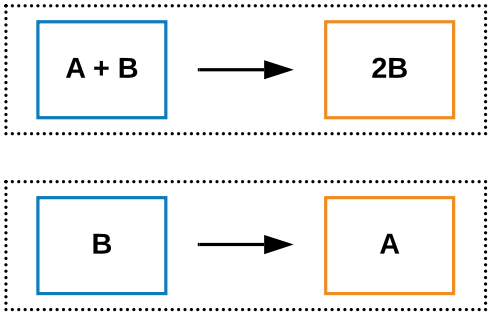

Figure 14: Standard Reaction Diagram of Example 10. The nodes of the graph are the *complexes* of the network, the edges represent the reactions.

By graph theory, we know that we can represent a graph using an *incidence matrix* $c \times e$ M, in which c and e represent respectively the nodes and the edges of the graph. Each M_{ce} will have 1 if the node c is a source of the edge e, -1 if the node c is the sink of the edge e, 0 if e is not connected to c. Then, considering Figure 14, we obtain the following incidence matrix:

$$\begin{array}{c} \\ A+B \\ 2B \\ B \\ A \end{array} \begin{array}{c} \begin{array}{cc} R_\alpha & R_\beta \end{array} \\ \begin{pmatrix} 1 & 0 \\ -1 & 0 \\ 0 & 1 \\ 0 & -1 \end{pmatrix} \end{array}.$$

On the graph 14 we define:

- the *nodes of the graph* are the group of chemical species that are participating as reactants or product. In Figure 14, the nodes are: A+B, $2B$, B and A. We can notice that we consider as node, not each species singularly, but the species which are connected by an arrow (identifying a reaction). Moreover, we define two nodes *strongly connected* if there is a path from one node to the other and and also a another path from the second back to the first. Conventionally, each node is strongly linked to itself;

- the *linkage classes* represent the distinct group of reactions, in which the graph is divided. In Figure 14, we have two linkage classes. A linkage

class is considered *strong* if all the nodes, included in it, are strongly connected;

- the *terminal strong-linkage class* is a linkage class in which no node reacts to a node in another strong-linkage class. If a node belongs to a terminal linkage class, it is called *terminal node*, otherwise a node belonging to a non-terminal linkage class is called *non-terminal*. In Figure 14, the non-terminal node are *A+B* and *B*, the terminal ones are *2B* and *A*.

Each species involved in the CRN can be represented as a vector. In 10 the two species involved are A and B and the vectors are: $A = [1, 0]$ and $B = [0, 1]$. The graph nodes defined previously are the *complexes* of the network and to each of them we can associate a vector. We obtain a set of the *complexes vectors* $A + B = [1, 1]$, $2B = [0, 2]$, $B = [0, 1]$ and $A = [1, 0]$.

The next level of representation of CRN is to associate to it the *reaction vectors*, obtained by subtracting the reactant complex vector to the product complex vector. Considering the example 10, we have: $2B - (A + B) = [-1, 1]$ and $A - B = [1, -1]$.

A reaction network has *rank* s, a positive integer number, if there exists a linearly independent set of s reaction vectors for the network and there exists no linearly independent set of $s + 1$ reaction vectors. Thus, the rank is the number of the elements in the largest linearly independent set of reaction vectors for the network. Considering the example 10, the two reaction vectors are linearly dependent, hence the rank of the network is equal to 1. Intuitively, this is explained by noting that the first reaction produces the species B and, vice versa, the second reaction produces the species A.

To use a formal procedure to determine the CRN rank, we can build the stoichiometric matrix $r \times N$ (Γ), where r represents the reactions of CRN and N is the species involved in the CRN and compute the rank of the obtained matrix. In this case, we have the matrix:

$$\begin{array}{c} \\ R_\alpha \\ R_\beta \end{array} \begin{array}{c} \begin{array}{cc} A & B \end{array} \\ \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix} \end{array}.$$

whose the rank is 1.

At this point, we notice that there two possible abstractions for the under study CRN:

- as a graph, then we can use an incidence matrix, studying how the complexes interact among them;
- as a stoichiometric matrix, studying how the single species are involved in the reactions.

Then, the incidence matrix gives us information about the structure of the CRN, while the stoichiometric matrix represents the dynamical characteristics of the system. According to these parameters, now we can define the *deficiency* of the network, a non-negative integer index, representing the amount of *linear independence* among the reactions of the network [71]. It is defined as:

$$\mathtt{deficiency} = \mathtt{dim}(\mathtt{Ker}(M_{ce}) - \mathtt{rank}(\Gamma),$$

where $\mathtt{dim}(\mathtt{Ker}(M_{ce})$ is dimension of the kernel of the incidence matrix (M_{ce}) and Γ is the stoichiometrix matrix. In [36], it is proved the following equivalence:

$$\mathtt{dim}(\mathtt{Ker}(M_{ce}) = n - l,$$

where n is the number of nodes and l is the number of linkage classes. Then, with regards to example 10, we have:

$$\mathtt{deficiency} = n - l - \mathtt{rank}(\Gamma) = 4 - 2 - 1 = 1.$$

Hence, the deficiency is equal to 1. The lowest value of the deficiency, which is zero, coincides to the higher extent of linear independence among the chemical reactions; instead, the highest value corresponds to a lower linear independence. Intuitively, the deficiency represents how much the reactions and the species influence each other in the CRN.

Before to introduce the Deficiency theorem, we introduce four other concepts:

- *composition*: it is the vector **c**, containing all the concentration of the chemical species involved in the network, in a specific time;
- *weakly reversible*: a reaction network is weakly reversible if whenever there is a reaction going from a complex x to a complex y, there is a path of reactions going back from y to x, where the paths need to follow the arrow directions;

- *stoichiometric subspace*: given a CRN with N species, the stoichiometric subspace is the set of vectors in $\mathbb{R}^N$, consisting of all possible linear combinations of the reaction vectors of the network, and it follows that the dimension of the stoichiometric subspace is equal to the rank of the stoichiometric matrix;

- *stoichiometric compatibility class*: suppose that $S \subset \mathbb{R}^N$ is the stoichiometric subspace of a CRN and $\mathbf{c}$ and $\mathbf{c}' \in \mathbb{R}^N$ are two composition of the same CRN. We say that $\mathbf{c}$ and $\mathbf{c}'$ are *stoichiometrically compatible* if $\mathbf{c}' - \mathbf{c}$ lies in S. Then, the stoichiometric compatibility class is the set of all possible compositions containing the composition $\mathbf{c}$.

Now we are ready to introduce the *Deficiency One Theorem* [30], which gives us information about uniqueness and existence of positive steady states.

Theorem 1 (Deficiency One Theorem)**.** *Consider a mass action system for which the underlying reaction network has l linkage classes, each containing just one terminal strong linkage class. Suppose that the deficiency δ of the network and the deficiencies of the individual linkage classes θ satisfy the following conditions:*

1. $\delta_\theta \leqslant 1$, $\theta = 1, 2, ...l$
2. $\sum_{\theta=1}^{l} \delta_\theta = \delta$.

Then no matter what (positive) values the rate constants take, the corresponding differential equations can admit no more than one steady state within a positive stoichiometric compatibility class. If the network is weakly reversible, the differential equations admit precisely one steady state in each positive stoiochiometric compatibility class.

In [70], the *Deficiency One theorem* is applied to the robustness concept. In this case, the authors consider a biological system absolute robustness for a species if the concentration of that species is identical in every positive steady state that the system might admit. To verify if a biological system has a species robust, for any perturbation, we can apply the following theorem:

Theorem 2 (Shinar and Feinberg's Absolute Concentration Robustness)**.** *A mass action system can be considered robust if it admits a positive steady state, the underlying reaction network has a deficiency equal to 1 and there are distinct nonterminal complexes that differ only in a single species.*

Considering the example 10, the CRN deficiency is equal to 1 and the nodes $A + B$ and B are non-terminal nodes (since they belong to two non-terminal complexes), which differs only for one species, A. Hence, we can state that this CRN is robust in A.

3.4.2 *The Deficiency Zero Theorem and the sensitivity analysis*

Using the same terminology already presented in Section 3.4.1, the authors define the *Deficiency Zero theorem* as described in [30]:

Theorem 3 (Deficiency Zero Theorem)**.** *For any reaction network of deficiency zero the following statements hold true:*

1. *if the network is not weakly reversible, then, for arbitrary kinetics (not necessarily mass action), the differential equations for the corresponding reaction network cannot admit positive steady state;*
2. *if the network is not weakly reversible then, for arbitrary kinetics, the differential equations for the corresponding reaction network cannot admit a cyclic composition trajectory along which all species concentrations are positive;*
3. *if the network is weakly reversible then, for mass action kinetics, the differential equations for the corresponding reaction network have the following properties:*
   - *There exists within each positive stoichiometric compatibility class precisely one steady state;*
   - *the steady state is asymptotically stable;*
   - *there is no nontrivial cyclic composition trajectory along which all species concentrations are positive.*

Thus, if the system is weakly reversible and its deficiency is equal to zero, we can affirm that the CRN has one positive steady state. As explained in [71], we can have some information about robustness only looking at the structure of the CRN. If the system deficiency is equal to zero, the system cannot be robust but it is more or less sensitive to the introduced perturbations in the environment.

Theorem 4 (Deficiency Zero and Absolute concentration robustness)**.** *Consider a mass-action system in which the underlying reaction network is conservative and*

has a deficiency of zero. Then, no matter what values the rate constants take, there is no species relative to which the system exhibits absolute concentration robustness.

What we can know about the model, when the deficiency is equal to zero, is described in [69]. From Theorem 4, we know that the system cannot be robust, but we can still study how much the individual chemical species involved in the CRN are sensitive to change of initial parameters. What emerges from this work, in fact, is an approach that allows us to calculate the *intrinsic lower bound* on the species sensitivities, which we will apply it to the Becker-Döring equations, explained in details in Chapter 6.

4

FORMALIZATION OF INITIAL CONCENTRATION ROBUSTNESS

In the first part of this Chapter, we propose a new formal definition of robustness notion against perturbations to the initial concentrations of species, based on Petri nets. We demonstrate the validity of our definition by applying it to the models of four different robust biochemical networks.

In the second part, we prove that our definition of *absolute initial concentration robustness* is an instance of the general Definition 12 given by Rizk et al. and described into detail in Chapter 3.

4.1 INITIAL CONCENTRATION ROBUSTNESS

Given a biochemical network, like a signalling pathway, our idea is to verify whether, even by varying the initial concentrations of some chemical species, the *output* of the chemical reactions remains either constant or, at least, bounded within a given interval of values. We will assume the initial concentration of the input molecules of the pathway to vary within certain given intervals, while the initial concentrations of all the other molecules (that are neither input, nor output) to be fixed. Under these assumptions, we define the property of robustness of the system and we formalize it by using Petri nets.

First, we introduce some auxiliary definitions. We extend the concept of marking. Recall that in Section 2.2.1 we defined the initial marking as an assignment of a fixed value to each place p. Now, we generalize the idea of initial marking by considering a marking as an assignment of an *interval of values* to each place p of the Petri net. We first define the domain of intervals.

Definition 14 (Intervals)**.** *We define the interval domain*

$$\mathcal{I} = \{[n, m] \mid n, m \in \mathbb{R}_{\geqslant 0} \cup \{+\infty\} \text{ and } n \leqslant m\}.$$

Moreover we say that $x \in [n, m]$ *iff* $n \leqslant x \leqslant m$.

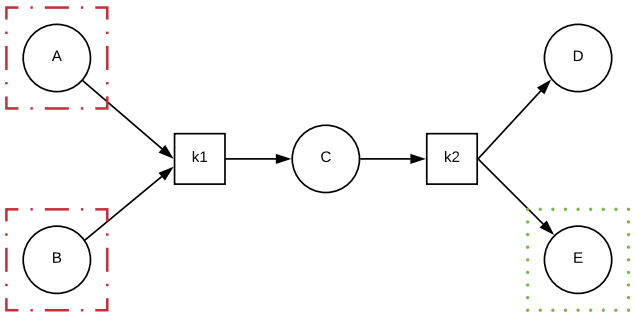

Figure 15: Example of Petri nets, in which A and B are marked as input of the system (red dot-line) and E is marked as output (green dots).

We now define interval markings.

Definition 15 (Interval marking). *An interval marking is a function* $m_{[\,]} : P \to I$. *We call* $M_{[\,]}$ *the domain of all interval markings.*

A non-trivial interval marking (i.e., an interval marking in which at least one interval is non-trivial) represents an infinite set of markings, one for each possible combination of values of the non-trival intervals. Therefore, given an interval marking, we relate it with the markings as in the original Petri nets formalism in the following way:

$$\text{Given a } m \in M \text{ and } m_{[\,]} \in M_{[\,]}, \; m \in m_{[\,]} \text{ iff } \forall p \in P, m(p) \in m_{[\,]}(p).$$

In a Petri net PN we assume that there exists at least one place p that we consider as the *input* of the network. In addition, we assume that there exists also a (unique) place p that we consider the *output* of the net, as shown in Figure 15 as example.

Within this framework, we can give our formal definition of absolute concentration robustness.

Definition 16 (α-Robustness). *A Petri net* PN *with output place* O *is defined as* α-robust with respect to a given interval marking $m_{[\,]}$ *iff* $\exists k \in \mathbb{R}$ *such that* $\forall m \in m_{[\,]}$, *the marking* m' *corresponding to the steady state reachable from* m, *is such that*

$$m'(O) \in [\, k - \frac{\alpha}{2}, k + \frac{\alpha}{2}] \,.$$

Given the previous definition, it can be observed that:

- the wider are the intervals of the initial interval marking, the more robust is the network. Indeed, it means that the system is able to absorb a higher level of perturbations.
- the smaller is the value of α, the more robust is the system.

Thus, in our framework we identify always two places as the input and the output, which we formally defined as:

Definition 17 (Input)**.** *Given a Petri net* PN*, with an interval marking* $m_{[\,]}$ *and an input place* I*. The interval marking of* I*, denoted as* $m(I)$*, is defined as* $m - n > 0$*.*

Definition 18 (Output)**.** *Given a Petri net* PN*, with an interval marking* $m_{[\,]}$ *and an output place* O*. The interval marking of* O*, denoted as* $m(O)$*, is defined as* $m - n = 0$ *and the the marking reached at the steady state as*

$$m'(O) \in [\, k - \frac{\alpha}{2}, k + \frac{\alpha}{2}] \,.$$

Here, we have given a general definition that can be modified in different ways. For example, rather than considering the marking at the steady states, it is possible to consider the marking reached at a given time T, or when the system terminates its execution (no transition is enabled).

It is worth noting that our definition is general enough to capture several notions of robustness available in literature. For example, by considering the initial intervals $[1, \infty]$ for the initial concentration of the input species and $\alpha = 0$ we obtain a formal definition for the robustness notions 1, described in Section 3.4.

A simple example of robust biochemical network is given by the following two reactions:

$$A + B \xrightarrow{k_1} 2\,B \qquad B \xrightarrow{k_2} A$$

The Petri net representation of the network is shown in Figure 16 (on the left with the initial marking, on the right with the steady state marking). In this case, the steady state is such that

$$m'(A) = \frac{k_2}{k_1} \qquad m'(B) = \theta - \frac{k_2}{k_1}$$

where θ is the sum of initial concentrations of A and B. If A is the output of the system, then its concentration at the steady state does not depend on

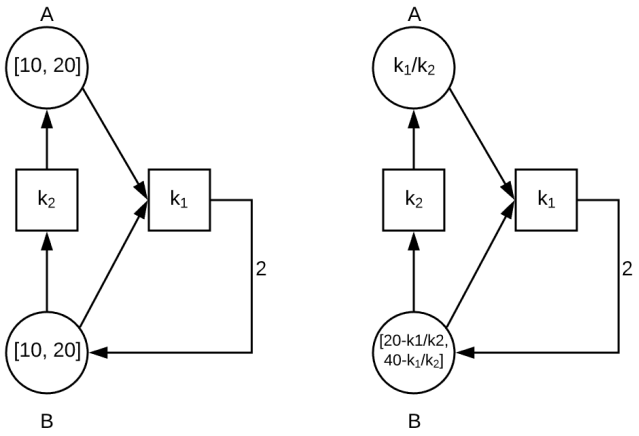

Figure 16: Example of robust biochemical network, considering the species A as output of the system.

the initial quantity of the (input) chemical species A and B (0-robustness with $k = \frac{k_2}{k_1}$).

If we consider $[10, 20]$ as the initial interval for both A and B, we obtain that θ will be in $[20, 40]$. So, for B as the output we obtain:

$$m'(B) \in [20 - \frac{k_2}{k_1}, 40 - \frac{k_2}{k_1}]$$

Thus, for output B we have α-robustness with $\alpha = 20$, suggesting that B is not independent on the initial concentrations of A and B.

In Figure 17 we can see a network that is never robust neither considering A as output, nor B. The chemical reactions of the CRN are: $A \xrightarrow{k_1} B$, $B \xrightarrow{k_2} A$. In this case, the concentrations of A and B at the steady state are both always influenced by the input values. Intuitively, we can motivate this behavior, considering that this CRN will reach a different steady state depending on the initial concentrations of the species. Moreover, we find that A is α-robust for $\alpha = 18$, B for $\alpha = 23$.

4.1.1 *Relative initial concentration robustness*

To compare the α-robustness of different systems or the α-robustness of the same system, but with different perturbations, we have to introduce another

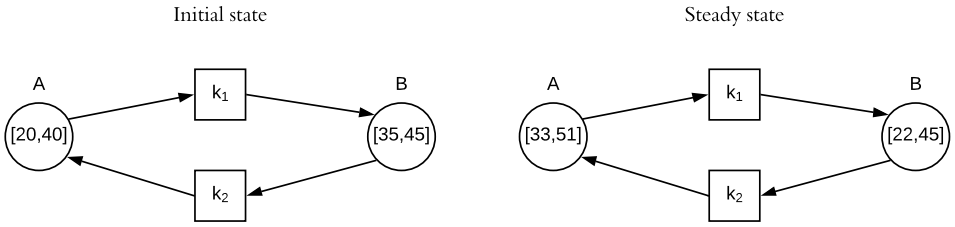

Figure 17: Example of non robust network. In this case we choose $k_1 = 2$ and $k_2 = 3$.

notion: the *relative* β*-robustness*. First of all, we introduce the concept of normalization of α-robustness defined as:

Definition 19 (Normalized α-robustness)**.** *Let* PN, α *an* k *be as in Definition 16. The normalized* α*-robustness of the output* O*, denoted* n_O*, is defined as* $\frac{\alpha}{k}$.

Definition 20 (Normalized Interval Marking)**.** *Let* PN, α *an* k *be as in Definition 16. The normalized interval marking of the input* I*, denoted* n_I*, is defined as* $\frac{m-n}{k}$.

Therefore, we can state the definition of relative initial concentration robustness as follows:

Definition 21 (Relative β-robustness)**.** *Let* PN *be as in Definition 16. The relative initial concentration robustness, denoted as* β*-robustness, is defined as:* $\frac{n_O}{n_I}$*, where* n_O *and* n_I *are respectively the normalized* α*-robustness and the normalized interval marking of* I.

4.1.1.1 *Example of relative robustness*

Considering Example 17. The species A and B are the input and the output respectively. The initial conditions of the systems are: $A = [20, 40]$ and $B = [35, 45]$. At the steady state, the concentration of B is in the interval $[22, 45]$.

First, We calculate the normalized α-robustness, as described in Definition 19, then we obtain:

$$n_O = \frac{\alpha}{k} = \frac{23}{33.5} = 0.68.$$

The normalized input values is calculated as follows:

$$n_I = \frac{m_I}{k} = \frac{20}{30} = 0.66.$$

Finally, the relative β-robustness is calculated as follows:

$$\beta - \text{robustness} = \frac{n_O}{n_I} = \frac{0.68}{0.66} = 1.03.$$

4.2 VALIDATING THE DEFINITION OF ROBUSTNESS

To validate our definition of robustness, we consider four examples of biological networks. The first two, the two component *EnvZ/OmpR* osmoregulatory signalling system and the isocitrate dehydrogenase regulatory system of *E. coli*, show an absolute concentration robustness (the α parameter of Definition 16 will be equal to 0). The third example models the bacterial chemotaxis of *E. coli*. The last example models an enzyme kinetics at saturation behaviour, inspired from the Lotka-Volterra reactions [34, 73], which shows a concentration robustness that it is not absolute (in this case the α parameter of Definition 16 will be greater than 0).

As illustrated in Definition 16, the aim of this work is to verify robustness at the steady state. Therefore, as first step, we should calculate analytically the steady state of the model, depicted in details in Section 2.1.4, or calculating the system deficiency depicted by Definitions 3.4.1 and 3.4.2. As an alternative, we could also verify robustness at any time t.

Concerning Example 4.2.1, we show the steady state calculation, step by step. For the other examples, we omit it.

4.2.1 *EnvZ/OmpR osmoregulatory signalling system*

In bacteria and in particular in E. coli, the *EnvZ/OmpR* system has the function to regulate the expression of two porins, *OmpF* and *OmpC*, which are proteins having many roles in the cell, as for example nutrients transportation, elimination of toxins and many others [12].

The regulatory system consists of two components. The first one is the *histine kinase EnvZ*, a particular kind of protein having the role to transmit information, adding and removing a phosphate to an aspartame acid, usually on the other component of the signalling pathway, the *response regulator OmpR*, which mediates a response of the cell to changes in its environment. The role of *EnvZ*

Table 1: The initial concentrations, the rates and the chemical reactions of *EnvZ/OmpR* system. The concentration of X and Y, marked by the symbol $\diamond$, can vary to prove the robustness in Y_P.

Initial concentrations	Rates	Chemical reactions
$X = 25\ \diamond$	$k_1, k_2, k_3, k_4 = 0.5$	$XD \underset{k_2}{\overset{k_1}{\rightleftharpoons}} X$
$Y = 150\ \diamond$	$k_5, k_{11} = 0.1$	$XT \underset{k_4}{\overset{k_3}{\rightleftharpoons}} X$
$XT = 0$	$k_6, k_9 = 0.02$	$XT \xrightarrow{k_5} X_P$
$X_P = 0$	$k_7, k_8, k_{10} = 0.5$	$X_P + Y \underset{k_7}{\overset{k_6}{\rightleftharpoons}} X_PY$
$X_PY = 0$		$X_PY \xrightarrow{k_8} X + Y_P$
$Y_P = 10$		$XD + Y_P \xrightarrow{k_9} XDY_P$
$XDY_P = 0$		$XDY_P \xrightarrow{k_{10}} XD + Y_P$
$XD = 50$		$XDY_P \xrightarrow{k_{11}} XD + Y$

is bifunctional because it phosphorylates and dephosphorylates *OmpR*: the model predicts that when *EnvZ* is much less abundant than *OmpR*, or when the concentration of this species is sufficiently high, the steady state level of $OmpR_P$ (the phosporylated form of *OmpR*) is insensitive to variations in the concentration of *Envz* and *OmpR*.

4.2.1.1 *Modeling and simulation of the EnvZ/OmpR system in E.coli.*

The main components of this chemical network are *EnvZ* and *OmpR* [12, 71], denoted in Table 1 respectively as X and Y. *Envz* phosporylates *OmpR* (Y_P) and itself (X_P), by binding and breaking down *ATP*. In this sequence of chemical reactions, in fact, *ATP* and *ADP* act as cofactor (denoted as T and D).

In order to check whether the system satisfies our definition of robustness we build the Petri nets model shown in Figure 18, where X and Y are considered as input and Y_P as output.

To study the equilibrium configuration, we need to calculate analytically the steady state by setting the time-derivative of Y_p to zero. Then, considering the

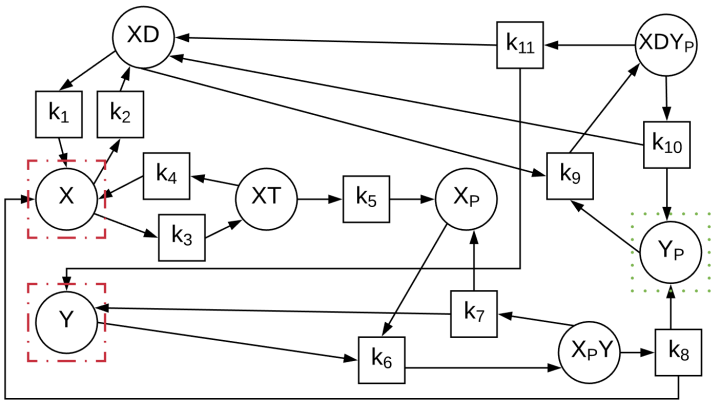

Figure 18: The Petri nets model for the reaction network of the *EnvZ/OmpR* system. The input of the network are X and Y(red dot line), the output is the concentration of Y_P(green dots).

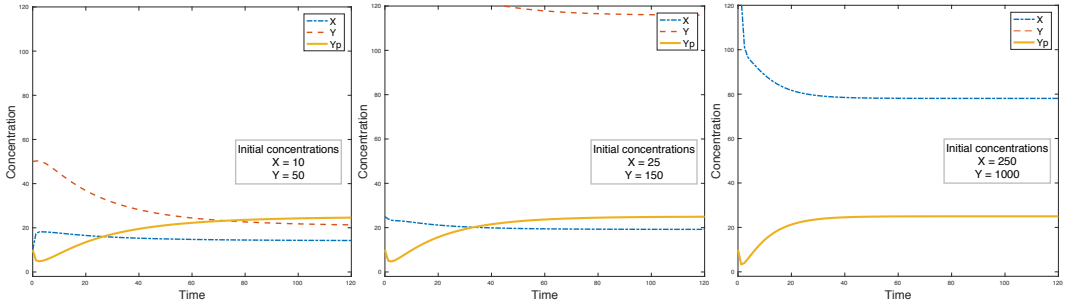

Figure 19: Graphical results of the simulation of the *EnvZ/OmpR* system. We vary the concentrations of X and Y to show robustness in Y_P. Note that in the third case the curve of Y is out of the graph.

reactions described in Table 1, we apply the law of mass action to obtain the set of differential equations representing the system, as follows:

$$
\begin{cases}
\frac{d[X]}{dt} = k_1[XD] - k_2[X] + k_3[XT] - k_4[X] + k_8[X_pY] \\
\frac{d[Y]}{dt} = k_7[XpY] - k_6[X_p] + k_3[XT] + k_{11}[XDY_p] \\
\frac{d[XT]}{dt} = k_4[X] - k_3[XT] - k_5[XT] \\
\frac{d[X_p]}{dt} = k_5[XT] - k_6[X_p][Y] + k_7[X_pY] + k_8[X_pY] - k_{10}[XDY_p] \\
\frac{d[X_pY]}{dt} = k_6[X_p][Y] - k_7[X_pY] - k_8[X_pY] \\
\frac{d[Y_p]}{dt} = k_8[X_pY] - k_9[XD][Y_p] + k_{10}[XDY_p] \\
\frac{d[XDY_p]}{dt} = k_9[XD][Y_p] - k_{10}[XDY_p] - k_{11}[XDY_p] \\
\frac{d[XD]}{dt} = k_2[X] - k_1[XD] - k_9[XD][Y_p] + k_{10}[XDY_p] + k_{11}[XDY_p].
\end{cases}
\tag{11}
$$

Hence, we calculate the steady state as follows:

$$\frac{d[Yp]}{dt} = 0$$
$$k_8[X_pY] - k_9[XD][Y_p] + k_{10}[XDY_p] = 0$$
$$k_8[X_pY] + k_{10}[XDY_p] = k_9[XD][Y_p]$$
$$[Y_p]_{ss} = \frac{k_8[X_pY] + k_{10}[XDY_p]}{k_9[XD]}$$

At the steady state, the concentration of Y_P does not depend on the input chemical species, thus, the system satisfies 0-robustness (absolute concentration robustness) for the widest intervals ($[1, \infty]$) of initial concentrations. Obviously, applying the relative α-robustness, described in Definition 21, we obtain that the CRN shows β-robustness with $\beta = 0$.

To illustrate the robustness of this system we show some simulation results obtained by using Dizzy [61]: a simulator of chemical reactions. Simulation results are in Figure 19, where it is shown that the concentration of Y_p is constant even varying the initial concentrations of the input species X and Y.

Moreover, note that in this case, we can also apply Theorem 2: the deficiency of the network is 1 and the sufficient conditions required by the theorem to assure robustness in Y_p can be verified (see [71] for details).

4.2.2 *The isocitrate dehydrogenase regulatory system*

In the literature, metabolic and regulatory pathways that contain multifunctional proteins, as the *Envz/OmpR* system described above, have frequently been observed to exhibit robustness, thanks to their ability to perform their tasks even in presence of internal and external perturbations. Among several examples, there is the isocitrate dehydrogenase regulatory system (*IDHKP-IDH)* of E. coli [25].

This system controls the partioning of carbon flux and it is useful when the bacterium of E. coli grows on substances, like for example acetate, which contains only a small quantity of carbon. Without this regulation system, in fact, the organism would not have enough carbon available for biosynthesis of cell constituents [71].

Table 2: The initial concentrations, the rates and the chemical reactions of *IDHKP-IDH* system. The concentration of E and I_P, marked by the symbol $\diamond$, can vary to prove the robustness in I.

Initial concentrations	Rates	Chemical reactions
$E = 0.001\ \diamond$	$k_1, k_4 = 0.02$	$E + I_P \underset{k_2}{\overset{k_1}{\rightleftharpoons}} EI_P$
$I=100$	$k_2, k_3, k_5, k_{11} = 0.5$	$EI_P \xrightarrow{k_3} E + I$
$I_P = 10000\ \diamond$	$k_6 = 0.1$	$EI_P + I \underset{k_5}{\overset{k_4}{\rightleftharpoons}} EI_PI$
$EI_P = 10$		$EI_PI \xrightarrow{k_6} EI_P + I_P$

4.2.2.1 *Modeling and simulation of the IDHKP-IDH system.*

The isocitrate dehydrogenase regulatory system works regulating the phosphorylation level of the TCA cycle enzyme isocitrate dehydrogenase (*IDH*), denoted as I in Table 2. The TCA cycle enzyme is a series of chemical reactions used by different microorganisms to produce energy. The protein I, in its active form, has the role of regulating how much carbon will flow through the system, while it is inactive in its phosporylated form I_P. The enzyme E is bifunctional: it phosporylates and dephosporylates I.

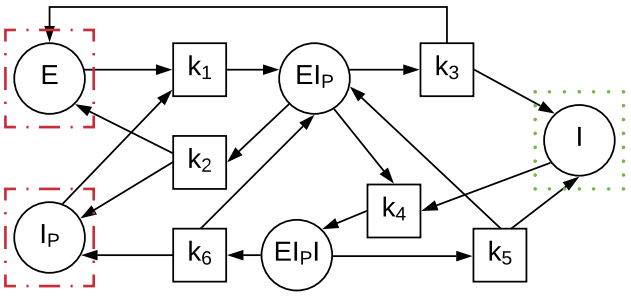

Figure 20: The Petri nets model for the reaction network of the *IDHKP/IDH* system. The input of the network are E and I_P(red dot line), the output is the concentration of I(green dots).

We build the Petri net model of the chemical networks and we identify I_P and E as the input of the model and I as the output, as shown in Figure 20. By studying the equilibrium configuration, we find that at the steady state the concentration of I is completely independent on the concentration of I_P and E.

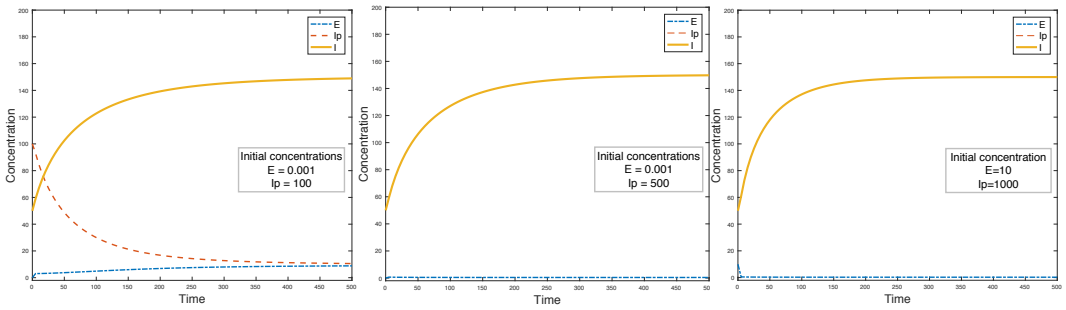

Figure 21: Graphical results of the simulation of the *IDHKP-IDH* system. We change the concentration of E and I_P to test robustness in I.

Thus, the system shows robustness in this species: choosing a wide range of possible initial concentrations for the input, we obtain a constant value for I, as it is possible to notice from the simulation results shown in Figure 21. Therefore, we verified 0-robustness for the widest interval range of initial concentrations of the input species. As in the case of the EnvZ/OmpR system, we have that Theorem 2 could be applied to prove absolute concentration robustness of this system, and the relative β-robustness is not significant.

4.2.3 *Bacterial chemotaxis*

In nature, one of the most important examples for robustness property is the bacterial chemotaxis. Bacterial chemotaxis is the process in which bacteria sense and move along gradients of specific chemicals, like for example sugar or amino acids (as serine and aspartame) [1]. Despite the physical limitations faced by the bacteria, these organisms can detect if the distribution of molecules of attractant (or repellents) changes in the environment and they use this information to guide their motion, composed of *runs*, in which the bacteria keep a constant direction, and *tumbles* randomly changing direction. The bacterium compares the current attractant concentration to the concentration in the past and, if it detects a positive increment, this reduces the tumbling frequency. After a while, even if the attractant is still present in the environment, the tumbling frequency increases and returns to the same level as before the attractant was added. This phenomenon is an example of *exact adaptation*,

because the amount of attractant (or repellents) concentration do not influence the bacterium response to the ambient change.

4.2.3.1 *Modeling and simulation of chemotaxis of E. coli.*

Table 3: The initial concentrations, the rates and the chemical reactions of chemotaxis phenomenon of the E. coli. The concentration of the attractant L, marked by the symbol $\diamond$, can vary to prove the robustness in $CheY_P$.

Initial concentration	Rates	Chemical reaction
X = 10	$k_1, k_{14} = 1.15$	$X \underset{k_2}{\overset{k_1}{\rightleftharpoons}} X^*$
X^* = 10	$k_2, k_{13} = 0.25$	$X^* + CheY \xrightarrow{k_3} CheYp + X$
L = 0 $\diamond$	$k_3 = 0.1$	$CheYp + Z \xrightarrow{k_4} CheY + Z$
CheY = 10	$k_4 = 10$	$CheYp \xrightarrow{k_5} CheY$
Z = 1	$k_5 = 0.002$	$L + X^* \xrightarrow{k_6} L + XY$
CheYp = 1	$k_6 = 10000$	$CheBp \xrightarrow{k_7} CheB$
XL = 0	$k_7, k_9, k_{10}, k_{12} = 1$	$XY \xrightarrow{k_8} XL$
X^*m = 1	$k_{11} = 0.08$	$CheR + XL \xrightarrow{k_9} X^*m + CheR$
CheR = 1000	$k_{15} = 0.18$	$X^*m + CheB \xrightarrow{k_{11}} Bp + X^*m$
XY = 0		$X^*m + CheBp \xrightarrow{k_{12}} X^* + Bp$
CheB = 2		$X^*m \underset{k_{14}}{\overset{k_{13}}{\rightleftharpoons}} Xm$
CheBp = 0		$X^*m + CheY \xrightarrow{k_{15}} CheYp + Xm$
Xm = 0		

From the literature [1, 10] we know that E. coli bacterium has receptors, each of them bounded to a protein kinase, constituting a group generally called X. Rapidly, this group can pass from inactive (X) into active (X^*) state, which modifying, in turn, the state of a regulator protein diffused in the cell, $CheY$, which becomes $CheY_P$ with the addition of a phosphoryl group. The protein $CheY$ is the main responsible of tumbles: in fact, higher is the $CheY_P$ concentration, higher is the tumbling frequency.

During this process, the binding between X and other molecules (attractants) reduces its probability to reach the active state that, consequently, reduces also

the probability to attach a phosporyl group to *CheY*, which has the effect of lowering the tumbling frequency.

Since the attractants (L) reduce the activity of X, there is the *methylation* mechanism to activate again the chemical group. An enzyme *CheR* adds at constant rate a methyl group to X, which becomes X_m and when it is in its active form, X^*_m, it increases the activity of the cell and the tumbles, because it influences directly the concentration of $CheY_p$. Continually, the methyl group

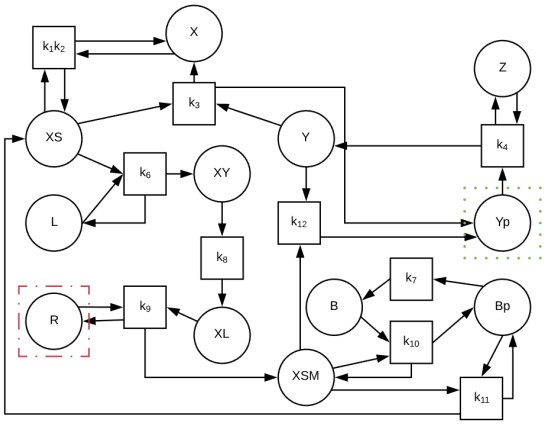

Figure 22: The Petri nets model for the reaction network of the bacterial chemotaxis network. The input of the network are L (red dot line), the output is the concentration of Yp (green dots).

is removed by the enzyme *CheB*, which is influenced by X that, adding a phosphoryl group to *CheB*, makes it more active, constituting a *negative feedback loop*: higher is the activity of X, higher is the activity of phosporylated *CheB*, which on the contrary reduces the activity of X. The exact adaptation is achieved because of the feedback circuit: the increased methylation of X precisely balances the reduction in activity caused by the attractant.

In Table 3 we summarize the chemical reactions of the chemotaxis network, its respective rates and the initial concentrations.

Applying Definition 16, we build the Petri nets of the reaction network, as in Figure 22, we find that calculating the steady state for $CheY_p$, it does not depend on L, hence we obtain an absolute concentration robustness in these species. Thus, for our Definition 16, we obtain $\alpha = 0$, because even changing the input of L, it does not influence the output of the network.

Moreover, we implement the described model of chemotaxis of E. coli [1], using the chemical kinetics stochastic simulation software Dizzy. In particular, to model how *CheR* works at saturation, as described above, we add some supplementary reactions and chemical species (as XL and XY), to limit the reaction speed of *CheR*, because it is not possible to use a constant rate, since the software Dizzy is based on mass action law.

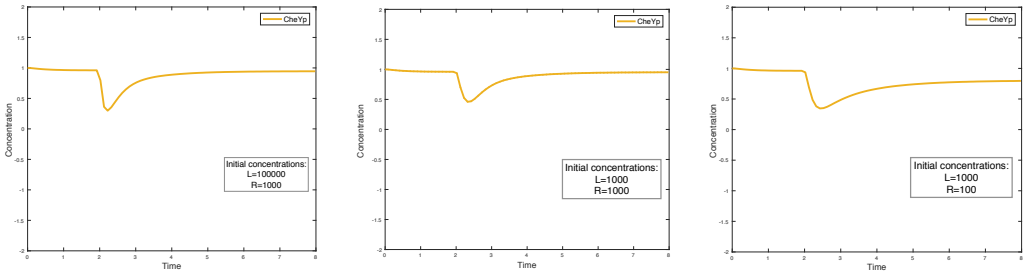

Figure 23: Graphical results of the simulation of the bacterial chemotaxis. To show how robustness is preserved, we change the concentration of the attractant L, to study how this influences $CheY_P$.

We simulate the chemical reactions network to verify the robustness, changing the initial concentration of the attractant. As shown in Figure 23, even modifying the input of L, after a reduction in concentration, $CheY_P$ returns exactly to the initial value, and this means that the tumbling frequency recurs again at the same level as before the addition of the attractant. As described in [1], the bacterial chemotaxis could not show exact adaptation, changing some parameters of the mathematical model. In Figure 23, in fact, varying the concentration of the enzyme *CheR*, the concentration of $CheY_P$ does not return to the initial value and the system turns out to be α-robust with $\alpha = 0.3$. Considering the input and the output of the CRN, we can apply the relative α-robustness, and it turns out that the CRN is robust for $\beta = 0.35$.

Applying Theorem 2, we find that the deficiency is 3, hence the theorem is not proved, but with our definition, we are able to describe the adaptation and the absolute robustness of the bacterial chemotaxis.

Table 4: The initial concentrations, the rates and the chemical reactions of enzyme activity at saturation model. The concentration of P, marked by the symbol $\diamond$, can vary to prove the robustness in X.

Initial concentrations	Rates	Chemical reactions
R = 1000	$k_1 = 100$	$R + X \xrightarrow{k_1} X + X + Z$
X = 30	$k_2 = 10$	$X \xrightarrow{k_2} W$
Z = 0	$k_3 = 0.5$	$Z \xrightarrow{k_3} R$
P = 1 $\diamond$	$k_4 = 0.01$	$X + P \xrightarrow{k_4} P + P$
C = 0	$k_5 = 0.5$	$P \xrightarrow{k_5} C$
W = 0		

4.2.4 *Enzyme activity at saturation*

The well-known Lotka-Volterra reactions [49, 51] can be interpreted as abstract chemical reactions and, in fact, they have been proposed to investigate the oscillatory dynamics of autocatalytic enzymes. Similarly, the logistic equation [76] is a model of population growth that is commonly used also in the context of biochemical reaction kinetics. It describes the growth of a population by taking the amount of available environmental resources into account (the *carrying capacity* of the environment) and it is used also to model enzyme dynamics at saturation. In this Section we consider an abstract model of enzyme activity inspired by the Lotka-Volterra reactions and the logistic equation.

4.2.4.1 *Modeling and simulation of enzyme activity at saturation model.*

We consider an abstract chemical reaction network in which an enzyme R produces a molecule X. To guarantee the mass conservation, we add to this idealized example the species Z, which has the role to preserve the concentration of R.

The production of X is autocatalytic (the more X are present, the higher is the production rate), but the concentration of enzymes R is limited. Hence, the enzyme activity can easily reach saturation. This reaction system is of the kind typically modeled by the logistic equation. It is expected to reach a dynamic equilibrium in which the concentration of X does not depend on its initial concentration, but only on the concentration of R. We add to this system an

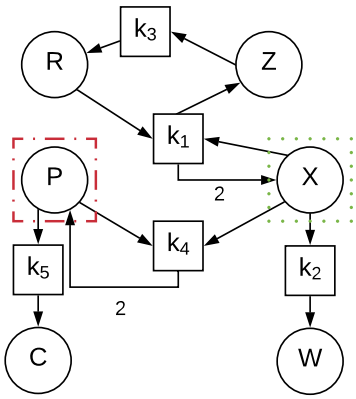

Figure 24: The Petri nets model for enzyme activity at saturation system. The input of the network is P (red dot line), the output is X (green dots).

additional molecular species P acting as a "predator" for X (as in Volterra's equations). What happens is that X can be consumed and transformed into P, and the reaction performing this action is autocatalytic (i.e., stimulated by P itself). In this model it can be interesting to investigate how the initial concentration of P influences the steady state concentration of X.

Building the Petri nets model of the reactions network, as shown in Figure 26, we identify P as the input and X as the output of the network. At the steady state, it emerges that the concentration of X is always constant and its constant value only loosely depends on the concentration of P. Indeed, even varying the concentration of the molecular species P in a wide interval, the concentration of X at the steady state assumes a value in a very small interval. It is worth noting that this kind of robustness of the system was not captured by the previous definitions presented in the literature. Instead, our definition is able to express not only absolute concentration robustness but also weaker levels of robustness by tuning the α parameter and the amplitude of the intervals of the input species.

In this case, to apply Definition 16, we choose an interval marking for $P = [1, 1000]$ and we find, by the means of simulations, that the concentration of X is in the range $[50, 47]$, as shown in Figure 25. Therefore, the system is α-robust with $\alpha = 3$. Computing the relative α-robustness, the system is robust with $\beta = 0.03$.

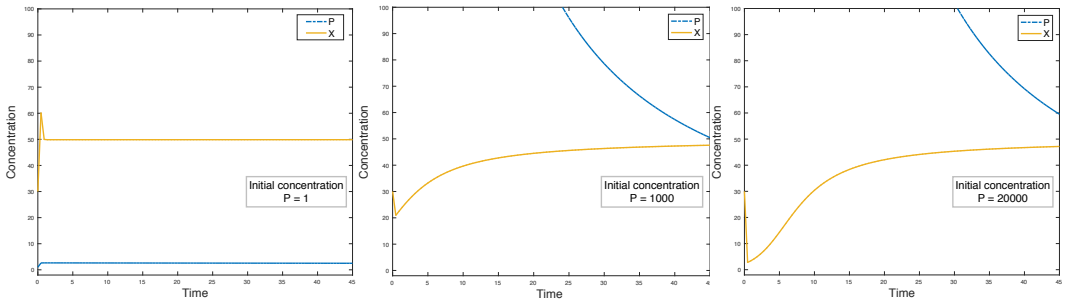

Figure 25: Graphical results of the enzyme activity at saturation model. We change the concentration of the P to test robustness in X.

4.3 INITIAL CONCENTRATION ROBUSTNESS IN THE GENERAL RIZK'S FRAMEWORK

Our definition of robustness focuses purposely on the initial concentrations. This specific problem is analyzable in the more general context presented in [64], already shown in Chapter 3.

As explained before, the methodology proposed by Rizk et al. in [64, 65] describes the ability of a system to maintain specific functionalities against perturbations. The robustness of a system is measured as the *distance* of the system behaviour under perturbations from its reference behaviour expressed as a temporal logic formula. The distance is computed by using a notion of *violation degree* measuring how much the temporal logic formula should be changed in order to match traces of perturbed behaviours obtained, for instance, through simulations.

The approach proposed by Rizk et al. is very general, both in the description of the reference behaviour and as regards perturbations. On the contrary, in our work, we focus on *concentration robustness*, namely on the influence of the initial concentrations of species on what will be the steady state of the system. What we proposed with α-robustness is a notion, which extends the notion of *absolute concentration robustness* considered in [70, 71].

Our definition of robustness is simpler and much less general than the one considered by Rizk et al.. However, it is conceived with the aim of enabling further studies on sufficient conditions that could allow robustness to be assesses by avoiding (or significantly reducing) the number of simulations to be per-

formed. This could be obtained, for instance, by adapting conditions already considered in the context of monotonicity analysis [4], as we will see in the next chapters.

4.3.1 *Initial concentration robustness as an LTL formula*

In [64], as shown in Chapter 3, a given behavior concerning a given set of perturbations P is evaluated considering a numerical trace, describing the system's evolution with time. Instead, as described by our Definition 16, our approach focuses on the evaluation of initial concentration robustness: we vary the concentration of at least one chemical species (input) and verifying, at the steady state, if the concentration of an output species is included in an interval. In line with that, the interval marking, defined in 15, is equivalent to P.

To match the two frameworks, we restrict the analysis of the system behavior at the equilibrium. We show that Definition 16 is an instance of Definition 12, expressed using the Rizk et al. framework shown in [64].

Theorem 5 (Definition 16 in the context of Definition 12)**.** *Given a Petri Net* PN *with output place* O*, an initial marking* $m_{[\,]} \in M_{[\,]}$ *and a continuous probability distribution* $\mathrm{prob}(m)$*, defined on* $m_{[\,]}$ *such that the integral of the pdf is normalized to 1.* PN *is* α*-robust w.r.t. to* $m_{[\,]}$ *iff there exists an interval* $[min, max] \in \mathbb{R}$ *such that* $R^s_{\phi,P} = 1$*, with* $\phi = F(G([O] \geqslant min \wedge [O] \leqslant max))$*,* P *equivalent to* $m_{[\,]}$ *and* $max - min = \alpha$*.*

Proof. Let us consider Definition 12:

$$R^s_{\phi,P} = \int_{p \in P} \mathrm{prob}(p)\, sd(T_p, \phi)\, dp.$$

Since the set of perturbations P is equivalent to the initial marking $m_{[\,]}$, as expressed in Theorem 5, the definition becomes:

$$R^s_{\phi,m_{[\,]}} = \int_{m \in m_{[\,]}} \mathrm{prob}(m)\, sd(T_m, \phi)\, dm,$$

in which, by Definition of sd 11 is:

$$R^s_{\phi,m_{[\,]}} = \int_{m \in m_{[\,]}} \mathrm{prob}(m) \frac{1}{1+vd(T_m,\phi)}\, dm.$$

By Definition of vd 10, we obtain:

$$R^s_{\phi,m_{[\,]}} = \int_{m \in m_{[\,]}} \mathrm{prob}(m) \frac{1}{1+dist(\phi, D_{T_m,\phi(y)})}\, dm$$

In order to prove the implication $\implies$ of the iff in the statement of the theorem, we assume that PN is α-robust and we have to show that there exists an interval $[min, max]$ for which $R^s_{\phi, m_{[\,]}} = 1$.

Assuming that PN is α-robust, hence $\exists\ k \in \mathbb{R} \cdot\ \forall\ m \in m_{[\,]} \cdot\ m_{ss}[O] \in [k - \frac{\alpha}{2}, k + \frac{\alpha}{2}]$. We choose $min = k - \frac{\alpha}{2}$ and $max = k + \frac{\alpha}{2}$. With these values for min and max, considering all the possible perturbations m, the distance between the formula ϕ and the domain $D_{T_m, \phi(y)}$ is always 0 because the domain $D_{T_m, \phi(y)}$ is defined as $D_{T_m, \phi(y)} = \{y_1 \leqslant m_{ss}[O] \wedge y_2 \geqslant m_{ss}[O]\}$. If the distance is equal to 0, $vd(T_m, \phi) = 1$, hence $sd(T_m, \phi) = 1$. Since the function $prob(m)$ is a continuous probability distribution, where $\int_{m \in m_{[\,]}} prob(m) dm = 1$, we obtain:

$$R^s_{\phi, m_{[\,]}} = 1.$$

In order to prove the opposite implication, we assume there exists an interval $[min, max] \in \mathbb{R}$ and we want to show that it is α-robust with $\alpha = max - min$, then $R^s_{\phi, m_{[\,]}} = 1$.

This assumption implies $sd = 1$ for all perturbations, and, going backwards, $vd = 0$. If the distance between the formula ϕ and the domain $D_{T_m, \phi(y)}$ is equal to 0, it means that the formula ϕ is satisfied, hence, at the steady state, the output concentration $[O]$ is included in an interval $[min, max]$, where its range is $\alpha = max - min$. Thus, considering $k = min + \frac{\alpha}{2}$, the system PN is α-robust for $[k - \frac{\alpha}{2}, k + \frac{\alpha}{2}]$. □

In [19], it is presented BIOCHAM (BIOCHemical Abstract Machine), a software environment for modeling biochemical systems, where it is already implemented the function *robustness* as mathematically described in Section 3.3.1. The implementation of function `robustness`, in BIOCHAM, to test this property at the steady state is the following :

Listing 1: Function robustness implemented in BIOCHAM

```
1   robustness(F(G([O] >= min /\ [O] <= max)), [In], [min -> x, max ->
        y])
```

where

- `robustness` is the function, implemented in BIOCHAM [19], computing the robustness measure;

- `In` is the concentration of the input species, which is perturbed;
- `O` is the concentration of the output species;
- `F` and `G` are two specific temporal operators, respectively meaning *future* and *globally*;
- the expression `[min -> x, max -> y]` represents the assignment of the interval limits;
- the expression `F(G([O] >= min ∧ [O] <= max))` represents the system behavior that has to be evaluated. The formula expresses that, when the system reaches the steady state, the concentration of output species is within a range.

To verify Definition 16 is an instance of Definition 12, we apply the LTL formula of initial concentration robustness described in Section 3.3.1 to the same examples shown in Section 4.2.

4.3.1.1 *Examples of equivalence between robustness definitions*

We show step by step the equivalence between Definition 16 and the general notion of robustness in Definition 12, considering Example 4.2.4, explained in Section 4.2.

The first step is to build the Petri Net of the CRN, choosing one or more places as the input of the network and one as the output. As shown in Figure 26, the molecular species P and X are selected as the input and the output respectively. Hence, we perturbed the initial concentration of the input, to verify if, at the steady state, the output concentration is included in an interval $[min, max]$.

We choose an interval marking for the Petri Net: let $PL = \{P, R, Z, C, W, X\}$ be the set of places, we vary only the initial concentration of the species P, which we choose as the input of the CRN, leaving the others equal to 0 or constant. Performing all the possible simulations, changing the initial value of P, we find out that the concentration of species X, the output of the CRN, at the steady state is always included in the interval $[47, 50]$. Thus, considering Definition 16, the system is α-robust for $\alpha = 3$.

At this point, to match Definition 16 with Definition 12 we calculate the domain $D_{T_m, \phi(y)}$, analysing the numerical trace depicting the time evolution of

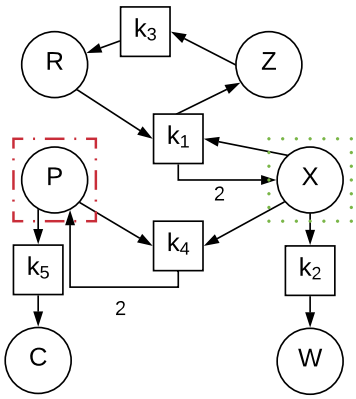

Figure 26: The Petri nets model for enzyme activity at saturation system. The input of the network is *P* (red dot line), the output is *X*(green dots).

the species X and considering all the possible perturbations of P. We evaluate the formula $\phi = (F(G([X] \geqslant 0 \wedge [X] \leqslant 61)))$, where the values 0 and 61 are the interval limits, found by simulations, and we proceed to calculate the violation degree and the satisfaction degree. The $vd = 0$, because the distance between the domain and the formula ϕ is null, hence the value of the satisfaction degree is $sd = 1$, because the formula is totally satisfied. Now, we can apply the robustness formula 12 that is $R^s_{\phi,m_{[\,]}} = 1$. Then, we obtain:

Listing 2: Function robustness implemented in BIOCHAM

```
robustness(F(G([X] >= min /\ [X] <= max)), [P], [min -> 0, max -> 61])
```

For the other examples, presented in Section 4.2, we validate the match between the two Definitions, using only BIOCHAM:

- *EnvZ/OmpR* 4.2.1: in this model, the inputs are X and Y and the output is Y_p. In BIOCHAM, we calculate the validity domain of the output to find the minimum (min) and the maximum (max) values. Then, we compute:

Listing 3: Function robustness implemented in BIOCHAM

```
robustness(F(G([Yp] >= min /\ [Yp] <= max)), [X, Y], [min ->
    0.560474, max -> 25])
```

We find that the robustness is equal to 1.

- *IDHKP-IDH* 4.2.2: the output of this example is I, the inputs species are E and I_p. Then, we compute:

Listing 4: Function robustness implemented in BIOCHAM

```
robustness(F(G([I] >= min /\ [I] <= max)), [E, Ip], [min ->
    96.1, max -> 112])
```

we obtain that the robustness is equal to 1.

- *Bacterial chemotaxis* 4.2.3: R and Y_p are respectively the input and the output of this system. then, we compute:

Listing 5: Function robustness implemented in BIOCHAM

```
robustness(F(G([Yp] >= min /\ [Yp] <= max)), [R], [min -> 0.9,
    max -> 1])
```

As for the examples before, also in this case the robustness is equal to 1.

5 MONOTONICITY IN CHEMICAL REACTION NETWORKS

The dynamical behavior of a biological network is complex to investigate, in particular, when a system consists of multiple components, interacting with each other through signaling pathways. As introduced in Chapter 2, one of the possible approaches is performing simulations based on the Ordinary Differential Equations (ODEs) models of the kinetics of the reaction. Besides, the exact parameters are often not well-known. Then, we need to investigate the system performing many simulations and by examining all the possible combinations of the chemical species concentrations.

One of the possible approaches to reduce the computational cost of simulations is to find some structural and dynamical characteristics that can detect internal properties of the system, such as robustness and monotonicity. Indeed, in this way, we could obtain qualitative results about the system, without making assumptions on the structure of ODEs involved in the CRN.

In the first part of this Chapter, we focus on the monotonicity in chemical reaction networks, showing the works presented in [4, 5, 24]. In [24] the principal aspects of monotone systems theory are presented, starting from the given definition developed by M. W. Hirsch. In [4, 5], instead, it is proposed a new approach to study the monotonicity in chemical systems. Using a graphical representation, the authors are able to investigate global monotonicity property analyzing only the structure of the dynamical systems. In this context, indeed, it is possible to analyze the qualitative features of the CRN, without any simulations.

In our case, as described in Chapter 4, we are interested in the verification of the initial concentration robustness, in which the possible perturbations affect the initial concentrations of the chemical species. To verify this property, we need to simulate the system for all the possible initial values of the input (which is the perturbed chemical species). Thus, in this context, we show how to reduce drastically the number of simulations, studying the CRN structure. More in detail, in the second part of this Chapter, we propose a graphical approach that is able to predict if the concentration of the output at the steady

state is monotonic with respect to the initial concentration of the input. If this property is verified, we are able to simulate the model only on the extreme values of the input concentration range.

In the end of this Chapter, we verify our sufficient condition of Input - Output monotonicity to two examples:

- Michaelis-Menten kinetics;
- ERK signalling pathway.

5.1 BASIC DEFINITIONS OF MONOTONICITY IN DYNAMICAL SYSTEMS

A dynamical system, such as a chemical reaction network, is monotone if the flow originated by the differential equations preserves a partial order [24].

More in detail, a monotone dynamical system of the form $\dot{x} = f(x)$, with $f : X \longrightarrow \mathbb{R}^n$, is a continuous semiflow Φ on a closed set $X \subset \mathbb{R}^n$ equipped with a compatible partial order $\succeq$, such that the partial order is preserved by the flow:

$$\forall x, y \in X : x \succeq y \implies \Phi_t(x) \succeq \Phi_t(y), \forall t \in \mathbb{R}_+.$$

As defined in [5], we assume that the partial order $\succeq$ defined on X has the following axioms:

- Reflexivity: $\forall x \in X$, $x \succeq x$;
- Transitivity: $\forall x_1, x_2, x_3 \in X$, $x_1 \succeq x_2$ and $x_2 \succeq x_3 \implies x_1 \succeq x_3$;
- Antisymmetry: $\forall x_1, x_2 \in X$, $x_1 \succeq x_2$ and $x_2 \succeq x_1 \implies x_1 = x_2$.

The partial order is closed, hence if $x_n \longrightarrow x$ and $y_n \longrightarrow y$ as $n \longrightarrow \infty$ and $x_n \succeq y_n$ for all n, then also $x \succeq y$. Geometrically, we can define the partial order on a convex cone $K \subset \mathbb{R}^n$, where K is a nonempty closed set with $K + K \subset K$, $\mathbb{R}_+ K \subset K$ and $K \cap (-K) = \{0\}$. We say that $x_1 \succeq x_2$ iff $x_1 - x_2 \in K$.

A system is *monotone* if for all $x_1 \succeq x_2$ and all $t \geqslant 0$, $x(t, x_1) \succeq x(t, x_2)$, where $x(t, x_i)$ represents the solution at time t with initial condition x_i and it is *strongly monotone* if all $x_1 \succ x_2 \implies x(t, x_1) \gg x(t, x_2)$ for all $t > 0$.

If the partial order is defined on a positive orthant, $K = \mathbb{R}^n_{\geqslant 0}$, to check the monotonicity is equivalent to verify that the Jacobian matrix $\frac{\partial f}{\partial x}$ has nonnegative off-diagonal entries, inspecting that the signs of entries never change sign.

In other cases, to check monotonicity is nontrivial. In literature, one of the most interesting results is the work presented in [5], where the authors show how to check global monotonicity in CRN, using a graphical approach.

5.1.1 *Graph-theoretic characterizations of global monotonicity in CRN*

The qualitative behavior of chemical reaction networks can be very difficult to understand, first of all because of the complexity of biological systems under study, usually described by a set of non-linear differential equations. For this reason one of the most common approach is to try to have some information about the dynamical behavior of the system, studying only its structure and without the entire execution of the CRN. This approach is common in Systems biology, as described also in Sections 3.4.1 and 3.4.2, where we introduce the Feinberg's theorems about deficiency.

In line with what has been said, in [4, 5] it is presented a theorem able to investigate the structural monotonicity of a chemical reaction network, using a graphical approach, providing easily verifiable conditions concerning asymptotic dynamics.

Before explaining in detail the sufficient condition, we proceed recalling some assumptions, crucial for the application of the method in [4, 5], recalling the definition of chemical reaction networks in Section 2.1.3:

- *Exclusion of auto-catalytic reactions*, that are reactions in which a chemical species appears both as reactant and as product.
- *The reaction rate depends monotonically on the concentrations of the species*, hence increasing the concentrations of chemical species, it increases also the rate of the perturbed reaction. In particular for irreversible reactions, we assume:

$$\frac{\partial R_i(S)}{\partial S_j} = \begin{cases} \geqslant 0 & \text{if} \quad \alpha_{ij} > 0 \\ = 0 & \text{if} \quad \alpha_{ij} = 0 \end{cases}$$

For reversible reactions, the assumption is:

$$\frac{\partial R_i(S)}{\partial S_j} = \begin{cases} \geqslant 0 & \text{if} \quad \alpha_{ij} > 0 \quad \text{and} \quad \beta_{ij} = 0 \\ \leqslant 0 & \text{if} \quad \beta_{ij} > 0 \quad \text{and} \quad \alpha_{ij} = 0 \\ = 0 & \text{if} \quad \alpha_{ij} = 0 \quad \text{and} \quad \beta_{ij} = 0 \end{cases}$$

- *Absence of a chemical species stops its reaction*: if any of the reactants of an irreversible or reversible reaction are missing, then the corresponding reaction does not take place;
- *Systems are described by differential equations*, as already defined in Section 2.1.3.
- *All stoichiometry classes are compact sets*, which implies that the solution of the system are bounded.

Before presenting the theorem in [4, 5], we need to give some auxiliary definition.

The first step to apply the theorem is to represent the CRN as a *species-reaction graph*, known shortly as *SR-graph*: a bipartite graph in which each node is associated to a chemical species or a reaction involved in the system, and the edges, linking species and reaction, describe what is the behavior of the species towards the respective reaction. If the species is a reactant, it helps the reaction, hence between the species and the reaction there is a positive edge. Otherwise, if the species is the product of the reaction, they are linked by a negative edge. Mathematically, the SR-graph is defined as follows:

Definition 1 (SR-graph)**.** *Given a finite set of reactions* R *over a set of species* S*, the associated SR-graph is defined by the quadruple* $\langle V_S, V_R, E_+, E_- \rangle$*, where* V_S *is a finite set of nodes, each one associated to a species,* V_R *is the set of nodes, disjoint from* V_S*, representing the reactions (either irreversible or reversible: in the latter case, the forward and backward reactions are taken into account only once in the graph). Whenever a certain reaction* R_i *belongs to the network:*

$$\sum_{j \in S} \alpha_{ij} S_j \rightleftharpoons \sum_{j \in S} \beta_{ij} S_j$$

the relations $E_+ \subseteq (V_S \times V_R)$ *and* $E_- \subseteq (V_S \times V_R)$ *are defined as follows:*

- $(S_j, R_i) \in E_+$ *if* $\alpha_{ij} > 0$*, with* $S_j \in V_S$ *and* $R_i \in V_R$*;*

- $(R_i, S_j) \in E_-$ *if* $\beta_{ij} > 0$, *with* $R_i \in V_R$ *and* $S_j \in V_S$;

Intuitively, we draw a positive edge between every node $S \in V_S$ and every node $R \in V_R$ if S is a reactant of the reaction R and, thus it contributes to the reaction. On the contrary, we draw a negative edge between every node $R \in V_R$ and every node $S \in V_S$, because that species is not "helping" the reaction.

In this context, we can notice that when we are modeling a reversible reaction, we can choose one of the two possible orientations because the product of the sign of the edges is equivalent.

To make an example of how to build the SR-graph, we consider the well-known Michaelis–Menten kinetics, which models enzyme kinetics:

$$E + S \underset{k_2}{\overset{k_1}{\rightleftharpoons}} ES \xrightarrow{k_3} E + P \tag{12}$$

where the enzyme E, binding the substrate S, forms a complex ES, which releases the product P regenerating the original enzyme.

In the chemical reaction network in 12, we have four node-species that are E, S, ES and P. We consider only one orientation in reversible reactions, hence we have only two node-reactions, labelled with R_1 and R_2: $E + S \xrightarrow{R_1} ES$ and $ES \xrightarrow{R_2} E + P$. The obtained SR-graph is shown in Figure 27. As we can notice, the SR-graph is graphically equivalent to a Petri net. In both formalisms, we can express the relations among species and reactions.

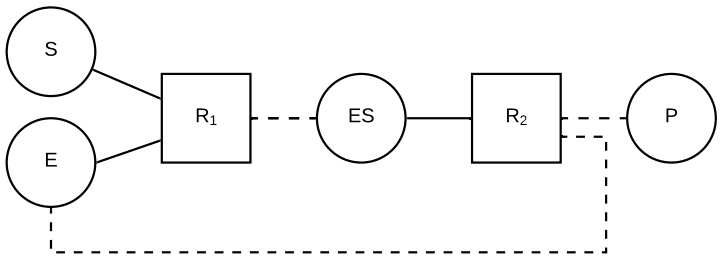

Figure 27: SR-graph of the chemical reaction network 12, representing the Michaelis-Menten kinetics. In this graph, the circles represent the node-species, the squares represent the node-reactions, the dotted edges and not dotted edges are respectively the negative and positive ones.

From the SR-graph, the authors derive the *reaction graph*, shortly called R-graph, in which only the relation among reactions are considered. Indeed,

R-graph has only the nodes representing the reactions. It is formally defined as:

Definition 2 (R-graph). *The R-graph is defined by the triple* $(V_R, \bar{E}_+, \bar{E}_-)$*, where* V_R *is the finite set of nodes associated to the reactions,* $\bar{E}_+$ *and* $\bar{E}_-$ *are respectively the positive and negative edges defined as:*

- $(R_i, R_j) \in \bar{E}_+$*, with* $i \neq j$*, whenever there exists in the SR-graph a species* $S_k \in V_S$ *such that* (S_k, R_i) *and* (S_k, R_j) *both belonging either to* $E_+ \cup E_-$*, but having opposite signs;*

- $(R_i, R_j) \in \bar{E}_-$*, with* $i \neq j$*, whenever there exists in the SR-graph a species* $S_k \in V_S$ *such that* (S_k, R_i) *and* (S_k, R_j) *both belonging either to* E_+ *or* E_-*.*

Intuitively, given two reactions, we draw a positive edge between them if they "cooperate" each other, hence, for example, the product of a reaction is among the reactants of the other reaction. Instead, we draw a negative edge between two reactions if both share the same reactants, hence they "compete".

To make an example of R-graph, we consider again the reaction network 12 and its associated SR-graph, in Figure 27.

In the SR-graph, we have to consider all the edges that link two reactions. In Example 27, we have the edges $ES - R_2$ and $ES - R_1$, which are respectively positive and negative, and the edges $E - R_1$ and $E - R_2$, respectively positive and negative. Therefore, according to Definition 2, we draw a positive edge between R_1 and R_2, as in Figure 28.

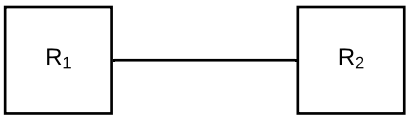

Figure 28: R-graph of the chemical reaction network 12, representing the Michaelis-Menten kinetics. In general in the R-graph, we have only one kind of nodes (the squares), which represent the reactions, and the dotted edges and not dotted edges are respectively the negative and positive ones. In this example, between the reactions R_1 and R_2 there is a positive edge because they "cooperate" each other.

On an arbitrary SR- or R-graph we can have some additional definitions, as described in [4, 5]:

- *simple loop*: it is a path connecting nodes via edges, in which the first and the last node coincide, and no node or edge occurs twice, with the exception of the first and last node;
- *length*: in a simple loop, it is the total number of negative edges;
- *positive simple loop*: a SR- or a R- graph has a positive simple loop when, any simple loop presented in the graph, has an even number of edges;
- *e-loop* and *o-loop*: Let L be a simple loop in the SR-graph. We say that L is an *e-loop* if letting λ be half of its length and σ the product of the signs of all its edges, it holds that $(-1)^{\lambda} = \sigma$. Otherwise, we say that L is an *o-loop*.

After these auxiliary definitions, we can state the two main results of [4, 5]:

Theorem 6 (R-graph positive loop property)**.** *The R-graph has the positive loop property iff the following two conditions are met:*

- *all simple loops in the SR-graph are e-loops;*
- *in the SR-graph, each node in* V_S *is linked at most two nodes in* V_R.

Theorem 7 (Global Monotonicity)**.** *A dynamical system is monotone with respect to an order induced by an orthant cone iff the associate R-graph has the positive loop property.*

Considering again the SR-graph in Figure 27: both requirements for global monotonicity are met. It is presented only one loop $E - R_2 - ES - R_1 - E$ and it is an *e-loop* because it has an even number of negative edges, and all the nodes-species are linked to at most two nodes-reactions. Thus, according to Theorem 7, the system in Figure 12 is globally monotonic. The proofs of the theorems are in [5].

5.2 DEFINITION OF INPUT-OUTPUT MONOTONICITY

In Chapter 4, we formalized a notion of initial concentration robustness, defined as the property of a system to preserve its functions despite the presence of perturbations affecting the initial concentrations. To represent the perturbations, we use intervals as initial parameters. Thus, to study the network

dynamics to verify robustness, we should simulate the model considering all the possible combinations of the values included in the range and performing many simulations. Then, to reduce the computational cost we decide to study monotonicity in chemical reaction networks between the input and the output of the CRN. Indeed, if the monotonicity is verified, we are able to test the model only on the extreme values of the input concentration range, reducing the number of simulations drastically.

Theorem 7 does not help us in this sense because we want to study the monotonicity between an input and an output species. Therefore, we give a new definition of monotonicity, namely the Input-Output monotonicity, and we prove that starting from [5], we can study the monotonicity relation of an input and an output species.

Consider a set of reactions R, over a set of species $S = S_1, ...S_n$.

The following two definitions describe the concept of monotonicity that we are interested in. Our properties of monotonicity describe whether the output species react in a monotone way to the increase of the input concentration.

We consider two initial states $S^0, \overline{S^0}$ such that $S^0_I > \overline{S^0_I}$ for one particular species I (the *input species*), and $S^0_k = \overline{S^0_k}$ for all other species $k \neq I$.

With $S_i(t)$ we indicate the solution of the ODEs for the species S_i with initial value S^0, and with $\overline{S_i}(t)$ the solution with initial value $\overline{S^0}$.

Definition 3 (Positive Input-Output Monotonicity)**.** *Given a set of reactions* R*, species* O *is* positively monotonic *with respect to* I *in* R *if and only if,* $\forall \overline{I}(t) \geqslant I(t)$*,* $\overline{O}(t) \geqslant O(t)$*, for every time* $t \in \mathbb{R}_{\geqslant 0}$.

Definition 4 (Negative Input-Output Monotonicity)**.** *Given a set of reactions* R*, species* O *is* negatively monotonic *with respect to* I *in* R *if and only if,* $\forall \overline{I}(t) \geqslant I(t)$*,* $\overline{O}(t) \leqslant O(t)$*, for every time* $t \in \mathbb{R}_{\geqslant 0}$.

Example 1. Consider a network R consisting of the following chemical reactions:

$$A + B \xrightarrow{k_1} C \qquad (R_1)$$
$$D + B \xrightarrow{k_2} E \qquad (R_2)$$

The differential equations describing the behavior of R are the following, according to the mass action kinetics:

$$\begin{cases} \frac{d[A]}{dt} = -k_1[A][B] \\ \frac{d[B]}{dt} = -k_1[A][B] - k_2[B][D] \\ \frac{d[C]}{dt} = +k_1[A][B] \\ \frac{d[D]}{dt} = -k_2[B][D] \\ \frac{d[E]}{dt} = +k_2[B][D] \end{cases}$$

Consider A as the input species and C as the output species. Assume now that the initial concentrations of species B, D, C and E are fixed, but that the initial concentration of species A can vary from 10 to 1000. In principle, in order to study the dynamics of the concentration of C we would need to perform many simulations, sampling possible initial values of the concentration of A. However, if species C is *positively monotonic* w.r.t. A, then just two simulations are necessary: one with $A = 10$ and one with $A = 1000$. The dynamics of C in all the other (intermediate) cases is included in the results we obtained from these two simulations. A similar simplification could be done by assessing the *negatively monotonicity* of E w.r.t. A. In addition, if the species to vary were two, say A and B with the latter varying from 20 to 200, and if species C was also *positively monotonic* w.r.t. B, then just two simulations would be necessary to study the behavior of C: one with $A = 10$ and $B = 20$ and another with $A = 1000$ and $B = 200$. Finally, if we could prove that species E is *positively monotonic* w.r.t. B then, to study its behavior we would need at most four different simulations, $A = 10$ and $B = 20$, $A = 10$ and $B = 200$, $A = 100$ and $B = 20$, and $A = 100$ and $B = 200$.

It is worth noting that interesting weaker notions of monotonicity could be defined. For example, for some reaction networks it could be interesting to study steady-state monotonicity, defined (in its positive formulation) as follows.

Definition 5. O *in* R *is* steady-state positively monotonic *with respect to* I *if and only if* $I < \overline{I}$ *implies*

$$\lim_{t\to\infty} \overline{O}(t) \geqslant \lim_{t\to\infty} O(t).$$

5.3 ASSESSMENT OF MONOTONICITY

According to Definition 2, in the R-graph, we intuitively draw a positive edge if the two reactions cooperate, otherwise we draw a negative edge if they compete. Therefore, considering again the chemical reaction network 1, the two reactions, namely R_1 and R_2, have in common a reactant, B, hence by definition they compete and the corresponding R-graph is shown in Figure 29.

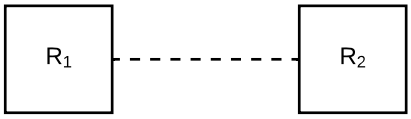

Figure 29: R-graph of the chemical reaction network 1. In this example, between the reactions R_1 and R_2 there is a negative edge because they "compete" each other: the species B is a reactant of both reactions.

In the following, we assume that the R-graph is connected: this is a natural assumption because otherwise the reaction system could be split into two systems with disjoint variables and completely independent behaviors.

Following [5], we introduce some definitions related to a labeling of the R-graph.

Definition 6. *A* consistent labeling *of a signed graph* $\langle V, E_+, E_- \rangle$ *is a labeling* $s : V \to \{+, -\}$ *in which vertices* $i, j \in V$ *have the same label if* $(i, j) \in E_+$*, and opposite labels if* $(i, j) \in E_-$.

It is well-known [37] that a consistent labeling can be constructed if and only if the graph has the positive loop property. In that case, a labeling can be constructed with Algorithm 1.

Let $\Gamma \in \mathbb{R}^{|S| \times |R|}$ be the *stoichiometry matrix* of R, i.e., the matrix such that Γ_{ij} is the amount of species S_i produced (if with a positive sign) or consumed (if with a negative sign) by reaction R_j.

Input : A connected reaction graph $\langle V_R, E_+, E_- \rangle$ with the positive loop property
Output : A consistent labeling s
Label an arbitrary reaction with $+$;
while *there are still unlabeled vertices* **do**
  for each R_j s.t. there exists R_i with label $+$ and $(R_i, R_j) \in E_+$, assign $+$ to R_j;
  for each R_j s.t. there exists R_i with label $-$ and $(R_i, R_j) \in E_-$, assign $+$ to R_j;
  for each R_j s.t. there exists R_i with label $+$ and $(R_i, R_j) \in E_-$, assign $-$ to R_j;
  for each R_j s.t. there exists R_i with label $-$ and $(R_i, R_j) \in E_+$, assign $-$ to R_j;
end

Algorithmus 1 : Consistent labeling of an R-graph.

In our running example (Figure 1),

$$\Gamma = \begin{array}{c} \\ A \\ B \\ C \\ D \\ E \end{array} \begin{array}{c} \begin{array}{cc} R_1 & R_2 \end{array} \\ \begin{pmatrix} -1 & 0 \\ -1 & -1 \\ 1 & 0 \\ 0 & -1 \\ 0 & 1 \end{pmatrix} \end{array}.$$

Let $R(\cdot) : \mathbb{R}^{|S|} \to \mathbb{R}^{|R|}$ be the function that gives the rates of all reactions as a function of the amounts of species present at time i, and DR be its Jacobian matrix. Assume that $DR_{ji}\Gamma_{ij} \leqslant 0$ for each i, j: i.e., every reaction has a rate which increases if more reactants are present, where this assumption is given by the law of mass action, for which the rate of a reaction is proportional to the product of reactant's concentrations.

Our main result is the following.

Theorem 1. *Let a set of chemical reactions* G *be given, with a species input* I*, marked as* S_{i_1}*, and a species output* O*, marked as* S_{i_2} *(with* $i_2 \neq i_1$*). If the following three conditions hold*

1. *the R-graph of* G *has the positive loop property and hence admits a consistent labeling* s*;*

2. *The input species* I (S_{i_1}) *participates in only one reaction* R_{j_1}

3. *The output species* O (S_{i_2}) *participates in only one reaction* R_{j_2}

then, the species O *is positively monotone with respect to* I *if* $\Gamma_{i_1 j_1} s(j_1)$ *and* $\Gamma_{i_2 j_2} s(j_2)$ *have opposite signs, and negatively monotone if they have the same sign.*

Proof. We quickly recall the proof of [5, Theorem 1]. Let $\Sigma \in \mathbb{R}^{|R|\times|R|}$ be the diagonal matrix such that $\Sigma_{jj} = 1$ for all reactions R_j such that $s(R_j) = +$ and $\Sigma_{jj} = -1$ for all reactions R_j such that $s(R_j) = -$.

Let $x(t) = \int_0^t R(S(\tau))d\tau$ be the vector such that x_i is the extent of the ith reaction.

The vector $x(t)$ solves, with initial condition $x(0) = 0$, the system of differential equations

$$\frac{d}{dt}x(t) = R(S^0 + \Gamma x(t)), \tag{13}$$

whose Jacobian is $J = DR \cdot \Gamma$.

If the R-graph of the system has the positive loop property, then it is proved in [5] that $\Sigma J \Sigma$ has non-negative off-diagonal elements, i.e., $(\Sigma J \Sigma)_{ij} \geqslant 0$ for all $i \neq j$. Thanks to a result in [40], this property implies that the dynamical system (13) is orthant-monotone with respect to the orthant $\{x : \Sigma x \geqslant 0\}$, i.e., if we call $x(t), \hat{x}(t)$ two solutions of (13) with initial conditions $x^0, \hat{x}^0$, then $\Sigma(\hat{x}^0 - x^0) \geqslant 0$ implies $\Sigma(\hat{x}(t) - x(t)) \geqslant 0$ for all $t > 0$.

We shall use a modification of this proof to show our result. We assume without loss of generality that $\Gamma_{i_1 j_1} s(j_1) < 0$, as otherwise we can switch the sign of each label, obtaining a new consistent labeling $-s$.

Define the dynamical system

$$\begin{cases} \frac{d}{dt}y(t) = R(S^0 + ze_{i_1} + \Gamma x(t)), \\ \frac{d}{dt}z(t) = 0, \end{cases} \tag{14}$$

where $e_{i_1} \in \mathbb{R}^{|S|}$ is the vector that has 1 in the i_1th component and 0 in all others.

Direct verification shows that the solution of this system with initial value $y(0) = 0, z(0) = 0$ is $y(t) = x(t), z(t) = 0$, whereas the solution $\overline{y}(t), \overline{z}(t)$ with initial value $\overline{y}(0) = 0, \overline{z}(0) = \overline{S_{i_1}} - S_{i_1} > 0$ is $\overline{y}(t) = \overline{x}(t), \overline{z}(t) = \overline{S_{i_1}} - S_{i_1}$,

where the quantity $\overline{x}(t)$ is defined analogously to (13) but with initial value $\overline{S^0} = S^0 + \overline{z}(0)e_{i_1}$. Define

$$\hat{\Sigma} = \begin{bmatrix} \Sigma & 0 \\ 0 & 1 \end{bmatrix}, \quad \hat{J} = \begin{bmatrix} J & \nu \\ 0 & 0 \end{bmatrix},$$

where $\nu = DR \cdot e_{i_1}$ is the i_1th column of DR, and has only one nonzero entry DR_{j_1,i_1} with sign $s(R_{j_1})$ (due to the assumption that $DR_{ji}\Gamma_{ij} \leqslant 0$ for all i, j). The matrix $\hat{J}$ is the Jacobian of (14), and the matrix $\hat{\Sigma}\hat{J}\hat{\Sigma}$ has non-negative off-diagonal elements. Thus the dynamical system (14) is orthant-monotone, and $\overline{z}(0) - z(0) > 0$, $\overline{y}(0) - y(0) = 0$ implies that $\Sigma(\overline{y}(t) - y(t)) \geqslant 0$ for all t.

The quantities of the output species $S_{i_2}(t), \overline{S}_{i_2}(t)$ at time t with the two initial conditions $S^0, \overline{S^0}$ are given by

$$S_{i_2}(t) = S^0_{i_2} + e^T_{i_2} \cdot \Gamma x(t), \quad \overline{S}_{i_2}(t) = \overline{S^0}_{i_2} + e^T_{i_2} \cdot \Gamma \overline{x}(t).$$

The row vector $e^T_{i_2}\Gamma$ is the i_2th row of Γ, and under our Condition 3 it has only one nonzero entry $\Gamma_{i_2 j_2}$. So

$$\overline{S}_{i_2}(t) - S_{i_2}(t) = \overline{S^0}_{i_2} - S^0_{i_2} + \Gamma_{i_2 j_2}(\overline{x}_{j_2}(t) - x_{j_2}(t)).$$

The term $\overline{S^0}_{i_2} - S^0_{i_2}$ is zero since $i_1 \neq i_2$, and the rest is zero (when $\overline{x}_{j_2}(t) = x_{j_2}(t)$) or has sign $\Gamma_{i_2 j_2} s(j_2)$. □

Remark 1. An attentive reader may have noticed that this proof can be extended with minimal changes to the more general case in which S_{i_1} (or S_{i_2}) participates in more than one reaction, as long as all nonzero entries in the i_1th (or i_2th) row of $\Gamma\Sigma$ have the same sign. However, it follows from the labeling rules that whenever $\Gamma\Sigma$ has two nonzero elements on the same row they must have opposite signs, so this more general case is never encountered in practice.

We continue our running example. Considering again the labeled R-graph in Figure 30, we can now apply our theorem and state that the chemical species $S_{i_2} = E$ is negatively monotonic w.r.t the input species $S_{i_1} = A$. This result is confirmed also by simulation in Figure 31: by increasing the initial concentration of A, the concentration of E decreases (at all times). In the simulations, initial concentrations of B, C, D and E are $B_0 = 20$, $C_0 = 0$, $D_0 = 10$, and $E_0 = 0$.

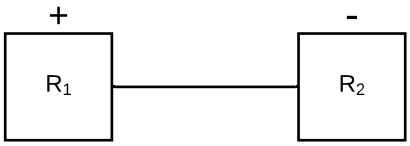

Figure 30: Labeled R-graph of the chemical reaction network 1. In this graph, the squares representing the node-reactions are labeled with a sign "+" or "-", according to Algorithm 1.

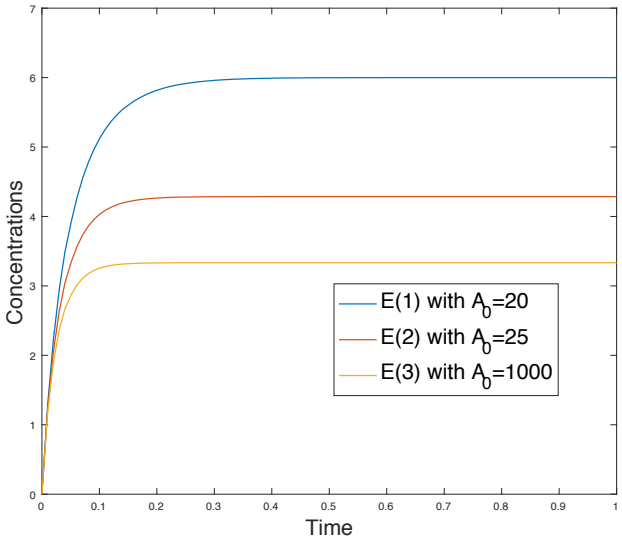

Figure 31: Simulation results of Example 1. In this case, we increase the initial concentration of A and compare the concentrations of the output E. The results show that E is negatively monotonic with respect to the variation of the input.

5.4 EXAMPLES OF APPLICATION OF INPUT-OUTPUT MONOTONICITY THEOREM

We show the application of Theorem 1 to two examples: in the first case, we apply our methodology to the CRN of Michaelis Menten kinetics and, in the second case, we use a more complex network of ERK signalling pathway.

Michaelis-Menten kinetics example. We can study the monotonicity relation between an input and output species, when both of them are involved in only one reaction. Therefore, in Example 12, we can study the monotonicity between the species S, which we can consider as the input of the CRN, and the

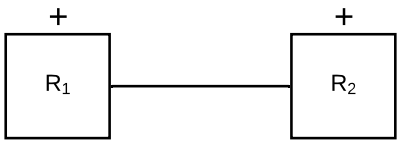

Figure 32: Labeled R-graph of CRN 12, representing Michaelis-Menten kinetics. Since the signs are both positive on the node-reactions, we can say that the output of the CRN, the species P, is positively monotonic w.r.t the input of the CRN, the species S.

species P, as the output, since they are involved respectively in reactions R_1 and R_2.

We already presented the R-graph of this CRN in Figure 28. Since the R-graph has the positive loop property, we can proceed with labeling, using Algorithm 1. Then, we obtain the labeled R-graph, represented in Figure 32, in which the signs on the node-reactions are both positive. Then, we compute the stoichiometric matrix of the system as follows:

$$\Gamma = \begin{array}{c} \\ E \\ S \\ ES \\ P \end{array} \begin{array}{c} \begin{array}{cc} R_1 & R_2 \end{array} \\ \begin{pmatrix} -1 & 1 \\ -1 & 0 \\ 1 & -1 \\ 0 & 1 \end{pmatrix} \end{array}.$$

We calculate the product $\Gamma_{i_1 j_1} s(j_1)$ and $\Gamma_{i_2 j_2} s(j_2)$, which have opposite signs. Then, we can affirm that the output of this CRN, the chemical species P, is positively monotonic with respect to the species S, which means that if we increase the initial concentration of the input (S), we obtain that the concentration of the output (P) increases at the steady state.

This result is also confirmed by simulations. Fixed the chemical rates $(k_1 = 0.1, k_2 = 1000, k_3 = 0.3)$, we increase the initial concentration of the input S in the range $[100, 2000]$ and we obtain that, at the steady state, the concentration of P increases, as we can notice in Figure 33. Indeed, in Figure 33, we show on the horizontal axis all the initial values of S, which we choose as initial concentration of the simulation, and on the vertical axis, we show that the value of P increases at the steady state.

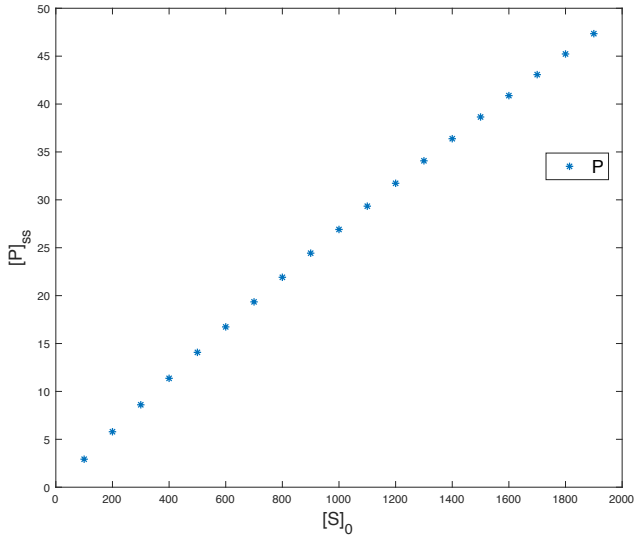

Figure 33: Simulation results of Example 12, representing Michaelis-Menten kinetics. To show how the concentration of the species P is positively monotonic with respect to the species S, we plot on the horizontal axis the initial concentration of S, in a range $[100, 2000]$ and on the vertical axis the concentration of P at the steady state.

Signalling pathway example. A signalling pathway is a long succession of chemical reactions, having a starting point that triggers the other connected processes. An initial stimulus, perceived by a *transductor* (a sort of the first messenger), activates the sequence amplifying the signal for the next reaction. Many biochemical processes are associated with signalling pathways as protein activation, repression, and expression of genes: if anomalies occur, we are facing an error propagation, that could give rise to some diseases, like cancer and diabetes, among the others.

One of the most important examples is the ERK pathway, which regulates growth, survival, proliferation, and differentiation of cells [68]. It consists of many fast phosphorylation events, representing the attachment of a phosphoryl group, which has the function to spread the signal along all the enzymatic cascade. For this example, we will focus only on a particular section of the

entire pathway, which we will denote as ERK$*$, as shown in Figure 34. The reactions involved are the following:

$$\begin{aligned} &\mathrm{Raf} \underset{R_{19}}{\overset{R_{18}}{\rightleftharpoons}} \mathrm{PRaf} \\ &\mathrm{Mek1} \underset{R_{27}}{\overset{R_{21}}{\rightleftharpoons}} \mathrm{PMek1} \\ &\mathrm{PMek1} \underset{R_{25}}{\overset{R_{23}}{\rightleftharpoons}} \mathrm{PPMek1} \end{aligned} \tag{15}$$

Considering Figure 34, the species PRaf is involved as *catalyst promoter* of

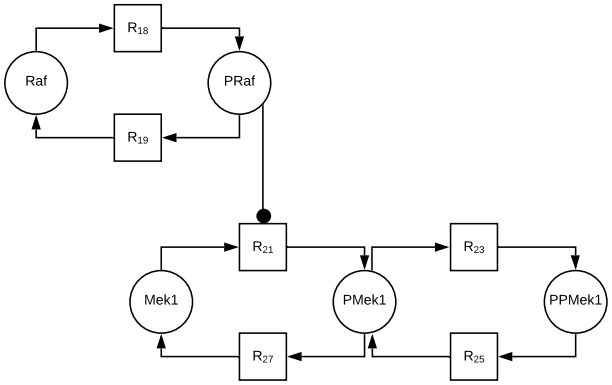

Figure 34: The subgraph of ERK signalling pathway. The particular edge characterized by the black circle represents that the species PRaf is involved as catalyst promoter in the following reaction R_{21}. Then, PRaf helps the production of PMek1.

the reaction R_{21}, which means that its concentration positively influences the production of the species PMek1. A catalytic species increases the reaction rate of the reaction in which it is involved, and during this process, it acts as a constant since its concentration is not consumed. Then, we translate this process in another equivalent CRN, making explicit the function of the enzyme PRaf on the reaction R_{21}. We summarize in Table 5 the reactions involved in the ERK$*$, the coefficient rates and the initial conditions.

Since the reactions follow one another in a chain, we assume that PRaf reaches the steady-state before the PPMek1 species, which represents the final product of the CRN, and we verify the assumption by simulations, as shown

Table 5: The initial concentrations, the rates and the chemical reactions of ERK∗ system.

Initial concentrations	Rates	Chemical reactions
Raf = 10	$k_{18} = 0.1445$	$\text{Raf} \underset{R_{19}}{\overset{R_{18}}{\rightleftharpoons}} \text{PRaf}$
PRaf = 0	$k_{19} = 0.37$	$\text{PRaf} + \text{Mek1} \xrightarrow{R_{21}} \text{PMek1+PRaf}$
Mek1 = 1	$k_{21} = 0.02$	$\text{PMek1} \xrightarrow{R_{27}} \text{Mek1}$
PMek1 = 0	$k_{27} = 0.07$	$\text{PMek1} \underset{R_{25}}{\overset{R_{23}}{\rightleftharpoons}} \text{PPMek1}$
PPMek1 = 0	$k_{23} = 667.957$	
	$k_{25} = 0.13$	

in Figure 35. Then, we subdivide the network into two distinct blocks, represented in Figure 36. On the new configuration obtained, we apply Theorem 1 separately on the two sub-networks, choosing Raf and PRaf respectively the input and the output of the first block, and Mek1 and PPMek1 the input and the output of the second one. For the first block, the Input-Output monotonic-

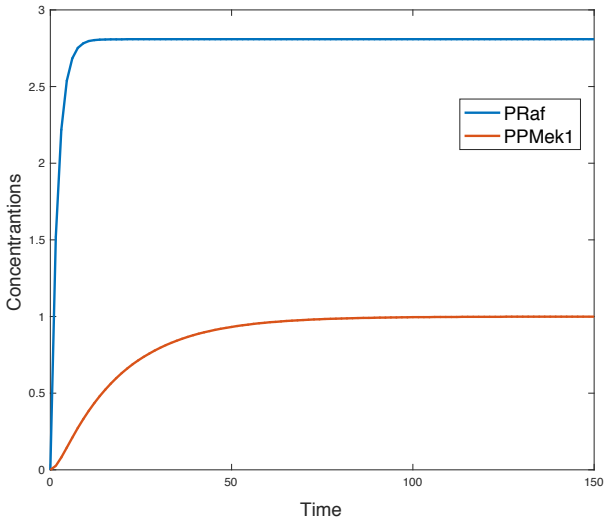

Figure 35: Simulation results of Example 15, representing ERK signalling pathway. We show how the species PRaf reaches the steady state before than species PPMek1.

ity is trivially applied, since we have only one reaction, and we obtain that

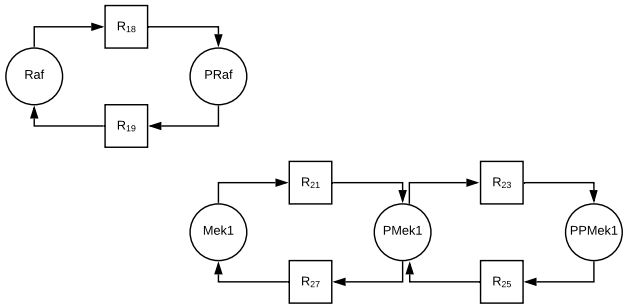

Figure 36: A "modified" subgraph of ERK signalling pathway, denoted as ERK$*$. In this equivalent version of signalling pathway we divide the CRN in two sub-networks, leaving its behavior unchanged.

PRaf is positively monotonic with respect to Raf. For the second sub-network we proceed applying first Algorithm 1, and we obtain the labelled R-graph, represented in Figure 37. We compute the stoichiometric matrix as follows:

$$\Gamma = \begin{array}{c} \\ \text{Mek1} \\ \text{PPMek1} \end{array} \begin{array}{c} \begin{array}{cc} R_1 & R_2 \end{array} \\ \begin{pmatrix} -1 & 0 \\ 0 & +1 \end{pmatrix} \end{array}.$$

We calculate the product $\Gamma_{i_1 j_1} s(j_1)$ and $\Gamma_{i_2 j_2} s(j_2)$, which have opposite signs. Then, we can affirm that the output of this CRN, the chemical species PPMek1, is positively monotonic with respect to the species Mek1, which means that if we increase the initial concentration of the input (Mek1), we obtain that the concentration of the output (PPMek1) increases at the steady state. As described above, PRaf cooperate with reaction R_{21}, increasing its reaction rate and acting as constant. Then we can conclude that, since PRaf is positively monotonic with respect to Raf and PPmek1 is positively monotonic with respect to Mek1, PPmek1 is positively monotonic also with respect to Raf and we show the result in Figure 38.

In general, for signaling pathways, it is always possible to apply a *modular approach*, which divides the reactions into distinct blocks. This particular reaction network, indeed, consists of a long series of chemical processes, in which each event is activated by the previous one.

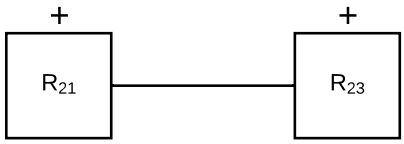

Figure 37: Labeled R-graph of the second sub-network of CRN 15, representing ERK signalling pathway. Since the signs are both positive on the node-reactions, we can say that the output of the CRN, the species PPMek1, is positively monotonic w.r.t the input of the CRN, the species Mek1.

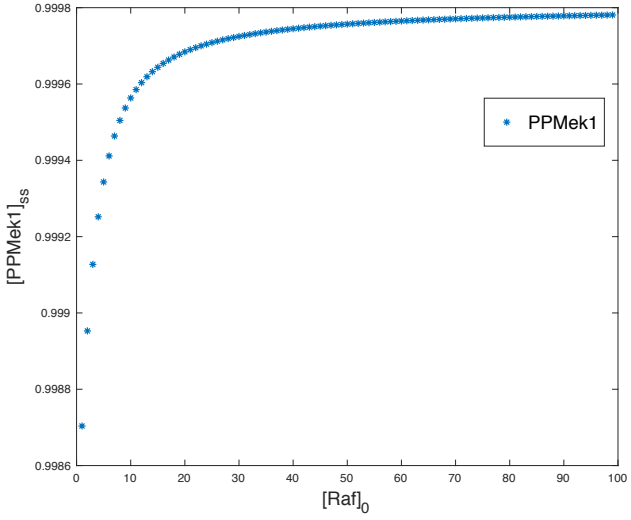

Figure 38: Simulation results of Example 15, representing ERK signalling pathway. To show how the concentration of the species PPMek1 is positively monotonic with respect to the species Mek1, we plot on the horizontal axis the initial concentration of Mek1, in a range $[1, 100]$ and on the vertical axis the concentration of PPMek1 at the steady state.

6

ROBUSTNESS IN THE BECKER-DÖRING MODEL

The Becker-Döring equations were theorized for the first time in 1935, by the two authors who gave rise to the name of the model [13], who proposed an infinite system of ordinary differential equations as a model for the time evolution of the distribution of cluster sizes for such a system [8]. Later, in 1977, Burton started to consider this system to explain condensation phenomena at different pressures [18], indeed, they consist of two main processes: the *aggregation* and the *fragmentation* of clusters of particles. The applications of this kinetic model are multiple and include many fields, like physics, chemistry, and biology. In particular, concerning biology, it has been applied to study protein aggregation, polymerization, and the formation of intracellular concentration gradients.

As shown in [67], the clustering processes can be very robust to concentration fluctuations which suggests that it exhibits robustness. Then, we proceed to verify this property using different approaches.

In the first part of the Chapter, we introduce in detail the mathematical properties of Becker-Döring equations, showing the analytical solution of the steady state. In the second part, we apply the *Deficiency Zero Theorem* [30], described in Section 3.4, to verify the robustness property of this model.

6.1 THE BECKER-DÖRING EQUATIONS

The Becker-Döring equations (BD) describe two principal phenomena, namely the *coagulation* and the *fragmentation* of clusters of particles, based on two processes:

1. a *monomer* (or *elementary particle*) is a cluster characterized by a size i equal to 1. Hitting a cluster of size $i \geqslant 1$, it gives rise to a coagulation phenomenon, producing a cluster of dimension $i+1$;
2. a cluster of size $i \geqslant 2$ can be subjected to a spontaneous fragmentation, splitting itself in a cluster of size $i-1$ and a monomer.

They can be summarized by the following set of kinetic reactions:

$$C_1 + C_i \underset{b_{i+1}}{\overset{a_i}{\rightleftharpoons}} C_{i+1} \tag{16}$$

where C_i denotes clusters consisting of i particles, which is also the cluster size, while kinetic coefficients a_i and $b_i + 1$ stand, respectively, for the rate of aggregation and fragmentation, and for the moment we consider them as fixed values. We can notice that at the basis of this model there are three fundamental assumptions:

- only a monomer coalesces to give rise a cluster;
- a cluster releases spontaneously a monomer;
- at the initial stage of the system, only $C_1(0)$, namely the initial concentration of monomers, is different from 0, hence all the clusters, with size $i \geqslant 2$, develop successively.

From the last assumption, it follows that the initial concentration of C_1 determines the mass of the system, hence the largest dimension of the cluster that can be formed. Indeed, if we add 5 molecules in an ideal closed pot, the maximum cluster will have size 5, and the only possible clusters will be C_2, C_3, C_4, and C_5. Then, in a model with the initial concentration of monomers $C_1(0) = 5$, we will have 10 reactions (considering for each cluster its process of fragmentation and coagulation) as follows:

$$\begin{aligned} C_1 + C_1 &\underset{b}{\overset{a}{\rightleftharpoons}} C_2 \\ C_1 + C_2 &\underset{b}{\overset{a}{\rightleftharpoons}} C_3 \\ C_1 + C_3 &\underset{b}{\overset{a}{\rightleftharpoons}} C_4 \\ C_1 + C_4 &\underset{b}{\overset{a}{\rightleftharpoons}} C_5 \end{aligned} \tag{17}$$

In a deterministic framework, applying the law of mass action 1, the BD model is a set of differential equations, representing the time evolution of each

concentrations of clusters of size i. Considering the example 17, we obtain the following differential equations:

$$\begin{cases} \frac{d[C_1]}{dt} = -2a[C_1]^2 + 2b[C_2] - a[C_1][C_2] + b[C_3] - a[C_1][C_3] + b[C_4] - a[C_1][C_4] + b[C_5] \\ \frac{d[C_2]}{dt} = +a[C_1]^2 - b[C_2] - a[C_1][C_2] + b[C_3] \\ \frac{d[C_3]}{dt} = +a[C_1][C_2] - b[C_3] - a[C_1][C_3] + b[C_4] \\ \frac{d[C_4]}{dt} = +a[C_1][C_3] - b[C_4] - a[C_1][C_4] + b[C_5] \\ \frac{d[C_5]}{dt} = +a[C_1][C_4] - b[C_5] \end{cases}$$

As we can notice it is possible to generalize the differential equations obtaining two differential equations, one describing the time evolution of the cluster C_1 and the other describing the time evolution of a generic cluster C_i, as follows:

$$\begin{cases} \frac{d[C_1]}{dt} = -J_1 - \sum_{i \geqslant 1}^{C_1(0)} J_i \\ \frac{d[C_i]}{dt} = J_{i-1} - J_i \end{cases} \tag{18}$$

for every $i \geqslant 2$, where J_i is the flux:

$$J_i = a[C_1][C_i] - b[C_{i+1}] \tag{19}$$

Analyzing the formula 19, we can recognize the generic process of coagulation, as a second order reaction, and the generic process of fragmentation, as a first order linear reaction. Looking at this formula, we can notice that the expression for the biggest cluster is different from the others one. In the previous example, indeed, the biggest cluster is C_5, which can only participate in fragmentation, otherwise other interactions it would create cluster bigger than 5, which is the total mass of the system. This is evident, also considering the fluxes.

Considering again a system with $C_1(0) = 5$, there are 4 fluxes, as follows:

$$\begin{aligned} J_1 &= a[C_1]^2 - b[C_2] \\ J_2 &= a[C_1][C_2] - b[C_3] \\ J_3 &= a[C_1][C_3] - b[C_4] \\ J_4 &= a[C_1][C_4] - b[C_5]. \end{aligned} \tag{20}$$

As we can notice, there is not a flux J_5 because the cluster of dimension 5 can be only involved in a spontaneous fragmentation, this is because of the mass conservation.

The mass of the system (ρ) is conserved, hence the considered system has not sink or source. From the generic formula of differential equations 18 we can deduce that the mass of system depends on the the initial condition of the system and has the following form:

$$\sum_{i \geqslant 1} iC_i(t) = \rho \tag{21}$$

6.2 VERIFICATION OF ROBUSTNESS PROPERTY IN BECKER-DÖRING MODEL

The Becker-Döring equations can be used to study different biological phenomena, as the formation of gradient concentrations, which we can define as the measurement of the variation of quantity of molecules from one area to another in the same system. This phenomenon is experimentally observable in unicellular and multicellular organisms, and it is involved in various processes, such as cellular differentiation, as described in [67, 78].

In the work [67], the authors develop a theoretical model describing cluster aggregation-fragmentation in subcellular systems, based on the Becker-Döring equations, and show that, for particular conditions, the concentration gradient can be robust to relevant biological fluctuations. Therefore, we proceed to verify formally the robustness of the system based on Becker-Döring equations, applying different approaches that we can summarize as follows:

- *Simulations*;
- *Petri net formalization*;

- *Analytical study of the steady state solution*;
- *Application of Feinberg's theorems*.

6.2.1 *Petri net and simulations*

First of all, we formalize the input and the output of the network, applying our definition 16. In our work, we identify as input and output of the network the concentration of monomer C_1, in agreement with the assumptions we mentioned in Section 6.1. Indeed, C_1 is the only cluster present at the initial state of the system, and it is the cluster involved in every aggregation and fragmentation process. Then, we represent the system as a Petri net, as shown in Figure 39.

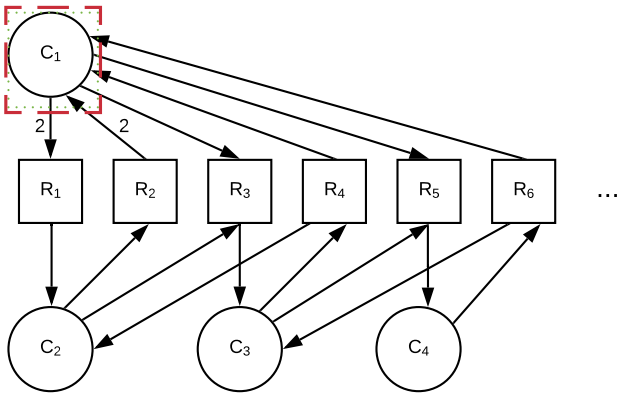

Figure 39: The Petri net for the BD model. C_1 is both the input and the output of the network. The encountered problems are that the Petri net is potentially infinite and, by changing the initial conditions of the network, we obtain different Petri nets.

The Petri net formalization helps us to understand two features of this model:

1. the Petri net is potentially infinite, because every time we change the initial concentration of the system we change the set of differential equations, describing the system;
2. we cannot apply Theorem 1, because the sufficient condition is not verified since both the input and the output are linked to more than one reaction.

As consequence of the last item, we cannot use, in simulations, only the minimum and the maximum of the range of the interval of values describing the initial concentration of monomers. Hence we start performing many simulations, perturbating the initial concentration of C_1. In particular, we change the mass of the system, in a wide range of values, to study the concentration of monomers at the steady state. As we can notice in Figure 40, we find out that, assuming as constant the coefficient rates $a = 1$ and $b = 1$, the concentration of C_1 seems to tend to 1 at the steady state, even with very different initial conditions.

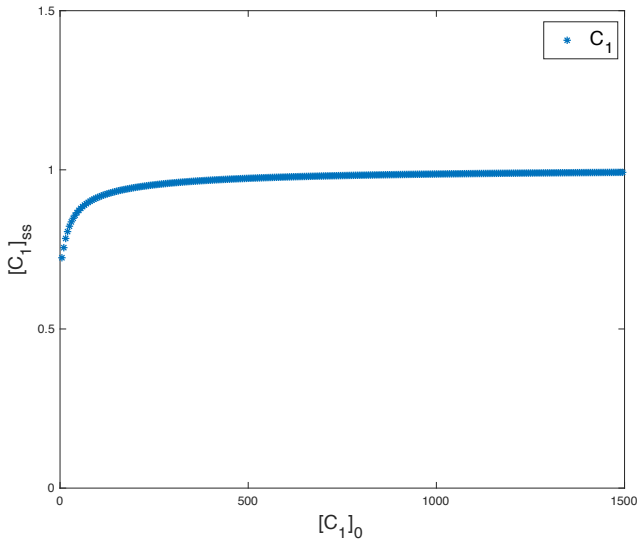

Figure 40: Simulations result of Becker-Döring model. We plot on the horizontal axis the initial concentration of C_1, in a range $[5, 1500]$, and on the vertical axis the concentration of C_1 at the steady state. We assume the the coefficient rates a and b are constant and equal to 1.

This preliminary result convince us to study analytically the solution of the steady state formula.

6.2.2 *Analysis of the Steady state*

In the simulation result shown in Figure 40, the concentration of monomer C_1 appears to be robust at the steady state. Indeed, we change, in a wide range of values, the input of the system (C_1) and we notice that the output (C_1) tends

to 1 at the steady state, considering the coefficient rates a and b constant and equal to 1.

Then, in line with the other works presented in literature [8, 47, 57], our main result in the analytical study of steady state is the following:

Theorem 8 (Monomers Steady State)**.** *Let* a *and* b *be the coefficient rates of coagulation and fragmentation process in the Becker-Döring system,* ρ *the mass of the system and* $[C_1]_{ss}$ *the concentration of monomers at the steady state. Then, as* $\rho \to \infty$, $[C_1]_{ss} \to \frac{b}{a}$.

Proof. As described in Section 6.1, the crucial assumption of the Becker-Döring model is mass conservation, which therefore depends on the initial concentration of monomers, hence the mass ρ at the initial state will remain the same at the steady state, that is the sum of the fluxes of the species goes to zero. Then, recalling Formula 21, we can write:

$$\rho(0) = \rho(\infty) = \sum_{i=1}^{k} i \cdot C_i, \tag{22}$$

where k is the maximum number of molecules in the system, hence it is $k = \rho$. Generalizing the fluxes formulas, as in example 20, we deduce the general steady state formula for a cluster of dimension i, as follows:

$$[C_i]_{ss} = \left(\frac{a}{b}\right)^{i-1} [C_1]_{ss}^{i}. \tag{23}$$

Replacing (23) in (22), we obtain:

$$\begin{aligned} \rho &= \sum_{i=1}^{k} i \left(\frac{a}{b}\right)^{i-1} [C_1]_{ss}^{i} \\ &= \frac{b}{a} \sum_{i=1}^{k} i \left(\frac{a}{b}[C_1]_{ss}\right)^{i} \\ &= \frac{b}{a} \frac{k(\frac{a}{b}C_1)^{k+2} - (k+1)(\frac{a}{b}C_1)^{k+1} + (\frac{a}{b}C_1)}{(1-\frac{a}{b}C_1)^2}. \end{aligned} \tag{24}$$

We want to study the asymptotic behaviour of C_1, then we define:

$$\lim_{t\to\infty} C_1(t) := x_k.$$

Equating (24) with its general form, we get:

$$k = x + \frac{a}{b}2x^2 + ... + \frac{a^{k-1}}{b^{k-1}}kx^k = \frac{b}{a}\cdot\frac{k\cdot\frac{a}{b}x^{(k+2)} - (k+1)\cdot\frac{a}{b}x^{(k+1)} + \frac{a}{b}x}{(1-\frac{a}{b}x)^2}.$$

We can rearrange the previous expression as follows:

$$1 = \frac{x}{k} + \frac{a}{b}\frac{2}{k}x^2 + ... + \frac{a^{k-1}}{b^{k-1}}x^k = \frac{b}{a}\cdot\frac{k\cdot\frac{a}{b}x^{(k+2)} - (k+1)\cdot\frac{a}{b}x^{(k+1)} + \frac{a}{b}x}{k\cdot(1-\frac{a}{b}x)^2}, \tag{25}$$

and we define:

$$\begin{aligned} g_k(x) &= \frac{x}{k} + \frac{a}{b}\frac{2}{k}x^2 + ... + \frac{a^{k-1}}{b^{k-1}}x^k, \\ &= \frac{b}{a}\cdot\frac{k\cdot\frac{a}{b}x^{(k+2)} - (k+1)\cdot\frac{a}{b}x^{(k+1)} + \frac{a}{b}x}{k\cdot(1-\frac{a}{b}x)^2}. \end{aligned}$$

We will need both expressions because they make it simpler to observe different properties.

- $g_k(x)$ is an increasing function of x
- with simple algebraic manipulation, we get:

  $$g_k(x) = \frac{1}{k}\cdot\frac{b}{a}\left(\frac{a}{b}x + \frac{a^2}{b^2}2x^2 + ... + k\frac{a^k}{b^k}x^k\right)$$

  from which it is easy to see that:

  $$g_k\left(\frac{b}{a}\right) = \frac{1}{k}\cdot\frac{b}{a}(1+2+...+k) = \frac{b}{a}\cdot\frac{k+1}{2}.$$

- Following the first two items, we notice that there is one and only one solution for the equation $g_k(x) = 1, \forall k$, in the range $\left[0, \frac{b}{a}\right]$.

- Looking at the other expression for the function $g_k(x)$, we find that

$$\lim_{k\to\infty} g_k(x) = 0, \quad \forall x \in \left[0, \frac{b}{a}\right).$$

- Because of the previous limit, if we now take a generic $x^* < \frac{b}{a}$, then, for k large enough we will have $g_k(x^*) < g_k(x_k) \equiv 1$ Since the function is increasing and monotonic with respect to x, then:

$$\frac{g_k(x_k) - g_k(x^*)}{x_k - x^*} > 0 \;\Rightarrow\; x_k > x^*.$$

This shows that, for k large enough, $x_k > x^*, \forall x^* < \frac{b}{a}$. Therefore $x_k \geqslant \frac{b}{a}$.

□

For an intuitive visualization of what just said, in Figure 41 we show a plot of the $g_k(x)$ in the particular case where $a = b$.

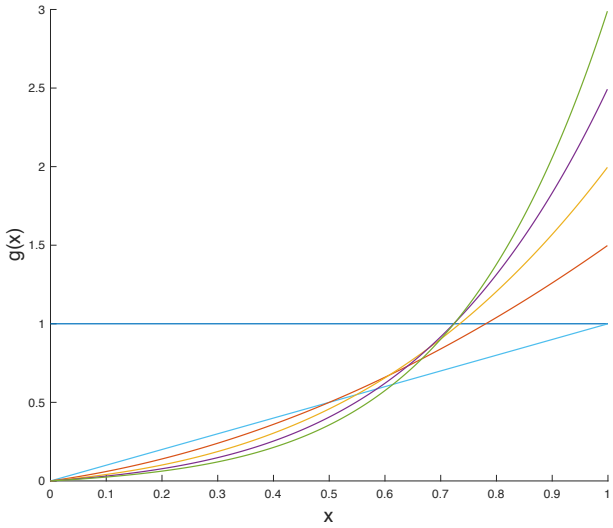

Figure 41: The plot of the function $g(x)$. Graphically, we see that the curves of function $g(x)$ will be closer to zero, increasing the value of k.

Our result is shown also in Figure 40 and Figure 42, respectively for $a = b$ and $a \neq b$. As we can notice, through the analytical solution of the steady state formula, we can avoid performing many simulations to study all possible behaviors of the system considering different values for a and b.

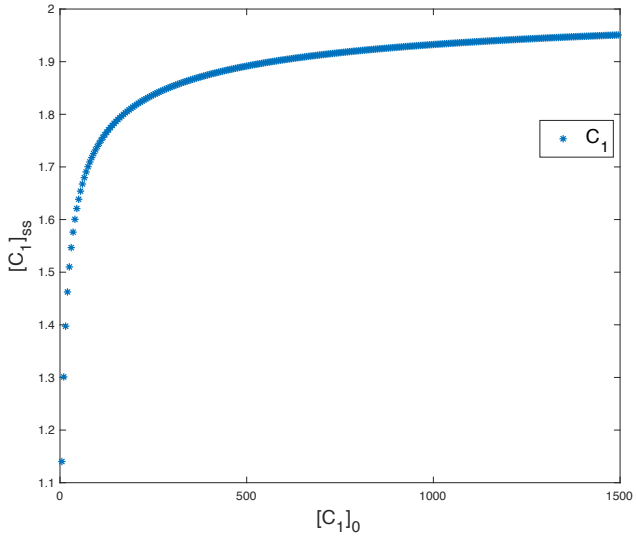

Figure 42: Simulations result of Becker-Döring model. We plot on the horizontal axis the initial concentration of C_1, in a range $[5, 1500]$, and on the vertical axis the concentration of C_1 at the steady state. The coefficient rates are $a = 1$ and $b = 2$.

6.2.3 *Application of Deficiency Theorems*

To verify robustness in the Becker-Döring model, we apply the Deficiency One Theorem 3.4.1, as described in detail in Section 3.4.

The first step is to fix an initial concentration for the monomers, such as $C_1(0) = 5$. Then we obtain the following set of reactions:

$$\begin{aligned} C_1 + C_1 &\underset{b}{\overset{a}{\rightleftharpoons}} C_2 \\ C_1 + C_2 &\underset{b}{\overset{a}{\rightleftharpoons}} C_3 \\ C_1 + C_3 &\underset{b}{\overset{a}{\rightleftharpoons}} C_4 \\ C_1 + C_4 &\underset{b}{\overset{a}{\rightleftharpoons}} C_5 \end{aligned} \tag{26}$$

Before to calculate the deficiency δ, we define the number of nodes and the linkage classes, and the rank of the matrix $r \times N$, where r represents the reactions and N the species involved in the network. In this case, we consider as a species each possible cluster C_i.

- *Nodes* (n) are one or more chemical species involved in forward and backward reaction. In this case, the number of nodes is 8, and they are: $C_1 + C_1$, C_2, $C_1 + C_2$, C_3, $C_1 + C_3$, C_4,$C_1 + C_4$, C_5;
- *linkage classes* (l) are the "groups" of reactions which compose the network, then we have 4 linkage classes;
- Considering the following matrix, we obtain that the *rank* (r) is 4.

$$\begin{array}{c} \\ R_1 \\ R_{1b} \\ R_2 \\ R_{2b} \\ R_3 \\ R_{3b} \\ R_4 \\ R_{4b} \end{array} \begin{array}{c} \begin{array}{ccccc} C_1 & C_2 & C_3 & C_4 & C_5 \end{array} \\ \begin{bmatrix} -2 & +1 & 0 & 0 & 0 \\ +2 & -1 & 0 & 0 & 0 \\ -1 & -1 & +1 & 0 & 0 \\ +1 & +1 & -1 & 0 & 0 \\ -1 & 0 & -1 & +1 & 0 \\ +1 & 0 & +1 & -1 & 0 \\ -1 & 0 & 0 & -1 & +1 \\ +1 & 0 & 0 & +1 & -1 \end{bmatrix} \end{array}$$

Then, we can calculate the deficiency δ, defined as $\delta = n - l - r$, and we find out that the $\delta = 8 - 4 - 4 = 0$. Therefore, applying Theorem 4, described in Section 4, we know that a system with $\delta = 0$ cannot be robust.

In [69], the authors explain that in a biological network, some chemical species can be more sensitive to the presence of fluctuations. Then, the authors present lower bounds on the individual species-concentration sensitivities that derive from reaction network structure alone, independently from kinetic parameters or even of the particular equilibrium state at which sensitivities are calculated. In this context, the main result of [69] states that the less sensitive species are the monomers or the molecules, which tend to react with other species through multiple interactions. Below, we informally present an example of the approach, applying it to Becker-Döring system.

We assume that the chemical reaction network, governed by the Becker-Döring equation, is a mass-action system and it takes place in an ideal closed pot, with fixed temperature and volume. Then, after a certain period, it will reach a steady state, identified as c. If we open the vessel and introduce other

species quantities, hypothetically, the system will be in another equilibrium, namely $c*$, and, consequently, all the species involved will arrive at a different steady state.

In the system, we make a distinction between monomer and species, which respectively identify an elementary substance e and a compound of monomers s, where e denotes a single element of the set of monomers E and s represents a single species of the set of involved species S.

Applying this specification to our system 26, we consider that all the clusters C_i consist of i copies of C_1, then we obtain:

- $S = \{C_1, C_2, C_3, C_4, C_5\}$;
- $E = \{C_1\}$.

For each monomer $e \in E$ and each species $s \in S$, we specify how many monomers form a species, denoting the e-content of species s, namely M_e^s. In our example, considering C_1 as the only monomer, we specify how many copies of C_1 there are in each species. Then, we obtain:

- $M_{C_1}^{C_1} = 1$,
- $M_{C_1}^{C_2} = 2$,
- $M_{C_1}^{C_3} = 3$,
- $M_{C_1}^{C_4} = 4$,
- $M_{C_1}^{C_5} = 5$.

For each pair of monomers $e, e' \in E$, we denote by $M_e^{max}(e')$, the maximal e-content that can be found as we search over all species that contain the element e'. Running our example, we set:

- $M_{C_1}^{max}(C_1) = 5$.

If the maximal e-content is high for the specific monomer under study, it means that this monomer is involved in many other species. Intuitively, since C_1 is the only monomer present in the system and it forms all the species, its maximal e-content is the possible maximum size of the cluster.

In the end, we define as *monomer degree* the sum of all maximal e-content, and it is given by:

$$\deg(e) = \sum_{e' \in E} M_e^{max}(e').$$

Thus, for our CRN, we have:

- $\deg(C_1) = 5$,

and as we can notice the monomer degree is the value, which identifies how a particular substance is free to combine with many other elements. The *lower bound* $\Lambda^s(c*)$ on the species sensitivities is defined as follows:

$$\Lambda^s(c*) \geqslant \max_{e \in E} \left\{ \frac{M_e^s}{\deg(e)} \right\}.$$

Thus, in our example it is:

- $\Lambda^{C_1}(c*) \geqslant \frac{1}{5}$.

Increasing the initial concentration of monomer C_1, we will obtain different results. For example, choosing $C_1 = 10$, the degree and the lower bound will be respectively:

- $\deg(C_1) = 10$,
- $\Lambda^{C_1}(c*) \geqslant \frac{1}{10}$.

Intuitively, if we change the initial concentration of monomers, two processes will occur in the system:

1. the monomer degree of C_1 will increase, in relation with the initial concentration. Then, C_1 will interact with many species in the system;
2. the sensitivity of the species and monomers in the system will decrease because the effects of perturbation will attenuated through proopagation across many species or buffered within species cointaining multiple instance of elements.

7 CONCLUSIONS

Two main characteristics define living cells: an intrinsic structural complexity, according to which there is modularity inside the cell itself, and the ability to interact with other networks, working as a system. Besides, at various frequencies and timescales, internal and external fluctuations can alter specific functions or traits of biological systems, causing genetic mutations, loss of structural integrity, and diseases among the others. Nevertheless, many biological networks can maintain their functionalities despite perturbations: this distinct property is known as *robustness* [44, 45].

Robust traits are pervasive in biology: they pertain to various structural levels, such as gene expression, protein folding, metabolic flux, species persistence. For this reason, the study of this characteristic is essential for many biologists, whose aim is to understand the performance and functions of a biological system [77].

In general, it is difficult to investigate biological systems because they are characterized by non-linearity and non-intuitive behaviors. They can be studied by performing wet-lab (*in vitro*) experiments, or through mathematical or computational (*in silico*) methods on a pathway model (often expressed in terms of ODEs or Markov chains). In particular, robustness is very arduous to analyse because there exist many different notions of this concept, studied using simulations or identifying significant topological structures of the systems [30]. However, this last approach can only be used in a very restrictive context, useful only for studying specific biological systems.

For this reason, in this thesis, we propose a new definition to formally describe a specific notion of robustness, the *initial concentration robustness*. This has the purpose of analyzing how perturbations in the initial concentrations of the involved chemical species (identified as inputs) can alter the system behavior at the steady state. Therefore, we developed a theoretical framework, based on Petri net formalism, and we applied it to different known models. Using simulations, we validated Definition 16 by applying it to systems already studied in literature, such as the EnvZ/OmpR osmoregulatory signalling sys-

tem 4.2.1, the isocitrate dehydrogenase regulatory system 4.2.2, the bacterial chemotaxis 4.2.3 and a model of the enzyme activity at saturation 4.2.4.

To understand the behavior of a biological system, we should simulate it considering all the possible combinations of the initial values, which implies a huge computational effort. To face this issue, we found a sufficient condition that allows knowing if the concentration of an output species is monotonic concerning the concentration of an input species (which is the perturbed substance). By monotonic, we mean that increasing (or decreasing) the concentration of the input, the concentration of the output, at each time step, increases (or decreases) consequently. If the sufficient condition is met (namely the monotonicity is found), we can significantly reduce the number of simulations, testing the model only on the extreme values of the input concentration range. Then, we test Theorem 1 on two examples: Michaelis-Menten kinetics and ERK signalling pathway.

Finally, we apply our theoretical framework to the case study of Becker-Döring equations [7], a model which describes condensations phenomena at different pressures. We focused on the robustness property of this model, studying analytically its steady state. Concerning this, we prove that the concentration of monomer tends to the ratio of coefficients at the steady state, and we verify this statement by simulations. Then, we applied known results in literature [69], to prove how the clusters presented in the model show different sensitivity to the perturbations on the initial concentration.

7.1 FUTURE WORK

There are several improvements that can be introduced to extend this research and that we leave as future work.

Stochasticity. One of the key aspects of the complex biological systems is their randomness. A deterministic approach makes it possible to infer the temporal behavior of the species involved as a result of a system of differential equations that come from the law of mass action (mainly) or other kinetic laws. However, we could extend our theoretical framework on the robustness in chemical reaction networks to stochastic systems.

Topological features. Our definition can always be verified at the cost of performing many simulations. This could be done on a large database of biologi-

cal systems in order to obtain a large training and validation set for a machine learning algorithm able to guess the robustness of a system, as in part it is described in [17]. Then, techniques of explainable AI could be applied in order to infer some topological features that the biological systems recognised as robust have in common.

Applicability to new specific problems. The framework we developed so far has been studied to be as general as possible, in order for it to be applied of many biological systems to investigate the initial concentration robustness. However, studying new biological systems, it would be interesting to extend this framework or to find a new theoretical model to infer robustness or other interesting biological properties.

Analysis of Becker-Döring equations with real rates. In this thesis, we analyse the Becker-Döring equations assuming that the coefficient rates of agglomeration and fragmentation are constant real values. Moreover, we could improve this analysis introducing some physical aspects of the model. As described in [8, 57, 72], the size and the shape of clusters involved in the reactions influence the dynamical properties of the system.

BIBLIOGRAPHY

[1] Uri Alon. *An introduction to systems biology: design principles of biological circuits*. CRC press, 2006.

[2] Uri Alon, Michael G Surette, Naama Barkai, and Stanislas Leibler. «Robustness in bacterial chemotaxis.» In: *Nature* 397.6715 (1999), p. 168.

[3] Rajeev Alur, Costas Courcoubetis, Thomas A Henzinger, and Pei-Hsin Ho. «Hybrid automata: An algorithmic approach to the specification and verification of hybrid systems.» In: *Hybrid systems*. Springer, 1993, pp. 209–229.

[4] David Angeli, Patrick De Leenheer, and Eduardo D Sontag. «On the structural monotonicity of chemical reaction networks.» In: *Decision and Control, 2006 45th IEEE Conference on*. IEEE. 2006, pp. 7–12.

[5] David Angeli, Patrick De Leenheer, and Eduardo Sontag. «Graph-theoretic characterizations of monotonicity of chemical networks in reaction coordinates.» In: *Journal of mathematical biology* 61.4 (2010), pp. 581–616.

[6] Christel Baier and Joost-Pieter Katoen. *Principles of model checking*. MIT press, 2008.

[7] John M Ball and Jack Carr. «Asymptotic behaviour of solutions to the Becker-Döring equations for arbitrary initial data.» In: *Proceedings of the Royal Society of Edinburgh Section A: Mathematics* 108.1-2 (1988), pp. 109–116.

[8] John M Ball, Jack Carr, and Oliver Penrose. «The Becker-Döring cluster equations: basic properties and asymptotic behaviour of solutions.» In: *Communications in mathematical physics* 104.4 (1986), pp. 657–692.

[9] Roberto Barbuti, Andrea Maggiolo-Schettini, Paolo Milazzo, Giovanni Pardini, and Luca Tesei. «Spatial P systems.» In: *Natural Computing* 10.1 (2011), pp. 3–16.

[10] Naama Barkai and Stan Leibler. «Robustness in simple biochemical networks.» In: *Nature* 387.6636 (1997), p. 913.

[11] David J. Barnes and Dominique Chu. *Introduction to Modeling for Biosciences*. 1st. Springer Publishing Company, Incorporated, 2010. ISBN: 1849963258, 9781849963251.

[12] Eric Batchelor and Mark Goulian. «Robustness and the cycle of phosphorylation and dephosphorylation in a two-component regulatory system.» In: *Proceedings of the National Academy of Sciences* 100.2 (2003), pp. 691–696.

[13] Richard Becker and Werner Döring. «Kinetische behandlung der keimbildung in übersättigten dämpfen.» In: *Annalen der Physik* 416.8 (1935), pp. 719–752.

[14] Behnam Behinaein, Karen Rudie, and Waheed Sangrar. «Structural analysis of Petri nets for modeling and analyzing signaling pathways.» In: *Electrical and Computer Engineering (CCECE), 2014 IEEE 27th Canadian Conference on*. IEEE. 2014, pp. 1–5.

[15] Andrea Bernini, Linda Brodo, Pierpaolo Degano, Moreno Falaschi, and Diana Hermith. «Process calculi for biological processes.» In: *Natural Computing* 17.2 (2018), pp. 345–373.

[16] Franco Blanchini and Elisa Franco. «Structurally robust biological networks.» In: *BMC systems biology* 5.1 (2011), p. 74.

[17] Pasquale Bove, Alessio Micheli, Paolo Milazzo, and Marco Podda. *Prediction of dynamical properties of biochemical pathways with Graph Neural Networks*. in press.

[18] JJ Burton. «Nucleation theory.» In: *Statistical Mechanics*. Springer, 1977, pp. 195–234.

[19] Laurence Calzone, François Fages, and Sylvain Soliman. «BIOCHAM: an environment for modeling biological systems and formalizing experimental knowledge.» In: *Bioinformatics* 22.14 (2006), pp. 1805–1807.

[20] Karen Chan, Andrea Saltelli, and Stefano Tarantola. «Sensitivity analysis of model output: variance-based methods make the difference.» In: *Winter Simulation Conference*. 1997, pp. 261–268.

[21] M.S. Child. *Molecular Collision Theory*. Dover Books on Chemistry Series. Dover Publications, 1996. ISBN: 9780486694375. URL: https://books.google.it/books?id=uBZEU5BS5i4C.

[22] Federica Ciocchetta and Maria Luisa Guerriero. «Modelling biological compartments in Bio-PEPA.» In: *Electronic Notes in Theoretical Computer Science* 227 (2009), pp. 77–95.

[23] Vincent Danos, Jérôme Feret, Walter Fontana, and Jean Krivine. «Abstract interpretation of cellular signalling networks.» In: *International Workshop on Verification, Model Checking, and Abstract Interpretation*. Springer. 2008, pp. 83–97.

[24] Patrick De Leenheer, David Angeli, and Eduardo D Sontag. «A tutorial on monotone systems-with an application to chemical reaction networks.» In: *Proc. 16th Int. Symp. Mathematical Theory of Networks and Systems (MTNS)*. Citeseer. 2004, pp. 2965–2970.

[25] Joseph P Dexter and Jeremy Gunawardena. «Dimerization and bifunctionality confer robustness to the isocitrate dehydrogenase regulatory system in Escherichia coli.» In: *Journal of Biological Chemistry* 288.8 (2013), pp. 5770–5778.

[26] Jacqueline M Dresch, Xiaozhou Liu, David N Arnosti, and Ahmet Ay. «Thermodynamic modeling of transcription: sensitivity analysis differentiates biological mechanism from mathematical model-induced effects.» In: *BMC systems biology* 4.1 (2010), p. 142.

[27] Andrzej Ehrenfeucht and Grzegorz Rozenberg. «Reaction Systems: a Formal Framework for Processes Based on Biochemical Interactions.» In: *Electronic Communications of the EASST* 26 (2010).

[28] E Allen Emerson. «Temporal and modal logic.» In: *Formal Models and Semantics*. Elsevier, 1990, pp. 995–1072.

[29] P. Érdi and J. Tóth. *Mathematical Models of Chemical Reactions: Theory and Applications of Deterministic and Stochastic Models*. Nonlinear science : theory and applications. Manchester University Press, 1989. ISBN: 9780719022081. URL: https://books.google.it/books?id=iDu8AAAAIAAJ.

[30] Martin Feinberg. «Chemical reaction network structure and the stability of complex isothermal reactors—I. The deficiency zero and deficiency one theorems.» In: *Chemical Engineering Science* 42.10 (1987), pp. 2229–2268.

[31] Martin Feinberg. *Foundations of Chemical Reaction Network Theory*. Vol. 202. Springer, 2019.

[32] Richard P Feynman. «The Feynman lectures on physics; vol. i.» In: *American Journal of Physics* 33.9 (1965), pp. 750–752.

[33] David Gilbert and Monika Heiner. «From Petri nets to differential equations–an integrative approach for biochemical network analysis.» In: *International Conference on Application and Theory of Petri Nets*. Springer. 2006, pp. 181–200.

[34] Daniel T Gillespie. «Exact stochastic simulation of coupled chemical reactions.» In: *The journal of physical chemistry* 81.25 (1977), pp. 2340–2361.

[35] Roberta Gori, Paolo Milazzo, and Lucia Nasti. «Towards an Efficient Verification Method for Monotonicity Properties of Chemical Reaction Networks.» In: *10th International conference on bioinformatics models, methods and algorithms (BIOINFORMATICS 2019)*. 2019.

[36] Jeremy Gunawardena. «Chemical reaction network theory for in-silico biologists.» In: *Notes available for download at http://vcp. med. harvard. edu/-papers/crnt. pdf* (2003).

[37] Frank Harary. «The number of linear, directed, rooted, and connected graphs.» In: *Transactions of the American Mathematical Society* 78.2 (1955), pp. 445–463.

[38] Monika Heiner, David Gilbert, and Robin Donaldson. «Petri nets for systems and synthetic biology.» In: *International school on formal methods for the design of computer, communication and software systems*. Springer. 2008, pp. 215–264.

[39] Thomas A Henzinger. «The theory of hybrid automata.» In: *Verification of Digital and Hybrid Systems*. Springer, 2000, pp. 265–292.

[40] Morris W Hirsch and Hal Smith. «Monotone dynamical systems.» In: *Handbook of differential equations: ordinary differential equations*. Vol. 2. Elsevier, 2006, pp. 239–357.

[41] Brian P Ingalls. *Mathematical modeling in systems biology: an introduction*. MIT press, 2013.

[42] Bertrand Iooss and Paul Lemaître. «A review on global sensitivity analysis methods.» In: *Uncertainty management in simulation-optimization of complex systems*. Springer, 2015, pp. 101–122.

[43] H. Kitano. *Systems biology: towards systems-level understanding of biological systems. Kitano H ed, Foundations of Systems Biology*. 2002.

[44] H. Kitano. «Biological robustness.» In: *Nature Reviews Genetics* 5.11 (2004), pp. 826–837.

[45] Hiroaki Kitano. «Towards a theory of biological robustness.» In: *Molecular systems biology* 3.1 (2007).

[46] Ina Koch. «Petri nets in systems biology.» In: *Software & Systems Modeling* 14.2 (2015), pp. 703–710.

[47] Markus Kreer. «Classical Becker-Döring cluster equations: Rigorous results on metastability and long-time behaviour.» In: *Annalen der Physik* 505.4 (1993), pp. 398–417.

[48] Xiangfang Li, Oluwaseyi Omotere, Lijun Qian, and Edward R Dougherty. «Review of stochastic hybrid systems with applications in biological systems modeling and analysis.» In: *EURASIP J. on Bioinformatics and Systems Biology* 2017.1 (2017), p. 8.

[49] Peter Linz. *Analytical and numerical methods for Volterra equations*. Siam, 1985.

[50] Bing Liu, Soonho Kong, Sicun Gao, Paolo Zuliani, and Edmund M Clarke. «Towards personalized prostate cancer therapy using delta-reachability analysis.» In: *Proceedings of the 18th International Conference on Hybrid Systems: Computation and Control*. ACM. 2015, pp. 227–232.

[51] Alfred J Lotka. «Contribution to the theory of periodic reactions.» In: *The Journal of Physical Chemistry* 14.3 (1910), pp. 271–274.

[52] Tadao Murata. «Petri nets: Properties, analysis and applications.» In: *Proceedings of the IEEE* 77.4 (1989), pp. 541–580.

[53] Lucia Nasti, Roberta Gori, and Paolo Milazzo. «Formalizing a notion of concentration robustness for biochemical networks.» In: *Federation of International Conferences on Software Technologies: Applications and Foundations*. Springer. 2018, pp. 81–97.

[54] Lucia Nasti and Paolo Milazzo. «A computational model of Internet addiction phenomena in social networks.» In: (2017).

[55] Lucia Nasti and Paolo Milazzo. «A Hybrid Automata model of social networking addiction.» In: *Journal of Logical and Algebraic Methods in Programming* 100 (2018), pp. 215–229.

[56] Gheorghe Paun. *Membrane computing: an introduction*. Springer Science & Business Media, 2012.

[57] Oliver Penrose. «Metastable states for the Becker-Döring cluster equations.» In: *Communications in Mathematical Physics* 124.4 (1989), pp. 515–541.

[58] Mario J Pérez-Jiménez and Francisco José Romero-Campero. «A study of the robustness of the EGFR signalling cascade using continuous membrane systems.» In: *Int. Conference on the Interplay Between Natural and Artificial Computation*. Springer. 2005, pp. 268–278.

[59] Francesca Pianosi, Keith Beven, Jim Freer, Jim W Hall, Jonathan Rougier, David B Stephenson, and Thorsten Wagener. «Sensitivity analysis of environmental models: A systematic review with practical workflow.» In: *Environmental Modelling & Software* 79 (2016), pp. 214–232.

[60] Corrado Priami and Paola Quaglia. «Beta binders for biological interactions.» In: *International Conference on Computational Methods in Systems Biology*. Springer. 2004, pp. 20–33.

[61] Stephen Ramsey, David Orrell, and Hamid Bolouri. «Dizzy: stochastic simulation of large-scale genetic regulatory networks.» In: *Journal of bioinformatics and computational biology* 3.02 (2005), pp. 415–436.

[62] Aviv Regev, William Silverman, and Ehud Shapiro. «Representation and simulation of biochemical processes using the π-calculus process algebra.» In: *Biocomputing 2001*. World Scientific, 2000, pp. 459–470.

[63] Aviv Regev, Ekaterina M Panina, William Silverman, Luca Cardelli, and Ehud Shapiro. «BioAmbients: an abstraction for biological compartments.» In: *Theoretical Computer Science* 325.1 (2004), pp. 141–167.

[64] Aurélien Rizk, Gregory Batt, François Fages, and Sylvain Soliman. «A general computational method for robustness analysis with applications to synthetic gene networks.» In: *Bioinformatics* 25.12 (2009), pp. i169–i178.

[65] Aurélien Rizk, Grégory Batt, François Fages, and Sylvain Soliman. «Continuous valuations of temporal logic specifications with applications to parameter optimization and robustness measures.» In: *Theoretical Computer Science* 412.26 (2011), pp. 2827–2839.

[66] Andrea Saltelli, Stefano Tarantola, and KP-S Chan. «A quantitative model-independent method for global sensitivity analysis of model output.» In: *Technometrics* 41.1 (1999), pp. 39–56.

[67] Timothy E Saunders. «Aggregation-fragmentation model of robust concentration gradient formation.» In: *Physical Review E* 91.2 (2015), p. 022704.

[68] Marcel Schilling, Thomas Maiwald, Stefan Hengl, Dominic Winter, Clemens Kreutz, Walter Kolch, Wolf D Lehmann, Jens Timmer, and Ursula Klingmüller. «Theoretical and experimental analysis links isoform-specific ERK signalling to cell fate decisions.» In: *Molecular systems biology* 5.1 (2009).

[69] Guy Shinar, Uri Alon, and Martin Feinberg. «Sensitivity and robustness in chemical reaction networks.» In: *SIAM Journal on Applied Mathematics* 69.4 (2009), pp. 977–998.

[70] Guy Shinar and Martin Feinberg. «Structural sources of robustness in biochemical reaction networks.» In: *Science* 327.5971 (2010), pp. 1389–1391.

[71] Guy Shinar and Martin Feinberg. «Design principles for robust biochemical reaction networks: what works, what cannot work, and what might almost work.» In: *Mathematical biosciences* 231.1 (2011), pp. 39–48.

[72] Marshall Slemrod. «The Becker-Döring equations.» In: *Modeling in applied sciences*. Springer, 2000, pp. 149–171.

[73] Kei-ichi Tainaka. «Stationary pattern of vortices or strings in biological systems: lattice version of the Lotka-Volterra model.» In: *Ph. Review Letters* 63.24 (1989), p. 2688.

[74] *Web page with the scripts used to simulate Becker-Döring equations*. `http://groups.di.unipi.it/msvbio/software/Becker-Doring-Simulation.html`.

[75] *Web page with the scripts used to verify the Input-Output Monotonicity on graphs*. `http://groups.di.unipi.it/msvbio/software/IOMonotonicity.html`.

[76] Eric W. Weisstein. *Logistic Equation*. From MathWorld–A Wolfram Web Resource, `http://mathworld.wolfram.com/LogisticEquation.html`.

[77] James Michael Whitacre. «Biological robustness: paradigms, mechanisms, and systems principles.» In: *Frontiers in genetics* 3 (2012), p. 67.

[78] Lewis Wolpert. «Positional information and the spatial pattern of cellular differentiation.» In: *Journal of theoretical biology* 25.1 (1969), pp. 1–47.

[79] Yamamoto Yosaburō. «Diciotto aprile del ventesimo anno Shōwa.» In: (1945).

[80] Farkhondeh Khorashadi Zadeh, Fanny Sarrazin, Jiri Nossent, Francesca Pianosi, AV Griensven, T Wagener, and W Bauwens. «Comparison of the PAWN and Sobol'sensitivity analysis methods for a highly-parameterized hydrological model using SWAT.» In: *Proceedings of the 36th IAHR World Congress*. 2015.

[81] Zhike Zi. «Sensitivity analysis approaches applied to systems biology models.» In: *IET systems biology* 5.6 (2011), pp. 336–346.